A posteriori Fehlerschätzung für die Stokes-Gleichungen

von

Jörg Peters

Diplomarbeit in Mathematik

vorgelegt der

FAKULTÄT FÜR MATHEMATIK, INFORMATIK UND NATURWISSENSCHAFTEN

der Rheinisch-Westfälischen Technischen Hochschule Aachen

im April 2003

angefertigt im

Institut für Geometrie und Praktische Mathematik (IGPM),
Lehrstuhl für Numerische Mathematik

Prof. Dr. Arnold Reusken

Ich versichere, daß ich diese Diplomarbeit selbständig verfaßt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtliche oder sinngemäße Wiedergaben aus anderen Quellen sind kenntlich gemacht und durch Zitate belegt.

Aachen, den 24. April 2003

Jörg Peters

Einleitung

In dem seit jeher weiten Überschneidungsgebiet von Natur- und Ingenieur- Wissenschaft einerseits und der Mathematik andererseits ist die numerische Simulation physikalischer Prozesse ein wichtiges Hilfsmittel. Aus einem physikalischen Modell ergibt sich in vielen Fällen ein System partieller Differentialgleichungen als mathematische Beschreibung, dessen Lösungen – falls solche überhaupt existieren – nicht als einfache Funktionen der Eingangsdaten darstellbar sind.
In der vorliegenden Arbeit werden die Stokes-Gleichungen behandelt. Sie ergeben sich aus einem physikalischen Modell, das die Bewegung zäher Fluide beschreibt.

Bei der approximativen Lösung mit Finite-Elemente-Verfahren entstehen große, dünnbesetzte Gleichungssysteme, die hohe Speicherplatz-Anforderungen stellen. Verwendet man etwa ein uniformes Simplexgitter zur Unterteilung des Rechengebietes Ω, so wächst die Anzahl der Simplexe um den Faktor 2^n, wenn man zur Berechnung einer genaueren Approximation alle Kantenlängen halbiert. (n ist die Dimension von Ω.)
Um dieses starke Wachstum der Problemgröße zu vermeiden, kann man versuchen, Ω nur lokal in Bereichen zu verfeinern, die einen großen Beitrag zum Diskretisierungsfehler leisten.
Die Frage, wie solche Bereiche *a posteriori*, d. h. unter Verwendung einer schon vorhandenen Näherungslösung der Stokes-Gleichungen, erkannt werden können, ist das Thema der vorliegenden Arbeit. Da die Ergebnisse in das am IGPM entwickelte Finite-Elemente Paket DROPS eingehen, wird neben der theoretischen Analyse der Aufgabe auch auf die praktische Verwendbarkeit der Fehlerschätzer geachtet.

Kapitel 1 beginnt mit einer kurzen Einordnung der Stokes-Gleichungen im Rahmen der Fluiddynamik. Die mathematischen Eigenschaften der Existenz, Eindeutigkeit und Regularität von Lösungen der Stokes-Gleichungen werden über die Frage nach stetiger Invertierbarkeit des zugehörenden Differentialoperators untersucht. Dabei werden mehrere praktisch auftretende Randbedingungen behandelt. Ferner wird die für alle folgenden Kapitel wichtige, schwache Formulierung als Sattelpunktproblem analysiert.

Ab Kapitel 2 orientiert sich die Struktur der vorliegenden Arbeit an der Abfolge des adaptiven Zyklus zum numerischen Lösen stationärer Gleichungen, der in

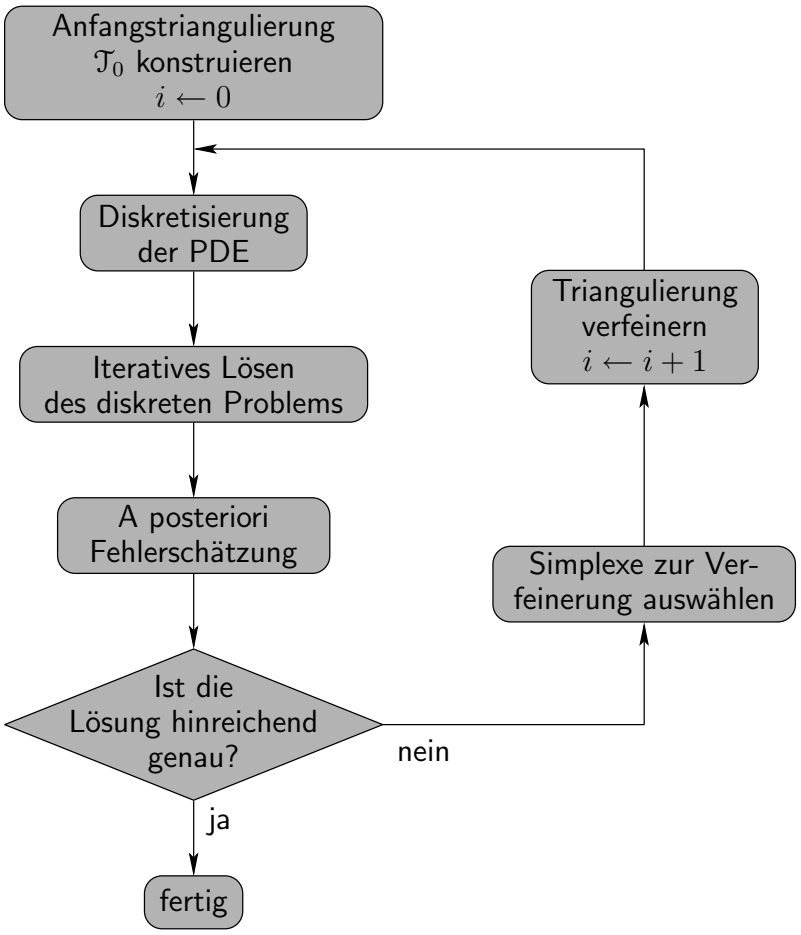

Abbildung 1: Adaptiver Zyklus

Abbildung 1 auf Seite vi dargestellt wird.
Kapitel 2 behandelt die Diskretisierung der Sattelpunktaufgabe mit der Methode der finiten Elemente. Die LBB-Bedingungen werden als Lösbarkeitskriterium der diskreten Aufgabe für die in DROPS verwendeten Taylor-Hood-Elemente genau untersucht. Es wird eine hinreichende Bedingung für die Gültigkeit der LBB-Bedingungen hergeleitet, die von der Art der verwendeten Triangulierungen abhängt.
Außerdem werden kurz die Schur-Komplement-Methode und das inexakte Uzawa-Verfahren als iterative Löser für das lineare Gleichungssystem vorgestellt.

Kapitel 3 beginnt mit einer Diskussion der Eigenschaften, die ein a posteriori Fehlerschätzer besitzen sollte. Danach wird die Theorie von Verfürth aus [39] auf die Stokes-Gleichungen spezialisiert und auf Ausströmungs-Randbedingungen erweitert. Es ergibt sich ein Fehlerschätzer für die $H^1(\Omega) \times L_2(\Omega)$-Norm und die $L_2(\Omega) \times H^{-1}(\Omega)$-Norm auf der Basis von Residuumsabschätzungen.

In Abschnitt 3.3 wird ein Fehlerschätzer konstruiert, der auf lokalen Stokes-Aufgaben basiert. Als zugrundeliegendes Gebiet genügt wegen der Verwendung von Ausströmungs-Randbedingungen ein einzelnes Simplex. Es wird gezeigt, daß die lokalen Ansatzräume den LBB-Bedingungen genügen.

In Kapitel 4 wird die „dual-weighted-residual"-Methode (DWR) aus [10] vorgestellt und auf die Stokes-Gleichungen angewendet. Sie ermöglicht eine alternative Analyse der Residuumsschätzer aus Kapitel 3. Als Verallgemeinerung dieser Verfahren ermöglicht es die DWR-Methode, den Fehler einer als Funktional gegebenen Zielgröße j zu schätzen. Durch die Wahl eines speziellen Funktionals wird ein weiterer Fehlerschätzer für die $H^1(\Omega) \times L_2(\Omega)$-Norm entwickelt.

Kapitel 5 liefert eine kurze Darstellung zweier Markierungsstrategien zur Steuerung des adaptiven Zyklus.

In Kapitel 6 werden die Residuumsschätzer anhand von drei numerischen Simulationen praktisch untersucht. Zunächst wird der Effizienzindex analysiert; in einem weiteren Experiment wird der Einfluß der Markierungsstrategie auf den adaptiven Zyklus dargestellt. Als letztes wird das Driven-Cavity-Problem adaptiv gelöst. In allen Fällen zeigt sich, daß die Fehlerschätzer den zu Beginn von Kapitel 3 diskutierten Bedingungen genügen.

Da DROPS im Rahmen des interdisziplinären Sonderforschungsbereichs 540 „Modellgestützte experimentelle Analyse kinetischer Phänomene in mehrphasigen fluiden Reaktionssystemen" eingesetzt wird, enthält die vorliegende Arbeit in Anhang A eine Einführung in die grundlegenden Konzepte der Fluiddynamik.

Inhaltsverzeichnis

Kapitel 1

Die Stokes-Gleichungen

Die Stokes-Gleichungen ergeben sich aus einem Modell der Bewegung zäher Fluide, das in Anhang A erläutert wird. In diesem Kapitel werden einige ihrer Eigenschaften zur späteren Verwendung gesammelt. Nach der Niederschrift der klassischen Formulierung wird eine schwache Formulierung angegeben, die als Sattelpunktproblem verstanden werden kann und sich als ein sogenanntes *gut gestelltes Problem der mathematischen Physik*[1] erweisen wird.

1.1 Klassische Formulierung

Die Bewegung eines inkompressiblen Fluids in einem Gebiet $\Omega \subseteq \mathbb{R}^n$ ($n \in \mathbb{N}$) kann durch Angabe des Geschwindigkeitsfeldes $u : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}^n$ für alle Zeitpunkte beschrieben werden. Beschränkt man sich auf Bewegungen, die zeitlich einen Gleichgewichtszustand erreicht haben, also *stationär* sind, so reicht ein zeitunabhängiges Geschwindigkeitsfeld $u : \Omega \longrightarrow \mathbb{R}^n$ aus. Dieses genügt wegen der Überlegungen in Anhang A den *inkompressiblen Navier-Stokes-Gleichungen* (A.34) mit $\mathrm{D}_t u = 0$:

$$\begin{aligned} (u\,\mathrm{D})u &= -\frac{1}{\rho}\,\mathrm{D}p + \nu\Delta u + f, \\ \mathrm{D}\cdot u &= 0. \end{aligned}$$

Darin ist $f : \Omega \longrightarrow \mathbb{R}^n$ eine Kraftdichte, $\nu \in \mathbb{R}$, $\nu > 0$, die sogenannte kinematische Zähigkeit des Fluids und $\rho \in \mathbb{R}$, $\rho > 0$, dessen Dichte. Den Faktor $\frac{1}{\rho}$ bei $\mathrm{D}p$ kann man in den Druck $p : \Omega \longrightarrow \mathbb{R}$ ohne Probleme aufnehmen. Das heißt, $(u, \tilde{p})^T$ mit $\tilde{p} := \frac{p}{\rho}$ löst genau dann $(u\,\mathrm{D})u = -\mathrm{D}p + \nu\Delta u + f$, wenn die ursprüngliche Gleichung von $(u, p)^T$ erfüllt wird. Lediglich die Ausströmungs- und Gleit-Randbedingungen (siehe Abschnitt 1.1.1) sind von dieser Normalisierung betroffen; dort tritt statt der Zähigkeit die kinematische Zähigkeit ν auf.

[1] nach J. Hadamard, siehe Abschnitt 1.2

Zu den Stokes-Gleichungen gelangt man, indem man den in u nichtlinearen Term $N(u) := (u\,\mathrm{D})u$ aus den Navier-Stokes-Gleichungen wegläßt. Man untersucht also die um $u_0 \equiv 0$ linearisierten Navier-Stokes-Gleichungen, denn es ist $N(u) = N(u_0) + \frac{\mathrm{d}}{\mathrm{d}u}N(u_0)(u-u_0) + \text{h. o. t.}$ mit $\frac{\mathrm{d}}{\mathrm{d}u}N(u_0) := \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}N(u_0+tu) = (u_0\,\mathrm{D})u + (u\,\mathrm{D})u_0$; die Linearisierung von N um $u_0 \equiv 0$ ist also Null.
Diese Vereinfachung führt dazu, daß die Stokes-Gleichungen als physikalisches Modell nur einen eingeschränkteren Gültigkeitsbereich als die Navier-Stokes-Gleichungen haben – mathematisch betrachtet sind sie wegen der Linearisierung eine Näherung für Strömungen mit „kleinen“ Geschwindigkeiten $u \approx 0$.
Physikalisch gesehen beschreibt $N(u)$ im stationären Fall die Trägheitskräfte im Fluid, so daß man davon ausgehen kann, daß die Stokes-Gleichungen eine gute Näherung des Fluidverhaltens liefern, solange die Trägheitskräfte im Vergleich zu den inneren Reibungskräften, den Druckkräften und den äußeren Kräften „klein“ sind. Wie in Abschnitt A.8 ausgeführt, gilt dies, wenn die *Reynoldszahl*[2] $\mathrm{Re} = \frac{U\cdot L}{\nu}$ gegen Null strebt. Dabei repräsentiert L eine typische Längenskala in Ω und U eine typische Geschwindigkeit.

Definition 1.1 (stationäre Stokes-Gleichungen). Ist $f : \Omega \longrightarrow \mathbb{R}^n$, so heißt das System

$$-\nu\Delta u(x) + \mathrm{D}p(x) = f(x) \quad \text{für alle } x \in \Omega, \tag{1.1a}$$

$$\mathrm{D}\cdot u(x) = 0 \quad \text{für alle } x \in \Omega \tag{1.1b}$$

von $n+1$ partiellen Differentialgleichungen die stationären Stokes-Gleichungen für das unbestimmte Geschwindigkeitsfeld $u : \Omega \longrightarrow \mathbb{R}^n$ und den Druck $p : \Omega \longrightarrow \mathbb{R}$. Dabei wirkt Δ komponentenweise bezüglich kartesischer Koordinaten für x und $u(x)$, also $\Delta u = (\Delta u_1, \ldots, \Delta u_n)^T$ mit dem Laplaceoperator $\Delta u_i = \sum_{j=1}^n \frac{\partial^2 u_i}{\partial x_j^2}$.
Der Term $-\nu\Delta u(x)$ repräsentiert die Diffusion des Fluids, Gleichung (1.1b) die Inkompressibilität.

Obwohl (1.1) ein System partieller Differentialgleichungen ist, die nicht einmal alle zweite Ordnung haben, ist es, wie nun gezeigt wird, in einem allgemeinen Sinn elliptisch. Dies ermöglicht in Abschnitt 1.1.2 die Anwendung der Regularitätstheorie aus [3].

Definition 1.2 (ADN-Elliptizität). Sei

$$\sum_{j=1}^{N} L_{i,j}(x,\,\mathrm{D})v_j(x) = \hat{f}_i(x) \quad \text{für alle } i = 1,\ldots,N \tag{1.2}$$

ein System linearer, partieller Differentialgleichungen über einem Gebiet $\Omega \subseteq \mathbb{R}^n$, bei dem $L_{i,j}(x,\xi)$ für alle $i,j = 1,\ldots,N$ ein Polynom in $\xi \in \mathbb{R}^n$ ist. Mit $s_1,\ldots,s_N$,

[2] Falls die Kraftdichte $f \not\equiv 0$ ist, müssen weitere Parameter berücksichtigt werden – siehe Abschnitt A.8.

$t_1, \ldots, t_N$ werden ganze Zahlen bezeichnet, die folgende Eigenschaften besitzen:

$$\begin{aligned} s_i \leq 0 \quad &\text{für alle } i = 1, \ldots, N, \\ \deg_\xi L_{i,j}(x,\xi) \leq s_i + t_j, \quad &\text{falls } s_i + t_j \geq 0, \\ L_{i,j} \equiv 0, \quad &\text{falls } s_i + t_j < 0. \end{aligned}$$

Der Hauptteil $L'_{i,j}$ von $L_{i,j}$ bezüglich der s_i und t_j besteht aus den Summanden mit $\deg_\xi L_{i,j}(x,\xi) = s_i + t_j$ $(i,j = 1, \ldots, N)$. Dann heißt (1.2) ADN-elliptisch[3] (bezüglich der s_i und t_j im Punkt x), wenn

$$\det \left(L'_{i,j}(x,\xi)\right)_{i,j=1,\ldots,N} \neq 0 \quad \text{für alle } \xi \in \mathbb{R}^N \setminus \{0\} \tag{1.3}$$

gilt.

Setzt man in Definition 1.2 $N = n+1$, $v_i = u_i$ $(i = 1, \ldots, n)$, $v_{n+1} = p$, $\hat{f}_i = f_i$ $(i = 1, \ldots, n)$, $\hat{f}_{n+1} \equiv 0$ und

$$L_{i,j}(x,\xi) = \begin{cases} -\nu|\xi|^2, & \text{falls} \quad i = j, i \leq n, \\ \xi_j, & \text{falls} \quad i = n+1, j \leq n, \\ \xi_i, & \text{falls} \quad j = n+1, i \leq n, \\ 0 & \text{sonst} \end{cases} \tag{1.4}$$

ein, so läßt sich (1.1) in der Form

$$\left(L_{i,j}(x, \mathrm{D})\right)_{i,j=1,\ldots,n+1} v = \hat{f}$$

darstellen. Wählt man dazu

$$s_i = 0,\ t_j = 2 \ (i,j = 1, \ldots, n) \text{ und } s_{n+1} = -1,\ t_{n+1} = 1, \tag{1.5}$$

so erhält man $L' = L$ für (1.1). Außerdem gilt folgender

Satz 1.3. *Der lineare Differentialoperator L der Stokesgleichungen erfüllt*

$$\det L'(x,\xi) = (-1)^n \nu^{n-1} |\xi|^{2n} \quad \textit{für alle } x \in \Omega, \xi \in \mathbb{R}^{n+1}. \tag{1.6}$$

Insbesondere sind die Stokes-Gleichungen ADN-elliptisch.

Beweis. Die ADN-Elliptizität folgt sofort aus Definition 1.2 und der ersten Aussage des Satzes.
Sei $\xi \in \mathbb{R}^{n+1} \setminus \{0\}$ beliebig. Die ersten n Zeilen von L' haben genau zwei von Null verschiedene Einträge: den Diagonaleintrag $-\nu|\xi|^2$ und eine Komponente von ξ in der letzten Spalte. Führt man nun mit den Zeilen $i = 1, \ldots, n$ die Operation

[3] Nach S. Agmon, A. Douglis, L. Nirenberg; siehe [2], [3].

„Addiere das $\frac{\xi_i}{\nu|\xi|^2}$-fache von Zeile i zu Zeile $n+1$!“ durch, so erhält man die obere Dreiecksmatrix

$$M = \begin{pmatrix} -\nu|\xi|^2 & & & \xi_1 \\ & \ddots & & \vdots \\ & & -\nu|\xi|^2 & \xi_n \\ 0 & \cdots & 0 & \frac{1}{\nu|\xi|^2}(\xi_1^2 + \cdots + \xi_n^2) \end{pmatrix},$$

deren Determinante offensichtlich $\det M = (-1)^n \nu^{n-1} |\xi|^{2n}$ lautet. Da die verwendeten Zeilenoperationen den Wert der Determinante nicht verändern, gilt $\det L' = \det M$, was diesen Teil des Beweises abschließt.
Ist $\xi = 0$, so ist $L'(x, \xi)$ die Nullmatrix, so daß die behauptete Formel auch in diesem Fall wahr ist. □

1.1.1 Randbedingungen

Zur vollständigen Beschreibung einer Strömungsaufgabe gehört neben der Differentialgleichung die Angabe von Randbedingungen.

Definition 1.4 (Randbedingungen). Sei $\partial\Omega \in C^1$, $x \in \partial\Omega$ beliebig; $n(x) \in \mathbb{R}^n$ bezeichnet die äußere Einheitsnormale an $\partial\Omega$ in x. Die Menge $\{t_i(x) \in \mathbb{R}^n \mid i = 1, \ldots, n-1\}$ sei linear unabhängig und enthalte nur Tangentialvektoren an $\partial\Omega$ in x.

1. Ist eine Funktion $u_\mathrm{D} : \partial\Omega \longrightarrow \mathbb{R}^n$ gegeben, so heißt die Gleichung $u(x) = u_\mathrm{D}(x)$ *Dirichlet-Randbedingung*. Man nennt

   $$\Gamma_\mathrm{D} = \{x \in \partial\Omega \mid u \text{ genügt in } x \text{ der Dirichlet-Randbedingung.}\}$$

   Dirichletrand von Ω. (Um diese Randbedingung zu stellen, wird keine Regularitätsbedingung an $\partial\Omega$ benötigt.)

2. Die Gleichung $\frac{\partial}{\partial n} u(x) = 0$ heißt *Neumann-Randbedingung* im Punkt x.

   $$\Gamma_\mathrm{N} = \{x \in \partial\Omega \mid u \text{ genügt in } x \text{ der Neumann-Randbedingung.}\}$$

   wird Neumannrand von Ω genannt.

3. Die Gleichung $\nu \frac{\partial}{\partial n} u(x) - p(x) n = f_\mathrm{A}$ wird als *Ausströmungs-Randbedingung* im Punkt x bezeichnet. $f_\mathrm{A} : \partial\Omega \longrightarrow \mathbb{R}^n$ beschreibt den Drucksprung beim Verlassen des Gebietes.

   $$\Gamma_\mathrm{A} = \left\{x \in \partial\Omega \;\middle|\; (u,p)^T \text{ genügt in } x \text{ der Ausströmungs-Randbedingung.}\right\}$$

   wird Ausströmungsrand von Ω genannt.

4. Die folgenden n Gleichungen heißen *Gleit-Randbedingungen* im Punkt x:

$$u(x) \cdot n(x) = 0$$
$$t_i(x)^T T(x) n(x) = 0 \quad (i = 1, \ldots, n-1).$$

Dabei ist $T(x) = -p(x) I_{n\times n} + \nu \left(\mathrm{D}u(x)^T + \mathrm{D}u(x) \right)$ der durch (A.33) definierte und an die Druckmodifikation auf Seite 1 angepaßte Spannungstensor des Fluids[4]. Auch die Randpunkte, die diesen Gleichungen genügen, werden benannt:

$$\Gamma_{\mathrm{G}} = \{ x \in \partial\Omega \mid u \text{ genügt in } x \text{ der Gleit-Randbedingung.} \}.$$

Jede dieser Randbedingungen besitzt (wenigstens) eine physikalische Interpretation, die in Abschnitt A.7 kurz erläutert wird. Insbesondere wird durch die Stokes-Gleichungen und genau eine der oben aufgeführten Randbedingungen die physikalische Situation in x vollständig beschrieben, so daß im weiteren Verlauf dieser Arbeit stets die Situation $\partial\Omega = \Gamma_{\mathrm{D}} \dot\cup \Gamma_{\mathrm{N}} \dot\cup \Gamma_{\mathrm{A}} \dot\cup \Gamma_{\mathrm{G}}$ vorausgesetzt wird.
Trotz der interessanten physikalischen Interpretation werden im größten Teil der vorliegenden Arbeit die mathematischen Eigenschaften des Fluidmodells im Vordergrund stehen. Daß auch diese „gut“ sind, deutete sich bereits durch der ADN-Elliptizität der Stokes-Gleichungen an. Jetzt wird dargelegt, daß auch die gerade vorgestellten Randbedingungen zur ADN-Theorie passen. Sie erfüllen alle die sogenannte komplementäre Randbedingung, welche zur Anwendung der elliptischen Regularitätstheorie aus [2] bzw. [3] notwendig ist.

Definition 1.5 (komplementäre Randbedingung). Sei $\Gamma \subseteq \partial\Omega$, $\Gamma \in C^1$, und

$$\sum_{j=1}^{N} B_{h,j}(x, \mathrm{D}) u_j(x) = \phi_h(x) \quad \text{für alle } h = 1, \ldots, m;\ x \in \Gamma, \tag{1.7}$$

ein System von $m = \frac{1}{2} deg_\xi(\det L')$ Rand-Differentialgleichungen, in denen alle Koeffizienten $B_{h,j}(x, \xi)$ Polynome in ξ sind. Zusätzlich zu den Gewichten t_j ($j = 1, \ldots, N$) aus Definition 1.2 gebe es ein weiteres System r_h ($h = 1, \ldots, m$) ganzer Zahlen, das die Eigenschaft

$$\begin{aligned} \deg_\xi B_{h,j}(x, \xi) \le r_h + t_j, \quad &\text{falls} \quad r_h + t_j \ge 0, \\ B_{h,j} \equiv 0, \quad &\text{falls} \quad r_h + t_j < 0 \end{aligned}$$

für alle $h = 1, \ldots, m$ und $j = 1, \ldots, N$ besitzt. Der Hauptteil $B'_{h,j}$ von $B_{h,j}$ bezüglich der r_h und t_j besteht aus den Termen von $B_{h,j}$, die genau $\deg_\xi B_{h,j}(x, \xi) = r_h + t_j$ für $h = 1, \ldots, m$ und $j = 1, \ldots, N$ erfüllen.

[4] Man beachte, daß $p(x) t_i(x)^T I_{n\times n} n(x) = 0$ ($i = 1, \ldots, n-1$) gilt, so daß diese Randbedingung nur von u abhängt.

Des weiteren sei $x \in \Gamma$ ein beliebiger Punkt und $n(x)$ die äußere Einheitsnormale zu Γ sowie $\zeta \neq 0$ ein beliebiger Tangentialvektor an Γ im Punkt x. Die m Nullstellen der charakteristischen Gleichung $\det L'(x, \zeta + \tau n) = 0$, aufgefaßt als Polynom in τ, die einen positiven Imaginärteil haben, werden $\tau_1^+, \ldots, \tau_m^+$ genannt.[5] Man setze

$$M^+(x, \zeta, \tau) = \prod_{h=1}^{m} \big(\tau - \tau_h^+(x, \zeta)\big)$$

und benenne die Komplementärmatrix[6] von $L'(x, \zeta + \tau n)$ mit $\tilde{L}$. Dann erfüllt B die komplementäre Randbedingung, wenn folgende Bedingung an die (Zeilen der) Matrix $P(x, \zeta, \tau) = B'(x, \zeta + \tau n)\tilde{L}(x, \zeta + \tau n)$ wahr ist:

> Für beliebiges $c \in \mathbb{C}^m$ folgt aus $c^T P(x, \zeta, \tau) \equiv 0 \mod (M^+(x, \zeta, \tau))$ die Aussage $c = 0$.

Da Definition 1.5 auf die Komplementärmatrix von L' zurückgreift, wird diese als erstes berechnet.

Lemma 1.6 (Komplementärmatrix von L). *Sind $x \in \Omega$ und $\xi \in \mathbb{R}^n$ beliebig, so lautet die Komplementärmatrix zu $L'(x, \xi)$*

$$\tilde{L}_{i,j}(x, \xi) = (-1)^{n-1} \nu^{n-2} |\xi|^{2n-4} \begin{cases} |\xi|^2 - \xi_i^2, & \text{falls} \quad i = j, i \leq n, \\ -\xi_i \xi_j, & \text{falls} \quad i \neq j, 1 \leq i, j \leq n, \\ -\nu|\xi|^2 \xi_j, & \text{falls} \quad i = n+1, j \leq n, \\ -\nu|\xi|^2 \xi_i, & \text{falls} \quad j = n+1, i \leq n, \\ -\nu^2 |\xi|^4, & \text{falls} \quad i = j = n+1. \end{cases}$$

Beweis. Falls $\xi = 0$ ist, folgt $L'(x, \xi) = 0$; dann ist per Definition $\tilde{L}(x, \xi) = 0$, ergo stimmt die behauptete Formel.
Sei also jetzt $\xi \neq 0$. Wegen Satz 1.3 ist $L'(x, \xi)$ invertierbar. Für jede invertierbare Matrix $A \in \mathbb{R}^{(n+1)\times(n+1)}$ gilt aufgrund der Cramerschen Regel $\det(A) A^{-1} = \tilde{A}$. Man kann somit $\tilde{A}$ bestimmen, indem man A durch elementare Zeilenumformungen in die Gestalt $\det(A) I_{(n+1)\times(n+1)}$ bringt und die verwendeten Zeilenoperationen parallel auf $I_{(n+1)\times(n+1)}$ anwendet.
Im Beweis von Satz 1.3 wurde L' bereits in die obere Dreiecksmatrix M umgewandelt. Die verwendeten Zeilenoperationen ergeben auf $I_{(n+1)\times(n+1)}$ angewendet die Matrix

$$T = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ \frac{\xi_1}{\nu|\xi|^2} & \cdots & \frac{\xi_n}{\nu|\xi|^2} & 1 \end{pmatrix},$$

[5] In [3] wird diese Anzahl durch eine zusätzliche Bedingung an L (siehe [3, Seite 39]) gesichert; aus Lemma 1.7 folgt hier sofort, das genau m Nullstellen mit positivem Imaginärteil existieren.

[6] „Matrix der Adjunkten“ – für $A \in \mathbb{R}^{n\times n}$ ist $\tilde{A}_{j,i} = (-1)^{i+j} \det(A_{\widehat{i,j}})$. Dabei entsteht $A_{\widehat{i,j}}$ aus A durch Streichen der i-ten Zeile und j-ten Spalte.

d. h., es gilt $TL' = M$. Um Spalte $n+1$ von M zu eliminieren, werden folgende Zeilenoperationen mit M durchgeführt: *Für $i = 1, \ldots, n$: „Addiere das $-\nu\xi_i$-fache von Zeile $n+1$ zu Zeile i“*. M geht auf diese Weise in die Diagonalmatrix

$$(M_2)_{i,j} = \begin{cases} -\nu|\xi|^2, & \text{falls} \quad i = j, i \leq n, \\ \frac{1}{\nu}, & \text{falls} \quad i = j = n+1, \\ 0 & \text{sonst} \end{cases}$$

über, während T zu der vollbesetzten Matrix T_2 wird:

$$(T_2)_{i,j} = \begin{cases} 1 - \frac{\xi_i^2}{|\xi|^2}, & \text{falls} \quad i = j, i \leq n, \\ -\frac{\xi_i \xi_j}{|\xi|^2}, & \text{falls} \quad i \neq j, 1 \leq i, j \leq n, \\ \frac{\xi_j}{\nu|\xi|^2}, & \text{falls} \quad i = n+1, j \leq n, \\ -\nu\xi_i, & \text{falls} \quad j = n+1, i \leq n, \\ 1, & \text{falls} \quad i = j = n+1. \end{cases}$$

Multipliziert man noch Zeile $n+1$ von M_2 und T_2 mit $-\nu^2|\xi|^2$, so erhält man die Matrix $M_3 = -\nu|\xi|^2 I_{(n+1)\times(n+1)}$, die sich nur um den Faktor $(-1)^{n-1}\nu^{n-2}|\xi|^{2n-2}$ von $\det(L')I_{(n+1)\times(n+1)}$ unterscheidet. Um diesen Faktor unterscheidet sich auch T_3, das Resultat der Zeilenoperation auf T_2, von der für $\tilde{L}$ behaupteten Matrix, was den Beweis abschließt. □

Lemma 1.7. *Das Polynom $M^+(x, \zeta, \tau)$ aus Definition 1.5 lautet für die Stokes-Gleichungen*

$$M^+(x, \zeta, \tau) = (\tau - i|\zeta|)^n.$$

Beweis. Aufgrund von Satz 1.3 gilt mit den Bezeichnungen aus Definition 1.5

$$\begin{aligned} \det(L'(x, \zeta + \tau n)) &= (-1)^n \nu^{n-1} |\zeta + \tau n|^{2n} \\ &= (-1)^n \nu^{n-1} \big((\zeta + \tau n)(\zeta + \tau n)\big)^n \\ &= (-1)^n \nu^{n-1} (|\zeta|^2 + \tau^2)^n. \end{aligned} \tag{1.8}$$

Da $|\zeta|^2 + \tau^2 = (|\zeta| + i\tau)(|\zeta| - i\tau)$ gilt, besitzt (1.8) genau $i|\zeta|$ als n-fache Nullstelle mit positivem Imaginärteil, d. h. $\tau_1^+ = \cdots = \tau_n^+ = i|\zeta|$. □

Im folgenden wird nachgewiesen, daß L und B_{D}, der zur Dirichletrandbedingung gehörende Differentialoperator, in dem frei gewählten Punkt $x \in \Gamma_{\mathrm{D}}$ die komplementäre Randbedingung erfüllen.[7] Dazu wird ohne Einschränkung der Allgemeinheit ein kartesisches Koordinatensystem verwendet, in dem die äußere Normale $n(x) = e_n$, dem n-ten Standardbasisvektor des $\mathbb{R}^n$ entspricht. $\zeta \in \mathbb{R}^n$, $\zeta \neq 0$, bezeichne einen beliebigen Tangentialvektor an Γ_{D} in x. Aufgrund von Satz 1.3

[7] Die Dirichlet-Randbedingung aus Definition 1.4 lautet damit explizit $B_{\mathrm{D}}(u, p)^T = u_{\mathrm{D}}$.

gilt für die Stokes-Gleichungen $m = n$, so daß der Rand-Differentialoperator B_{D} folgender Matrix entspricht:

$$B_{\mathrm{D}}(x, \mathrm{D}) = \begin{pmatrix} 1 & & 0 \\ & \ddots & \vdots \\ & 1 & 0 \end{pmatrix} \in \mathbb{R}^{n\times(n+1)}.$$

Als Wert für die Gewichte r_h wird $r_1 = \cdots = r_n = -2$ gewählt, so daß zusammen mit (1.5) sofort $B'_{\mathrm{D}} = B_{\mathrm{D}}$ verifiziert werden kann. Offensichtlich besteht $P_{\mathrm{D}} = B'_{\mathrm{D}}\tilde{L}$ aus den oberen n Zeilen von $\tilde{L}$.

Bemerkung 1.8. Die Wahl der Gewichte r_h, s_i und t_j erfolgt im Hinblick auf zwei Kriterien:

1. Die Differentialgleichung soll ADN-elliptisch sein und der Rand-Differentialoperator der komplementären Randbedingung genügen.

2. In Vorausschau auf Regularitätssatz 1.14 sollen die r_h möglichst nahe an 0 liegen, die s_i und t_j sollen möglichst groß sein; dann ergibt Satz 1.14 einen großen Gewinn an Regularität unter schwachen Voraussetzungen.

Satz 1.9 (komplementäre Randbedingung für $B_{\mathbf{D}}$). *B_D erfüllt bezüglich L die komplementäre Randbedingung aus Definition 1.5.*

Beweis. Es wird

> „Für beliebiges $c \in \mathbb{C}^n$ folgt aus $c^T P_{\mathrm{D}}(x, \zeta, \tau) \equiv 0 \mod (M^+(x, \zeta, \tau))$ die Aussage $c = 0$.“

gezeigt. Seien also $c \in \mathbb{R}^n$ und $q(\tau) \in \mathbb{C}[\tau]^n$ bis auf die Bedingung

$$c^T P_{\mathrm{D}} = q^T M^+ \tag{1.9}$$

beliebig gewählt. Daraus wird nun $c = 0 \in \mathbb{R}^n$ gefolgert.
Zunächst wird mit Lemma 1.6 die vorletzte Spalte von $c^T P_{\mathrm{D}}(\tau)$ dargestellt; man beachte, daß $n = e_n$ sowie $\zeta \cdot n = 0$ gelten[8]:

$$\begin{aligned} \sum_{i=1}^{n} c_i (P_{\mathrm{D}})_{i,n}(\tau) &= \alpha|\zeta + \tau n|^{2n-4} \Bigg(\sum_{i=1}^{n-1} c_i \big(-(\zeta_i + \tau n_i)(\zeta_n + \tau n_n) \big) \\ &\qquad + c_n \big(|\zeta + \tau n|^2 - (\zeta_n + \tau n_n)^2 \big) \Bigg) \qquad (1.10) \\ &= \alpha(|\zeta|^2 + \tau^2)^{n-2} \Bigg(-\tau \sum_{i=1}^{n-1} c_i \zeta_i + c_n |\zeta|^2 \Bigg), \end{aligned}$$

[8] Daraus folgt $\zeta_n = 0$.

wobei α für die von Null verschiedene, reelle Zahl $(-1)^{n-1}\nu^{n-2}$ steht. Setzt man (1.10) in (1.9) ein und faktorisiert nach der gemeinsamen $(n-2)$-fachen Nullstelle $i|\zeta|$, so erhält man

$$\alpha(\tau + i|\zeta|)^{n-2} \underbrace{\left(c_n|\zeta|^2 - \tau \sum_{i=1}^{n-1} c_i\zeta_i\right)}_{=A} = q_n(\tau)(\tau - i|\zeta|)^2. \tag{1.11}$$

Angenommen, A wäre nicht das Nullpolynom. Dann wäre auch q_n nicht das Nullpolynom und würde folglich von $(\tau+i|\zeta|)^{n-2}$ geteilt. Faktorisierte man (1.11) nach diesem Polynom, so wäre $\deg_\tau(\alpha A) \leq 1$ und der Grad des rechten Terms wenigstens zwei. Man erhielte also einen Widerspruch; das heißt $A \equiv 0$. Insbesondere gilt

$$c_n = 0 \quad \text{und} \quad \sum_{i=1}^{n-1} c_i\zeta_i = 0. \tag{1.12}$$

Sei nun $1 \leq j < n$ beliebig. Aufgrund von Lemma 1.6 ergibt sich unter Verwendung von (1.12)

$$\begin{aligned}
\sum_{i=1}^{n} c_i(P_{\mathrm{D}})_{i,j}(\tau) &= \alpha|\zeta + \tau n|^{2n-4}\Bigg(\sum_{i=1,i\neq j}^{n-1} c_i\big(-(\zeta_i + \tau n_i)(\zeta_j + \tau n_j)\big) \\
&\qquad + c_j\big(|\zeta + \tau n|^2 - (\zeta_j + \tau n_j)^2\big) - c_n(\zeta_n + \tau n_n)(\zeta_j + \tau n_j)\Bigg) \\
&= \alpha(|\zeta|^2 + \tau^2)^{n-2}\Bigg(-\zeta_j \sum_{i=1,i\neq j}^{n-1} c_i\zeta_i + c_j|\zeta|^2 + c_j\tau^2 - \zeta_j c_j \zeta_j\Bigg) \\
&= \alpha(|\zeta|^2 + \tau^2)^{n-1}c_j.
\end{aligned}$$

Wie schon einmal setzt man in Gleichung (1.9) ein und faktorisiert nach $(t - i|\zeta|)^{n-1}$:

$$\alpha(\tau + i|\zeta|)^{n-1}c_j = q_j(\tau)(\tau - i|\zeta|).$$

Da der rechte Term auf jeden Fall die Nullstelle $i|\zeta|$ besitzt, ist dies auch im linken Term der Fall. Daher ist $c_j = 0$. Insgesamt folgt $c_1 = \cdots = c_n = 0$, was den Beweis abschließt. □

Um zu beweisen, daß die Neumann-Randbedingung der komplementären Randbedingung genügt, wird ein beliebiges $x \in \Gamma_{\mathrm{N}}$ untersucht; ansonsten wird die obige Notation beibehalten. Deshalb gilt $\frac{\partial u}{\partial n} = \frac{\partial u}{\partial x_n} = \mathrm{D}_n u$, was zu

$$B_{\mathrm{N}}(x, \mathrm{D}) = \begin{pmatrix} \mathrm{D}_n & & 0 \\ & \ddots & \vdots \\ & \mathrm{D}_n & 0 \end{pmatrix} \in \mathbb{R}^{n\times(n+1)}$$

als Matrix des Neumann-Randoperators führt. Mit den Gewichten $r_1 = \cdots = r_n = -1$ folgt $B'_{\mathrm{N}} = B_{\mathrm{N}}$. Setzt man $P_{\mathrm{N}} = B'_{\mathrm{N}}\tilde{L}$, so ergibt sich $P_{\mathrm{N}}(x,\zeta,\tau) = \tau P_{\mathrm{D}}(x,\zeta,\tau)$. Man überzeugt sich leicht davon, daß der zusätzliche Faktor τ im Beweis von Satz 1.9 nicht zu Schwierigkeiten führt.

Satz 1.10 (komplementäre Randbedingung für B_{N}). *B_N erfüllt bezüglich L die komplementäre Randbedingung aus Definition 1.5.*

Beweis. Der Beweis wird auf den von Satz 1.9 zurückgeführt. Mit dem Ansatz $c^T P_{\mathrm{N}} = q^T M^+$ gilt wegen $P_{\mathrm{N}} = \tau P_{\mathrm{D}}$ und $\tau \nmid M^+$ notwendigerweise $\tau \mid q(\tau)$. Also folgt mit $\tilde{q} = \frac{1}{\tau} q \in \mathbb{C}[\tau]^n$ die Äquivalenz $c^T P_{\mathrm{N}} = q^T M^+ \iff c^T p_{\mathrm{D}} = \tilde{q}^T M^+$. Aus letzterem ergibt sich, wie im Beweis von Satz 1.9 gesehen, $c = 0$. □

Auch die Ausströmungs-Randbedingung ist komplementär zu L; sie ist bemerkenswert, weil in ihr der Druck auftritt. Unter Weiterverwendung obiger Notation lautet die Matrix des Randoperators in einem beliebigen $x \in \Gamma_{\mathrm{A}}$

$$B_{\mathrm{A}}(x,\,\mathrm{D}) = \nu \begin{pmatrix} \mathrm{D}_n & & 0 \\ & \ddots & \vdots \\ & \mathrm{D}_n & \nu^{-1} \end{pmatrix} \in \mathbb{R}^{n\times(n+1)}.$$

Benutzt man dieselben Gewichte wie bei B_{N}, findet man $B'_{\mathrm{A}} = B_{\mathrm{A}}$. $P_{\mathrm{A}} = B'_{\mathrm{A}}\tilde{L}$ unterscheidet sich nur in der letzten Zeile von $P_{\mathrm{N}}(x,\zeta,\tau)$. Es gilt

Satz 1.11 (komplementäre Randbedingung für B_{A}). *B_A erfüllt bezüglich L die komplementäre Randbedingung aus Definition 1.5.*

Beweis. Man geht analog zum Beweis von Satz 1.9 vor. Anstelle von (1.10) ergibt sich

$$\sum_{i=1}^{n} c_i (P_{\mathrm{A}})_{i,n}(\tau) = -\alpha(|\zeta|^2 + \tau^2)^{n-2}\tau^2 \Bigg(\sum_{i=1}^{n-1} c_i\zeta_i + c_n\tau \Bigg). \tag{1.13}$$

Nachdem man $(\tau - i|\zeta|)^{n-2}$ herausfaktorisiert hat, zeigt ein Vergleich mit $q_n(\tau)(\tau - i|\zeta|)^2$, daß der Term in der großen Klammer in (1.13) das Nullpolynom ist, was (1.12) impliziert.
Mit $c_n = 0$ ist $c^T P_{\mathrm{A}} = c^T P_{\mathrm{N}}$ und daher $c^T P_{\mathrm{A}} = q^T M^+ \iff c^T P_{\mathrm{N}} = q^T M^+$. Aus letzterem folgt wie im Beweis von Satz 1.10, daß $c = 0$ ist. □

Für die Gleit-Randbedingung ist die Matrix des zugehörigen Randoperators etwas komplizierter. Trotzdem kann mit einem ähnlichen Beweis gezeigt werden, daß auch für beliebige $x \in \Gamma_{\mathrm{G}}$ die komplementäre Randbedingung erfüllt ist. In [13] wird dies für die Raumdimension $n = 3$ mit den Gewichten $r_1 = r_2 = -1$, $r_3 = -2$ nachgerechnet. Es gilt also der

Satz 1.12 (komplementäre Randbedingung für B_G). *Ist $n = 3$, so erfüllt B_G bezüglich L die komplementäre Randbedingung aus Definition 1.5.*[9]

Nach der Bereitstellung des Differentialoperators L und der Rand-Differentialoperatoren B kann die in dieser Arbeit behandelte Aufgabe klassisch formuliert werden.

Aufgabe 1.13 (klassische Stokes-RWA). Sei $\Omega \in \mathbb{R}^n$ ein beschränktes Gebiet mit Lipschitzrand $\partial\Omega = \Gamma_D \dot\cup \Gamma_N \dot\cup \Gamma_A \dot\cup \Gamma_G$, wobei $|\Gamma_D| > 0$ gelten möge.[10] (Γ_N, Γ_A und Γ_G dürfen leer sein.) Ferner seien $f : \Omega \longrightarrow \mathbb{R}^n$, $u_D : \Gamma_D \longrightarrow \mathbb{R}^n$, $u_N : \Gamma_N \longrightarrow \mathbb{R}^n$ und $f_A : \Gamma_A \longrightarrow \mathbb{R}^n$ gegeben. Dann lautet die *klassische Stokes-Randwertaufgabe* (Stokes-RWA):

> Finde eine Funktion $u : \overline{\Omega} \longrightarrow \mathbb{R}^n \in C^2(\Omega) \cap C^0(\overline{\Omega}) \cap C^1(\Gamma_N \cup \Gamma_A \cup \Gamma_G)$ sowie eine Funktion $p : \overline{\Omega} \longrightarrow \mathbb{R} \in C^1(\Omega) \cap C^0(\Gamma_A)$, die den Gleichungen
>
> $$-\nu\Delta u(x) + \mathrm{D}p(x) = f(x) \quad \text{für alle } x \in \Omega, \tag{1.14a}$$
> $$\mathrm{D} \cdot u(x) = 0 \quad \text{für alle } x \in \Omega \tag{1.14b}$$
> $$u(x) = u_D(x) \quad \text{für alle } x \in \Gamma_D, \tag{1.14c}$$
> $$\frac{\partial u}{\partial n}(x) = u_N(x) \quad \text{für alle } x \in \Gamma_N, \tag{1.14d}$$
> $$\nu\frac{\partial u}{\partial n}(x) - p(x)n(x) = f_A(x) \quad \text{für alle } x \in \Gamma_A, \tag{1.14e}$$
> $$\left.\begin{aligned} u(x) \cdot n(x) &= 0 \\ t_i(x)^T T(x) n(x) &= 0 (i = 1, \ldots, n-1) \end{aligned}\right\} \quad \text{für alle } x \in \Gamma_G \tag{1.14f}$$
>
> genügen.

1.1.2 Regularitätstheorie

In [2] und [3] wird untersucht, welche Differentierbarkeitseigenschaften mögliche Lösungen des Systems

$$Lv = \hat{f} \quad \text{in} \quad \Omega, \quad Bv = \phi \quad \text{auf} \quad \partial\Omega \tag{1.15}$$

in Abhängigkeit von der Regularität der Daten besitzen, wenn L ADN-elliptisch ist und B die komplementäre Randbedingung erfüllt. Als Daten werden dabei $\partial\Omega$, $\hat{f}$, ϕ sowie die Koeffizientenfunktionen von L und B angesehen. Obwohl sich beide Arbeiten hauptsächlich mit Schauder-Abschätzungen[11] befassen, gibt es

[9] Vermutlich gilt dies in jeder Raumdimension.

[10] Dies ist nützlich, damit $a(.,.)$ elliptisch ist – siehe Satz 1.31.

[11] nach J. Schauder – Abschätzungen bezüglich der Höldernormen

einige Sätze über Abschätzungen in Sobolevnormen[12]. Seien

$$l_1 = \max_{h=1,\dots,m}\{0, r_h + 1\}, \quad l \in \mathbb{N},\ l \geq l_1, \quad t' = \max_{j=1,\dots,N}\{t_j\}$$

gegeben und $\partial\Omega \in C^{l+t',\,0}$, $\hat{f}_i \in H^{l-s_i}(\Omega)$, $\phi_h \in H^{l-r_h-\frac{1}{2}}(\partial\Omega)$. Liegen die Koeffizienten von $L_{i,j}$ in $C^{l-s_i}(\overline{\Omega})$ und die von $B_{h,j}$ in $C^{l-r_h}(\partial\Omega)$, so gilt aufgrund von [3, Satz 10.5] der folgende

Satz 1.14 (höhere Regularität bis zum Rand). *Es gibt eine Konstante $C > 0$, die nur von Ω und den Koeffizientenfunktionen von L und B abhängt, so daß für jede Lösung v von (1.15), die $\|v_j\|_{l_1+t_j} < \infty$ $(j = 1, \dots, N)$ erfüllt, $\|v_j\|_{l+t_j}$ endlich ist, und folgende Abschätzung gilt:*

$$\|v_j\|_{l+t_j} \leq C\left(\sum_i \|\hat{f}_i\|_{l-s_i} + \sum_h \|\phi_h\|_{l-r_h-\frac{1}{2}} + \sum_j \|u_j\|_0\right). \tag{1.16}$$

Falls v die einzige Lösung von (1.15) mit $\|v_j\|_{l_1+t_j} < \infty$ $(j = 1, \dots, N)$ ist, kann der letzte Summand in (1.16) entfallen.

Bemerkung 1.15. In [3] werden auch Regularitätssätze für $U \subseteq \Omega$ bewiesen, wenn nur auf $\overline{U} \cap \partial\Omega$ ein Rand-Differentialoperator, der die komplementäre Randbedingung erfüllt, gegeben ist ([3, Satz 10.6], [2, Satz 7.3]). Die Sobolevabschätzungen erfordern allerdings modifizierte Sobolevnormen und werden deshalb hier nicht vorgestellt.

Nun wird Satz 1.14 exemplarisch auf die klassische Stokes-Randwertaufgabe 1.13 und den Fall $\Gamma_{\mathrm{D}} = \partial\Omega$ angewendet. Es wird sogar zugelassen, daß anstelle von (1.14b) die Gleichung $\mathrm{D} \cdot u = g$ mit einer Funktion $g : \Omega \longrightarrow \mathbb{R}$ gilt.
Es folgt $l_1 = 0$, $t' = 2$, wozu man also alle $l \in \mathbb{N}_0$ untersuchen kann. Die Koeffizienten von $L_{i,j}$ und $B_{h,j}$ sind konstant, also von der Klasse C^∞, und beschränken nicht die Wahl von l. Von f_i $(i = 1, \dots, n)$ wird $f_i \in H^l(\Omega)$ gefordert. Für $f_{n+1} = g$ wird $f_{n+1} \in H^{l+1}(\Omega)$ vorausgesetzt. An $\phi_h \equiv (u_{\mathrm{D}})_h$ $(h = 1, \dots, n)$ wird die Bedingung $u_{\mathrm{D}} \in H^{l+\frac{3}{2}}(\partial\Omega)$ gestellt. Dann liefert Satz 1.14 für ein beschränktes Gebiet $\Omega \in C^{l+2}$ den

Satz 1.16 (Regularität der Stokes-RWA). *Es gibt eine Konstante $C > 0$, die nur von Ω und ν abhängt, so daß für jede Lösung von Aufgabe 1.13, die $\|u\|_2 < \infty$ und $\|p\|_1 < \infty$ erfüllt, $\|u\|_{l+2}$ und $\|p\|_{l+1}$ endlich sind, und folgende Abschätzungen gelten:*

$$\|u\|_{l+2} \leq C\left(\sum_{i=1}^{n} \|f_i\|_l + \|g\|_{l+1} + \|u_D\|_{l+\frac{3}{2}} + \|u\|_0 + \|p\|_0\right), \tag{1.17}$$

$$\|p\|_{l+1} \leq C\left(\sum_{i=1}^{n} \|f_i\|_l + \|g\|_{l+1} + \|u_D\|_{l+\frac{3}{2}} + \|u\|_0 + \|p\|_0\right). \tag{1.18}$$

[12] Zur Schreibweise siehe Notation 1.17 auf Seite 14.

Falls u und p die einzige Lösung von Aufgabe 1.13 mit $\|u\|_2 < \infty$, $\|p\|_1 < \infty$ sind, so kann der Term $\|u\|_0 + \|p\|_0$ in (1.17) *und* (1.18) *entfallen.*

Entsprechende Regularitätssätze für die anderen in Abschnitt 1.1.1 diskutierten Randbedingungen lassen sich mit Hilfe von Satz 1.14 ebenfalls aufstellen.
Galdi zeigt in [20]für den Fall von Dirichlet-Randbedingungen, daß man sich von der Einschränkung $\|u\|_2 < \infty$ und $\|p\|_1 < \infty$ in Satz 1.16 befreien kann. Es genügt, $\|u\|_1 < \infty$ und $\|p\|_0 < \infty$ zu fordern.[13]

1.2 Variationsformulierung

Wie bei der skalaren Poissongleichung $\Delta u = f$ stellt sich bei den Stokes-Gleichungen das Problem, daß zwar von jedem $u \in C^2$ $\Delta u \in C^0$ berechnet werden kann, jedoch die Umkehrung, also das Lösen der Gleichung, nicht für jedes $f \in C^0$ möglich ist[14]. Mithin sind die klassischen Stokes-Gleichungen für die natürlich auftretenden rechten Seiten f kein *gut gestelltes Problem der mathematischen Physik*, denn ein solches besitzt folgende Eigenschaften:

1. Es existiert eine Lösung der Aufgabe.
2. Diese Lösung ist im gegebenen Definitionsbereich eindeutig.
3. Die Lösung hängt stetig von den Daten der Aufgabenstellung ab.

Diese Eigenschaften haben auch für numerische Berechnungen große Bedeutung; insbesondere Punkt 3 sollte erfüllt sein, damit numerische Ergebnisse durch unvermeidliche Rundungsfehler von Computern nicht per se unbrauchbar sind. Eine Möglichkeit das Existenzproblem zu beheben, eröffnet die Distributionentheorie. Dort hat jede partielle Differentialgleichung mit konstanten Koeffizienten aufgrund des Satzes von Malgrange-Ehrenpreis eine Distributionen-Lösung. Allerdings hat diese starke Existenzaussage ihren Preis: Distributionen sind stetige, lineare Funktionale auf Räumen sogenannter Testfunktionen, und sie können nur in bestimmten Fällen mit Funktionen identifiziert werden. (Das sind genau die regulären Distributionen.)
Als erfolgreicher Mittelweg haben sich Funktionen mit *schwachen Ableitungen* erwiesen, die die *Sobolevräume* bilden. Dabei wird der Ableitungsbegriff der Distributionentheorie auf einem Funktionenraum verwendet. Die in dieser Arbeit verwendeten Sobolevräume liegen sogar alle in L_2. Bezüglich geeigneter innerer Produkte sind sie selbst Hilberträume, in denen bestimmte Mengen von C^∞-Funktionen dicht liegen. Es gilt für beliebige $k \in \mathbb{N}$

$$H^k(\Omega) = \overline{C^\infty(\Omega)}^{\|\cdot\|_k}$$

[13] siehe [20, Abschnitt IV.6]
[14] Bei der Poissonaufgabe benötigt man z. B. $f \in C^{0,\alpha}$, $\alpha > 0$.

mit der Sobolevnorm $\|\cdot\|_k$, die in Notation 1.17 vorgestellt wird. Der auf diesem Hilbertraum verwendete Ableitungsbegriff stimmt für $f \in H^k(\Omega) \cap C^k(\Omega)$ mit der klassischen Ableitung überein; ansonsten gilt für beliebige Testfunktionen $v \in C_c^\infty(\Omega)$ und Multiindizes α mit $|\alpha| \leq k$ (siehe Notation 1.17) die Identität

$$\int_\Omega v \, \mathrm{D}^\alpha f = (-1)^{|\alpha|} \int_\Omega f \, \mathrm{D}^\alpha v.$$

Zu Details über Sobolevräume wird auf die umfangreiche Literatur verwiesen ([19], [1], [42], [4]); jetzt wird die hier verwendete Notation vorgestellt.

Notation 1.17 *(Sobolevräume, Dualität).* Sei $k \in \mathbb{N}_0$, $\alpha, \beta \in \mathbb{Z}^n$, $x, y \in \mathbb{R}^n$ und $U \subseteq \Omega$.

- $x \cdot y$ bezeichnet das euklidisches Skalarprodukt; die euklidische Norm wird $|x|$ genannt. $|U|$ ist das Lebesguemaß von U.
- α heißt Multiindex. Addition, Multiplikation und relationale Operatoren werden komponentenweise angewendet. Es ist

$$x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}, \quad \mathrm{D}^\alpha = \mathrm{D}^{\alpha_1} \mathrm{D}^{\alpha_2} \cdots \mathrm{D}^{\alpha_n} \ (\alpha \geq 0), \quad |\alpha| = \sum_{i=1}^n |\alpha_i|,$$

$$\alpha! = \prod_{i=1}^n \alpha_i, \quad \binom{\alpha}{\beta} = \frac{\alpha!}{\beta!(\alpha-\beta)!}.$$

- Sind $u, v \in H^k(U)$, so lauten das Skalarprodukt $(u,v)_{k;U}$ und die Norm $\|u\|_{k;U}$ von $H^k(U)$ folgendermaßen:

$$(u,v)_{k;U}^{\mathrm{semi}} = \sum_{|\alpha|=k, 0 \leq \alpha} \int_U \mathrm{D}^\alpha u(x) \cdot \mathrm{D}^\alpha v(x) \, \mathrm{d}x, \quad |u|_{k;U} = \sqrt{(u,u)_{k;U}^{\mathrm{semi}}},$$

$$(u,v)_{k;U} = \sum_{|\alpha| \leq k, 0 \leq \alpha} \int_U \mathrm{D}^\alpha u(x) \cdot \mathrm{D}^\alpha v(x) \, \mathrm{d}x, \quad \|u\|_{k;U} = \sqrt{(u,u)_{k;U}}.$$

  Falls $k = 0$ oder $U = \Omega$ ist, so wird der entsprechende Index weggelassen.

- Der Dualraum von $H^k(U)$ wird mit $H^{-k}(U) = H^k(U)'$ bezeichnet. Er trägt die Norm $\|u'\|_{-k;U} = \sup\{\langle u', u\rangle_{H^{-k}(U)\times H^k(U)} \,|\, u \in H^k(U), \|u\|_{k;U} = 1\}$, die das Dualitätsprodukt $\langle u', u\rangle_{H^{-k}(U)\times H^k(U)} = u'(u)$ $(u \in H^k(U), u' \in H^{-k}(U))$ verwendet.
- Der Raum der beschränkten, linearen Operatoren, die den Hilbertraum X in den Hilbertraum Y abbilden, wird $\mathcal{L}[X,Y]$ genannt. Auf ihm wird die Norm $\|M\|_{\mathcal{L}[X,Y]} = \sup\{\|Mx\|_Y \mid \|x\|_X = 1\}$ definiert. Man setzt $\mathrm{GL}[X,Y] = \{M \in \mathcal{L}[X,Y] \mid M^{-1}$ existiert und $M \in \mathcal{L}[Y,X].\}$. Ist $M \in \mathcal{L}[X,Y]$, so heißt $M' \in \mathcal{L}[Y',X']$ mit $\langle M'y', x\rangle = \langle y', Mx\rangle$ für alle $y' \in Y'$, $x \in X$ der zu M duale Operator.

- Die Menge der beschränkten $\mathbb{R}$-Bilinearformen auf $X \times Y$ wird $\mathcal{B}[X, Y]$ genannt. Zu $b \in \mathcal{B}[X, Y]$ gibt es stets die Bilinearform $b' \in \mathcal{B}[Y, X]$, die so definiert ist: Für alle $x \in X$, $y \in Y$ setzt man $b'(y, x) = b(x, y)$. Falls $b' = b$ gilt (insbesondere $X = Y$), so heißt b symmetrisch.

1.2.1 Sattelpunktaufgaben

Es wird sich zeigen, daß sich die Stokes-Gleichungen als Sattelpunktaufgabe schreiben lassen. Bevor das in Abschnitt 1.2.2 durchgeführt wird, wird eine abstrakte Sattelpunktaufgabe definiert und deren Eigenschaften angegeben. Anschaulich sind Sattelpunkte in der Analysis Punkte auf einem Funktionsgraphen, die in einer Koordinatenrichtung eine Minimalstelle und in einer anderen ein Maximum sind. Betrachtet man anstelle einzelner Koordinatenrichtungen je ganze Funktionenräume, so befindet man sich in dem für Differentialgleichungen nützlichen Rahmen.
Für den Rest des Abschnittes 1.2.1 seien V und Q zwei Hilberträume („die Koordinatenachsen“). Ihr Produkt $X = V \times Q$ wird durch $(x, y)_X = (u, p)_V + (v, q)_Q$ $(x = (u, p)^T, y = (v, q)^T)$ auch zu einem Hilbertraum. Es werden nun zwei beliebige Bilinearformen $a \in \mathcal{B}[V, V]$, $b \in \mathcal{B}[V, Q]$ und zwei Funktionale $f \in V'$, $g \in Q'$ untersucht.

Aufgabe 1.18 (Sattelpunktaufgabe). Die (unendlich vielen) Gleichungen

$$a(u, v) + b(v, p) = f(v) \quad \text{für alle } v \in V, \tag{1.19a}$$
$$b(u, q) = g(q) \quad \text{für alle } q \in Q \tag{1.19b}$$

heißen Sattelpunktaufgabe. Es wird ein Tupel $(u, p)^T \in V \times Q$ gesucht, das (1.19) erfüllt. Setzt man für alle $x = (u, p)^T, y = (v, q)^T \in X$ $l(x, y) = a(u, v) + b(u, q) + b(v, p)$, so ist offensichtlich $l \in \mathcal{B}[X, X]$. Ebenso erhält man durch $r(x) = f(u) + g(p)$ ein $r \in X'$. Dann nennt man auch

$$l(x, y) = r(y) \quad \text{für alle } y \in X \tag{1.20}$$

Sattelpunktaufgabe. Dabei soll ein $x \in X$ bestimmt werden, das (1.20) erfüllt.

Die beiden Definitionen von Sattelpunktaufgabe sind äquivalent, denn es gilt

Lemma 1.19. *Ist $x = (u, p)^T \in X$, so gilt: (u, p) löst (1.19). $\iff$ x löst (1.20).*

Beweis. Sei $y = (v, q)^T \in X$ beliebig. Löst $x = (u, p)^T \in X$ (1.19), dann erfüllen x und y auch die Gleichung aus (1.20), denn diese ist lediglich die Summe der Gleichungen aus (1.19a) und (1.19b).
Löst umgekehrt $x = (u, p)^T \in X$ (1.20), so gilt die Gleichung aus (1.20) insbesondere für $(v, 0)^T$ bzw. $(0, q)^T$, was die Gleichungen (1.19a) und (1.19b) impliziert. □

Um zu verdeutlichen, warum Aufgabe 1.18 Sattelpunktaufgabe genannt wird, kann man das Funktional $J : X \longrightarrow \mathbb{R}$,

$$J(\begin{pmatrix} u \\ p \end{pmatrix}) = a(u,u) + 2b(u,p) - 2f(u) - 2g(p) = l(\begin{pmatrix} u \\ p \end{pmatrix}, \begin{pmatrix} u \\ p \end{pmatrix}) - 2r(\begin{pmatrix} u \\ p \end{pmatrix}) \quad (1.21)$$

betrachten, denn es hat unter bestimmten Voraussetzungen Lösungen von (1.19) als Sattelpunkte. Man bezeichnet eine Bilinearform $m \in \mathcal{B}[X,X]$ als X-*elliptisch*, wenn für eine Zahl $E \in \mathbb{R}$, $E > 0$, gilt: Jedes $x \in X$ erfüllt $m(x,x) \geq E\|x\|_X^2$.

Satz 1.20 (Sattelpunkte von J). *Ist a symmetrisch und V-elliptisch, so löst $x = (u,p)^T \in X$ genau dann* (1.19)*, wenn*

$$J(\begin{pmatrix} u \\ q \end{pmatrix}) \leq J(x) \leq J(\begin{pmatrix} v \\ p \end{pmatrix}) \quad (1.22)$$

für alle $(v,q)^T \in X$ wahr ist. Dies ist auch äquivalent zu

$$J(x) = \max_{q \in Q} \min_{v \in V} J\big((v,q)^T\big).$$

Beweis. Die Max-Min-Charakterisierung wird in [27, Kapitel 12] bewiesen. Sei zunächst $(u,p)^T \in X$ eine Lösung von (1.19). Man rechnet leicht[15] $J((v,p)^T) - J((u,p)^T) = a(u-v,u-v) - 2\big(a(u,u-v) + b(u-v,p) - f(u-v)\big)$ nach. Die große Klammer verschwindet wegen (1.19a) und wegen der V-Elliptizität von a ist $a(u-v,u-v) \geq E\|u-v\|_V^2 \geq 0$, was die rechte Ungleichung in (1.22) beweist. Die linke Ungleichung folgt aus $J((u,p)^T) - J((u,q)^T) = 2(b(u,p-q) - g(p-q)) = 0$, wobei die letzte Gleichheit ist eine Konsequenz von (1.19b) ist.

Nun erfülle $(u,p)^T \in X$ die Ungleichungskette (1.22) für alle $(v,q)^T \in X$; $(v,q)^T \in X$ sei beliebig. Aus (1.22) folgt dann für $q_\pm = p \mp q$: $0 \leq J((u,p)^T) - J((u,q_\pm)^T) = \pm 2\big(b(u,q) - g(q)\big)$. Also ist (1.19b) wahr.

Statt von (1.22) auf (1.19a) zu folgern, wird die logische Kontraposition bewiesen: Existiert ein $(v,q)^T \in X$, das nicht (1.19a) genügt, dann gibt es ein $(w,r)^T \in X$, für das nicht (1.22) gilt. Sei also für $(v,q)^T \in X$ o. B. d. A. $a(u,v) + b(v,p) - f(v) > 0$. Setzt man $(w,r)^T = (u+tv,p)$, $t \in \mathbb{R}$, in (1.21) ein, so erhält man

$$\begin{aligned} J(\begin{pmatrix} w \\ p \end{pmatrix}) - J(\begin{pmatrix} u \\ p \end{pmatrix}) &= t^2 a(v,v) + t\,2\big(a(u,v) + b(v,p) - f(v)\big) \\ &= C_1 t^2 + C_2 t = C_1 t\Big(t + \frac{C_2}{C_1}\Big). \end{aligned} \quad (1.23)$$

Da aufgrund der Voraussetzungen $C_1, C_2 > 0$ gilt, ist (1.23) bei $t = -\frac{C_2}{2C_1}$ strikt negativ, so daß die rechte Ungleichung von (1.22) verletzt ist. Dies vervollständigt den Beweis des Lemmas. □

[15] Addition der „nahrhaften Null“ $2a(u,u) - 2a(u,v) - 2a(u,u) + 2a(u,v)$, Symmetrie und „zweites Binom“.

Lösbarkeit der Sattelpunktaufgabe

Um die Lösbarkeit der Sattelpunktaufgabe zu untersuchen, werden die zu den Bilinearformen gehörenden Operatoren verwendet, weil man statt unendlich vieler Gleichungen dann nur zwei (Operator-) Gleichungen betrachten muß.

Definition 1.21 (Operator zu einer Bilinearform). Seien H_1, H_2 Hilberträume und $m \in \mathcal{B}[H_1, H_2]$. Dann heißt

$$M : H_1 \longrightarrow H_2' : h_1 \longmapsto \big(H_2 \longrightarrow \mathbb{R} : h_2 \longmapsto m(h_1, h_2)\big)$$

der zu m gehörende Operator. Der zu m' gehörende Operator wird als M' bezeichnet.

Die Notation in Definition 1.21 ist mit dem üblichen Dualitätsbegriff kompatibel, denn es gilt

Lemma 1.22. *Seien H_1, H_2 Hilberträume und $m \in \mathcal{B}[H_1, H_2]$. Dann gilt:*

1. *$M \in \mathcal{L}[H_1, H_2']$ und*

$$\|M\|_{\mathcal{L}[H_1,H_2']} = \inf\{C \in \mathbb{R} \mid |m(h_1, h_2)| \le C\|h_1\|_{H_1}\|h_2\|_{H_2} \text{ für alle } h_1 \in H_1,\ h_2 \in H_2\}.$$

2. *Der zu m' gehörende Operator M' ist zu M dual. Es gilt zusätzlich : $m = m' \iff M = M'$.*

Beweis. Elementare Funktionalanalysis ([27], [4]). □

Ab jetzt wird mit A, B, B', L der zu a, b, b', l gehörende Operator gekennzeichnet Aufgabe 1.19 und Aufgabe 1.20 sind dann äquivalent zu folgenden Operatorgleichungen:

$$Au + B'p = f, \tag{1.24a}$$
$$Bu = g, \tag{1.24b}$$

beziehungsweise

$$L \begin{pmatrix} u \\ p \end{pmatrix} = r. \tag{1.25}$$

Man kann (1.24) und (1.25) mit den Blockoperatoren

$$\begin{pmatrix} A & B' \\ B & 0 \end{pmatrix} \quad \text{bzw.} \quad \begin{pmatrix} A + B' & B \end{pmatrix}$$

schreiben. Unter leichtem Mißbrauch der Notation werden beide mit L bezeichnet. Unter welchen Bedingungen ist der Operator L stetig invertierbar? Das Lemma von Lax und Milgram erweist sich als zu stumpf, um diese Frage zu beantworten, denn, falls $Q \neq \{0\}$ gilt, so ist l nicht X-elliptisch: Läßt man l auf $(0, q)^T \in X$, $\|q\|_Q = 1$, wirken, dann erhält man $l((0, q)^T, (0, q)^T) = 0$. Dennoch wird dieses fundamentale Lemma hier zitiert.

Lemma 1.23 (Lax und Milgram). *Ist $m \in \mathcal{B}[H,H]$, wobei H ein Hilbertraum ist, H-elliptisch mit Elliptizitätskonstante $E > 0$, so gilt: $M^{-1} \in \mathcal{L}[H',H]$ und $\|M\|_{\mathcal{L}[H',H]} \leq E^{-1}$.*

Beweis. Zum Beispiel in [19, funktionalanalytisch], [26, mit Banachschen Fixpunktsatz]. □

Lemma 1.24 (stetige Invertierbarkeit). *Ist $M \in \mathcal{L}[H_1,H_2]$ mit Hilberträumen H_1, H_2, so sind äquivalent:*

1. *M^{-1} existiert und $M^{-1} \in \mathcal{L}[H_2,H_1]$.*
2. *Es gilt*
$$\inf_{h_1 \in H_1,\, \|h_1\|_{H_1}=1} \|Mh_1\|_{H_2} = \varepsilon > 0$$
und für jedes $h_2' \in H_2'$ mit $\|h_2'\|_{H_2'} = 1$ ist $\|M'h_2'\|_{H_1'} > 0$.

Zudem ist $\|M^{-1}\| = \varepsilon^{-1}$, falls eine der beiden Aussagen zutrifft.

Beweis. „1 $\Longrightarrow$ 2“: Neben den Voraussetzungen des Satzes sei also Aussage 1 gegeben. Da M bijektiv und beschränkt ist, gilt

$$\begin{aligned}\inf_{h_1 \in H_1, \|h_1\|_{H_1}=1} \|Mh_1\|_{H_2} &= \inf_{h_1 \in H_1, h_1 \neq 0} \frac{\|Mh_1\|_{H_2}}{\|h_1\|_{H_1}} = \inf_{h_2 \in H_2, h_2 \neq 0} \frac{\|h_2\|_{H_2}}{\|M^{-1}h_2\|_{H_1}} \\ &= \left(\sup_{h_2 \in H_2, h_2 \neq 0} \frac{\|M^{-1}h_2\|_{H_1}}{\|h_2\|_{H_2}} \right)^{-1} = \|M^{-1}\|^{-1} > 0, \quad (1.26)\end{aligned}$$

was den ersten Teil von 2 und die Zusatzaussage $\varepsilon^{-1} = \|M^{-1}\|$ beweist. M' ist stetig invertierbar, da auch M dies ist, und es gilt $\|(M')^{-1}\| = \|(M^{-1})'\| = \|M^{-1}\| = \varepsilon > 0$. Deshalb kann man die Berechnung (1.26) auch für M' durchführen, was $\inf_{h_2' \in H_2', \|h_2'\|_{H_2'}=1} \|M'h_2'\|_{H_1'} = \varepsilon$ liefert. Dies umfaßt den zweiten Teil von Aussage 2.

„2 $\Longrightarrow$ 1“: Sei nun Aussage 2 gegeben, dann ist M injektiv, denn für ein beliebiges $0 \neq h_1 \in H_1$ folgt aus dem ersten Teil von 2, daß $\|Mh_1\|_{H_2} \geq \varepsilon \|h_1\|_{H_1} > 0$ gilt, d. h. $\ker M = \{0\}$. Also existiert $M^{-1} : M(H_1) \longrightarrow H_1$ und ist bijektiv. Die Linearität rechnet man leicht nach.

Nun wird die Surjektivität von M nachgewiesen. $H = M(H_1)$ ist abgeschlossen, denn sei $(b_k)_{k \in \mathbb{N}} \in H^{\mathbb{N}}$ eine beliebige Folge mit Grenzwert $b \in H_2$. Dazu existiert die Folge $(x_k)_{k \in \mathbb{N}} \in H_1^{\mathbb{N}}$ mit $Mx_k = b_k$, die wegen des ersten Teils von Aussage 2 eine Cauchyfolge ist: $\|x_l - x_k\|_{H_1} \leq \varepsilon^{-1} \|b_l - b_k\|_{H_2}$ $(k,l \in \mathbb{N})$. Ihr Grenzwert sei $x \in H_1$. Da M auf H_1 stetig ist, erhält man $(b_k =) Mx_k \to Mx$, $k \to \infty$, also $b = Mx \in H$. Wegen der Abgeschlossenheit von H kann man H_2 in die H_2-orthogonalen Teilräume $H_2 = H \oplus H^{\perp}$ zerlegen. Der zweite Teil von Aussage 2 impliziert $\ker M' = \{0\}$. Angenommen, es gäbe ein $0 \neq h \in H^{\perp}$. Wegen des

Rieszschen Hilbertraumisomorphismus $J : H_2 \longrightarrow H_2' : f \longmapsto (f, \cdot)_{H_2}$ wäre dann $h' = Jh \neq 0$. Also würde der Widerspruch $0 \neq \langle M'h', h\rangle = \langle h', Mh\rangle = (h, Mh)_{H_2} = 0$ folgen, so daß man insgesamt $H^\perp = \{0\}$ erhält. Somit ist $M(H_1) = H_2$.
Es wird demonstriert, daß M^{-1} stetig ist. Dazu sei $0 \neq h_2 \in M(H_1)$ beliebig. Teil eins der Aussage 2 ergibt $\|h_2\|_{H_2} = \|MM^{-1}h_2\|_{H_2} \geq \varepsilon\|M^{-1}h_2\|_{H_1}$, so daß $\|M^{-1}\| \leq \varepsilon^{-1}$ folgt. (Auch $h_2 = 0$ erfüllt diese Ungleichung.) Damit ist der Beweis abgeschlossen; die Zusatzaussage folgt jetzt aus Aussage 1. □

Bemerkung 1.25. Ist M der zu $m \in \mathcal{B}[H_1, H_2]$ gehörende Operator, so entspricht die zweite Aussage von Lemma 1.24 wegen $\|Mh_1\|_{H_2'} = \sup\{|\langle Mh_1, h_2\rangle| \;\big|\; h_2 \in H_2, \|h_2\|_{H_2} = 1\}$ den *Babuška-Bedingungen*

$$\inf\Big\{\sup\{|m(h_1,h_2)| \mid h_2 \in H_2, \|h_2\|_{H_2} = 1\} \;\Big|\; h_1 \in H_1, \|h_1\|_{H_1} = 1\Big\} = \varepsilon > 0, \tag{1.27a}$$

$$\sup\{|m(h_1,h_2)| \mid h_1 \in H_1, \|h_1\|_{H_1} = 1\} > 0 \quad \text{für alle } h_2 \in H_2, \|h_2\|_{H_2} = 1. \tag{1.27b}$$

Da $\|M^{-1}\| = \|M'^{-1}\|$ gilt, kann man anstelle von $\|M'h_2'\|_{H_1'} > 0$ in Lemma 1.24 äquivalent $\|M'h_2'\|_{H_1'} = \varepsilon > 0$ schreiben. In diesem Fall darf der erste Teil von Aussage 2 zu $\|Mh_1\|_{H_2} > 0$ für alle $h_1 \in H_1$, $\|h_1\|_{H_1} = 1$, abgeschwächt werden.[16] Die Babuška-Bedingungen (1.27) werden dann zu

$$\sup\{|m(h_1,h_2)| \mid h_2 \in H_2, \|h_2\|_{H_2} = 1\} > 0 \quad \text{für alle } h_1 \in H_1, \|h_1\|_{H_1} = 1, \tag{1.28a}$$

$$\inf\Big\{\sup\{|m(h_1,h_2)| \mid h_1 \in H_1, \|h_1\|_{H_1} = 1\} \;\Big|\; h_2 \in H_2, \|h_2\|_{H_2} = 1\Big\} = \varepsilon > 0. \tag{1.28b}$$

Mit Hilfe von Lemma 1.24 wird die Umkehrbarkeit von L charakterisiert. Anhand des Matrixbeispiels

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \end{pmatrix}, \quad L = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

wird deutlich, daß weder A noch B invertierbar sein müssen, damit L dies ist. Faßt man (1.25) formal als restringierte Minimierungsaufgabe nach Lagrange auf, so ist anschaulich klar, daß A nur auf der Teilmenge von V umkehrbar zu sein braucht, die der Nebenbedingung $Bv = g$ genügt, d. h. auf $\tilde{v} + \ker B$, wobei $B\tilde{v} = g$ ist. Diese Darstellung legt nahe, daß $\tilde{v}$ nur in $V/\ker B$ gesucht werden muß, so daß B nur auf diesem Quotienten stetig invertierbar sein muß. Diese Anschauung läßt sich so präzisieren:

[16] Man wendet Lemma 1.24 auf M' an.

1. Man definiert $V_0 = \ker B \leq V$. Da B stetig ist, ist V_0 abgeschlossen.

2. Wegen der Abgeschlossenheit von V_0 existiert die V-orthogonale Zerlegung $V = V_0 \oplus V_\perp$ mit den Orthoprojektoren $P_0 \in \mathcal{L}[V, V_0]$, $P_0|_{V_0} = id$, $P_0|_{V_\perp} = 0$ und $P_\perp \in \mathcal{L}[V, V_\perp]$, $P_\perp|_{V_0} = 0$, $P_\perp|_{V_\perp} = id$. Via des Rieszschen Hilbertraumisomorphismus $J_V : V \longrightarrow V' : f \longmapsto (f, \cdot)_V$ induziert sie eine Zerlegung $V' = V_0' \oplus V_\perp'$ von V'. Sie ist V'-orthogonal und V_0' ist abgeschlossen, weil für beliebige $h_1, h_2 \in V$ $(h_1, h_2)_V = (J_V h_1, J_V h_2)_{V'}$ gilt. Das heißt, die zugehörigen Orthoprojektoren sind P_0' und $P_\perp'$.

3. Mit Hilfe von Punkt 1 und 2 erhält man eine orthogonale Zerlegung von A, B und B': Man setzt

$$A_{0,0} : V_0 \longrightarrow V_0',\ A_{0,0} = P_0' \circ A|_{V_0}, \quad A_{\perp,\perp} : V_\perp \longrightarrow V_\perp',\ A_{\perp,\perp} = P_\perp' \circ A|_{V_\perp},$$
$$A_{0,\perp} : V_0 \longrightarrow V_\perp',\ A_{0,\perp} = P_0' \circ A|_{V_0}, \quad A_{\perp,0} : V_\perp \longrightarrow V_0',\ A_{\perp,0} = P_\perp' \circ A|_{V_\perp}.$$

   All diese Operatoren sind per definitionem stetig und linear.

   Entsprechend wird B zerlegt: $B_0 \in \mathcal{L}[V_0, Q']$, $B_0 \equiv 0$, $B_\perp \in \mathcal{L}[V_\perp, Q']$, $B_\perp v_\perp = Bv$, was $B_0' \equiv 0$, $B_\perp' \in \mathcal{L}[Q, V_\perp']$, $B_\perp' q = B'q$, nach sich zieht.

   Zerlegt man noch f durch Einschränken auf V_0 und $V_\perp$ in f_0 und $f_\perp$, findet man die zu (1.24) äquivalente Gleichung

$$\begin{pmatrix} A_{0,0} & A_{\perp,0} & 0 \\ A_{0,\perp} & A_{\perp,\perp} & B_\perp' \\ 0 & B_\perp & 0 \end{pmatrix} \begin{pmatrix} u_0 \\ u_\perp \\ p \end{pmatrix} = \begin{pmatrix} f_0 \\ f_\perp \\ g \end{pmatrix}. \tag{1.29}$$

Satz 1.26 (Lösbarkeit der Sattelpunktaufgabe). *Die Aussagen*

$$L^{-1} \in \mathcal{L}[X', X], \tag{1.30}$$
$$A_{0,0}^{-1} \in \mathcal{L}[V', V] \quad \textit{und} \quad B_\perp^{-1} \in \mathcal{L}[Q', V] \tag{1.31}$$

sind äquivalent. Ist eine der beiden erfüllt, so hat man die Abschätzung

$$\|L^{-1}\|_{\mathcal{L}[X',X]} \leq \sqrt{3}\max\Big\{\|A_{0,0}^{-1}\|\left(1 + \|A\|\,\|B_\perp^{-1}\|\right), \|B_\perp^{-1}\|\big(1+ \|A\|\left(\|A_{0,0}^{-1}\| + \|B_\perp^{-1}\| + \|B_\perp^{-1}\|\,\|A\|\,\|A_{0,0}^{-1}\|\right)\big)\Big\}. \tag{1.32}$$

Beweis. Zuerst gelte (1.31), d. h. (1.30) ist zu zeigen. Betrachtet man (1.29), so kann man die Zeilen in der Reihenfolge 3, 1, 2 auflösen und erhält

$$\begin{aligned} u_\perp = B_\perp^{-1} g, \quad u_0 = A_{0,0}^{-1}\left(f_0 - A_{0,\perp} B_\perp^{-1} g\right), \\ p = (B_\perp')^{-1}\left(f_\perp - A_{\perp,\perp} B_\perp^{-1} g - A_{\perp,0} A_{0,0}^{-1}(f_0 - A_{0,\perp} B_\perp^{-1} g)\right). \end{aligned} \tag{1.33}$$

Da $(B_\perp')^{-1} = (B_\perp^{-1})'$ ist, sind die auftretenden Operatoren nach Voraussetzung alle linear und stetig, so daß (1.30) gezeigt ist.

Nun sei (1.30) wahr; daraus wird (1.31) gefolgert. Sei $f_0 \in V_0'$ beliebig, dann hat (1.29) mit der rechten Seite $(f_0, 0, 0)^T$ eine eindeutige Lösung $(u_0, u_\perp, p)^T \in X$. Wegen $B_\perp u_\perp = 0$ und weil $B_\perp$ nach Konstruktion injektiv ist, gilt $u_\perp \in V_\perp \cap V_0 = \{0\}$, also $u_\perp \equiv 0$. Somit folgt aus Zeile 1 von (1.29), daß $A_{0,0}u_0 = f_0$ für jedes $f_0 \in V_0'$ eindeutig lösbar ist, d. h. $A_{0,0} \in \mathcal{L}[V_0, V_0']$ ist bijektiv. Nach dem Satz über die offene Abbildung ([4, Kap. 5]) folgt daraus schon $A_{0,0}^{-1} \in \mathcal{L}[V_0', V_0]$.
Betrachtet man zu einem beliebigen $g \in Q'$ die Gleichung (1.29) mit der rechten Seite $(0, 0, g)^T$, so ergibt sich analog $B_\perp^{-1} \in \mathcal{L}[Q', V_\perp]$.
Die Abschätzung (1.32) erhält man aus der expliziten Darstellung der Lösung zu Beginn des Beweises. Zunächst gilt für $L^{-1}r = (v_0, v_\perp, q)^T \in V_0 \times V_\perp \times Q$ mit $r \in X'$, $\|r\|_{X'} = 1$, beliebig, daß $\|L^{-1}r\|_X = \|(\|v_0\|_V, \|v_\perp\|_V, \|q\|_Q)^T\|_2 \leq \sqrt{3}\|(\|v_0\|_V, \|v_\perp\|_V, \|q\|_Q)^T\|_\infty$ ist. Mit Hilfe der Dreiecksungleichung und der Tatsache, daß $\|A_{0,0}\|$, $\|A_{0,\perp}\|$, $\|A_{\perp,0}\|$ und $\|A_{\perp,\perp}\|$ kleiner als $\|A\|$ sind, folgert man aus (1.33) sofort (1.32). □

Bemerkung 1.27. Formuliert man (1.31) mit Hilfe der Bilinearformen, so ergeben sich die *LBB-Bedingungen*[17]

$$\inf\big\{\sup\{|a(u,v)| \mid v \in V_0, \|v\|_V = 1\} \;\big|\; u \in V, \|u\|_V = 1\big\} = \alpha > 0, \tag{1.34a}$$

$$\sup\{|a(u,v)| \mid u \in V_0, \|u\|_V = 1\} > 0 \quad \text{für alle } v \in V_0, \|v\|_V = 1, \tag{1.34b}$$

$$\inf\big\{\sup\{|b(u,q)| \mid u \in V, \|u\|_V = 1\} \;\big|\; q \in Q, \|q\|_Q = 1\big\} = \beta > 0, \tag{1.34c}$$

wobei (1.34c) Ungleichung (1.28b) entspricht. Gemäß Lemma 1.24 erwartet man eine weitere Bedingung an b: $\sup\{\, |b(u,q)| \,\big|\, q \in Q, \|q\|_Q = 1\,\} > 0$ für alle $u \in V_\perp$, $\|u\|_V = 1$. Diese ist bei Sattelpunktaufgaben stets erfüllt, denn angenommen, es wäre $\sup\{|b(u,q)| \mid q \in Q, \|q\|_Q = 1\} = 0$ für ein $u \in V_\perp$ mit $\|u\|_V = 1$. Das hieße $u \in \ker B = V_0$, also $u = 0$, was einen Widerspruch darstellen würde.
Daß in (1.34c) das Supremum über $V \geq V_\perp$ anstelle von $V_\perp$ gebildet wird, ändert dessen Wert nicht, da $\|v\|^2 = \|P_0 v\|^2 + \|P_\perp v\|^2$ und $b(v,q) = b(P_\perp v, q)$ für alle $v \in V$, $q \in Q$ gilt.

1.2.2 Anwendung auf die Stokes-Gleichungen

Sei $\Omega \subseteq \mathbb{R}^n$ ein beschränktes Gebiet mit Lipschitzrand $\Gamma = \partial\Omega$. Der Einfachheit halber wird $\Gamma_N = \Gamma_G = \emptyset$ und $|\Gamma_D| > 0$ angenommen. Für die Stokes-Gleichungen wird die Theorie aus Abschnitt 1.2.1 auf folgende Weise spezialisiert:

$$V = H_0^1(\Omega) = \big\{u \in H^1(\Omega) \mid u \equiv 0 \quad \text{auf} \quad \Gamma_D\big\}, \tag{1.35}$$

$$Q = \begin{cases} L_2(\Omega), & \text{falls} \quad |\Gamma_A| > 0, \\ L_2^0(\Omega) = \big\{q \in L_2(\Omega) \big|\ \int_\Omega q = 0\big\} & \text{sonst.} \end{cases} \tag{1.36}$$

[17] nach O. A. Ladyzhenskaya, I. Babuška, F. Brezzi

Die Bilinearformen werden für die Diskussion inhomogener Randbedingungen auf Oberräumen von V und Q definiert. Man kann sie ohne weiteres auf V und Q einschränken:

$$a:\ H^1(\Omega)\times H^1(\Omega)\longrightarrow\mathbb{R}:\ (u,v)^T\longmapsto\nu\sum_{i=1}^{n}\int_\Omega \mathrm{D}u_i\cdot\mathrm{D}u_i, \tag{1.37}$$

$$b:\ H^1(\Omega)\times L_2(\Omega)\longrightarrow\mathbb{R}:\ (u,p)^T\longmapsto-\int_\Omega p\,\mathrm{D}\cdot u, \tag{1.38}$$

$$f\in V'\quad\text{beliebig},\quad g=0\,(\in Q'). \tag{1.39}$$

Lemma 1.28 (Stetigkeit von a und b). *Es ist $a\in\mathcal{B}[V,V]$ und $|a(u,v)|\le\nu\|u\|_1\|v\|_1$ für alle $u,v\in H^1(\Omega)$ und $a'=a$. Außerdem gilt $b\in\mathcal{B}[V,Q]$, $|b(v,q)|\le\|u\|_1\|q\|$ für alle $v\in H^1(\Omega)$, $q\in L_2(\Omega)$.*

Beweis. Durch Einsetzen in die Definitionen und die Cauchy-Schwarzsche Ungleichung. □

Aufgabe 1.18 heißt *schwache Form* der Stokes-Gleichungen mit homogenen Randwerten. Die zugehörigen Operatorgleichungen (1.24) bzw. (1.25) werden genauso genannt. Diese Bezeichnung wird so gerechtfertigt:

Satz 1.29 (klassische und schwache Formulierung). *Sei $u\in V\cap C^2(\Omega)\cap C^1(\Omega\cup\Gamma_A)\cap C^0(\Omega\cup\Gamma_D)$, $p\in Q\cap C^1(\Omega)$ und $\tilde f\in C^0(\Omega)$ mit $f:Q\longrightarrow\mathbb{R}$, $f(q)=\int_\Omega\tilde f\,q\in Q'$. Dann sind (1.14) (mit $f=\tilde f$, $u_D\equiv0$,$f_A\equiv0$) und (1.19) äquivalent.*

Beweis. Zunächst sei $(u,p)^T$ eine Lösung der klassischen Formulierung (1.14) mit homogenen Dirichlet- und Ausströmungs- Randbedingungen. Betrachtet man ein beliebiges $v\in C^\infty(\Omega)\cap V$, so erhält man durch partielle Integration von (1.14a) unter Verwendung der Randbedingungen

$$\begin{aligned}f(v)&=\int_\Omega\tilde f v=\int_\Omega-\nu\Delta u\cdot v+\int_\Omega v\cdot\mathrm{D}p\\&=-\nu\sum_{i=1}^{n}\oint_{\Gamma_A}n\cdot(\mathrm{D}u_i)v_i+\nu\sum_{i=1}^{n}\int_\Omega\mathrm{D}u_i\cdot\mathrm{D}v_i+\oint_{\Gamma_A}n\cdot vp-\int_\Omega(\mathrm{D}\cdot v)\,p\\&=a(u,v)+b(v,p),\end{aligned}$$

wodurch (1.19a) erwiesen ist, weil $C^\infty(\Omega)\cap V$ in V dicht liegt. Analog ergibt (1.14b) nach Multiplikation mit einem beliebigen $q\in Q$ und anschließender Integration, daß (1.19b) erfüllt ist.
Um die entgegengesetzte Implikation zu beweisen sei jetzt (1.19) erfüllt. Man betrachte zunächst zu beliebigem $v_i\in C_c^\infty(\Omega)$ die vektorwertige Testfunktion

$v = (0, \ldots, 0, v_i, 0, \ldots, 0)^T$ mit v_i an i-ter Stelle. Bei der partiellen Integration von (1.19a) treten dann keine Randterme auf, d. h.

$$0 = a(u,v) + b(v,p) - f(v) = \left(-\nu\Delta u_i + (\,\mathrm{D}p)_i - \tilde{f}_i, v_i\right)_\Omega .$$

Da der linke Faktor des L_2-Skalarproduktes L_2-orthogonal zu $C_c^\infty(\Omega)$ ist und diese Menge dicht in $L_2(\Omega)$ liegt, ist er gleich $0 \in L_2(\Omega)$. Da er nach Voraussetzung sogar stetig ist, beweist dies die punktweise Gültigkeit der i-ten Komponente von (1.14a).
Nach Voraussetzung an V gilt $u = 0$ auf Γ_D, was (1.14c) mit $u_\mathrm{D} \equiv 0$ verifiziert. Um (1.14e) zu beweisen, betrachte man ein $v \in V$ wie zuvor jedoch mit $v_i \in C^\infty(\Omega) \cap V$. Dann fallen bei der partiellen Integration von (1.19a) nicht alle Randterme weg, sondern man erhält

$$\begin{aligned} 0 &= a(u,v) + b(v,p) - f(v) \\ &= \left(-\nu\Delta u_i + (\,\mathrm{D}p)_i - \tilde{f}_i, v_i\right)_\Omega + \nu \oint_{\Gamma_\mathrm{A}} n \cdot \mathrm{D}u_i v_i - \oint_{\Gamma_\mathrm{A}} n_i v_i p \\ &= (\nu n \cdot \mathrm{D}u_i - n_i p, v_i)_{\Gamma_\mathrm{A}} , \end{aligned}$$

wobei der letzte Schritt durch die bereits nachgewiesene L_2-Orthogonalität der Lösung zu V begründet ist. Da die C^∞-Funktionen auch in $L_2(\Gamma_\mathrm{A})$ dicht liegen, ergibt sich (1.14e) mit $f_\mathrm{A} \equiv 0$.
Aus $0 = b(u,q) = -(\,\mathrm{D}\cdot u, q)$ für alle $q \in Q$ folgt, daß $\mathrm{D}\cdot u \in L_2(\Omega)$ L_2-orthogonal zu Q ist. Falls $Q = L_2(\Omega)$ ist, ist u also fast überall Null, d. h. als stetige Funktion identisch Null. Im Fall reiner Dirichlet-Randbedingungen gilt $\mathrm{D}\cdot u \perp L_2^0(\Omega)$, somit ist $u \equiv C$ konstant. Der Wert der Konstante ergibt sich aus der Kompatibilitätsbedingung (siehe Abschnitt A.7) für reine Dirichletrandwerte: $0 = \int_{\partial\Omega} n \cdot u = \int_\Omega \mathrm{D}\cdot u = C|\Omega|$, ergo $u \equiv 0$. □

Inhomogene Randbedingungen

Randwerte werden bei Sobolevfunktionen immer im Spursinn verstanden: Ist Γ eine Teilmannigfaltigkeit von Ω, z. B. $\Gamma = \partial\Omega$, so existieren ein *Spuroperator* $T_{\Omega,\Gamma} \in \mathfrak{L}[H^1(\Omega), H^{1/2}(\Gamma)]$, der $Tu = u|_\Gamma$ für alle $u \in C^0(\overline{\Omega}) \cap H^1(\Omega)$ erfüllt, und ein *Fortsetzungsoperator* $E_{\Gamma,\Omega} \in \mathfrak{L}[H^{1/2}(\Gamma), H^1(\Omega)]$ mit $T \circ E = id_\Gamma$. Das kann in [42, Kap. I,§8] nachgelesen werden.

Dirichlet-Randbedingungen werden bei der schwachen Formulierung direkt in der Definition von V festgelegt, deshalb heißen sie auch *Zwangsbedingungen*. Will man also $u_\mathrm{D} : \Gamma_\mathrm{D} \longrightarrow \mathbb{R}^n$, $u_\mathrm{D} \neq 0$, vorschreiben so, verwendet man $V_\mathrm{D} = \{u \in H^1(\Omega) \mid Tu = u_\mathrm{D}\}$ und erhält die Aufgabe

$$a_\mathrm{D}(u,v) + b(v,p) = f(v) \quad \text{für alle } v \in V, \tag{1.40a}$$

$$b_\mathrm{D}(u,q) = 0 \quad \text{für alle } q \in Q \tag{1.40b}$$

mit $a_\mathrm{D} : V_\mathrm{D} \times V \longrightarrow \mathbb{R}$, $a_\mathrm{D}(u,v) = a(u,v)$ und $b_\mathrm{D} : V_\mathrm{D} \times Q \longrightarrow \mathbb{R}$, $b_\mathrm{D}(u,q) = b(u,q)$. Dabei wird $(u,p)^T \in V_\mathrm{D} \times Q$ gesucht. Man beachte, daß V_D kein linearer Raum und a_D weder bilinear noch symmetrisch ist. Auch b_D ist natürlich nicht linear in der ersten Komponente.
V und V_D sind jedoch parallele, affine Teilräume in $H^1(\Omega)$, d. h. $V_\mathrm{D} = V + \{E_{\Gamma_\mathrm{D},\Omega} u_\mathrm{D}\}$. Setzt man die Darstellung $(u_0,p)^T + (E_{\Gamma_\mathrm{D},\Omega} u_\mathrm{D}, 0)^T$, $(u_0,p)^T \in V \times Q$, einer Lösung von (1.40) dort ein und bringt alle Terme mit $E_{\Gamma_\mathrm{D},\Omega} u_\mathrm{D}$ auf die rechte Seite, so erhält man (1.19) für $(u_0,p)^T \in X$. Dazu muß man die rechte Seite $f(v) - a(E_{\Gamma_\mathrm{D},\Omega} u_\mathrm{D}, v)$ bzw. $-b(E_{\Gamma_\mathrm{D},\Omega} u_\mathrm{D}, q)$ wählen. Umgekehrt erhält man zu einer Lösung von (1.19) mit der gerade angegebenen rechten Seite eine Lösung von (1.40).

Inhomogene Ausströmungs-Randbedingungen werden durch zusätzliche Terme in den Stokes-Gleichungen gestellt und werden deshalb von Lösungen unabhängig von den genauen Funktionenräumen V und Q erfüllt. Deshalb heißen sie *natürliche Randbedingungen*. Im ersten Teil des Beweises von Satz 1.29 heben sich die Randintegrale des Diffusions- und Druckterms über Γ_A genau auf. Bei inhomogenen Randbedingungen erhält man stattdessen ein Randintegral von f_A, was auf folgende Aufgabe führt:

$$a(u,v) + b(v,p) = f(v) + \oint_{\Gamma_\mathrm{A}} f_\mathrm{A} T_{\Omega,\Gamma_\mathrm{A}} v \quad \text{für alle } v \in V, \tag{1.41a}$$

$$b(u,q) = 0 \quad \text{für alle } q \in Q. \tag{1.41b}$$

Daß jede hinreichend reguläre Lösung dieser Aufgabe (1.14e) erfüllt, zeigt der Beweis von Satz 1.29.

Aufgabe 1.30 (Stokes-RWA). Die Stokes-RWA mit beliebigen Dirichlet- und Ausströmungs- Randbedingungen ist durch Aufgabe 1.18 mit (1.35) bis (1.38) gegeben. Als rechte Seite tritt

$$\bar{f} : V \longrightarrow \mathbb{R} : v \longmapsto f(v) - a(E_{\Gamma_\mathrm{D},\Omega} u_\mathrm{D}, v) + \oint_{\Gamma_\mathrm{A}} f_\mathrm{A} T_{\Omega,\Gamma_\mathrm{A}} v$$

in Gleichung (1.19a) auf. Damit $\bar{f} \in V'$ gilt, wird $f \in V'$, $u_\mathrm{D} \in H^{1/2}(\Gamma_\mathrm{D})$, $f_\mathrm{A} \in L_2(\Gamma_\mathrm{A})$ vorausgesetzt.[18] Die rechte Seite von (1.19b) lautet

$$\bar{g} : Q \longrightarrow \mathbb{R} : q \longmapsto -b(E_{\Gamma_\mathrm{D},\Omega} u_\mathrm{D}, q).$$

Es wird ein Paar $(u,p)^T \in X$ gesucht, das (1.19) erfüllt. Dies wird als homogener Anteil der Lösung bezeichnet. Bei Referenzen auf die Lösung der vorliegenden Aufgabe ist stets der homogene Anteil $(u,p)^T$ gemeint. $(u,p)^T + (E_{\Gamma_\mathrm{D},\Omega} u_\mathrm{D}, 0)^T \in V_\mathrm{D} \times Q$ heißt (volle) Lösung der inhomogenen Stokes-RWA. Für die Herleitung der Fehlerschätzer in Kapitel 3 und 4 wird davon ausgegangen, daß die Repräsentation $f(v) = \int_\Omega fv$, $f \in L_2(\Omega)$, gilt.

[18] Vermutlich genügt $f_\mathrm{A} \in H^{1/2}(\Gamma_\mathrm{A})'$ anstelle von $f_\mathrm{A} \in L_2(\Gamma_\mathrm{A})$.

Nachweis der LBB-Bedingungen

Aufgrund von Lemma 1.28 genügt es, die LBB-Bedingungen (1.34) für die Stokes-RWA nachzuweisen, um diese als gut gestelltes Problem der mathematischen Physik zu erkennen. Denn die Existenz eines stetigen Lösungsoperators L^{-1} sichert die Gültigkeit aller drei geforderten Eigenschaften.

Satz 1.31 (LBB-Bedingungen an a). *Die Bilinearform a ist V-elliptisch und erfüllt* (1.34a) *und* (1.34b).

Beweis. Da $|\Gamma_{\mathrm{D}}| > 0$ ist (siehe Seite 21), gilt die Ungleichung von Poincaré [19]

$$\|u\|^2 \leq C(\Omega) \sum_{i=1}^{n} \| \mathrm{D}u_i\|_0^2$$

mit einer gebietsabhängigen Konstante $C(\Omega) > 0$ für alle $u \in V$. Also gilt nach Definition von a für beliebige $u \in V$

$$a(u,u) = \nu \sum_{i=1}^{n} (\, \mathrm{D}u_i,\, \mathrm{D}u_i) \geq \frac{\nu}{C(\Omega)+1} \|u\|_1^2.$$

Damit ist die V-Elliptizität von a bewiesen. Zum Beweis von (1.34a) betrachtet man ein beliebiges $u \in V$ mit $\|u\|_1 = 1$. Es ist $\sup\{|a(u,v)|\,|\, v \in V, \|v\|_V = 1\} \geq a(u,u) \geq \frac{\nu}{C(\Omega)+1}$ unabhängig von u, so daß auch das Infimum größer als die rechte Konstante ist. Da a symmetrisch ist, folgt (1.34b) unmittelbar aus (1.34a). □

Bemerkung 1.32. Die Elliptizitätskonstante E von a geht über Lemma 1.24 bzw. (1.34a) in die Norm von L^{-1} ein, die später für Schätzungen der Effizienz von Fehlerschätzern interessant ist. Für den Fall $\Gamma_{\mathrm{D}} = \partial\Omega$ beweist Galdi in [20, Kap. II.4] $C(\Omega) \leq \frac{d^2}{\pi^2}$, falls Ω in einem Streifen der Breite d liegt.
Im allgemeinen Fall kann man die skalare, schwache Eigenwertaufgabe $(\, \mathrm{D}u,\, \mathrm{D}v) - \lambda\,(u,v) = 0$ für alle $v \in H = \{v \in H^1(\Omega)|\; T_{\Omega,\Gamma_{\mathrm{D}}} v = 0\}$ lösen; u wird in derselben Menge gesucht. Für den kleinsten Eigenwert $\lambda_{\min}$ gilt dann: $C(\Omega) = \lambda_{\min}^{-1}$ ist die optimale *Poincare-Konstante*. Dies sieht man sofort ein, wenn man sich klar macht, daß die Eigenvektoren eine (Orthonormal-) Basis von H bilden und daß die Eigenwert-Gleichung der Poincare-Ungleichung entspricht, wenn man diese als Gleichung liest (siehe auch [20, Kap. II.4]).

[19] Der einfachste Beweis für reine Dirichlet-Randbedingungen via partieller Integration steht vermutlich in [42, Kap. 1, §7]. [42, Kap. 1, Satz 7.7] kann verwendet werden, um die hier verwendete Behauptung zu beweisen; die letzte Summe in Gleichung (8) kann entfallen, weil sie nur benutzt wird, um zu zeigen, daß das konstante Polynom ϕ (für $l = 1$) in $W_2^l(\Omega)$ identisch Null ist. Führt man den Beweis mit V aus dieser Arbeit durch, so ist $\phi \in V$ als konstantes Polynom wegen $|\Gamma_{\mathrm{D}}| > 0$ identisch Null.

Satz 1.33. *Ungleichung* (1.34c) *ist genau dann erfüllt, wenn es eine Konstante* $\beta > 0$ *gibt, so daß für jedes* $p \in Q$ *ein* $u \in V$ *mit*

$$\mathrm{D} \cdot u = p, \quad \|u\|_1 \leq \beta^{-1}\|p\| \tag{1.42}$$

existiert.

Beweis. Sei $p \in Q$, $\|p\| = 1$, beliebig und $u \in V$ die Lösung von (1.42). Dann ist $\sup_{v\in V,\|v\|_1=1}|b(v,p)| \geq |b(\frac{u}{\|u\|_1},p)| = \|p\|^2\|u\|_1^{-1} \geq \beta\|p\|$, so daß das Infimum über $p \in Q$ mit $\|p\| = 1$ in (1.34c) durch $\beta > 0$ nach unten beschränkt ist.
Umgekehrt impliziert (1.34c) nach Bemerkung 1.27 bereits, daß $B^{-1} \in \mathfrak{L}[Q', V]$ existiert. Zu beliebigem $p \in Q$ ist dann $u = B^{-1}p \in V$ mit $\beta^{-1} = \|B^{-1}\|$ eine Lösung von (1.42). □

In [20, Kap. III.3] wird (1.42) detailliert untersucht[20]. Für reine Dirichlet-Randbedingungen gilt (1.42) aufgrund von

Satz 1.34 ([20, Kap. III, Satz 3.1]). *Sei* Ω *ein beschränktes Gebiet im* $\mathbb{R}^n$, $n \geq 2$, *so daß* $\Omega = \cup_{k=1}^N \Omega_k$, *wobei jedes* Ω_k *sternförmig bezüglich einer offenen Kugel* B_k *mit* $\overline{B_k} \subset \Omega_k$ *sei. Zum Beispiel erfülle* Ω *die Kegeleigenschaft. Dann, gegeben* $p \in L_q(\Omega)$, *das* $\int_\Omega p = 0$ *erfüllt, existiert wenigstens eine Lösung* u *von* (1.42). *Weiterhin läßt die Konstante* β^{-1} *aus* (1.42) *folgende Abschätzung zu:*

$$\beta^{-1} \leq c_0 C \left(\frac{\mathrm{diam}(\Omega)}{R_0}\right)^n \left(1 + \frac{\mathrm{diam}(\Omega)}{R_0}\right),$$

worin R_0 *der kleinste Radius der Kugeln* B_k, $c_0 = c_0(n,q)$ *und*

$$C = \max_k \left(1 + \frac{|\Omega_k|}{|\Omega_k \cap D_k|}\right) \prod_{i=1}^{k-1} (1 + |F_i|^{1/q-1}|D_i - \Omega_i|^{1-1/q})$$

mit $D_i = \cup_{s=i+1}^N \Omega_s$ *und* $F_i = \Omega_i \cap D_i$, $i = 1, \ldots, N-1$, *ist. Schließlich hat* u *in* Ω *kompakten Träger, falls* p *dies auch hat.*

Galdi [20] bemerkt außerdem, daß die Konstante β in allen Gebieten dieselbe ist, die durch Homothetie und Abbildungen aus $SO(n)$ ineinander überführt werden können.
Ist $|\Gamma_\mathrm{A}| > 0$, so hat Q eine Dimension mehr als im Dirichletfall, denn die konstanten Funktionen kommen hinzu. Damit b (1.42) erfüllt, erwartet man, daß auch V um wenigstens eine Dimension größer wird.

Satz 1.35 (LBB-Bedingungen für $|\Gamma_\mathbf{A}| > 0$). *Im Fall* $|\Gamma_A| > 0$ *gilt* (1.42) *mit* $Q = L_2(\Omega)$ *und* $V = H_0^1(\Omega) = \{u \in H^1(\Omega) | \, T_{\Omega,\Gamma_D} u = 0\}$.

[20] $\mathrm{D} \cdot u = p$ stellt sich als ein überraschend diffiziles Problem heraus; weitere Literatur dazu siehe ebenfalls [20, Kap. III.7].

Beweis. Es wird folgende Aussage aus [20, Kap. III, Übung 3.4] verwendet: Die Aufgabe $\mathrm{D} \cdot u = p$, $T_{\Omega,\partial\Omega} u = a$ hat für jedes Paar $f \in L_2(\Omega)$, $a \in H^{1/2}(\partial\Omega)$, das die Kompatibilitätsbedingung $\int_\Omega f = \oint_{\partial\Omega} n \cdot a$ erfüllt, wenigstens eine Lösung, die der Abschätzung $\|u\|_1 \leq C(\|f\| + \|a\|_{1/2;\,\partial\Omega})$ genügt. Die Konstante C hängt weder von f noch von a ab.
Da $|\Gamma_\mathrm{A}| > 0$ ist, existiert eine Funktion $a \in H^{1/2}(\partial\Omega)$ mit $\operatorname{supp} a \subset \Gamma_\mathrm{A}$ und $\oint_{\Gamma_\mathrm{A}} n \cdot a = k > 0$.[21] Sei nun $p \in L_2(\Omega)$ beliebig. Da $L_2 = L_2^0 \oplus \operatorname{span}\{1\}$ eine orthogonale Zerlegung ist, kann man $p = p_0 + p_\perp$ mit $p_0 \in L_2^0$, $p_\perp \in \mathbb{R}$ und $\|p\|^2 = \|p_0\|^2 + \|p_\perp\|^2$ schreiben. Daraus folgt $\int_\Omega p = p_\perp |\Omega|$, so daß $a_\lambda = \lambda a \in H^{1/2}(\Gamma_\mathrm{A})$ mit $\lambda = \frac{p_\perp |\Omega|}{k} \in \mathbb{R}$ und P zusammen die Kompatibilitätsbedingung erfüllen. Es bleibt also noch zu zeigen, daß $\|a_\lambda\|_{1/2;\,\Gamma_\mathrm{A}} < C\|p\|$ gilt, um die Abschätzung in (1.42) nachzuweisen. Es ist $\|a_\lambda\|_{1/2;\,\Gamma_\mathrm{A}} = |\lambda| \|a\|_{1/2;\,\Gamma_\mathrm{A}} = |p_\perp|\,|\Omega|\,k^{-1} \|a\|_{1/2;\,\Gamma_\mathrm{A}} = \|p_\perp\| \sqrt{|\Omega|}\, k^{-1} \|a\|_{1/2;\,\Gamma_\mathrm{A}} \leq C\|p\|$, denn a wird nicht variiert. Damit ist alles bewiesen. □

[21] Zum Beispiel kann man $1 \cdot n$ auf einer kompakt in Γ_A enthaltenen Teilmenge glätten.

Kapitel 2

Diskretisierung der Variationsaufgabe

Es ist in der Praxis fast immer unmöglich, Aufgabe 1.18 durch Angabe einer expliziten Funktion exakt zu lösen. Deshalb ersetzt man die unendlich-dimensionalen Räume V, Q durch endlich-dimensionale Räume V_h, Q_h, $h > 0$, und erhält im Fall der Stokes-Gleichungen ein (großes) lineares Gleichungssystem anstelle von Aufgabe 1.18. Dieser Vorgang heißt *Galerkindiskretisierung*. Im folgenden wird die Lösbarkeit der diskreten Aufgabe untersucht und beleuchtet, inwiefern ihre Lösung gegen die von 1.18 konvergiert, wenn die *Gitterweite* h gegen Null strebt. In theoretischen Arbeiten über die Stokes-Gleichungen (z. B. [19], [20]) wird meist verwendet, daß Aufgabe 1.30 $H^1_{0,\mathrm{div}}(\Omega)$-elliptisch ist und sogar zu einem Tupel von Poisson-Gleichungen degeneriert. Das typische Vorgehen zur Diskretisierung solcher Aufgaben besteht in *konformer Diskretisierung*, d. h., man verwendet einen endlich-dimensionalen *Teil*raum $\tilde{X}_h \leq H^1_{0,\mathrm{div}}(\Omega)$ zur Diskretisierung und das *Lemma von Céa* sowie *Interpolationsungleichungen* zur Abschätzung des Diskretisierungsfehlers. Leider erweist es sich als schwierig, einfache Funktionen mit kleinem Träger zu finden, die in $H^1_{0,\mathrm{div}}$ liegen.[1] Im folgenden wird die Sattelpunktaufgabe auf X diskretisiert.

Da die diskrete Sattelpunktaufgabe

$$a(u_h, v_h) + b(v_h, p_h) = f(v_h) \quad \text{für alle } v_h \in V_h, \tag{2.1a}$$

$$b(u_h, q_h) = g(q_h) \quad \text{für alle } q_h \in Q_h \tag{2.1b}$$

dieselbe Struktur wie Aufgabe 1.18 hat, kann man die LBB-Bedingungen zur Untersuchung der Lösbarkeit heranziehen, indem man in (1.34) V_h und Q_h verwendet. Es wird die Bezeichnung $L_h : X_h \longrightarrow X_h' : x_h \longmapsto (Lx_h)|_{X_h}$ $(X_h = V_h \times Q_h)$ für die Diskretisierung von L verwendet, so daß (2.1) auch als

$$L_h x_h = r|_{X_h} \tag{2.2}$$

geschrieben werden kann.

[1] In dieser Richtung siehe z. B. [34], [36].

2.1 Abstrakte a priori Fehlerschätzung

Lemma 2.1. *Ist $V_h \leq V$, so ist die Bilinearform a der Stokes-Gleichungen V_h-elliptisch. Insbesondere erfüllt a für beliebige $Q_h \leq Q$ die Gleichungen* (1.34a) *und* (1.34b) *mit V_h und Q_h.*

Beweis. Die Elliptizität nachzuweisen ist trivial, da a V-elliptisch ist und $V_h \leq V$ gilt.
Der Beweis der LBB-Bedingungen (1.34a) und (1.34b) erfolgt dann analog zu Satz 1.31. □

Wären die Stokes-Gleichungen nur $H^1_{0,\mathrm{div}}$-elliptisch, so müßte statt des vorigen Lemmas eine aufwendigere Analyse erfolgen, weil im allgemeinen $V_h \setminus H^1_{0,\mathrm{div}} \neq \{0\}$ ist.
Der Nachweis von (1.34c) für die diskrete Aufgabe wird erst in Abschnitt 2.2.3 geführt, weil diese Bedingung im Gegensatz zu (1.34a) und (1.34b) sehr von der genauen Wahl der Räume V_h und Q_h abhängt. Vorerst wird die Gültigkeit von (1.34c) mit der von h unabhängigen Konstante $\beta_{\mathrm{diskret}} > 0$ vorausgesetzt, um den Diskretisierungsfehler abzuschätzen.

Satz 2.2 (Céa). *Seien $V_h \leq V$, $Q_h \leq Q$ so gegeben, daß L und L_h die LBB-Bedingungen erfüllen. Dann besitzen* (1.19) *und* (2.1) *jeweils genau die Lösung $x = (u,p)^T \in X$ und $x_h = (u_h, p_h)^T \in X_h$; diese erfüllen die Ungleichung*

$$\|x - x_h\|_X \leq C \inf_{y_h \in X_h} \|x - y_h\|_X$$

mit der Konstante $C = (1 + \|L_h^{-1}\|_{\mathcal{L}[X_h', X_h]} \|L\|_{\mathcal{L}[X, X']})$.

Beweis. Da die LBB-Bedingungen in beiden Fällen erfüllt sind, besitzen (1.19) und (2.1) je genau eine Lösung $x \in X$, $x_h \in X_h$. Betrachtet man ein beliebiges $y_h \in X_h$, so gilt aufgrund der Definition der Sattelpunktaufgaben, daß der Diskretisierungsfehler *galerkinorthogonal* zu y_h steht: $l(x - x_h, y_h) = 0$. Schreibt man diese Schlüsselrelation mit den zugehörigen Operatoren und subtrahiert Ly_h, so erhält man $L_h(x_h - y_h) = \big(L(x_h - y_h)\big)|_{X_h} = \big(L(x - y_h)\big)|_{X_h}$. Da L_h die LBB-Bedingungen erfüllt, kann es stetig invertiert werden, also

$$x_h - y_h = L_h^{-1}\big(L(x - y_h)\big)|_{X_h}.$$

Dies beweist mit der Dreiecksungleichung zusammen wegen $\|x - x_h\|_X \leq \|x - y_h\|_X + \|x_h - y_h\|_X \leq (1 + \|L_h^{-1}\|\,\|L\|)\|x - y_h\|_X$ durch Bilden des Infimums den Satz. □

Bemerkung 2.3. Da die Distanz von x_h zu x unter allen Elementen von X_h bis auf die (multiplikative) Konstante kleinstmöglich ist, nennt man die Galerkindiskretisierung *quasi-optimal*.

Man kann über (2.1) die sog. Ritz-Projektion $R_h : X \longrightarrow X_h$ definieren (siehe [27]). Satz 2.2 sagt dann, daß R_h zwar im allgemeinen ein schiefer Projektor ist, jedoch der Winkel zwischen x und x_h gleichmäßig von $\frac{\pi}{2}$ weg beschränkt bleibt, wenn C von h unabhängig ist.

Zusammen mit Approximationsaussagen über X_h bezüglich X erhält man aus Satz 2.2 Konvergenz in der X-Norm, also $\sqrt{\|u\|_1^2 + \|p\|^2}$ im Fall der Stokes-Gleichungen. Verwendet man die zu (1.19) duale Aufgabe, kann man Abschätzungen des Diskretisierungsfehlers in anderen Normen beweisen.
Hat man eine stetige, dichte Einbettung $\iota : X \hookrightarrow Y$, wobei Y ebenfalls ein Hilbertraum sei, so erhält man mit $\iota' : Y' \hookrightarrow X'$ und der Identifizierung von Y mit Y' via des Riesz-Isomorphismus einen *Gelfandschen Dreier* $X \hookrightarrow Y = Y' \hookrightarrow X'$. Damit kann man Fehlerabschätzungen in der Norm von Y gewinnen, denn ein Operator $O \in \mathcal{L}[X', X]$ (z. B. L^{-1}) liegt via der vorigen Inklusionen in anderen Klassen, z. B. in $\mathcal{L}[Y, Y]$. Eine ähnliche Technik kann auch auf a posteriori Fehlerschätzer angewendet werden, siehe Abschnitt 3.2.1.

Satz 2.4 (Aubin, Nitsche). *Sei $X \hookrightarrow Y \hookrightarrow X'$ ein Gelfandscher Dreier aus Hilberträumen, $X_h \leq X$; L bzw. L_h mögen die LBB-Bedingungen erfüllen. Die Lösungen der Aufgaben* (1.19) *und* (2.1) *werden wieder mit $x \in X$ und $x_h \in X_h$ bezeichnet. Außerdem wird für $y \in Y$ mit $\phi_y \in X$ die Lösung der zu* (1.19) *dualen Aufgabe $L'\phi_y = (y, \cdot)_Y$ bezeichnet. Dann existiert mit der Konstanten $C = \|L\|_{\mathcal{L}[X,X']}$ folgende Abschätzung des Diskretisierungsfehlers in der Y-Norm:*

$$\|x - x_h\|_Y \leq C\|x - x_h\|_X \sup_{y \in Y,\ \|y\|_Y = 1} \inf_{z_h \in X_h} \|\phi_y - z_h\|_X.$$

Beweis. Alle $x \in X$ erfüllen $\|x\|_Y = \|(x, \cdot)_Y\|_{Y'} = \sup_{y \in Y,\ \|y\|_Y = 1} |(x, y)_Y|$, also insbesondere $x - x_h$:

$$\|x - x_h\|_Y = \sup_{y \in Y,\ \|y\|_Y = 1} |(x - x_h, y)_Y|. \tag{2.3}$$

Da die stetige Invertierbarkeit von L' zu der von L äquivalent ist, kann die duale Aufgabe für alle rechten Seiten $y \in Y$ gelöst werden. Die Lösbarkeit von (1.19) und (2.1) folgt wieder aus den LBB-Bedingungen. Setzt man $x - x_h$ als Testfunktion in die duale Aufgabe ein, so ergibt sich $\langle L'\phi_y, x - x_h \rangle_{X',X} = (y, x - x_h)_Y$. Die Galerkinorthogonalität des Diskretisierungsfehlers für (1.19) führt für beliebige $z_h \in X_h$ zu $\langle L(x - x_h), z_h \rangle_{X',X} = 0$, was unter Verwendung der Definition des dualen Operators insgesamt $\langle L(x - x_h), \phi_y - z_h \rangle_{X',X} = (y, x - x_h)_Y$ liefert. Also gilt $|(y, x - x_h)_Y| \leq \|L\| \|x - x_h\|_X \inf_{z_h \in X_h} \|\phi_y - z_h\|_X$, was in (2.3) eingesetzt den Satz beweist. □

Der Satz kann also ohne weitere Regularitätsannahmen gezeigt werden. Um jedoch in dem sup inf-Term h-Potenzen zu finden, muß die zu (1.19) duale Aufgabe für alle rechten Seiten $y \in Y$ eine gewisse Regularität besitzen.

Bemerkung 2.5. Man kann sich fragen, warum im Satz 2.4 die duale Aufgabe in Erscheinung tritt. Im wesentlichen liegt das daran, daß man Y mit seinem Dualraum identifiziert hat und eigentlich die duale Norm abschätzt. Um den Operator L wieder ins Spiel zu bringen, muß eine Aufgabe für y formuliert werden, wo der Operator auf der „falschen“ Seite des Dualitätsproduktes steht.

2.2 Finite Elemente

Der vorliegende Abschnitt dient in erster Linie dazu, die Notation zu präzisieren, denn über die verwendeten Konzepte gibt es sehr viel Literatur, die hier nur zitiert wird. [15], [16] und [43] sind Standardwerke und gut lesbar. [26] kann ebenfalls empfohlen werden.

Definition 2.6 (geometrische Grundkonzepte). Ein nicht-entartetes s-Simplex $[x_0, \ldots, x_s]$ ist ein s-Tupel von Punkten $x_i \in \mathbb{R}^n$ $(s \leq n)$ in allgemeiner Lage. Da meist keine Verwechselungsgefahr besteht, bezeichnet $[x_0, \ldots, x_s]$ auch die Menge $\{\sum_{i=0}^{s} \lambda_i x_i \mid \sum_{i=1}^{s} \lambda_i = 1,\ \lambda_0, \ldots, \lambda_s > 0\}$, also die offene, konvexe Hülle der *Ecken* x_i. Als Gleichheit von Simplexen wird die Gleichheit der q-Tupel verwendet; ist nur die Gleichheit der Mengen gemeint, so wird $\approx$ geschrieben. 0-, 1-, 2- und 3-Simplexe tragen die Namen Punkt, Strecke, Dreieck, Tetraeder. Das Simplex $[0, e_1, e_2, \ldots, e_s]$ heißt Referenz-s-Simplex, wobei die e_i die ersten s Standard-Basisvektoren des $\mathbb{R}^n$ sind.
Ist $S = [x_0, \ldots, x_s]$ ein Simplex, so heißt R ein Randsimplex von S, $R \leq S$, falls Indizes $0 \leq i_0 < i_1 < \cdots < i_r \leq s$ existieren, so daß $R = [x_{i_0}, \ldots, x_{i_r}]$ gilt. Falls R dann ein 0-, 1-, $(s-1)$- Simplex ist, nennt man es Ecke, Kante, Facette von S. Es werden die Bezeichnungen $\mathcal{N}(S) = \{R \leq S \mid R \text{ ist eine Ecke von } S.\}$ und $\mathcal{F}(S) = \{R \leq S \mid R \text{ ist eine Facette von } S.\}$ eingeführt.
Mit $h_S = \operatorname{diam} S$ wird der Durchmesser, mit ρ_S der Inkugeldurchmesser von S bezeichnet. $\delta(S) = \frac{h_S}{\rho_S}$ ist das Entartungsmaß von S.

Bemerkung 2.7. Ein Simplex S ist genau dann entartet, wenn $\delta_S = \infty$ gilt. Für gleichförmige s-Simplexe ist $\delta_S = \sqrt{s(s+1)/2}$ minimal.

Ist $M \in \mathbb{R}^{n \times n}$ eine reguläre Matrix und $b \in \mathbb{R}^n$, so ist $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n : x \longmapsto Mx + b$ eine *affine Abbildung*. Die Inverse existiert und ist wegen $F^{-1}y = M^{-1}y + (-M^{-1}b)$ ebenfalls affin. Im gegenwärtigen Zusammenhang sind affine Abbildungen interessant, weil es zu zwei nicht-entarteten s-Simplexen $S = [x_0, \ldots, x_s]$, $T = [y_0, \ldots, y_s]$ genau eine affine Abbildung F gibt, so daß $F(S) = [Fx_0, \ldots, Fx_s] = T$ gilt. Man kann deshalb viele Probleme mit Simplexen durch die Untersuchung des entsprechenden Referenzsimplexes und eine affine Transformation behandeln. Einige wichtige Beziehungen zwischen S, T und F lauten so:

Lemma 2.8 (affine Abbildungen). *Mit den Bezeichnungen des vorangehenden Absatzes gilt:*

$$\|M\|_2 \leq \frac{h_T}{\rho_S}, \quad \|M^{-1}\|_2 \leq \frac{h_S}{\rho_T}, \quad |\det M| = \frac{|T|}{|S|}.$$

Falls $M = sO$ mit $s > 0$ und einer orthogonalen Matrix O gilt, so nennt man F eine *Ähnlichkeitsabbildung*. In diesem Fall erfüllen S und T die Beziehung $\delta_S = \delta_T$, das heißt, alle Simplexe derselben Ähnlichkeitsklasse besitzen dasselbe Entartungsmaß; Ähnlichkeit ist eine Äquivalenzrelation.

Definition 2.9 (Triangulierungen). Eine konsistente Triangulierung bzw. ein Simplizialkomplex $\mathcal{T}$ (im $\mathbb{R}^n$) ist eine endliche Menge von nicht-entarteten Simplexen mit den Eigenschaften:

1. Aus $S \in \mathcal{T}$ und $T \leq S$ folgt $T \in \mathcal{T}$.

2. Aus $S, T \in \mathcal{T}$ und $S \not\approx T$ folgt $S \cap T = \emptyset$.

Oft wird mit $\mathcal{T}$ auch der zugrundeliegende, abgeschlossene topologische Raum $\cup_{S\in\mathcal{T}} S \subset \mathbb{R}^n$ bezeichnet; solche abgeschlossenen topologischen Räume heißen Polyeder. Gilt in einer konsistenten Triangulierung zusätzlich für beliebige $R, S, T \in \mathcal{T}$ die Implikation $R \subseteq \overline{S} \cap \overline{T} \implies (R \leq S) \wedge (R \leq T)$, so heißt $\mathcal{T}$ konsistent numeriert.
Analog zu Definition 2.6 werden $\mathcal{N}(\mathcal{T}) = \{S \in \mathcal{T} \mid S \text{ ist ein Punkt.}\} \subseteq \mathcal{T}$ und $\mathcal{F}(\mathcal{T}) = \{S \in \mathcal{T} \mid S \text{ ist ein } (n-1)\text{-Simplex.}\} \subseteq \mathcal{T}$ als Bezeichnungen verwendet. Dabei wird davon ausgegangen, daß $\mathcal{T}$ wenigstens ein n-Simplex enthält. Allgemein wird für $1 \leq m \leq n$ die Menge aller m-Simplexe in $\mathcal{T}$ $\mathcal{T}^{(m)}$ genannt.
Die bereits erwähnte Gitterweite erhält für Triangulierungen die Definition $h = h(\mathcal{T}) = \max_{S\in\mathcal{T}} h_S$, und das Entartungsmaß einer Triangulierung lautet $\delta(\mathcal{T}) = \max_{S\in\mathcal{T}} \delta_S$.

Bemerkung 2.10. Das Besondere an konsistent numerierten Triangulierungen ist folgendes: Wenn sich zwei abgeschlossene Simplexe schneiden, so ist der Schnitt nicht nur ein gemeinsamer Randsimplex, sondern dann induzieren beide Simplexe die gleiche Eckpunktreihenfolge auf diesem Randsimplex. Diese Eigenschaft wird in dem Finite-Elemente-Paket DROPS [2] zur Implementierung eines effizienten Verfeinerungsalgorithmus für Triangulierungen verwendet.
Jede konsistente Triangulierung $\mathcal{T}$ kann konsistent numeriert werden, indem man die Punkte $P \in \mathcal{T}$ linear ordnet (z. B. mit natürlichen Zahlen abzählt) und die Reihenfolge der Eckpunkte in jedem Simplex $S \in \mathcal{T}$ dieser Ordnung anpaßt. Das besondere des Verfeinerungsalgorithmus in DROPS liegt darin, daß von DROPS

[2] DROPS wird im IGPM entwickelt. Eines der Anwendungsgebiete ist die Simulation von mehrphasigen Fluidsystemen. Es wird in C++ programmiert.

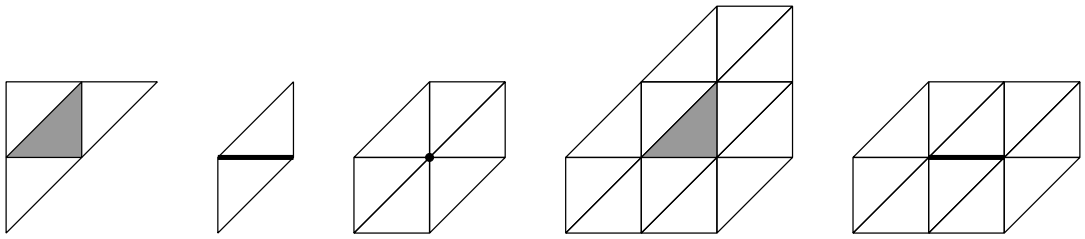

Abbildung 2.1: Die Umgebungen aus Notation 2.11 im $\mathbb{R}^2$. S ist ein Dreieck, F eine Strecke, x ein Punkt.

erzeugte Verfeinerungen einer konsistent numerierten Triangulierung ohne weitere Nachbearbeitung ebenfalls konsistent numeriert sind.

Obwohl Polyeder eine einfache Struktur aufweisen, sind sie im allgemeinen keine Lipschitzgebiete, da sie nicht unbedingt nur auf einer Seite ihres Randes zu liegen brauchen.

Notation 2.11. Sei $\Omega \subseteq \mathbb{R}^n$ ein beschränktes Gebiet mit Lipschitzrand, dessen Abschluß durch $\mathfrak{T}$ konsistent trianguliert ist. Mit

$$\mathcal{N}_\Omega(\mathfrak{T}) = \mathcal{N}(\mathfrak{T}) \cap \Omega, \quad \mathcal{N}_\mathrm{D}(\mathfrak{T}) = \mathcal{N}(\mathfrak{T}) \cap \Gamma_\mathrm{D}, \quad \mathcal{N}_\mathrm{A}(\mathfrak{T}) = \mathcal{N}(\mathfrak{T}) \cap \Gamma_\mathrm{A},$$
$$\mathcal{F}_\Omega(\mathfrak{T}) = \{S \in \mathcal{F}(\mathfrak{T}) \mid S \subset \Omega\}, \quad \mathcal{F}_\mathrm{D}(\mathfrak{T}) = \{S \in \mathcal{F}(\mathfrak{T}) \mid S \subset \Gamma_\mathrm{D}\},$$
$$\mathcal{F}_\mathrm{A}(\mathfrak{T}) = \{S \in \mathcal{F}(\mathfrak{T}) \mid S \subset \Gamma_\mathrm{A}\}$$

werden Teilmengen von $\mathcal{N}$ und $\mathcal{F}$ benannt. Zu jedem n-Simplex $S \in \mathfrak{T}$, jedem $(n-1)$-Simplex $F \in \mathfrak{T}$ und jedem Punkt $x \in \mathfrak{T}$ sind weitere Teilmengen von $\overline{\Omega}$ nützlich, die sogar als konsistente Teilkomplexe von $\mathfrak{T}$ aufgefaßt werden können:

$$\omega_S = \bigcup_{\mathcal{F}(S)\cap\mathcal{F}(T)\neq\emptyset} \overline{T}, \quad \omega_F = \bigcup_{F\in\mathcal{F}(T)} \overline{T}, \quad \omega_x = \bigcup_{x\in\mathcal{N}(T)} \overline{T},$$
$$\tilde{\omega}_S = \bigcup_{\mathcal{N}(S)\cap\mathcal{N}(T)\neq\emptyset} \overline{T}, \quad \tilde{\omega}_F = \bigcup_{\mathcal{N}(F)\cap\mathcal{N}(T)\neq\emptyset} \overline{T}.$$

Um den Umfang technischer Details nicht grenzenlos wachsen zu lassen, werden in dieser Arbeit nur Gebiete Ω behandelt, deren Abschluß ein Polyeder mit Lipschitzrand ist. Ferner sollen die Randbereiche mit Dirichlet- und Ausströmungs-Randbedingungen im nachfolgenden Sinn von jeder betrachteten Triangulierung $\mathfrak{T}$ aufgelöst werden:

Bedingung 2.12. Für jedes $S \in \mathfrak{T}$ folgt aus $S \cap \Gamma_\mathrm{D} \neq \emptyset$ die Aussage $S \subseteq \Gamma_\mathrm{D}$; ebenso folgt aus $S \cap \Gamma_\mathrm{A} \neq \emptyset$, daß $S \subseteq \Gamma_\mathrm{A}$ gilt.

Lemma 2.13 (Partitionen). *Unter der Bedingung 2.12 erlaubt $\mathfrak{T}$ die Partitionen $\mathcal{N}(\mathfrak{T}) = \mathcal{N}_\Omega \dot\cup \mathcal{N}_D \dot\cup \mathcal{N}_A$ und $\mathcal{F}(\mathfrak{T}) = \mathcal{F}_\Omega \dot\cup \mathcal{F}_D \dot\cup \mathcal{F}_A$.*

Bedingung 2.12 bedeutet im wesentlichen Einschränkungen an grobe Triangulierungen, weil auch sie die Bereiche mit unterschiedlichen Randbedingungen auflösen müssen – andererseits sind kompliziertere Konfigurationen mit DROPS derzeit nicht handhabbar[3] und könnten deshalb auch nicht numerisch getestet werden.
Zu den Verallgemeinerungen der abstrakten Fehlerschätzungen aus Abschnitt 2.1 dieses Kapitels und der Projektionsabschätzungen aus Abschnitt 2.2.1 auf den Fall krummliniger Ränder wird auf die eingangs erwähnte Literatur verwiesen.

Da letztendlich das asymptotische Verhalten des Diskretisierungsfehlers für $h \to 0$ interessant ist, werden später Familien $\{\mathcal{T}_h\}_{h>0}$ von Triangulierungen beliebig kleiner Gitterweite untersucht. In den relevanten Fehlerschätzungen tritt neben der Gitterweite auch das Entartungsmaß auf. Diese Abhängigkeit kann beseitigt werden, wenn eine endliche obere Schranke δ der Entartungsmaße aller $\mathcal{T}_h$ existiert. Eine positive untere Schranke ist durch Bemerkung 2.7 immer gegeben. Man verwendet diese

Definition 2.14 (reguläre Familien). Sei $\Omega \in \mathbb{R}^n$ ein Polyeder. Eine Familie $\{\mathcal{T}_h\}_{h>0}$ von konsistenten Triangulierungen von Ω heißt regulär, wenn sie $\delta = \sup_{h>0} \delta(\mathcal{T}_h) < \infty$ erfüllt.

Lemma 2.15. *Sei $\{\mathcal{T}_h\}_{h>0}$ eine Familie von konsistenten Triangulierungen des Polyeders $\Omega \in \mathbb{R}^n$. Dann gibt es Konstanten $C_1, C_2, C_3 \in \mathbb{N}$, die nur von δ abhängen, so daß für jede der Triangulierungen $\mathcal{T}_h$ gilt: Für jedes n-Simplex $S \in \mathcal{T}_h$ ist $|\{T \in \mathcal{T}_h \mid \mathcal{F}(S) \cap \mathcal{F}(T) \neq \emptyset\}| \leq C_1$. Jedes $(n-1)$-Simplex $F \in \mathcal{T}_h$ genügt $|\{T \in \mathcal{T}_h \mid F \in \mathcal{F}(T)\}| \leq C_2$. Die Ungleichung $|\{T \in \mathcal{T}_h \mid x \in \mathcal{N}(T)\}| \leq C_3$ wird von jedem $x \in \mathcal{T}_h$ erfüllt.*
Die Mengen ω aus Notation 2.11 enthalten also eine unabhängig von h beschränkte Anzahl von n-Simplexen aus $\mathcal{T}_h$.

Jetzt wird Notation für Funktionenräume bereitgestellt. $\mathcal{P}_k$ bezeichnet den Raum der Polynome auf $\mathbb{R}^n$, deren Grad kleiner oder gleich k ist.

Definition 2.16 (Finite-Elemente-Räume). Sei Ω ein beschränktes Gebiet mit konsistenter Triangulierung $\mathcal{T}$, das Bedingung 2.12 erfüllt. Dann sind

$$S^{k,-1} = \{f : \Omega \longrightarrow \mathbb{R} \mid f|_S \in \mathcal{P}_k \quad \text{für alle } S \in \mathcal{T}\},$$
$$S^k = S^{k,-1} \cap C^0(\Omega), \quad S_0^k = \left\{f \in S^k \mid f|_{\Gamma_D} \equiv 0\right\}$$

die hier verwendeten Finite-Elemente-Räume. Für vektorwertige Funktionen lautet die Definition bis auf den Wertebereich genauso. Durch Vorgabe von Funktionswerten auf dem Hauptgitter der Stufe k wird genau eine Funktion aus S^k festgelegt. Das Hauptgitter[4] $G_k(\mathcal{T})$ der Stufe k besteht dabei aus allen Punk-

[3] Das Einsatzgebiet von DROPS liegt im Bereich von Mehrphasensystemen, wo die Geometrie des Gebietsrandes nicht im Vordergrund steht. Die weitaus schwierigeren geometrischen Probleme an der Phasengrenze müssen mit Verfahren wie der Levelset- oder Phasenfeld-Methode gelöst werden, die nicht für den Gebietsrand eingesetzt werden.

[4] principal lattice; siehe [23]

ten $x \in \overline{\Omega}$, die auf folgende Weise geschrieben werden können: Es existieren ein n-Simplex $[x_0, \ldots, x_n] \in \mathcal{T}$ und baryzentrische Koordinaten $\lambda_j \in \{0, \frac{1}{k}, \ldots, 1\}$ $(j = 0, \ldots, n)$, so daß die Gleichungen $x = \sum_{j=0}^{n} \lambda_j x_j$ und $\sum_{j=0}^{n} \lambda_j = 1$ gelten. Daraus ergibt sich der lineare Standard-Interpolationsoperator $I^{\mathrm{S}} : C(\Omega) \longrightarrow S^k$; $I^{\mathrm{S}} f$ ist die eindeutig bestimmte Funktion in S^k, die $I^{\mathrm{S}} f = f$ auf $G_k(\mathcal{T})$ erfüllt.
Zu den Punkten des Hauptgitters existiert eine Lagrangebasis von S^k; sie wird auch Knotenbasis genannt. Finite Elemente, die ausschließlich auf der Vorgabe von Funktionswerten basieren, heißen dementsprechend Lagrange-Elemente.

Um zu Überprüfen, ob die gerade definierten Funktionenräume in H^1 liegen, wendet man partielle Integration über Simplexe an und erhält

Lemma 2.17 (Teilräume von $H^1(\Omega)$). *Ist $\Omega \subset \mathbb{R}^n$ ein durch $\mathcal{T}$ konsistent trianguliertes Gebiet mit Lipschitzrand und $f \in C^0(\Omega)$ eine Funktion, die auf jedem n-Simplex $S \in \mathcal{T}$ $f|_{\overline{S}} \in C^1(\overline{S})$ erfüllt, so folgt $f \in H^1(\Omega)$.*

Korollar 2.18. *Unter den Voraussetzungen von Lemma 2.17 gilt für jedes $k \in \mathbb{N}$ $S^k \leq H^1(\Omega)$.*

Satz 2.19 (Interpolationsfehler). *Gegeben seien natürliche Zahlen $0 \leq l \leq m \leq k+1$, $l \in \{0, 1\}$. Dann gibt es eine Konstante $C > 0$, die nur von Ω und δ abhängt, so daß jede Funktion $f \in H^m(\Omega) \cap C^0(\Omega)$ die Ungleichung*

$$|f - I^S f|_l \leq C h^{m-l} |f|_m$$

erfüllt. Darin ist I^S der Standard-Interpolationsoperator zu $S^k(\mathcal{T})$.

Der *Einbettungssatz von Sobolev* garantiert $H^l \hookrightarrow C^0$ für $l > \frac{n}{2}$. Während man damit im Fall $n \in \{2, 3\}$ bei der Poissongleichung noch zufrieden sein kann, weil man aus Satz 2.19 ohne $u \in H^2$ sowieso keine h-Potenzen in der Fehlerschätzung mit dem Lemma von Céa erhält, ist die Regularitätsforderung bei den Stokes-Gleichungen noch gravierender. Um nämlich den Druck $p \in L^2$ mit I^{S} interpolieren zu können, muß man auch $p \in H^2$ fordern.
Zum Glück ist Standard-Interpolation nicht das letzte Wort – Clément sowie Scott und Zhang und andere entwickelten ab den 1970er Jahren Interpolationsmethoden für „rauhe“ Funktionen.

2.2.1 Der Interpolationsoperator von Scott und Zhang

Funktionen aus Sobolevräumen können auch unter wesentlich milderen Voraussetzungen als Stetigkeit polynomial approximiert werden. Hier wird die Methode aus [31] vorgestellt. Für den Rest dieses Abschnittes wird vorausgesetzt, daß $\Omega \subset \mathbb{R}^n$ ein beschränktes Gebiet mit Lipschitzrand ist, dessen Abschluß durch $\mathcal{T}$ konsistent trianguliert wird. $k \geq 0$ ist die Ordnung des verwendeten Finite-Elemente-Raumes $S^k(\mathcal{T})$.
Jedem Punkt a_i des Hauptgitters $G^k(\mathcal{T})$ wird auf folgende Weise ein Trägersimplex $\sigma_i \in \mathcal{T}$ zugeordnet:

1. Falls es ein n-Simplex $S \in \mathfrak{T}$, $a_i \in S$, gibt, so setzt man $\sigma_i = S$.
2. Ansonsten wählt man ein $(n-1)$-Simplex $F \in \mathfrak{T}$ mit $a_i \in \overline{F}$ als σ_i. Falls $a_i \in \partial\Omega$ gilt, sei auch $\sigma_i \subset \partial\Omega$. Falls $a_i \in \Gamma_{\mathrm{D}}$ gilt, sei auch $\sigma_i \subseteq \Gamma_{\mathrm{D}}$. Diese Wahlen sind wegen Bedingung 2.12 stets möglich.

Dann faßt man die Punkte des Hauptgitters so zusammen: $a_{i,1} = a_i$, $\{a_{i,j}\}_{j=1}^{n_0} = G^k(\mathfrak{T}) \cap \overline{\sigma_i}$. Die Knotenbasis $\{\phi_{i,j}\}$ von $S^k(\overline{\sigma_i})$ hat eine $L_2(\sigma_i)$-Dualbasis $\{\psi_{i,j}\}$, die $\int_{\sigma_i} \phi_{i,j}\psi_{i,l} = \delta_{j,l}$ erfüllt. Man setzt nun $\psi_i = \psi_{i,1}$ und definiert den Operator

$$I^{\mathrm{Z}} : H^m(\Omega) \longrightarrow S^k(\mathfrak{T}) : f \longmapsto \sum_i \phi_i \int_{\sigma_i} \psi_i f. \tag{2.4}$$

Falls der Grad der zur Interpolation verwendeten Polynome eine Rolle spielt, wird $I^{\mathrm{Z},k}$ geschrieben. Da für $(n-1)$-Simplexe F der Spuroperator $T_{\Omega,F}$ existiert, ist I^{Z} für alle Sobolevräume H^m mit $m > \frac{1}{2}$ wohldefiniert. Die Hauptresultate über I^{Z} lauten zusammengefaßt so:

Satz 2.20 (lokale Interpolation mit I^{Z}). *Seien $0 \le l \le m \le k+1$ natürliche Zahlen und $m > \frac{1}{2}$. Ferner seien $S, F \in \mathfrak{T}$ ein beliebiges n-Simplex sowie eine beliebige Facette. Dann gilt für alle $f \in H^m(\tilde{\omega}_s)$*

$$\|f - I^Z f\|_{l;\,S} \le C h_S^{m-l} |f|_{m;\,\tilde{\omega}_S}$$

mit einer Konstante $C > 0$. C hängt nur von n, k und $\delta(\mathfrak{T})$ ab. Im Fall $m > l$ erfüllen alle $f \in H^m(\tilde{\omega}_F)$ mit einer Konstanten $C_2 > 0$, die dieselben Eigenschaften wie C besitzt, die Ungleichung

$$\|f - I^Z f\|_{l;\,F} \le C h_S^{m-l-\frac{1}{2}} |f|_{m;\,\tilde{\omega}_F}.$$

Satz 2.21 (Interpolation mit I^{Z}). *Seien $0 \le l \le m \le k+1$ natürliche Zahlen und $m > \frac{1}{2}$. Dann ist I^Z ein linearer Projektor von $H^m(\Omega)$ nach $S^k(\mathfrak{T})$. Es gibt eine Konstante $C > 0$, so daß für alle $f \in H^m(\Omega)$ gilt:*

$$\left(\sum_{S \in \mathfrak{T}^{(n)}} h_S^{2(l-m)} |f - I^Z f|_{l;\,S}^2 \right)^{\frac{1}{2}} \le C \|f\|_m. \tag{2.5}$$

Ist $S^k(\mathfrak{T}) \le H^l(\Omega)$, so folgt $I^Z \in \mathfrak{L}[H^m(\Omega), H^l(\Omega)]$ mit

$$\|f - I^Z f\|_l \le C h^{m-l} \|f\|_m \quad \text{für alle } f \in H^m(\Omega). \tag{2.6}$$

Falls Γ_D in $\partial\Omega$ abgeschlossen ist, so respektiert I^Z homogene Dirichlet-Randwerte, also $I^Z : H_0^m \longrightarrow S_0^k$.

Da auch bei der Approximation klassisch differenzierbarer Funktionen etwa mittels Taylorentwicklungen keine höheren Potenzen der Gitterweite als in (2.6) erzielt werden können, ist I^{Z} von *optimaler Ordnung*.

2.2.2 A priori Fehlerschätzung, Konvergenzanalyse

Jetzt stehen die Funktionenräume zur Verfügung, mit denen die diskrete Aufgabe (2.1) in dieser Arbeit formuliert wird. Man wählt $V_h = (S_0^{n_V}(\mathcal{T}))^n$. Ferner setzt man $Q_h = S^{n_Q}(\mathcal{T})$, falls $|\Gamma_{\mathrm{A}}| > 0$ gilt, ansonsten $Q_h = S^{n_Q}(\mathcal{T})/\mathbb{R}$. Dabei sind n_V und n_Q natürliche Zahlen.
Für jede Wahl von n_Q enthält S^{n_Q} die konstanten Funktionen, so daß man zu $p \in S^{n_Q}$ immer den Repräsentanten $p - |\Omega|^{-1} \int_\Omega p \in S^{n_Q} \cap L_2^0$ bilden kann.
Die Räume V_h, Q_h heißen *Taylor-Hood-Elemente* oder profaner $\mathcal{P}_{n_V}\mathcal{P}_{n_Q}$*-Elemente*.
Um den Diskretisierungsfehler zu beschränken, muß in Satz 2.2 der inf-Term analysiert werden. Dazu verwendet man eine Projektion der kontinuierlichen Lösung $x \in X$ von Aufgabe 1.30 auf X_h, so daß Satz 2.21 ins Spiel kommt. Ist $|\Gamma_{\mathrm{A}}| > 0$, so erhält man für den Projektor

$$I : X \longrightarrow X_h : (u,p)^T \longmapsto (I^{\mathrm{Z}}u_1, \ldots, I^{\mathrm{Z}}u_n, I^{\mathrm{Z}}p)^T \tag{2.7}$$

aus Satz 2.21 für jedes $x = (u,p)^T \in X \cap (H^2(\Omega) \times H^1(\Omega))$ die Fehlerabschätzung

$$\|x - Ix\|_X \leq Ch \left(\|u\|_2^2 + \|p\|_1^2\right)^{\frac{1}{2}}. \tag{2.8}$$

Die Scott-Zhang-Interpolationsoperatoren zu den Geschwindigkeitskomponenten bilden in S^{n_V} ab, die zum Druck in S^{n_Q}. Gilt $\partial\Omega = \Gamma_{\mathrm{D}}$, so wird in (2.7) die Interpolation für den Druck durch $(id - \pi_0)I^{\mathrm{Z}}p$ ersetzt. π_0 bezeichnet den $L_2(\Omega)$-Projektor auf die konstante Funktion 1. Aufgrund der Identität $(id - (id - \pi_0)I^{\mathrm{Z}})p = (id - \pi_0)(id - I^{\mathrm{Z}})p$ für alle $p \in S^{n_Q} \cap L_2^0(\Omega)$ erhält man auch in diesem Fall eine zu (2.8) analoge Abschätzung des Interpolationsfehlers.
Für diese Abschätzungen des Projektionsfehlers wird $n_V \geq 1$ und $n_Q \geq 0$ benötigt. Man erhält für Geschwindigkeit und Druck dieselbe h-Potenz in der Abschätzung, wenn $n_V = n_Q + 1$ gilt. Um zu verhindern, daß $Q_h/\mathbb{R}$ trivial ist, wählt man $n_Q \geq 1$. Mithin lautet das „einfachste“ brauchbare Taylor-Hood-Element $\mathcal{P}_2\mathcal{P}_1$.

Satz 2.22 (lineare Konvergenz). $\{\mathcal{T}_h\}_{h>0}$ *sei eine reguläre Familie konsistenter Triangulierungen von* Ω, $x = (u,p)^T \in X \cap (H^2(\Omega) \times H^1(\Omega))$ *die Lösung von Aufgabe 1.30 und* $x_h \in X_h$ *die Lösung der Diskretisierung* (2.1) *auf dem Taylor-Hood-Raum* X_h *zu* $\mathcal{T}_h$. *Dann gilt die folgende Abschätzung des Diskretisierungsfehlers mit einer von* h *unabhängigen Konstante* $C > 0$*:*

$$\|x - x_h\|_X \leq Ch \left(\|u\|_2^2 + \|p\|_1^2\right)^{\frac{1}{2}}.$$

Beweis. Man setze in Satz 2.2 die Projektion $Ix \in X_h$ ein, um das Infimum abzuschätzen. Aus (2.8) folgt dann sofort die Behauptung. □

Mit einem Dichtheitsargument kann man sich von der zusätzlichen Regularitätsbedingung in Satz 2.22 befreien. Das Ergebnis befriedigt aus theoretischer Sicht,

weil das Galerkinverfahren tatsächlich konvergiert. Aus praktischer Sicht nützt der nächste Satz wenig, weil man keine Aussage über die Konvergenzgeschwindigkeit erhält.

Satz 2.23 (Konvergenz des Galerkinverfahrens). *Es mögen die Voraussetzungen von Satz 2.22 ohne $x \in H^2(\Omega) \times H^1(\Omega)$ gelten. Dann erfüllt die Galerkinmethode* $\lim_{h\to 0^+} \|x - x_h\|_X = 0$.

Beweis. Zur Lösung $x \in X$ von Aufgabe 1.30 wähle man ein $y \in X \cap (H^2(\Omega) \times H^1(\Omega))$ mit $\|x - y\|_X \leq \varepsilon$. Das ist für jedes $\varepsilon > 0$ möglich, weil $H^2(\Omega) \times H^1(\Omega)$ dicht in X liegt. Nach (2.8) folgt für hinreichend kleines $h > 0$ $\|y - Iy\|_X \leq \varepsilon$. Aufgrund der Dreiecksungleichung gilt $\|x - Iy\|_X \leq \|x - y\|_X + \|y - Iy\|_X \leq 2\varepsilon$, so daß Satz 2.2 $\|x - x_h\|_X \leq 2C\varepsilon$ besagt. Das beweist den Satz. □

Bemerkung 2.24. Anstelle von (2.6) würde man gerne (2.5) verwenden, um (2.8) zu formulieren. Denn in (2.5) zeigt sich bereits, daß adaptiv verfeinerte Triangulierungen sinnvoll sind. In diese Fehlerschätzung geht nämlich nicht der für adaptive Triangulierungen künstliche Parameter h, sondern die Durchmesser h_S der einzelnen Simplexe ein.

Es sei $Y = L_2(\Omega) \times L_2(\Omega)$. Man kann versuchen, aus Satz 2.4 mit dem Gelfandschen Dreier $H_0^1(\Omega) \times L_2(\Omega) \hookrightarrow Y \hookrightarrow H_0^{-1}(\Omega) \times L_2(\Omega)$ eine Fehlerschätzung in der Y-Norm $\sqrt{\|u\|^2 + \|p\|^2}$ zu gewinnen. Diese ist „schwächer“ als die X-Norm, so daß man eine bessere Konvergenzaussage erwarten kann. Allerdings zeigt sich, daß man $L'^{-1} \in \mathfrak{L}[Y, H^2(\Omega) \times H^1(\Omega)]$ als Regularitätsbedingung fordern muß. Das ist aufgrund der Theorie aus Abschnitt 1.1.2 sogar für beliebig glatte Gebiete unrealistisch. Denn man erhält für p nur die Glattheit der rechten Seite g, also im vorliegenden Fall $p \in L^2(\Omega)$. Durch einen algebraischen Trick[5] kann man jedoch unter dieser schwächeren Bedingung immer noch bessere Konvergenz für die Geschwindigkeitskomponenten beweisen.

Satz 2.25 (quadratische Konvergenz der Geschwindigkeit). *$\{\mathfrak{T}_h\}_{h>0}$ sei eine reguläre Familie konsistenter Triangulierungen von Ω, $x = (u, p)^T \in X \cap (H^2(\Omega) \times H^1(\Omega))$ die Lösung von Aufgabe 1.30 und $x_h \in X_h$ die Lösung der Diskretisierung* (2.1) *auf dem Taylor-Hood-Raum X_h zu $\mathfrak{T}_h$. Ferner erfülle L' die Regularitätsbedingung*

$$L'^{-1} \in \mathfrak{L}[Y, \left(X \cap H^2(\Omega) \times L_2(\Omega), \sqrt{\|u\|_2^2 + \|p\|^2} \right)]. \tag{2.9}$$

Dann gilt die folgende Abschätzung des Diskretisierungsfehlers mit einer von h unabhängigen Konstante $C > 0$:

$$\|u - u_h\| \leq Ch^2 \sqrt{\|u\|_2^2 + \|p\|_1^2}.$$

[5] Salopp etwa: Alles Störende wird durch Übergang zu einem geeigneten Faktorraum beseitigt.

Beweis. Der bereits erwähnte Gelfandsche Dreier $V \times Q \hookrightarrow L_2(\Omega) \times Q \hookrightarrow H^{-1}(\Omega) \times Q$ geht durch Faktorisieren nach Q in den für die Poissongleichung bekannten Dreier $V \hookrightarrow L_2(\Omega) \hookrightarrow H^{-1}(\Omega)$ über. Die Einschränkung von L auf die Geschwindigkeitskomponenten ergibt den Operator $\tilde{L} \in \mathcal{L}[V \times Q/Q, V' \times Q'/Q']$. Nun wird Satz 2.4 angewendet, was Abschätzungen in den Normen der Faktorräume ergibt. Diese werden jetzt berechnet. Für beliebiges $y + Q = (w, q + Q)^T \in Y$ erhält man

$$\|y + Q\|_{Y/Q} = \inf_{r \in Q} \|y + r\|_Y = \inf_{r \in Q} \sqrt{\|w\|^2 + \|q + r\|_Q^2} = \|w\|.$$

Die X/Q-Norm eines beliebigen $y + Q = (w, q + Q)^T \in X/Q$ lautet durch eine analoge Rechnung überprüfbar $\|x + Q\|_{X/Q} = \|w\|_1$. Somit sagt Satz 2.4

$$\|u - u_h\| \leq C\|u - u_h\|_1 \sup_{y+Q \in Y/Q,\, \|y+Q\|_{Y/Q}=1} \inf_{z_h \in V_h} \|\phi_{y+Q} - z_h\|_1 \tag{2.10}$$

aus. $\phi_{y+Q} \in V_h$ ist hier der Geschwindigkeitsanteil der Lösung von $\tilde{L}'\phi_{y+Q} = y+Q$. Nach Satz 2.22 gilt $\|u-u_h\|_1 \leq Ch\sqrt{\|u\|_2^2 + \|p\|_1^2}$, was eine erste h-Potenz liefert. Um die zweite h-Potenz zu erhalten, wird der sup inf-Term analysiert. Man betrachte eine beliebige Restklasse $y + Q \in Y/Q$ mit $\|y + Q\|_{Y/Q} = 1$. Da L'^{-1} in $H^2(\Omega) \times L_2(\Omega)$ abbildet, ist $\phi_{y+Q} \in H^2(\Omega)$. Durch $z_h = I^Z\phi_{y+Q}$ erhält man mittels Satz 2.21 die Ungleichung $\inf_{z_h \in V_h} \|\phi_{y+Q} - z_h\|_1 \leq Ch\|\phi_{y+Q}\|_2$. Die Regularitätsbedingung impliziert, daß $\tilde{L}'^{-1} \in \mathcal{L}[Y/Q, (X \cap (H^2 \times L_2)/Q)]$ gilt. Somit ist $\|\phi_{y+Q}\|_2 \leq C\|y + Q\|_{Y/Q} \leq C$. Das beweist die quadratische Konvergenz der Geschwindigkeitskomponenten. □

Warum ist (2.9) eine Regularitätsbedingung? Man überlegt sich unter Zuhilfenahme des Gelfandschen Dreiers leicht, daß $L'^{-1} \in \mathcal{L}[Y, X]$ gilt. Aber (2.9) fordert, daß sogar Stetigkeit bezüglich der Norm $\sqrt{\|u\|_2^2 + \|p\|^2}$ im Bildraum vorliegt.

Bemerkung 2.26 *(Inhomogene Randwerte).* Bei der Formulierung von Aufgabe 1.30 tritt der Fortsetzungsoperator $E_{\Gamma_D,\Omega}$ auf. Für die Numerik benötigt man eine leicht berechenbare Version dieses Operators; für relativ glatte Daten bietet sich der Standard-Interpolationsoperator an. Sind die Daten nicht stetig auf Γ_D, so könnte man auf Interpolation nach Scott und Zhang zurückgreifen. Allerdings müssen dann die Integrale mit einem Verfahren berechnet werden, das nicht auf punktweiser Auswertung beruht. Es treten auch weitere Schwierigkeiten auf: Im allgemeinen ist $Iu_D \neq u_D$, weil $u_D \notin S^k|_{\Gamma_D}$ ist. Das diskrete Problem paßt also nicht mehr genau zum kontinuierlichen. Zwar kann dieser Fehler mit den Verallgemeinerungen des Lemmas von Céa (*Lemma von Strang*) behandelt werden, doch wird diese Fehlerquelle hier ignoriert. Die Testprobleme haben weitgehend glatte Randdaten, wie z. B. quadratische Einströmprofile, Ausströmen gegen konstanten Druck oder homogene Randbedingungen, so daß der Fehler auf feineren Triangulierungen vernachlässigt werden kann.

Des weiteren wird angemerkt, daß die Sätze 2.22, 2.23 und 2.25 auch für die volle Lösung der inhomogenen Aufgabe 1.30 gelten, wenn die Fortsetzung der Dirichlet-Randwerte hinreichend glatt ist: Dann erhält man durch Interpolation der Fortsetzung eine Näherung, die ebenfalls linear bzw. quadratisch in h konvergiert.

2.2.3 Stabilität der Taylor-Hood-Elemente

Abschnitt 2.2.2 legt nahe, die Finite-Elemente-Paare $\mathcal{P}_{k+1}\mathcal{P}_k$, $k \geq 1$ zu untersuchen. Da in DROPS $\mathcal{P}_2\mathcal{P}_1$-Elemente verwendet werden, wird in diesem Abschnitt die Stabilität, also die LBB-Bedingung

$$\inf\Big\{\sup\left\{|b(u_h, q_h)| \mid u_h \in V_h, \|u_h\|_V = 1\right\} \Big| \; q_h \in Q_h, \|q_h\|_Q = 1\Big\} = \beta_{\text{diskret}} > 0, \tag{2.11}$$

für das diskrete Problem, nachgewiesen. Ungleichung (2.11) entspricht (1.34c) bei der diskreten Aufgabe. Wie immer sei $\{\mathcal{T}_h\}_{h>0}$ eine reguläre Familie von konsistenten Triangulierungen des beschränkten Lipschitzgebietes $\Omega \subset \mathbb{R}^n$.
Zwar erspart man sich durch die Diskretisierung der Sattelpunktaufgabe das Problem, divergenzfreie Finite-Elemente-Räume zu konstruieren, doch der Nachweis der LBB-Bedingung ist auch nicht einfach. Viele Bücher, die das Thema behandeln, befassen sich nur mit $\Omega \subset \mathbb{R}^2$ (etwa [23]), und auch diese Beweise sind langwierig.

Hier wird ein Resultat von Stenberg aus [32] verwendet. Es beruht auf einer Zerlegung von $\mathcal{T}_h$ in *Makroelemente*, auf denen dann lokale finite Elemente definiert werden. Das Hauptresultat aus [32] besagt dann, daß (2.11) erfüllt ist, wenn auf jedem dieser Finite-Elemente-Räume eine einfache, algebraische Bedingung gilt. In [32] wird dabei stets eine Raumdimension von zwei oder drei vorausgesetzt. Da es bei dem im folgenden gegebenen Nachweis der algebraischen Bedingung nicht auf die Raumdimension ankommt, wird angenommen, daß die hier genannte, offensichtliche Verallgemeinerung des Hauptresultats aus [32] gilt. Es bleibt zu bemerken, daß der für DROPS relevante Fall $n = 3$ durch [32] und den folgenden Beweis explizit abgedeckt wird. Leider behandelt Stenberg ausschließlich Dirichlet-Randbedingungen. Obwohl eine Erweiterung seiner Methode auf Ausströmungs-Randbedingungen möglich zu sein scheint, wird für den Rest des Abschnittes $\Gamma_{\text{D}} = \partial\Omega$ vorausgesetzt, um nicht alle Resultate aus [32] hier nachrechnen zu müssen.
Sei nun eine Zerlegung $\mathcal{M}_h$ von $\mathcal{T}_h$ in Makroelemente gegeben. Dabei heißt $M \subset \Omega$ ein Makroelement, wenn es eine Teilmenge $T \subset \mathcal{T}_h$ gibt, die eine konsistente Triangulierung von M ist. Zusätzlich sei M zusammenhängend und enthalte wenigstens zwei n-Simplexe aus $\mathcal{T}_h$. An diese Zerlegung werden weitere Bedingungen gestellt.

Bedingung 2.27 (Makroelemente).

1. Für jedes $h > 0$ ist $\Omega \subseteq \cup_{M \in \mathcal{M}_h} M$ gegeben.

2. Alle $M \in \mathcal{M}_h$, $h > 0$ beliebig, liegen in endlich vielen Äquivalenzklassen im Sinne von Definition 2.29.

3. Jedes $(n-1)$-Simplex $F \in \mathcal{F}(\mathcal{T}_h)$ liegt für beliebiges $h > 0$ im Inneren von mindestens einem und höchstens L der Makroelemente aus $\mathcal{M}_h$. L hängt nicht von h ab.

Bemerkung 2.28. Man beachte, daß die Makroelemente nicht wie in [23] disjunkt sein müssen. Das bedeutet eine wesentlich größere Freiheit bei der Wahl der Triangulierungen $\mathcal{T}_h$ und des Verfeinerungsalgorithmusses.

Definition 2.29 (Äquivalenz von Makroelementen). Ein Makroelement M heißt genau dann äquivalent zu einem Referenzmakroelement $\hat{M}$, wenn eine stetige, bijektive Abbildung $F_M : \hat{M} \longrightarrow M$ mit den nachfolgenden Eigenschaften existiert: $F_M(\hat{M}) = M$. Wenn $\{S_i\} \subseteq \mathcal{T}_h$ die Menge der Simplexe mit $S_i \subset \hat{M}$ darstellt, so ist $\{F_M(S_i)\} \subset \mathcal{T}_h$ die Menge aller Simplexe, die in M liegen. Ferner ist $F_M|_{S_i} = F_{F_M(S_i)} \circ F_{S_i}^{-1}$, wobei $F_{F_M(S_i)}$ bzw. F_{S_i} die Abbildungen sind, die den Referenzsimplex auf $F_M(S_i)$ bzw. S_i abbilden.

Hier werden die Makroelemente

$$\mathcal{M}_h = \{\omega_x \mid x \in \mathcal{N}_\Omega(\mathcal{T}_h)\} \tag{2.12}$$

gewählt, also die Nachbarschaften der inneren Ecken von $\mathcal{T}_h$. Damit diese Wahl Bedingung 2.27 genügt, müssen die Triangulierungen zwei leichte Einschränkungen hinnehmen.
Triangulierungen, die nicht Bedingung 2.30 gehorchen, können natürlich dennoch zu stabilen Finite-Elemente-Paaren führen. Allerdings kann die Stabilität dann nicht mit den Methoden aus [32] bewiesen werden kann.

Bedingung 2.30 (Triangulierungen).

1. $\mathcal{N}_\Omega(\mathcal{T}_h)$ ist für jedes $h > 0$ nicht leer. Das ist eine sehr milde Einschränkung, die allenfalls grobe Ausgangstriangulierungen betrifft.

2. Jedes n-Simplex $S \in \mathcal{T}_h$, $h > 0$ beliebig, hat eine Ecke $x \leq S$ in $\mathcal{N}_\Omega(\mathcal{T}_h)$. Es existieren ab $n = 2$ beliebig feine Triangulierungen, die diese Bedingung verletzen. Allerdings gibt es ein konstruktives Verfahren, um die Gültigkeit dieser Einschränkung zu sichern. Um die Präsentation des Stabilitätssatzes nicht zu sehr hinauszuzögern, wird es in einem Unterabschnitt auf Seite 47 angegeben.

Bemerkung 2.31. Falls Punkt zwei der vorherigen Bedingung gilt, so auch ihr erster Teilpunkt. Denn konsistente Triangulierungen eines Gebietes $\Omega \subset \mathbb{R}^n$ enthalten per Definition wenigstens ein n-Simplex. Dies besitzt nach Punkt zwei eine Ecke in Ω.

Satz 2.32. *Falls die reguläre Familie* $\{\mathcal{T}_h\}_{h>0}$ *Bedingung 2.30 erfüllt, so gilt für* $\{\mathcal{M}_h\}_{h>0}$ *aus* (2.12) *Bedingung 2.27.*

Beweis. Es wird die reguläre Familie $\{\mathcal{T}_h\}_{h>0}$ untersucht. Sie möge Bedingung 2.30 erfüllen. Jetzt werden die Punkte aus Bedingung 2.27 nachgewiesen.
Zu 1: Man betrachte zu einem beliebigen $h > 0$ ein beliebiges n-Simplex $S \in \mathcal{T}_h$. Wegen 2.30.2 existiert eine Ecke $x \leq S$ mit $x \in \Omega$. Weil Ω offen ist, folgt $x \in \mathcal{N}_\Omega(\mathcal{T}_h)$. Nach Notation 2.11 hat man $\overline{S} \subset \omega_x$. Da S beliebig gewählt ist, ergibt sich daraus die zweite Inklusion in

$$\Omega \subset \bigcup_{S \in \mathcal{T}_h^{(n)}} \overline{S} \subseteq \bigcup_{x \in \mathcal{N}_\Omega(\mathcal{T}_h)} \omega_x.$$

Das beweist den ersten Punkt.

Zu 3: Zuerst wird gezeigt, daß jedes $F \in \mathcal{F}_\Omega(\mathcal{T}_h)$, $h > 0$ beliebig, im Inneren wenigstens eines Makroelementes liegt. Jede Ecke $x \leq F$ von F liegt in $\mathcal{N}_\Omega$. Da wenigstens ein n-Simplex S mit $F \leq S$ existiert, folgt $F \subset \overline{S} \subset \omega_x$. Somit bleibt $\partial\omega_x \cap F = \emptyset$ zu zeigen. $\mathcal{T}_h$ induziert eine konsistente Triangulierung von ω_x durch einen *Teilkomplex* $\mathcal{T} \leq \mathcal{T}_h$. Aufgrund von $S \in \mathcal{T}$ sind $x, F \in \mathcal{T}$. Es wird verwendet, daß der Rand $\partial\mathcal{T}$ eines Simplizialkomplexes immer ein Teilkomplex ist[6]. Angenommen, es wäre $\partial\omega_x \cap F \neq \emptyset$, so gäbe es ein $R \in \partial\mathcal{T}$ mit $F \cap R \neq \emptyset$. Wegen Definition 2.9 erhielte man $F = R \subset \partial\omega_x$. Es gälte sogar $\overline{F} \subseteq \partial\omega_x$, weil der Rand einer Menge stets abgeschlossen ist. Somit wäre $x \in \partial\omega_x$. Man überlegt sich aber schnell, daß x im Innern von ω_x liegt, wenn $x \in \mathcal{N}_\Omega$ wahr ist. Also erhielte man einen Widerspruch. Das demonstriert die erste Hälfte von Punkt 3.
Es wird jetzt nachgewiesen, daß zu einem beliebigen $(n-1)$-Simplex $F \in \mathcal{F}_\Omega(\mathcal{T}_h)$ höchstens $n+2$ Punkte $x \in \mathcal{N}_\Omega(\mathcal{T}_h)$ existieren, die $F \subseteq \omega_x$ verifizieren. Sei $F \subset \omega_x$. Es gibt maximal zwei n-Simplexe $S_1, S_2 \in \mathcal{T}_h$, in deren Rand sich F befindet. Da $F \subseteq \omega_x$ gilt, ist mindestens ein S_i in ω_x. Denn ist x eine der Ecken von F, so auch von S_1 und $S_1 \subset \omega_x$. Ansonsten ist $T = [x, f_1, \ldots, f_n]$ ein n-Simplex in ω_x mit $F \leq T$, also $T \in \{S_1, S_2\}$. Dabei sind die f_i die Ecken von F.
Nun gilt für ein n-Simplex $S \in \mathcal{T}_h$ per Definition von ω_y

$$S \subset \omega_y \iff y \text{ ist Ecke von } S.$$

Somit ist x eine der insgesamt höchstens $n + 2$ Ecken von S_1 und S_2. Mithin ist Punkt 3 vollständig bewiesen.

Zu 2: Der Nachweis dieser Eigenschaft zerfällt in drei Teile – als erstes wird eine Äquivalenzrelation zwischen Makroelementen eingeführt. Danach wird gezeigt, daß nur endlich viele Äquivalenzklassen existieren, und zum Schluß wird die letzte Bedingung aus Definition 2.29 überprüft.

[6] Zur algebraischen Topologie siehe z. B. [33].

1. Es sei $x \in \mathcal{N}_\Omega(\mathcal{T}_h)$ beliebig. $\mathcal{T}_h$ induziert die konsistente Triangulierung $\mathcal{T}_x \subseteq \mathcal{T}_h$ auf ω_x. Man setze $\nu = |\mathcal{N}(\mathcal{T}_x)| \in \mathbb{N}$. Sei nun eine Abzählung von $\mathcal{N}(\mathcal{T}_x)$ mit natürlichen Zahlen durch $\{x_1, x_2, \ldots, x_\nu\}$ gegeben. Durch diese Abzählung läßt sich jedes m-Simplex $S = [x_{i_1}, \ldots, x_{i_m}] \in \mathcal{T}_x$ eindeutig mit dem Multiindex $\alpha_S = (i_1, \ldots, i_m) \in \{1, \ldots, \nu\}^m$ identifizieren. Dann heißt $\mathcal{A}(\mathcal{T}_x) = \{\alpha_S \mid S \in \mathcal{T}_x\}$ der zu $\mathcal{T}_x$ und zur verwendeten Abzählung der Ecken gehörende *abstrakte Simplizialkomplex.*[7] Ist $y \in \mathcal{N}_\Omega(\mathcal{T}_h)$ beliebig, so heißen ω_x und ω_y äquivalent, wenn es Abzählungen ihrer Eckenmengen gibt, so daß $\mathcal{A}(\mathcal{T}_x) = \mathcal{A}(\mathcal{T}_y)$ erfüllt ist. Man überzeugt sich ohne Schwierigkeiten davon, daß diese Beziehung tatsächlich eine reflexive, symmetrische, transitive Relation auf $\cup_{h>0}\mathcal{M}_h$ formt. Lediglich bei der Transitivität muß man etwas aufpassen, weil der „Mittelmann“ auf zwei unterschiedliche Weisen numeriert sein kann. Da es sich aber um Abzählungen derselben Menge handelt, können sie durch eine Permutation ineinander überführt werden.

2. Wie viele Äquivalenzklassen gibt es in $\cup_{h>0}\mathcal{M}_h$ unter der Relation aus Punkt 1? Da die Familie $\{\mathcal{M}_h\}_{h>0}$ regulär ist, existiert nach Lemma 2.15 eine natürliche Zahl C, so daß für jedes $h > 0$ und jedes $x \in \mathcal{N}(\mathcal{T}_h)$ die Abschätzung $|\{S \subseteq M \mid S \in \mathcal{T}_h\}| \leq C$ richtig ist.[8] Insbesondere gilt $|\mathcal{N}(\mathcal{T}_x)| \leq C$. Die Zahl der abstrakten Simplizialkomplexe mit höchstens C Ecken ist endlich: Maximal $\sum_{\nu=0}^{n} C^{\nu+1}$ abstrakte Simplexe mit einer Dimension kleiner gleich n existieren, weshalb die Anzahl möglicher abstrakter Simplizialkomplexe durch das Maß $2^{\sum_{\nu=0}^{n} C^{\nu+1}}$ der Potenzmenge beschränkt ist. Da jeder abstrakte Simplizialkomplex eine Äquivalenzklasse repräsentiert, ist damit dieser Unterpunkt bewiesen.

3. Nun wird eine Abbildung $F_{M,M'} : M \longrightarrow M'$ zwischen zwei beliebigen im Sinne von Punkt 1 äquivalenten Makroelementen $M = \omega_x \in \cup_{h>0}\mathcal{M}_h$, $M' = \omega_{x'} \in \cup_{h>0}\mathcal{M}_h$ angegeben, die die Eigenschaften aus Definition 2.29 besitzt. Es wird vorausgesetzt, daß die Ecken von M und M' so numeriert sind, daß $\mathcal{A}(\mathcal{T}_x) = \mathcal{A}(\mathcal{T}_{x'})$ gilt.

   Ist $x_i \in \mathcal{N}(\mathcal{T}_x)$ so setzt man $F_{M,M'}(x_i) = x'_i$. Sei $S = [x_{i_1}, \ldots, x_{i_m}] \in \mathcal{T}_x$ ein beliebiges m-Simplex und $y \in S$ mit den *baryzentrischen Koordinaten* $(\lambda_i)_{i=1}^m$. Dann ist via $F_{M,M'}(y) = \sum_{i=1}^m \lambda_i x'_i$ auf ganz M eine Funktion definiert, deren Bild in M' liegt.

   Offensichtlich gilt $F_{M,M'}(\mathcal{N}(\mathcal{T}_x)) = \mathcal{N}(\mathcal{T}_{x'})$. Da die abstrakten Simplizialkomplexe von M und M' übereinstimmen, erfüllt jedes m-Simplex $S = [x_{i_1}, \ldots, x_{i_m}] \in \mathcal{T}_x$ die Aussage

$$[F_{M,M'}(x_{i_1}), \ldots, F_{M,M'}(x_{i_m})] = [x'_{i_1}, \ldots, x'_{i_m}] \in \mathcal{T}_{x'}. \tag{2.13}$$

[7] Abstrakte Simplizialkomplexe sind eine Verallgemeinerung gerichteter Graphen.

[8] Man verwendet zusätzlich die Aussage, daß die Anzahl der Simplexe in $\overline{[x_1, \ldots, x_m]} = 2^{m+1} - 1$ ist.

Somit ist $F_{M,M'}$ eine *simpliziale Abbildung* im Sinne von [33, Definition 3.1.16] zwischen $\mathcal{T}_x$ und $\mathcal{T}_{x'}$. Aufgrund von [33, Satz 3.1.18] induziert sie eine stetige Abbildung $|F_{M,M'}| : M \longrightarrow M'$ und wegen (2.13) wird jedes nicht-entartete m-Simplex aus $\mathcal{T}_x$ auf ein nicht-entartetes m-Simplex aus $\mathcal{T}_{x'}$ abgebildet. Deshalb kann man [33, Bemerkung 3.1.19 (b)] anwenden: $|F_{M,M'}|$ ist bijektiv und stückweise affin.

Da $|F_{M,M'}|$ auf jedem n-Simplex in $\mathcal{T}_x$ affin ist, kann man, wie in Definition 2.29 als letztes gefordert wird, die Einschränkung von $|F_{M,M'}|$ auf S als Verkettung der Affinitäten schreiben, die das Referenzsimplex auf S bzw. $F_{M,M'}(S)$ abbilden. Damit ist nachgewiesen, daß M und M' äquivalent nach Definition 2.29 sind.

□

Auf den Makroelementen $M \in \mathcal{M}_h$ werden lokale finite Elemente definiert:

$$V_{M,0} = \left\{ v \in H^1(M) \;\middle|\; v|_{\partial M} \equiv 0, v \in S^2(M) \right\}, \tag{2.14}$$

$$Q_M = \left\{ p \in C^0(M) \;\middle|\; p \in S^1(M) \right\}, \tag{2.15}$$

$$N_M = \{ p \in P_M \mid (\,\mathrm{D} \cdot v, p)_M = 0 \quad \text{für alle } v \in V_{M,0} \}. \tag{2.16}$$

M trägt die durch $\mathcal{T}_h$ induzierte konsistente Triangulierung. Man beachte, daß in der Definition von N_M die Gleichung $b(v,p) = 0$ steht. N_M ist der Annihilator von $V_{M,0}$ in P_M bezüglich b; diese Teilmenge von P_M darf nach Aussage des Hauptresultates aus [32] nicht zu groß sein. Dies präzisiert die Anschauung, daß der Finite-Elemente-Raum des Druckes nicht zu groß sein darf, damit die Divergenzfreiheit nicht zu viele Einschränkungen an die Geschwindigkeit liefert.

Satz 2.33 ([32, Satz 3.1]). *Ist $\mathcal{M}_h$ eine Familie von Makroelement-Zerlegungen von $\{\mathcal{T}_h\}$, die Bedingung 2.27 erfüllt, so gilt:*
Falls N_M für jedes $M \in \mathcal{M}_h$, $h > 0$, nur aus den konstanten Funktionen besteht, so gilt die diskrete LBB-Bedingung (2.11) *mit einer von h unabhängigen Konstante $\beta_{diskret} > 0$.*

Bevor nachgewiesen wird, daß die in der vorliegenden Arbeit verwendete Diskretisierung tatsächlich stabil ist, wird eine Aussage über ein spezielles Quadraturschema vorgestellt.

Lemma 2.34 (Quadraturformel). *Es wird ein beliebiges n-Simplex S mit Ecken x_i und Kantenmittelpunkten y_j betrachtet. Es existieren zwei nur von n abhängige Gewichte $w_1, w_2 \in \mathbb{R}$, so daß für jedes $p \in \mathcal{P}_2(S)$ die Integrationsformel*

$$\int_S p = Qp = |S| \left(w_1 \sum_i p(x_i) + w_2 \sum_j p(y_j) \right)$$

gilt. Überdies ist stets $w_2 > 0$.

Beweis. Als erstes wird das Lemma für das Referenz-n-Simplex S bewiesen. $\mathcal{P}_2$ besitzt eine an $G_2(S)$ angepaßte Lagrange-Basis $\{p_i, q_j\}$, das heißt $p_i(x_k) = \delta_{i,k}$, $p_i(y_j) = 0$, $q_j(y_k) = \delta_{j,k}$, $q_j(x_i) = 0$ für alle Indizes i, j, k. Jedes Polynom $f \in \mathcal{P}_2$ hat darin die Darstellung $f = \sum_i f(x_i)p + \sum_j f(y_j)q$. Die Existenz einer auf $\mathcal{P}_2$ exakten Quadraturformel mit den behaupteten Stützstellen ergibt sich sofort durch Integration dieser Darstellung. Die Gewichte lauten $a_i = \int_S p_i$ bzw. $b_j = \int_S q_j$.
Seien nun x_i und x_k zwei Ecken von S. Dann existiert eine affine, bijektive Abbildung $F : S \longrightarrow S$, die x_i auf x_k abbildet. Ihre Determinante hat den Betrag 1, da die Maße von Definitionsbereich und Bild identisch sind (Lemma 2.8). Ferner gilt $p_i = p_k \circ F$, so daß der Transformationssatz $a_i = \int_S p_i = \int_S p_k = a_k$ liefert. Die a_i sind also paarweise gleich und es wird $w_1 = a_1$ gesetzt.
Für die Gewichte der Kantenmittelpunkte kann eine analoge Schlußfolgerung durchgeführt werden. Es sei $w_2 = b_1$.
Der Schritt, mittels affiner Abbildungen und des Transformationssatzes eine Quadraturformel vom Referenzsimplex auf ein beliebiges Simplex gleicher Dimension zu übertragen, wird hier nicht näher erläutert (siehe z. B. [16]).
Betrachtet man auf dem Referenzsimplex $S = [0, e_1, \ldots, e_n]$ die Funktion $p(x) = x_1(1-x_1) \in \mathcal{P}_2$, so stellt man fest: p verschwindet in allen Ecken von S; außerdem ist $p(x) > 0$ auf S. Mithin erhält man aus $0 < \int_S p = Q(p) = w_2 \sum_j p(y_j) = w_1 C$ mit $C > 0$, daß $w_2 > 0$ wahr ist. Also ist das Lemma bewiesen. □

Bemerkung 2.35. Für Tetraeder gilt z. B. $w_1 = -\frac{1}{20}$, $w_2 = \frac{1}{5}$; insbesondere ist $w_1 < 0$. Deswegen und wegen der unnötig hohen Zahl an Stützstellen wird diese Quadraturformel in DROPS nicht eingesetzt.

Satz 2.36 (N_M für $\mathcal{P}_2\mathcal{P}_1$ – *Elemente*). *Falls $\{\mathcal{T}_h\}_{h>0}$ eine reguläre Familie konsistenter Triangulierungen von Ω ist, welche die Bedingung 2.30 erfüllt, so besteht N_M bei den $\mathcal{P}_2\mathcal{P}_1$-Elementen nur aus den konstanten Funktionen. $M \in \mathcal{M}_h$, $h > 0$ steht dabei für ein beliebiges Makroelement.*

Beweis. Es wird ein beliebiges Makroelement $M = \omega_x \in \mathcal{M}_h$ mit beliebigem $h > 0$ und eine beliebige Funktion $p \in N_M$ betrachtet. Die Beweisidee lautet, zu p eine Funktion $v \in V_{M,0}$ zu konstruieren, mit der $\mathrm{D}p = 0$ nachgewiesen wird. v wird durch die Angabe der Funktionswerte auf $G_k(M)$ eindeutig beschrieben. Es sei $v(x) = 0$.
Da $S^1(M) \leq H^1(M)$ gilt, kann man die Definitionsgleichung von N_M partiell integrieren: $0 = (\mathrm{D} \cdot v, p)_M = -(v, \mathrm{D}p)_M = -\int_M v \cdot \mathrm{D}p$. $\mathrm{D}p$ ist stückweise konstant, so das $(v \cdot \mathrm{D}p)|_S \in \mathcal{P}_2$ liegt. $S \subset M$ bezeichnet ein beliebiges n-Simplex und $g_S = \mathrm{D}p|_S$ sei der konstante Wert von $\mathrm{D}p$ auf S.
Nach Lemma 2.34 existiert auf S eine Quadraturformel, deren Stützstellen die Ecken und Kantenmittelpunkte von S sind. Dabei lautet das Gewicht in den Ecken w_1, in den Kantenmittelpunkten $w_2 > 0$. Sie integriert auf $\mathcal{P}_2$ exakt. $\{x_i\}$

bezeichnet die Menge der Kantenmittelpunkte von Kanten $K_i \in \mathcal{T}_h$, die im Inneren von ω_x liegen. Das sind genau jene Kanten mit x als Ecke. Nun ist v auf ∂M und in x Null, so daß

$$\int_M v \cdot \mathrm{D}p = \sum_{S \subset M, S \in \mathcal{T}^{(n)}} \int_S v \cdot \mathrm{D}p = \sum_{S \subset M, S \in \mathcal{T}^{(n)}} \sum_{x_i \in \overline{S}} w_2 \, v(x_i) \cdot g_S$$

gilt. Jetzt werden die Werte von v in den x_i festgelegt. Mit r_i wird der Einheitsvektor in Richtung $x_i - x$ bezeichnet. Wegen $p \in H^1(M)$ ist $r_i \cdot \mathrm{D}p$ auf der gemeinsamen Facette F zweier beliebiger n-Simplexe aus $\mathcal{T}_h$, die in M liegen, stetig, wenn sich die Kante zu r_i im Rand von F befindet. Deshalb ist $r_i \cdot \mathrm{D}p$ auf K_i stetig mit dem konstanten Wert $k_i = r_i \cdot g(S)$ ($S \geq K$ beliebig). Man wählt $v(x_i) = -k_i r_i$ und erhält aus den vorangehenden Gleichungen

$$0 = -\int_M v \cdot \mathrm{D}p = \sum_{S \subset M, S \in \mathcal{T}^{(n)}} |S| w_2 \sum_{x_i \in \overline{S}} k_i r_i \cdot g_S = \sum_{S \subset M, S \in \mathcal{T}^{(n)}} |S| w_2 \sum_{x_i \in \overline{S}} (r_i \cdot g_S)^2.$$

Also ist in allen Kantenmittelpunkten x_i der Wert $k_i = 0$. Für jedes n-Simplex $S \subset M$ sind die n Kanten, die x als Ecke haben, linear unabhängig. Somit ist der auf S konstante Vektor g_S identisch Null ($r_i \cdot g_S = k_i = 0$). Dies beweist den Satz, denn aus $\mathrm{D}p = 0$ folgt die Konstanz von p. □

Folgerung 2.37 (Stabilität). *Unter den Voraussetzungen des vorigen Satzes erfüllt die Bilinearform b mit $\mathcal{P}_2\mathcal{P}_1$-Elementen unabhängig von h die diskrete LBB-Bedingung* (2.11).

Beweis. Wende Satz 2.33 auf das Resultat von Satz 2.36 an. □

Das Erzeugen geeigneter Triangulierungen

In diesem Abschnitt wird untersucht, wie gravierend der zweite Punkt in Bedingung 2.30 die Wahl und Verfeinerung von Triangulierungen behindert. Es wird sich herausstellen, daß in DROPS ein einziger Vorbereitungsschritt ausreicht, um eine beliebige konsistente Triangulierung $\mathcal{T}$ von Ω in eine „ähnliche“ $\mathcal{T}'$ umzuwandeln, die folgende Eigenschaften hat: $\mathcal{T}'$ ist eine konsistente Triangulierung von Ω. $\mathcal{T}'$ und alle mit DROPS erzeugten Verfeinerungen genügen Bedingung 2.30. Alle Simplexe aus $\mathcal{T}$, die schon eine Ecke im Inneren von Ω haben, sind auch in $\mathcal{T}'$.
Die erste Triangulierung in Abbildung 2.2 besitzt keine inneren Ecken. Wie die mittlere Triangulierung in Abbildung 2.2 demonstriert, kann Punkt zwei aus Bedingung 2.30 von beliebig feinen Triangulierungen verletzt werden: Die kräftig gezeichneten Dreiecke haben keine Ecke im Inneren des triangulierten Quadrates. Tatsächlich gibt es auch kein $x \in \mathcal{N}_\Omega(\mathcal{T})$, so daß eine der hervorgehobenen Diagonalen in ω_x wäre. Teil (3) zeigt schließlich eine nach Bedingung 2.30 erlaubte Triangulierung.

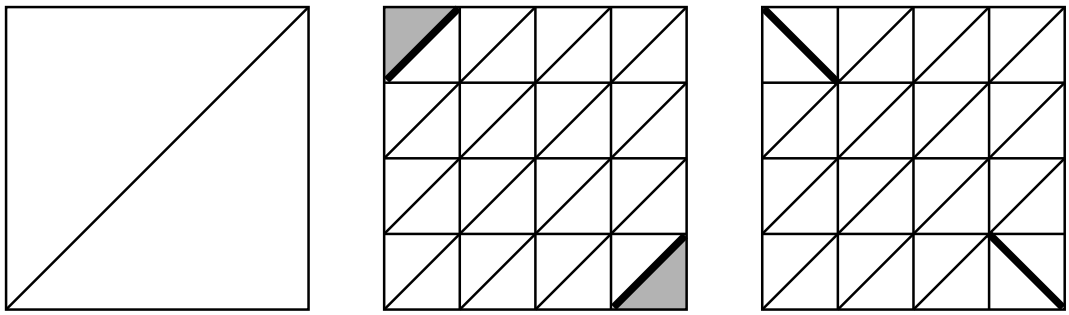

Abbildung 2.2: Triangulierungen: (1) ohne innere Ecken (2) mit Dreiecken, die zwei Kanten in $\partial\Omega$ haben (3) welche die Bedingung 2.30 erfüllt

Durch eine leichte Modifikation der *baryzentrischen Unterteilung* aus der algebraischen Topologie erhält man ein Verfahren, um diesen Mißstand zu beheben. Es ist als Algorithmus 2.38 angegeben. Dessen Eigenschaften werden in Satz 2.39 illustriert.

Algorithmus 2.38 (Vorbereiten einer Triangulierung).
Vorbedingung: $\mathcal{T}$ ist eine konsistente Triangulierung von Ω.
Nachbedingung: siehe Satz 2.39

1: $\mathcal{T}' \leftarrow \mathcal{T}$
2: **for all** $S = [x_0, \ldots, x_n] \in \mathcal{T}$ mit mehr als einer Facette in $\partial\Omega$ **do**
3: $\quad b \leftarrow \frac{1}{n+1} \sum_{i=0}^{n} x_i$
4: $\quad$ Entferne S aus $\mathcal{T}'$!
5: $\quad \mathcal{T}' \leftarrow \mathcal{T}' \cup \{[b]\}$
6: $\quad$ **for all** $R = [x_{i_1}, \ldots, x_{i_m}] \leq S$, $1 \leq m < n$ **do**
7: $\quad\quad \mathcal{T}' \leftarrow \mathcal{T}' \cup \{[b, x_{i_1}, \ldots, x_{i_m}]\}$
8: $\quad$ **end for**
9: **end for**

Satz 2.39 (Korrektheit von Algorithmus 2.38). *Ist $\mathcal{T}$ eine konsistente Triangulierung von Ω und $\mathcal{T}'$ das Ergebnis der Anwendung von Algorithmus 2.38 auf $\mathcal{T}$ so gilt: $\mathcal{T}'$ ist eine konsistente Triangulierung von Ω, die die Bedingung 2.30 erfüllt. Alle $S \in \mathcal{T}$ mit einer Ecke in Ω befinden sich auch in $\mathcal{T}'$. Für jedes n-Simplex in $\mathcal{T}$, der Punkt zwei verletzt, entstehen genau n neue n-Simplexe. Eine Anwendung von Algorithmus 2.38 auf eine Triangulierung, die der Bedingung 2.30 genügt, bewirkt keine Veränderung.*

Beweis. Sei $\mathcal{T}$ eine konsistente Triangulierung von Ω. Falls $\mathcal{T}$ die Bedingung 2.30 erfüllt, gilt dies nach Zeile 1 auch für $\mathcal{T}'$, so daß die Iteration in Zeile 2 nicht ausgeführt wird. Algorithmus 2.38 terminiert dann mit $\mathcal{T}' = \mathcal{T}$.
Sei $S \in \mathcal{T}'$ ein beliebiges n-Simplex. Falls $S \in \mathcal{T}$ gilt, hat S nach Voraussetzung eine Ecke in $\mathcal{N}_\Omega$. Ansonsten ist S in Zeile 7 mit $m = n - 1$ aus einem $T \in \mathcal{T}$

erzeugt worden. Insbesondere hat S die Ecke $b \in T \subseteq \Omega$. Also ist Punkt zwei aus Bedingung 2.30 gegeben. Mit Bemerkung 2.31 ergibt sich die Gültigkeit von Bedingung 2.30.
Aus $\mathcal{T}'$ werden genau die n-Simplexe ohne Ecke in Ω entfernt. Daher befinden sich alle Simplexe aus $\mathcal{T}$, die eine Ecke in Ω besitzen, in $\mathcal{T}$.

Es bleibt zu zeigen, daß $\mathcal{T}'$ konsistent ist. Dazu betrachtet man ein beliebiges $S \in \mathcal{T}$, das von Algorithmus 2.38 durch Zeile 5 und die Iteration in Zeile 6 zerlegt wurde. Es reicht aus nachzuweisen, daß die Menge der Simplexe aus den Zeilen 5 und 7 Punkt zwei von Definition 2.9 erfüllt. Denn diese Simplexe sind Teilmengen von S und haben deshalb wegen der Konsistenz von $\mathcal{T}$ leeren Schnitt mit Simplexen aus $\mathcal{T}$ und den Kindern modifizierter n-Simplexe $S \neq T \in \mathcal{T}$.
Zu diesem technischen Detail wird auf [33, Satz 3.2.2, Beweisteile 1, 2] verwiesen. Damit ist der Satz bewiesen. □

Eine wichtige Frage ist, ob Verfeinerungen von $\mathcal{T}'$ ebenfalls noch Bedingung 2.30 erfüllen. Denn die mehrfache Anwendung der baryzentrischen Unterteilung auf ein Simplex führt zur Entartung, d. h., δ wird „größ“. Der folgende Satz identifiziert eine Klasse von Verfeinerungsalgorithmen, bei der es genügt, Algorithmus 2.38 zu Beginn auf die Ausgangstriangulierung anzuwenden, damit Bedingung 2.30 für alle Verfeinerungen wahr ist. Zu dieser Klasse gehört auch die in DROPS eingesetzte Methode.

Satz 2.40 (Verfeinerungen und Bedingung 2.30). *Falls $\mathcal{T}$ eine konsistente Triangulierung von Ω ist, die Bedingung 2.30 genügt und $\mathcal{T}'$ eine weitere konsistente Triangulierung von Ω darstellt, welche die Bedingung*

> *Für jedes n-Simplex $S \in \mathcal{T}'$ gilt: $S \in \mathcal{T}$ oder es existieren endlich viele Simplexe $S_i \in \mathcal{T}'$ ($S_1 = S$), so daß $T = \cup_i S_i \in \mathcal{T}$ wahr ist.*

erfüllt, so erfüllt $\mathcal{T}'$ ebenfalls Bedingung 2.30.

Im wesentlichen besagt die Bedingung in Satz 2.40, daß der Verfeinerungsalgorithmus, der $\mathcal{T}'$ aus $\mathcal{T}$ generiert hat, die Simplexe in $\mathcal{T}$ einzeln zerlegt. Der der Verfeinerungsalgorithmus in DROPS auf der Zerlegung einzelner Tetraeder in „Kinder“ beruht, reicht ein einziger Vorbereitungsschritt mit Algorithmus 2.38 aus, um Bedingung 2.30 auf allen nachfolgenden Triangulierungen zu garantieren.

Beweis. Sei $\mathcal{T}$ eine konsistente Triangulierung von Ω, die Bedingung 2.30 erfüllt und $\mathcal{T}'$ eine konsistente Triangulierung von Ω, welche die Bedingung in Satz 2.40 erfüllt. Angenommen, es gäbe ein n-Simplex $S \in \mathcal{T}'$, das keine Ecke in Ω besitzt. Dann wäre $S \notin \mathcal{T}$, weil $\mathcal{T}$ Bedingung 2.30 genügt. Setzte man $S_1 = S$, so würden endlich viele $S_i \in \mathcal{T}'$ existieren, so daß $T = \cup_i S_i \in \mathcal{T}$ läge. Somit wäre $\overline{T} \cap \partial\Omega$ höchstens der Abschluß einer Facette $F \leq T$, denn $\mathcal{T}$ erfüllt Bedingung 2.30. Folglich wäre $S \subset \overline{F}$ und daher entartet. Der Widerspruch beweist, daß jedes n-Simplex in $\mathcal{T}'$ wenigstens eine Ecke in Ω hat. Damit ist der Beweis vollständig. □

2.3 Lösungsverfahren für die diskrete Aufgabe

Um auf einem Computer mit Funktionen aus X_h zu rechnen, geht man zweckmäßigerweise zu den Koordinatenspalten bezüglich der Knotenbasis und den Matrizen als Basisdarstellung der Operatoren über. Matrizen und Koordinatenspalten werden im folgenden fett gedruckt, um sie von den Operatoren und Vektoren zu unterscheiden. Ferner wird der Index h, der die Einschränkung der Operatoren auf die endlich-dimensionalen Ansatz- bzw. Testfunktionen-Räume repräsentiert, bei den Matrizen in diesem Abschnitt weggelassen. Dadurch werden die Namen der auftretenden Objekte etwas besser lesbar.
Seien $n_{X,h}$, $n_{V,h}$, $n_{Q,h}$ die Dimensionen von X_h, V_h, Q_h. ($n_{X,h} = n_{V,h} + n_{Q,h}$) Die Lagrange-Basen von Q_h, V_h und X_h werden wie folgt bezeichnet:

$$\mathcal{B}_{Q,h} = (b_{Q,1}, \ldots, b_{Q,n_{Q,h}}), \tag{2.17}$$

$$\mathcal{B}_{V,h} = (b_{V,1}, \ldots, b_{V,n_{V,h}}), \tag{2.18}$$

$$\mathcal{B}_{X,h} = \mathcal{B}_{V,h} \sqcup \mathcal{B}_{Q,h} = (b_{X,1}, \ldots, b_{X,n_{X,h}}) \tag{2.19}$$

Definition 2.41 (Steifigkeitsmatrix, Massenmatrix). Die Darstellung $\mathbf{L} \in \mathbb{R}^{n_{X,h} \times n_{X,h}}$ von L_h bezüglich der Basis $\mathcal{B}_X$, $L_{h_{i,j}} = l(b_{X,j}, b_{X,i}) = \langle L b_{X,j}, b_{X,i} \rangle$ für alle $i, j \in \{1, \ldots, n_{X,h}\}$, heißt Steifigkeitsmatrix. $\mathbf{M} \in \mathbb{R}^{n_{X,h} \times n_{X,h}}$ mit $M_{i,j} = (b_{X,j}, b_{X,i}) = \int_\Omega b_{X,j} b_{X,i}$ für alle $i, j \in \{1, \ldots, n_{X,h}\}$ wird die Massenmatrix genannt. $\mathbf{r} \in \mathbb{R}^{n_{X,h}}, r_i = r(b_{X,i})$, heißt Lastvektor.

Die Begriffe aus Definition 2.41 stammen aus der Ingenieursliteratur. Sie sind allesamt Bezeichnungen für Basisdarstellungen bezüglich der *Knotenbasis* $\mathcal{B}_{X,h}$ der gewählten Finiten Elemente.
Analog dazu besitzen auch A_h, B_h, B'_h Basisdarstellungen $\mathbf{A}$, $\mathbf{B}$, $\mathbf{B}'$. Naturgemäß besteht zwischen einem Operator und seiner Basisdarstellung ein enger Zusammenhang:

Lemma 2.42 (Eigenschaften der Basisdarstellungen). $\mathbf{A}$ *ist eine symmetrische, positiv definite Matrix. Ferner gilt* $\mathbf{B}' = \mathbf{B}^T$.

Beweis. trivial. □

Bemerkung 2.43 *(Massenmatrix).* $\mathbf{M}$ ist die Basisdarstellung des zur Bilinearform $(\cdot, \cdot)_{L_2(\Omega)}$ gehörenden Operators. Es gilt also $(x, y) = \mathbf{x}^T \mathbf{M} \mathbf{y}$ für alle $x, y \in X_h$. Die Massenmatrix stimmt nur dann mit der Einheitsmatrix überein, wenn $\mathcal{B}_{X,h}$ eine L_2-Orthonormalbasis von X_h ist.
Daß die Massenmatrix überhaupt erwähnt wird, liegt daran, daß sie als Vorkonditionierer für Aufgaben, in denen die Form a auftritt, interessant ist.

Durch die Wahl von Finite-Elemente-Räumen sind alle in Lemma 2.42 auftretenden Matrizen *dünnbesetzt*, denn die Funktionen in X_H besitzen alle nur einen

„kleinen“ Täger. Um das Speichern der vielen Nulleinträge zu vermeiden, wird in DROPS das *„compressed row storage“*-Format eingesetzt. Es besteht aus drei Tupeln:

1. *val* enthält alle von Null verschiedenen Einträge aus $\mathbf{L}$ in der Reihenfolge, wie sie beim zeilenweisen Lesen von oben nach unten erscheinen.

2. c ist ein Tupel von Indizes mit $\dim c = \dim val$. Für alle $1 \leq j \leq \dim val$ gilt: val_j steht in $\mathbf{L}$ in Spalte j.

3. r ist ein Tupel von Indizes. Es hat $(n_{X,h} + 1)$ Komponenten. Es gilt für alle Indizes $1 \leq i \leq n_{X,h}, r_i \leq k \leq r_{i+1}$: val_k steht in Zeile i.

Trotz dieser Speicherung können Matrix-Vektor-Produkte mit $\mathbf{L}$ und $\mathbf{L}^T$ effizient berechnet werden.
Tatsächlich werden in DROPS nur $\mathbf{A}$ und $\mathbf{B}$ gespeichert. $\mathbf{L}$ wird durch ein Objekt repräsentiert, das die Multiplikation mit $\mathbf{L}$ durch Multiplikation mit $\mathbf{A}$, $\mathbf{B}$ und $\mathbf{B}^T$ realisiert. Dies ist für den Anwender transparent.

Die diskretisierten Stokes-Gleichungen stellen also ein lineares Gleichungssystem hoher Dimension mit der Matrix $\mathbf{L}$ dar. Wie auch bei der kontinuierlichen Aufgabe (1.24) und (1.25) liegt $\mathbf{L}$ in Blockgestalt vor:

$$\mathbf{Lx} = \mathbf{r} \iff \begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} \tag{2.20}$$

Da $\mathbf{A}$ symmetrisch und positiv definit ist, bietet sich als erstes Lösungsverfahren die *Methode des Schur-Komplementes* an, welche durch Anwendung des Gauß-Algorithmus auf die Blöcke in (2.20) hergeleitet werden kann. Multipliziert man die erste Zeile von (2.20) von links mit $\mathbf{BA}^{-1}$, so erhält man mit der Bezeichnung *Schur-Komplement* für $\mathbf{S} = \mathbf{BA}^{-1}\mathbf{B}^T$ die Gleichung $\mathbf{Sp} = \mathbf{BA}^{-1}\mathbf{f} - \mathbf{g}$ aus (2.20). Sie kann nach $\mathbf{p}$ aufgelöst werden, denn wie man sich leicht überlegt, ist $\mathbf{S}$ symmetrisch positiv definit.

Algorithmus 2.44 (Schur-Komplement-Methode).

1: Man löse $\mathbf{Aw} = \mathbf{f}$.
2: Man löse $\mathbf{Sp} = \mathbf{Bw} - \mathbf{g}$.
3: Man löse $\mathbf{Av} = \mathbf{f} - \mathbf{B}^T\mathbf{p}$.

Auf die Systeme in Zeile 1 und 3 kann ein iteratives Verfahren für symmetrisch positiv definite Probleme angewendet werden, z. B. PCG. Als Vorkonditionierung kommen die üblichen Verfahren für die Poisson-Gleichung in Frage, da $\mathbf{A}$ im wesentlichen ein Tupel von Laplace-Operatoren repräsentiert.
Die Gleichung in Zeile 2 kann ebenfalls mit dem CG-Algorithmus gelöst werden. Einen Vorkonditionierer zu konstruieren, erweist sich als schwierig, weil man keine

explizite Darstellung von $\mathbf{S}$ hat. Denn ein Faktor von $\mathbf{S}$ lautet $\mathbf{A}^{-1}$. Sogar in iterativen Verfahren ist die Multiplikation mit $\mathbf{S}$ also eine sehr teure Operation: Es muß nämlich ein Gleichungssystem mit $\mathbf{A}$ gelöst werden. Dies muß mit sehr hoher Genauigkeit im Vergleich zum Lösen von $\mathbf{Sp} = \mathbf{Bw} - \mathbf{g}$ geschehen, da die Multiplikation mit $\mathbf{S}$ beim iterativen Lösen von Zeile 2 recht oft wiederholt wird. Der Fehler beim Lösen des inneren Systems akkumuliert sich dabei.

Ein anderes populäres Verfahren zur Lösung von (2.20) vermeidet einige dieser Probleme.

Algorithmus 2.45 (Inexakte Uzawa-Methode).

Vorbedingung: Startvektoren $\mathbf{v}_0$, $\mathbf{p}_0$, Vorkonditionierer $\tilde{\mathbf{A}}$, $\tilde{\mathbf{S}}$ für $\mathbf{A}$, $\mathbf{S}$.

1: $i \leftarrow 0$
2: **repeat**
3: $\quad \mathbf{v}_{i+1} \leftarrow \mathbf{v}_i + \tilde{\mathbf{A}}^{-1}(\mathbf{f} - \mathbf{A}\mathbf{v}_i - \mathbf{B}^T\mathbf{p}_i)$
4: $\quad \mathbf{p}_{i+1} \leftarrow \mathbf{p}_i + \tilde{\mathbf{S}}^{-1}(\mathbf{B}\mathbf{v}_{i+1} - \mathbf{g})$
5: $\quad i \leftarrow i + 1$
6: **until** Die Genauigkeit von $(\mathbf{v}_i, \mathbf{p}_i)^T$ ist hoch genug.

In der vorliegenden Arbeit ist der Korrekturterm in Zeile 3 das Ergebnis mehrerer Anwendungen eines Poisson-Lösers (PCG, Mehrgitter) aus DROPS auf $\mathbf{Aw} = \mathbf{f} - \mathbf{A}\mathbf{v}_i - \mathbf{B}^T\mathbf{p}_i$ mit Startvektor $\mathbf{0}$. Für den Korrekturterm in Zeile 4 wird $\tilde{\mathbf{S}} = \mathbf{M}$, also die Massenmatrix verwendet.
In [25] wird folgendes Resultat bezüglich der Konvergenz von Algorithmus 2.45 bewiesen:

Satz 2.46 (Konvergenz der inexakten Uzawa-Methode). *Seien $\tilde{\mathbf{A}}$, $\tilde{\mathbf{S}}$ symmetrisch positiv definite Vorkonditionierer von $\mathbf{A}$ und $\mathbf{S}$, mit der Eigenschaft, daß $\tilde{\mathbf{A}} - \mathbf{A}$ sowie $\tilde{\mathbf{S}} - \mathbf{S}$ positiv semidefinit sind. Ferner seien $\sigma_A, \sigma_S \in [0, 1)$ mit*

$$(1 - \sigma_A)\left(\tilde{\mathbf{A}}\mathbf{v}, \mathbf{v}\right) \leq (\mathbf{A}\mathbf{v}, \mathbf{v}) \quad \textit{für alle } \mathbf{v},$$
$$(1 - \sigma_S)\left(\tilde{\mathbf{S}}\mathbf{p}, \mathbf{p}\right) \leq (\mathbf{S}\mathbf{p}, \mathbf{p}) \quad \textit{für alle } \mathbf{p}$$

gegeben. Solche Zahlen existieren, weil $\tilde{\mathbf{A}}$ und $\tilde{\mathbf{S}}$ positiv definit sind. Dann gilt mit der problemabhängigen Norm $[|\cdot|]$ die Ungleichung

$$[|(\mathbf{v} - \mathbf{v}_i, \mathbf{p} - \mathbf{p}_i)^T|] \leq \rho^i [|(\mathbf{v} - \mathbf{v}_0, \mathbf{p} - \mathbf{p}_0)^T|].$$

Der Parameter ρ lautet

$$\rho = \frac{1}{2}\left(\sigma_S(1 - \sigma_A) + \sqrt{\sigma_S^2(1 - \sigma_A)^2 + 4\sigma_A}\right) \leq 1 - \frac{1}{2}(1 - \sigma_S)(1 - \sigma_A).$$

Aus diesem Satz wird geschlossen, daß die Konvergenzrate des inexakten Uzawa-Verfahrens hoch ist, wenn gute Vorkonditionierer für $\mathbf{A}$ und $\mathbf{S}$ zur Verfügung stehen.

Bemerkung 2.47 *(CG-Verfahren für Sattelpunktprobleme).* In [44] wird die Konvergenz iterativer Verfahren für Sattelpunktprobleme untersucht. Dort finden sich zwei besonders interessante Ergebnisse. Neben $\mathbf{L}$ wird die Blockmatrix

$$\tilde{\mathbf{L}} = \begin{pmatrix} \tilde{\mathbf{A}} & \mathbf{0} \\ \mathbf{B} & \tilde{\mathbf{S}} \end{pmatrix}$$

betrachtet. Mit den Bezeichnungen $\mathbf{x} = (\mathbf{v}, \mathbf{p})^T$, $\mathbf{x}_i = (\mathbf{v}_i, \mathbf{p}_i)^T$, $\mathbf{r} = (\mathbf{f}, \mathbf{g})^T$ kann man die Zeilen 3 und 4 in Algorithmus 2.45 als

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \tilde{\mathbf{L}}^{-1}(\mathbf{L}\mathbf{x}_i - \mathbf{r}) \tag{2.21}$$

darstellen. Daher kann das inexakte Uzawa-Verfahren als *Richardson-Verfahren* zum Lösen von $\tilde{\mathbf{L}}^{-1}\mathbf{L}\mathbf{x} = \tilde{\mathbf{L}}^{-1}\mathbf{r}$ aufgefaßt werden.
Nimmt man an, daß die Vorkonditionierer $\tilde{\mathbf{A}}$ und $\tilde{\mathbf{S}}$ symmetrisch positiv definit sind und daß $\mathbf{A} - \tilde{\mathbf{A}}$ positiv definit ist, so wird in [44] folgendes bewiesen: $\tilde{\mathbf{L}}^{-1}\mathbf{L}$ ist bezüglich des Skalarproduktes

$$\left(\begin{pmatrix} \mathbf{u} \\ \mathbf{p} \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ \mathbf{q} \end{pmatrix} \right)_* = \Big((\mathbf{A} - \tilde{\mathbf{A}}\mathbf{u}, \mathbf{v} \Big) + \Big(\tilde{\mathbf{S}}\mathbf{p}, \mathbf{q} \Big)$$

symmetrisch positiv definit. Deshalb kann auf die vorkonditionierte Sattelpunktaufgabe $\tilde{\mathbf{L}}^{-1}\mathbf{L}\mathbf{x} = \tilde{\mathbf{L}}^{-1}\mathbf{r}$ das CG-Verfahren angewendet werden. Das ergibt theoretisch eine Methode mit wesentlich höherer Konvergenzrate als die Richardson-Iteration (2.21).

Kapitel 3

A posteriori Fehlerschätzung I – Residuumsverfahren

Für zuverlässige und effiziente numerische Simulationen sind Fehlerschätzungen von großer Bedeutung. In Abschnitt 2.2.2 wurden bereits *a priori* Abschätzungen des Diskretisierungsfehlers vorgestellt. Obwohl sie die Konvergenz der verwendeten Galerkinmethode demonstrieren, sind die Schranken für den Fehler oft pessimistisch. Außerdem können die a priori Abschätzungen aus 2.2.2 nicht eingesetzt werden, um die Triangulierung adaptiv zu verfeinern, weil sie nur Abschätzungen des Fehlers über dem gesamten Gebiet, z. B. in der $H^1(\Omega) \times L_2(\Omega)$-Norm, liefern. Daraus geht nicht hervor, ob der Diskretisierungsfehler in Teilgebieten von Ω besonders groß ist.
Abschätzungen des Diskretisierungsfehlers, in denen eine bereits vorliegende Lösung der diskreten Aufgabe verwendet wird, heißen *a posteriori Fehlerschätzungen.*

Bevor die ersten solchen Schätzer vorgestellt werden, werden einige Generalvoraussetzungen für dieses Kapitel getroffen: $\Omega \subseteq \mathbb{R}^n$ sei ein beschränktes Gebiet mit Lipschitzrand, das durch die reguläre Familie $\{\mathcal{T}_h\}_{h>0}$ von konsistenten Triangulierungen zerlegt wird. $\mathcal{T}$ bezeichnet eine beliebige Triangulierung aus dieser Familie. $x_0 = (u_0, p_0)^T \in X$ sei die Lösung der Stokes-RWA 1.30 und $x_{0,h} = (u_{0,h}, p_{0,h})^T \in X_h$ sei die Lösung der diskreten Stokes-RWA (2.1). $x_h = (u_h, p_h)^T$ sei ein beliebiges Element aus X_h. $X_h = V_h \times Q_h \leq X = V \times Q$ sei der Finite-Elemente-Raum aus $\mathcal{P}_{n_Q+1}\mathcal{P}_{n_Q}$-Elementen auf $\mathcal{T}$, also konkret

$$V_h = \left(S^{n_Q+1}(\mathcal{T})\right)^n, \quad Q_h = \begin{cases} S^{k_Q}(\mathcal{T}), & \text{falls} \quad |\Gamma_A| > 0, \\ S^{k_Q}(\mathcal{T}) \cap L_2^0(\Omega) & \text{sonst.} \end{cases}$$

Im folgenden werden lokale Fehlerschätzungen η_S ($S \in \mathcal{T}^{(n)}$) gesucht, die einen guten Indikator für die Größe des Diskretisierungsfehlers auf S liefern, also $\eta_S \approx \|x_{0,h} - x_0\|_{*;\omega_S}$. $\|\cdot\|_*$ ist eine noch näher zu bestimmende Norm. Dabei ist auf folgende Charakteristika zu achten:

1. *Lokalität* bedeutet, daß der Diskretisierungsfehler auf einzelnen n-Simplexen dargestellt werden kann. Diese Eigenschaft ist wichtig, um Triangulierungen adaptiv verfeinern zu können (siehe Kapitel 5).

2. *Zuverlässigkeit* steht für die Existenz einer Abschätzung des Diskretisierungsfehlers von der Form

$$\|x_{0,h} - x_0\|_{*,\Omega} \leq C \sqrt{\sum_{S \in \mathcal{T}^{(n)}} \eta_S^2}$$

   mit einer moderaten Konstante $C > 0$, die nicht von h abhängt. Diese Eigenschaft ermöglicht die Kontrolle des globalen Fehlers z. B. im Abbruchkriterium des adaptiven Zyklus aus Abbildung 1.

3. *Effizienz* heißt, daß der Diskretisierungsfehler durch η_S nach unten beschränkt wird:

$$\eta_S \leq C \|x_{0,h} - x_0\|_{*,\omega_S}.$$

   Die Konstante $C > 0$ hängt wieder nicht von h ab. Eine solche Ungleichung garantiert, daß der Diskretisierungsfehler $\|x_{0,h} - x_0\|_{*,\omega_S}$ groß ist, wenn η_S groß ist. Deshalb wird eine Verfeinerungsstrategie, die zuerst Gebiete verfeinert, auf denen η_S groß ist, vorrangig Gebiete mit großem Diskretisierungsfehler verfeinern.

4. *Moderater Rechenaufwand.* Damit der adaptive Zyklus ausgewogen ist, sollten alle seine Hauptschritte etwa gleich viel Rechenzeit erfordern.[1] Somit sollte die Berechnung aller η_S nicht länger dauern als etwa die Diskretisierung oder das Lösen der Gleichungssysteme.

Der weitere Aufbau des vorliegenden Kapitels lautet so: In Abschnitt 3.1 wird die den Fehlerschätzern dieses Kapitels zugrundeliegende Idee erläutert. Dazu wird in diesem Abschnitt Kapitel 2 aus [39] für lineare Differentialoperatoren spezialisiert. Wegen des anschaulichen Kerns geht dieser Abschnitt den konkreten a posteriori Fehlerschätzern voran, deren Herleitung in Abschnitt 3.2 etwas technischer ist. Dort werden Fehlerschätzer für die $H^1 \times L_2$-Norm und die $L_2 \times H^{-1}$-Norm hergeleitet. Im Abschnitt 3.3 wird eine Technik untersucht, bei der auf einem einzelnen n-Simplex eine lokale Stokes-Aufgabe gelöst wird, deren Norm als Fehlerschätzer verwendet werden kann.

[1] Sonst wäre es geschickter, die aufwendigen Arbeitsschritte möglichst nicht in jedem Zyklus auszuführen.

3.1 Allgemeine Theorie

Auf abstrakte Weise lautet die Stokes-RWA: Finde ein $x \in X$ mit $Lx = r$. Da es jetzt eine zentrale Rolle spielen wird, erhält das Residuum dieser Operatorgleichung eine eigene Bezeichnung:

$$R : X \longrightarrow X' : x \longmapsto Lx - r = L(x - x_0). \tag{3.1}$$

Analog wird das Residuum der diskreten Aufgabe benannt:

$$R_h : X_h \longrightarrow X_h' : x_h \longmapsto L_h x_h - r|_{X_h} = L(x_h - x_{0,h})|_{X_h}.$$

Da die diskrete Aufgabe durch konforme Diskretisierung entsteht, ist R_h nur eine Einschränkung von R.

Satz 3.1 (Residuum zur Fehlerschätzung). *Für alle $x \in X$ gilt*

$$\|L\|^{-1}_{\mathcal{L}[X,X']}\|R(x)\|_{X'} \leq \|x - x_0\|_X \leq \|L^{-1}\|_{\mathcal{L}[X',X]}\|R(x)\|_{X'}. \tag{3.2}$$

Beweis. Sei $x \in X$ beliebig. $R(x) = L(x - x_0)$, also $\|R(x)\|_{X'} = \|L(x - x_0)\|_{X'} \leq \|L\|\|x - x_0\|_X$. Die Stetigkeit von L beweist also die linke Ungleichung.
L hat eine stetige Inverse, so daß $x - x_0 = L^{-1}R(x)$ gilt. Somit erhält man die rechte Ungleichung aus der Stetigkeit von L^{-1} (Stabilität). □

Dieser einfache Satz, der als Aussage über das Residuum von (endlich dimensionalen) linearen Gleichungssystemen wohlbekannt ist, liefert die Grundlage für die Fehlerschätzer in diesem Kapitel. Er besagt, daß das die Dualnorm des Residuums global ein zuverlässiger, effizienter Fehlerschätzer ist.

Bemerkung 3.2. Verfürth verallgemeinert Satz 3.1 in [39, Kapitel 2] auf allgemeine Differentialoperatoren, die als Banachraum-wertige Funktionen stetig differenzierbar sind. Zur Herleitung des allgemeinen Resultates wird der Operator durch seine Ableitung approximiert, so daß das zu (3.2) analoge Resultat im allgemeinen nur noch in einer Umgebung der exakten Lösung gilt.
Im Fall eines linearen Differentialoperators geht sein Satz bis auf ein Detail in Satz 3.1 über: Aufgrund der Beweistechnik treten bei Verfürth auf der linken und rechten Seite von (3.2) zusätzlich die Faktoren $\frac{1}{2}$ bzw. 2 auf.

3.1.1 Kondition von L

In einführenden Vorlesungen über die Numerik linearer Gleichungssysteme lernt man, daß ein kleines Residuum einer Näherungslösung y von $Mx = b$ im allgemeinen nicht bedeutet, daß $\|y - x\|$ klein ist. Probleme treten bei Matrizen M mit großer Kondition auf. Deshalb werden hier die „Konstanten“ $\|L\|^{-1}$ und $\|L^{-1}\|$ aus (3.2) untersucht. Dabei ergibt sich eine Abschätzung von $\operatorname{cond} L = \|L\|\|L^{-1}\|$.

Lemma 1.28 liefert $\|A\| \leq \nu$, $\|B\| \leq 1$, und der Beweis von Satz 1.31 ergibt $\|A^{-1}\| \leq \nu^{-1}(1 + C_{\mathrm{P}})$. Darin ist $C_{\mathrm{P}} > 0$ die Konstante aus der Poincaré-Ungleichung.[2] Für $\|B_{\perp}^{-1}\|$ gilt wegen Lemma 1.24, Satz 1.33 und Satz 1.34 $\|B_{\perp}^{-1}\| \leq \beta^{-1}$ mit der Konstante β aus Satz 1.34. Damit kann Ungleichung (1.32) zu

$$\|L^{-1}\|_{\mathcal{L}[X',X]} \leq \sqrt{3}\max\{(1 + C_{\mathrm{P}})(\nu^{-1} + \beta^{-1}), \beta^{-1}(2 + C_{\mathrm{P}})(1 + \nu\beta^{-1})\}$$

vereinfacht werden. Wie beim Beweis dieser Ungleichung kann man auch $\|L\|$ durch die Normen von A und B abschätzen. Sei dazu $y \in X$, $\|y\|_X = 1$, beliebig. Dann folgt

$$\begin{aligned}\|Ly\|_{X'} &= \sqrt{\|Av\|^2 + \|B'q\|^2 + \|Bv\|^2} = \|(\|Av\|, \|B'q\|, \|Bv\|)^T\|_2 \\ &\leq \sqrt{3}\max\{\|Av\|, \|B'q\|, \|Bv\|\} \leq \sqrt{3}\max\{\nu, 1\},\end{aligned}$$

also $\|L\| \leq \sqrt{3}\max\{\nu, 1\}$. Über die Relation $1 \leq \|L\|\|L^{-1}\|$ erhält man die Schranken

$$\begin{aligned}&\frac{1}{\sqrt{3}}\min\{\nu^{-1}, 1\} \leq \|L^{-1}\| \leq \sqrt{3}\max\left\{(1+C_{\mathrm{P}})(\nu^{-1}+\beta^{-1}), \frac{(2 + C_{\mathrm{P}})(1 + \nu\beta^{-1})}{\beta}\right\}, \\ &\frac{1}{\sqrt{3}}\min\left\{\frac{1}{(1 + C_{\mathrm{P}})(\nu^{-1} + \beta^{-1})}, \frac{\beta}{(2 + C_{\mathrm{P}})(1 + \nu\beta^{-1})}\right\} \leq \|L\| \leq \sqrt{3}\max\{\nu, 1\}.\end{aligned}$$

Durch Multiplikation der rechten Terme erhält man zusätzlich eine Abschätzung der Kondition des Differentialoperators L:

$$\mathrm{cond}(L) \leq 3\max\{\nu, 1\}\max\left\{(1 + C_{\mathrm{P}})(\nu^{-1} + \beta^{-1}), \frac{(2 + C_{\mathrm{P}})(1 + \nu\beta^{-1})}{\beta}\right\}$$

Die Kondition hängt in einfacher Weise von der kinematischen Viskosität ν des Fluids ab. Über C_{P} finden die lineare Ausdehnung von Ω und die Art der Randbedingungen Eingang in die Abschätzung. β beschreibt die geometrische Kompliziertheit von Ω.

Bis jetzt wurden drei Grundprinzipien dieses Kapitels vorgestellt:

- Die Stetigkeit von L liefert untere Fehlerschranken, also Effizienz.
- Die Stabilität von L liefert obere Fehlerschranken, also Zuverlässigkeit.
- Das Residuum $R(x)$ einer Näherungslösung ist ein stetiges, lineares Funktional auf X als Testfunktionenraum. Seine X'-Norm kann zur a posteriori Fehlerschätzung verwendet werden.

[2] Siehe dazu Bemerkung 1.32.

Wegen

$$\|R(x)\|_{X'} = \sup_{y\in X,\, \|y\|_X=1} \langle R(x), y\rangle$$

ist die Ermittlung der X'-Norm eine unendlich-dimensionale Maximierungsaufgabe. Deshalb wird jetzt untersucht, wie mit moderatem Aufwand eine Lösung bzw. vernünftige Schranken einer Lösung berechnet werden können.

3.1.2 Abschätzung des Residuums durch Projektion

Die exakte Berechnung von $\|R(x)\|_{X'}$ ist im allgemeinen nicht möglich, weil X „zu groß“ ist. Stattdessen wählt man einen (endlich-dimensionalen) Finite-Elemente-Raum $\tilde{X}_h \leq X$ als Raum der Testfunktionen. Anstelle von $\|R(x)\|_{X'}$ wird nur $\|R(x)|_{\tilde{X}_h}\|_{\tilde{X}_h'} = \sup_{y\in\tilde{X}_h, \|y\|_X=1} \langle R(x), y\rangle$ ausgewertet. Diese Einschränkung, d. h. Projektion, von $R(x) \in X'$ auf $R(x)|_{\tilde{X}_h} \in \tilde{X}_h'$ ist das Grundprinzip des vorliegenden Abschnittes. Tatsächlich erhält man so eine endlich-dimensionale Maximierungsaufgabe, deren Lösung man in den Griff bekommen kann (siehe Abschnitt 3.2).
Die Wahl von $\tilde{X}_h$ ist ein wichtiges, wenn auch technisches Teilproblem: Nach der Berechnung von $x_{0,h}$ hat man das Residuum $R(x_{0,h})$ zur Verfügung, welches als Funktional auf X_h verschwindet. Damit $\|R(x_{0,h})\|_{X'}$ gut angenähert wird, muß $\tilde{X}_h$ unabhängig vom Diskretisierungsparameter h hinreichend verschieden von X_h sein, was durch Bedingung (3.3) in Satz 3.3 präzisiert wird. Sie besagt, daß die Projektion $R(x) \longrightarrow R(x)|_{\tilde{X}_h}$ auf dem Komplement von X_h in X stabil ist.
Da $\tilde{X}_h$ endlich-dimensional ist und $\dim X/X_h = \infty$ gilt, kann man nicht erwarten, daß sich etwa $\|R(x_{0,h})\|_{X'} \leq C\|R(x_{0,h})\|_{\tilde{X}_h'}$ beweisen läßt.[3] Es muß also mit zusätzlichen Fehlertermen gerechnet werden. Allerdings läßt sich erreichen, daß sie höhere Potenzen von h enthalten, so daß sie auf feineren Triangulierungen keine Rolle mehr spielen.
Um die Möglichkeit zu haben, etwa die Daten $\bar{f}$, $\bar{g}$, die in $R(x)$ auftreten, zu approximieren, wird in Satz 3.3 nicht $\|R(x)\|_{\tilde{X}_h'}$ sondern die Norm $\|\tilde{R}_h(x)\|_{\tilde{X}_h'}$ einer Approximation $\tilde{R}_h$ von R zur Abschätzung von $\|R(x)\|_{X'}$ verwendet. Die Konstruktion von $\tilde{R}_h$ ist im Vergleich zu der von $\tilde{X}_h$ einfach; es wird auf Abschnitt 3.1.3 verwiesen.

Satz 3.3 (Residuumsschätzung durch Projektion). *Sei $I_h \in \mathcal{L}[X, X_h]$ ein Projektor (für Testfunktionen) und $\tilde{X}_h \leq X$. $\tilde{R}_h \in \mathcal{L}[X_h, X']$ sei eine Approximation von R (bzw. R_h) auf X_h. Diese drei Objekte mögen für beliebige $x_h \in X_h$ der Stabilitätsungleichung*

$$\|(id_X - I_h)'\tilde{R}_h(x_h)\|_{X'} \leq C\|\tilde{R}_h(x_h)\|_{\tilde{X}_h'} \tag{3.3}$$

[3] Zum Beispiel könnte $R(x_{0,h}) \neq 0$, aber $\langle R(x_{0,h}), y\rangle = 0$ für alle $y \in X_h \oplus \tilde{X}_h$, auftreten.

mit einer von h unabhängigen Konstante $C > 0$ genügen. Dann wird das Residuum (3.1) *von beliebigen $x_h \in X_h$ durch*

$$\begin{aligned}\|R(x_h)\|_{X'} \leq C\|\tilde{R}_h(x_h)\|_{\tilde{X}_h'} + \|(id_X - I_h)'\big(R(x_h) - \tilde{R}_h(x_h)\big)\|_{X'} \\ + \|I_h\|_{\mathcal{L}[X,X_h]}\|R(x_h)\|_{X_h'},\end{aligned} \tag{3.4a}$$

$$\|\tilde{R}_h(x_h)\|_{\tilde{X}_h'} \leq \|R(x_h)\|_{\tilde{X}_h'} + \|R(x_h) - \tilde{R}_h(x_h)\|_{\tilde{X}_h'} \tag{3.4b}$$

nach oben und unten[4] *abgeschätzt.*

Beweis. Sei $x_h \in X_h$ beliebig. Ungleichung (3.4b) ergibt sich sofort durch die Anwendung der Dreiecksungleichung auf $\|\tilde{R}_h(x_h)\|_{\tilde{X}_h'} = \|R(x_h) + (\tilde{R}_h(x_h) - R(x_h))\|_{\tilde{X}_h'}$.
Zum Nachweis von (3.4a) sei ferner $x \in X$, $\|x\|_X = 1$, beliebig. Durch Addition zweier „nahrhafter Nullen“, Umstellen und Verwenden der Dualität erhält man

$$\begin{aligned}\langle R(x_h), x\rangle_{X'\times X} &= \langle R(x_h), x\rangle + \Big\langle \tilde{R}_h(x_h), x - I_h x\Big\rangle - \Big\langle \tilde{R}_h(x_h), x - I_h x\Big\rangle \\ &\quad + \langle R(x_h), I_h x\rangle - \langle R(x_h), I_h x\rangle \\ &= \Big\langle \tilde{R}_h(x_h), x - I_h x\Big\rangle + \Big\langle R(x_h) - \tilde{R}_h(x_h), x - I_h x\Big\rangle + \langle R(x_h), I_h x\rangle \\ &= \Big\langle (id_X - I_h)'\tilde{R}_h(x_h), x\Big\rangle + \Big\langle (id_X - I_h)'\big(R(x_h) - \tilde{R}_h(x_h)\big), x\Big\rangle \\ &\quad + \langle R(x_h), I_h x\rangle \\ &\leq \|(id_X - I_h)'\tilde{R}_h(x_h)\|_{X'} + \|(id_X - I_h)'\big(R(x_h) - \tilde{R}_h(x_h)\big)\|_{X'} \\ &\quad + \|R(x_h)\|_{X_h'}\|I_h\|_{\mathcal{L}[X,X_h]}.\end{aligned}$$

Wendet man auf den ersten Term der letzten Zeile (3.3) an, so erhält man durch die Bildung des Supremums über alle $x \in X$, $\|x\|_X = 1$, Ungleichung (3.4a). □

Bemerkung 3.4. In Satz 3.3 ergibt (3.4b) wegen

$$\|R(x_h)\|_{\tilde{X}_h'} = \sup_{y\in\tilde{X}_h, \|y\|_X=1} \langle R(x_h), y\rangle \leq \sup_{y\in X, \|y\|_X=1} \langle R(x_h), y\rangle = \|R(x_h)\|_{X'} \tag{3.5}$$

tatsächlich bis auf den Fehlerterm eine untere Schranke für $\|R(x_h)\|_{X'}$.
Außerdem gilt für die Lösung $x_{0,h}$ der diskreten Aufgabe $\|R(x_{0,h})\|_{X_h'} = 0$. Dies ist eine wichtige Konsequenz ihrer *Galerkin-Orthogonalität* zum Raum der Testfunktionen.

Der Projektor I_h wird im Beweis von Satz 3.3 auf beliebige Funktionen aus X angewendet, was einen Operator wie den aus Abschnitt 2.2.1 erforderlich macht. Somit kann ein weiteres Grundprinzip dieses Kapitels formuliert werden: *Man benötigt einen Projektionsoperator $X \longrightarrow X_h$ für „rauhe“ Funktionen.*

Bei der Wahl von I_h sind anhand der Randbedingungen zwei Fälle zu unterscheiden.

[4] siehe Bemerkung 3.4

- $|\Gamma_A| > 0$: Man definiert I_h als $I_h : X \longrightarrow X_h : (u,p)^T \longmapsto (I^{\mathrm{Z},n_V}u, I^{\mathrm{Z},n_Q}p)^T$.

- $\partial\Omega = \Gamma_D$: Die Geschwindigkeitskomponenten werden wie zuvor mit I^{Z,n_V} interpoliert. Q und Q_h enthalten jedoch nur Funktionen mit verschwindendem Integralmittel. Diese Eigenschaft wird von I^{Z,n_Q} im allgemeinen nicht respektiert. Abhilfe schafft ein „Korrekturschritt". Sei $\pi_0 : L_2(\Omega) \longrightarrow \mathbb{R}$ die $L_2(\Omega)$-Projektion auf die Funktion, die konstant 1 ist. Dann werden die Druckkomponenten $p \in Q$ durch $(id - \pi_0)I^{\mathrm{Z},n_Q}p$ approximiert. Es gilt

$$\pi_0\big((id - \pi_0)I^{\mathrm{Z},n_Q}p\big) = 0, \quad \text{also} \quad \int_\Omega (id - \pi_0)I^{\mathrm{Z},n_Q}p = 0$$

  für alle $p \in Q$. Setzt man die Interpolation für V und Q zu I_h zusammen, so folgt $I_h : X \longrightarrow X_h$.

Bemerkung 3.5 *(Interpolationsfehler).* Falls $|\Gamma_\mathrm{A}| > 0$ erfüllt ist, können die Resultate aus Abschnitt 2.2.1 direkt zur Abschätzung des Interpolationsfehlers $id - I_h$ herangezogen werden. Bei reinen Dirichlet-Randbedingungen gilt dies zunächst nur für die Anteile in V bzw. V_h. Für den Druckanteil $p \in Q$ findet man die Identität

$$p - (id - \pi_0)I^{\mathrm{Z},n_Q}p = p - \pi_0 p - (id - \pi_0)I^{\mathrm{Z},n_Q}p = (id - \pi_0)(id - I^{\mathrm{Z},n_Q})p,$$

weil $\pi_0 p = 0$ für alle $p \in Q$ gilt. Also können auch im Falle reiner Dirichlet-Randbedingungen die Approximationssätze aus Abschnitt 2.2.1 angewendet werden.
In [39] wird zur Druckinterpolation der Nulloperator verwendet. Das ist bei inhomogenen Randwerten nicht sinnvoll, weil in Lemma 3.11 bestimmte Fehlerterme nicht klein werden.

3.1.3 Konstruktion von $\tilde{X}_h$ und $\tilde{R}_h$

Hilfsmittel

Mit $\hat{S} = [0, e_1, \ldots, e_n]$, $\hat{F} = [0, e_1, \ldots, e_{n-1}]$ werden das Referenz-n-Simplex und die Referenz-Facette bezeichnet. Sind $S, F \in \mathcal{T}$ ein beliebiges n-Simplex und eine seiner Facetten, so sei F_S eine affine Abbildung mit $F_S(\hat{S}) = S$ und $F_S(\hat{F}) = F$. Die Einschränkung von F_S auf $\hat{F}$ heißt F_F.
Nun seien $V_{\hat{S}} \leq L_\infty(\hat{S})$, $V_{\hat{F}} \leq L_\infty(\hat{F})$ endlich-dimensionale Räume von Testfunktionen.

Definition 3.6 (Fortsetzungsoperator, Abschneidefunktion).
$\hat{P} : L_\infty(\hat{F}) \longrightarrow L_\infty(\hat{S}) : u(x_1, \ldots, x_n) \longmapsto (\hat{P}u : (x_1, \ldots, x_n) \mapsto u(x_1, \ldots, x_{n-1}))$ heißt die in e_n-Richtung konstante Fortsetzung von $\hat{F}$ auf $\hat{S}$.

Eine Funktion $\psi_{\hat{S}} \in C^\infty(\hat{S})$ heißt unter folgenden Bedingungen Abschneidefunktion auf $\hat{S}$:

$$0 \le \psi_{\hat{S}} \le 1, \quad \sup_{\hat{S}} \psi_{\hat{S}} = 1, \quad \psi_{\hat{S}}|_{\partial\hat{S}} \equiv 0, \quad \operatorname{supp} \psi_{\hat{S}} \subseteq \bar{S}.$$

Eine Funktion $\psi_{\hat{F}} \in C^\infty(\hat{S})$ heißt unter folgenden Bedingungen Abschneidefunktion auf $\hat{F}$:

$$0 \le \psi_{\hat{F}} \le 1, \quad \sup_{F} \psi_{\hat{F}} = 1, \quad \psi_{\hat{F}}|_{\partial\hat{S}\setminus\hat{F}} \equiv 0, \quad \{x_n \ge 0\} \cap \operatorname{supp} \psi_{\hat{F}} \subseteq \bar{\hat{S}}.$$

Mit Hilfe von F_S kann Definition 3.6 auf S und F übertragen werden:

$$\begin{gathered} V_S = \{u \circ F_S^{-1} \mid u \in V_{\hat{S}}\}, \quad V_F = \{v \circ F_F^{-1} \mid v \in V_{\hat{F}}\}, \\ P : L_\infty(F) \longrightarrow L_\infty(S) : v \longmapsto \big(\hat{P}(v \circ F_F)\big) \circ F_S^{-1}, \\ \psi_S = \psi_{\hat{S}} \circ F_S^{-1}, \quad \psi_F = \psi_{\hat{F}} \circ F_S^{-1}. \end{gathered}$$

Bemerkung 3.7 *(Bubblefunktionen).* Sind $\lambda_0, \ldots, \lambda_n$ die baryzentrischen Koordinaten bezüglich $\hat{S}$, so ist $\psi_{\hat{S}} = (n+1)^{(n+1)} \prod_{i=0}^n \lambda_i$ eine Abschneidefunktion auf $\hat{S}$. Ebenso ist $\psi_{\hat{F}} = n^n \prod_{i=0}^{n-1} \lambda_i$ eine Abschneidefunktion auf $\hat{F}$.

Es wird angenommen, daß jeder Facette $F \in \mathcal{F}(\mathcal{T})$ eine Einheitsnormale n_F zugeordnet ist, die auf $\partial\Omega$ mit der äußeren Normale an den Gebietsrand übereinstimmt. Ist f eine auf jedem $S \in \mathcal{T}^{(n)}$ stetige Funktion, so ist der Sprung

$$[f]_F(x) = \lim_{t \to 0^+} f(x + t n_F) - \lim_{t \to 0^+} f(x - t n_F)$$

für jedes $x \in F$ und jedes $F \in \mathcal{F}(\mathcal{T})$ wohldefiniert, wenn man wegen der Facetten in $\partial\Omega$ f durch Null auf das komplement von Ω fortsetzt.
Zu einem beliebigen Hilbertraum X von Funktionen auf einem Gebiet M und einer abgeschlossenen Menge $N \subseteq M$ wird $X|_N = \{f \in X \mid \operatorname{supp} f \subseteq N\}$ gesetzt.

Als richtige Wahl für $\tilde{X}_h$ in dem Sinne, daß (3.3) gilt, werden sich Räume von Funktionen erweisen, die sich auf n-Simplexen und deren Facetten als $\psi_S u$ bzw. $\psi_F v$ mit $u \in V_S$, $v \in V_F$ darstellen lassen. Dabei sind V_S und V_F Polynomräume kleinen Grades. Das nachfolgende technische Lemma dokumentiert den Einfluß der Abschneidefunktion und des Fortsetzungsoperators auf einige Normabschätzungen. Es findet sich in [39] als Lemma 3.3.

Lemma 3.8 (abgeschnittene Funktionen). *Es gibt Konstanten $C_1, \ldots, C_7 > 0$, die nur von $V_{\hat{S}}$, $V_{\hat{F}}$, $\psi_{\hat{S}}$, $\psi_{\hat{F}}$ und der Regularitätskonstante δ abhängen, mit*

denen alle $S \in \mathfrak{T}^{(n)}$, $F \in \mathcal{F}(S)$, $u \in V_S$, $v \in V_F$ folgende Ungleichungen erfüllen:

$$C_1\|u\|_S \le \sup_{w\in V_S} \|w\|_S^{-1} \int_S u\psi_S w \le \|u\|_S, \tag{3.6a}$$

$$C_2\|v\|_F \le \sup_{w\in V_F} \|w\|_F^{-1} \int_F v\psi_F w \le \|v\|_F, \tag{3.6b}$$

$$C_3 h_S^{-1}\|\psi_S u\|_S \le \|\,\mathrm{D}(\psi_S u)\|_S \le C_4 h_S^{-1}\|\psi_S u\|_S, \tag{3.6c}$$

$$C_5 h_S^{-1}\|\psi_F P v\|_S \le \|\,\mathrm{D}(\psi_F P v)\|_S \le C_6 h_S^{-1}\|\psi_F P v\|_S, \tag{3.6d}$$

$$\|\psi_F P v\|_S \le C_7 h_S^{\frac{1}{2}}\|v\|_F. \tag{3.6e}$$

Beweis. Siehe [39, Kap.3]. Alle Schritte sind Standardmethoden: Man transformiert auf $\hat{S}$ bzw. $\hat{F}$ und weist nach, daß die jeweils gegeneinander abgeschätzten Terme Normen auf $V_{\hat{S}}$ bzw. $V_{\hat{F}}$ definieren. Wegen $\dim V_{\hat{S}}$, $\dim V_{\hat{F}} < \infty$ sind diese Normen äquivalent. Schließlich transformiert man auf S bzw. F zurück. □

Neben den Standardresultaten (3.6c) bis (3.6e) sind vor allem die unteren Schranken in (3.6a) und (3.6b) interessant.[5] Sie besagen, daß die mit Abschneidefunktionen gewichteten $L_2(S)$- und $L_2(F)$-Skalarprodukte auf V_S und V_F stabil sind, d. h., sie erfüllen die Babuška-Bedingung (1.27). Deshalb kann man sie zur Darstellung von auf V_S und V_F stetigen linearen Funktionalen verwenden. Jedes solche Funktional wird eindeutig durch eine Funktion aus V_S bzw. V_F repräsentiert. Ebenfalls mit Standardmethoden beweist man folgendes

Lemma 3.9. *Es gibt eine Konstante $C_8 > 0$, die dieselben Eigenschaften wie $C_1, \dots, C_7$ aus Lemma 3.8 aufweist, so daß für alle $S \in \mathfrak{T}^{(n)}$, $F \in \mathcal{F}(S)$, $v \in V_F$ diese Ungleichung gilt:*

$$\|\psi_F P v\|_F \le C_8 h_F^{\frac{1}{2}}\|\,\mathrm{D}(\psi_F P v)\|_S.$$

Beweis. Seien $S \in \mathfrak{T}$ ein beliebiges n-Simplex und $F \in \mathcal{F}(S)$ eine Facette. Dazu existiert eine affine Abbildung F_S, die $\hat{F}$ auf F und $\hat{S}$ auf S abbildet. Man untersucht nun eine beliebige Funktion $v \in V_F$. Durch Vorschalten von F_S transformiert man auf $\hat{F}$: $\|\psi_F P v\|_F = D_F^{\frac{1}{2}}\|\psi_{\hat{F}}\hat{P}\hat{v}\|_{\hat{F}}$. Die Funktionaldeterminante D_F für das Oberflächenintegral genügt auf regulären Triangulierungen der Abschätzung $C_1 h_F^{n-1} \le D_F \le C_2 h_F^{n-1}$ mit Konstanten, die nur von δ abhängen.
Nun betrachtet man $\psi_{\hat{F}}\hat{P}\hat{v}$ als Funktion auf $\hat{S}$. Diese Fortsetzung ist nur stetig, wenn man auf $\hat{S}$ eine stärkere Norm wählt[6]: $\|\psi_F P v\|_F \le C D_F^{\frac{1}{2}}\|\psi_{\hat{F}}\hat{P}\hat{v}\|_{1;\,\hat{S}} \le C D_F^{\frac{1}{2}}|\psi_{\hat{F}}\hat{P}\hat{v}|_{1;\,\hat{S}}$. Der vorige Schritt rechtfertigt sich über die Poincaré-Ungleichung. Transformiert man nun via F_S^{-1} zurück, so erhält man mit der Funktionaldeterminante $C_3 h_S^n \le D_S \le C_5 h_S^n$ die Ungleichung

$$\|\psi_F P v\|_F \le C D_F^{\frac{1}{2}} D_S^{-\frac{1}{2}} h_S |\psi_F P v|_{1;\,S}.$$

[5] Die oberen Schranken folgen sofort aus der Cauchy-Schwarzschen Ungleichung.
[6] Siehe z. B. [19].

Nun gilt für $D_F^{\frac{1}{2}} D_S^{-\frac{1}{2}} h_S$ die Abschätzung $D_F^{\frac{1}{2}} D_S^{-\frac{1}{2}} h_S \leq C h_F^{\frac{n-1}{2}} h_F^{-\frac{n}{2}} h_F = C h_F^{\frac{1}{2}}$.[7] Damit ist das Lemma bewiesen. □

Konstruktion

Zunächst werden die Hauptgedanken bei der Konstruktion von $\tilde{X}_h$ und $\tilde{R}_h$ dargestellt. Dazu sei $x_h = (u_h, p_h)^T \in X_h$ beliebig.

1. $\langle R(x_h), y \rangle$ besteht nur aus L_2-Skalarprodukten von Daten aus Aufgabe 1.30, p_h, $\mathrm{D}p_h$ und $\mathrm{D}u_{h,i}$ $(i = 1, \ldots, n)$ mit der Testfunktion y und deren Ableitungen. Die Skalarprodukte können wegen der Additivität des Lebesgue-Integrals als $\int_\Omega \cdot = \sum_{S \in \mathcal{T}^{(n)}} \int_S \cdot$ geschrieben werden. Daraus wird sich unter anderem die Lokalität der Fehlerschätzer ergeben.

2. Durch partielle Integration auf den n-Simplexen kann man sich der Gradienten der Testfunktionen in $\langle R(x_h), y \rangle$ entledigen. $R(x_h)$ besteht als lineares Funktional jetzt nur noch aus einer Summe von L_2-Skalarprodukten über n-Simplexe und Facetten, in denen keine Ableitung von y mehr auftritt. Man erhält so einen deutlichen Hinweis auf eine mögliche Struktur von $\tilde{X}_h$:

   Wenn $\tilde{X}_h$ auf jedem n-Simplex S alle Funktionen der Form $\tilde{x} = \psi_S \cdot \tilde{y}$ mit $\tilde{y} \in V_S$ enthält, kann mit Hilfe von (3.6a) und (3.6b) die $\tilde{X}_h'$-Norm von $R(x_h)$ nach unten beschränkt werden. Genau dies ist zur Anwendung von Satz 3.3 notwendig. Analog gilt dies für Integrale über Facetten. Da die vorige Wahl von Testfunktionen untere Schranken ermöglicht, führt sie zur Effizienz der zu konstruierenden Schätzer.

3. Leider erfordert Lemma 3.8, daß dann alle Daten von $R(x_h)$ lokal in V_S bzw. V_F liegen. In Bezug auf x_h, $\mathrm{D}x_{h,i}$ $(i = 1, \ldots, n)$ erhält man weitere Hinweise, wie $\tilde{X}_h$ zu wählen ist: Auf jedem n-Simplex sollte V_S die in $R(x_h)$ auftretenden Ableitungen von x_h enthalten.

   Man kann jedoch nicht erwarten, daß die Daten f, $E_{\Gamma_\mathrm{D},\Omega} u_\mathrm{D}$, usw., also die rechten Seiten der Stokes-Gleichungen, lokal in V_S liegen. Hier kommt $\tilde{R}_h$ ins Spiel. Man ersetzt die rechte Seite einfach durch Approximationen (z. B. mit Hilfe von Interpolation) in V_S.

Bemerkung 3.10. Daß bei dem zu konstruierenden Fehlerschätzer wegen Punkt 2 die klassische Form der Stokes-Gleichungen auftritt, ist kein Zufall, sondern ein weiteres Grundprinzip dieses Kapitels.

Die Punkte 1 und 2 liefern für eine beliebige Testfunktion $x = (v, q)^T \in X$ die Darstellung:

[7] Da F eine Facette von S ist, sind h_S und h_F wegen der Regularität der betrachteten Triangulierungen gegeneinander abschätzbar.

$$
\begin{aligned}
\langle R(x_h), x\rangle &= a(u_h, v) + b(u_h, q) + b(v, p_h) - r(x) \qquad (3.7)\\
&= \sum_{S\in\mathcal{T}^{(n)}} \Bigg(\nu \sum_{i=1}^{n} \int_S \mathrm{D}u_{h,i} \cdot \mathrm{D}v_i - \int_S q\,\mathrm{D}\cdot u_h - \int_S p_h\,\mathrm{D}\cdot v \\
&\quad - \int_S fv + \nu \sum_{i=1}^{n} \int_S (\mathrm{D}E_{\Gamma_\mathrm{D},\Omega}u_\mathrm{D})_i \cdot \mathrm{D}v_i - \int_S q\,\mathrm{D}\cdot E_{\Gamma_\mathrm{D},\Omega}u_\mathrm{D}\Bigg) \\
&\quad - \sum_{F\in\mathcal{F}_\mathrm{A}(\mathcal{T})} \int_F f_\mathrm{A} T_{\Omega,\Gamma_\mathrm{A}} v \\
&= \sum_{S\in\mathcal{T}^{(n)}} \Bigg(\int_S \Big(-\nu\Delta u_h + \mathrm{D}p_h - f - \nu\Delta E_{\Gamma_\mathrm{D},\Omega}u_\mathrm{D}\Big)v - \int_S \Big(\mathrm{D}\cdot u_h \\
&\quad + \mathrm{D}\cdot E_{\Gamma_\mathrm{D},\Omega}u_\mathrm{D}\Big)q\Bigg) + \sum_{F\in\mathcal{F}_\mathrm{A}(\mathcal{T})} \int_F \left(\nu\frac{\partial u_h}{\partial n} - np_h - f_\mathrm{A} + \nu\frac{\partial E_{\Gamma_\mathrm{D},\Omega}u_\mathrm{D}}{\partial n}\right) T_{\Omega,\Gamma_\mathrm{A}} v \\
&\quad + \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} \int_F \left(\nu[\frac{\partial u_h}{\partial n}] + \nu[\frac{\partial E_{\Gamma_\mathrm{D},\Omega}u_\mathrm{D}}{\partial n}]\right) T_{\Omega,\Gamma_\mathrm{A}} v
\end{aligned}
$$

In den Integralen über die n-Simplexe treten erste und zweite Ableitungen von u_h und erste Ableitungen von p_h auf. Wegen des mit v getesteten Ausdrucks sollte gemäß Überlegung 3 für die Geschwindigkeitsanteile in $\tilde{V}_h$ ($\tilde{X}_h = \tilde{V}_h \times \tilde{Q}_h$) die Inklusion $\psi_S\mathcal{P}_m \leq \tilde{V}_h|_{\hat{S}}$ mit $m = \max\{n_V - 2, n_Q - 1\}$ gelten. Das mit q getestete Integral erfordert $\psi_S\mathcal{P}_{n_V-1} \leq \tilde{Q}_h|_{\hat{S}}$. Für $\mathcal{P}_2\mathcal{P}_1$-Elemente bedeutet das $\psi_S\mathcal{P}_0 \leq \tilde{V}_h|_{\hat{S}}$ sowie $\psi_S\mathcal{P}_1 \leq \tilde{Q}_h|_{\hat{S}}$. Die Integrale über die Facetten enthalten erste Ableitungen von u_h und nullte Ableitungen von p_h. Demzufolge wird $\psi_F\mathcal{P}_{m'} \leq \tilde{V}_h|_{\hat{F}}$ mit $m' = \max\{n_V - 1, n_Q\}$ benötigt. Für $\mathcal{P}_2\mathcal{P}_1$-Elemente heißt das $\psi_F\mathcal{P}_1 \leq \tilde{V}_h|_{\hat{F}}$. Konkret wird

$$\tilde{V}_h = \operatorname{span}\left\{\psi_S u, \psi_F P v \;\middle|\; S\in\mathcal{T}^{(n)}, F\in\mathcal{F}_\Omega\cup\mathcal{F}_\mathrm{A}; u\in\mathcal{P}_m, v\in\mathcal{P}_{m'}\right\}, \qquad (3.8a)$$

$$\tilde{Q}_h = \begin{cases} \operatorname{span}\left\{\psi_S q \;\middle|\; S\in\mathcal{T}^{(n)}, q\in\mathcal{P}_{n_V-1}\right\}, & \text{falls} \quad |\Gamma_\mathrm{A}| > 0, \\ \operatorname{span}\left\{\psi_S q \;\middle|\; S\in\mathcal{T}^{(n)}, q\in\mathcal{P}_{n_V-1}, \int_\Omega \psi_S q = 0\right\} & \text{sonst}, \end{cases} \qquad (3.8b)$$

$$\tilde{X}_h = \tilde{V}_h \times \tilde{Q}_h \qquad (3.8c)$$

gesetzt, was den vorigen Überlegungen entspricht. Man verifiziert jetzt leicht, daß diese Wahl von $\tilde{X}_h$ in (3.7) keine Schwierigkeiten bereitet. Das bedeutet insbesondere, daß $\tilde{X}_h \leq H^1(\Omega)\times H^1(\Omega)$ gilt. Diese Aussage ist eine einfache Folgerung aus Lemma 2.17.
Der Übersichtlichkeit wegen wird u_D mit seiner Fortsetzung $E_{\Gamma_\mathrm{D},\Omega}u_\mathrm{D}$ identifiziert. Des weiteren wird der Spuroperator $T_{\Omega,\Gamma_\mathrm{A}}$ für Testfunktionen bei Integralen über Facetten immer implizit angenommen und nicht mehr aufgeschrieben.

Um die Daten f, u_D und f_A zu approximieren, bietet sich der Interpolationsoperator von Scott und Zhang an: Gegeben seien natürlichen Zahlen $\tilde{m} \leq m$ und $\tilde{m}' \leq m'$. Dann werden die Daten in den Integralen über n-Simplexe, die mit v getestet werden, durch $I_S = I^{\mathrm{Z},\tilde{m}}$ genähert. Die mit q getesteten Divergenzterme werden mit $I_Q = I^{\mathrm{Z},n_V-1}$ approximiert, falls $|\Gamma_\mathrm{A}| > 0$ gilt. Bei reinen Dirichlet-Randbedingungen geht man gemäß der Konstruktion von I_h auf Seite 60 vor: $I_Q = (id - \pi_0)I^{\mathrm{Z},n_V-1}$. Auf die Daten in Integralen über Facetten wird $I_F = I^{Z,\tilde{m}'}$ angewendet. Dabei bezeichnet $I^{\mathrm{Z},k}$ den Scott-Zhang-Interpolationsoperator, der mit Polynomen aus $\mathcal{P}_k$ interpoliert.[8] Für $\tilde{R}_h$ erhält man gemäß Punkt 3 für alle $x = (v,q)^T \in X$ den Ausdruck

$$\begin{aligned}\left\langle \tilde{R}_h(x_h), y \right\rangle = \sum_{S\in\mathcal{T}^{(n)}} \Bigg(\int_S \big(-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S \Delta u_\mathrm{D}\big)v - \int_S \big(\mathrm{D}\cdot u_h \\ + I_Q\,\mathrm{D}\cdot u_\mathrm{D}\big)q \Bigg) + \sum_{F\in\mathcal{F}_\mathrm{A}(\mathcal{T})} \int_F \left(\nu\frac{\partial u_h}{\partial n} - np_h - I_F f_\mathrm{A} + \nu I_F \frac{\partial u_\mathrm{D}}{\partial n}\right)v \\ + \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} \int_F \left(\nu[\frac{\partial u_h}{\partial n}] + \nu I_F[\frac{\partial u_\mathrm{D}}{\partial n}]\right)v.\end{aligned} \tag{3.9}$$

Die Störungsterme in Satz 3.3 enthalten wegen der Ähnlichkeit von R und $\tilde{R}_h$ hauptsächlich den Interpolationsfehler von I_S, I_F und I_Q. Verfürth verwendet in [39] zur Approximation der Daten immer die L_2-Projektion auf $\mathcal{P}_0$. Aus theoretischer Sicht zeigt sich durch die Verwendung der Interpolation nach Scott und Zhang deutlicher, daß die Störterme von höherer Ordnung sind, wenn die Daten glatt sind. Denn dann liefert Satz 2.20 weitere h-Potenzen gegenüber den Haupttermen in Satz 3.3.
Andererseits soll der Fehlerschätzer gerade über Gebiete geringerer Glattheit von Lastvektor und Lösung Auskunft geben. Deshalb erscheint es nicht sinnvoll, $\tilde{X}_h$ nur um höherer Interpolierender der Daten Willen zu vergrößern.
In der Implementierung in DROPS wird davon ausgegangen, daß die Daten wenigstens stetig sind, weil zu ihrer Interpolation der Standard-Interpolationsoperator eingesetzt wird.[9]

3.2 Residuumsschätzer

Ziel des vorliegenden Abschnittes ist die Anwendung von Satz 3.3. Die diversen Dualraum-Normen werden zunächst in einer Reihe vorbereitender Lemmata berechnet. Die Fehlerterme in Satz 3.3 machen den Anfang:

[8] Siehe Abschnitt 2.2.1.

[9] Die L_2-Projektion der Daten nach Verfürth hat für die Implementierung keine Vorteile, da die Integration in DROPS nur näherungsweise, d. h. über Quadraturformeln, erfolgt.

Lemma 3.11 (Fehlerterme). *Mit den Benennungen aus Abschnitt 3.1.3 lauten die Fehlerterme in* (3.4)

$$\|R(x_h) - \tilde{R}_h(x_h)\|_{\tilde{X}_h'} \leq C\Bigg(\sum_{S\in\mathcal{T}^{(n)}} h_S^2 \|(I_S - id)(f + \nu u_D)\|_S^2 + \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} h_F \|(I_F - id)([\frac{\partial u_D}{\partial n}])\|_F^2 + \sum_{F\in\mathcal{F}_A(\mathcal{T})} h_F \|(I_F - id)(f_A - \nu\frac{\partial u_D}{\partial n})\|_F^2 + \|(I_Q - id)\,\mathrm{D}\cdot u_D\|_\Omega^2 \Bigg)^{\frac{1}{2}}, \quad (3.10)$$

$$\|(id - I_h)'(R(x_h) - \tilde{R}_h(x_h))\|_{X'} \leq C\Bigg(\sum_{S\in\mathcal{T}^{(n)}} h_S^2 \|(I_S - id)(f + \nu u_D)\|_S^2 + \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} h_F \|(I_F - id)([\frac{\partial u_D}{\partial n}])\|_F^2 + \sum_{F\in\mathcal{F}_A(\mathcal{T})} h_F \|(I_F - id)(f_A - \nu\frac{\partial u_D}{\partial n})\|_F^2 + \|(I_Q - id)\,\mathrm{D}\cdot u_D\|_\Omega^2 \Bigg)^{\frac{1}{2}}. \quad (3.11)$$

Ferner erfüllt die Lösung $x_{0,h}$ der diskreten Aufgabe die Gleichung $\|R(x_{0,h})\|_{X_h'} = 0$.

Beweis. Zu (3.10): Sei $x = (v, q)^T \in \tilde{X}_h$, $\|x\|_X = 1$, beliebig. Dann ist

$$\Big\langle R(x_h) - \tilde{R}_h(x_h), x \Big\rangle = \sum_{S\in\mathcal{T}^{(n)}} \int_S \big((I_S - id)(f + \nu u_\mathrm{D})v + ((I_Q - id)\,\mathrm{D}\cdot u_\mathrm{D})q\big) + \sum_{F\in\mathcal{F}_\mathrm{A}(\mathcal{T})} \int_F (I_F - id)(f_\mathrm{A} - \nu\frac{\partial u_\mathrm{D}}{\partial n})v - \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} \int_F (I_F - id)([\frac{\partial u_\mathrm{D}}{\partial n}])v. \quad (3.12)$$

Die Integrale werden nun als L_2-Skalarprodukte mit der Cauchy-Schwarzschen Ungleichung abgeschätzt. Wegen (3.6c) und (3.6d) gilt $\|v\|_S \leq Ch_S\|v\|_{1;S}$. Beiträge über Facetten liefern nur die Anteile der Form $\psi_F P\tilde{v}$ von v. Entsprechend Lemma 3.9 wird $\|v\|_F \leq Ch_F^{\frac{1}{2}}\|v\|_{1;\omega_F}$ verwendet. Man erhält

$$\Big\langle R(x_h) - \tilde{R}_h(x_h), x \Big\rangle \leq C \sum_{S\in\mathcal{T}^{(n)}} \Big(\|(I_S - id)(f + \nu u_\mathrm{D})\|_S h_S \|v\|_{1;S} + \|(I_Q - id)\,\mathrm{D}\cdot u_\mathrm{D}\|_S \|q\|_S \Big) + C \sum_{F\in\mathcal{F}_\mathrm{A}(\mathcal{T})} \|(I_F - id)(f_\mathrm{A} - \nu\frac{\partial u_\mathrm{D}}{\partial n})\|_F h_F^{\frac{1}{2}} \|v\|_{1;\omega_F} + C \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} \|(I_F - id)([\frac{\partial u_\mathrm{D}}{\partial n}])\|_F h_F^{\frac{1}{2}} \|v\|_{1;\omega_F}.$$

Jedes n-Simplex S tritt in den vorstehenden Normen insgesamt höchstens $(n+1)$mal auf.[10] Durch Anwendung der diskreten Cauchy-Schwarzschen Ungleichung werden die Normen der Testfunktion von den übrigen Termen getrennt. Es gilt dann $(\sum\|x\|^2_{X;S})^{\frac{1}{2}} \le (n+1)\|x\|_X$. Deshalb ergibt sich

$$\begin{aligned}\Big\langle R(x_h) - \tilde{R}_h(x_h), x\Big\rangle &\le C\Big(\sum_{S\in\mathcal{T}^{(n)}} h_S^2\|(I_S - id)(f+\nu u_{\mathrm{D}})\|_S^2\Big)^{\frac{1}{2}}\|v\|_1\\ &+ C\|(I_Q - id)\,\mathrm{D}\cdot u_{\mathrm{D}}\|_\Omega\|q\|_S + C\Big(\sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} h_F\|(I_F - id)([\frac{\partial u_{\mathrm{D}}}{\partial n}])\|_F^2\Big)^{\frac{1}{2}}\|v\|_1\\ &+ C\Big(\sum_{F\in\mathcal{F}_{\mathrm{A}}(\mathcal{T})} h_F\|(I_F - id)(f_{\mathrm{A}} - \nu\frac{\partial u_{\mathrm{D}}}{\partial n})\|_F^2\Big)^{\frac{1}{2}}\|v\|_1.\end{aligned}$$

Man beachte, daß $\|v\|_1, \|q\| \le \|x\|_X = 1$ gilt. Durch Bildung des Supremums über $x \in \tilde{X}_h$, $\|x\|_X = 1$, ist (3.10) nachgewiesen.

Zu (3.11): Hier wird $x = (v,q)^T \in X$, $\|x\|_X = 1$, beliebig untersucht. Wegen

$$\Big\langle (id - I_h)'(R(x_h) - \tilde{R}_h(x_h)), x\Big\rangle = \Big\langle R(x_h) - \tilde{R}_h(x_h), x - I_h x\Big\rangle$$

entspricht der zu untersuchende Term (3.12). Allerdings ist die Testfunktion nun $x - I_h x \in X$, so daß nicht mehr die spezielle Struktur von $\tilde{X}_h$ ausgenutzt werden kann. Stattdessen verwendet man die lokalen Interpolationsaussagen aus Satz 2.20.

Wie im vorigen Beweisteil werden als erstes die Integrale über die einzelnen Simplexe mit der Cauchy-Schwarzschen Ungleichung abgeschätzt und anschließend $\|x - I_h x\|_S$ bzw. $\|x - I_h x\|_F$ mittels Satz 2.20 beschränkt. Dabei treten Normen über $\tilde{\omega}_S$ und $\tilde{\omega}_F$ auf. In diesen Mengen ist nach Lemma 2.15 höchstens eine nur von δ abhängige Anzahl i von Simplexen aus $\mathcal{T}^{(n)}$ enthalten. Deshalb gilt wieder $(\sum\|x\|^2_{X;S})^{\frac{1}{2}} \le i\|x\|_X$ und man erhält

$$\begin{aligned}\Big\langle (id - I_h)'(R(x_h) - \tilde{R}_h(x_h)), x\Big\rangle &\le C\Big(\sum_{S\in\mathcal{T}^{(n)}} h_S^2\|(I_S - id)(f+\nu u_{\mathrm{D}})\|_S^2\Big)^{\frac{1}{2}}\|v\|_1\\ &+ C\|(I_Q - id)\,\mathrm{D}\cdot u_{\mathrm{D}}\|_\Omega\|q\|_S + C\Big(\sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} h_F\|(I_F - id)([\frac{\partial u_{\mathrm{D}}}{\partial n}])\|_F^2\Big)^{\frac{1}{2}}\|v\|_1\\ &+ C\Big(\sum_{F\in\mathcal{F}_{\mathrm{A}}(\mathcal{T})} h_F\|(I_F - id)(f_{\mathrm{A}} - \nu\frac{\partial u_{\mathrm{D}}}{\partial n})\|_F^2\Big)^{\frac{1}{2}}\|v\|_1.\end{aligned}$$

[10] In der ersten Summe sowie als Teilmenge der ω_F.

Wegen

$$\|v\|_1, \|q\| \leq \|x\|_X = 1$$

hat man eine obere Schranke für $\sup_{x\in X,\, \|x\|_X=1} \left\langle (id - I_h)'(R(x_h) - \tilde{R}_h(x_h)), x \right\rangle$ gefunden.

Die Aussage über das Residuum der diskreten Lösung ist offensichtlich richtig. Das komplettiert den Beweis. □

Es lohnt sich, die Fehlerabschätzungen in Lemma 3.11 für homogene Dirichlet-Randbedingungen zu betrachten, denn die Terme sind deutlich kürzer. Das Auftreten von $\|(I_Q - id)\,\mathrm{D}\cdot u_\mathrm{D}\|_\Omega$ fällt unangenehm auf, weil keine Potenzen der Gitterweite auftreten. Damit dieser Term vernachlässigt werden kann, muß eine divergenzfreie Fortsetzung von u_D verwendet werden oder u_D sollte sehr glatt sein. Letzteres wird bei den Simulationen in Kapitel 6 der Fall sein. Ist u_D stückweise polynomial von hinreichend kleinem Grad, so verschwinden die Ausdrücke in u_D aufgrund der Projektionseigenschaften von I_S, I_Q und I_F.

Als nächstes wird der Hauptterm in Satz 3.3 untersucht. $\|\tilde{R}_h(x_h)\|_{\tilde{X}_h'}$ ist der eigentliche a posteriori Fehlerschätzer. Um die Berechnung dieser Dualnorm zu vereinfachen, wird in numerischen Simulationen der *Residuumsfehlerschätzer*

$$\begin{aligned}\eta_{R,S} = \Bigg(& h_S^2 \|-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S \Delta u_\mathrm{D}\|_S^2 + \|\,\mathrm{D}\cdot u_h + I_Q\,\mathrm{D}\cdot u_\mathrm{D}\|_S^2 \\ & + \frac{1}{2} \sum_{F\leq S, F\in\mathcal{F}_\Omega(\mathcal{T})} h_F \|[\nu \frac{\partial u_h}{\partial n} + \nu I_F \frac{\partial u_\mathrm{D}}{\partial n}]\|_F^2 \\ & + \sum_{F\leq S, F\in\mathcal{F}_\mathrm{A}(\mathcal{T})} h_F \|\nu \frac{\partial u_h}{\partial n} - np_h - I_F f_\mathrm{A} + \nu \frac{\partial u_\mathrm{D}}{\partial n}\|_F^2 \Bigg)^{\frac{1}{2}} \end{aligned} \tag{3.13}$$

eingesetzt. Aufgrund seiner Definition ist $\eta_{R,S}$ ein lokaler Fehlerschätzer. Er gestattet obere und untere Schranken für $\|\tilde{R}_h(x_h)\|_{\tilde{X}_h'}$ wie Lemma 3.12 und Lemma 3.13 demonstrieren. Das bedeutet Zuverlässigkeit und Effizienz.

Lemma 3.12 (obere Schranke). *Es gibt eine von h unabhängige Konstante $C > 0$, so daß gilt:*

$$\|\tilde{R}_h(x_h)\|_{\tilde{X}_h'} \leq C \sqrt{\sum_{S\in\mathcal{T}^{(n)}} \eta_{R,S}^2}.$$

Beweis. Man betrachtet ein beliebiges $x = (v, q)^T \in \tilde{X}_h$ mit $\|x\|_X = 1$. Zuerst wird die Summation über Facetten in (3.9) umsortiert: Jedes $F \in \mathcal{F}_\Omega(\mathcal{T})$ ist im Rand von genau zwei n-Simplexen. Auf diese wird jeweils die Hälfte des Integrals über F verteilt. Zu jeder Facette $F \in \mathcal{F}_\mathrm{A}(\mathcal{T})$ gibt es genau ein n-Simplex S mit

$F \leq S$. Dort wird F mitgezählt. Man erhält so für (3.9) die Formel

$$\begin{aligned}\left\langle \tilde{R}_h(x_h), x \right\rangle = \sum_{S\in\mathcal{T}^{(n)}} \Bigg(\int_S \big(-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S \Delta u_{\mathrm{D}}\big)v \\ - \int_S \big(\mathrm{D}\cdot u_h + I_Q\,\mathrm{D}\cdot u_{\mathrm{D}}\big)q + \sum_{F\leq S, F\in\mathcal{F}_{\mathrm{A}}(\mathcal{T})} \int_F \big(\nu\frac{\partial u_h}{\partial n} - np_h - I_F f_{\mathrm{A}} + \nu I_F \frac{\partial u_{\mathrm{D}}}{\partial n}\big)v \\ + \frac{1}{2}\sum_{F\in\mathcal{F}_\Omega(S)} \int_F [\nu\frac{\partial u_h}{\partial n} + \nu I_F \frac{\partial u_{\mathrm{D}}}{\partial n}]v \Bigg). \quad (3.14)\end{aligned}$$

Jetzt wird auf die einzelnen Integrale die Cauchy-Schwarzsche Ungleichung angewendet. Aufgrund von (3.6c) und (3.6d) gilt auf jedem n-Simplex S die Abschätzung $\|v\|_S \leq Ch_S\|v\|_{1;S}$. Auf den Facetten kommt Lemma 3.9 zum Einsatz: $\|v\|_F \leq Ch_F^{\frac{1}{2}}\|v\|_{1;\omega_F}$. Das ist möglich, weil alle Funktionen der Form $(\psi_s\tilde{v}, *)^T \in \tilde{X}_h$ auf allen Facetten verschwinden. Man erhält

$$\begin{aligned}\left\langle \tilde{R}_h(x_h), x \right\rangle \leq C \sum_{S\in\mathcal{T}^{(n)}} \Bigg(\|-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S\Delta u_{\mathrm{D}}\|_S h_S \|v\|_{1;S} \\ + \|\,\mathrm{D}\cdot u_h + I_Q\,\mathrm{D}\cdot u_{\mathrm{D}}\big\|_S \|q\|_S + \frac{1}{2}\sum_{F\in\mathcal{F}_\Omega(S)} \|[\nu\frac{\partial u_h}{\partial n} + \nu I_F\frac{\partial u_{\mathrm{D}}}{\partial n}]\|_F h_F^{\frac{1}{2}}\|v\|_{1;\omega_F} \\ + \sum_{F\leq S, F\in\mathcal{F}_{\mathrm{A}}(\mathcal{T})} \|\nu\frac{\partial u_h}{\partial n} - np_h - I_F f_{\mathrm{A}} + \nu I_F\frac{\partial u_{\mathrm{D}}}{\partial n}\|_F h_F^{\frac{1}{2}}\|v\|_{1;\omega_F} \Bigg). \quad (3.15)\end{aligned}$$

Aufgrund von $\|v\|_{1;S}, \|q\|_S, \|v\|_{1;\omega_F} \leq \|x\|_{X;\omega_S}$ können die Normen der Testfunktion zusammengefaßt werden. Dann wird die diskrete Cauchy-Schwarzsche Ungleichung verwendet. Da ω_S höchstens $n+2$ n-Simplexe aus $\mathcal{T}$ enthält, folgert man $\sqrt{\sum_{S\in\mathcal{T}^{(n)}}\|x\|^2_{X;\omega_S}} \leq \sqrt{n+2}\|x\|_X$. Deshalb ergibt sich

$$\begin{aligned}\left\langle \tilde{R}_h(x_h), x \right\rangle \leq C \Bigg(\sum_{S\in\mathcal{T}^{(n)}} \Bigg(h_S^2 \|-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S\Delta u_{\mathrm{D}}\|_S \\ + \|\,\mathrm{D}\cdot u_h + I_Q\,\mathrm{D}\cdot u_{\mathrm{D}}\big\|_S + \frac{1}{2}\sum_{F\in\mathcal{F}_\Omega(S)} h_F \|[\nu\frac{\partial u_h}{\partial n} + \nu I_F\frac{\partial u_{\mathrm{D}}}{\partial n}]\|_F \\ + \sum_{F\leq S, F\in\mathcal{F}_{\mathrm{A}}(\mathcal{T})} h_F \|\nu\frac{\partial u_h}{\partial n} - np_h - I_F f_{\mathrm{A}} + \nu I_F\frac{\partial u_{\mathrm{D}}}{\partial n}\|_F \Bigg)\Bigg)^{\frac{1}{2}} \|x\|_X.\end{aligned}$$

$x \in \tilde{X}_h$ war bis auf die Länge beliebig, so daß

$$\|\tilde{R}_h(x_h)\|_{\tilde{X}_h'} = \sup_{x\in\tilde{X}_h, \|x\|_X=1} \left\langle \tilde{R}_h(x_h), x \right\rangle$$

durch die vorherige Ungleichung beschränkt wird. In dieser findet sich im zweiten Klammerpaar von außen genau $\eta_{R,S}$ wieder, was den Beweis abschließt. □

Lemma 3.13 (untere Schranken). *Es existieren von h unabhängige Konstanten $C_1, C_2 > 0$, so daß gilt:*

$$\eta_{R,S} \leq C_1 \sup_{y \in \tilde{X}_h|\omega_S, \|y\|_X = 1} \left\langle \tilde{R}_h(x_h), y \right\rangle = C_1 \|\tilde{R}_h(x_h)\|_{\tilde{X}_h|'_{\omega_S}}, \tag{3.16}$$

$$\sqrt{\sum_{S \in \mathcal{T}(n)} \eta_{R,S}^2} \leq C_2 \|\tilde{R}_h(x_h)\|_{\tilde{X}'_h}. \tag{3.17}$$

Wegen (3.16) *ist $\eta_{R,S}$ sogar lokal ein effizienter Fehlerschätzer.*

Beweis. Zu (3.16): Sei $S \in \mathcal{T}$ ein beliebiges n-Simplex. $\eta_{R,S}^2$ besteht im wesentlichen aus vier Summanden. Es wird gezeigt, daß jeder davon durch $C\|\tilde{R}_h(x_h)\|_{\tilde{X}_h|'_\omega}$ nach oben beschränkt ist. Dabei gilt für $\omega \in \{S, \omega_S, \omega_F\}$ die Bezeichnung

$$\tilde{X}_h|_\omega = \left\{ x \in \tilde{X}_h \mid \operatorname{supp} x \subseteq \omega \right\}.$$

Im folgenden kommen die Gültigkeit von (3.6a) und (3.6b) sowie die Konstruktion von $\tilde{X}_h$ zum Tragen.

Erster Summand: Per Konstruktion von $\tilde{X}_h$ liegt $-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S \Delta u_\mathrm{D}$ auf S in $\mathcal{P}_m$. Deshalb kann (3.6a) auf seine $L_2(S)$-Norm angewendet werden. Im zweiten Schritt kommt folgende Ungleichung zum Zug: $2\|v\|_S^2 \geq 2\|\psi_S v\|_S^2 \geq C(1 + h_s^2)\|\psi_S v\|_{1;S}^2$; der erste Schritt folgt aus den Eigenschaften von ψ_S, der zweite ist eine Anwendung von (3.6c). Man erhält

$$\begin{aligned} h_S &\|-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S \Delta u_\mathrm{D}\|_S \\ &\leq C \sup_{v \in \mathcal{P}_m} h_S \|v\|_S^{-1} \int_S (-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S \Delta u_\mathrm{D})\psi_S v \\ &\leq C \sup_{v \in \mathcal{P}_m} \|\psi_S v\|_{1;S}^{-1} \int_S (-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S \Delta u_\mathrm{D})\psi_S v. \end{aligned}$$

Es gilt $x_v = (\psi_S v, 0)^T \in \tilde{X}_h|_S$ für alle $v \in \mathcal{P}_m$. Für die Norm von x_v findet man $\|x_v\|_X = \|\psi_S v\|_{1;S}$. Setzt man x_v in (3.9) als Testfunktion ein, so fallen alle Randintegrale weg, weil x_v auf $\Omega \setminus S$ Null ist. Ebenso fallen die mit der Druckkomponente von x_v multiplizierten Terme weg. Unter Wiederaufnahme der vorherigen Ungleichungskette erhält man

$$\begin{aligned} \sup_{v \in \mathcal{P}_m} &\|\psi_S v\|_{1;S}^{-1} \int_S (-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S \Delta u_\mathrm{D})\psi_S v \\ &\leq C \sup_{v \in \mathcal{P}_m} \|x_v\|_X^{-1} \left\langle \tilde{R}_h(x_h), x_v \right\rangle \\ &\leq C \sup_{x \in \tilde{X}_h|_S,\ \|x\|_X = 1} \left\langle \tilde{R}_h(x_h), x_v \right\rangle = \|\tilde{R}_h(x_h)\|_{\tilde{X}_h|'_S}. \end{aligned}$$

Der letzte Schritt beruht auf der Vergrößerung der Menge, über die das Supremum gebildet wird.

Zweiter Summand: Den Divergenzterm in (3.13) abzuschätzen, ist etwas einfacher. Aufgrund der Definition von $\tilde{X}_h$ und $\tilde{R}_h$ folgt $(\mathrm{D}\cdot u_h + I_Q\,\mathrm{D}\cdot u_\mathrm{D})|_S \in \mathcal{P}_{n_V-1}$, so daß (3.6a) benutzt werden kann. Setzt man zu $q \in \mathcal{P}_m$ die Testfunktion $x_q = (0, \psi_S q)^T \in \tilde{X}_h|_S$ in (3.9) ein, so bleibt genau der Divergenzterm übrig. Es gilt ferner $\|\psi_S q\|_S = \|x_q\|_X$. Also folgert man:

$$\begin{aligned}\|\,\mathrm{D}\cdot u_h + I_Q\,\mathrm{D}\cdot u_\mathrm{D}\|_S &\le C \sup_{q\in\mathcal{P}_m} \|q\|_S^{-1} \int_S (\mathrm{D}\cdot u_h + I_Q\,\mathrm{D}\cdot u_\mathrm{D})\psi_S q \\ &\le C \sup_{q\in\mathcal{P}_m} \|\psi_S q\|_S^{-1} \left\langle \tilde{R}_h(x_h), x_q \right\rangle \\ &\le C \sup_{x\in\tilde{X}_h|_S,\ \|x\|_X=1} \left\langle \tilde{R}_h(x_h), x \right\rangle = \|\tilde{R}_h(x_h)\|_{\tilde{X}_h|_S'}.\end{aligned}$$

Dritter Summand: Es genügt, eine beliebige Facette $F \le S$, $F \in \mathcal{F}_\Omega(\mathcal{T})$ zu analysieren. Wie schon zuvor gilt: $[\nu\frac{\partial}{\partial n}u_h + \nu I_F\frac{\partial}{\partial n}u_\mathrm{D}]_F$ ist in $\mathcal{P}_{m'}$ enthalten. Als erstes wird eine Ungleichung über die Testfunktion $x_v = (\psi_F P v, 0)^T \in \tilde{X}_h|_{\omega_F}$ mit beliebigem $0 \neq v \in \mathcal{P}_{m'}$ bewiesen. Wendet man nacheinander die linke und rechte Ungleichung aus (3.6d) auf v an, so findet man $\|\psi_F P v\|_{1;S} \le C\sqrt{1+h_S^2}\|\,\mathrm{D}(\psi_F P v)\|_S \le C h_S^{-1}\|\psi_F P v\|_S$, da h o. B. d. A. nach oben beschränkt ist. Zusammen mit (3.6e) ergibt sich $\|\psi_F P v\|_{1;S} \le C h^{-\frac{1}{2}}\|v\|_F$.
Nun wird (3.6b) benutzt, danach die gerade hergeleitete Ungleichung. Das auftretende $L_2(F)$-Skalarprodukt kann als Anwendung von $\tilde{R}_h(x_h)$ auf die Testfunktion x_v verstanden werden. Man beachte ferner $(Pv)|_F \equiv v_F$.

$$\begin{aligned}h_F^{\frac{1}{2}}\|[\nu\frac{\partial}{\partial n}u_h + \nu I_F\frac{\partial}{\partial n}u_\mathrm{D}]_F\|_F &\le C \sup_{v\in\mathcal{P}_{m'}} h^{\frac{1}{2}}\|v\|_F^{-1} \int_F [\nu\frac{\partial}{\partial n}u_h + \nu I_F\frac{\partial}{\partial n}u_\mathrm{D}]|_F \psi_F P v \\ &\le C \sup_{v\in\mathcal{P}_{m'}} \|x_v\|_X^{-1} \Bigg(\left\langle \tilde{R}_h(x_h), x_v \right\rangle \\ &\quad - \sum_{S\le\omega_F} \int_S (-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S \Delta u_\mathrm{D})\psi_F P v \Bigg)\end{aligned}$$

Der zusätzliche Term in der letzten Zeile der Ungleichung entspricht dem ersten Summanden, der im vorliegenden Beweis untersucht wurde. Deshalb kann er für jedes der maximal zwei n-Simplexe mit $S \le \omega_F$ gegen $\|\tilde{R}_h(x_h)\|_{\tilde{X}_h|_S'}$ abgeschätzt werden. Durch eine Vergrößerung des Testfunktionenraumes zu $\tilde{X}_h|_{\omega_F}$ erhält man eine eventuell noch größere obere Schranke für diesen Term.
Der verbleibende Term mit dem Dualitätsprodukt wird wie bei den vorigen Abschätzungen behandelt: Man wendet die Cauchy-Schwarzsche Ungleichung an

und vergrößert die Menge, über die das Supremum gebildet wird:

$$\begin{aligned} h_F^{\frac{1}{2}} \|[\nu \frac{\partial}{\partial n} u_h + \nu I_F \frac{\partial}{\partial n} u_{\mathrm{D}}]_F\|_F & \\ &\leq C \sup_{x \in \tilde{X}_h|_{\omega_F}} \|x\|_X^{-1} \left\langle \tilde{R}_h(x_h), x \right\rangle + 2C \|\tilde{R}_h(x_h)\|_{\tilde{X}_h|'_{\omega_F}} \\ &\leq C \|\tilde{R}_h(x_h)\|_{\tilde{X}_h|'_{\omega_F}}. \end{aligned}$$

Vierter Summand: Es wird eine beliebige Facette $F \leq S$ mit $F \in \mathcal{F}_{\mathrm{A}}(\mathcal{T})$ untersucht. Nach Konstruktion von $\tilde{R}_h$ liegt $(\nu \frac{\partial}{\partial n} u_h - n p_h - I_F f_{\mathrm{A}} + \nu \frac{\partial}{\partial n} u_{\mathrm{D}})|_F$ in $\mathcal{P}_{m'}$. Deshalb kann der Beweis für den dritten Summanden von $\eta_{R,S}$ unverändert übernommen werden. Man erhält

$$h_F^{\frac{1}{2}} \|\nu \frac{\partial}{\partial n} u_h - n p_h - I_F f_{\mathrm{A}} + \nu \frac{\partial}{\partial n} u_{\mathrm{D}}\|_F \leq C \|\tilde{R}_h(x_h)\|_{\tilde{X}_h|'_{\omega_F}}.$$

Jetzt werden die bisher bewiesenen Abschätzungen zusammengesetzt. Zählt man die Integrale über die Facetten einzeln, so besteht $\eta_{R,S}^2$ aus höchstens $n+3$ Summanden. Die Äquivalenz der 1-Norm und 2-Norm im $\mathbb{R}^{n+3}$ ergibt sofort die lokale untere Schranke (3.16), wenn man berücksichtigt, daß $S, \omega_F \subseteq \omega_S$ für S und alle seine Facetten F gilt.[11]
Um (3.17) zu zeigen, addiert man die lokalen Abschätzungen. Dabei muß man jedoch vorsichtig vorgehen, d. h., man kann nicht einfach die lokalen Dualnormen addieren. Dann müßte nämlich eine Summe von Suprema gegen das Supremum einer Summe abgeschätzt werden, wobei die Norm der Testfunktion unabhängig von h beschränkt bleiben muß.
Stattdessen steht folgender Weg offen: $\tilde{R}_h(x_h) \in X'$ kann via des Rieszschen Darstellungssatzes eindeutig in der Form $(\tilde{x}, \cdot)_X$ mit einem $\tilde{x} \in X$ geschrieben werden. Es gilt $\|\tilde{R}_h(x_h)\|_{X'} = \|\tilde{x}\|_X$. Der entscheidende Punkt ist, daß man von X' zu X gelangt, so daß man leichter die lokalen Abschätzungen addieren kann. Mit der Cauchy-Schwarzschen Ungleichung erhält man

$$\|\tilde{R}_h(x_h)\|_{\tilde{X}_h|_{\omega_S}} = \sup_{x \in \tilde{X}_h|_{\omega_S}, \|x\|_X = 1} (\tilde{x}|_{\omega_S}, x)_X \leq \sup_{x \in \tilde{X}_h|_{\omega_S}, \|x\|_X = 1} \|\tilde{x}|_{\omega_S}\|_X = \|\tilde{x}\|_{X|_{\omega_S}}.$$

Aufgrund von (3.16) folgt

$$\sum_{S \in \mathcal{T}^{(n)}} \eta_{R,S}^2 \leq C_1^2 \sum_{S \in \mathcal{T}^{(n)}} \|\tilde{x}\|_{X|_{\omega_S}}^2 \leq (n+2) C_1^2 \|\tilde{x}\|_X^2,$$

weil jedes n-Simplex in maximal $n+2$ Umgebungen ω_S auftritt. Nun kann man zu der Norm von X' zurückkehren: $\|\tilde{x}\|_X = \|\tilde{R}_h(x_h)\|_{X'}$, was den Beweis von Ungleichung (3.17) vervollständigt. □

[11] Insbesondere sind $\tilde{X}_h|_S$ und $\tilde{X}_h|_{\omega_F}$ Teilräume von $\tilde{X}_h|_{\omega_S}$.

Bemerkung 3.14 *(lokale Fehlerschranken).* Im Falle der Stokes-Gleichungen können lokale untere Fehlerschranken einfach durch Testfunktionen mit kleinem Träger im zu untersuchenden Teilgebiet bewiesen werden. Eine heuristische Erklärung dafür, daß dies eine „gute“ Abschätzung ergibt, lautet so: Der Diskretisierungsfehler $e_h = x_h - x_0$ genügt der Gleichung $Le_h = R(x_h)$, die wiederum die Stokes-Gleichungen darstellt. Also verhält sich e_h ebenso diffusiv wie ein zähes Fluid.
Lokale Abschätzungen des Fehlers auf $\omega \subseteq \Omega$ nach oben sind dagegen nicht zu gewinnen. Denn der Fehleranteil, der durch die Gleichung $Le_h = R(x_h)$ von $\Omega \setminus \omega$ nach Ω transportiert wird, kann den Fehler, der mit Testfunktionen $\operatorname{supp} f \subseteq \omega$ gemessen werden kann, stets dominieren. Besonders deutlich ist dies bei Transportgleichungen, weil dann eine Zerlegung von e_h in e_{lokal} und $e_{transport}$ vorliegt. $e_{transport}$ kann mit lokalen Testfunktionen nicht gut approximiert werden.

Zu guter Letzt wird die Stabilitätsbedingung (3.3) nachgewiesen.

Lemma 3.15 (Stabilität). *Mit den von h unabhängigen Konstanten $C_1, C_2 > 0$ erhält man für alle $x_h = (u_h, p_h)^T \in X_h$*

$$\|(id - I_h)' \tilde{R}_h(x_h)\|_{X'} \leq C_1 \Big(\sum_{S \in \mathcal{T}^{(n)}} \eta_{R,S}^2 \Big)^{\frac{1}{2}}$$

und deshalb zusammen mit (3.17) *die Ungleichung*

$$\|(id - I_h)' \tilde{R}_h(x_h)\|_{X'} \leq C_2 \|\tilde{R}_h(x_h)\|_{\tilde{X}_h'}.$$

Beweis. Der Beweis der ersten Ungleichung ist dem Beweis von Lemma 3.12 sehr ähnlich. Sei $x = (v, q)^T \in X$, $\|x\|_X = 1$, beliebig. Durch Sortieren der Terme in (3.9) erhält man wie in (3.14) die folgende Darstellung von $\Big\langle \tilde{R}_h(x_h), (id - I_h)x \Big\rangle$:

$$\begin{aligned}\Big\langle \tilde{R}_h(x_h), (id - I_h)x \Big\rangle = \sum_{S \in \mathcal{T}^{(n)}} \Bigg(\int_S \big(-\nu \Delta u_h + \mathrm{D} p_h - I_S f - \nu I_S \Delta u_{\mathrm{D}}\big)(id - I_h)v \\ - \int_S \big(\mathrm{D} \cdot u_h + I_Q \, \mathrm{D} \cdot u_{\mathrm{D}}\big)(id - I_h)q + \frac{1}{2} \sum_{F \in \mathcal{F}_\Omega(S)} \int_F [\nu \frac{\partial u_h}{\partial n} + \nu I_F \frac{\partial u_{\mathrm{D}}}{\partial n}](id - I_h)v \\ + \sum_{F \leq S, F \in \mathcal{F}_{\mathrm{A}}(\mathcal{T})} \int_F \big(\nu \frac{\partial u_h}{\partial n} - n p_h - I_F f_{\mathrm{A}} + \nu I_F \frac{\partial u_{\mathrm{D}}}{\partial n}\big)(id - I_h)v \Bigg).\end{aligned}$$

Auf sämtliche Integrale wird die Cauchy-Schwarzsche Ungleichung angewendet. Um wie im Beweis von Lemma 3.12 fortzufahren, werden die Integrale der Testfunktionen abgeschätzt. Während in jenem Beweis die speziellen Eigenschaften von $\tilde{X}_h$ für die folgende Abschätzung verwendet wurden, kommen hier die Interpolationseigenschaften von I_h zum Zug. Da I_h über den Interpolationsoperator

von Scott und Zhang definiert ist, liefert Satz 2.20 die für

$$\begin{aligned}\Big\langle \tilde{R}_h(x_h), (id - I_h)x \Big\rangle \\ \leq C \sum_{S\in\mathcal{T}^{(n)}} \Big(&\|-\nu\Delta u_h + \mathrm{D}p_h - I_S f - \nu I_S \Delta u_{\mathrm{D}}\|_S h_S \|v\|_{1;\tilde{\omega}_S} \\ + \|\,\mathrm{D}\cdot u_h + I_Q\,\mathrm{D}\cdot u_{\mathrm{D}}\big\|_S \|q\|_{\tilde{\omega}_S} &+ \frac{1}{2}\sum_{F\in\mathcal{F}_\Omega(S)} \|[\nu\frac{\partial u_h}{\partial n} + \nu I_F \frac{\partial u_{\mathrm{D}}}{\partial n}]\|_F h_F^{\frac{1}{2}} \|v\|_{1;\tilde{\omega}_F} \\ &+ \sum_{F\leq S, F\in\mathcal{F}_{\mathrm{A}}(\mathcal{T})} \|\nu\frac{\partial u_h}{\partial n} - np_h - I_F f_{\mathrm{A}} + \nu I_F \frac{\partial u_{\mathrm{D}}}{\partial n}\|_F h_F^{\frac{1}{2}} \|v\|_{1;\tilde{\omega}_F} \Big)\end{aligned}$$

notwendigen Ungleichungen. Der einzige Unterschied zu (3.15) besteht darin, daß die Integrationsgebiete $\tilde{\omega}_S$ und $\tilde{\omega}_F$ hier etwas größer sind als die entsprechenden in (3.15). Dennoch enthalten sie nach Lemma 2.15 stets weniger als eine von h unabhängige maximale Anzahl von Simplexen aus $\mathcal{T}$, so daß man jetzt wie im Beweis von 3.12 durch Anwendung der diskreten Cauchy-Schwarzschen Ungleichung das gewünschte Resultat erzielt.

Die zweite behauptete Ungleichung folgt sofort aus der gerade bewiesenen Ungleichung in Verbindung mit Lemma 3.13. □

Die vorangehenden Überlegungen ergeben das Hauptresultat dieses Kapitels über a posteriori Fehlerschätzung in der Norm von X:

Satz 3.16 (Der Residuumsschätzer $\eta_{R,S}$). *Für den Diskretisierungsfehler von $x_h = x_{0,h}$ gelten die folgenden a posteriori Schätzungen: Die Unleichung*

$$\begin{aligned}\|x_0 - x_{0,h}\|_X \leq C_1 \sqrt{\sum_{S\in\mathcal{T}^{(n)}} \eta_{R,S}^2} + \quad C_2 \Big(&\sum_{S\in\mathcal{T}^{(n)}} h_S^2 \|(I_S - id)(f + \nu u_D)\|_S^2 \\ &+ \|(I_Q - id)\,\mathrm{D}\cdot u_D\|_\Omega^2 + \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} h_F \|(I_F - id)([\frac{\partial u_D}{\partial n}])\|_F^2 \\ &+ \sum_{F\in\mathcal{F}_A(\mathcal{T})} h_F \|(I_F - id)(f_A - \nu\frac{\partial u_D}{\partial n})\|_F^2 \Big)^{\frac{1}{2}}.\end{aligned} \tag{3.18}$$

bedeutet, daß $\eta_{R,S}$ bis auf die Terme höherer Ordnung zuverlässig ist. Globale

Effizienz ergibt sich aus

$$\sqrt{\sum_{S\in\mathcal{T}^{(n)}}\eta_{R,S}^2} \le C_3\|x_0-x_{0,h}\|_X + \quad C_4\Bigg(\sum_{S\in\mathcal{T}^{(n)}} h_S^2\|(I_S-id)(f+\nu u_D)\|_S^2 + \|(I_Q-id)\,\mathrm{D}\cdot u_D\|_\Omega^2 + \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} h_F\|(I_F-id)([\frac{\partial u_D}{\partial n}])\|_F^2 + \sum_{F\in\mathcal{F}_A(\mathcal{T})} h_F\|(I_F-id)(f_A-\nu\frac{\partial u_D}{\partial n})\|_F^2\Bigg)^{\frac{1}{2}} \tag{3.19}$$

wieder bis auf Terme höherer Ordnung. Die letzte Unleichung besagt, daß $\eta_{R,S}$ auch lokal effizient ist:

$$\eta_{R,S} \le C_5\|x_0-x_{0,h}\|_{X|_{\omega_S}} + C_6\Bigg(\sum_{T\subseteq\omega_S,T\in\mathcal{T}^{(n)}} h_T^2\|(I_S-id)(f+\nu u_D)\|_T^2 + \|(I_Q-id)\,\mathrm{D}\cdot u_D\|_{\omega_S}^2 + \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})\cap\mathcal{F}(S)} h_F\|(I_F-id)([\frac{\partial u_D}{\partial n}])\|_F^2 + \sum_{F\in\mathcal{F}_A(\mathcal{T})\cap\mathcal{F}(S)} h_F\|(I_F-id)(f_A-\nu\frac{\partial u_D}{\partial n})\|_F^2\Bigg)^{\frac{1}{2}}. \tag{3.20}$$

Die Konstanten $C_1,\dots,C_6$ hängen nicht von h ab.

Beweis. Wegen Lemma 3.15 sind die Voraussetzungen von Satz 3.3 erfüllt. Setzt man dort die Gleichungen aus Lemma 3.11, 3.12, 3.13 ein, so erhält man Abschätzungen von $\|R(x_h)\|_{X'}$. Diese ergeben mit Satz 3.1 die Abschätzungen (3.18) und (3.19).
Um Ungleichung (3.20) zu zeigen, wird eine lokale Version der unteren Schranke aus (3.2) benötigt. Sei dazu $x\in X$ mit $\operatorname{supp} x\subseteq\omega\subseteq\Omega$ und $\|x\|_X=1$ beliebig. Dann gilt $\langle R(x_h),x\rangle = \big\langle L\big((x_h-x_0)\big|_\omega\big),x\big\rangle$, da alle Integrale in der Variationsformulierung nur über ω Beiträge liefern. Also folgt $\|x_h-x_0\|_{X|_\omega}\le \|L\|^{-1}\|R(x_h)\|_{X|_\omega'}$. Hier bedeutet wie bereits zuvor $X|_\omega=\{x\in X\mid \operatorname{supp} x\subseteq\omega\}$. Um das Residuum $R(x_h)|_\omega$ nach unten zu beschränken kann (3.4b) aus Satz 3.3 verwendet werden. Diese Ungleichung hängt nicht von der Stabilitätsbedingung (3.3) ab, so daß nur noch der Fehlerterm in (3.20) überprüft werden muß.
Sei $S\in\mathcal{T}^{(n)}$ beliebig. Da $\eta_{R,S}$ laut (3.16) eine untere Schranke auf ω_S liefert, setzt man $\omega=\omega_S$. Wie sieht eine Abschätzung für $\|(id-I_h)'(R(x_h)-\tilde{R}_h(x_h))\|_{X|_S'}$ aus? Verwendet man im Beweis von (3.10) Testfunktionen $x\in\tilde{X}_h$ mit $\operatorname{supp} x\subseteq\omega_S$, so verschwinden die Beiträge aller Simplexe zu (3.16), die nicht in ω_S liegen. Das beweist (3.20) und somit ist die gesamte Behauptung nachgewiesen. □

$\eta_{R,S}$ erweist sich somit als lokal und bis auf die Fehlerterme höherer Ordnung auch als zuverlässig und effizient. Wie hoch der Rechenaufwand ist, wird in Kapitel 6 durch die Experimente illustriert.

3.2.1 Fehlerschätzung in schwächeren Normen

Wie schon bei der a priori Fehlerschätzung in Abschnitt 2.1 kann man mit a posteriori Methoden den Fehler in schwächeren Normen als der von X schätzen. Hier werden zunächst die abstrakten Resultate aus [39, Kap. 2] vorgestellt. Die Anwendung auf die Stokes-Gleichungen wird danach diskutiert.
In Analogie zu Satz 2.4 wird die duale Aufgabe verwendet, um eine Abschätzung zu gewinnen. Im folgenden seien X_+, X_- Hilberträume mit der Eigenschaft $X_+ \hookrightarrow X \hookrightarrow X_-$. Die Pfeile repräsentieren stetige, dichte Inklusionen.[12]

Satz 3.17 (Fehler in der Norm von X_-). *Sei $L' \in \mathrm{GL}[X_+, X'_-]$ und $x_0 \in X_+$. Für alle $x \in X$ gilt*

$$\|L'\|^{-1}_{\mathcal{L}[X_+,X'_-]}\|R(x)\|_{X'_+} \le \|x - x_0\|_{X_-} \le \|L'^{-1}\|_{\mathcal{L}[X'_-,X_+]}\|R(x)\|_{X'_+}.$$

Beweis. Seien $x \in X$, $y \in X_+$ beliebig. Dann erhält man[13] $\langle R(x), y\rangle_{X'_+\times X_+} = \langle L(x - x_0), y\rangle_{X'_+\times X_+} = \langle L'y, x - x_0\rangle_{X'_-\times X_-}$.
Bildet man das Supremum über alle $y \in X_+$ mit $\|y\|_{X_+} = 1$, so erhält man die linke Ungleichung der Behauptung. Setzt man $z = L'y$ und bildet das Supremum über alle $z \in X'_-$ mit $\|z\|_{X'_-} = 1$, so ergibt sich $\|x - x_0\|_{X''_-} = \|x - x_0\|_{X_-} \le \|R(x)\|_{X'_+}\|L'^{-1}\|_{\mathcal{L}[X'_-,X_+]}$. Das beweist die rechte Abschätzung der Behauptung. □

In [39, Kap.2] findet man folgende Übertragung von Satz 3.3 auf die Situation in Satz 3.17:

Satz 3.18. *Sei $I_h \in \mathcal{L}[X_+, X_h]$ ein Projektor (für Testfunktionen) und $\tilde{X}_{+,h} \le X_+$. $\tilde{R}_h : X_h \longrightarrow X'_+$ eine Approximation von R auf X_h. Diese drei Objekte mögen für beliebige $x_h \in X_h$ der Stabilitätsgleichung*

$$\|(id - I_h)'\tilde{R}_h(x_h)\|_{X'_+} \le C\|\tilde{R}_h(x_h)\|_{\tilde{X}'_{+,h}}$$

mit einer von h unabhängigen Konstante $C > 0$ genügen. Dann kann das Residuum (3.1) von beliebigen $x_h \in X_h$ durch

$$\|R(x_h)\|_{X'_+} \le C\|\tilde{R}_h(x_h)\|_{\tilde{X}'_{+,h}} + \|(id - I_h)'(R(x_h) - \tilde{R}_h(x_h))\|_{X'_+} + \|id_{X_+}\|_{\mathcal{L}[X_+,X]}\|I_h\|_{\mathcal{L}[X,X_h]}\|R(x_h)\|_{X'_h},$$

$$\|\tilde{R}_h(x_h)\|_{\tilde{X}'_{+,h}} \le C\|\tilde{R}_h(x_h)\|_{\tilde{X}'_{+,h}} + \|R(x_h) - \tilde{R}_h(x_h)\|_{\tilde{X}'_{+,h}}$$

nach oben und unten abgeschätzt werden.[14]

[12] Dieses Tripel ist im allgemeinen kein Gelfandscher Dreier.
[13] Man beachte, daß Hilberträume reflexiv sind.
[14] Man beachte, daß X_h die X-Norm und $\tilde{X}_{+,h}$ die X_+-Norm trägt.

Für die Stokes-Gleichungen wählt man $X_- = L_2(\Omega)^n \times (H^1(\Omega) \cap Q)'$, $X_+ = (H^2(\Omega)^n \cap V) \times (H^1(\Omega) \cap Q)$. Es wird angenommen, daß $L = L'$ die Regularitätsbedingung aus Satz 3.17 erfüllt, also $L' \in \mathrm{GL}[X_+, X'_-]$.
Der Interpolationsoperator I_h wird genau wie in Abschnitt 3.1.3 definiert. Dies ist auf X_+ ohne Probleme möglich.

Um $\tilde{X}_{+,h} \leq X_+$ sicherzustellen, müssen Testfunktionen in $H^2(\Omega) \times H^1(\Omega)$ konstruiert werden. Die Druckanteile der Funktionen in $\tilde{X}_h$ erfüllen diese Bedingungen bereits aufgrund von Lemma 2.17: $\tilde{Q}_h \leq H^1(\Omega)$. Zur Konstruktion von Geschwindigkeits-Testfunktionen in $H^2(\Omega)$ kommt eine Verallgemeinerung des Lemmas 2.17 zum Einsatz: Ist $f \in C^1(\Omega)$ eine Funktion, die auf jedem n-Simplex $S \in \mathfrak{T}$ die Bedingung $f|_{\bar{S}} \in C^2(\bar{S})$ erfüllt, so folgt $f \in H^2(\Omega)$.
Sei zu $S \in \mathfrak{T}^{(n)}$ eine Abschneidefunktion $\tilde{\psi}_S$ gegeben und $u|_S \in \mathcal{P}_m$. Dann gilt für $\psi_S = \tilde{\psi}_S^2$, daß $\psi_S u \in \tilde{V}_h \leq H^1(\Omega)$ ist. Nach der Produktregel gilt für $\mathrm{D}(\psi_S u)$: $\mathrm{D}(\psi_S u) = \tilde{\psi}_S u \,\mathrm{D}\tilde{\psi}_S + \psi_S \,\mathrm{D}u = \tilde{\psi}_S(u \,\mathrm{D}\tilde{\psi}_S + \tilde{\psi}_S \,\mathrm{D}u)$. Die Ableitung ist auf S und $\Omega \setminus S$ stetig. Durch den Faktor $\tilde{\psi}_S$ ist sie auch auf ∂S stetig ($\equiv 0$). Somit gilt $\psi_S u \in C^1(\Omega)$, ergo $\psi_S u \in H^2(\Omega)$.
Es bleibt noch zu zeigen, daß die Funktionen der Form $\psi_F P v \in \tilde{V}_h$ ($F \in \mathfrak{F}(\mathfrak{T})$, $v|_F \in \mathcal{P}_{m'}$) in C^1 liegen. Zunächst wird $\psi_F = \tilde{\psi}_F^2$ mit einer Abschneidefunktion $\tilde{\psi}_F$ definiert. Wie zuvor erhält man, daß $\mathrm{D}(\psi_F P v)$ auf $\Omega \setminus \bar{F}$ stetig ist. Doch auf $\bar{F}$ ist $\mathrm{D}(\psi_F P v)$ im allgemeinen nicht stetig, denn P entsteht durch eine affine Transformation von $\hat{P}$, der konstanten Fortsetzung von $\hat{F} \longrightarrow \hat{S}$ in $\hat{e}_n$-Richtung. Sind S, T die n-Simplexe mit $F \leq S$, $F \leq T$, so wird $\hat{S}$ durch F_S und F_T nicht unbedingt gleich stark in $\hat{e}_n$-Richtung verzerrt. Deshalb ist $\mathrm{D}(\psi_F P v)$ im allgemeinen in der zu F orthogonalen Richtung nicht stetig.
Vermeidet man die affine Transformation und schneidet zusätzlich auf $\bar{F} \setminus F$ ab, so ist $\mathrm{D}(\psi_F P v)$ auch auf $\bar{F}$ stetig. Man erhält dann $\psi_F P v \in C^1(\Omega)$, also $\psi_F P v \in H^2(\Omega)$. Dies erreicht man, indem man die Fortsetzung P modifiziert. Zur genauen Konstruktion wird auf [39, Remark 3.6] verwiesen.
Lemma 3.8 bleibt auch für die modifizierte Fortsetzung $\tilde{P}$ gültig. Verfürth bemerkt in [39], daß die folgende Verallgemeinerung der Abschätzungen (3.6c) und (3.6d) gilt ($l \in \{1, 2\}$):

$$\tilde{C}_3 h_S^{-l} \|\psi_T u\|_{0;S} \leq \|\,\mathrm{D}^l(\psi_T u)\|_{0;S} \leq \tilde{C}_4 h_S^{-l} \|\psi_T u\|_{0;S}, \tag{3.21}$$

$$\tilde{C}_5 h_S^{-l} \|\psi_F \tilde{P} v\|_{0;S} \leq \|\,\mathrm{D}^l(\psi_F \tilde{P} v)\|_{0;S} \leq \tilde{C}_6 h_S^{-l} \|\psi_F \tilde{P} v\|_{0;S}. \tag{3.22}$$

Verwendet man die im vorangehenden Abschnitt definierten Abschneidefunktionen und $\tilde{P}$ statt P in (3.8), so erhält man $\tilde{X}_{+,h} \leq X_+$. Die Interpolationsoperatoren I_S, I_F und I_Q werden beibehalten, ebenso die Definitionen (3.9) und (3.13) von $\tilde{R}_h$ und $\eta_{R,S}$. Es ergibt sich folgendes Analogon zu Satz 3.16:

Satz 3.19 (Schätzung in der $L_2(\Omega) \times H^{-1}(\Omega)$-Norm). *Es sei $x_0 \in X_+$, und L' möge die Regularitätsbedingung $L' \in \mathrm{GL}[X_+, X'_-]$ erfüllen. Für den Diskretisierungsfehler von $x_h = x_{0,h}$ gelten die folgenden a posteriori Schätzungen: Die*

Ungleichung

$$\begin{aligned}\|x_0 - x_{0,h}\|_{X_-} \leq C_1 \sqrt{\sum_{S\in\mathcal{T}^{(n)}} h_S^2 \eta_{R,S}^2} + \quad C_2 \Bigg(\sum_{S\in\mathcal{T}^{(n)}} \Big(h_S^4 \|(I_S - id)(f + \nu u_D)\|_S^2 \\ + h_S^2 \|(I_Q - id)\,\mathrm{D}\cdot u_D\|_S^2 \Big) + \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} h_F^3 \|(I_F - id)([\frac{\partial u_D}{\partial n}])\|_F^2 \\ + \sum_{F\in\mathcal{F}_A(\mathcal{T})} h_F^3 \|(I_F - id)(f_A - \nu\frac{\partial u_D}{\partial n})\|_F^2 \Bigg)^{\frac{1}{2}}\end{aligned} \tag{3.23}$$

impliziert die globale Zuverlässigkeit von $h_S\eta_{R,S}$, wenn man von den Fehlertermen absieht. Globale Effizienz ergibt sich aus

$$\begin{aligned}\sqrt{\sum_{S\in\mathcal{T}^{(n)}} h_S^2 \eta_{R,S}^2} \leq C_3 \|x_0 - x_{0,h}\|_{X_-} + \quad C_4 \Bigg(\sum_{S\in\mathcal{T}^{(n)}} \Big(h_S^4 \|(I_S - id)(f + \nu u_D)\|_S^2 \\ + h_S^2 \|(I_Q - id)\,\mathrm{D}\cdot u_D\|_S^2 \Big) + \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})} h_F^3 \|(I_F - id)([\frac{\partial u_D}{\partial n}])\|_F^2 \\ + \sum_{F\in\mathcal{F}_A(\mathcal{T})} h_F^3 \|(I_F - id)(f_A - \nu\frac{\partial u_D}{\partial n})\|_F^2 \Bigg)^{\frac{1}{2}}\end{aligned} \tag{3.24}$$

bis auf die Fehlerterme. Als letztes gilt die lokale Effizienz der Größe $h_S\eta_{R,S}$, wenn man die Fehlerterme außer acht läßt:

$$\begin{aligned}h_S\eta_{R,S} \leq C_5 \|x_0 - x_{0,h}\|_{X_-|\omega_S} + C_6 \Bigg(\sum_{T\subseteq\omega_S, T\in\mathcal{T}^{(n)}} \Big(h_T^4 \|(I_S - id)(f + \nu u_D)\|_T^2 \\ + h_T^2 \|(I_Q - id)\,\mathrm{D}\cdot u_D\|_T^2 \Big) + \sum_{F\in\mathcal{F}_\Omega(\mathcal{T})\cap\mathcal{F}(S)} h_F^3 \|(I_F - id)([\frac{\partial u_D}{\partial n}])\|_F^2 \\ + \sum_{F\in\mathcal{F}_A(\mathcal{T})\cap\mathcal{F}(S)} h_F^3 \|(I_F - id)(f_A - \nu\frac{\partial u_D}{\partial n})\|_F^2 \Bigg)^{\frac{1}{2}}.\end{aligned} \tag{3.25}$$

Die Konstanten $C_1, \ldots, C_6$ hängen nicht von h ab.

Beweis. Es wird eine Skizze angegeben. In den Sätzen 3.3 und 3.18 treten dieselben Terme auf. Sie werden lediglich in der Norm von X_+ statt der von X gemessen. (Die einzige Ausnahme bildet $\|R(x_{0,h})\|_{X_h'}$. Dieser Ausdruck ist in beiden Fällen Null, da $x_{0,h}$ die diskrete Aufgabe löst.) Deshalb müssen nur die Beweise

von Lemma 3.11 bis Lemma 3.15 leicht modifiziert werden. Grundsätzlich gilt: Die Funktionen in X_+ besitzen eine um eins höhere Differenzierbarkeitsstufe als die Funktionen in X. Daher gewinnt man aus den Approximationssätzen in Abschnitt 2.2.1 eine Potenz von h_S. Statt (3.6c) und (3.6d) werden (3.21) und (3.22) eingesetzt. Dadurch verliert/gewinnt man ebenfalls eine h_S-Potenz.
Im Beweis von Lemma 3.11, Lemma 3.12 und der ersten Ungleichung in Lemma 3.15 wird anschließend die diskrete Cauchy-Schwarzsche Ungleichung angewendet, was die zusätzlichen h_S-Potenzen quadriert. Bei der lokalen unteren Schranke (3.16) benötigt man wegen der linken Ungleichung in (3.21) und (3.22) im Vergleich zu den ursprünglichen Abschätzungen eine zusätzliche Potenz von h_S, die beim Zusammensetzen von (3.17) quadriert wird.
Der Beweis der zweiten Ungleichung in Lemma 3.15 gelingt durch das Zusammenfügen der modifizierten ersten Ungleichung dieses Lemmas mit der angepaßten Ungleichung (3.17).
Durch Anwendung von Satz 3.18 und Satz 3.17 ergeben sich dann alle hier behaupteten Abschätzungen. □

3.3 Schätzer mit lokalen Stokes-Aufgaben

Anstatt wie in Satz 3.3 $\tilde{R}_h(x_h)$ auf einen Raum finiter Elemente höherer Ordnung zu projizieren, um die Norm von $R(x_h)$ abzuschätzen, kann man $\tilde{R}_h(x_h)$ für eine *Defektkorrektur* verwenden, indem man die Stokes-Gleichungen auf einem weiteren Raum $\hat{X}_h$ mit $\tilde{R}_h(x_h)$ als rechter Seite löst. Man erhält eine Korrektur $\hat{x}_h \in \hat{X}_h$, deren Norm als Maß für den Fehler verwendet werden kann.

Satz 3.20 (Residuumsschätzung durch Defektkorrektur). *Es seien die Voraussetzungen von Satz 3.3 erfüllt. Ferner sei $\hat{X}_h \leq X$ mit $\tilde{X}_h \leq \hat{X}_h$ gegeben und $\hat{L} \in \mathrm{GL}[\hat{X}_h, \hat{X}_h']$ eine Approximation des Differentialoperators L. Für beliebige $x_h \in X_h$ mögen die Räume $\tilde{X}_h$ und $\hat{X}_h$ die Ungleichung*

$$\|\tilde{R}_h(x_h)\|_{\hat{X}_h'} \leq C_1 \|\tilde{R}_h(x_h)\|_{\tilde{X}_h'} \tag{3.26}$$

mit einer von h unabhängigen Konstante $C_1 > 0$ erfüllen. Dann liefert die eindeutige Lösung $\hat{x}_h \in \hat{X}_h$ von

$$\left\langle \hat{L}\hat{x}_h, x \right\rangle = \left\langle \tilde{R}_h(x_h), x \right\rangle \quad \textit{für alle } x \in \hat{X}_h \tag{3.27}$$

($x_h \in X_h$ beliebig) die untere und obere Schranke

$$\|\hat{L}\|^{-1}_{\mathcal{L}[\hat{X}_h, \hat{X}_h']} \|\tilde{R}_h(x_h)\|_{\tilde{X}_h'} \leq \|\hat{x}_h\|_{\hat{X}_h} \leq C_1 \|\hat{L}^{-1}\|_{\mathcal{L}[\hat{X}_h', \hat{X}_h]} \|\tilde{R}_h(x_h)\|. \tag{3.28}$$

Beweis. Linke Ungleichung: Wegen (3.27) gilt

$$\|\hat{L}\hat{x}_h\|_{\hat{X}_h'} = \|\tilde{R}_h(x_h)\|_{\hat{X}_h'}. \tag{3.29}$$

Auf $\|\hat{L}\hat{x}_h\|_{\hat{X}_h'}$ wendet man $\|\hat{L}\hat{x}_h\|_{\hat{X}_h'} \leq \|\hat{L}\|_{\mathcal{L}[\hat{X}_h,\hat{X}_h']}\|\hat{x}_h\|_{\hat{X}_h}$ an. Durch Bildung des Supremums über $\tilde{X}_h \leq \hat{X}_h$ erhält man für $\tilde{R}_h(x_h)$ die Abschätzung $\|\tilde{R}_h(x_h)\|_{\hat{X}_h'} \geq \|\tilde{R}_h(x_h)\|_{\tilde{X}_h'}$. Setzt man diese zwei Ungleichungen in (3.29) ein, so erhält man die linke Abschätzung in (3.28).
Rechte Ungleichung: Nach Voraussetzung ist $\hat{L}$ auf $\hat{X}_h$ stetig invertierbar, so daß $\hat{x}_h = \hat{L}^{-1}(\tilde{R}_h(x_h))|_{\hat{X}_h}$ gilt. Deshalb beweist die Ungleichungskette

$$\|\hat{x}_h\|_{\hat{X}_h} \leq \|\hat{L}^{-1}\|_{\mathcal{L}[\hat{X}_h',\hat{X}_h]}\|\tilde{R}_h(x_h)\|_{\hat{X}_h'} \leq C\|\hat{L}^{-1}\|_{\mathcal{L}[\hat{X}_h',\hat{X}_h]}\|\tilde{R}_h(x_h)\|_{\tilde{X}_h}$$

die Behauptung des Satzes. Im letzten Schritt wurde (3.26) angewendet. □

Mit Hilfe von Satz 3.20 werden a posteriori Fehlerschätzer, die auf dem Lösen lokaler Stokes-Aufgaben beruhen, in der Theorie auf die Residuumsfehlerschätzer zurückgeführt. Damit die aus $\hat{L}$ resultierenden Gleichungssysteme möglichst geringe Dimension haben, ist man daran interessiert, daß $\hat{X}_h$ ein möglichst kleiner Raum ist.
Konkret wird $\hat{X}_h$ so definiert: Es sei $S \in \mathcal{T}$ ein beliebiges n-Simplex.

$$\begin{aligned}
\hat{V}_h &= \operatorname{span}\{\psi_S u, (\psi_F P v)|_S \mid F \leq \mathcal{F}(S) \cap (\mathcal{F}_\Omega(\mathcal{T}) \cup \mathcal{F}_A(\mathcal{T})), u \in \mathcal{P}_{m''}, v \in \mathcal{P}_{m'}\}, \\
\hat{Q}_h &= \operatorname{span}\{\psi_S q \mid q \in \mathcal{P}_{n_V-1}\} \text{ und} \\
\hat{X}_h &= \hat{V}_h \cap \hat{Q}_h
\end{aligned} \tag{3.30}$$

definieren den Ansatz- bzw. Testfunktionenraum für das lokale Stokes-Problem. $m' = \max\{n_V - 1, n_Q\}$ wird aufgrund der Analyse auf Seite 65 wie dort gewählt. $m'' = \max\{m, n+1+(n_V-1)-1\}$ mit m wie auf Seite 65 stellt sicher, daß $\psi_S Dp_h$ in $\hat{V}_h$ liegt, da ψ_S, ψ_F die "Bubble"-Abschneidefunktionen aus Bemerkung 3.7 sind. Diese Eigenschaft wird für den Nachweis der LBB-Bedingungen in Satz 3.23 benötigt. Alle $x \in \hat{X}_h$ erfüllen offensichtlich $\operatorname{supp} x \subseteq \bar{S}$ und sind per Definition glatt.
Nun wird $\hat{L}$ via

$$\hat{L} : \hat{X}_h \longrightarrow \hat{X}_h' : x \longmapsto (Lx)|_{\hat{X}_h} \tag{3.31}$$

als die Einschränkung von L auf $\hat{X}_h$ festgelegt. Somit ist $\hat{L} \in \mathrm{GL}[\hat{X}_h, \hat{X}_h']$, wenn $\hat{X}_h$ und b die diskrete Version der LBB-Bedingungen (1.34c) erfüllen.
Im folgenden wird die Lösung $\hat{x}_{0,h} \in \hat{X}_h$ von

$$\left\langle \hat{L}\hat{x}_{0,h}, x \right\rangle = \left\langle \tilde{R}_h(x_h), x \right\rangle \quad \text{für alle } x \in \hat{X}_h \tag{3.32}$$

als Fehlerschätzer untersucht. $\tilde{R}_h$ ist durch (3.9) gegeben, $x_h \in X_h$ beliebig. Aufgrund von Satz 3.23 existiert $\hat{x}_{0,h} \in \hat{X}_h$ und ist eindeutig.

Bemerkung 3.21 *(Randbedingungen).* Da jedes n-Simplex $S \in \mathcal{T}$ o. B. d. A. mindestens eine Ecke x in Ω besitzt, erfüllt die lokale Stokes-Aufgabe auf den an x liegenden Facetten von S natürliche Randbedingungen. Deshalb entfällt für $q \in \hat{Q}_h$ die Bedingung $\int_\Omega q = 0$.
Anstatt Satz 3.20 anzuwenden, wird der a posteriori Fehlerschätzer

$$\eta_{L,S} = \|\hat{x}_{0,h}\|_{X|_S} \tag{3.33}$$

direkt mit $\eta_{R,S}$ verglichen.

Satz 3.22 (Fehlerschätzung mit lokalen Stokes-Aufgaben). *Es gibt zwei von h unabhängige Konstanten $C_1, C_2 < 0$, so daß auf jedem n-Simplex $S \in \mathcal{T}$ die Ungleichungen*

$$\eta_{L,S} \le C_1 \eta_{R,S}, \tag{3.34}$$

$$\eta_{R,S} \le C_2 \Big(\sum_{T \in \mathcal{T}^{(n)},\, T \subseteq \omega_S} \eta_{L,T}^2 \Big)^{\frac{1}{2}} \tag{3.35}$$

erfüllt sind.

Beweis. Seien $S \in \mathcal{T}$ ein beliebiges n-Simplex und $\hat{x}_h \in \hat{X}_h$ die Lösung von (3.32) mit der rechten Seite $\tilde{R}_h(x_h)$, $x_h \in X_h$.
Zu (3.34): Es genügt, $\|\tilde{R}_h(x_h)\|_{\hat{X}_h'} \le C \eta_{R,S}$ zu beweisen, denn wie im Beweis von Satz 3.20 gilt $\eta_{L,S} = \|\hat{x}_h\|_{X|_S} \le \|\hat{L}^{-1}\| \|\tilde{R}_h(x_h)\|_{\hat{X}_h'}$. Dazu sei $x = (v,q)^T \in \hat{X}_h$ mit $\|x\|_X = 1$ beliebig. Durch Anwenden der Cauchy-Schwarzschen Ungleichung und der Abschätzungen (3.6c) und (3.6d) sowie des Lemmas 3.9 erhält man ähnlich wie im Beweis von Lemma 3.12

$$\begin{aligned}
\Big\langle \tilde{R}_h(x_h), x \Big\rangle &= \int_S \big(-\nu \Delta u_h + \mathrm{D} p_h - I_S f - \nu I_S \Delta u_{\mathrm{D}}\big) v \\
&\quad - \int_S \big(\mathrm{D} \cdot u_h + I_Q \, \mathrm{D} \cdot u_{\mathrm{D}} \big) q + \frac{1}{2} \sum_{F \in \mathcal{F}_\Omega(S)} \int_F [\nu \frac{\partial u_h}{\partial n} + \nu I_F \frac{\partial u_{\mathrm{D}}}{\partial n}] v \\
&\quad + \sum_{F \le S,\, F \in \mathcal{F}_{\mathrm{A}}(\mathcal{T})} \int_F \big(\nu \frac{\partial u_h}{\partial n} - n p_h - I_F f_{\mathrm{A}} + \nu I_F \frac{\partial u_{\mathrm{D}}}{\partial n}\big) v \\
&\le C \Big(h_S \| -\nu \Delta u_h + \mathrm{D} p_h - I_S f - \nu I_S \Delta u_{\mathrm{D}} \|_S \|v\|_{1;\,S} \\
&\quad + \| \mathrm{D} \cdot u_h + I_Q \, \mathrm{D} \cdot u_{\mathrm{D}} \big\|_S \|q\|_S \\
&\quad + \frac{1}{2} \sum_{F \in \mathcal{F}_\Omega(S)} h_F^{\frac{1}{2}} \| [\nu \frac{\partial u_h}{\partial n} + \nu I_F \frac{\partial u_{\mathrm{D}}}{\partial n}] \|_F \|v\|_{1;\,S} \\
&\quad + \sum_{F \le S,\, F \in \mathcal{F}_{\mathrm{A}}(\mathcal{T})} h_F^{\frac{1}{2}} \| \nu \frac{\partial u_h}{\partial n} - n p_h - I_F f_{\mathrm{A}} + \nu I_F \frac{\partial u_{\mathrm{D}}}{\partial n} \|_F \|v\|_{1;\,S} \Big).
\end{aligned} \tag{3.36}$$

Mittels der diskreten Cauchy-Schwarzschen Ungleichung und der Definition von $\eta_{R,S}$ erhält man $\left\langle \tilde{R}_h(x_h), x \right\rangle \leq C\eta_{R,S}\|x\|_X$, was durch Bilden des Supremums $\|\tilde{R}_h(x_h)\|_{\hat{X}_h'} \leq C\eta_{R,S}$ beweist.

Zu (3.35): Wegen $\|\hat{L}\|^{-1}\|\tilde{R}_h(x_h)\|_{\hat{X}_h'} \leq \|\hat{x}\|_{\hat{X}_h} = \eta_{L,S}$ reicht es aus, $\eta_{R,S} \leq C\sum_{T\in\mathfrak{T}^{(n)},\, T\subseteq\omega_S}\|\tilde{R}_h(x_h)\|_{\hat{X}_h'(T)}$ nachzuweisen. $\hat{X}_h(T)$ bezeichnet den Raum $\hat{X}_h$ für $S = T$ in (3.30). Nach Lemma 3.13 gilt $\eta_{R,S} \leq C\|\tilde{R}_h(x_h)\|_{\tilde{X}_h|_{\omega_S}'}$. Deshalb ist Ungleichung (3.35) bewiesen, wenn $\|\tilde{R}_h(x_h)\|_{\tilde{X}_h|_{\omega_S}'} \leq \sum_{T\in\mathfrak{T}^{(n)},\, T\subseteq\omega_S}\|\tilde{R}_h(x_h)\|_{\hat{X}_h'(T)}$ gezeigt wird.

Die entscheidende Beobachtung dazu lautet $\tilde{X}_h|_{\omega_S} \leq \bigoplus_{T\in\mathfrak{T}^{(n)},\, T\subseteq\omega_S}\hat{X}_h(T)$. Ein beliebiges $x \in \tilde{X}_h|_{\omega_S}$ kann somit als $x = \sum_{T\in\mathfrak{T}^{(n)},\, T\subseteq\omega_S}\pi_T(x)$ mit $\pi_T(x) \in \tilde{X}_h(T)$ geschrieben werden. Es ist ferner $\|x\|_X^2 = \sum_{T\in\mathfrak{T}^{(n)},\, T\subseteq\omega_S}\|\pi_T(x)\|_X^2$, so daß zusammen mit der Sublinearität des Supremums

$$\begin{aligned}
\|\tilde{R}_h(x_h)\|_{\tilde{X}_h|_{\omega_S}} &= \sup_{x\in\tilde{X}_h|_{\omega_S},\, \|x\|_X=1} \left\langle \tilde{R}_h(x_h), x \right\rangle \\
&= \sup_{x\in\tilde{X}_h|_{\omega_S},\, \|x\|_X=1} \sum_{T\in\mathfrak{T}^{(n)},\, T\subseteq\omega_S} \left\langle \tilde{R}_h(x_h), \pi_T(x) \right\rangle \\
&= \sum_{T\in\mathfrak{T}^{(n)},\, T\subseteq\omega_S} \sup_{x\in\hat{X}_h|_{\omega_S},\, \|x\|_X=1} \left\langle \tilde{R}_h(x_h), x \right\rangle \\
&= \sum_{T\in\mathfrak{T}^{(n)},\, T\subseteq\omega_S} \|\tilde{R}_h(x_h)\|_{\hat{X}_h'(T)}
\end{aligned}$$

folgt. Damit ist die Behauptung bewiesen. □

Die Frage nach der stetigen Invertierbarkeit von $\hat{L}$ (unabhängig von S und h) wird durch die V-Elliptizität von a und den folgenden Satz positiv beantwortet.

Satz 3.23 (Stabilität von $\hat{X}_h$). *Es existiert eine von h unabhängige Konstante $C > 0$, so daß*

$$\sup_{u\in\hat{V}_h} \|u\|_{1;S}^{-1} \int_S (-p)\,\mathrm{D}\cdot u \geq C\|p\|_S$$

für alle $p \in \hat{Q}_h$ gilt.

Beweis. Sei $0 \neq p = \psi_S q \in \hat{Q}_h$ beliebig. Nach Konstruktion von $\hat{V}_h$ ist $\{\psi_S v \mid v \in \mathcal{P}_{m''}\} \leq \hat{V}$ und es gilt $\mathrm{D}p \in \mathcal{P}_{m''}$. Somit erhält man aus (3.6c) und (3.6a)

$$\|p\|_S \leq Ch_S\|\,\mathrm{D}p\|_S \leq h_S \sup_{v\in\mathcal{P}_{m''}} \|v\|_S^{-1} \int_S \mathrm{D}p\,\psi_S v. \tag{3.37}$$

Partielle Integration ergibt $\int_S \mathrm{D}p\ \psi_S v = -\int_S p\,\mathrm{D}\cdot(\psi_S v)$, weil $p = \psi_S q$ auf ∂S verschwindet.

Ungleichung 3.6c ermöglicht für beliebige $v \in \mathcal{P}_{m''}$ die Ungleichungskette

$$\|\psi_S v\|_{1;S}^2 = \|\psi_S v\|_S^2 + \|\operatorname{D}(\psi_S v)\|_S^2 \leq C(1+h^{-2})\|v\|_S^2,$$

also $\|v\|_S^{-1} \leq C(1+h^{-2})^{\frac{1}{2}}\|\psi_S v\|_{1;S}^{-1}$. Das führt mit (3.37) zu

$$\|p\|_S \leq C(h_S^2+1)^{\frac{1}{2}} \sup_{v\in\mathcal{P}_{m''}} \|\psi_S v\|_1^{-1} \int_S (-p)\operatorname{D}\cdot(\psi_S v) \leq C \sup_{u\in\hat{V}_h} \|u\|_1^{-1} \int_S (-p)\operatorname{D}\cdot u.$$

Im letzten Schritt wird verwendet, daß h_S nach oben beschränkt ist[15]; zusätzlich wird die Menge, über die das Supremum gebildet wird, vergrößert. □

Obwohl die Gebiete, auf denen die lokalen Probleme gelöst werden, nur aus einem n-Simplex bestehen, hat $\hat{V}_h$ eine relativ hohe Dimension; denn nur dann kann die LBB-Bedingung für die lokalen Probleme nachgewiesen werden.[16] Verwendet man $\mathcal{P}_2\mathcal{P}_1$-Elemente für ein dreidimensionales Grundgebiet, so lauten die Grade der Polynome in den lokalen Räumen $n_V = 2$, $n_Q = 1$, $n = 3$, $m = 0$, $m' = 1$, $m'' = 4$, durch $\dim \hat{V}_h = 3(36 + 4\cdot 4) = 156$, $\dim \hat{Q}_h = 4$ bedingt. Dies macht sowohl die lokale Diskretisierung als auch das Lösen der lokalen Grundsysteme zu sehr teuren Operationen. Im Vergleich dazu lauten die Dimensionen für $n = 2$: $\dim \hat{V}_h = 2(15 + 3\cdot 3) = 48$, $\dim \hat{Q}_h = 3$.
Da die lokale Methode vermutlich recht langsam wäre, und wegen des hohen Aufwandes bei der Implementierung wird von einem numerischen Test des Schätzers $\eta_{L,S}$ abgesehen, solange die Stabilität nicht mit kleineren Räumen $\hat{V}_h$ nachgewiesen werden kann.

[15] durch $\operatorname{diam}(\Omega)$

[16] Ein Problem, das z. B. bei der Poissongleichung nicht auftritt.

Kapitel 4

A posteriori Fehlerschätzung II – DWR-Verfahren

Die Verfahren aus Kapitel 3 dienen in Verbindung mit dem adaptiven Zyklus in Abbildung 1 dazu, effizient eine Näherungslösung von $Lx = r$ mit einer vorgegebenen Genauigkeit zu bestimmen. In diesem Kapitel wird eine etwas allgemeinere Aufgabe betrachtet.
Oft ist man nicht nur an der Lösung x der vorigen Differentialgleichung interessiert, sondern es sollen weitere physikalische Zielgrößen bestimmt werden. Beispiele in der Fluiddynamik sind etwa der Auftrieb eines ins Fluid getauchten Gegenstandes und die Komponenten Reibungskraft, die auf die Oberfläche eines solchen wirkt. Abstrakt gesehen handelt es sich bei solchen Zielgrößen um Funktionale auf dem Raum der Ansatzfunktionen: $j : X \longrightarrow \mathbb{R}$. Im vorliegenden Kapitel wird eine Methode vorgestellt, mit der a posteriori, d. h. unter Verwendung von Näherungslösungen für Aufgabe 1.30, der Fehler $|j(x) - j(x_0)|$ der Zielgröße j abgeschätzt werden kann. Man erhält wie in Kapitel 3 lokale Indikatoren η. Sie beschreiben die Größe des Fehlers als Produkt eines lokalen Residualterms der Differentialgleichung mit einem Gewichtungsfaktor, der dessen Einfluß auf den Fehler der Zielgröße J angibt.
In [9] wird das DWR-Verfahren („dual-weighted-residual“) anhand der Poisson-Gleichung dargestellt. [10] ist eine Übersichtsarbeit, in der die Methode in einem Rahmen für nichtlineare Differentialgleichungen (wie z. B. die Navier-Stokes-Gleichungen) und beliebige, differenzierbare Funktionale vorgestellt wird. Hier wird nur der Fall linearer Differentialgleichungen und stetiger, linearer Funktionale j betrachtet. Dies erlaubt eine klarere Darstellung der Methode und ermöglicht dennoch eine im Rahmen dieser Arbeit wichtige Anwendung: Der Residuumsfehlerschätzer $\eta_{R,S}$ wird mittels des DWR-Verfahrens analysiert und modifiziert. Dazu wird das stetige, lineare Funktional

$$j : X \longrightarrow \mathbb{R} : x \longmapsto \|x_{0,h} - x_0\|_X^{-1} \left(x_{0,h} - x_0, x\right)_X$$

untersucht. $x_0 \in X$ und $x_{0,h} \in X_h$ bezeichnen die Lösungen der Aufgabe 1.30

und ihrer Diskretisierung. Der Zusammenhang mit $\eta_{R,S}$ aus Kapitel 3 folgt aus der Gleichung $\|x_{0,h}-x_0\|_X = j(x_{0,h}-x_0)$. Sie zeigt, daß die durch $\eta_{R,S}$ geschätzte Größe $\|x_{0,h}-x_0\|_X$ als Wert des Funktionals j auftritt.

4.1 Allgemeine Theorie

Zur Darstellung des Grundprinzips der DWR-Methode wird jetzt ein abstrakter Rahmen angenommen: (In Abschnitt 4.2 erhalten die nachfolgenden Symbole wieder ihre zur Stokes-Aufgabe passende Bedeutung.) $X_h \leq X$ seien reelle Hilberträume, $r, j \in X'$ beliebige lineare, stetige Funktionale auf X. Zu $l \in \mathcal{B}[X, X]$ sei $L \in \mathcal{L}[X, X']$ der zugehörige stetige lineare Operator. Er soll $L^{-1} \in \mathcal{L}[X', X]$ erfüllen.
$x_0 \in X$ bezeichne die (eindeutige, stets existierende) Lösung von $Lx = r$ in X; $x_{0,h} \in X_h$ sei die ebenfalls stets eindeutig existierende Lösung von $(Lx)|_{X_h} = r|_{X_h}$ in X_h (konforme Diskretisierung). Genau wie in (3.1) wird zu $x \in X$ das Residuum $R(x) = Lx - r = L(x - x_0)$ definiert.
Das Ziel der folgenden Überlegung besteht in der (näherungsweisen) Berechnung von $j(x_0)$. Betrachtet man $j(x_{0,h})$ als Approximation des gesuchten Wertes, muß insbesondere der Fehler $|j(x)-j(x_0)|$ abgeschätzt werden. Zu diesem Zweck wird das Funktional j mit der Variationsgleichung $Lx = r$ in Verbindung gebracht. Formal geschieht dies über die restringierte Optimierungsaufgabe

$$\text{Berechne } \inf\{j(x)|x \in X, Lx = r\}! \tag{4.1}$$

Da die Menge, deren Infimum gebildet wird, nach Voraussetzung genau das Element $j(x_0)$ enthält, ist das Infimum ein Minimum und somit der gesuchte Wert $j(x_0)$. Der Vorteil der Formulierung (4.1) liegt in der Anwendbarkeit der Multiplikatorenregel von Lagrange, um die Aufgabe zu lösen. Das geschieht in

Satz 4.1 (Lagrangesche Multiplikatoren). *Unter den Voraussetzungen des Abschnittes sind für jedes $x \in X$ die Aussagen*

1. *j(x) löst* (4.1).
2. *Es ist $Lx = r$ und es existiert ein $z \in X$ mit $L'z = j$.*

äquivalent. Sind x_0, z_0 die Lösungen der Gleichungen unter Punkt 2, so gilt die grundlegende Dualitätsgleichung

$$r(z_0) = \langle Lx_0, z_0\rangle = \langle L'z_0, x_0\rangle = j(x_0). \tag{4.2}$$

Beweis. „1. $\Rightarrow$ 2.“: Wie bereits erwähnt, ist $j(x_0)$ die Lösung von (4.1). Daher ist $Lx_0 = r$ erfüllt.
Da $L^{-1} \in \mathcal{L}[X', X]$ ist, gilt $L'^{-1} \in \mathcal{L}[X', X]$. $z_0 = L'^{-1}j$ löst die zweite Gleichung von Punkt 2.

„2. $\Rightarrow$ 1.“: Die Lösung $x_0 \in X$ von $Lx = r$ ist eindeutig, so daß das Infimum aus (4.1) genau $j(x_0)$ lautet.
Es bleibt noch der Nachweis der Dualitätsgleichung. Es gilt $r(y) = \langle Lx_0, y\rangle$ für alle $y \in X$. Durch die Wahl $y = z_0$ ergibt sich die linke Gleichung. Analog gilt für die Lösung $z_0 \in X$ von $L'z = j$: $j(y) = \langle L'z_0, y\rangle$ für alle $y \in X$, also insbesondere $j(x_0) = \langle L'z_0, x_0\rangle$. Die mittlere Gleichung folgt unmittelbar aus der Definition des zu L dualen Operators L' (und der Reflexivität des Hilbertraumes X). □

Bemerkung 4.2. Formal wird zur Lösung von (4.1) die Lagrangesche Multiplikatorenregel wie folgt angewendet: Man bestimmt die Sattelpunkte des Lagrangefunktionals $\mathfrak{L} : X \times X \longrightarrow R : (x, z)^T \longmapsto r(z) + j(x) - l(x, z)$ durch Nullsetzen der „partiellen Ableitungen“ nach x und z

$$0 = \frac{\partial \mathfrak{L}}{\partial x} = r - Lx, \quad 0 = \frac{\partial \mathfrak{L}}{\partial z} = j - l(\cdot, z) = j - L'z.$$

Mit Hilfe von Satz 4.1 werden die Eigenschaften von j über die Lösung $z_0 \in X$ von $L'z = j$ zugänglich. Die Lösung von $(L'z)|_{X_h} = j|_{X_h}$ in X_h wird mit $z_{0,h}$ bezeichnet und die Diskretisierungsfehler heißen $e = x_{0,h} - x_0$ sowie $e' = z_{0,h} - z_0$. Zusätzlich wird das Residuum $\rho : X \longrightarrow X' : z \longmapsto L'z - j = L'(z - z_0)$ eingeführt.
Bei der Ermittlung von $j(x_0)$ hilft Gleichung (4.2) nicht weiter. Doch mit ihrer Hilfe läßt sich der Fehler $j(x_{0,h}) - j(x_0) = j(e)$ beschreiben, der bei der Verwendung von $j(x_{0,h})$ als Näherung für $j(x_0)$ auftritt. Für die Fehler der diskreten Lösung gilt Galerkin-Orthogonalität, d. h.

$$(Le)|_{X_h} = 0 = (L'e')|_{X_h}. \tag{4.3}$$

Wegen dieser Eigenschaft überträgt sich die Dualität in (4.2) auf e und e':

$$\begin{aligned} j(e) = \langle L'z_0, e\rangle = \langle Le, z_0\rangle &= -\langle Le, e'\rangle \\ &= -\langle L'e', e\rangle = \langle L'e', x_0\rangle = \langle Lx_0, e'\rangle = r(e'). \end{aligned} \tag{4.4}$$

Gleichung (4.3) wird für die Unformungen 3 und 5 der Gleichungskette verwendet. Formuliert man (4.4) über die Residuen, so ergibt sich

Satz 4.3 (DWR-Fehlerschätzung). *Mit den Bezeichnungen dieses Abschnittes gilt die Fehlerdarstellung*

$$j(e) = \min_{y_h \in X_h} \langle R(x_{0,h}), z_0 - y_h\rangle = \min_{y_h \in X_h} \langle \rho(z_{0,h}), x_0 - y_h\rangle = r(e'). \tag{4.5}$$

Die Minima werden bei $z_{0,h}$ bzw. $x_{0,h}$ angenommen, weil sie bei beliebigen $y_h \in X_h$ angenommen werden.

Beweis. Sei $y_h \in X_h$ beliebig.
Erste Gleichung: Wegen (4.4) gilt $j(e) = \langle Le, z_0 \rangle$. Addiert man nun eine Null in Form von $\langle Le, -y_h \rangle = 0$ (Gleichung (4.3)), so folgt

$$j(e) = \langle R(x_{0,h}), z_0 - y_h \rangle ,$$

und durch Bilden des Minimums über $y_h \in X_h$ erhält man die Gültigkeit des ersten Gleichheitszeichens.
Dritte Gleichung: Der Beweis verläuft analog zum vorherigen.
Das mittlere Gleichheitszeichen ist eine direkte Konsequenz von (4.4), denn es gilt $j(e) = r(e')$. Damit ist der Beweis vollständig. □

Bemerkung 4.4.

- Wie der vorangehende Beweis demonstriert, ist es nicht notwendig, das Minimum in (4.5) zu bilden. In der Theorie für nichtlineare Differentialgleichungen[1] kann man es in der Fehlerdarstellung jedoch nicht vermeiden.

- Die Abkürzung „DWR“ steht für „dual-weighted-residual“. Gleichung (4.5) verdeutlicht den Grund dieser Benennung: Der Fehler $j(e)$ wird wie in Kapitel 3 durch das Residuum $R(x_{0,h})$ beschrieben. Dieses wird hier jedoch auf die spezielle Testfunktion z_0 angewendet. Die Lösung des dualen Problems gewichtet also das Residuum.

- Im Vergleich mit der in Kapitel 3 grundlegenden Ungleichung (3.2) bietet (4.5) vor allem folgenden Vorteil: Es treten keine Konstanten wie etwa $\|L^{-1}\|$ auf; (4.5) liefert eine exakte Darstellung des Fehlers. Je nachdem, wie genau z_0 approximiert wird, läßt sich mit DWR-Verfahren ein *asymptotisch exakter* Fehlerschätzer konstruieren. Das heißt, daß der *Effizienzindex*

  $$I_{\text{eff}} = \frac{\text{Schätzung von } |j(e)|}{|j(e)|}$$

  für $h \to 0$ gegen 1 konvergiert.[2]

Um zu einem numerisch berechenbaren Fehlerschätzer zu gelangen, sind die zu Beginn von Kapitel 3 formulierten Charakteristika (siehe Seite 55) zu beachten. Lokalität kann erreicht werden, solange $R(x_{0,h}) \in X'$ über L_2-Skalarprodukte dargestellt werden kann. Denn in diesem Fall können die Integrale über Ω und $\partial\Omega$ durch Summation von einzelnen Integralen über die $S \in \mathfrak{T}^{(n)}$ und die zugehörigen Facetten ermittelt werden. Diese Einzelintegrale liefern die lokale Schätzgröße. Die Summation über alle $S \in \mathfrak{T}$ ergibt dann die globale Schätzung.
Bevor Zuverlässigkeit und Effizienz im Rahmen der Stokes-Gleichungen in Abschnitt 4.2 untersucht werden, werden Näherungsmethoden für z_0 diskutiert. Diese hängen eng mit der Forderung nach moderatem Rechenaufwand zusammen.

[1] Siehe [10].

[2] Siehe [10, Kap.5]. Der zur Berechnung eines asymptotisch exakten Schätzers nötige Aufwand ist in der Praxis meist nicht akzeptabel.

4.1.1 Approximation der dualen Aufgabe

Im allgemeinen steht die Lösung $z_0 \in X$ von $L'z = j$ nicht zur Verfügung. Damit unter Verwendung von Satz 4.3 ein numerisch berechenbarer Fehlerschätzer konstruiert werden kann, muß z_0 approximiert werden.
Gleichung (4.5) weist darauf hin, daß die naheliegende Idee, z_0 durch $z_{0,h} \in X_h$ zu ersetzen, nicht zum Ziel führt. Denn wegen

$$\min_{y_h \in X_h} \langle R(x_{0,h}), z_{0,h} - y_h \rangle = 0$$

wird der so gebildete Fehlerschätzer unbrauchbar.
In [10, Kap.5] werden zwei Approximationsverfahren für z_0 vorgeschlagen, die den in Kapitel 3 verwendeten Fehlerschätzertypen ähneln.

1. *Globale Approximation höherer Ordnung:* Die Gleichung $L'z = j$ wird auf einem Finite-Elemente-Raum höherer Ordnung als X_h oder auf einer feineren Triangulierung gelöst. Deren Lösung wird dann statt z_0 in die Fehlerdarstellung eingesetzt. Wie eine numerische Simulation für die Poissongleichung nahelegt ([10, Kap.5.1, Tabelle 3]), sind mit solchen Approximationen asymptotisch exakte Fehlerschätzungen möglich.

   Für lineare Probleme sind die Kosten der zusätzlichen Diskretisierung von L' und das Lösen der Gleichung viel zu hoch. Deshalb wird diese Strategie hier nicht verfolgt.[3]

2. *Lokale Approximation höherer Ordnung:* In [10] wird davon ausgegangen, daß eine elementweise Projektion $I_h^+ z_{0,h}$ auf einen Ansatzraum X_h^+, der aus finiten Elementen höherer Ordnung besteht, eine verbesserte Approximation von z_0 liefert. In (4.5) setzt man dann $I_h^+ z_{0,h} - z_{0,h}$ für $z_0 - z_{0,h}$ ein. Zur Durchführung dieser Idee wird neben $\mathcal{T}$ eine gröbere Triangulierung $\mathcal{T}'$ benötigt, die mit $\mathcal{T}$ als Verfeinerung der Bedingung aus Satz 2.40 genügt.

   Ist die Approximation $f_1 \in S^k(\mathcal{T})$ einer Funktion f gegeben, so kann man jene mittels des Standard-Interpolationsoperators I_h^+ von $S^{k+1}(\mathcal{T}')$ auf $\tilde{f} = I_h^+ f_1 \in S^{k+1}(\mathcal{T}')$ projizieren. Da im allgemeinen $\mathcal{T}' \neq \mathcal{T}$ gilt, ist dann auch $\tilde{f} \neq f_1$. Allerdings ist $\tilde{f}$ von höherer Ordnung als f_1 und erlaubt somit in Interpolationssätzen Abschätzungen mit höheren h-Potenzen, falls f hinreichend regulär ist.

Eingedenk der Stabilitätsungleichung (3.3) könnte man auch versuchen, $z_{0,h}$ auf $\tilde{X}_h$ (zur Triangulierung $\mathcal{T}$) zu projizieren und $z_0 - y_h$ durch die Näherung $\tilde{I} z_{0,h} - z_{0,h}$ zu ersetzen. Allerdings ist nicht klar, wie der Projektor $\tilde{I}$ aussieht, da wegen der Abschneidefunktionen kein offensichtlicher Interpolationsoperator zur

[3] Bei nichtlinearen Differentialgleichungen kann anstelle von L' eine Linearisierung von L' untersucht werden. In diesem Kontext kann das Aufstellen und Lösen eines weiteren linearen Problems pro Iteration des adaptiven Zyklus akzeptabel sein.

Verfügung steht. Im Vergleich zu $I_h^+ z_{0,h}$ wird ohnehin kein Effizienzvorteil erwartet, weil auch für diese Approximation die Lösung $z_{0,h}$ der diskreten, dualen Aufgabe benötigt wird. Da es im Vergleich zu $\tilde{X}_h$ einfacher ist, $\mathcal{P}_3\mathcal{P}_2$-Elemente in DROPS zu implementieren, bleibt die praktische Verwendung von $\tilde{X}_h$ Gegenstand einer zukünftigen Untersuchung.
$z_0 - z_{0,h}$ durch die Lösung der lokalen Stokes-Aufgaben wie in Abschnitt 3.3 anzunähern, wird nicht der Forderung nach moderatem Rechenaufwand gerecht, solange keine „kleinen“, lokalen Testfunktionenräume zur Verfügung stehen. Bei der DWR-Methode ist die Annäherung von z_0 im lediglich eine Teilaufgabe, was zu einer wesentlich höheren Komplexität der DWR-Methode im Vergleich zur Berechnung der $\eta_{L,S}$ führen würde. Letzteres ist, wie in Abschnitt 3.3 ausgeführt, sehr aufwendig.

4.2 Anwendung auf die Stokes-Gleichungen

Nun wird wieder der zu den Stokes-Gleichungen gehörende Operator L untersucht. Zusätzlich zu den Generalvoraussetzungen aus Kapitel 3 (Seite 55) werden folgende Benennungen vereinbart: $(w_0, s_0)^T = z_0 \in X$ ist die Lösung von $L'z = j$ in X. Dabei ist $j \in X'$ vorerst beliebig. $(w_{0,h}, s_{0,h})^T = z_{0,h} \in X_h$ erfüllt $(L'z_{0,h})|_{X_h} = j|_{X_h}$, löst also die konform diskretisierte lokale Gleichung. Das Residuum der lokalen Aufgabe ist $\rho_h : X \longrightarrow X' : z \longmapsto L'z - j = L'(z - z_0)$.

Um einen a posteriori Fehlerschätzer für die Stokes-Gleichungen zu gewinnen, wird als nächstes wird Satz 4.3 angewendet. Das Residuum $R(x_{0,h})$ wird in (3.7) dargestellt, was wieder zur Lokalität des Schätzers führt. Wie vor der Definition von $\eta_{R,S}$ werden die Integrale über Facetten auf die ihnen benachbarten n-Simplexe aufgeteilt. Für beliebige $S \in \mathcal{T}^{(n)}$ gilt mit der Definition

$$\begin{aligned}
\eta_{\mathrm{DWR},S} = &\int_S \big(-\nu\Delta u_{0,h} + \mathrm{D}p_{0,h} - f - \nu\Delta u_{\mathrm{D}}\big)(w_0 - w_{0,h}) \\
&- \int_S \big(\mathrm{D}\cdot u_{0,h} + \mathrm{D}\cdot u_{\mathrm{D}}\big)(s_0 - s_{0,h}) \\
&+ \frac{1}{2}\sum_{F\le S, F\in\mathcal{F}_\Omega(\mathcal{T})} \int_F [\nu\frac{\partial u_{0,h}}{\partial n} + \nu\frac{\partial u_{\mathrm{D}}}{\partial n}](w_0 - w_{0,h}) \\
&+ \sum_{F\le S, F\in\mathcal{F}_{\mathrm{A}}(\mathcal{T})} \int_F \left(\nu\frac{\partial u_{0,h}}{\partial n} - np_{0,h} - f_{\mathrm{A}} + \nu\frac{\partial u_{\mathrm{D}}}{\partial n}\right)(w_0 - w_{0,h})
\end{aligned} \tag{4.6}$$

die Gleichung

$$\langle R(x_{0,h}), z_0 - z_{0,h}\rangle = \sum_{S\in\mathcal{T}^{(n)}} \eta_{\mathrm{DWR},S}. \tag{4.7}$$

Satz 4.5 (DWR-Fehlerschätzungen für die Stokes-Gleichungen). *Unter den Voraussetzungen des vorliegenden Abschnittes gilt die a posteriori Fehlerdarstellung*

$$|j(x_{o,h}) - j(x_0)| = |\sum_{S\in\mathcal{T}^{(n)}} \eta_{DWR,S}|. \tag{4.8}$$

Beweis. Die Stokes-Aufgabe erfüllt alle Voraussetzungen aus Abschnitt 4.1, so daß Satz 4.3 verwendet werden kann. Dieser ergibt mit (4.6) und (4.7) sofort die Behauptung. □

Neben der Darstellung (4.8) des globalen Fehlers, kommt als lokale Schätzgröße $|\eta_{\mathrm{DWR},S}|$ in Frage. Dieser Ausdruck wird in (4.8) nicht verwendet, um durch die Dreiecksungleichung die globale Fehlerschätzung nicht unnötig zu pessimieren. Man beachte, daß in der Abschätzung keine multiplikativen Konstanten auftreten (dafür allerdings z_0).

4.2.1 Zusammenhang mit $\eta_{R,S}$

Wie zu Beginn dieses Kapitels angekündigt, wird nun gezeigt, daß die Theorie aus Kapitel 3 in gewisser Weise einen Spezialfall des DWR-Verfahrens repräsentiert. Dazu wird als Zielgröße j das stetige lineare Funktional

$$j : X \longrightarrow \mathbb{R} : x \longmapsto \|x_{0,h} - x_0\|_X^{-1} \left(x_{0,h} - x_0, x\right)_X \tag{4.9}$$

gewählt. Man erhält $j(x_{0,h}) - j(x_0) = \|x_{0,h} - x_0\|_X$; deshalb kann die in Satz 3.16 via $\eta_{R,S}$ abgeschätzte Größe $\|x_{0,h} - x_0\|_X$ mit Hilfe von Satz 4.5 untersucht werden. Man findet dabei den über

$$\begin{aligned}\tilde{\eta}_{R,S} = \Bigg(& h_S^2 \|-\nu\Delta u_{0,h} + \mathrm{D}p_{0,h} - f - \nu\Delta u_{\mathrm{D}}\|_S^2 + \|\,\mathrm{D}\cdot u_{0,h} + \mathrm{D}\cdot u_{\mathrm{D}}\|_S^2 \\ & + \frac{1}{2} \sum_{F\le S, F\in\mathcal{F}_\Omega(\mathcal{T})} h_F \|[\nu\frac{\partial u_{0,h}}{\partial n} + \nu\frac{\partial u_{\mathrm{D}}}{\partial n}]\|_F^2 \\ & + \sum_{F\le S, F\in\mathcal{F}_{\mathrm{A}}(\mathcal{T})} h_F \|\nu\frac{\partial u_{0,h}}{\partial n} - np_{0,h} - f_{\mathrm{A}} + \nu\frac{\partial u_{\mathrm{D}}}{\partial n}\|_F^2 \Bigg)^{\frac{1}{2}}\end{aligned}$$

definierten Fehlerschätzer. Für ihn liefert Satz 4.5 die folgende globale Zuverlässigkeitsaussage, die der entsprechenden Aussage über $\eta_{R,S}$ in Satz 3.16 ähnelt:

Satz 4.6 ($\tilde{\eta}_{R,S}$ und das DWR-Verfahren). *Für den a posteriori Fehlerschätzer $\tilde{\eta}_{R,S}$ gilt die Abschätzung*

$$\|x_{0,h} - x_0\|_X = j(x_{0,h} - x_0) \le C \sqrt{\sum_{S\in\mathcal{T}^{(n)}} \tilde{\eta}_{R,S}^2}.$$

In $C > 0$ gehen die Konstanten des Interpolationssatzes 2.20 und die Norm von L'^{-1} ein. C ist unabhängig von h.

Beweis. Als erstes erhält man aus Satz 4.5 die Fehlerdarstellung (4.8). Auf die in $\eta_{\mathrm{DWR},S}$ auftretenden L_2-Skalarprodukte wird die Cauchy-Schwarzsche Ungleichung angewendet. Ergänzt man dann bei jedem Summanden einen Faktor 1 in der Form $h_S^{-1}h_S$ bzw. $h_F^{-\frac{1}{2}}h_F^{\frac{1}{2}}$ und benutzt die diskrete Cauchy-Schwarzsche Ungleichung, so folgt aus (4.8) insgesamt die Abschätzung

$$|j(x_{0,h}-x_0)| \leq \sqrt{\sum_{S\in\mathcal{T}^{(n)}} \tilde{\eta}_{R,S}^2}\sqrt{\sum_{S\in\mathcal{T}^{(n)}} \gamma_S^2}$$

mit den Gewichten

$$\gamma_S^2 = h_S^{-2}\|w_0-w_{0,h}\|_S^2 + \|s_0-s_{0,h}\|_S^2 + \frac{1}{2}\sum_{F\leq S, F\in\mathcal{F}_\Omega(\mathcal{T})} h_F^{-1}\|w_0-w_{0,h}\|_F^2 + \sum_{F\leq S, F\in\mathcal{F}_\mathrm{A}(\mathcal{T})} h_F^{-1}\|w_0-w_{0,h}\|_F^2.$$

Aufgrund von Bemerkung 4.4 dürfen in γ_S^2 die Terme $w_0-w_{0,h}$ und $s_0-s_{0,h}$ durch die entsprechenden Komponenten von z_0-Iz_0 ersetzt werden, ohne daß sich der Wert von γ_S ändert. I ist dabei der Interpolationsoperator, der in der a priori Konvergenzanalyse auf Seite 38 definiert wird.
Jetzt kann γ_S^2 mittels des Interpolationssatzes 2.20 durch

$$\gamma_S^2 \leq C_i^2(n+3)\|z_0\|_{X|_{\tilde{\omega}_S}}^2$$

abgeschätzt werden. $C_i>0$ ist eine von h unabhängige Interpolationskonstante. Wie in den Beweisen in Abschnitt 3.2 argumentiert man, daß alle $\tilde{\omega}_S$ nur eine von der Regularitätskonstante δ abhängende maximale Anzahl $n_{\max}<\infty$ von n-Simplexen aus $\mathcal{T}$ enthalten. Also folgt

$$\sqrt{\sum_{S\in\mathcal{T}^{(n)}}\omega_S^2} \leq C_i\sqrt{n_{\max}(n+3)}\|z_0\|_X \leq C_i\sqrt{n_{\max}(n+3)}\|L'^{-1}\|\,\|j\|_{X'}. \tag{4.10}$$

Mittels der Cauchy-Schwarzschen Ungleichung erhält man

$$\|j\|_{X'} = \sup_{x\in X,\|x\|_X=1}\|x_{0,h}-x_0\|_X^{-1}\,(x_{0,h}-x_0,x)_X \leq \frac{\|x_{0,h}-x_0\|_X}{\|x_{0,h}-x_0\|_X}\,1 = 1,$$

womit der Satz bewiesen ist. □

Der vorangehende Beweis beruht auf der Idee, für die Lösung des dualen Problems eine a priori Konvergenzanalyse durchzuführen. Die Ungleichung (4.10) veranschaulicht noch einmal den Namen des DWR-Verfahrens, weil sich zeigt, daß z_0 die lokalen Residuumsterme $\tilde{\eta}_{R,S}$ gewichtet. $\tilde{\eta}_{R,S}$ entspricht dabei bis auf die Approximation der Daten dem Residuumsschätzer $\eta_{R,S}$ aus Abschnitt 3.2.

Aufgrund von (4.10) kann man heuristisch begründen, daß die DWR-Fehlerschätzung eine Verbesserung[4] gegenüber den Residuumsmethoden aus Kapitel 3 darstellt, sobald der Lösungsoperator L'^{-1} der dualen Aufgabe eine große Norm hat. Denn diese Norm geht in den Effizienzindex von $\eta_{\mathrm{DWR},S}$ nicht direkt ein.
Satz 4.6 demonstriert ferner, daß bei exakter Integration der Daten die Fehlerterme in der oberen Schranke von Satz 3.16 wegfallen können. Untere Fehlerschranken, also lokale oder globale Effizienz gemäß der Charakteristika auf Seite 55, lassen sich allerdings mit den Mitteln aus Kapitel 3 für $\tilde{\eta}_{R,S}$ nicht so einfach nachweisen, weil Lemma 3.8 nicht mehr angewendet werden kann. Eventuell läßt sich dieses Problem durch eine zusätzliche Projektion auf $\tilde{X}_h$ (unter Inkaufnahme neuer Fehlerterme) beheben.

Untersucht man mit derselben Methode wie in Satz 4.6 das Funktional

$$j_2 : X \longrightarrow \mathbb{R} : x \longmapsto \|x_{0,h} - x_0\|^{-1}_{L_2\times H^{-1}} \left(x_{0,h} - x_0, x\right)_{L_2\times H^{-1}},$$

so erhält man Fehlerschätzungen, die denen aus Abschnitt 3.2.1 entsprechen. Allerdings benötigt man weitergehende Approximationsaussagen als die in Abschnitt 2.2.1.

4.2.2 DWR-Fehlerschätzung in der Norm von X

Durch Satz 4.5 und das Funktional j aus (4.9) wird das DWR-Verfahren für Fehlerschätzung in der X-Norm theoretisch beschrieben. Um $\eta_{\mathrm{DWR},S}$ praktisch auswerten zu können, muß zunächst die Abhängigkeit von der exakten Lösung z_0 des dualen Problems beseitigt werden. Außerdem hängt j von der exakten Lösung x_0 der Stokes-Aufgabe ab.
Die beiden exakten Lösungen werden, wie in Abschnitt 4.1.1 erläutert, durch die lokalen Projektionen von $x_{0,h}$ bzw. $z_{0,h}$ auf Ansatzräume höherer Ordnung ersetzt. Ist $\mathcal{T}'$ eine gröbere Triangulierung als $\mathcal{T}$, welche mit $\mathcal{T}$ zusammen die Bedingung aus Satz 2.40 erfüllt, so werden $x_{0,h}$ und $z_{0,h}$ auf

$$X_h^+(\mathcal{T}') = U_h \times P_h \quad \text{mit } U_h = \left(S^{n_V+1}(\mathcal{T}')\right)^n,$$

$$P_h = \begin{cases} S^{n_Q+1}(\mathcal{T}'), \text{ falls } |\Gamma_A| > 0, \\ S^{n_Q+1}(\mathcal{T}') \cap L_2^0(\Omega), \text{ sonst} \end{cases}$$

projiziert. Als Projektor dient der Standard-Interpolationsoperator

$$I^+ : X_h \longrightarrow X_h^+ : (u,p)^T \longmapsto (I^{\mathrm{S},n_V+1}u, I^{\mathrm{S},n_Q+1}p)^T$$

von X_h^+, wobei im Fall reiner Dirichlet-Randbedingungen für die Druckkomponenten der Interpolator $(1-\pi_0)I^{\mathrm{S},n_Q+1}$ eingesetzt wird. Ersetzt man j durch die

[4] Es handelt sich um eine Verbesserung in dem Sinn, daß der Effizienzindex des DWR-Verfahren näher an 1 liegt.

Näherung

$$j^+ : X \longrightarrow \mathbb{R} : x \longmapsto \|x_{0,h} - I^+ x_{0,h}\|_X \left(x_{0,h} - I^+ x_{0,h}, x\right)_X$$

und löst mit dieser rechten Seite die duale Aufgabe

$$(L'z)|_{X_h} = j^+|_{X_h},$$

deren Lösung wieder mit $z_{0,h} \in X_h$ bezeichnet wird, so erhält man den a posteriori Fehlerschätzer

$$\begin{aligned}\eta_{D,S}^2 = &\int_S (-\nu\Delta u_{0,h} + \mathrm{D}p_{0,h} - f - \nu\Delta u_{\mathrm{D}})(w_{0,h}^+ - w_{0,h}) \\ &- \int_S (\mathrm{D}\cdot u_{0,h} + \mathrm{D}\cdot u_{\mathrm{D}})(s_{0,h}^+ - s_{0,h}) \\ &+ \frac{1}{2} \sum_{F\le S, F\in\mathcal{F}_\Omega(\mathcal{T})} \int_F [\nu\frac{\partial u_{0,h}}{\partial n} + \nu\frac{\partial u_{\mathrm{D}}}{\partial n}](w_{0,h}^+ - w_{0,h}) \\ &+ \sum_{F\le S, F\in\mathcal{F}_{\mathrm{A}}(\mathcal{T})} \int_F \left(\nu\frac{\partial u_{0,h}}{\partial n} - np_{0,h} - f_{\mathrm{A}} + \nu\frac{\partial u_{\mathrm{D}}}{\partial n}\right)(w_{0,h}^+ - w_{0,h}).\end{aligned} \tag{4.11}$$

Dabei ist $(w_{0,h}^+, s_{0,h}^+)^T = I^+ z_{0,h}$.
In der vorliegenden Literatur ([9], [10]) gibt es keine theoretischen Beweise dafür, daß Fehlerschätzer dieses Typs tatsächlich zuverlässig und effizient sind. Insofern sollten sie als Fehlerindikatoren bezeichnet werden.
In [9, Kap.4, Lemma 4.1] wird lediglich für die Poissongleichung auf $[-1,1]^2$ gezeigt, daß für einen speziellen Schätzer η und seine Modifikation $\tilde{\eta}$, bei der z_0 durch eine höhere Finite-Elemente-Methode approximiert wird, folgendes gilt: Auf einer mit η bis zur Toleranz $\varepsilon > 0$ verfeinerten Triangulierung ist $|\eta(x_{0,h}) - \tilde{\eta}(x_{o,h})| = o(\varepsilon)$ erfüllt. Der Satz wäre deutlich interessanter, wenn man diese Aussage für die mit $\tilde{\eta}$ erzeugte Triangulierung beweisen könnte.

Dennoch legen numerische Simulationen mit DWR-Fehlerschätzern in [10, Kap.8–12] nahe, daß diese tatsächlich zuverlässig und effizient sind. In verschiedenen Experimenten, sogar bei nichtlinearen und zeitabhängigen Differentialgleichungen, zeigt sich außerdem die große Allgemeinheit der DWR-Methode.
Die Charakterisierung von $\eta_{D,S}$ erfolgt in der vorliegenden Arbeit ebenfalls lediglich über die Experimente in Kapitel 6.

Bemerkung 4.7. Man beachte, daß aufgrund der Symmetrie von L im Fall der Stokes-Gleichungen zum Lösen von $L'z = j^+$ nur die rechte Seite neu diskretisiert werden muß. Die Basisdarstellung von L' entspricht der von L. Des weiteren kann der Aufwand beim Lösen der dualen Aufgabe verringert werden, wenn man davon ausgeht, daß die aus ihr gewonnenen Gewichtungsfaktoren für das eigentliche Residuum nicht exakt bekannt sein müssen. Dann kann die duale Aufgabe mit geringerer Genauigkeit gelöst werden, um Rechenzeit zu sparen.

Kapitel 5

Markierungsstrategien

Die Fehlerschätzer aus den vorangehenden Kapiteln liefern zu jedem n-Simplex $S \in \mathcal{T}$ eine reelle Zahl: $\{(S, \eta_S)\}_{S \in \mathcal{T}^{(n)}}$. η_S wird als eine Schätzung des Diskretisierungsfehlers auf S oder auf einer kleinen Umgebung von S angesehen. Das vorliegende Kapitel beschreibt Strategien zur Auswahl von Simplexen in $\mathcal{T}$, auf denen der Diskretisierungsfehler „zu groß“ ist, denn solche Simplexe müssen verfeinert werden. Die entsprechenden Methoden heißen *Markierungsstrategien*.

Bemerkung 5.1. Zur Lösung stationärer Probleme genügt eine Markierungsstrategie, die Simplexe zur Verfeinerung vorsieht: Man startet mit einer groben Triangulierung $\mathcal{T}_0$ und entscheidet in jeder Iteration des adaptiven Zyklus in Abbildung 1, welche n-Simplexe im Schritt $\mathcal{T}_i \curvearrowright \mathcal{T}_{i+1}$ verfeinert werden sollen.
Bei explizit zeitabhängigen Differentialgleichungen ist es notwendig, daß die Markierungsstrategie Simplexe auch so kennzeichnen kann, daß sie beim Übergang von $\mathcal{T}_i$ zu $\mathcal{T}_{i+1}$ entfernt werden. Denn bei instationären Gleichungen kann es vorkommen, daß sich ein lokal stark verfeinerter Bereich wie z. B. eine Schockwelle durch Ω bewegt. Ohne die Möglichkeit, überflüssige Simplexe zu entfernen, erhielte man einen viel zu großen Bereich von Ω, der sehr fein trianguliert wäre. Das würde in numerischen Simulationen zu Speicherplatz- und Effizienzproblemen führen.

In [6] wird heuristisch begründet, daß für die Poisson-Gleichung mit P_1-Elementen unter allen Triangulierungen mit einer gegebenen Anzahl von n-Simplexen auf denjenigen der Diskretisierungsfehler minimal ist, auf denen die Fehlerschätzungen η_S equilibriert sind: $\eta_S = \eta_T$ für alle n-Simplexe $S,\ T \in \mathcal{T}$.
Dörfler zeigt in [18], daß die Strategie aus Abschnitt 5.2 für die Poisson-Gleichung zu einer Reduktion des Fehlers um einen konstanten Faktor in $(0, 1)$ pro Iteration des adaptiven Zyklus führt, wenn die Anfangstriangulierung hinreichend fein ist. Es besteht noch immer die Schwierigkeit festzustellen, wie fein die erste Triangulierung sein muß.
Theoretisch fundierte Markierungsstrategien für allgemeinere Aufgaben als die Poissongleichung sind in der vorliegenden Literatur rar. Die beiden vorigen Ab-

sätze beschreiben im Grunde den Kern der in DROPS vorhandenen Markierungsstrategien, die in den folgenden Abschnitten erläutert werden. Insgesamt gibt es zu Markierungsstrategien viele offene Fragen.

5.1 Schwellenwert-Methode

Dies ist vermutlich das einfachst denkbare Verfahren. Trotzdem erzielt man damit in der Praxis zufriedenstellende Ergebnisse, wie Verfürth in [39] bemerkt.

Algorithmus 5.2 (Schwellenwert-Methode).
Vorbedingung: Schwellenwert $w > 0$, Fehlerschätzung $\{(S, \eta_S)\}_{S \in \mathcal{T}^{(n)}}$.
Nachbedingung: Funktion $M : \mathcal{T}^{(n)} \longrightarrow \{0, 1\}$; $S \in \mathcal{T}^{(n)}$ ist genau dann zu verfeinern, wenn $M(S) = 1$ ist.

1: **for all** $S \in \mathcal{T}^{(n)}$ **do**
2: **if** $\eta_S > w$ **then**
3: $M(S) \leftarrow 1$
4: **else**
5: $M(S) \leftarrow 0$
6: **end if**
7: **end for**

Die Schwierigkeit dieser Methode liegt in der Wahl eines geeigneten Wertes für w. Populär ist z. B.

$$w = \frac{1}{2} \max_{S \in \mathcal{T}^{(n)}} \eta_S.$$

Durch die Angabe eines zusätzlichen Schwellenwertes $w_2 < w$ kann man auf analoge Weise die Vergröberung von Triangulierungen handhaben.

5.2 Fehleranteil-Methode

Dieses Verfahren stammt aus [18]. Es stellt auf folgende (heuristische) Weise eine Verbesserung der Schwellenwert-Methode dar: Man benennt mit

$$E = \sqrt{\sum_{S \in \mathcal{T}^{(n)}} \eta_S^2} \quad \text{bzw.} \quad E_M = \sqrt{\sum_{S \in \mathcal{T}^{(n)}, M(S)=1} \eta_S^2}$$

die Schätzung des Diskretisierungsfehlers bzw. den geschätzten Diskretisierungsfehler auf den zur Verfeinerung markierten n-Simplexen. Anschaulich ist dann plausibel: Nur wenn $E_M \geq \gamma E$ mit einer Konstante $\gamma \in (0, 1)$ gilt, kann man erwarten, daß der Diskretisierungsfehler in jeder Iteration des adaptiven Zyklus um einen konstanten Faktor reduziert wird. Bei der Schwellenwert-Methode ist

es nicht ohne weiteres möglich, etwas über den Anteil $\frac{E_M}{E}$ des „markierten Fehlers“ am Gesamtfehler zu sagen. Der folgende Algorithmus garantiert eine solche Aussage.

Algorithmus 5.3 (Fehleranteil-Methode).

Vorbedingung: $\gamma \in (0,1)$, Fehlerschätzung $\{(S, \eta_S)\}_{S \in \mathcal{T}^{(n)}}$

Nachbedingung: Funktion $M : \mathcal{T}^{(n)} \longrightarrow \{0,1\}$; $S \in \mathcal{T}^{(n)}$ ist genau dann zu verfeinern, wenn $M(S) = 1$ gilt; $\sqrt{E_{M,\mathrm{sq}}} \geq \gamma \sqrt{E_{\mathrm{sq}}}$.

1: $E_{\mathrm{sq}} \leftarrow \sum_{S \in \mathcal{T}^{(n)}} \eta_S^2$
2: $E_{M,\mathrm{sq}} \leftarrow 0$
3: Sei $\big((S, \eta_S)_i\big)_{i=1}^{|\mathcal{T}^{(n)}|}$ die Anordnung von $\{(S, \eta_S)\}_{S \in \mathcal{T}^{(n)}}$ unter der Ordnungsrelation $(S, \eta_S) \geq (T, \eta_T) \iff \eta_S \geq \eta_T$.
4: $i \leftarrow 1$
5: **while** $i \leq |\mathcal{T}^{(n)}|$ und $E_{M,\mathrm{sq}} < \gamma^2 E_{\mathrm{sq}}$ **do**
6: $\quad M(S_i) \leftarrow 1$
7: $\quad E_{M,\mathrm{sq}} \leftarrow E_{M,\mathrm{sq}} + \eta_{S,i}^2$
8: $\quad i \leftarrow i + 1$
9: **end while**
10: **while** $i \leq |\mathcal{T}^{(n)}|$ **do**
11: $\quad M(S_i) \leftarrow 0$
12: $\quad i \leftarrow i + 1$
13: **end while**

In DROPS wird die Fehleranteil-Methode als das Standardverfahren eingesetzt, allerdings steht auch die Schwellenwert-Methode zur Verfügung.
Ein weiteres in DROPS implementiertes Verfahren, bei dem versucht wird, den Fehler gleichmäßig auf Ω zu verteilen, erwies sich bei Experimenten mit der Poisson-Gleichung als wenig geeignet. Es hängt sehr empfindlich von seinen Parametern ab und tendiert zur Erzeugung uniformer Verfeinerungen. Eine Beschreibung findet sich in [25].
Will man erreichen, daß der adaptive Zyklus mit möglichst wenigen Verfeinerungsschritten auskommt, kann man aufwendigere Markierungsstrategien verfolgen. Es wird auf [39, Kapitel 4.1] verwiesen.

Kapitel 6

Numerische Experimente

In diesem Kapitel werden die Ergebnisse einiger numerischer Experimente mit dem Finite-Elemente-Paket DROPS präsentiert. Es wird am Lehrstuhl für Mathematik des Instituts für Geometrie und praktische Mathematik an der RWTH-Aachen entwickelt. Eine Übersicht der implementierten Algorithmen findet man in [25] und den darin enthaltenen Literaturhinweisen. Inzwischen sind Teile von DROPS auch parallelisiert worden. Das heißt, daß numerische Experimente auf einem Mehrprozessor-Computer durchgeführt werden können (siehe [24]).
Die folgenden Experimente werden mit der seriellen Version von DROPS auf einem Personalcomputer mit einem $1,2$GHz Athlon-Prozessor und 768MB Arbeitsspeicher durchgeführt. Als Betriebssystem wird SuSe-Linux 8.1 im Mehrbenutzermodus verwendet. Allerdings ist das System – abgesehen von den üblichen Hintergrundprozessen – ruhig. DROPS wird mit GCC-3.2 (-W -Wall -pedantic -O2 -funroll-loops -march=athlon -fomit-frame-pointer -finline-limit=2000) compiliert.
Falls nicht anders vermerkt, sind alle Zeiten in den folgenden Tabellen in Sekunden angegeben; es handelt sich stets um das arithmetische Mittel.

6.1 Zuverlässigkeit und Effizienz

Um die Zuverlässigkeit und Effizienz der Fehlerschätzer im Sinne der Definition auf Seite 55 zu untersuchen, wird eine Aufgabe mit bekannter Lösung benötigt, damit der tatsächliche Diskretisierungsfehler bestimmt werden kann. Hier wird Aufgabe 1.30 auf $\Omega = [0,\ \frac{\pi}{4}]^3$ mit $\partial\Omega = \Gamma_{\mathrm{D}}$ und $\nu = 1$ betrachtet. Die Funktionen

$$u(x,y,z) = \frac{1}{3}\begin{pmatrix} \sin(x)\sin(y)\sin(z) \\ -\cos(x)\cos(y)\sin(z) \\ 2\,\cos(x)\sin(y)\cos(z) \end{pmatrix}, \tag{6.1}$$

$$p(x,y,z) = \cos(x)\sin(y)\sin(z) + C \tag{6.2}$$

lösen mit den Daten

$$f(x,y,z) = \begin{pmatrix} 0 \\ 0 \\ 3\,\cos(x)\sin(y)\cos(z) \end{pmatrix}, \quad g \equiv 0 \tag{6.3}$$

die inhomogene Aufgabe 1.30, wenn als Dirichlet-Randbedingung $u_D = u|_{\partial\Omega}$ verwendet wird und die Konstante C so gewählt wird, daß $\int_\Omega p = 0$ ist. Offensichtlich gilt $(u,p)^T \in C^\infty(\Omega)$.
Zur Diskretisierung werden $\mathcal{P}_2\mathcal{P}_1$-Elemente auf den Triangulierungen verwendet, die durch *reguläre Verfeinerung*[1] aus der wie folgt konstruierten Ausgangstriangulierung $\mathcal{T}_0$ entstehen: Man zerlege die Seitenflächen von Ω durch Einfügen je einer Flächendiagonale in Dreiecke und verbinde die Eckpunkte jedes Dreiecks mit dem Schwerpunkt von Ω. Diese Triangulierung enthält 12 Tetraeder.

Bemerkung 6.1. Im Gegensatz zur Standardzerlegung des Würfels nach Kuhn (siehe [12]), die in DROPS sonst verwendet wird, erfüllt die soeben beschriebene Triangulierung die Bedingung 2.30. Deshalb sind die $\mathcal{P}_2\mathcal{P}_1$-Elemente auf $\mathcal{T}_0$ und auf allen von DROPS erzeugten Verfeinerungen von $\mathcal{T}_0$ nach Folgerung 2.37 und Satz 2.40 unabhängig von der Gitterweite h stabil.

Die Lösung des linearen Gleichungssystems (2.20) wird mit dem inexakten Uzawa-Verfahren aus Algorithmus 2.45 bestimmt. Der Vorkonditionierer für $\mathbf{A}$ besteht aus der Anwendung einer kleinen Zahl SSOR-vorkonditionierter CG-Schritte auf $\mathbf{Av} = \mathbf{f} - \mathbf{Av}_i - \mathbf{B}^T\mathbf{p}_i$ mit dem Startvektor $\mathbf{0}$. Die Anwendung einiger CG-Schritte auf $\mathbf{Mq} = \mathbf{Bv}_{i+1} - \mathbf{g}$ ergibt die Vorkonditionierung für $\mathbf{S}$. $\mathbf{M}$ ist die Massenmatrix. Als Startvektor wird ebenfalls $\mathbf{0}$ benutzt. Die Uzawa-Iteration wird beendet, sobald die euklidische Norm des Residuums 10^{-7} unterschreitet.

Im folgenden werden der Residuumsschätzer $\eta_{R,S}$ aus Satz 3.16 und seine Modifikation zur Schätzung der $L_2(\Omega) \times H^{-1}(\Omega)$-Norm aus Satz 3.19 getestet. Um die Fehlerschätzer unabhängig von den anderen Komponenten des adaptiven Zyklus in Abbildung 1 untersuchen zu können, wird in jeder Iteration uniform verfeinert. Das bedeutet, daß alle Tetraeder der feinsten Triangulierung nach der regulären Verfeinerungsregel (siehe Abbildung 6.1) in 8 „Kinder“ zerlegt werden.
Tabelle 6.1 enthält zu jeder Iteration i Informationen über die Triangulierung, auf der die Aufgabe gelöst wird: i, F_v, F_p, AT und h sind die Anzahl der Iterationen des adaptiven Zyklus, die Anzahl der Geschwindigkeits- bzw. Druck-Freiheitsgrade, die Anzahl der Tetraeder in der feinsten Triangulierung sowie deren Gitterweite.
Die Zahl der Freiheitsgrade wächst wie erwartet in geometrischer Progression, die Gitterweite wird in jedem Schritt halbiert. Auf der feinsten Triangulierung belegt DROPS etwa 320MB Speicher, so daß eine weitere Verfeinerung mit den gegebenen Ressourcen nicht erreicht werden kann. Dennoch sind noch circa 55 Prozent

[1] siehe [11]

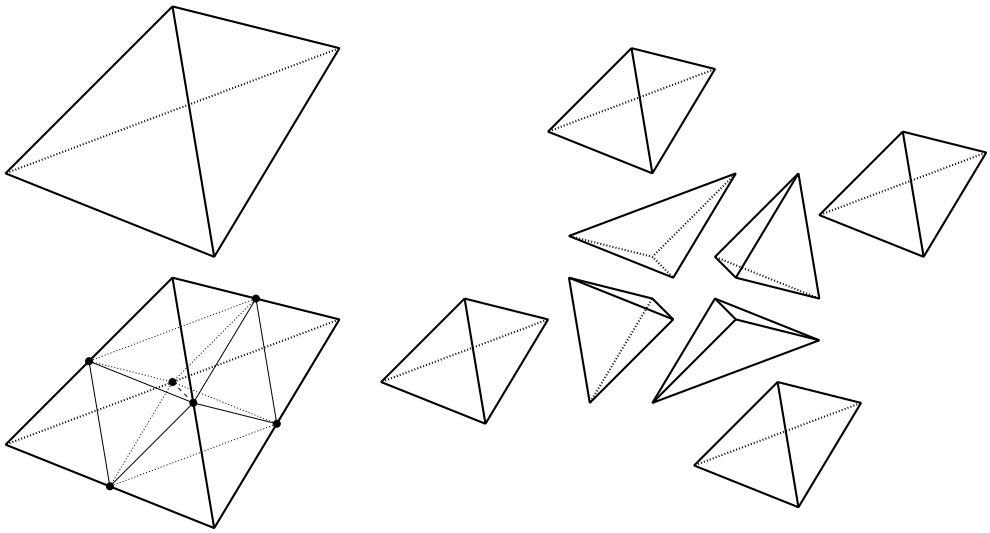

Abbildung 6.1: Reguläre Verfeinerung eines Tetraeders

i	F_v	F_p	AT	h
1	27	9	12	$1,11$
2	273	35	96	$0,555$
3	2565	189	768	$0,278$
4	22413	1241	6144	$0,139$
5	187677	9009	49152	$0,0694$

Tabelle 6.1: Triangulierungen bei uniformer Verfeinerung

des Arbeitsspeichers frei. Diese für uniforme Verfeinerung typische, schlechte Ausnutzung des Arbeitsspeichers ist neben der erhöhten Genauigkeit beim adaptiven Lösen von Differentialgleichungen ein weiterer Anreiz für die Verwendung adaptiver Verfahren.

In Tabelle 6.2 wird das Zeitverhalten der Komponenten des adaptiven Zyklus dargestellt. Es sind: i die Anzahl der Iterationen des adaptiven Zyklus, t_{ges} die Gesamtdauer von Schritt i, bei der allerdings von t_1 und t_2 nur t_1 berücksichtigt wird, t_v die Laufzeit des Verfeinerungsalgorithmus, t_d die Dauer der Diskretisierung, t_l die Laufzeit des Lösers, t_1 die Dauer der Berechnung aller $\eta_{R,S}$ und t_2 die Zeit zum Schätzen der $L_2(\Omega) \times H^{-1}(\Omega)$-Norm.
Man sieht, daß der Zeitbedarf des Uzawa-Lösers alle anderen Schritte des adaptiven Zyklus dominiert und daß der Verfeinerungsalgorithmus hier praktisch keine Rolle für die Gesamtdauer des adaptiven Zyklus spielt. Die beiden Residuumsschätzer demonstrieren einen moderaten Zeitbedarf. Er liegt in derselben Größenordnung wie die Dauer der Diskretisierung.

i	t_{ges}	t_v	t_d	t_l	t_1	t_2
1	0,01	0,00	0,01	0,00	0,00	0,00
2	0,10	0,00	0,01	0,07	0.02	0,03
3	3,25	0,00	0,12	3,03	0,10	0,19
4	39,6	0.03	1,14	37,6	0,78	1,53
5	545	0.20	9,72	529	6,14	12,2

Tabelle 6.2: Zeitverhalten des adaptiven Zyklus

i	η_X	η_Y	e_X	e_u	e_p	I_X	$\tilde{I}_Y$
1	0,229	0,255	0,0375	0,00134	0,0264	6,11	12,2
2	0,0568	0,0308	0,00767	0,000162	0,00467	7,41	16,7
3	0.0147	0.00398	0,00181	$1,81e{-}5$	0,00102	8,10	19,8
4	0.00377	0.000510	0,000444	$2,02e{-}6$	0,000244	8,51	21,2
5	0.000954	$6.47e{-}5$	0,000111	$2,59e{-}7$	$6,02e{-}5$	8,59	21,8

Tabelle 6.3: Effizienz der Residuumsschätzer

Die Fehlerschätzungen werden in Tabelle 6.3 gezeigt. i ist Anzahl der Iterationen des adaptiven Zyklus, η_X und η_Y sind die mit dem Residuumsschätzer berechneten globalen Schätzungen des Diskretisierungsfehlers in der $H^1(\Omega) \times L_2(\Omega)$-Norm gemäß Satz 3.16 und in der $L_2(\Omega) \times H^{-1}(\Omega)$-Norm nach Satz 3.19. e_X ist der Diskretisierungsfehler in der $H^1(\Omega) \times L_2(\Omega)$-Norm, e_u und e_p sind die Diskretisierungsfehler der Geschwindigkeitskomponenten und des Drucks in der $L_2(\Omega)$-Norm. Die letzten zwei Spalten enthalten die Effizienzindizes

$$I_X = \frac{\eta_X}{e_X}, \quad \tilde{I}_Y = \frac{\eta_Y}{(e_u^2 + h^2 e_p^2)^{\frac{1}{2}}}.$$

Man beachte, daß im Nenner von $\tilde{I}_Y$ nicht die $H^{-1}(\Omega)$-Norm des Drucks sondern eine Approximation auftritt.

Wie bei einer glatten Lösung zu erwarten ist, erhält man mit $\mathcal{P}_2\mathcal{P}_1$-Elementen quadratische Konvergenz in der Norm von X und kubische bzw. quadratische Konvergenz in der $L_2(\Omega)$-Norm der Geschwindigkeit bzw. des Drucks. Der Effizienzindex stabilisiert sich für beide Schätzer ab $i = 3$ in der Größenordnung von 10 bzw. 20; auf gröberen Triangulierungen darf der Einfluß der Fehlerterme in Satz 3.16 und 3.19 nicht vernachlässigt werden.

Der $L_2(\Omega) \times H^{-1}(\Omega)$-Schätzer weist einen ungünstigeren Effizienzindex als $\eta_{R,S}$ auf. Das legt der Beweis von Satz 3.19 nahe, weil im Vergleich zu Satz 3.16 einmal mehr interpoliert wird.

ν	η_X	e_X	I_X	AU
$0,5$	$0,00378$	$0,000781$	$4,84$	873
$1,0$	$0,00377$	$0,000444$	$8,49$	81
$2,0$	$0,00377$	$0,000307$	$12,3$	147
$4,0$	0.00377	$0,000263$	$14,3$	254

Tabelle 6.4: Kondition und Effizienz

6.1.1 Kondition von L und Effizienzindex

Aufgrund von Satz 3.1 gehen $\|L\|$ und $\|L^{-1}\|$ in den Effizienzindex von $\eta_{R,S}$ ein. Dies wird demonstriert, indem in der Aufgabe aus Abschnitt 6.1 die kinematische Zähigkeit ν variiert wird. Das folgende Experiment wird auf der Triangulierung mit $i = 4$ aus Tabelle 6.1 durchgeführt.

In Tabelle 6.4 stehen die kinematische Zähigkeit ν, der mittels $\eta_{R,S}$ geschätzte Diskretisierungsfehler e_X in der $H^1(\Omega) \times L_2(\Omega)$-Norm und der Effizienzindex I_X. Zusätzlich wird die Anzahl AU der Iterationen genannt, die der Uzawa-Löser zur Reduktion des Residuums bis unter 10^{-7} benötigt.

Man sieht deutlich, daß sich der Effizienzindex mit zunehmender Zähigkeit verschlechtert. Ebenso läßt die Konvergenzgeschwindigkeit des Uzawa-Lösers erheblich nach.

Daß η_X konstant ist, erklärt sich darüber, wie sich die Lösungen der Stokes-Gleichungen mit den Daten (6.3) verhalten. Sei nun $(u_1, p)^T$ die Lösung der Stokes-Gleichungen (6.1). Dann ist für beliebiges $\nu > 0$ mit $u_\nu = \nu^{-1} u_1$ durch $(u_\nu, p)^T$ ebenfalls eine Lösung der Stokes-Gleichungen mit den rechten Seiten (6.3) gegeben. (Natürlich sind die Randwerte mit ν^{-1} skaliert.) L ist auf der Menge $M = \left\{(u_\nu, p)^T \mid \nu > 0\right\}$ konstant. In $\eta_{R,S}$ tritt nun bis auf die Approximation von Daten die starke Form von L auf (siehe (3.13)), so daß sich die Invarianz von L auf $\eta_{R,S}$ überträgt. Da das Residuum aller Lösungen aus M dasselbe ist, liegt es nahe, daß alle Fehlerschätzer auf der Basis von Satz 3.1 eine ähnliche Invarianzeigenschaft besitzen.

Aus dem Verhalten von AU kann man schließen, daß die Massenmatrix nur in der Nähe von $\nu = 1$ ein guter Vorkonditionierer für das Schur-Komplement $\mathbf{S}$ ist. Da $\mathbf{A}$ die kinematische Viskosität ν als skalaren Faktor enthält, kann mit dem skalierten Vorkonditionierer νM die Konvergenz der Uzawa-Iteration verbessert werden. Man erhält mit ihm die Iterationszahlen 81, 81, 81, 83 anstelle der Spalte „AU“ in Tabelle 6.4, ohne daß sich die Werte in den anderen Spalten ändern. Dies ist eine deutliche Verbesserung.

Abbildung 6.2 zeigt das Geschwindigkeitsfeld der Lösung in der $\{x = \frac{\pi}{4}\}$-Ebene mit Blick in Richtung von $(-1, 0, 0)^T$.

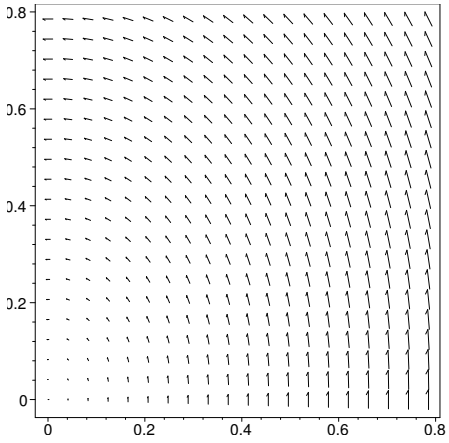

Abbildung 6.2: Geschwindigkeitsfeld in der $\{x = \frac{\pi}{4}\}$-Ebene

6.2 Adaptivität

A posteriori Fehlerschätzer bilden in der Numerik die Grundlage für die adaptiven Markierungsstrategien aus Kapitel 5. Nachdem im vorigen Abschnitt die Eigenschaften der Residuumsschätzer aus Kapitel 3 untersucht worden sind, wird jetzt der Einfluß der Markierungsstrategie auf den adaptiven Zyklus in Abbildung 1 dargestellt. Dazu wird die Stokes-Aufgabe

$$\begin{aligned} -\Delta u + \mathrm{D}p &= 0 \qquad \text{in } \Omega, \\ \mathrm{D}\cdot u &= 0 \qquad \text{in } \Omega, \\ u &= u_{\mathrm{D}} \quad \text{auf } \partial\Omega \end{aligned} \tag{6.4}$$

auf dem L-Gebiet $\Omega = (-1,1)^3 \setminus \big([0,1]\times[-1,0]\times[0,1]\big)$ numerisch gelöst (siehe Abbildung 6.3). Die Funktionen

$$u = r^{\alpha}\begin{pmatrix} (1+\alpha)\sin(\phi)\psi(\phi) + \cos(\phi)\frac{\partial\psi}{\partial\phi}(\phi) \\ \sin(\phi)\frac{\partial\psi}{\partial\phi}(\phi) - (1+\alpha)\cos(\phi)\psi(\phi) \\ 0 \end{pmatrix}$$

$$p = -r^{\alpha-1}(1-\alpha)^{-1}\big((1+\alpha)^2\frac{\partial\psi}{\partial\phi}(\phi) + \frac{\partial^3\psi}{\phi^3}(\phi)\big)$$

mit

$$\begin{aligned} \psi(\phi) = (1+\alpha)^{-1}\sin((1+\alpha)\phi)\cos(\frac{3\pi}{2}\alpha) - \cos((1+\alpha)\phi) \\ - (1-\alpha)^{-1}\sin((1-\alpha)\phi)\cos(\frac{3\pi}{2}\alpha) + \cos((1-\alpha)\phi), \end{aligned}$$

$$\alpha = \frac{856399}{1572864}$$

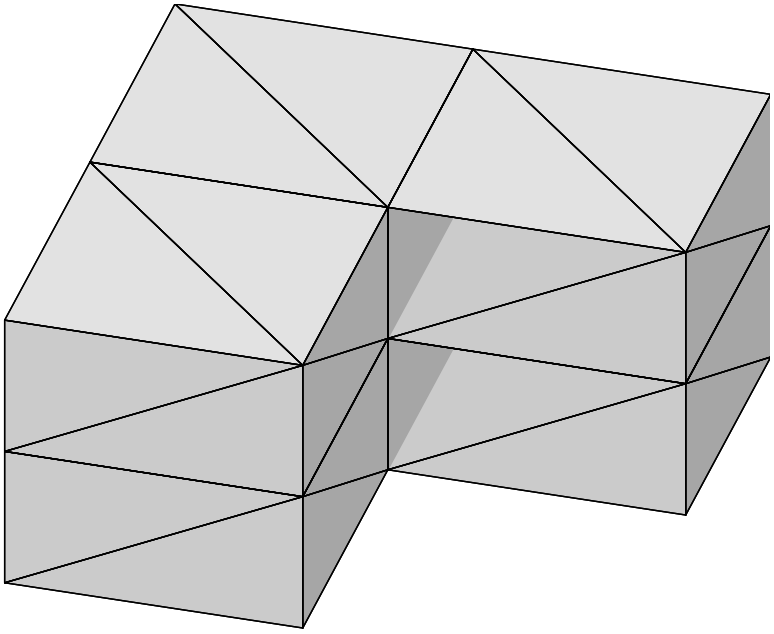

Abbildung 6.3: Rechengebiet für Abschnitt 6.2

lösen (6.4) im schwachen Sinne, wenn man $u_D = u|_{\partial\Omega}$ wählt. r, ϕ, z sind die Zylinderkoordinaten mit

$$x = r\cos(\phi), \quad y = r\sin(\phi), \quad z = z.$$

In den Abbildungen 6.4 und 6.5 sind u und p über der $\{z = \frac{1}{2}\}$-Ebene dargestellt. Auf der Innenseite

$$\Gamma_0 = \{0\} \times [-1, 0] \times [0, 1] \cup [0, 1] \times \{0\} \times [0, 1]$$

von Ω verschwindet u; dort liegen also homogene Randbedingungen vor. Obwohl $p \in L_2(\Omega)$ gilt, hat der Druck eine Singularität auf der z-Achse.[2] Deshalb ist die Verwendung eines adaptiven Verfahrens zur numerischen Lösung der vorliegenden Aufgabe angebracht.

Bemerkung 6.2. Obwohl die theoretische Lösung der Aufgabe bekannt ist, kann der Diskretisierungsfehler des Drucks mit DROPS nicht berechnet werden, weil die zur Verfügung stehenden Quadraturformeln in der Nähe der Singularität versagen.

Die Anfangstriangulierung von Ω wird aus 6 Quadern erzeugt, die dann einzeln in je 6 Tetraeder zerlegt werden so daß insgesamt eine konsistent numerierte Triangulierung entsteht. Wie in Abschnitt 6.1 werden zur Diskretisierung $\mathcal{P}_2\mathcal{P}_1$-Elemente verwendet. Das daraus entstehende lineare Gleichungssystem wird mit

[2] Seine $L_2(\Omega)$-Norm kann in Zylinderkoordinaten gut abgeschätzt werden.

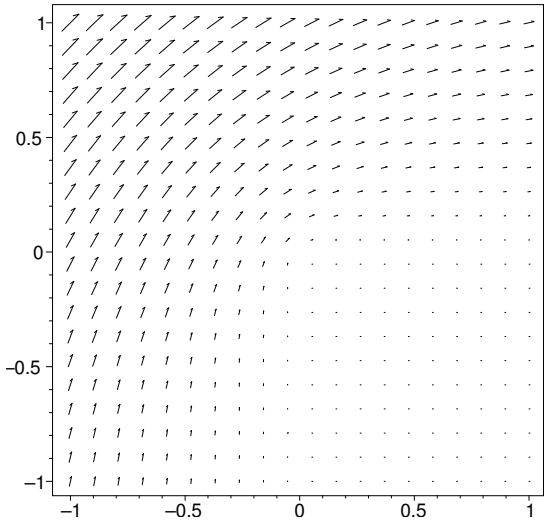

Abbildung 6.4: Exaktes Geschwindigkeitsfeld in der $\{z = \frac{1}{2}\}$-Ebene

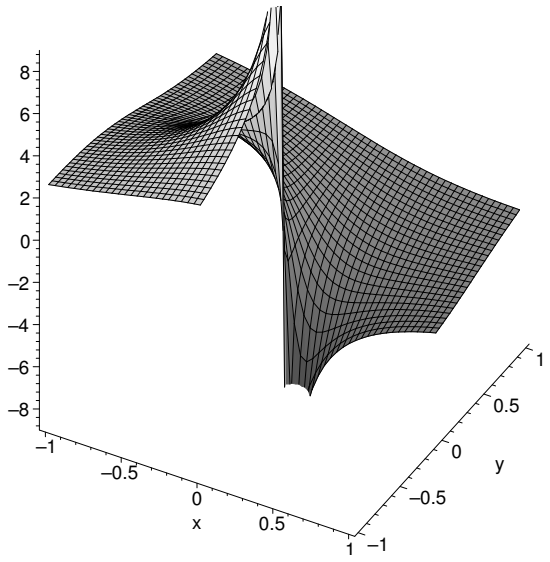

Abbildung 6.5: Exakter Druck in der $\{z = \frac{1}{2}\}$-Ebene

dem inexakten Uzawa-Verfahren gelöst. Entsprechend der Bemerkungen über die Vorkonditionierung mit der Massenmatrix am Ende des vorherigen Abschnitts wird in einer kleinen Testrechnung der Skalierungsfaktor $\tau = 1,6$ für M ermittelt. Die Fehleranteil-Methode aus Kapitel 5 wird nun für verschiedene Wahlen ihres Parameters γ untersucht; je größer $\gamma \in [0, 1]$ ist, desto mehr Tetraeder werden zur Verfeinerung markiert.

Als Vergleichsexperiment dient eine Rechnung mit uniformer Verfeinerung, deren Resultat in Tabelle 6.5 dargestellt ist. Es sind: i die Zahl der Iterationen des adaptiven Zyklus, L das Level der feinsten Triangulierung[3], AT die Anzahl der Tetraeder in der feinsten Triangulierung, F_u und F_p die Anzahl der Freiheitsgrade für Geschwindigkeit und Druck, t_g die Gesamtdauer des adaptiven Zyklus, t_l die zum Lösen des größten Gleichungssystems aufgewendete Zeit, t_s der Zeitaufwand des Fehlerschätzers auf der feinsten Triangulierung, inklusive der Dauer der Markierungsstrategie, e_u der $L_2(\Omega)$-Diskretisierungsfehler der Geschwindigkeit und η_X die Schätzung des Diskretisierungsfehlers in der $H^1(\Omega) \times L_2(\Omega)$-Norm.

Während der Rechnung belegt DROPS 106MB Arbeitsspeicher, die nächst feinere, uniforme Triangulierung führt während der Diskretisierung zu einer Out-of-Memory-Situation, so daß DROPS terminiert wird. Es ergibt sich also eine schlechte Ausnutzung der Möglichkeiten des Rechners.

Tabelle 6.6 enthält die Ergebnisse des adaptiven Experiments. Es werden dieselben Daten wie in Tabelle 6.5 gezeigt; die zusätzliche Spalte γ enthält den

[3]Ein Tetraeder ist in Level L, wenn er durch genau Lfaches Verfeinern eines Tetraeders der Anfangstriangulierung entstanden ist, oder wenn er durch weniger als L Verfeinerungen eines Tetraeders aus der Anfangstriangulierung entstanden ist und selber nicht verfeinert ist.

i	L	AT	F_u	F_p	t_g	t_l	t_s	e_u	η_x
4	3	18432	65565	3825	353	309	2,57	0,0179	4,12

Tabelle 6.5: Uniforme Vergleichsrechnung

γ	i	L	AT	F_u	F_p	t_g	t_l	t_s	e_u	η_x
0,4	45	5	4263	14703	949	590	34,0	0,56	0,0104	4,08
0,6	17	5	5933	16356	1278	249	49,8	0,76	0,00936	3,81
0,8	9	5	7316	25725	1576	176	69,4	0,94	0,00864	3,86
0,8	11	7	38709	142374	7667	1965	1114	4,77	0,00198	1,80

Tabelle 6.6: Adaptives Experiment

Parameter der Fehleranteil-Methode. Als Abbruchkriterium dient für die ersten drei Experimente das Unterschreiten der Fehlerschätzung von 4,12 aus der uniformen Testrechnung, so daß die Zeilen mit Tabelle 6.5 verglichen werden können. Es zeichnen sich zwei Optimierungsziele bei der Wahl von γ ab. Soll die Anzahl der Freiheitsgrade minimiert werden, muß γ klein gewählt werden; der adaptive Zyklus wird dann sehr oft mit stets nur kleinen Modifikationen der Triangulierung ausgeführt, so daß im Fall $\gamma = 0,4$ etwa 23 Prozent der Tetraeder und Freiheitsgrade des uniformen Testlaufs ausreichen, um dieselbe Genauigkeit zu erreichen. Soll die Gesamtlaufzeit des adaptiven Zyklus minimiert werden, ist γ größer zu wählen; allerdings sind die erzeugten Triangulierungen dann größer als erforderlich.
Die letzte Zeile von Tabelle 6.5 zeigt ein Experiment, bei dem mit $\gamma = 0,8$ weiter verfeinert wird. Die Rechnung wird beendet, weil die Konvergenzrate des Uzawa-Verfahrens nachläßt. Die letzte Iteration des adaptiven Zyklus belegt 244MB des Arbeitsspeichers, so daß von der Problemgröße her gesehen, noch ein oder zwei weitere Iterationen möglich wären.

Bemerkung 6.3 *(Singularitäten).* Das vorliegende Testproblem weist eine eindimensionale Singularität auf, es ist sogar in z-Richtung konstant. Bei punktförmigen Singularitäten sind noch größere Unterschiede in der Anzahl der benötigten Freiheitsgrade zwischen uniformer und adaptiver Rechnung zu erwarten. Das gilt auch für den Fall einer „Beinahe-Singularität“ (großer, aber prinzipiell beschränkter Gradient der Lösung), wenn diese auf der adaptiv verfeinerten Triangulierung aufgelöst werden kann.

Die Triangulierung zur letzten Zeile von Tabelle 6.6 wird in Abbildung 6.6 gezeigt; Abbildung 6.7 zeigt das numerisch ermittelte Geschwindigkeitsfeld auf einer groben Triangulierung. Es ist in beiden Fällen der Schnitt mit der $\{z = \frac{1}{2}\}$-Ebene zu sehen.

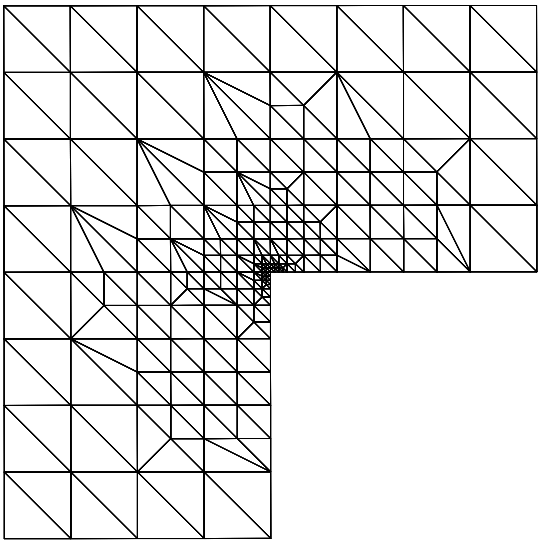

Abbildung 6.6: Adaptive Triangulierung in der $\{z = \frac{1}{2}\}$-Ebene

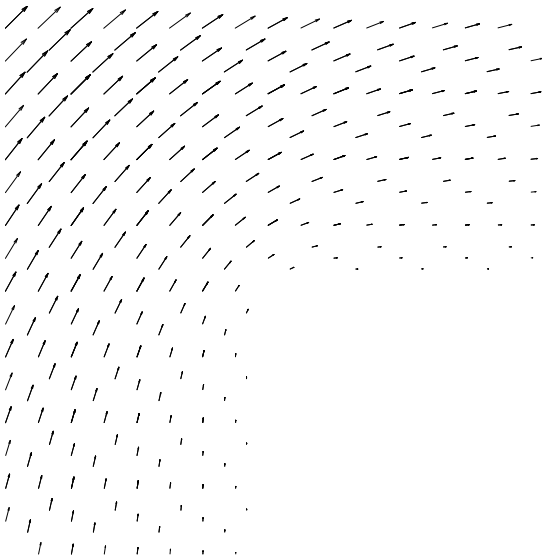

Abbildung 6.7: Geschwindigkeitsfeld in der $\{z = \frac{1}{2}\}$-Ebene

6.3 Driven-Cavity

Physikalisch gesehen, beschreibt das Driven-Cavity-Problem die Bewegung einer Flüssigkeit in einem würfelförmigen Becken, über das ein Deckel tangential abgezogen wird. Er treibt das Fluid an. Bei kleinen Reynoldszahlen können die vollen Bewegungsgleichungen der Flüssigkeit durch die Stokes-Gleichungen angenähert werden.
Sei $\Omega = (0,\,1)^3$ der Einheitswürfel mit dem „Deckel“ $\Sigma_0 = (0,\,1)^2 \times \{1\}$ in der $\{z = 1\}$-Ebene. Dann wird das Driven-Cavity-Problem durch die inhomogene Stokes-Randwertaufgabe 1.30 mit den Daten

$$f \equiv \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad g \equiv 0, \quad u_{\mathrm{D}} = \begin{cases} 0, & \text{auf } \partial\Omega \setminus \Sigma_0, \\ \phi & \text{auf } \Sigma_0 \end{cases}$$

mit

$$\phi(x,y,z) = \begin{cases} (1,0,0)^T, & \text{für } 0,1 \le x,y \le 0,9 \\ \frac{0{,}5-\max\{|x-0{,}5|,|y-0{,}5|\}}{0{,}1}(1,0,0)^T & \text{sonst} \end{cases}$$

mathematisch beschrieben. Die geglätteten Randbedingungen auf Σ_0 sind notwendig, weil u_{D} sonst nicht in $H^{\frac{1}{2}}(\partial\Omega)$ liegt. In numerischen Simulationen würde sich die Randbedingung ohne Glättung durch unendlich hohen Druck an der Ein- und Ausströmungskante von Ω äußern. Die Anfangstriangulierung besteht aus $2\times2\times2$ Teilwürfeln, die durch das Einfügen einer Raumdiagonale trianguliert

i	t_d	t_l	t_s	AU	F_u	F_p	η_X
1	$0,01$	$0,16$	$0,01$	463	81	27	$51,77$
2	$0,01$	$0,46$	$0,03$	360	243	48	$48,57$
3	$0,04$	$1,04$	$0,06$	348	546	86	$32,84$
4	$0,11$	$3,55$	$0,14$	314	1335	168	$25,11$
5	$0,19$	$6,57$	$0,22$	271	2247	246	$17,09$
6	$0,50$	$17,9$	$0,53$	228	5430	535	$10,11$
7	$1,60$	$52,0$	$1,49$	193	16335	1338	$6,131$
8	$3,35$	$89,5$	$3,02$	158	33129	2637	$3,782$
9	$8,14$	229	$7,03$	162	78903	5823	$2,546$
10	$18,5$	1840	$14,6$	548	170760	11076	$1,562$

Tabelle 6.7: Driven-Cavity adaptiv

werden. Auf den Triangulierungen werden $\mathcal{P}_2\mathcal{P}_1$-Elemente benutzt. Als iterativer Löser kommt wieder das Uzawa-Verfahren mit denselben Parametern wie in Abschnitt 6.1 zum Einsatz, so daß zusammen mit dem Residuums-Fehlerschätzer $\eta_{R,S}$ und der Fehleranteil-Strategie aus 5.2 alle Komponenten des adaptiven Zyklus beschrieben sind. Der Anteil γ des zu markierenden Fehlers (siehe 5.3) wird auf $0,8$ gesetzt.

Damit erhält man die Resultate[4] in Tabelle 6.7. Mit i werden die Schritte des adaptiven Zyklus gezählt. Die nächsten drei Spalten enthalten die Zeiten, welche zur Diskretisierung, zum Lösen und zum Schätzen des Fehlers benötigt werden. Die Anzahl der Uzawa-Iterationen steht in Spalte AU; F_u und F_p sind die Anzahlen der Geschwindigkeits- und Druck- Freiheitsgrade. η_X schließlich ist die globale $H^1(\Omega) \times L_2(\Omega)$-Schätzung des Fehlers.

Die Fehleranteil-Strategie mit $\gamma = 0,8$ führt in diesem Beispiel zu einer kontinuierlichen Verringerung des Fehlers. Aufgrund der hohen Iterationszahlen wird die meiste Zeit des adaptiven Zyklus vom Uzawa-Löser benötigt. Der Fehlerschätzer beansprucht etwa so viel Zeit wie das Diskretisieren, was als moderat angesehen wird.

Hingegen ist es angebracht, die Uzawa-Methode zu beschleunigen. Hierzu kommen das CG-Verfahren für Sattelpunktaufgaben gemäß Bemerkung 2.47 und die Verwendung von Interpolanten der Lösung auf der alten Triangulierung als Startvektor in Frage.

Die Abbildungen 6.8 und 6.9 zeigen zwei Schnitte durch die feinste adaptiv erzeugte Triangulierung. In Abbildung 6.10 sind Isolinien des Drucks und das Geschwindigkeitsfeld in der $\{y = \frac{1}{2}\}$-Ebene zu sehen.

[4]Diese Berechnung wird auf einem Computer mit einem 600MHz Pentium III Prozessor durchgeführt.

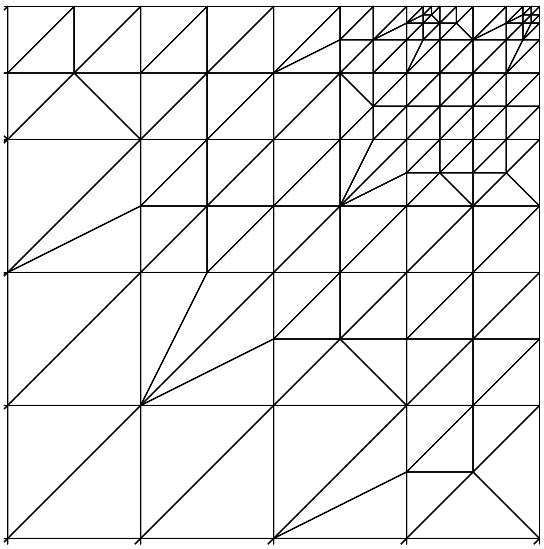

Abbildung 6.8: Driven Cavity: Triangulierung in der $\{y = 0.5\}$-Ebene; Ausschnitt oben rechts

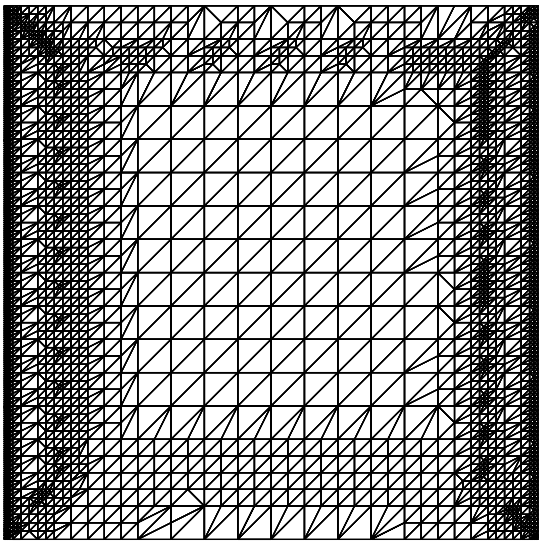

Abbildung 6.9: Driven Cavity: Triangulierung in der $\{z = 1\}$-Ebene.

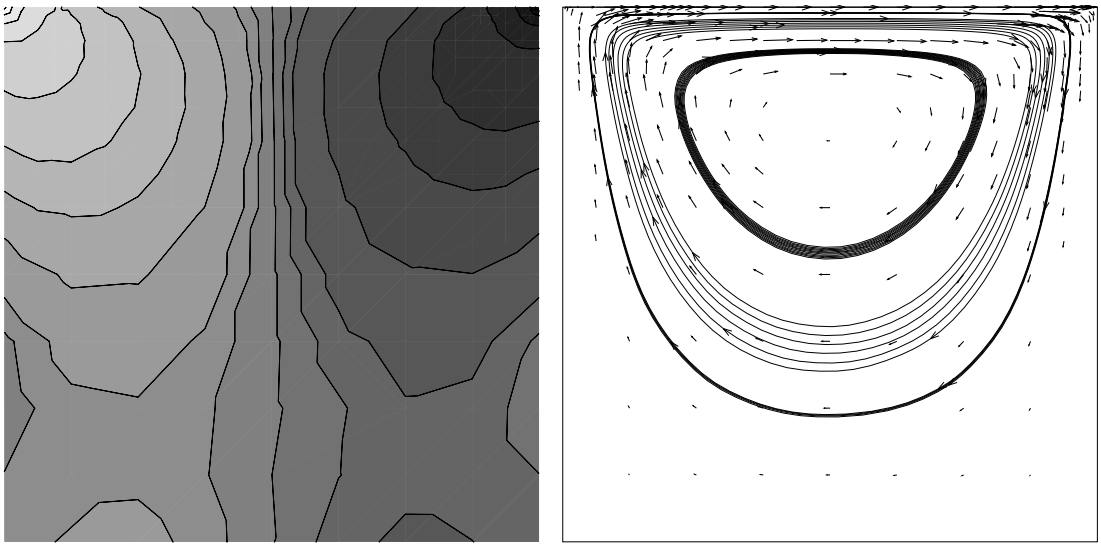

Abbildung 6.10: Driven Cavity: Isobaren und Geschwindigkeitsfeld in der $\{y = 0.5\}$-Ebene.

6.4 Zusammenfassung und Ausblick

In den Kapiteln 1 und 2 wurden die mathematischen Eigenschaften der Stokes-Gleichungen und ihrer Diskretisierung ausführlich analysiert. Zu diesen Themen gibt es keine großen offenen Fragen mehr. Hingegen zeigte sich bei den Experimenten in diesem Kapitel, daß die Lösungsverfahren in DROPS beschleunigt werden müssen. (Die Schur-Komplement-Methode weist eine ähnliche Geschwindigkeit wie das Uzawa-Verfahren auf.) Dazu kommen z. B. die Konzepte aus [44] in Frage.

Die Residuumsschätzer aus Kapitel 3 wurden theoretisch vollständig analysiert und auch für Ausströmungs-Randbedingungen formuliert. Ihre Leistungsfähigkeit stellte sich in den numerischen Experimenten unter Beweis. Es bleibt die Frage, ob die Schätzer auf der Basis lokaler Stokes-Aufgaben, die in Abschnitt 3.3 konstruiert wurden, so modifiziert werden können, daß die lokalen Aufgaben eine deutlich kleinere Dimension erhalten. Dabei soll die Lösbarkeit der lokalen Probleme erhalten bleiben. Das würde sie für eine Implementierung attraktiver erscheinen lassen.

Die in Kapitel 4 vorgestellte DWR-Methode ist ein Kandidat für Fehlerschätzung bei verschiedenen wichtigen Verallgemeinerungen der Aufgabenstellung der vorliegenden Arbeit. So sind etwa im Rahmen des SFB 540 Fehlerschätzungsmethoden für die vollen Navier-Stokes-Gleichungen erforderlich. Außerdem könnte durch die Wahl der richtigen Zielgröße für die DWR-Methode eventuell die Verfolgung der Phasengrenze bei der Simulation mehrphasiger, fluider Systeme mit höherer Genauigkeit erfolgen. Gleichzeitig existieren zur Theorie der DWR-Methode einige offene Fragen. Zum Beispiel gibt es in der vorliegenden Literatur so gut wie keine theoretischen Beweise für die Zuverlässigkeit und Effizienz der DWR-Methode, wenn man statt der exakten dualen Lösung nur eine Approximation zur Verfügung hat.
Leider konnte die bereits begonnene Implementierung des in Abschnitt 4.2.2 beschriebenen DWR-Fehlerschätzers aus Zeitmangel und wegen des Fehlens bestimmter Nachbarschaftsrelationen für Tetraeder in DROPS nicht mehr fertiggestellt werden. Diese Arbeit sollte fortgesetzt und numerisch getestet werden.

Weitere offene Fragen stellen sich in Bezug auf Markierungsstrategien zur Steuerung des adaptiven Zyklus. Die in Kapitel 5 beschriebenen Verfahren sind in der vorliegenden Literatur populär. Dennoch gibt es nicht viele Konvergenzanalysen für den adaptiven Zyklus als ganzes. Auf diesem Gebiet besteht ebenfalls Forschungsbedarf.

Anhang A

Physikalische Grundlagen der Fluiddynamik

Die klassische Hydrodynamik ist ein Teilgebiet der Kontinuumsmechanik. Man geht davon aus, daß Materie praktisch stetig verteilt und homogen in der Struktur ist. Dies schließt nicht aus, daß das Medium, welches das Fluid [1] enthält, anisotrop ist.
Im folgenden werden zunächst die Grundlagen der Kinematik gelegt, dann werden die Navier-Stokes-Gleichungen hergeleitet. Zum Schluß wird das Konzept der dynamischen Ähnlichkeit (Reynoldszahl, Froudezahl) erläutert.

A.1 Kinematik

Die Aufgabe besteht darin, die Bewegung eines Fluids während eines Zeitraumes $[t_0, t_e) \subseteq \mathbb{R}$ zu beschreiben. Da jedes Fluid auf mikroskopischer Ebene aus Atomen oder Molekülen besteht, wird folgender Ansatz gewählt:

Definition A.1. Sei $\Omega_0 \subseteq R^n$ das vom Fluid zur Zeit t_0 okkupierte Gebiet. Dann gibt die stetige Funktion

$$\varphi : \Omega_0 \times [t_0, t_e) \longrightarrow \mathbb{R}^n : (X, t) \longrightarrow \varphi(X, t)$$

die Position des Fluidteilchens, das sich zur Zeit t_0 bei X befindet, zur Zeit t an. $\varphi(\cdot, t)$ sei für jedes t ein Homöomorphismus, der $\varphi(\cdot, t) = id_{\Omega_0}$ erfüllt.
Ferner sei

$$\Omega(t) := \varphi(\Omega_0, t)$$

das von dem Fluid zur Zeit t eingenommene Gebiet und zu $X \in \Omega_0$

$$x(t) := \varphi(X, t) \quad \text{für alle } t \in [t_0, t_e)$$

[1] Sowohl Flüssigkeiten als auch Gase werden hier als Fluid bezeichnet.

die Zeitparametrisierung der Bahn von X. Die folgende Teilmenge von $\mathbb{R}^n \times \mathbb{R}$ wird Raum-Zeitzylinder genannt:

$$C(\Omega_0, t_0, t_e) = \bigcup_{t \in [t_0, t_e)} \Big(\Omega(t) \times \{ t \} \Big)$$

Bemerkung A.2.

- Es gilt $\varphi(\partial\Omega_0, t) = \partial\Omega(t)$, weil φ ein Homöomorphismus zwischen Ω_0 und $\Omega(t)$ ist. Somit kann auf diese Weise kein Phasenübergang wie zum Beispiel das Kochen einer Flüssigkeit beschrieben werden, denn dabei entsteht „neuer Rand“.

- Für den Rest dieses Kapitels wird angenommen, daß φ so glatt ist, daß die auftretenden Ableitungen existieren und stetig sind.

Es gibt zwei häufig benutzte Bezugssystemtypen, um die Ortskoordinaten von Fluidteilchen zu beschreiben:

Definition A.3 (Bezugssysteme).

1. *Nach Euler* Man betrachtet das Fluid von einem $\mathbb{R}^n$ als Inertialsystem aus, das als ruhend gilt. Positionen in diesem Raum werden als Minuskeln geschrieben.

2. *Nach Lagrange* Man wählt zu einem beliebigen $X \in \Omega_0$ das Koordinatensystem, in dem dieser Punkt zu jeder Zeit $t \in [t_0, t_e)$ im Ursprung ruht. Aus Sicht von A.3.1 befindet sich der Ursprung zur Zeit t bei $x = \varphi(X, t)$, was bedeutet, daß in Lagrangeschen Koordinaten Trägheitskräfte auftreten, wenn $\varphi(X, \cdot)$ nicht eine lineare Funktion ist.

Obwohl in der Punktmechanik die Lagrangeschen Bezugssysteme eine große Rolle spielen, wird im weiteren für Ortskoordinaten ein Eulersches Bezugssystem verwendet werden, wenn nichts anderes gesagt wird.
In Kinematik und Dynamik spielen Zeitableitungen eine wichtige Rolle, deshalb werden diese jetzt für obige Bezugssysteme untersucht.

Definition A.4. Sei $f : \{(x,t) \mid t \in [t_0, t_e), x \in \Omega(t)\} \longrightarrow \mathbb{R}$ eine stetig differenzierbare Funktion. Die partielle Zeitableitung

$$\frac{\partial f}{\partial x}(x,t) \equiv \mathrm{D}_t f(x,t)$$

heißt räumliche Ableitung von f; sie beschreibt die zeitliche Änderung an der Stelle x in Eulerschen Koordinaten.

Betrachtet man f via $x = \varphi(X,t)$ als Funktion der Lagrangeschen Koordinate $X \in \Omega_0$, so nennt man die zeitliche Änderung von f bei X materielle Ableitung. Unter Mißbrauch der Notation wird sie als

$$\frac{\mathrm{d}f}{\mathrm{d}t}(x,t) := \frac{\mathrm{d}f}{\mathrm{d}t}(\varphi(X,t),t)$$

geschrieben.

Der Zusammenhang zwischen räumlicher und materieller Ableitung ergibt sich aus der Kettenregel, doch damit er in seiner üblichen Form niedergeschrieben werden kann, benötigt man folgende

Definition A.5 (Geschwindigkeitsfeld). $\varphi(X,t)$ ist offensichtlich die Flußabbildung des Vektorfeldes

$$v(x,t) := \mathrm{D}_t\varphi(X,t) \quad \text{mit} \quad x = \varphi(X,t),$$

welches man das Geschwindigkeitsfeld des Fluids nennt. Es ist wohldefiniert, weil φ in den räumlichen Koordinaten für jedes t ein Homöomorphismus ist.

In Anwendungen ist man fast immer an dem Geschwindigkeitsfeld v und nicht am Fluß φ interessiert. Deshalb wird später nur eine Differentialgleichung für v hergeleitet. Aus v läßt sich natürlich φ durch Lösen der gewöhnlichen Differentialgleichung

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi(X,t) = v(\varphi(X,t),t)$$

für jedes $X \in \Omega_0$ gewinnen.
Aus der Kettenregel folgt somit für die Stelle $x = \varphi(X,t)$

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}t}(x,t) &= \frac{\mathrm{d}f}{\mathrm{d}t}(\varphi(X,t),t)\\ &= \mathrm{D}_t f(x,t) + \mathrm{d}_x f(x,t)\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}(X,t)\right)\\ &= \mathrm{D}_t f(x,t) + \mathrm{d}_x f(x,t)\,(v(x,t))\\ &= \mathrm{D}_t f(x,t) + (v(x,t)\,\mathrm{D}_x) f(x,t),\end{aligned} \tag{A.1}$$

wobei man die letzte Umformung entweder durch eine kurze Rechnung überprüft oder einsieht, daß es egal ist, ob man die volle Linearisierung von f in Richtung v auswertet oder f nur in Richtung v linearisiert.
Der Term $(v\,\mathrm{D}_x)f$ läßt sich als Einfluß der Bewegung oder Konvektion von X in Eulerschen Koordinaten interpretieren.

A.2 Transportsatz

Es wird jetzt untersucht, wie sich das Integral einer Funktion über eine vom Fluß φ transportierte Menge zeitlich ändert. Die Bedeutung dieser Ableitung zeigt sich, wenn man als Funktion die Volumendichte einer physikalischen Größe, die einzelnen Fluidteilchen zugeordnet ist, zum Beispiel Masse oder Impuls, wählt — in Verbindung mit Erhaltungssätzen ergeben sich Bestimmungsgleichungen für die Dichtefunktion.
Ab jetzt sei zu $U_0 \subseteq \Omega_0$ die messbare Menge $U(t) \subseteq \Omega(t)$ analog zu $\Omega(t)$ in Definition A.1 gegeben. Ferner sei φ für jedes $t \in [t_0, t_\mathrm{e})$ stetig differenzierbar und fast überall nicht entartet, das heißt, es gebe zwei positive Zahlen c und C, so daß fast überall $c \leq det\, \mathrm{d}_X\varphi(X,t) \leq C$ erfüllt ist.

Satz A.6 (Transportsatz). *Ist $f : \mathbb{R}^n \times \mathbb{R} \longrightarrow \mathcal{B}$ eine für alle $t \in [t_0, t_e)$ integrierbare Funktion, die stetig differenzierbar ist und $\mathcal{B}$ ein Banachraum, so gilt*

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{U(t)} f(x,t)\,\mathrm{d}x &= \int_{U(t)} \frac{\mathrm{d}}{\mathrm{d}t} f(x,t) + f(x,t)\,\mathrm{D}_x v(x,t)\,\mathrm{d}x \\ &= \int_{U(t)} \mathrm{D}_t f(x,t) + (v\,\mathrm{D}_x) f(x,t) + f(x,t)\,\mathrm{D}_x v\,\mathrm{d}x.\end{aligned}$$

Beweis. Man wendet den Transformationssatz mit $x = \varphi(X,t)$, d. h. $\mathrm{d}x = \det(\mathrm{d}_X\varphi(X,t))\,\mathrm{d}X$, an und differenziert anschließend:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{U(t)} f(x,t)\,\mathrm{d}x &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{U_0} f(\varphi(X,t),t)\det(\mathrm{d}_X\varphi(X,t))\,\mathrm{d}X \\ &= \int_{U_0} \frac{\mathrm{d}}{\mathrm{d}t}\big(f(\varphi(X,t),t)\det(\mathrm{d}_x\varphi(X,t))\big)\,\mathrm{d}X \\ &= \int_{U_0} \frac{\mathrm{d}}{\mathrm{d}t} f(\varphi(X,t),t)\det(\mathrm{d}_x\varphi(X,t)) \\ &\qquad + f(\varphi(X,t),t)\frac{\mathrm{d}}{\mathrm{d}t}\det(\mathrm{d}_X\varphi(X,t))\,\mathrm{d}X \\ &= \int_{U_0} \Big(\frac{\mathrm{d}}{\mathrm{d}t} f(\varphi(X,t),t) + f(\varphi(X,t),t)\,\mathrm{D}_x v(\varphi(X,t),t)\Big) \\ &\qquad \det(\mathrm{d}_X\varphi(X,t))\,\mathrm{d}X.\end{aligned}$$

Der letzte Schritt wird durch Lemma A.7 gerechtfertigt, was die nochmalige Anwendung des Transformationssatzes ermöglicht:

$$= \int_{U(t)} \frac{\mathrm{d}}{\mathrm{d}t} f(x,t) + f(\varphi(X,t),t)\,\mathrm{D}_x v(\varphi(X,t),t)\,\mathrm{d}x$$

Wandelt man die materielle Ableitung nun noch wie in Gleichung A.1 in eine räumliche um, so ergibt sich die zweite Gleichung des Satzes. □

Lemma A.7 (Ableitung von Determinanten).

1. *Ist* $A : \mathbb{R} \longrightarrow GL_n(\mathbb{R})$ *stetig differenzierbar, so gilt*

$$\frac{\mathrm{d}}{\mathrm{d}t} \det A(t) = \operatorname{tr}\left(\frac{\mathrm{d}}{\mathrm{d}t} A(t) A(t)^{-1}\right) \det A(t).$$

2. *Für alle Stellen, an denen* $\mathrm{d}_X \varphi(X,t)$ *regulär ist, gilt*

$$\frac{\mathrm{d}}{\mathrm{d}t} \det\big(\mathrm{d}_X \varphi(X,t)\big) = \mathrm{D}_x v(\varphi(X,t),t) \det\big(\mathrm{d}_X \varphi(X,t)\big).$$

Beweis. Zu 1: Mit

$$\big(a_{i,j}\big)_{i,j=1,\ldots,n} = \big(A_j\big)_{j=1,\ldots,n} = A(t)$$

werden die Einträge bzw. die Spalten von $A(t)$ bezeichnet, mit

$$\big(b_{i,j}\big)_{i,j=1,\ldots,n} = \big(B_j\big)_{j=1,\ldots,n} = A^{-1}(t)$$

die Spalten von $A^{-1}(t)$.
Dann ist bekanntermaßen

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} \det A(t) &= \sum_{j=1}^{n} \det\begin{pmatrix} A_1 & \ldots & \frac{\mathrm{d}}{\mathrm{d}t} A_j & \ldots & A_n \end{pmatrix} \\ &\qquad \text{(Entwicklungssatz von Laplace)} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} (-1)^{i+j} \frac{\mathrm{d}}{\mathrm{d}t} A_j \det A_{\widetilde{i,j}},\end{aligned} \tag{A.2}$$

wobei $A_{\widetilde{i,j}}$ die Teilmatrix von $A(t)$ ist, die durch Streichen der i-ten Zeile und j-ten Spalte entsteht. Die Cramersche Regel liefert

$$b_{j,i} \det A(t) = (-1)^{i+j} \det A_{\widetilde{i,j}},$$

was in A.2 zu

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} \det A(t) &= \det A(t) \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\mathrm{d}}{\mathrm{d}t} a_{i,j} b_{j,i} \\ &= \det A(t) \sum_{i=1}^{n} \left(\frac{\mathrm{d}A}{\mathrm{d}t}(t) A^{-1}\right)_{i,i} = \det A(t) \operatorname{tr}\left(\frac{\mathrm{d}A}{\mathrm{d}t}(t) A(t)^{-1}\right)\end{aligned}$$

führt.
Zu 2: Man verwende in Aussage 1 die Gleichung $A(t) := \mathrm{d}_X \varphi(X,t)$. Es folgt

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} A(t) &= \mathrm{d}_X \left(\frac{\mathrm{d}}{\mathrm{d}t} \varphi(X,t)\right) = \mathrm{d}_X v(\varphi(X,t),t) \\ &\qquad \text{(Kettenregel)} \\ &= \mathrm{d}_x v(\varphi(X,t),t)\, \mathrm{d}_X \varphi(X,t) = \mathrm{d}_x v(\varphi(X,t),t) A(t),\end{aligned}$$

was durch Einsetzen in Aussage 1 Teil 2 des Lemmas bedingt. □

A.3 Maße und Dichten

Bevor der Transportsatz zum ersten Einsatz gelangt, muß noch etwas über sogenannte „set-functions“, also Funktionen, die auf der Potenzmenge einer anderen Menge erklärt sind, gesagt werden, denn in einem Fluid ist zum Beispiel die Eigenschaft Masse für Teilvolumina des Fluids, nicht aber für einzelne Punkte, sinnvoll erklärt.
Das angemessene mathematische Konzept findet sich in dem Begriff des Maßes. Da der Transportsatz Aussagen über Integrale ermöglicht, sollten die auftretenden Maße am besten als Integrale darstellbar sein.

Satz A.8 (Radon und Nikodym). *Sei $(\Omega, \mathcal{S}, \mu)$ ein σ-endlicher Maßraum und ν ein endliches, vorzeichenbehaftetes Maß auf $\mathcal{S}$. Dann sind äquivalent:*

1. $\mu(A) = 0 \implies \nu(A) = 0$ *für alle* $A \in \mathcal{S}$.
2. *Das Maß ν ist bezüglich μ absolut stetig.*
3. *Es existiert eine Funktion $f \in \mathcal{L}_1(\mu)$, so daß*

$$\nu(A) = \int_A f \, \mathrm{d}\nu \quad \text{für alle } A \in \mathcal{S}$$

gilt. f ist ν-fast-überall eindeutig.

Dabei heißt σ-endlich, daß jede Menge $A \in \mathcal{S}$ als höchstens abzählbare Vereinigung von Mengen mit endlichem Maß dargestellt werden kann.
Solange also eine physikalische Größe ν auf „Objekten ohne Volumen“ nur den Wert Null annimmt, kann sie als Integraldichte f dargestellt werden, indem man für μ das Lebesguemaß $\mathcal{L}^n$ des $\mathbb{R}^n$ wählt.

A.4 Kontinuitätsgleichung und Massenerhaltung

In Hinblick auf den vorigen Abschnitt wird angenommen, daß nur endliche Massen betrachtet werden und außerdem

$$\mathcal{L}^n(A) = 0 \implies m(A) = 0 \quad \text{für alle } A \subseteq \mathbb{R}^n,\ A \text{ messbar},$$

gilt. Dabei bezeichnet $m(A)$ die Masse von A. Wegen Satz A.8 existiert die sogenannte Massendichte ρ mit der Eigenschaft

$$m(A) = \int_A \rho(x,t) \, \mathrm{d}x.$$

In der nichtrelativistischen Physik gibt es den

Satz A.9 (Massenerhaltung). *Die Masse des Gebietes $U(t)$ aus Abschnitt A.2 ist zeitlich konstant, also*

$$\frac{\mathrm{d}}{\mathrm{d}t} m(U(t)) = 0 \quad \textit{für alle } t \in [t_0, t_e).$$

Wendet man auf diese Identität den Transportsatz an, so erhält man die Kontinuitätsgleichung in Integralform:

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}t} m(U(t)) &= \frac{\mathrm{d}}{\mathrm{d}t} \int_{U(t)} \rho(x,t)\,\mathrm{d}x \\ &= \int_{U(t)} \mathrm{D}_t\rho(x,t) + (v\,\mathrm{D}_x)\rho(x,t) + \rho(x,t)\,\mathrm{D}_x v\,\mathrm{d}x. \end{aligned} \tag{A.3}$$

Da $\varphi(\cdot, t)$ ein Homöomorphismus ist, kann durch geeignete Wahl von U_0 zum Zeitpunkt t jede Teilmenge von $\Omega(t)$ in der Form $U(t)$ geschrieben werden — deshalb kann Gleichung A.3 lokalisiert werden, was auf ihre differentielle Form

$$\begin{aligned} 0 =\ & \mathrm{D}_t\rho(x,t) + (v\,\mathrm{D}_x)\rho(x,t) + \rho(x,t)\,\mathrm{D}_x v \\ & \text{(Produktregel)} \\ =\ & \mathrm{D}_t\rho(x,t) + \mathrm{D}_x\big(\rho(x,t)v(x,t)\big) \quad \text{für alle } (x,t) \in C(\Omega_0, t_0, t_\mathrm{e}) \end{aligned} \tag{A.4}$$

führt.

Bemerkung A.10. Nachdem man die Produktregel angewendet hat (in der Integralform), liefert der Satz von Gauß folgende Version des Transportsatzes für skalarwertige Funktionen:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{U(t)} f(x,t)\,\mathrm{d}x = \int_{U(t)} \mathrm{D}_t\rho(x,t)\,\mathrm{d}x + \oint_{\partial U(t)} \rho(x,t) v n(x,t)\,\mathrm{d}\sigma(x). \tag{A.5}$$

$n(x,t)$ ist die äußere Einheitsnormale auf $\partial U(t)$ an der Stelle (x,t). Betrachtet man nun noch einmal das Prinzip der Massenerhaltung, so ergibt sich hieraus: Der Massengewinn über die Oberfläche eines Gebietes wird genau durch die Dichteänderung im Inneren dieses Gebietes kompensiert.

Im weiteren werden nur Fluide mit konstanter Dichte untersucht, was den Erhaltungssatz auf

$$\mathrm{D}_x v(x,t) = 0 \quad \text{für alle } (x,t) \in C(\Omega_0, t_0, t_\mathrm{e}), \tag{A.6}$$

$$\rho(x,t) \equiv \rho \geq 0 \tag{A.7}$$

reduziert. Diese Vereinfachung kann unter folgenden Bedingungen gerechtfertigt werden:

- Die auftretenden Geschwindigkeiten sind im Vergleich zur Schallgeschwindigkeit des Fluids klein; bei fast allen Flüssigkeiten liegt sie zwischen $1000\frac{m}{s}$ und $1500\frac{m}{s}$, bei Gasen in der Größenordnung von $200\frac{m}{s}$–$400\frac{m}{s}$.

- Insbesondere bei Gasen dürfen Druck und Temperatur nur sehr geringen Schwankungen unterworfen sein. Die Dichte von Flüssigkeiten hängt nur schwach von der Temperatur ab — andererseits führen große Temperaturunterschiede auch bei ihnen zu Problemen, wie in Abschnitt A.6.2 deutlich werden wird.

Bemerkung A.11 *(Mehrphasensysteme).* Im Falle mehrphasiger, nicht-mischbarer fluider Systeme gilt Gleichung A.6 in jeder Phase, doch ist ρ auf der Phasengrenze noch nicht einmal stetig. Es muß deshalb eine Bewegungsgleichung für die Phasengrenze aufgestellt werden, auf der Randbedingungen von der Form

$$u_i(x,t) n_{\text{Phasengrenze}}(x,t) = 0$$

eine Trennung der Fluide gewährleisten.

A.5 Dynamik

Die Herleitung der Bewegungsgleichungen basiert im wesentlichen auf den drei Newtonschen Axiomen und den Erhaltungssätzen für Impuls und Drehimpuls. Dabei treten Kräfte auf, die sich grob in zwei Kategorien einteilen lassen:

1. *Volumenkräfte* können als Kraftdichten, die auf jedes Fluidvolumen wirken, beschrieben werden; in diese Kategorie fallen zum Beispiel Gravitation, die Corioliskraft aufgrund der Erdrotation oder die Wirkung externer Magnetfelder.

2. *Kontaktkräfte* wirken direkt zwischen einzelnen Fluidpartikeln aufgrund elektromagnetischer Wechselwirkung. Dies unmittelbar zu simulieren ist weder theoretisch noch numerisch für makroskopische Fluidmengen durchführbar. Daher widmet sich ihnen der Rest dieses Abschnittes.

Axiom A.12 (Euler-Cauchy-Prinzip). *Die Kontaktkräfte im Inneren eines Fluids sind durch das Vektorfeld*

$$\hat{t} : C(\Omega_0, t_0, t_e) \times S^n \longrightarrow \mathbb{R}^n : (x,t,n) \longmapsto \hat{t}(x,t,n) \tag{A.8}$$

wie folgt gegeben:

$$\oint_{\partial U(t)} \hat{t}(x,t,n(x,t))\,\mathrm{d}\sigma(x) \quad (U(t) \subset \Omega(t)) \tag{A.9}$$

ist die Kraft, die das Fluid in $\Omega(t) \setminus U(t)$ auf $U(t)$ ausübt. Analog ergibt

$$\oint_{\partial U(t)} \hat{t}(x,t,n(x,t)) \wedge (x - x_0)\,\mathrm{d}\sigma(x) \quad (U(t) \subset \Omega(t), x_0 \in \mathbb{R}^n) \tag{A.10}$$

das von $\Omega(t) \setminus U(t)$ auf $U(t)$ ausgeübte Drehmoment bezüglich der Drehachse x_0. $\hat{t}$ heißt Spannungsvektor.

Als nächstes wird ein Ausdruck für den Impuls bzw. Drehimpuls von $U(t)$ angegeben:

$$I(U(t)) := \int_{U(t)} \rho(x,t)v(x,t)\,\mathrm{d}x \quad \text{(Impuls)}, \tag{A.11}$$

$$L(U(t)) := \int_{U(t)} \rho(x,t)(x-x_0)\wedge v(x,t)\,\mathrm{d}x \quad \text{(Drehimpuls)}. \tag{A.12}$$

Axiom A.13 (Isaac Newton). *Die zeitliche Änderung des Impulses eines Körpers ist gleich der Summe der auf ihn einwirkenden Kräfte.*

Axiom A.13 und der Transportsatz ergeben zusammen

$$\begin{aligned}\oint_{U(t)} \hat{t}(x,t,n(x,t))\,\mathrm{d}\sigma(x) + \int_{U(t)} f(x,t)\rho(x,t)\,\mathrm{d}x \\ = \frac{\mathrm{d}}{\mathrm{d}t} I(U(t)) \\ = \int_{U(t)} \mathrm{D}_t(\rho v) + (v\,\mathrm{D}_x)(\rho v) + \rho v\,\mathrm{D}_x v\,\mathrm{d}x.\end{aligned} \tag{A.13}$$

Da nicht klar ist, wie man das auftretende Oberflächenintegral in ein Volumenintegral umwandeln kann, erscheint diese Gleichung nicht als nützlich. Aus grundlegenden Bewegungsgesetzen ergeben sich jedoch Bedingungen an die Struktur des Spannungsvektors, die jetzt untersucht werden.

A.5.1 Spannungstensor

Als erstes erhält man aus dem dritten Newtonschen Axiom (actio gleich reactio) das

Lemma A.14. *Es gilt*

$$-\hat{t}(x,t,n) = \hat{t}(x,t,-n) \quad \textit{für alle } (x,t,n) \in C(\Omega_0,t_0,t_e)\times S^n.$$

Beweis. Seien $A(t), B(t) \subseteq \Omega(t)$ mit gemeinsamem Rand $r(t) = \partial A(t) \cap \partial B(t)$. Dann verhalten sich die Kontaktkräfte, $F_{\mathrm{A}\,\to\,\mathrm{B}}$ und $F_{\mathrm{B}\,\to\,\mathrm{A}}$, die die Gebiete aufeinander ausüben, so:

$$\begin{aligned}\oint_{r(t)} -\hat{t}(x,t,n(x,t))\,\mathrm{d}\sigma(x) = -F_{\mathrm{B}\,\to\,\mathrm{A}} \\ = F_{\mathrm{A}\,\to\,\mathrm{B}} = \oint_{r(t)} \hat{t}(x,t,-n(x,t))\,\mathrm{d}\sigma(x).\end{aligned} \tag{A.14}$$

Wählt man nun einen beliebigen Punkt $(x,t,n) \in C(\Omega_0,t_0,t_\mathrm{e})\times S^n$, so gibt es eine Schar von offenen Kreiszylindern K_r, $r \geq 0$, mit folgenden Eigenschaften:

1. $K_r \subseteq \Omega(t)$.
2. x liegt im Schwerpunkt von K_r.
3. n ist die Symmetrieachse von K_r.
4. Die Höhe und der Basiskreisradius von K_r sind gleich r.

Schneidet man K_r mit einer Hyperebene orthogonal zu n durch x und nennt die zwei Teile $A(t)$ und $B(t)$, dann ist deren gemeinsamer Rand $r(t)$ eine $(n-1)$-dimensionale Vollkugel.
In A.14 stimmt die Normalenrichtung in beiden Integranden konstant mit n bzw. $-n$ überein. Für $r \longrightarrow 0$ konvergieren daher beide Integralmittel in

$$\frac{1}{|r(t)|} \oint_{r(t)} -\hat{t}(x,t,n)\,\mathrm{d}\sigma(x) = \frac{1}{|r(t)|} \oint_{r(t)} \hat{t}(x,t,-n)\,\mathrm{d}\sigma(x)$$

nach Lebesgues Satz über die Ableitung (siehe [19]) gegen den Wert des jeweiligen Integranden in (x,t), was den Beweis vervollständigt. □

Definition A.15. Sei $U(t) \subseteq \Omega(t)$ ein offenes, beschränktes, konvexes Gebiet, $x_0 \in U(t)$ und $r \geq 0$. Dann sei

$$U_{r,x_0} := r(U(t) - x_0) + x_0$$

die mit r um x_0 skalierte Menge $U(t)$.

Als nächstes wird gezeigt, daß die Kontaktkräfte lokal im Gleichgewicht sind. Was das genau bedeutet, erklärt

Lemma A.16. *Ist $U(t) \subseteq \Omega(t)$ ein offenes, beschränktes, konvexes Gebiet und $x_0 \in U(t)$, so gilt*

$$\left| \frac{1}{|\partial U_{r,x_0}|} \oint_{\partial U_{r,x_0}} \hat{t}(x,t,n)\,\mathrm{d}\sigma(x) \right| \longrightarrow 0 \quad \textit{für} \quad r \longrightarrow 0.$$

Beweis. Sei $U(t) \subseteq \Omega(t)$ ein offenes, beschränktes, konvexes Gebiet und $x_0 \in U(t)$. Man rechnet leicht $|U_{r,x_0}| = r^n |U(t)|$ und $|\partial U_{r,x_0}| = r^{n-1}|\partial U(t)|$ nach. Außerdem gilt $U_{r,x_0} \subseteq U(t)$, wenn $r \leq 1$ ist. Dann liefert die Bewegungsgleichung A.13 nach Division durch $|\partial U_{r,x_0}|$

$$\begin{aligned}
&\left| \frac{1}{|\partial U_{r,x_0}|} \oint_{\partial U_{r,x_0}} \hat{t}(x,t,n(x,t))\,\mathrm{d}\sigma(x) \right| \\
&\qquad = \left| \frac{1}{|\partial U_{r,x_0}|} \int_{U_{r,x_0}} \underbrace{-f\rho + \mathrm{D}_t(\rho v) + (v\,\mathrm{D}_x)(\rho v) + \rho v(\mathrm{D}_x v)}_{=:F}\,\mathrm{d}x \right| \\
&\qquad \leq \frac{|U_{r,x_0}|}{|\partial U_{r,x_0}|} \|F\|_{\infty,U(t)} \\
&\qquad = r\frac{|U(t)|}{|\partial U(t)|} \|F\|_{\infty,U(t)} \longrightarrow 0 \quad \text{für} \quad r \longrightarrow 0,
\end{aligned}$$

was den Beweis abschließt. □

Auf der Grundlage der vorangehenden Lemmata läßt sich zeigen, daß der Spannungsvektor „linear“ von der Normalenrichtung n abhängt, was dazu führt, daß auf das Oberflächenintegral in A.13 der Satz von Gauß angewendet werden kann.

Satz A.17. *Es gibt eine stetige Funktion* $T : C(\Omega_0, t_0, t_e) \times S^n \longrightarrow \mathbb{R}^{n\times n}$*, so daß für alle* $(x, t, n) \in C(\Omega_0, t_0, t_e) \times S^n$

$$\hat{t}(x, t, n) = T(x, t)n$$

erfüllt ist. Man nennt T *den Spannungstensor.*

Beweis. Der Spannungstensor wird für beliebige $(x, t) \in C(\Omega_0, t_0, t_\mathrm{e})$ durch $\mathbb{R}$-lineare Fortsetzung der folgenden Zuordnung definiert:

$$T(x, t) : e_i \longmapsto \hat{t}(x, t, e_i) \quad \text{für} \quad i = 1, \ldots, n.$$

T wird so offensichtlich zu einer in (x, t, n) stetigen Funktion. Man fixiere nun einen beliebigen Punkt $(x_0, t, n_0) \in C(\Omega_0, t_0, t_\mathrm{e}) \times S^n$, der $n_0 > 0$ (komponentenweise) erfüllt, sowie ein beliebiges $\varepsilon > 0$.
Sei S_r, $r \geq 0$, eine Schar von n-Simplizes mit den Eigenschaften:

1. x_0 ist der Schwerpunkt von S_r, $r \geq 0$.
2. Die Seite $F_{r,0}$ von S_r hat für alle $r \geq 0$ die äußere Normale n_0.
3. Die Seiten $F_{r,i}$, $i = 1, \ldots, n$ haben für alle $r \geq 0$ die äußeren Normalen $-e_i$.
4. Der Umkreisradius von S_r ist für alle $r \geq 0$ gleich r.

Da $\hat{t}$ stetig ist, existiert ein $r_0 > 0$, so daß jedes $0 \leq r < r_0$ die Ungleichungen

$$\left| \int_{F_{r,0}} \hat{t}(x, t, n_0)\, \mathrm{d}\sigma(x) - |F_{r,0}|\hat{t}(x_0, t, n_0) \right| \leq \varepsilon |F_{r,0}|, \tag{A.15}$$

$$\left| \int_{F_{r,i}} \hat{t}(x, t, -e_i)\, \mathrm{d}\sigma(x) - |F_{r,i}|\hat{t}(x_0, t, -e_i) \right| \leq \varepsilon |F_{r,i}| \quad \text{für} \quad i = 1, \ldots, n \tag{A.16}$$

erfüllt, und zusätzlich $S_{r_0} \subseteq U(t)$ wahr ist. Beachtet man, daß man

$$\oint_{\partial S_r} \hat{t}(x, t, n(x, t))\, \mathrm{d}\sigma(x) = \oint_{F_{r,0}} \hat{t}(x, t, n_0)\, \mathrm{d}\sigma(x) + \sum_{i=1}^{n} \oint_{F_{r,i}} \hat{t}(x, t, -e_i)\, \mathrm{d}\sigma(x)$$

schreiben kann, so folgt aus A.15 und A.16 mittels Dreiecksungleichung

$$\left| \oint_{\partial S_r} \hat{t}(x, t, n(x, t))\, \mathrm{d}\sigma(x) - |F_{r,0}|\hat{t}(x_0, t, n_0) - \sum_{i=1}^{n} |F_{r,i}|\hat{t}(x_0, t, -e_i) \right| \leq (n + 1)\varepsilon \max_{i=0,\ldots,n} (|F_{r,i}|). \tag{A.17}$$

Im nächsten Schritt werden die in A.17 auftretenden Maße der $F_{r,i}$ gegen $|F_{r,0}|$ abgeschätzt. Für die Seiten des Simplexes S_r findet man für $i = 1, \ldots, n$

$$|F_{r,i}| = |F_{r,0}|. \tag{A.18}$$

Wendet man A.18 und Lemma A.14 auf A.17 an, so ergibt sich

$$\left| \oint_{\partial S_r} \hat{t}(x,t,n(x,t))\,\mathrm{d}\sigma(x) - |F_{r,0}| \Bigg(\hat{t}(x_0,t,n_0) - \underbrace{\sum_{i=1}^{n} (n_0)_i \hat{t}(x_0,t,e_i)}_{=T(x_0,t)n_0} \Bigg) \right| \leq (n+1)\varepsilon |F_{r,0}|, \tag{A.19}$$

was per Dreiecksungleichung und Division durch $|F_{r,0}|$ zu

$$\left| \hat{t}(x_0,t,n_0) - T(x_0,t)n_0 \right| \leq (n+1)\varepsilon + \frac{1}{|F_{r,0}|} \left| \oint_{\partial S_r} \hat{t}(x,t,n(x,t))\,\mathrm{d}\sigma(x) \right| \tag{A.20}$$

wird. Um Lemma A.16 auf den rechten Term von A.20 anwenden zu können, wird mittels A.18

$$|\partial S_r| = |F_{r,0}| + \sum_{i=1}^{n} |F_{r,i}| = |F_{r,0}| \left(1 + \sum_{i=1}^{n} (n_0)_i \right) = |F_{r,0}| C(n_0)$$

abgeschätzt, wobei $C(n_0) > 0$ eine Konstante ist, die nur von n_0 abhängt. Somit erhält man

$$\begin{aligned} \left| \hat{t}(x_0,t,n_0) - T(x_0,t)n_0 \right| &\leq (n+1)\varepsilon + \frac{C(n_0)}{|\partial S_r|} \left| \oint_{\partial S_r} \hat{t}(x,t,n(x,t))\,\mathrm{d}\sigma(x) \right| \\ &= (n+1+C(n_0))\varepsilon \end{aligned} \tag{A.21}$$

für alle hinreichend kleinen r. Damit ist die Gleichheit von $\hat{t}$ und T für alle $n_0 \in S^n$, $n_0 > 0$, bewiesen.
Dasselbe Argument läßt sich für alle n_0 anwenden, die in jeder Komponente von Null verschieden sind, indem man die Koordinatenachsen, deren Komponente negativ ist, umkehrt. In dem entstehenden Koordinatensystem gilt wieder $n_0 > 0$. Es fehlt noch der Beweis der Gleichheit auf dem Schnitt der Koordinatenebenen mit der S^n: Sowohl $\hat{t}$ als auch T sind für alle $n_0 \in S^n$ definiert und stetig. Da sie nach dem Vorangegangenen sogar auf einer offenen Teilmenge von S^n übereinstimmen, deren Abschluß die gesamte S^n ist, folgt aus der Eindeutigkeit der stetigen Fortsetzung auf den Abschluß die Gleichheit beider Funktionen und damit die Behauptung des Satzes. □

Bemerkung A.18. Im Gegensatz zum Spannungsvektor ist der Spannungstensor sogar auf $C(\Omega_0, t_0, t_\mathrm{e}) \times \mathbb{R}^n$ erklärt.

Prinzipiell kann man jetzt die Bewegungsgleichung A.13 lokalisieren, doch vorher wird bewiesen, daß der Spannungstensor symmetrisch ist, weil diese Eigenschaft in Abschnitt A.6.2 benötigt wird und ihr Beweis analog zum gerade Gesehenen verläuft.

A.5.2 Symmetrie des Spannungstensors

Wie für den Linearimpuls existiert ein zweites Newtonsches Axiom auch für den Drehimpuls L, der in A.12 definiert wurde. Mit Hilfe des Transportsatzes ergibt sich eine zu A.13 analoge „Drehbewegungsgleichung“ bezüglich einer Drehachse in $x_0 \in \mathbb{R}^n$.

Notation A.19. Ist x die räumliche Variable und $x_0 \in \mathbb{R}^n$ der Punkt auf den die Drehung bezogen wird, so wird $\tilde{x} := x - x_0$ geschrieben.

$$\begin{aligned}
&\oint_{U(t)} \hat{t}(x,t,n(x,t)) \wedge (x-x_0)\,\mathrm{d}\sigma(x) + \int_{U(t)} \rho(x,t)(x-x_0) \wedge f(x,t)\,\mathrm{d}x \\
&= \frac{\mathrm{d}}{\mathrm{d}t} L(U(t)) \\
&= \int_{U(t)} \mathrm{D}_t(\rho(x-x_0)\wedge v) + (v\,\mathrm{D}_x)(\rho(x-x_0)\wedge v) + \rho(x-x_0)\wedge v\,\mathrm{D}_x v\,\mathrm{d}x.
\end{aligned} \tag{A.22}$$

Es läßt sich beweisen, daß sich die Drehmomente lokal im Gleichgewicht befinden.

Lemma A.20. *Ist $U(t) \subseteq \Omega(t)$ ein offenes, beschränktes, konvexes Gebiet und $x_0 \in U(t)$, so gilt*

$$\left| \frac{1}{|\partial U_{r,x_0}|} \oint_{\partial U_{r,x_0}} \hat{t}(x,t,n) \wedge \tilde{x}\,\mathrm{d}\sigma(x) \right| \longrightarrow 0 \quad \textit{für} \quad r \longrightarrow 0.$$

Beweis. Sei $U(t) \subseteq \Omega(t)$ ein offenes, beschränktes, konvexes Gebiet und $x_0 \in U(t)$. Aus der Drehmomentgleichung A.22 folgt dann

$$\begin{aligned}
\left| \oint_{U(t)} \hat{t}(x,t,n(x,t)) \wedge \tilde{x}\,\mathrm{d}\sigma(x) \right| = \Bigg| \int_{U(t)} \rho(x,t)\tilde{x} \wedge f(x,t) + \underbrace{\mathrm{D}_t(\tilde{x}\wedge(\rho v))}_{=\tilde{x}\wedge \mathrm{D}_t(\rho v)} \\
+ \underbrace{(v\,\mathrm{D}_x)(\tilde{x}\wedge(\rho v))}_{=:A} + \tilde{x}\wedge(\rho v)\,\mathrm{D}_x v\,\mathrm{d}x \Bigg|.
\end{aligned}$$

Wegen

$$A = \big((v\,\mathrm{D}_x)\tilde{x}\big) \wedge (\rho v) + \tilde{x} \wedge \big((v\,\mathrm{D}_x)(\rho v)\big) = \tilde{x} \wedge \big((v\,\mathrm{D}_x)(\rho v)\big)$$

erhält man mit F aus dem Beweis von Lemma A.16

$$\begin{aligned}\left| \oint_{U(t)} \hat{t}(x,t,n(x,t)) \wedge \tilde{x}\,\mathrm{d}\sigma(x) \right| &= \left| \int_{U(t)} \tilde{x} \wedge F(x,t,v(x,t))\,\mathrm{d}x \right| \\ &\leq |U_{r,x_0}| \|\tilde{x} \wedge F\|_{\infty,U(t)} \\ &\leq |U_{r,x_0}| \|F\|_{\infty,U(t)} \|\tilde{x}\|_{\infty,U(t)} \\ &= |U_{r,x_0}| \|F\|_{\infty,U(t)} C(U(t))\, r \longrightarrow 0 \quad \text{für} \quad r \longrightarrow 0.\end{aligned}$$

Da man den Grenzwert auch noch nach der Division durch $|U_{r,x_0}|$ durchführen kann, ist die Behauptung bewiesen. □

Obwohl sich das Resultat des nächsten Satzes von A.17 unterscheidet, gleichen sich die Beweismethoden — die Integralform einer Bewegungsgleichung wird auf ein „infinitesimales Volumenelement“ angewendet.

Satz A.21. *Für jedes $(x,t) \in C(\Omega_0, t_0, t_e)$ gilt: Der Spannungstensor ist symmetrisch.*

Beweis. Die Symmetrie wird für einen beliebig zu wählenden Punkt $(x_0,t) \in C(\Omega_0, t_0, t_\mathrm{e})$ gezeigt. Es wird $B_r := \{x \in \mathbb{R}^n \mid \|x - x_0\|_2 < r\}$ definiert, und

$$\big(t_{i,j}\big)_{i,j=1,\ldots,n} = \big(T_j\big)_{j=1,\ldots,n} = T(x,t)$$

seien die Komponenten bzw. die Spalten von $T(x,t)$.

$$\begin{aligned}\oint_{\partial B_r} \hat{t}(x,t,n(x,t)) \wedge \tilde{x}\,\mathrm{d}\sigma(x) &= \oint_{\partial B_r} (T(x,t)n) \wedge \tilde{x}\,\mathrm{d}\sigma(x) \\ &= \oint_{\partial B_r} \sum_{j=1}^{n} n_j T_j \wedge \tilde{x}\,\mathrm{d}\sigma(x) \\ &\quad \text{(Satz von Gauß)} \\ &= \int_{B_r} \sum_{j=1}^{n} \mathrm{D}_j (T_j \wedge \tilde{x})\,\mathrm{d}x \\ &= \int_{B_r} \sum_{j=1}^{n} \left(\left(\sum_{i=1}^{n} e_i\, \mathrm{D}_j t_{i,j} \right) \wedge \tilde{x} - T_j \wedge e_j \right) \mathrm{d}x \\ &= \int_{B_r} (\,\mathrm{D}_x T) \wedge \tilde{x}\,\mathrm{d}x - \int_{B_r} \tilde{A}(x,t)\,\mathrm{d}x.\end{aligned} \tag{A.23}$$

Dabei ist

$$\tilde{A}(x,t) := \sum_{i,j=1}^{n} t_{i,j} e_i \wedge e_j.$$

Die Gleichung A.23 wird nach dem Integral über $\tilde{A}$ aufgelöst und durch $|B_r|$ geteilt, was auf

$$\frac{1}{|B_r|}\int_{B_r}\tilde{A}(x,t)\,\mathrm{d}x = \frac{1}{|B_r|}\int_{B_r}(\,\mathrm{D}_x T)\wedge\tilde{x}\,\mathrm{d}x - \frac{1}{|B_r|}\oint\hat{t}\wedge\tilde{x}\,\mathrm{d}\sigma(x) \qquad (A.24)$$

führt. Aufgrund von Lebesgues Satz über die Ableitung gilt für die beiden ersten Terme in A.24

$$\frac{1}{|B_r|}\int_{B_r}\tilde{A}(x,t)\,\mathrm{d}x \longrightarrow \tilde{A}(x_0,t) \quad \text{für} \quad r\longrightarrow 0,$$
$$\frac{1}{|B_r|}\int_{B_r}(\,\mathrm{D}_x T)\wedge\tilde{x}\,\mathrm{d}x \longrightarrow (\,\mathrm{D}_x T)\wedge(x_0-x_0)=0 \quad \text{für} \quad r\longrightarrow 0,$$

und für den letzten gilt wegen Lemma A.20

$$\frac{1}{|B_r|}\oint\hat{t}\wedge\tilde{x}\,\mathrm{d}\sigma(x)\longrightarrow 0 \quad \text{für} \quad r\longrightarrow 0.$$

Somit ergibt sich

$$\tilde{A}(x_0,t)=\sum_{i,j=1}^{n}t_{i,j}e_i\wedge e_j=0.$$

Nun läßt sich $T(x_0,t)$ eindeutig als $S+A$ schreiben, wobei $S=\frac{1}{2}(T+T^T)$ symmetrisch und $A=\frac{1}{2}(T-T^T)$ antisymmetrisch ist. Wegen

$$A=\frac{1}{2}\sum_{i,j=1}^{n}t_{i,j}e_i\otimes e_j-\frac{1}{2}\sum_{i,j=1}^{n}t_{i,j}e_j\otimes e_i$$
$$=\sum_{i,j=1}^{n}\frac{1}{2}(e_i\otimes e_j-e_j\otimes e_i)=\sum_{i,j=1}^{n}e_i\wedge e_j=\tilde{A}(x_0,t)$$

ist der antisymmetrische Anteil von $T(x_0,t)$ Null, was den Satz beweist. □

Mit dem Wissen aus Satz A.17 kann jetzt die Bewegungsgleichung A.13 lokalisiert werden. Der Term

$$\oint_{\partial U(t)}\hat{t}(x,t,n(x,t))\,\mathrm{d}\sigma(x)=\oint_{\partial U(t)}T(x,t)n(x,t)\,\mathrm{d}\sigma(x)$$
$$=\int_{U(t)}\mathrm{D}_x T(x,t)\,\mathrm{d}x$$

wird mit dem Satz von Gauß in ein Volumenintegral umgewandelt; dabei ist $\mathrm{D}_x T$ als $(\,\mathrm{D}_x T_{1,\cdot},\ldots,\mathrm{D}_x T_{n,\cdot})^T$, also als Spalte, die die Divergenzen der Zeilen von T enthält, definiert. Das liegt daran, daß für gewöhnlich im Satz von Gauß ein Skalarprodukt von Spalten steht, während das Matrix-Vektor-Produkt hier von der

Form Zeile × Spalte ist. Deshalb erhält man die grundlegende Bewegungsgleichung

$$\mathrm{D}_t(\rho v) + (v\,\mathrm{D}_x)(\rho v) + \rho v\,\mathrm{D}_x v = \mathrm{D}_x T + \rho f. \tag{A.25}$$

Wenn man sich auf inkompressible Flüsse ($\mathrm{D}_x v = 0$) und konstante Dichte einschränkt, wird sie zu

$$\rho\,(\mathrm{D}_t v + (v\,\mathrm{D}_x)v) = \mathrm{D}_x T + \rho f. \tag{A.26}$$

A.6 Modelle für den Spannungstensor

Nachdem im vorhergehenden wesentliche Eigenschaften des Spannungstensors hergeleitet wurden, können jetzt Modelle angegeben werden, die diesen genügen. Experimente der Hydrostatik legen folgendes nahe:

Axiom A.22 (hydrostatischer Druck). *In ruhenden Fluiden gibt es eine Funktion $p : \Omega_0 \longrightarrow \mathbb{R} : x \longmapsto p(x)$, deren Integral $\oint_F p(x) n \,\mathrm{d}\sigma(x)$ über eine Hyperfläche $F \subset \Omega_0$ die Kraft angibt, die das Fluid auf der von n abgewendeten Seite auf F ausübt. p heißt der Druck.*

Man beachte, daß die Kraft, die der Druck erzeugt immer orthogonal zur betrachteten Oberfläche wirkt.

A.6.1 Eulergleichungen

Das einfachste Modell für den Spannungstensor bewegter Fluide wäre, daß der Druck zeitabhängig ist, aber ansonsten keine weiteren Kontaktkräfte auftreten.

$$\Longrightarrow T(x,t) := -p(x,t) I_{n\times n}.$$

Auf diese Weise erhält man aus A.25 die Eulergleichungen

$$\mathrm{D}_t(\rho v) + (v\,\mathrm{D}_x)(\rho v) + \rho v\,\mathrm{D}_x v = -\,\mathrm{D}_x p + \rho f. \tag{A.27}$$

Fluide, die dieser Bewegungsgleichung gehorchen heißen ideal, da sie keine innere Reibung besitzen. Obwohl diese Bewegungsgleichungen in vielen Anwendungsgebieten eine gute Approximation des realen Fluidverhaltens liefern (z. B. bei Gasen), sind in dieser Arbeit Fluide von Interesse, deren Kontaktkräfte nur einem komplexeren Ansatz genügen.

A.6.2 Navier-Stokes-Gleichungen

Auch die Navier-Stokes-Gleichungen sollen den hydrostatischen Fall korrekt beschreiben; dort ist der Spannungstensor isotrop:

Definition A.23 (isotrope Tensoren). Eine $p+q$-lineare Abbildung

$$L : \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{p \text{ Faktoren}} \times \underbrace{\mathbb{R}^{1\times n} \times \cdots \times \mathbb{R}^{1\times n}}_{q \text{ Faktoren}} \longrightarrow \mathbb{R}$$

heißt p-fach kontravarianter und q-fach kovarianter Tensor oder Tensor der Stufe (p, q); L heißt isotrop, wenn für alle orthogonalen Abbildungen $M \in O(n)$ gilt:

$$L(Mx_1, \ldots, Mx_p, y_{p+1}M^T, \ldots, y_{p+q}M^T) = L(x_1, \ldots, x_p, y_{p+1}, \ldots, y_{p+q})$$
$$\text{für alle } x_1, \ldots, x_p \in \mathbb{R}^{n\times 1}, y_{p+1}, \ldots, y_{p+q} \in \mathbb{R}^{1\times n}.$$

Bemerkung A.24.

- Die 1-fach kovarianten und 1-fach kontravarianten Tensoren $m(x, y)$ stehen zu den linearen Abbildungen $M : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ durch

  $$m(x, y) = y(Mx) \quad \text{für alle } x \in \mathbb{R}^n,\ y \in \mathbb{R}^{1\times n}$$

  in Bijektion.

- Isotropie eines Tensors bedeutet, daß er keine Raumrichtung auszeichnet — auf den Bahnen der orthogonalen Gruppe in seinem Definitionsbereich ist sein Wert konstant.

Als erstes wird T nun in den isotropen Teil

$$-p(x,t)I_{n\times n} := \frac{1}{n}\operatorname{tr}(T(x,t))I_{n\times n} \tag{A.28}$$

und den Rest

$$\tilde{T} := T - \frac{1}{n}\operatorname{tr}(T)I_{n\times n} \tag{A.29}$$

zerlegt. Mit derselben Technik wie in Satz A.30 kann man für Tensoren der Stufe $(1, 1)$ zeigen, daß die Vielfachen der Identität die einzigen isotropen Tensoren der Stufe $(1, 1)$ sind. Man beachte $\operatorname{tr}(\tilde{T}) = 0$, was zusammen mit Lemma A.25 die Bezeichnung mechanischer Druck für p und die physikalische Sinnhaftigkeit dieser Zerlegung rechtfertigt.

Lemma A.25. *Ist $(x_0, t) \in C(\Omega_0, t_0, t_e)$, so gilt für die Normalkomponente der Kontaktkraft*

$$\frac{1}{|\partial B_r(x_0)|}\oint_{\partial B_r(x_0)} n^T T n \,\mathrm{d}\sigma(x) \longrightarrow \frac{1}{n}\operatorname{tr}(T(x_0,t)) \quad \textit{für} \quad r \longrightarrow 0.$$

Der „Druckanteil“ der Kontaktkräfte beruht also im wesentlichen auf der Spur von T.

Beweis. Sei $(x_0, t) \in C(\Omega_0, t_0, t_e)$, dann folgt mit der Transformation $x \longmapsto x - x_0$ und dem Satz von Gauß

$$\begin{aligned} \frac{1}{|\partial B_r|} \oint_{\partial B_r(x_0)} n^T T n \, \mathrm{d}\sigma(x) &= \frac{1}{|\partial B_r|} \oint_{\partial B_r(0)} n^T T(x + x_0, t) n \, \mathrm{d}\sigma(x) \\ &= \frac{|B_r|}{|\partial B_r|} \frac{1}{|B_r|} \int_{B_r(0)} \mathrm{D}_x (T(x + x_0, t) n) \, \mathrm{d}x. \end{aligned} \tag{A.30}$$

Nach der Anwendung des Transformationssatzes hat der Integrand die Form $\frac{x^T}{r} T(x + x_0, t) \frac{x}{r}$, so daß

$$\begin{aligned} \mathrm{D}_x (T(x + x_0, t) n) &= \frac{1}{r} (\mathrm{D}_x T(x + x_0, t))^T x + \frac{1}{r} \sum_{i,j=1}^{n} t_{i,j}(x + x_0, t) \, \mathrm{D}_i x_j \\ &= \frac{1}{r} (\mathrm{D}_x T(x + x_0, t))^T x + \frac{1}{r} \operatorname{tr}(T(x + x_0, t)) \end{aligned}$$

gilt. Zusammen mit $\frac{|B_r|}{|\partial B_r|} = \frac{r}{n}$ ergibt dies in A.30

$$\begin{aligned} &\frac{1}{|\partial B_r|} \oint_{\partial B_r(x_0)} n^T T n \, \mathrm{d}\sigma(x) \\ &\qquad = \frac{1}{n} \frac{1}{|B_r|} \int_{B_r(0)} (\mathrm{D}_x T(x + x_0, t))^T x + \operatorname{tr}(T(x + x_0, t)) \, \mathrm{d}x. \end{aligned}$$

Da der Integrand stetig ist, konvergiert das Integralmittel im rechten Term für $r \longrightarrow 0$ gegen den Wert des Integranden bei $x = 0$, was das Lemma beweist. □

Der Tensor $\tilde{T}$ heißt deviatorischer Spannungstensor und die durch ihn beschriebenen Kräfte wirken nur in bewegten Fluiden. Da T und $pI_{n\times n}$ symmetrisch sind, ist $\tilde{T}$ dies auch.
Im folgenden wird angenommen, daß $\tilde{T}$ linear von $\mathrm{d}_x v$ abhängt. Das zusätzliche Argument v von $\tilde{T}$ hätte auch schon bei der Herleitung der Eigenschaften des Spannungsvektors $\hat{t}$ berücksichtigt werden können, aber da es zu keiner Änderung der Sätze in Abschnitt A.5 führt, konnte so die Notation etwas bequemer bleiben. Präzise ausgedrückt heißt das: Es gibt einen Tensor $L(n, y, u, w)$ der Stufe $(2, 2)$, so daß

$$T n = \sum_{i,k,l=1}^{n} e_i L(n, e_k, e_i^T, e_l^T) (\mathrm{d}_x v)_{k,l} \tag{A.31}$$

gilt.

Bemerkung A.26. Gleichung A.31 stammt aus der Annahme, daß die Kontaktkräfte in einer Flüssigkeit das Resultat eines Impulstransports durch die Bewegung von Fluidteilchen gegeneinander sind (Kraft $\simeq$ Impulsänderung). Entsprechend dem Prototyp

$$u_t + \mathrm{D}_x (F \circ u)(x, t) = 0$$

einer Transportgleichung, geht man davon aus, daß die Flußfunktion F von der transportierten Größe abhängt. Da eine uniforme Geschwindigkeit nicht zu Wechselwirkung zwischen Teilchen führt, nimmt man eine Abhängigkeit von $\mathrm{d}_x v$ an — der einfachste Zusammenhang ist natürlich eine lineare Flußfunktion.

Um zu den Navier-Stokes-Gleichungen zu gelangen, geht man davon aus, daß L ein isotroper Tensor ist, was nach Satz A.30 eine drastische Einschränkung der Struktur von L nach sich zieht:

$$\tilde{T}_{i,j} = \big(\alpha\delta_{i,j}\delta_{k,l} + \beta\delta_{i,k}\delta_{j,l} + \gamma\delta_{i,l}\delta_{j,k}\big)(\,\mathrm{d}_x v)_{k,l} \quad \text{mit} \quad \alpha,\beta,\gamma \in \mathbb{R}.$$

Wegen der Symmetrie von $\tilde{T}$ setzt man $\beta = \gamma$.

$$\Longrightarrow\ \tilde{T} = \alpha \operatorname{tr}(\,\mathrm{d}_x v) I_{n\times n} + 2\beta \operatorname{Sym}(\,\mathrm{d}_x v) \quad \text{mit} \quad \alpha,\beta \in \mathbb{R},$$

wobei $\operatorname{Sym} M = \frac{1}{2}(M + M^T)$ der Symmetrisierungsoperator ist. Man beachte, daß $\operatorname{tr}(\operatorname{Sym} M) = \operatorname{tr} M$ für alle $M \in \mathbb{R}^{n\times n}$ gilt; zusammen mit der Spurfreiheit von $\tilde{T}$ ergibt sich

$$0 = \operatorname{tr}(\tilde{T}) = \alpha n \,\mathrm{D}_x v + 2\beta \,\mathrm{D}_x v.$$

Ist $\mathrm{D}_x v \neq 0$, so erhält man $\alpha = -\frac{2}{n}\beta$, doch auch falls $\mathrm{D}_x v = 0$ ist, gilt

$$\tilde{T} = 2\beta\Big(\operatorname{Sym}(\,\mathrm{d}_x v) - \frac{1}{n}\,\mathrm{D}_x v I_{n\times n}\Big) \quad \text{mit} \quad \beta \in \mathbb{R}. \tag{A.32}$$

Der Koeffizient $\mu = \beta$ heißt Zähigkeit. Insgesamt lautet der Spannungstensor somit

$$T(x,t) = -p(x,t) I_{n\times n} + 2\mu\Big(\operatorname{Sym}(\,\mathrm{d}_x v) - \frac{1}{n}\,\mathrm{D}_x v I_{n\times n}\Big). \tag{A.33}$$

Eine einfache Rechnung liefert

$$\begin{aligned}
\mathrm{D}_x T =\ & \mathrm{D}_x\Big(-pI_{n\times n} - \frac{2\mu}{n}\,\mathrm{D}_x v I_{n\times n} + \mu\,(\,\mathrm{d}_x v + \,\mathrm{d}_x v^T)\Big) \\
=\ & -\mathrm{D}_x p - \frac{2\mu}{n}\begin{pmatrix} \mathrm{D}_1\,\mathrm{D}_x v \\ \vdots \\ \mathrm{D}_n\,\mathrm{D}_x v \end{pmatrix} + \mu\Delta v + \mu\begin{pmatrix} \mathrm{D}_1\,\mathrm{D}_x v \\ \vdots \\ \mathrm{D}_n\,\mathrm{D}_x v \end{pmatrix} \\
=\ & -\mathrm{D}_x p + \mu\Big(\frac{n-2}{n}\,\mathrm{D}_x(\,\mathrm{D}_x v) + \Delta v\Big).
\end{aligned}$$

Dieser Ausdruck führt auf folgende

Aufgabe A.27 (Inkompressible Navier-Stokes-Gleichungen). Zu einem Gebiet Ω_0, einem Zeitintervall $[t_0, t_\mathrm{e})$, einer Dichte $\rho = \rho_0 = \mathit{const.}$, einer Zähigkeit $\mu = \mu_0 = \mathit{const.}$ und einer äußeren Kraftdichte f bestimme man Funktionen $v(x,t)$ und $p(x,t)$ in geeigneten Funktionenklassen mit geeigneten Rand- und Anfangswerten für v, so daß für jedes $(x,t) \in C(\Omega_0, t_0, t_\mathrm{e})$

$$\begin{aligned}
\mathrm{D}_x v(x,t) &= 0, \\
\rho\,(\,\mathrm{D}_t v + (v\,\mathrm{D}_x)v) &= -\mathrm{D}_x p + \mu\Delta v + \rho f
\end{aligned} \tag{A.34}$$

gilt. Das sind $n+1$ Gleichungen für ebensoviele Unbestimmte.

Teilt man die Bewegungsgleichung durch ρ, so erhält man $\mathrm{D}_t v + (v\,\mathrm{D}_x)v + \frac{1}{\rho}\,\mathrm{D}_x p - \frac{\mu}{\rho}\Delta v = f$; der Faktor $\nu = \frac{\mu}{\rho}$ wird kinematische Zähigkeit genannt. Er hat die Einheit $\frac{\text{Fläche}}{\text{Zeit}}$ und ist somit als Diffusionskoeffizient (für Linearimpuls) interpretierbar.

Bemerkung A.28. Die Zähigkeit vieler Fluide ist relativ stark von der Temperatur abhängig, was bei Modellen mit großen Temperaturdifferenzen nicht vernachlässigt werden darf. Bei einem Druck von 100 kPa gilt für Wasser laut [30]: $\mu_{0^\circ\mathrm{C}} = 1793\,\mu\mathrm{Pa\,s}$, $\mu_{100^\circ\mathrm{C}} = 281,8\,\mu\mathrm{Pa\,s}$.

Isotrope Tensoren

In diesem Abschnitt wird der bereits verwendete Satz über isotrope Tensoren der Stufe $(2,2)$ bewiesen.

Notation A.29. $[e_1, \ldots, e_n]$ bezeichnet die Standardbasis des $\mathbb{R}^n$, $[e^1, \ldots, e^n]$ die dazu duale (Standard-) Basis des $\mathbb{R}^{1\times n}$. $L = e_i \otimes e_j \otimes e^k \otimes e^l$ ist der Tensor der Stufe $(2,2)$, der $L(e_i, e_j, e^k, e^l) = 1$ und $L(e_a, e_b, e^c, e^d) = 0$ für alle $(a,b,c,d) \neq (i,j,k,l)$ erfüllt.
Da sie bei den folgenden Rechnungen teilweise nützlich ist, möge ab jetzt die Einsteinsche Summationskonvention gelten, nach der über diagonal stehende, doppelt auftretende Indizes summiert wird.

Satz A.30. *Die Menge*

$$\{\delta^{ij}\delta_{kl} e_i \otimes e_j \otimes e^k \otimes e^l,\ \delta^i_k \delta^j_l e_i \otimes e_j \otimes e^k \otimes e^l,\ \delta^i_l \delta^j_k e_i \otimes e_j \otimes e^k \otimes e^l\}$$

ist eine Basis des Raumes, welcher aus den isotropen Tensoren der Stufe $(2,2)$ *besteht.*

Beweis. Sei L ein beliebiger isotroper Tensor der Stufe $(2,2)$ und seien $i,j,k,l \in \{1,\ldots,n\}$ beliebige Zahlen. Zunächst wird mit ausgewählten orthogonalen Abbildungen die Struktur von L eingegrenzt.

1. Sei $O_m \in O(n)$ die Spiegelung, die $O_m e_m = -e_m$ erfüllt. Dann gilt $L(O_m e_i, O_m e_j, e^k O_m^T, e^l O_m^T) = -L(e_i, e_j, e^k, e^l)$, sobald $m \in \{i,j,k,l\}$ ist und m in dem Tupel (i,j,k,l) in ungerader Anzahl auftritt. Wegen der Isotropie von L erhält man so $-L(e_i, e_j, e^k, e^l) = L(e_i, e_j, e^k, e^l)$, das heißt $L(e_i, e_j, e^k, e^l) = 0$.

2. Es bleiben noch $e_i \otimes e_i \otimes e^j \otimes e^j$, $e_i \otimes e_j \otimes e^i \otimes e^j$, $e_i \otimes e_j \otimes e^j \otimes e^i$ und $e_i \otimes e_i \otimes e^i \otimes e^i$ $(i \neq j)$ als Basisvektoren mit möglicherweise von Null verschiedenen Linearfaktoren in L. Sei im folgenden $i \neq j$. Es wird $L(e_i, e_i, e^j, e^j) = L(e_1, e_1, e^2, e^2)$ gezeigt. Ist $(i,j) = (1,2)$, so bleibt nichts zu tun. Sonst sei

$O \in O(n)$ die Abbildung, die durch die nachfolgend angebene Permutation von Basisvektoren eindeutig bestimmt ist:

$$O = \begin{cases} (2\,j), & i = 1 \wedge j \neq 2, \\ (1\,i), & j = 2 \wedge i \neq 1, \\ (1\,i)(2\,j), & i \neq 1 \wedge j \neq 2. \end{cases}$$

Da diese Permutationen selbstinvers sind, ist es auch O, so daß

$$L(e_i, e_i, e^j, e^j) = L(Oe_i, Oe_i, e^j O^T, e^j O^T) = L(e_1, e_1, e^2, e^2)$$

folgt. Offensichtlich ergibt sich mit diesen orthogonalen Abbildungen auch $L(e_i, e_j, e^i, e^j) = L(e_1, e_2, e^1, e^2)$ und $L(e_i, e_j, e^j, e^i) = L(e_1, e_2, e^2, e^1)$.

3. Durch die selbstinversen, orthogonalen Abbildungen $O \in O(n)$, die die Basisvektoren e_1 und e_i $(i \neq 1)$ vertauschen, findet man, daß $L(e_i, e_i, e_i, e_i) = L(e_1, e_1, e_1, e_1)$ für alle i gilt.

Also liegen die isotropen Tensoren in einen höchstens vierdimensionalen Teilraum von $U = \mathrm{span}(\{\tilde{b}_1, \tilde{b}_2, \tilde{b}_3, \tilde{b}_4\})$, wobei

$$\begin{aligned} \tilde{b}_1 &= \sum_{i \neq j} e_i \otimes e_i \otimes e^j \otimes e^j, \\ \tilde{b}_2 &= \sum_{i \neq j} e_i \otimes e_j \otimes e^i \otimes e^j, \\ \tilde{b}_3 &= \sum_{i \neq j} e_i \otimes e_j \otimes e^j \otimes e^i, \\ \tilde{b}_4 &= \sum_{1=1}^{n} e_i \otimes e_i \otimes e^i \otimes e^i \end{aligned}$$

definiert wird. Die ersten drei dieser Tensoren weisen große Ähnlichkeit mit Produkten von Kroneckertensoren auf; tatsächlich sind

$$\begin{aligned} b_1 &= \tilde{b}_1 + \tilde{b}_4 = \delta^{ij}\delta_{kl} e_i \otimes e_j \otimes e^k \otimes e^l, \\ b_2 &= \tilde{b}_2 + \tilde{b}_4 = \delta^i_k \delta^j_l e_i \otimes e_j \otimes e^k \otimes e^l, \\ b_3 &= \tilde{b}_3 + \tilde{b}_4 = \delta^i_l \delta^j_k e_i \otimes e_j \otimes e^k \otimes e^l \end{aligned}$$

sogar isotrop:

$$\begin{aligned} b_1(Ox, Oy, uO^T, vO^T) &= \delta^{ij}\delta_{kl} O^i_p x^p O^j_q y^p u_r (O^{-1})^r_k v_s (O^{-1})^s_l \\ &= \delta^{ij}\delta_{kl} O^i_p x^p (O^T)^q_j y_q u_r (O^{-1})^r_k v^s O^l_s \\ &= O^i_p x^p (O^{-1})^q_i y_q u_r (O^{-1})^r_k v^s O^k_s \\ &= \delta^q_p x^p y_q \delta^r_s u_r v^s = \delta^{ij}\delta_{kl} x^i y^j u_k v_l \\ &= b_1(x, y, u, v) \quad \text{für alle } O \in O(n). \end{aligned}$$

Die Rechnungen zu b_2 und b_3 verlaufen analog. Als nächstes wird bewiesen, daß $\{b_1, b_2, b_3\}$ linear unabhängig ist. Dazu sei $S := ab_1 + bb_2 + cb_3 = 0$ mit $a, b, c \in \mathbb{R}$. Wegen $S(e_1, e_1, e^2, e^2) = a$, $S(e_1, e_2, e^1, e^2) = b$ und $S(e_1, e_2, e^2, e^1) = c$ ist $a = b = c = 0$, was lineare Unabhängigkeit bedeutet. Zum Schluß wird nachgewiesen, daß $\tilde{b}_4$ nicht isotrop ist: Offensichtlich gilt $\tilde{b}_4(e_1, e_1, e^1, e^1) = 1$. Man betrachte die Drehung O in der e_1e_2-Ebene, die wie folgt definiert ist:

$$O: e_1 \longmapsto \frac{1}{\sqrt{2}}(e_1 + e_2), e_2 \longmapsto \frac{1}{\sqrt{2}}(e_2 - e_1), e_i \longmapsto e_i \quad (i \neq 1, 2).$$

Damit folgt

$$\begin{aligned}\tilde{b}_4(Oe_1, Oe_1, e^1O^T, e^1O^T) &= \frac{1}{4}\tilde{b}_4(e_1 + e_2, e_1 + e_2, e^1 + e^2, e^1 + e^2) \\ &= \frac{1}{4}\left(\tilde{b}_4(e_1, e_1, e^1, e^1) + \tilde{b}_4(e_2, e_2, e^2, e^2)\right) \\ &= \frac{1}{2} \neq 1,\end{aligned}$$

was zeigt, daß $\tilde{b}_4$ nicht isotrop ist.
Mithin hat der Raum der isotropen Tensoren als Teilraum von U maximal die Dimension 3, so daß $\{b_1, b_2, b_3\}$ eine Basis für alle isotropen Tensoren darstellt. □

A.7 Randbedingungen

Da es sich als schwierig herausgestellt hat, Informationen über physikalisch sinnvolle Randwerte zu finden, folgt an dieser Stelle eine kurze Übersicht.
Anfangswerte für v können weitgehend beliebig vorgegeben werden, solange sie divergenzfrei sind. Bei stationären Gleichungen treten häufig folgende Randbedingungen an v auf:

- *Dirichlet-Randbedingungen* treten in zwei Situationen auf – zum einen in Einströmungsbereichen des Randes (Inflow) als inhomogene Randdaten des einströmenden Geschwindigkeitsfeldes – zum anderen als (meist) homogene Randdaten an Randstücken mit einer Haftbedingung (no-slip)[2]. Typisch für den letzten Fall sind flüssig-fest-Phasengrenzen. Werden auf ganz $\partial\Omega$ Dirichlet-Randbedingungen vorgeschrieben, so folgt aus Inkompressibilität und dem Satz von Gauß die *Kompatibilitätsbedingung* $\int_{\partial\Omega} u_{\mathrm{D}}(x)n(x)\,\mathrm{d}\sigma x = 0$.

- *Neumann-Randbedingungen* werden normalerweise nicht direkt verwendet, da sie nur eine „mathematische Interpretation“ haben: Das Strömungsverhalten ändert sich beim Überschreiten des Gebietsrandes nicht, was normalerweise nicht experimentell modelliert werden kann. Da sie mathematisch

[2]Das Fluid haftet am Rand des Gebietes und dieser ruht.

wohldefiniert sind, werden sie manchmal an weit vom zu modellierenden Phänomen entfernten Rändern als „Gleichgültigkeitsbedingung“ eingesetzt. Einströmen über solch ein Randstück, d. h. $v(x) \cdot n(x) < 0$, deutet meist auf fehlerhafte Modellbildung hin[3].

Beim Übergang zur schwachen Formulierung tritt außerdem das Randintegral $\oint_{\Gamma_N} npv_{\text{Test}} \, \mathrm{d}\sigma x$ auf, das verhindert, daß die Stokes-Gleichungen ein Sattelpunktproblem ergeben.

- *Ausströmungs-Randbedingungen* werden auf Randstücken verwendet, wo das Fluid aus dem Gebiet herausströmen soll (Outflow). Homogene Bedingungen dieser Art beschreiben Ausströmen in Vakuum, inhomogene Bedingungen das Ausströmen gegen einen vorgegebenen Druck (z. B. atmosphärischen Luftdruck). Wie bei Neumann-Randbedingungen weist Einströmen über einen solchen Rand auf Probleme im Modell hin. Die physikalisch korrekte (homogene) Bedingung lautet eigentlich $Tn = 0$ auf Γ_A, d. h. die Randbedingung 3 aus Definition 1.4 ist eine Näherung für den Fall, daß die Normalkomponente $\tilde{v}$ von v $\mathrm{D}\tilde{v} \approx 0$ erfüllt, denn dann ist $2\,\mathrm{Sym}(\,\mathrm{d}v)n \approx \frac{\partial v}{\partial n}$. Insofern dürfen solche Ränder nicht zu nah an den Bereichen des Modells liegen, die genau untersucht werden sollen.

  Ein Vorteil dieser Randbedingung liegt darin, daß sie auf natürliche Weise zu der schwachen Formulierung als Sattelpunktaufgabe paßt. Dazu findet man in [28] umfassende Betrachtungen.

- *Gleit-Randbedingungen* werden zur Modellierung von Phasengrenzen verwendet, die das Fluid nicht durchdringen kann und wo der Geschwindigkeitsgradient in Normalenrichtung keine tangentialen Spannungen erzeugt (Reibungsfreiheit). Es handelt sich um eine Näherung für schnell fließende Fluide mit geringer Zähigkeit an einer festen Berandung.

Die Beschreibung eines schnell fließenden Fluids geringer Zähigkeit in der Nähe von festen Berandungen stellt eine schwierige Aufgabe dar: Die vermutlich physikalisch korrekten Haft-Randbedingungen erfordern eine sehr feine Diskretisierung in Randnähe, da die Geschwindigkeit des Fluids sich rapide ändert. Andererseits sind die Gleit-Randbedingungen bei hohen Geschwindigkeiten eventuell eine zu große Vereinfachung. Zu sogenannten Wandgesetzen im Zusammenhang mit Turbolenzmodellen existieren viele Ansätze, deren Beschreibung an dieser Stelle zu weit führt.
An der Phasengrenze zwischen zwei Fluiden wird normalerweise folgendes Modell verwendet:

[3] $\frac{\partial u}{\partial n} = 0$ besagt, daß sich v in Richtung n nicht ändert – diese Aussage ergibt auf Inflowrändern keinen Sinn, weil sie im Grunde keine Strömung auf dem Rand vorschreibt.

1. Die Geschwindigkeiten am Rand gehen stetig ineinander über[4], also $v_1 - v_2 = 0$ auf Γ.

2. Auf Γ herrscht ein Kräftegleichgewicht zwischen der Normalkraft $\gamma(\frac{1}{R_1} + \cdots + \frac{1}{R_{n-1}})n$ aufgrund der Oberflächenspannung, wobei R_i $(i = 1, \ldots, n-1)$ die Hauptkrümmungsradien von Γ sind, und den Normalkräften aus den Fluidspannungen:

$$(T_1 - T_2)n = \gamma(\frac{1}{R_1} + \cdots + \frac{1}{R_{n-1}})n.$$

Dabei ist $\operatorname{sgn}(R_i) = 1$, wenn n auf die Seite von Γ zeigt, auf der der zugehörige Krümmungsmittelpunkt liegt.

Diese $2n$ Gleichungen beschreiben die je n Komponenten v_1 und v_2 auf Γ.

A.8 Dynamische Ähnlichkeit

Wenn Strömungsaufgaben experimentell untersucht werden, muß oft auf verkleinerte Modelle zurückgegriffen werden (z. B. bei Staudämmen, bei der Umströmung von Brückenauflagern). Es stellt sich die Frage, wie sich die Lösungen der Navier-Stokes-Gleichungen unter einer Skalierung der Parameter ρ, μ und der Maßstäbe (Skalen) für Länge L und Geschwindigkeit U verhalten. Der Einfachheit halber werden die stationären Gleichungen mit dem Kraftterm $f \equiv 0$ untersucht. Die Navier-Stokes-Gleichungen können durch folgende Transformation in dimensionslose Gleichungen umgewandelt werden:

$$x' = \frac{x}{L}, \quad t' = \frac{V}{L}t, \quad v'(x') = \frac{v(x)}{V}, \quad p'(x') = \frac{p(x)}{\rho V^2}. \tag{A.35}$$

Aufgrund dieser Regeln können die Differentialoperatoren in (A.34) umgeformt werden:

$$\begin{aligned} \mathrm{D}_x v(x) &= \mathrm{D}_x\big(V v'(\frac{x}{L})\big) = \frac{V}{L}\,\mathrm{D}_{x'} v'(x'), \\ \Delta_x v(x) &= \mathrm{D}_x\,\mathrm{D}_x\big(v(x)\big) = \frac{V}{L^2}\Delta_{x'} v'(x'), \quad \text{u. s. w.} \end{aligned}$$

Dies führt mit der Bezeichnung $R = \frac{\rho L V}{\mu}$ zu den transformierten Navier-Stokes-Gleichungen

$$\begin{aligned} \mathrm{D}_{x'} v'(x') &= 0, \\ (v'\,\mathrm{D}_{x'})v' &= -\,\mathrm{D}_{x'} p' + \frac{1}{R}\Delta_{x'} v'. \end{aligned} \tag{A.36}$$

[4]Das bedeutet, daß sich die Fluide nicht voneinander lösen.

Die dimensionslose Zahl R heißt *Reynoldszahl*[5] der Strömungsaufgabe. Haben Strömungsaufgaben auf geometrisch ähnlichen Gebieten in der dimensionslosen Form die gleichen Randbedingungen und die gleiche Reynoldszahl, so heißen sie *dynamisch ähnlich*. Lösungen solcher Aufgaben stehen via (A.35) zueinander in Korrespondenz, was den Einsatz von maßstäblich veränderten Modellen zur experimentellen Untersuchung von Strömungsaufgaben ermöglicht.
Physikalisch gesehen beschreibt die Reynoldszahl das Verhältnis der Kontaktkräfte zu den Inertialkräften, die auf das Fluid wirken. Insbesondere sind also die Stokes-Gleichungen, die in dieser Arbeit untersucht werden, physikalisch nur für Flüsse mit $R \ll 1$ anwendbar.

Bemerkung A.31. Es gibt noch eine Reihe weiterer dimensionsloser Parameter, die bei Strömungsexperimenten mit von Null verschiedenen Krafttermen berücksichtigt werden müssen (nach Lame, Prandtl, Strouhal und anderen). Die Wirkung eines konstanten Gravitationsfeldes der Stärke $f(x) \equiv \rho g$ hat zum Beispiel auf skalierte Modelle unterschiedlichen Einfluß. Beim Übergang von (A.34) zu (A.36) wird aus $f(x) = \rho g$ der Term $f'(x') = \frac{L}{V^2} g$. Dessen Kehrwert heißt *Froudezahl* und muß zusätzlich zur Reynoldszahl bei zwei Strömungsaufgaben übereinstimmen, wenn auch Gravitationseffekte berücksichtigt werden sollen.

[5] nach O. Reynolds, 1883; G. G. Stokes, 1851

Literaturverzeichnis

[1] Adams, Robert A.: *Pure and Applied Mathematics Series*. Bd. 65: *Sobolev Spaces*. Erste Auflage. 24/28 Oval Road, London NW 1, United Kingdom : Academic Press, 1975. – ISBN 0-12-044150-0

[2] Agmon, S. ; Douglis, A. ; Nirenberg, L.: Estimates Near the Boundary for Solutions of Elliptic Partial Differential Equations Satisfying General Boundary Conditions I. In: *Communications on Pure and Applied Mathematics* 12 (1959), S. 623–727

[3] Agmon, S. ; Douglis, A. ; Nirenberg, L.: Estimates Near the Boundary for Solutions of Elliptic Partial Differential Equations Satisfying General Boundary Conditions II. In: *Communications on Pure and Applied Mathematics* 17 (1964), S. 35–92

[4] Alt, Hans W.: *Lineare Funktionalanalysis*. Dritte Auflage. Springer-Verlag, 1999 (Springer Lehrbuch). – ISBN 3-540-65421-6

[5] Apel, Thomas: Interpolation of Non-Smooth Functions on Anisotropic Finite Element Meshes / Technische Universität Chemnitz-Zwickau, Sonderforschungsbereich 393 — Numerische Simulation auf massiv parallelen Rechnern. URL `http://www.tu-chemnitz.de/sfb393/`, März 1997 (97-06). – Preprint-Reihe des Chemnitzer SFB 393.

[6] Babuška, I. ; Rheinboldt, W. C.: A Posteriori Error Estimates for the Finite Element Method. In: *Int. J. Numer. Meth. Engrg.* 12 (1978), S. 1597–1615

[7] Bänsch, Eberhard ; Siebert, Kunibert G.: A Posteriori Error Estimation for Nonlinear Problems by Duality Techniques / Albert-Ludwigs-Universität Freiburg, Institut für Angewandte Mathematik. URL `http://web.mathematik.uni-freiburg.de/preprints/`, Dezember 1995 (95-30). – Preprintserie der mathematischen Fakultät.

[8] Batchelor, G. K.: *An Introduction to Fluid Dynamics*. Achte Auflage. Cambridge University Press, 1985. – ISBN 0-5210-09817-3

[9] Becker, Roland ; Rannacher, Rolf: A Feed-Back Approach to Error Control in Finite Element Methods: Basic Analysis and Examples / Universität Heidelberg. URL `http://www.iwr.uni-heidelberg.de/`, 1996 (1996-52). – Preprint.

[10] Becker, Roland ; Rannacher, Rolf: An Optimal Control Approach to A Posteriori Error Estimation in Finite Element Methods / Universität Heidelberg. URL `http://www.iwr.uni-heidelberg.de/sfb359/`, 2001 (2001-14). – Preprint des Sonderforschungsbereichs 359.

[11] Bey, Jürgen: *Finite-Volumen- und Mehrgitterverfahren für elliptische Randwertprobleme*, Eberhard-Karls-Universität Tübingen, Mathematische Fakultät, Dissertation, 1997

[12] Bey, Jürgen: Simplicial Grid Refinement: On Freudenthal' s Algorithm and the Optimal Number of Congruence Classes / Rheinisch-Westfälische Technische Hochschule Aachen, Institut für Geometrie und Praktische Mathematik. Februar 1998 (151). – IGPM-Report.

[13] Böhm, Walter: *Abschätzungen für instationäre, kompressible Navier-Stokes-Gleichungen bei gemischten Randwerten*, Saarbrücken, Diplomarbeit, 1989

[14] Bramble, James H.: *Pitman Research Notes in Mathematics Series*. Bd. 294: *Multigrid methods*. Zweite Auflage. Longman Scientific & Technical, 1995. – ISBN 0-582-23435-2

[15] Ciarlet, Jacques L. (Hrsg.): *Handbook of Numerical Analysis*. Bd. II: *Finite Element Methods (Part 1)*. Erste Auflage. Elsevier Science Publishers B. V. (North Holland), 1991. – ISBN 0-444-70365-9

[16] Ciarlet, Philippe G.: *Studies in Mathematics and Its Applications*. Bd. 4: *The Finite Element Method for Elliptic Problems*. Erste Auflage. North-Holland Publishing Company, 1978. – ISBN 0-444-85028-7

[17] Courant, Richard ; Hilbert, David: *Heidelberger Taschenbücher*. Bd. 30: *Methoden der Mathematischen Physik I*. Dritte Auflage. Springer-Verlag, 1968. – ISBN 3-540-04177-X

[18] Dörfler, W.: A Convergent Adaptive Algorithm for Poisson's Equation. In: *SIAM J. Numer. Anal.* 33 (1996), S. 1106–1124

[19] Evans, Lawrence C.: *Graduate Studies in Mathematics*. Bd. 19: *Partial Differential Equations*. Erste Auflage. American Mathematical Society, 1998. – ISBN 0-8218-0772-2

[20] Galdi, Giovanni P.: *Springer Tracts in Natural Philosophy*. Bd. 38: *An Introduction to the Mathematical Theory of the Navier-Stokes Equations, Volume I, Linearized Steady Problems*. Erste Auflage. Springer-Verlag, 1994. – ISBN 3-540-94172-X

[21] Galdi, Giovanni P.: *Springer Tracts in Natural Philosophy*. Bd. 39: *An Introduction to the Mathematical Theory of the Navier-Stokes Equations, Volume II, Nonlinear Steady Problems*. Erste Auflage. Springer-Verlag, 1994. – ISBN 3-540-94150-9

[22] Gehrtsen, Christian ; Vogel, Helmut: *Physik*. Achtzehnte Auflage. Springer-Verlag, 1995 (Springer Lehrbuch). – ISBN 3-540-59278-4

[23] Girault, Vivette ; Raviart, Pierre-Arnaud: *Springer Series in Computational Mathematics*. Bd. 5: *Finite Element Methods for Navier-Stokes Equations, Theory and Algorithms*. Erste Auflage. Springer-Verlag, 1986. – ISBN 3-540-15796-4

[24] Gross, Sven: *Parallelisierung eines adaptiven Verfahrens zur numerischen Lösung partieller Differentialgleichungen*, Rheinisch-Westfälische Technische Hochschule Aachen, Diplomarbeit, 2002

[25] Gross, Sven ; Peters, Jörg ; Reichelt, Volker ; Reusken, Arnold: The DROPS Package for Numerical Simulations of Incompressible Flows Using Parallel Adaptive Multigrid Techniques / Rheinisch-Westfälische Technische Hochschule Aachen, Institut für Geometrie und Praktische Mathematik. Februar 2002 (211). – IGPM-Report.

[26] Grossmann, Christian ; Roos, Hans-Görg: *Numerik partieller Differentialgleichungen*. Zweite Auflage. Teubner-Verlag, 1994 (Teubner Studienbücher Mathematik). – ISBN 3-519-12089-5

[27] Hackbusch, Wolfgang: *Theorie und Numerik elliptischer Differentialgleichungen*. Zweite Auflage. Teubner-Verlag, 1996 (Teubner Taschenbücher Mathematik). – ISBN 3-519-12074-7

[28] Heywood, John G. ; Rannacher, Rolf ; Turek, Stefan: Artificial Boundaries and Flux and Pressure Conditions for the Incompressible Navier-Stokes Equations,. In: *Int. J. Numer. Meth. Fluids* 22 (1996), S. 325–352. – URL `http://www.featflow.de/ture/papers.html`

[29] Köngeter, Jürgen ; Forkel, Christian: *Hydromechanik III*. Lehrstuhl und Institut für Wasserbau und Wasserwirtschaft, Rheinisch-Westfälische Technische Hochschule Aachen, 1999

[30] Lide, David R. (Hrsg.): *Handbook of Chemistry and Physics*. 76. Auflage. CRC Press, 1995. – ISBN 0-8493-0597-7

[31] Scott, L. R. ; Zhang, Shangyou: Finite Element Interpolation of Nonsmooth Functions Satisfying Boundary Conditions. In: *Mathematics of Computation* 54 (1990), Nr. 190, S. 482–493

[32] Stenberg, Rolf: A Technique for Analyzing Finite Element Methods for Incompressible Viscous Flow. In: *International Journal for Numerical Methods in Fluids* 11 (1990), S. 935–948. – URL `http://www.math.hut.fi/~rstenber/`

[33] Stöcker, Heiner: *Algebraische Topologie*. Zweite Auflage. B. G. Teubner, 1994 (Mathematische Leitfäden). – ISBN 3-519-12226-X

[34] Turek, Stefan: Multigrid techniques for a divergence-free finite element discretization. In: *East-West J. Numer. Math.* 2 (1994), Nr. 3, S. 229–255

[35] Turek, Stefan: Trends in Processor Technology and Their Impact on Numerics for PDE's / Ruprecht-Karls-Universität Heidelberg. URL `http://www.iwr.uni-heidelberg.de/sfb359/`, 1999 (1999-31). – Preprint des Sonderforschungsbereichs 359.

[36] Urban, Karsten: *Multiskalenverfahren fr das Stokes-Problem und angepaßte Wavelet-Basen*, Rheinisch-Westfälische Technische Hochschule Aachen, Dissertation, 1995

[37] Verfürth, Rüdiger: A Posteriori Error Estimators for the Stokes Equations. In: *Numerische Mathematik* 55 (1989), S. 309–325

[38] Verfürth, Rüdiger: A Posteriori Error Estimators for the Stokes Equations II non-conforming Discretizations. In: *Numerische Mathematik* 60 (1991), S. 235–249

[39] Verfürth, Rüdiger: *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. Erste Auflage. Wiley-Teubner, 1996 (Wiley-Teubner Series Advances in Numerical Mathematics). – ISBN 3-519-02605-8

[40] Verfürth, Rüdiger: On the Constants in Some Inverse Inequalities for Finite Element Functions / Ruhr-Universität Bochum, Lehrstuhl Mathematik XI. URL `http://www.ruhr-uni-bochum.de/num1/rv/`, Mai 1999. – Preprint.

[41] Waerden, Bartel L. van der: *Algebra I*. Neunte Auflage. Springer-Verlag, 1993. – ISBN 3-540-56799-2

[42] Wloka, Joseph: *Partielle Differentialgleichungen, Sobolevräume und Randwertaufgaben*. Erste Auflage. Teubner-Verlag, 1982 (Mathematische Leitfäden). – ISBN 3-519-02225-7

[43] Zienkiewiez, Olgierd C.: *Methode der finiten Elemente*. Zweite Auflage. Carl Hanser Verlag München, 1984. – ISBN 3-446-12525-6

[44] Zulehner, Walter: Analysis of Iterative Methods for Saddle Point Problems: A Unified Approach. In: *Mathematics of Computation* 71 (2001), S. 479–505