

VERGLEICH ZWEIER
FINITE-ELEMENT METHODEN FÜR
DIE STOKES GLEICHUNG

Diplomarbeit
Unter der Leitung von
Prof. Dr. Rüdiger Verfürth
Am Institut für Angewandte Mathematik
der Universität Zürich

Philippe Gressly

Mai 1993

Inhaltsverzeichnis

1	Zusammenfassung	1
2	Einleitung	2
2.1	Hintergrund	2
2.2	Das Problem	2
2.3	Regularitätsannahmen	3
3	Die Finite-Element Räume	3
3.1	Gradient der Geschwindigkeit	3
3.2	Variationsformulierung	4
3.3	Das Problem paßt in den abstrakten Rahmen	5
3.4	Dieses Problem hat eine eindeutige Lösung	8
4	Fehlerabschätzung	15
4.1	Behauptung	15
4.2	Beweis in fünf Schritten	15
5	Theorie zur Implementierung	25
5.1	Konventionen, Notationen	25
5.2	Ableitungen und Integrale	27
6	1.Programm	28
6.1	Vektorräume	28
6.2	Matrixschreibweise und Berechnung der Integrale	30
6.3	Implementierung	36
7	2.Programm	38
7.1	Vektorräume	39
7.2	Matrixschreibweise und Berechnung der Integrale	41
7.3	Implementierung	47

8 Vergleich der Programme	47
8.1 Speicherplatz	47
8.2 Geschwindigkeit und Genauigkeit	48
8.3 Auswertung	63
A Anhang	64
A.1 Lemmata	64
A.2 Notationen	72

Danksagung

Ganz herzlich bedanken möchte ich mich an dieser Stelle bei Prof. Dr. R. Verfürth für die Themengebung und dafür, daß er mir bei allen Problemen jederzeit geholfen hat.

Ein spezieller Dank geht auch an die Assistenten Carmen Heusser und Hermann Weber, die mich immer mit Rat und Tat unterstützten.

1 Zusammenfassung

Rolf Stenbergs [8] neue Finite-Elemente zur numerischen Lösung der Stokes-Gleichung (2.2) werden in das Programm **FEM-FLOW** [9] eingebaut und getestet.

In einem ersten Teil (Kapitel 2, 3, 4) wird die Theorie der Elemente noch einmal angegeben. Kapitel 2 und 3 befassen sich mit der Existenz der Lösung. Kapitel 4 gibt eine Fehlerabschätzung an.

Das Programm **FEM-FLOW** [9] stellt die Triangulierung und Verfeinerung von Gebieten $\Omega \subset \mathbb{R}^2$ zur Verfügung. Die Vektorräume für die Finiten-Elemente werden also auch auf den zweidimensionalen Fall ausgerichtet sein.

In diesen Finite-Element Räumen wählen wir für den Grad der Polynome $k = 1$; das bedeutet stückweise linearer Druck und stückweise konstante Geschwindigkeit. Der Gradient der Geschwindigkeit wird auch als Unbekannte eingeführt und wird mit $k = 1$ stückweise linear.

Stenberg [8] schlägt am Schluß seiner Arbeit eine Hybridisierungstechnik von Fraijs de Veubeke [10] [11] vor. Diese Arbeit befaßt sich damit, inwiefern sich die Hybridisierungstechnik im Falle $k = 1$ lohnt. Es werden zwei Programme geschrieben (mit und ohne Hybridisierung) und miteinander verglichen. Die Vektorräume für die beiden Finite-Element Methoden finden sich in Kapitel 6 (ohne Hybridisierung) und Kapitel 7 (mit Hybridisierung).

Die beiden Programme werden auf Problemen mit bekannter Lösung getestet. Das Kapitel 8 beinhaltet eine Zusammenstellung der Resultate. Dabei wurde auf Speicherplatz und Konvergenzgeschwindigkeit geachtet. Um einen *gerechten* Vergleich zu machen, wurde in beiden Programmen auf eine Vorkonditionierung verzichtet. Es wurde lediglich so skaliert, daß die Matrixeinträge bei der Verfeinerung ihre Größenordnung nicht ändern.

Es wird sich zeigen, daß sich die Hybridisierungstechnik im linearen Fall ($k = 1$) nicht lohnt. Man braucht etwas ($\approx 18\%$) mehr Speicherplatz und daher sind die Gleichungssysteme größer, was sich negativ auf die Effizienz auswirkt.

2 Einleitung

2.1 Hintergrund

Bei vielen industriellen Anwendungen kommen Strömungsprobleme vor. Die Physik stellt zu ihrer Lösung Differentialgleichungen auf.

Eine dieser Strömungsgleichungen ist die Stokes Gleichung (2.2).

Für ein Gebiet Ω sind Randwerte u_0 und äußere Kräfte f gegeben. Für konstante Dichte ρ kann die Massenerhaltung ($\partial_t \rho + \nabla \cdot (\rho u) = 0$) geschrieben werden als $\nabla \cdot u = 0$.

Gesucht sind dann zwei Funktionen u bzw. p , für die Geschwindigkeit bzw. den (mathematischen) Druck. Als numerische Resultate können explizit Funktionen u_h und p_h angegeben werden, die u und p *genügend* gut approximieren.

2.2 Das Problem

Gegeben ist ein abgeschlossenes polygonales Gebiet Ω in \mathbb{R}^2 . Ω ist unterteilt in Dreiecke. Wir bezeichnen mit \mathcal{C}_h die Menge aller Dreiecke und mit Γ_h die Menge der Kanten der Elemente. Wir bezeichnen mit T die Dreiecke aus \mathcal{C}_h und mit e die Kanten aus Γ_h .

Dabei sei \mathcal{C}_h regulär, d.h. es gibt eine Konstante σ , mit welcher für jede Verfeinerung von \mathcal{C}_h noch gilt

$$\frac{h_T}{\rho_T} \leq \sigma. \quad (2.1)$$

Hier bezeichnet h_T die längste Kante von T und ρ_T den Inkreisradius.

In der ganzen Arbeit ist h definiert als das Maximum über alle h_T .

Auf dem Gebiet Ω sei nun folgende Differentialgleichung (Stokes) gegeben:

$$\begin{array}{rcl} -\Delta u + \nabla p & = & f \quad \text{in } \Omega \\ \nabla \cdot u & = & 0 \quad \text{in } \Omega \\ u & = & u_0 \quad \text{auf } \partial\Omega \end{array} \quad (2.2)$$

Gegeben sind f und die Randbedingung u_0 . Gesucht ist die Geschwindigkeit u und der Druck p .

2.3 Regularitätsannahmen

Von f wird verlangt, daß es eine Funktion aus $L^2(\Omega)^2$ ist. Dabei bezeichnet $L^2(\Omega)^2$ die Menge aller Funktionen von Ω nach \mathbb{R}^2 mit der Eigenschaft, daß ihr Quadrat noch integrierbar ist.

Die Randfunktion u_0 sei aus $H^{1/2}(\partial\Omega)^2 := \{u_0 \in L^2(\Gamma)^2 : \exists v \in H^1(\Omega)^2 : v|_\Gamma = u_0\}$.

Für u_0 gelte noch

$$\int_{\partial\Omega} u_0 \cdot n \, ds = 0.$$

Der Druck p sei über Ω normiert. Das heißt:

$$\int_{\Omega} p = 0$$

Für eine optimale Fehlerabschätzung (Kapitel 4) sei noch die Annahme getroffen, daß Ω ein konvexes Gebiet ist. Das impliziert dann, daß aus $u_0 = 0$ in Gleichung (2.2) die Abschätzung

$$\|u\|_2 + \|p\|_1 \leq C \|f\|_0 \tag{2.3}$$

folgt.

3 Die Finite-Element Räume

3.1 Gradient der Geschwindigkeit

Wir führen den Gradienten der Geschwindigkeit als neue Variable ein.

$$\sigma := \nabla u$$

Somit liest sich unsere Gleichung (2.2) nun:

$$\begin{aligned} \sigma - \nabla u &= 0 && \text{in } \Omega \\ -\nabla \cdot \sigma + \nabla p &= f && \text{in } \Omega \\ \nabla \cdot u &= 0 && \text{in } \Omega \\ u &= u_0 && \text{auf } \partial\Omega \end{aligned} \tag{3.1}$$

Dabei ist:

$$\begin{aligned} \sigma &= \{\sigma_{ij}\}, \sigma_{ij} = \partial_j u_i \\ (\nabla \cdot \sigma)_i &= \sum_{j=1}^2 \partial_j \sigma_{ij} \quad i = 1, 2 \\ \nabla \cdot u &= \sum_{j=1}^2 \partial_j u_j \\ (\nabla u)_{ij} &= \partial_j u_i \end{aligned}$$

3.2 Variationsformulierung

Aus der modifizierten Stokes-Gleichung (3.1) erhalten wir durch Multiplikation von rechts mit τ, v bzw. q , Integration über Ω und Anwenden des Satzes von Gauß (A.1.5 S. 67) folgende Variationsformulierung:

$$\begin{aligned} (\sigma, \tau) + (\nabla \cdot \tau, u) &= \langle u_0, \tau \cdot n \rangle \\ -(\nabla \cdot \sigma, v) + (\nabla p, v) &= (f, v) \\ (u, \nabla q) &= \langle u_0 \cdot n, q \rangle \end{aligned} \tag{3.2}$$

Das gilt auch noch in den diskreten Räumen. Dort lautet die Variationsformulierung: Finde $(\sigma_h, u_h, p_h) \in H_h \times V_h \times P_h$, sodaß

$$\begin{aligned} \forall \tau \in H_h: \quad & (\sigma_h, \tau) + (\nabla \cdot \tau, u_h) = \langle u_0, \tau \cdot n \rangle \\ \forall v \in V_h: \quad & -(\nabla \cdot \sigma_h, v) + (\nabla p_h, v) = (f, v) \\ \forall q \in P_h: \quad & (u_h, \nabla q) = \langle u_0 \cdot n, q \rangle . \end{aligned} \tag{3.3}$$

Dabei sind (für $k \geq 1$) H_h , V_h und P_h folgendermaßen definiert:

$$\begin{aligned} H_h &:= \{\tau \in H(\nabla \cdot, (\Omega)) : \tau|_T \in P_k(T)^{2 \times 2}, T \in \mathcal{C}_h\} \\ V_h &:= \{v \in L^2(\Omega)^2 : v|_T \in P_{k-1}(T)^2, T \in \mathcal{C}_h\} \\ P_h &:= \{p \in C(\Omega) \cap L_0^2(\Omega) : p|_T \in P_k(T), T \in \mathcal{C}_h\} \\ H(\nabla \cdot, (\Omega)) &:= \{\tau \in L^2(\Omega)^{2 \times 2} : \nabla \cdot \tau \in L^2(\Omega)^2\}. \end{aligned}$$

In der Implementierung werden wir $k = 1$ wählen. Die Räume $P_k(T)$ bezeichnen Polynome $\phi : T \rightarrow \mathbb{R}$ vom Grad $\leq k$.

3.2.1 Bessere Geschwindigkeitsapproximation

Falls wir nun u_h, σ_h, p_h als Lösung von (3.3) berechnet haben, können wir mit folgender Technik noch eine bessere Approximation u_h^* an u erreichen:

$$V_h^* := \{v \in L^2(\Omega)^2 : \forall T \in \mathcal{C}_h : v|_T \in P_{k+1}(T)^2\}$$

Wir berechnen eine neue Approximation $u_h^* \in V_h^*$ an die Geschwindigkeit elementweise auf

jedem Dreieck $T \in \mathcal{C}_h$, indem wir folgendes Problem lösen:

$$\begin{aligned} \forall v \in (I - Q_h)V_h^*|_T: \quad (\nabla u_h^*, \nabla v)_T &= (\sigma_h, \nabla v)_T \\ \forall T \in \mathcal{C}_h \quad \quad \quad &: \quad Q_h u_h^*|_T = u_h|_T \end{aligned} \tag{3.4}$$

Dabei bezeichne $(\cdot, \cdot)_T$ das Skalarprodukt in $L^2(T)^{2 \times 2}$. Also $(f, g)_T := \int_T f \cdot g$.

Die Projektion $Q_h : L^2(\Omega)^2 \rightarrow V_h$ sei folgendermaßen definiert: $\forall \phi \in V_h$ gelte $(Q_h f, \phi) = (f, \phi)$; d.h. Q_h ist die L^2 -Projektion. *Bemerkung:* u und u^* haben den selben Laplace-Operator. Das ganze ist ein lokales Hilfsproblem, um eine bessere Approximation an u aus ∇u zu bekommen.

3.3 Das Problem paßt in den abstrakten Rahmen

Brezzi [2] behandelt abstrakte Sattelpunktsprobleme. Stenberg versucht seine Elemente im Rahmen dieser Sattelpunktsprobleme zu behandeln. Dazu (Eindeutigkeit der Lösung, Fehlerabschätzung von (3.3)) verwenden wir netzabhängige Normen. Seien also:

$$\| \sigma \|_{0,h}^2 := \| \sigma \|_0^2 + \sum_{e \in \Gamma_h} h_e \int_e |\sigma \cdot n|^2$$

für $\sigma \in L^2(\Omega)^{2 \times 2}$, sodass $\forall e \in \Gamma_h$ gilt $\sigma \cdot n \in L^2(e)^2$

$$\| u \|_{1,h}^2 := \sum_{T \in \mathcal{C}_h} \| \nabla u \|_{0,T}^2 + \sum_{e \in \Gamma_h} h_e^{-1} \int_e |[u]_e|^2$$

für $u \in L^2(\Omega)^2$, sodass $\forall T \in \mathcal{C}_h$ gilt $u|_T \in H^1(T)^2$

und

$$\| \sigma, u \|_h^2 := \| \sigma \|_{0,h}^2 + \| u \|_{1,h}^2$$

Hier bezeichnet n den Einheitsnormalenvektor an eine Kante e . Mit $[u]_e$ bezeichnen wir den Sprung von u über die entsprechende Kante. Liegt die Kante auf $\partial\Omega$, so bezeichnet $[u]_e$ jedoch den Wert von u selbst.

Nach dem Satz von Gauß und elementweiser Integration auf den Dreiecken erhalten wir folgende Ungleichung:

$$(\nabla \cdot \sigma, u) \leq \| \sigma \|_{0,h} \| u \|_{1,h} \tag{3.5}$$

für $\| \sigma \|_{0,h} < \infty$ und $\| u \|_{1,h} < \infty$.

Beweis

Die Höldersche Ungleichung lautet:

$$\sum_i a_i b_i \leq \sqrt{(\sum_i a_i^2)(\sum_i b_i^2)} \quad (*)$$

$$\begin{aligned}
(\nabla \cdot \sigma, u)_T &= \int_T \nabla \cdot \sigma \cdot u \\
&\stackrel{(A.1.5)}{=} - \int_T \sigma \cdot \nabla u + \int_{\partial T} u \cdot (\sigma \cdot n) \\
&\leq \left| \int_T \sigma \cdot \nabla u \right| + \sum_{e \in \partial T} \int_e |h_e^{-\frac{1}{2}} u \cdot h_e^{\frac{1}{2}} \sigma \cdot n| \\
&\stackrel{(*)}{\leq} \int_T |\sigma \cdot \nabla u| + \sqrt{(\sum_e \int_e |h_e^{-\frac{1}{2}} u|^2) \cdot (\sum_e \int_e |h_e^{\frac{1}{2}} \sigma \cdot n|^2)} \\
&\leq \|\sigma\|_{0,T} \cdot \|\nabla u\|_{0,T} + \sqrt{(\sum_e \int_e |h_e^{-\frac{1}{2}} u|^2) \cdot (\sum_e \int_e |h_e^{\frac{1}{2}} \sigma \cdot n|^2)} \\
&\stackrel{(*)}{\leq} \sqrt{\|\sigma\|_{0,T}^2 + \sum_e h_e \int_e |\sigma \cdot n|^2} \cdot \sqrt{\|\nabla u\|_{0,T}^2 + \sum_e h_e^{-1} \int_e |u|^2} \\
&= \|\sigma\|_{0,h} \cdot \|u\|_{1,h}.
\end{aligned}$$

□

Nach [1] gelten die Ungleichungen:

$$\inf_{\tau \in H_h} \|\sigma - \tau\|_{0,h} \leq Ch^r |\sigma|_r \quad 1/2 < r \leq k+1, \quad (3.6)$$

$$\inf_{v \in V_h} \|u - v\|_{1,h} \leq Ch^{s-1} |u|_s \quad 1 \leq s \leq k. \quad (3.7)$$

Beweis

Babuška [1] zeigt dieselben Abschätzungen mit stärkeren Normen.

Zudem sind die beiden Normen $\|\cdot\|_0$ und $\|\cdot\|_{0,h}$ äquivalent:

$$\exists C > 0 : \forall \sigma \in H_h : \|\sigma\|_0 \leq \|\sigma\|_{0,h} \leq C \|\sigma\|_0. \quad (3.8)$$

Beweis

Bezeichne mit Φ eine affine Abbildung, die das Referenzdreieck $\hat{T} := \{x \in \mathbb{R}^2 : x_1, x_2 \geq 0, x_1 + x_2 \leq 1\}$ auf das Dreieck T ($:= \Phi(\hat{T})$) abbildet. Es spielt in diesem Beweis keine Rolle, welche Ecke von \hat{T} auf welche Ecke von T abgebildet wird. φ sei die Einschränkung von Φ auf die Dreiecksseite $\hat{e} := [0, 1]$. Zudem bezeichne $e := \varphi(\hat{e})$ die entsprechende Dreiecksseite von T .

Wie üblich definiert, ist $g(t)$ die Gramsche Determinante bezüglich $\varphi(t)$. In unserem Fall ist $g = \det(D\varphi^t D\varphi) = \nabla\varphi^t \nabla\varphi = |\nabla\varphi|^2$. Es ist bekannt, daß

$$\int_e f = \int_{\hat{e}} (f \circ \varphi) \cdot \sqrt{g} = \int_{\hat{e}} (f \circ \varphi) |\nabla\varphi|.$$

Weil φ affin ist, so ist $\nabla\varphi$ konstant und es gilt $|\nabla\varphi| = \frac{|e|}{|\hat{e}|} = |e| = ch$:

$$\int_e f = ch \int_{\hat{e}} f \circ \varphi =: ch \int_{\hat{e}} \hat{f}$$

Danach gilt jetzt

$$\begin{aligned} \sum_{e \in \partial T} h_e \int_e |\sigma \cdot n|^2 &\leq C \sum_{\hat{e} \in \hat{T}} h_e^2 \int_{\hat{e}} |\hat{\sigma} \cdot \hat{n}|^2 \\ &\leq Ch_T^2 \|\hat{\sigma}\|_{0, \hat{T}}^2 \\ &\stackrel{[5]}{\leq} C \|\sigma\|_{0, T}^2 \end{aligned} \tag{3.9}$$

$$\begin{aligned} \Rightarrow \sum_{e \in \Gamma_h} h_e \int_e |\sigma \cdot n|^2 &\leq \sum_{T \in \mathcal{C}_h} \sum_{e \in \partial T} h_e \int_e |\sigma \cdot n|^2 \\ &\stackrel{(3.9)}{\leq} C \sum_{T \in \mathcal{C}_h} \|\sigma\|_{0, T}^2 \\ &\leq C \|\sigma\|_0^2 \end{aligned} \tag{3.10}$$

$$\begin{aligned} \Rightarrow \|\sigma\|_{0, h}^2 &= \|\sigma\|_0^2 + \sum_{e \in \Gamma_h} h_e \int_e |\sigma \cdot n|^2 \\ &\stackrel{(3.10)}{\leq} \|\sigma\|_0^2 + C \|\sigma\|_0^2 \\ &\leq C \|\sigma\|_0^2. \end{aligned}$$

□

Um das Problem (3.3) mit Brezzis Theorie der Sattelpunktsprobleme zu lösen (vgl. [2]), definieren wir zwei bilineare Abbildungen:

$$a(\sigma, u; \tau, v) := (\sigma, \tau) + (\nabla \cdot \tau, u) - (\nabla \cdot \sigma, v)$$

und

$$b(\tau, v; p) := (v, \nabla p)$$

Damit schreibt sich die Variationsformulierung (3.3) als:

Finde $(\sigma_h, u_h, p_h) \in H_h \times V_h \times P_h$, sodaß gilt

$$\begin{aligned} \forall (\tau, v) \in H_h \times V_h \quad a(\sigma_h, u_h; \tau, v) + b(\tau, v; p_h) &= (f, v) + \langle u_0, \tau \cdot n \rangle \\ \forall q \in P_h \quad b(\sigma_h, u_h; q) &= \langle u_0 \cdot n, q \rangle. \end{aligned} \tag{3.11}$$

3.4 Dieses Problem hat eine eindeutige Lösung

Dieses Problem (Gleichung 3.11) hat für *jede* rechte Seite eine eindeutige Lösung. (Die Eindeutigkeit wird in (3.4.3) gegeben.) Zum Beweis, daß das abstrakte Problem eindeutig lösbar ist, müssen wir zuerst zwei Lemmata beweisen. Zuerst eine Definition:

$$Z_h := \{(\sigma, u) \in H_h \times V_h : \forall q \in P_h : b(\sigma, u; q) = 0\}$$

3.4.1 Sup-Bedingungen für a

Es gelten die beiden Abschätzungen:

$$\exists C > 0 : \forall (\tau, v) \in Z_h : \sup_{\substack{(\sigma, u) \in Z_h \\ (\sigma, u) \neq 0}} \frac{a(\sigma, u; \tau, v)}{\|\sigma, u\|_h} \geq C \|\tau, v\|_h$$

und

$$\exists C > 0 : \forall (\sigma, u) \in Z_h : \sup_{\substack{(\tau, v) \in Z_h \\ (\tau, v) \neq 0}} \frac{a(\sigma, u; \tau, v)}{\|\tau, v\|_h} \geq C \|\sigma, u\|_h .$$

Beweis

zur ersten Abschätzung: Es genügt zu zeigen, daß

$$\begin{aligned} \exists C_1, C_2 > 0 \quad \forall (\tau, v) \in Z_h : \exists (\sigma, u) \in H_h \times V_h : \\ a(\sigma, u; \tau, v) \geq C_1 \|\tau, v\|_h^2 \end{aligned} \quad (3.12)$$

und

$$\|\sigma, u\|_h^2 \leq C_2 \|\tau, v\|_h \quad (3.13)$$

wobei C_1 und C_2 nicht von (τ, v) und (σ, u) abhängen.

Aus den beiden Gleichungen (3.12) und (3.13) folgt die erste Behauptung mittels Division der Gleichung (3.12) durch (3.13):

$$\Rightarrow \exists C > 0 : \forall (\tau, v) \in Z_h : \exists (\sigma, u) \in H_h \times V_h : \frac{a(\sigma, u; \tau, v)}{\|\sigma, u\|_h} \geq C \|\tau, v\|_h .$$

Um die beiden Ungleichungen (3.12,3.13) zu zeigen, sei jetzt $(\tau, v) \in Z_h$ beliebig.

Es ist bekannt [3] [4], daß (H_h, V_h) eine stabile Diskretisierung für den Laplace Operator liefert. Mit unseren netzabhängigen Normen kann die Stabilität folgendermaßen angegeben werden:

$$\exists C^* > 0 : \inf_{\substack{v \in V_h \\ v \neq 0}} \sup_{\substack{\gamma \in H_h \\ \gamma \neq 0}} \frac{(\nabla \cdot \gamma, v)}{\|v\|_{0,h}} \geq C^*$$

was äquivalent ist zu den Bedingungen:

$$\exists C^* > 0 : \forall v \in V_h : \exists \gamma \in H_h : (\nabla \cdot \gamma, v) \geq C^* \|v\|_{1,h}^2$$

und

$$\|\gamma\|_{0,h} \leq \|v\|_{1,h}.$$

(Den Beweis dieser Äquivalenz liefert Lemma (A.1.1 auf Seite 64).)

Wir definieren nun $(\sigma, u) := (\tau - \delta\gamma, v)$ mit γ wie oben und δ , sodaß $0 < \delta < 2\epsilon < 4C^*$. (ϵ ist damit auch definiert und wird gleich gebraucht.)

Zu (3.12):

$$\begin{aligned} a(\sigma, u; \tau, v) &= a(\tau - \delta\gamma, v; \tau, v) \\ &= \|\tau\|_0^2 - \delta(\gamma, \tau) + \delta(\nabla \cdot \gamma, v) \\ &\geq \|\tau\|_0^2 + \delta C \|v\|_{1,h}^2 - \delta \|\gamma\|_0 \|\tau\|_0 \\ &\geq \|\tau\|_0^2 + \delta C \|v\|_{1,h}^2 - \delta \|v\|_{1,h} \|\tau\|_0 \\ &\geq (1 - \frac{\delta}{2\epsilon}) \|\tau\|_0^2 + \delta(C - \frac{\epsilon}{2}) \|v\|_{1,h}^2 \\ &\geq C(\|\tau\|_0^2 + \|v\|_{1,h}^2) \\ &= C \|\tau\|_0^2 + C \|v\|_{1,h}^2 \\ &\geq C \|\tau\|_{0,h}^2 + C \|v\|_{1,h}^2 \\ &\geq C_1 \|\tau, v\|_h^2 \end{aligned}$$

Dabei gilt die erste Gleichung einfach wegen der Wahl von (σ, u) .

Die zweite Gleichheit gilt wegen der Definition der Linearform a .

Die erste Ungleichung folgt mit der Voraussetzung $(\nabla \cdot \gamma, v) \geq C^* \|v\|_{1,h}^2$ und der Hölderschen Ungleichung: $(\gamma, \tau) \leq \|\gamma\|_0 \|\tau\|_0$.

Die zweite Ungleichung gilt nach Voraussetzung $\|\gamma\|_{0,h} \leq \|v\|_{1,h}$ und aus der Tatsache, daß $\|\sigma\|_0 \leq \|\sigma\|_{0,h} \leq C \|\sigma\|_0$.

Die dritte Ungleichung folgt durch ausmultiplizieren und weil für $\epsilon > 0$ stets gilt:

$$\frac{1}{\epsilon} \frac{\|\tau\|_0}{\|v\|_{1,h}} + \epsilon \frac{\|v\|_{1,h}}{\|\tau\|_0} \geq 2$$

Die vierte Ungleichung folgt, da die beiden Vorfaktoren vor den Normen nach Voraussetzung an δ und ϵ echt größer als Null sind. Somit ist auch das Minimum größer als Null.

Die fünfte Ungleichung ist wahr wegen der Äquivalenz der $\|\cdot\|_0$ - und der $\|\cdot\|_{0,h}$ - Normen (vgl. 3.8).

Ein C_1 wie in der letzten Ungleichung existiert nach Definition der $\|\cdot, \cdot\|_h$ - Norm.

Daraus folgt die Behauptung (3.12).

Beweis der zweiten Unleichung (3.13):

$$\begin{aligned}
\| \sigma, u \|_h &\leq \| \sigma \|_{0,h} + \| u \|_{1,h} \\
&= \| \tau - \delta \gamma \|_{0,h} + \| v \|_{1,h} \\
&\leq \| \tau \|_{0,h} + \delta \| \gamma \|_{0,h} + \| v \|_{1,h} \\
&\leq \| \tau \|_{0,h} + \delta \| v \|_{1,h} + \| v \|_{1,h} \\
&= \| \tau \|_{0,h} + (\delta + 1) \| v \|_{1,h} \\
&\leq (\delta + 1)(\| \tau \|_{0,h} + \| v \|_{1,h}) \\
&\leq (\delta + 1)\sqrt{2} \| \tau, v \|_h \\
&= C_2 \| \tau, v \|_h .
\end{aligned}$$

□

Die zweite Abschätzung in Kapitel 3.4.1 folgt fast analog. Weil die Bilinearform $a(\cdot, \cdot)$ aber nicht symmetrisch ist, können (σ, u) und (τ, v) nicht einfach vertauscht werden. Sei also $(\sigma, u) \in Z_h$ beliebig. Wieder existieren γ und δ wie gehabt. Wir definieren jetzt $(\tau, v) := (\sigma + \delta\gamma, u)$. Wie im 1. Beweisteil gilt

$$\| \tau, v \|_h \leq C_2 \| \sigma, u \|_h$$

Aber bei der ersten Abschätzung ändert sich ein Vorzeichen:

$$\begin{aligned}
a(\sigma, u; \tau, v) &= a(\sigma, u; \sigma + \delta\gamma, u) \\
&\geq \| \sigma \|_0^2 + \delta(\gamma, \sigma) + \delta(\nabla \cdot \gamma, u)
\end{aligned}$$

Dies darf aber immer noch durch

$$\| \sigma \|_0^2 - \delta \| \gamma \|_0 \| \sigma \|_0 + \delta(\nabla \cdot \gamma, u)$$

nach unten abgeschätzt werden. Ab dieser Ungleichung geht die Analogie durch.

□

3.4.2 Sup-Bedingungen für b

Behauptung:

$$\exists C > 0 : \forall q \in P_h : \sup_{\substack{(\sigma, u) \in H_h \times V_h \\ (\sigma, u) \neq 0}} \frac{b(\sigma, u; q)}{\| \sigma, u \|_h} \geq C \| q \|_0$$

Beweis

Es genügt zu zeigen, daß

$$\exists C > 0 : \forall q \in P_h(\|q\|_0 = 1) : \sup_{\substack{(\sigma, u) \in H_h \times V_h \\ (\sigma, u) \neq 0}} \frac{b(\sigma, u; q)}{\|\sigma, u\|_h} \geq C \quad (3.14)$$

Die Behauptung folgt dann durch skalieren. Wir zeigen, daß

$$\forall q \in P_h(\|q\|_0 = 1) : \exists(\sigma, u) \in H_h \times V_h : \frac{b(\sigma, u; q)}{\|\sigma, u\|_h} \geq \tilde{C}_1 \eta_q \quad (3.15)$$

und

$$\frac{b(\sigma, u; q)}{\|\sigma, u\|_h} \geq 1 - \tilde{C}_2 \eta_q \quad (3.16)$$

Dabei hängt η_q wie folgt von q ab:

$$\eta_q := \sqrt{\sum_{T \in \mathcal{C}_h} h_T^2 \|\nabla q\|_{0,T}^2}$$

Aus (3.15) und (3.16) folgt

$$\sup_{\substack{(\sigma, u) \in H_h \times V_h \\ (\sigma, u) \neq 0}} \frac{b(\sigma, u; q)}{\|\sigma, q\|_h} \geq \max\{\tilde{C}_1 \eta_q, 1 - \tilde{C}_2 \eta_q\}.$$

Das Maximum von $\tilde{C}_1 \eta_q$ und $1 - \tilde{C}_2 \eta_q$ ist für alle positiven η_q größer als Null. Vergleiche dazu Abbildung (1) auf Seite 12:

Die Behauptung (3.14) folgt danach mit

$$C := \frac{\tilde{C}_1}{\tilde{C}_1 + \tilde{C}_2}.$$

Sei im folgenden $q \in P_h$ mit $\|q\|_0 = 1$ beliebig.

Beweis (von (3.15))

Für q gilt $\nabla q \in V_h$. Deshalb können wir eine Funktion u durch

$$u|_T := h_T^2 \nabla q|_T \quad \forall T \in \mathcal{C}_h$$

Abbildung 1: Das Maximum ist größer Null

definieren. Wir setzen nun $\sigma := 0$. Damit gilt

$$b(\sigma, u; q) = (u, \nabla q) = \sum_{T \in \mathcal{C}_h} h_T^2 \|\nabla q\|_{0,T}^2$$

Skalieren über das Referenzdreieck (vgl. Lemma (A.1.4) auf Seite 67) liefert die Ungleichung:

$$\|u\|_{1,h}^2 \leq C \sum_{T \in \mathcal{C}_h} h_T^{-2} \|u\|_{0,T}^2 = C \sum_{T \in \mathcal{C}_h} h_T^2 \|\nabla q\|_{0,T}^2$$

Da q beliebig war folgt, daß es für alle $q \in P_h$ ($\|q\|_0 = 1$) ein (σ, u) (nämlich $\sigma = 0$ und u wie oben) gibt, sodaß gilt:

$$\begin{aligned} \frac{b(\sigma, u; q)}{\|\sigma, u\|_h} &= \frac{\sum_{T \in \mathcal{C}_h} h_T^2 \|\nabla q\|_{0,T}^2}{\sqrt{\|u\|_{1,h}^2}} \\ &\geq \frac{\sum_{T \in \mathcal{C}_h} h_T^2 \|\nabla q\|_{0,T}^2}{\sqrt{C \sum_{T \in \mathcal{C}_h} h_T^2 \|\nabla q\|_{0,T}^2}} \\ &= \frac{\eta_q^2}{\sqrt{C} \eta_q} \\ &=: \tilde{C}_1 \eta_q \end{aligned}$$

Damit ist Gleichung (3.15) bewiesen.

Beweis von (3.16):

Aus [7] (Korollar 2.4.2°) wissen wir, daß es eine Funktion $v \in H_0^1(\Omega)^d$ gibt, mit

$$\nabla \cdot v = -q \quad \wedge \quad \|v\|_1 \leq C \|q\|_0 = C.$$

Sei nun \tilde{v} die Interpolierende von v auf V_h . Nach der Abschätzung über die Interpolierende (Ciarlet [5] Theorem 3.1.6 mit $k = m = 0$ und $p = q = 2$) gilt:

$$\|v - \tilde{v}\|_0 \leq Ch|v|_1.$$

Jetzt ist:

$$\begin{aligned} b(0, \tilde{v}; q) &= \int_{\Omega} \tilde{v} \nabla q \\ &= \int_{\Omega} v \nabla q + \int_{\Omega} (\tilde{v} - v) \nabla q \\ &= - \int_{\Omega} \nabla \cdot v q + \int_{\Omega} (\tilde{v} - v) \nabla q \\ &= \int_{\Omega} q^2 + \int_{\Omega} (\tilde{v} - v) \nabla q \\ &= 1 + \int_{\Omega} (\tilde{v} - v) \nabla q \\ &\geq 1 - \sum_{T \in \mathcal{C}_h} \|\tilde{v} - v\|_{0,T} \|\nabla q\|_{0,T} \\ &\geq 1 - \sum_{T \in \mathcal{C}_h} Ch_T |v|_{1,T} \|\nabla q\|_{0,T} \\ &\geq 1 - C |v|_1 \sqrt{\sum_{T \in \mathcal{C}_h} h_T^2 \|\nabla q\|_{0,T}^2} \\ &\geq 1 - \tilde{C}_2 \sqrt{\sum_{T \in \mathcal{C}_h} h_T^2 \|\nabla q\|_{0,T}^2} \\ &=: 1 - \tilde{C}_2 \eta_q \end{aligned}$$

Wir wählen nun

$$(\sigma, u) := \left(0, \frac{\tilde{v}}{\|\tilde{v}\|_{1,h}}\right)$$

Mit dieser Wahl ist

$$b(0, \tilde{v}; q) = \frac{b(\sigma, u; q)}{\|\sigma, u\|_h}$$

und die Aussage (3.16) ist bewiesen.

□

3.4.3 Schluß

Hier beweisen wir die Aussage von Kapitel 3.4; nämlich, daß die Gleichung (3.11) für jede rechte Seite eindeutig lösbar ist. Dazu verwenden wir folgenden Hilfsatz aus der linearen Algebra: Wenn A eine lineare Abbildung von \mathbb{R}^n nach \mathbb{R}^n ist, so gilt folgende Äquivalenz:

$$(A(x) = 0 \Rightarrow x = 0) \iff (A(x) = b \text{ ist } \forall b \text{ eindeutig lösbar.})$$

Es reicht also zu zeigen, daß das lineare (linear in (σ, u, p)) Gleichungssystem

$$\begin{aligned} \forall (\tau, v) \in H_h \times V_h : \quad a(\sigma, u; \tau, v) + b(\tau, v; p) &= 0 \\ \forall q \in P_h : \quad b(\sigma, u; q) &= 0 \end{aligned} \tag{3.17}$$

nur die Lösung $(\sigma, u, p) = (0, 0, 0)$ hat.

Beweis

Nach Lemma (3.4.1) folgt, daß es eine Konstante $C > 0$ gibt, sodaß $\forall (\sigma, u) \in H_h \times V_h$ gilt:

$$C \|\sigma, u\|_h \leq \sup_{\substack{(\tau, v) \in Z_h \\ (\tau, v) \neq 0}} \frac{a(\sigma, u; \tau, v)}{\|\tau, v\|_h} \tag{3.18}$$

Weil aber $\forall (\tau, v) \in Z_h : b(\tau, v; p) = 0$ gilt, so ist:

$$\sup_{(\tau, v) \in Z_h} \frac{a(\sigma, u; \tau, v)}{\|\tau, v\|_h} = \sup_{(\tau, v) \in Z_h} \frac{a(\sigma, u; \tau, v) + b(\tau, v; p)}{\|\tau, v\|_h} = 0$$

Es ist gleich Null, wegen der Voraussetzung (3.17). Jetzt folgt (3.18)

$$C \|\sigma, u\|_h \leq 0$$

und daraus $(\sigma, u) = 0$.

Wegen der Sup-Bedingung für b (Lemma 3.4.2) wissen wir: $\exists C > 0$:

$$C \|p\|_0 \leq \sup_{(\tau, v) \in H_h \times V_h} \frac{b(\tau, v; p)}{\|\tau, v\|_h} = \sup_{(\tau, v) \in H_h \times V_h} \frac{-a(\sigma, u; \tau, v)}{\|\tau, v\|_h} = 0.$$

Dabei gilt die erste Gleichung nach Voraussetzung (3.17), und die zweite Gleichung ist wahr, da $(\sigma, u) = 0$, was oben gezeigt wurde. Somit ist auch $p = 0$ und der Beweis zu Ende.

□

4 Fehlerabschätzung

Die folgenden Fehlerabschätzungen wurden für die Lösung u_h^* des Problems (3.4) auf Seite 5 gemacht. Dabei sei erwähnt, daß wir u_h^* aus zwei Gründen nicht berechnet haben. Zum ersten ist in meinen gerechneten Beispielen die Lösung u_h des Problems (3.3) schon sehr genau. Zweitens ist der Inhalt dieser Arbeit der Vergleich der beiden Programme. Dieser Vergleich kann ebenso gut auf den nicht “optimierten” Lösungen stattfinden.

4.1 Behauptung

Für eine Lösung $u_h^* \in V_h^*$ und $p_h \in P_h$ der Probleme (3.4) und (3.3) gilt:

$$\| u - u_h^* \|_{1,h} + \| p - p_h \|_0 \leq Ch^{k+1}(|u|_{k+2} + |p|_{k+1})$$

Wenn die Regularitätsannahme (2.3) erfüllt ist, so gilt sogar:

$$\text{für } k \geq 2 : \quad \| u - u_h^* \|_0 \leq Ch^{k+2}(|u|_{k+2} + |p|_{k+1}) \quad (4.1)$$

und

$$\text{für } k = 1 : \quad \| u - u_h^* \|_0 \leq Ch^2(|u|_3 + |p|_2)$$

Gilt zusätzlich die Bedingung $f \in V_h$, so ist die Abschätzung (4.1) auch für $k = 1$ richtig.

4.2 Beweis in fünf Schritten

4.2.1 1.Schritt

Im ersten Schritt zeigen wir für beliebige $\tilde{\sigma} \in H_h, \tilde{p} \in P_h$ die Abschätzung:

$$\begin{aligned} & \| \sigma_h - \tilde{\sigma} \|_{0,h} + \| u_h - Q_h u \|_{1,h} + \| p_h - \tilde{p} \|_0 \\ & \leq C \left[\| \sigma - \tilde{\sigma} \|_{0,h} + \| p - \tilde{p} \|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}} \right] \end{aligned} \quad (4.2)$$

Weil $\| \sigma - \sigma_h \|_{0,h} \leq \| \sigma - \tilde{\sigma} \|_{0,h} + \| \sigma_h - \tilde{\sigma} \|_{0,h}$ und $\| p - p_h \|_0 \leq \| p - \tilde{p} \|_0 + \| p_h - \tilde{p} \|_0$ folgt dann sofort auch

$$\begin{aligned}
& \| \sigma - \sigma_h \|_{0,h} + \| u_h - Q_h u \|_{1,h} + \| p - p_h \|_0 \\
& \leq \| \sigma - \tilde{\sigma} \|_{0,h} + \| \tilde{\sigma} - \sigma_h \|_{0,h} \\
& \quad + \| u_h - Q_h u \|_{1,h} + \| p - \tilde{p} \|_0 + \| p_h - \tilde{p} \|_0 \\
& \stackrel{(4.2)}{\leq} \| \sigma - \tilde{\sigma} \|_{0,h} + \| p - \tilde{p} \|_0 \\
& \quad + C(\| \sigma - \tilde{\sigma} \|_{0,h} + \| p - \tilde{p} \|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}}) \\
& \leq C(\| \sigma - \tilde{\sigma} \|_{0,h} + \| p - \tilde{p} \|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}})
\end{aligned} \tag{4.3}$$

Beweis von (4.2):

Mit [2] folgt, daß ein Tripel $(\tau, v, q) \in H_h \times V_h \times P_h$ existiert mit:

$$\| \tau \|_{0,h} + \| v \|_{1,h} + \| q \|_0 \leq C \tag{4.4}$$

und

$$\begin{aligned}
& \| \sigma_h - \tilde{\sigma} \|_{0,h} + \| u_h - Q_h u \|_{1,h} + \| p_h - \tilde{p} \|_0 \\
& \leq a(\sigma_h - \tilde{\sigma}, u_h - Q_h u; \tau, v) + b(\tau, v; p_h - \tilde{p}) + b(\sigma_h - \tilde{\sigma}, u_h - Q_h u; q).
\end{aligned} \tag{4.5}$$

Es gilt:

$$\begin{aligned}
& a(\sigma_h - \tilde{\sigma}, u_h - Q_h u; \tau, v) + b(\tau, v; p_h - \tilde{p}) + b(\sigma_h - \tilde{\sigma}, u_h - Q_h u; q) \\
& = a(\sigma - \tilde{\sigma}, u - Q_h u; \tau, v) + b(\tau, v; p - \tilde{p}) + b(\sigma - \tilde{\sigma}, u - Q_h u; q) \\
& \stackrel{\text{Def}}{=} (\sigma - \tilde{\sigma}, \tau) + (\nabla \cdot \tau, u - Q_h u) - (\nabla \cdot (\sigma - \tilde{\sigma}), v) + (v, \nabla(p_h - \tilde{p})) + (u_h - Q_h u, \nabla q)
\end{aligned} \tag{4.6}$$

Diese Gleichung gilt, erstens durch Einschreiben von σ , u und p :

$$\begin{aligned}
& a(\sigma_h - \tilde{\sigma}, u_h - Q_h u; \tau, v) + b(\tau, v; p_h - \tilde{p}) + b(\sigma_h - \tilde{\sigma}, u_h - Q_h u; q) \\
& = a(\sigma_h - \sigma + \sigma - \tilde{\sigma}, u_h - u + u - Q_h u; \tau, v) \\
& \quad + b(\tau, v; p_h - p + p - \tilde{p}) \\
& \quad + b(\sigma_h - \sigma + \sigma - \tilde{\sigma}, u_h - u + u - Q_h u; q) \\
& = a(\sigma_h - \sigma, u_h - u; \tau, v) + a(\sigma - \tilde{\sigma}, u - Q_h u; \tau, v) \\
& \quad + b(\tau, v; p_h - p) + b(\tau, v; p - \tilde{p}) \\
& \quad + b(\sigma_h - \sigma, u_h - u; q) + b(\sigma - \tilde{\sigma}, u - Q_h u; q)
\end{aligned}$$

$$\stackrel{(4.4,3.8)}{\implies} (\sigma - \tilde{\sigma}, \tau) \leq C \|\sigma - \tilde{\sigma}\|_{0,h}. \quad (4.9)$$

Mit der Ungleichung (3.5) erhalten wir als nächstes

$$\begin{aligned} (v, \nabla \cdot (\sigma - \tilde{\sigma})) &\stackrel{(3.5)}{\leq} \|v\|_{1,h} \cdot \|\sigma - \tilde{\sigma}\|_{0,h} \\ &\stackrel{(4.4)}{\leq} C \|\sigma - \tilde{\sigma}\|_{0,h}. \end{aligned} \quad (4.10)$$

Und den letzten Summanden schätzen wir wie folgt ab:

$$\begin{aligned} (v, \nabla(p - \tilde{p})) &= \sum_{T \in \mathcal{C}_h} \int_T v \nabla(p - \tilde{p}) \\ &\stackrel{(A.1.5)}{=} - \sum_{T \in \mathcal{C}_h} (\nabla \cdot v, p - \tilde{p})_T + \sum_{e \in \Gamma_h} \int_e ([v \cdot n]_e)(p - \tilde{p}) \\ &\leq (\sum_{T \in \mathcal{C}_h} \|\nabla \cdot v\|_{0,T}^2 + \sum_{e \in \Gamma_h} h_e^{-1} \int_e |[v \cdot n]_e|^2)^{1/2} \\ &\quad \cdot (\|p - \tilde{p}\|_0^2 + \sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{1/2} \\ &\leq C \|v\|_{1,h} (\|p - \tilde{p}\|_0^2 + \sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{1/2} \\ &\stackrel{(4.4)}{\leq} C (\|p - \tilde{p}\|_0^2 + \sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{1/2} \\ &\leq C (\|p - \tilde{p}\|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}}). \end{aligned} \quad (4.11)$$

Setzen wir nun die drei zuletzt berechneten Abschätzungen (4.9,4.10,4.11) in

$$\begin{aligned} \|\sigma_h - \tilde{\sigma}\|_{0,h} + \|u_h - Q_h u\|_{1,h} + \|p_h - \tilde{p}\|_0 &\stackrel{(4.5,4.6)}{\leq} a(\sigma - \tilde{\sigma}, u - Q_h u; \tau, v) \\ &\quad + b(\tau, v; p - \tilde{p}) \\ &\quad + b(\sigma - \tilde{\sigma}, u - Q_h u; q) \\ &= (\sigma - \tilde{\sigma}, \tau) + (\nabla \cdot \tau, u - Q_h u) \\ &\quad - (\nabla \cdot (\sigma - \tilde{\sigma}), v) + (v, \nabla(p - \tilde{p})) \\ &\quad + (u - Q_h u, \nabla q) \\ &\stackrel{(4.7,4.8)}{=} (\sigma - \tilde{\sigma}, \tau) - (\nabla \cdot (\sigma - \tilde{\sigma}), v) \\ &\quad + (v, \nabla(p - \tilde{p})) \end{aligned}$$

ein, so erhalten wir die behauptete Ungleichung (4.2) und der erste Schritt ist beendet.

□

4.2.2 2.Schritt

Jetzt wollen wir $\|Q_h(u_h^* - u)\|_0$ abschätzen. Und zwar durch:

$$\| Q_h(u_h^* - u) \|_0 \leq Ch(\| \sigma - \tilde{\sigma} \|_{0,h} + \| p - \tilde{p} \|_0 + (\sum_e h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}}) \quad (4.12)$$

Den Term h erhalten wir leider nur für $(k \geq 2) \vee (k = 1 \wedge f \in V_h)$. Für $k = 1 \wedge f \notin V_h$ erhalten wir nur

$$\| Q_h(u_h^* - u) \|_0 \leq C(\| \sigma - \tilde{\sigma} \|_{0,h} + \| p - \tilde{p} \|_0 + (\sum_e h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}})$$

Beweis

Zuerst bemerken wir, daß

$$Q_h u_h^* = u_h$$

ist, wegen der Definition (3.4) von u_h^* . Wegen der Regularitätsannahmen an u und p (vgl. (2.3) auf Seite 3) gilt für eine Lösung $(z, \gamma, q) \in H_0^1(\Omega)^2 \times L^2(\Omega)^{2 \times 2} \times L_0^2(\Omega)$ von

$$\begin{aligned} \gamma - \nabla z &= 0 && \in \Omega, \\ -\nabla \cdot \gamma + \nabla q &= u_h - Q_h u && \in \Omega, \\ \nabla \cdot z &= 0 && \in \Omega, \\ z &= 0 && \text{auf } \partial\Omega, \end{aligned} \quad (4.13)$$

und für konvexe Gebiete Ω die Abschätzung

$$\| z \|_2 + \| \gamma \|_1 + \| q \|_1 \leq C \| u_h - Q_h u \|_0 \quad (4.14)$$

Beweis

Diese Abschätzung gilt für $\| z \|_2$ und $\| q \|_1$ schon wegen der Regularitätsannahme (2.3). Und für $\| \gamma \|_1$ gilt sie weil

$$\| \gamma \|_1 = \| \nabla z \|_1 \leq \| z \|_2 \stackrel{(2.3)}{\leq} C \| u_h - Q_h u \|_0$$

Seien jetzt $\tilde{\gamma}$ und \tilde{q} die Interpolierenden zu γ und q in den Räumen H_h und P_h . Wegen der Definition von Q_h erhalten wir die Gleichungen

$$\begin{aligned} (\nabla \cdot \tilde{\gamma}, u) &= (\nabla \cdot \tilde{\gamma}, Q_h u) \\ (\nabla \tilde{q}, u) &= (\nabla \tilde{q}, Q_h u). \end{aligned} \quad (4.15)$$

Wenn wir nun die diskrete Variationsformulierung (3.3) von der kontinuierlichen (3.2) subtrahieren so erhalten wir:

$$\begin{aligned}
\forall \tau_h \in H_h & : (\sigma - \sigma_h, \tau_h) + (\nabla \cdot \tau_h, u - u_h) = 0 \\
\forall v_h \in V_h & : (-\nabla \cdot (\sigma - \sigma_h), v_h) + (\nabla(p - p_h), v_h) = 0 \\
\forall q_h \in P_h & : (u - u_h, \nabla q_h) = 0
\end{aligned} \tag{4.16}$$

$$\begin{aligned}
& \| u_h - Q_h u \|_0^2 \\
& = (u_h - Q_h u, u_h - Q_h u) \\
& = (u_h - Q_h u, -\nabla \cdot \gamma + \nabla q) \\
& = -(u_h - Q_h u, \nabla \cdot \gamma) + (u_h - Q_h u, \nabla q) \\
& \stackrel{(4.13)}{=} -(u_h - Q_h u, \nabla \cdot \gamma) + (u_h - Q_h u, \nabla q) \\
& \quad + (p - p_h, \nabla \cdot z) + (\gamma - \nabla z, \sigma - \sigma_h) \\
& = -(u_h - Q_h u, \nabla \cdot \gamma) + (u_h - Q_h u, \nabla q) \\
& \quad - (\nabla(p - p_h), z) + (\sigma - \sigma_h, \gamma) \\
& \quad + (\nabla \cdot (\sigma - \sigma_h), z) \\
& \stackrel{(4.16)}{=} -(u_h - Q_h u, \nabla \cdot \gamma) + (u_h - Q_h u, \nabla q) \\
& \quad - (\nabla(p - p_h), z) + (\sigma - \sigma_h, \gamma) \\
& \quad + (\nabla \cdot (\sigma - \sigma_h), z) \\
& \quad - (\sigma - \sigma_h, \tilde{\gamma}) - (\nabla \cdot \tilde{\gamma}, u - u_h) \\
& \quad - (\nabla \cdot (\sigma - \sigma_h), Q_h z) + (\nabla(p - p_h), Q_h z) \\
& \quad + (u - u_h, \nabla \tilde{q}) \\
& = (\sigma - \sigma_h, \gamma - \tilde{\gamma}) + (\nabla \cdot (\sigma - \sigma_h), z - Q_h z) - (\nabla(p - p_h), z - Q_h z) \\
& \quad + (\nabla \cdot \gamma, Q_h u - u_h) - (\nabla \cdot \tilde{\gamma}, u - u_h) - (\nabla q, Q_h u - u_h) + (\nabla \tilde{q}, u - u_h) \\
& \stackrel{(4.15)}{=} (\sigma - \sigma_h, \tilde{\gamma} - \gamma) + (\nabla \cdot (\sigma - \sigma_h), z - Q_h z) - (\nabla(p - p_h), z - Q_h z) \\
& \quad + (\nabla \cdot (\gamma - \tilde{\gamma}), Q_h u - u_h) - (\nabla(q - \tilde{q}), Q_h u - u_h) \\
& \leq C(\| \sigma - \sigma_h \|_{0,h} + \| u_h - Q_h u \|_{1,h} + \| p - p_h \|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}}) \\
& \quad \cdot (\| \gamma - \tilde{\gamma} \|_{0,h} + \| z - Q_h z \|_{1,h} + \| q - \tilde{q} \|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |q - \tilde{q}|^2)^{\frac{1}{2}}).
\end{aligned} \tag{4.17}$$

Wegen (3.6),(3.7) und [5] 3.1.6 gilt für $k \geq 2$ die Abschätzung:

$$\begin{aligned}
& \| \gamma - \tilde{\gamma} \|_{0,h} + \| z - Q_h z \|_{1,h} \\
& + \| q - \tilde{q} \|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |q - \tilde{q}|^2)^{\frac{1}{2}} \\
& \stackrel{(A.3)}{\leq} Ch^1 |\gamma|_1 + Ch^1 |z|_2 + Ch^1 |q|_1 \\
& \stackrel{(4.14)}{\leq} Ch \| u_h - Q_h u \|_0.
\end{aligned} \tag{4.18}$$

Wir kürzen auf beiden Seiten von Ungleichung (4.17) $\| u_h - Q_h u \|_0$ und erhalten mit (4.18):

$$\begin{aligned}
\| u_h - Q_h u \|_0 & \stackrel{(4.17)}{\leq} Ch (\| \sigma - \sigma_h \|_{0,h} + \| u_h - Q_h u \|_{1,h} \\
& + \| p - p_h \|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}}) \\
& \stackrel{(4.3)}{\leq} Ch (\| \sigma - \tilde{\sigma} \|_{0,h} + \| p - \tilde{p} \|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}}).
\end{aligned}$$

Für $k = 1$ und $f \in V_h$ verschwindet der Term:

$$(\nabla \cdot (\sigma - \sigma_h) + \nabla(p - p_h), z - Q_h z)$$

und somit erhalten wir das gleiche Resultat, da auch der Term $Ch^1 |z|_1$ in Ungleichung (4.18) verschwindet.

Für $k = 1$ und $f \notin V_h$ erhalten wir nur noch

$$\begin{aligned}
\| u_h - Q_h u \|_0 & \leq C (\| \sigma - \sigma_h \|_{0,h} + \| u_h - Q_h u \|_{1,h} + \| p - p_h \|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}}) \\
& \leq C (\| \sigma - \tilde{\sigma} \|_{0,h} + \| p - \tilde{p} \|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}}).
\end{aligned}$$

□

4.2.3 3.Schritt

Hier wollen wir die Normen $\| (I - Q_h)(u_h^* - u) \|_0$ und $\| (I - Q_h)(u_h^* - u) \|_{1,h}$ abschätzen. Zuerst bemerken wir, daß für $v \in P_0(T)^2$ gilt $(I - Q_h|_T)v|_T = 0$. Dabei steht I für die Identität.

Da auf dem Referenzdreieck \hat{T}

$$\| (I - Q_h)(u_h^* - u) \|_{0,\hat{T}} \leq C |(I - Q_h)(u_h^* - u)|_{1,\hat{T}}$$

gilt, so folgt mit dem Transformationssatz

$$\| (I - Q_h)(u_h^* - u) \|_{0,T} \leq Ch_T |(I - Q_h)(u_h^* - u)|_{1,T}. \tag{4.19}$$

Wir schreiben jetzt

$$\begin{aligned}
|(I - Q_h)(u_h^* - u)|_{1,T}^2 &\stackrel{\text{def}}{=} (\nabla(I - Q_h)(u_h^* - u), \nabla(I - Q_h)(u_h^* - u))_T \\
&= (\nabla(u_h^* - u), \nabla(I - Q_h)(u_h^* - u))_T \\
&\quad - (\nabla Q_h(u_h^* - u), \nabla(I - Q_h)(u_h^* - u))_T.
\end{aligned} \tag{4.20}$$

Jetzt folgt wegen (2.2) und (3.4)

$$\begin{aligned}
(\nabla(u_h^* - u), \nabla(I - Q_h)(u_h^* - u))_T &= (\sigma_h - \sigma, \nabla(I - Q_h)(u_h^* - u))_T \\
&\stackrel{\text{Hölder}}{\leq} \|\sigma - \sigma_h\|_{0,T} |(I - Q_h)(u_h^* - u)|_{1,T}.
\end{aligned} \tag{4.21}$$

Wegen der Inversen Abschätzung (A.1.2) gilt:

$$\begin{aligned}
&(\nabla Q_h(u_h^* - u), \nabla(I - Q_h)(u_h^* - u))_T \\
&\leq |Q_h(u_h^* - u)|_{1,T} |(I - Q_h)(u_h^* - u)|_{1,T} \\
&\leq Ch_T^{-1} \|Q_h(u_h^* - u)\|_{0,T} |(I - Q_h)(u_h^* - u)|_{1,T}.
\end{aligned} \tag{4.22}$$

Setzen wir die drei obigen Ungleichungen (4.20, 4.21, 4.22) zusammen, so erhalten wir

$$|(I - Q_h)(u_h^* - u)|_{1,T} \leq \|\sigma - \sigma_h\|_{0,T} + Ch_T^{-1} \|Q_h(u_h^* - u)\|_{0,T}.$$

Dies nun eingesetzt in Ungleichung (4.19) und über alle $T \in \mathcal{C}_h$ summiert ergibt die erste gesuchte Ungleichung

$$\|(I - Q_h)(u_h^* - u)\|_0 \leq C(h \|\sigma - \sigma_h\|_0 + \|Q_h(u_h^* - u)\|_0). \tag{4.23}$$

Zudem gilt die Ungleichung (Beweis gleich wie (A.3) auf Seite 66):

$$h_T^{-1} \int_{\partial T} |(I - Q_h)v|^2 \leq C |\nabla(I - Q_h)v|_{0,T}^2 \quad \forall v \in V_{h|T}^*.$$

Wieder die obigen Ungleichungen zusammengefaßt, so erhalten wir die zweite gesuchte Abschätzung:

$$\|(I - Q_h)(u_h^* - u)\|_{1,h} \leq C(\|\sigma - \sigma_h\|_0 + \|Q_h(u_h^* - u)\|_{1,h}). \tag{4.24}$$

□

4.2.4 4.Schritt

In diesem Schritt wollen wir $\|u - u_h^*\|_0$ und $\|u - u_h^*\|_{1,h}$ abschätzen.

$$\begin{aligned}
\|u - u_h^*\|_0 &= \|u - u_h^* - Q_h(u_h^* - u) + Q_h(u_h^* - u)\|_0 \\
&\leq \|u - u_h^* - Q_h(u_h^* - u)\|_0 + \|Q_h(u_h^* - u)\|_0 \\
&\stackrel{(4.23)}{\leq} C(h \|\sigma - \sigma_h\|_0 + \|Q_h(u_h^* - u)\|_0) + \|Q_h(u_h^* - u)\|_0 \\
&\leq C(h \|\sigma - \sigma_h\|_0 + \|Q_h(u_h^* - u)\|_0) \\
&\stackrel{(4.12)}{\leq} Ch^*(\|\sigma - \sigma_h\|_{0,h} + \|u_h - Q_h u\|_{1,h} \\
&\quad + \|p - p_h\|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}})
\end{aligned} \tag{4.25}$$

h^* wird gleich h gesetzt, wenn $k \geq 2$ ist oder wenn $k = 1$ und $f \in V_h$ liegt. Im Falle $k = 1$ und $f \notin V_h$ wird h^* gleich 1 gesetzt:

$$h^* := \begin{cases} h & \text{für } (k \geq 2) \vee (k = 1 \wedge f \in V_h) \\ 1 & \text{für } (k = 1 \wedge f \notin V_h) \end{cases}$$

Und zum zweiten schätzen wir $u - u_h^*$ in der $\|\cdot\|_{1,h}$ -Norm ab:

$$\begin{aligned}
\|u - u_h^*\|_{1,h} &= \|u - u_h^* - Q_h(u_h^* - u) + Q_h(u_h^* - u)\|_{1,h} \\
&\leq \|u - u_h^* - Q_h(u_h^* - u)\|_{1,h} + \|Q_h(u_h^* - u)\|_{1,h} \\
&\stackrel{(4.24)}{\leq} C(\|\sigma - \sigma_h\|_0 + \|Q_h(u_h^* - u)\|_{1,h}) + \|Q_h(u_h^* - u)\|_{1,h} \\
&\leq C(\|\sigma - \sigma_h\|_0 + \|Q_h(u_h^* - u)\|_{1,h}) \\
&\stackrel{(4.2)}{\leq} C(\|\sigma - \tilde{\sigma}\|_{0,h} + \|p - \tilde{p}\|_0 + (\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2)^{\frac{1}{2}})
\end{aligned} \tag{4.26}$$

□

4.2.5 5.Schritt und Schluß

In diesem Abschnitt werden die in den ersten vier Schritten gewonnenen Abschätzungen zusammengefaßt und die zuerst behaupteten Ungleichungen für die Fehlerabschätzung bewiesen.

Zuerst wird aber die Abschätzung für die Interpolierende aus Ciarlet [5] noch einmal angegeben. Seien also $\tilde{\sigma}$ aus H_h und \tilde{p} aus P_h nicht mehr beliebig, sondern die Interpolierenden an σ und p . Wie Ciarlet beweist, gelten die beiden Abschätzungen:

$$\| \sigma - \tilde{\sigma} \|_0 \leq Ch^{k+1} |\sigma|_{k+1}$$

und

$$\| p - \tilde{p} \|_0 \leq Ch^{k+1} |p|_{k+1}.$$

Es gilt noch die Abschätzung (vgl. A.3):

$$\left(\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2 \right)^{\frac{1}{2}} \leq Ch^{k+1} |p|_{k+1}.$$

Mit diesen Hilfsmitteln kann die erste Behauptung der Fehlerabschätzung angegeben werden:

$$\begin{aligned} \| u - u_h^* \|_{1,h} + \| p - p_h \|_0 &\stackrel{(4.26)}{\leq} C(\| \sigma - \tilde{\sigma} \|_{0,h} + \| p - \tilde{p} \|_0 + \left(\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2 \right)^{\frac{1}{2}}) \\ &\quad + \| p - \tilde{p} + \tilde{p} - p_h \|_0 \\ &\stackrel{(4.2)}{\leq} C(\| \sigma - \tilde{\sigma} \|_{0,h} + \| p - \tilde{p} \|_0 + \left(\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2 \right)^{\frac{1}{2}}) \\ &\stackrel{(3.8)}{\leq} C(\| \sigma - \tilde{\sigma} \|_0 + \| p - \tilde{p} \|_0 + \left(\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2 \right)^{\frac{1}{2}}) \\ &\leq Ch^{k+1} (|u|_{k+2} + |p|_{k+1}). \end{aligned}$$

Nun zur Abschätzung von $\| u - u_h^* \|_0$: Sei wieder $h^* = h$ für $(k \geq 2)$ oder $(k = 1 \wedge f \in V_h)$. Und $h^* = 1$ für $k = 1 \wedge f \notin V_h$.

$$\begin{aligned} \| u - u_h^* \|_0 &\stackrel{(4.25)}{\leq} Ch^*(\| \sigma - \sigma_h \|_{0,h} + \| u_h - Q_h u \|_{1,h} + \| p - p_h \|_0 + \left(\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2 \right)^{\frac{1}{2}}) \\ &\stackrel{(3.8)}{\leq} Ch^*(\| \sigma - \tilde{\sigma} \|_0 + \| \tilde{\sigma} - \sigma_h \|_0 + \| u_h - Q_h u \|_{1,h} \\ &\quad + \| p - \tilde{p} \|_0 + \| \tilde{p} - p_h \|_0 + \left(\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2 \right)^{\frac{1}{2}}) \\ &\stackrel{(4.2)}{\leq} Ch^*(\| \sigma - \tilde{\sigma} \|_0 + \| p - \tilde{p} \|_0 + \left(\sum_{e \in \Gamma_h} h_e \int_e |p - \tilde{p}|^2 \right)^{\frac{1}{2}}) \\ &\leq Ch^*(h^{k+1} |u|_{k+2} + h^{k+1} |p|_{k+1}). \end{aligned}$$

□

5 Theorie zur Implementierung

Das Problem (2.2) wird auf zwei Arten gelöst. Erstens soll die Variationsformulierung, wie sie in (3.3) steht, ausprogrammiert und in **FEM-FLOW** [9] eingebaut werden. Das geschieht in Kapitel 6. Stenberg [8] schlägt eine Hybridisierungstechnik von Fraijs de Veubeke [10] [11] vor. Mit dieser zweiten Methode könne Speicherplatz gespart werden. Dieses Problem wird in Kapitel 7 beschrieben und ausprogrammiert.

In beiden Programmbeschreibungen werden zuerst die diskreten Vektorräume beschrieben. Danach wird die Stokes-Matrix angegeben und die Integrale (=Einträge der Matrix) berechnet. In einem dritten Teil werden Besonderheiten der jeweiligen Implementierung durchleuchtet.

Die “Theorie zur Implementierung” wird sich zum einen mit der Definition von immer wiederkehrenden Größen (Notation) befassen, zum anderen werden elementare Integrale ausgerechnet.

5.1 Konventionen, Notationen

In **FEM-FLOW** [9] gibt es einige globale Variablen: Mit $ITMDPT(i, k)$, $ITNODE(i, k)$ sind globale Numerierungen der Kanten beziehungsweise der Knoten(=Eckpunkte) bezeichnet. $ITEDGE(i, k)$ gibt an, an welches Dreieck das Dreieck k über die i -te Kante grenzt. Falls das Dreieck k über seine i -te Kante an den Rand der Triangulierung grenzt, so ist $ITEDGE(i, k) = 0$. Die Variablen NE, NV und NT geben die Anzahl der Kanten, der Eckpunkte bzw. der Dreiecke in der Triangulierung an. VX, VY geben die effektiven Koordinaten der Eckpunkte an.

Der i -te Eckpunkt im Dreieck T_k sei (X_k^i, Y_k^i) mit:

$$\begin{aligned} X_k^i &:= VX(ITNODE(i, k)) \\ Y_k^i &:= VY(ITNODE(i, k)) \end{aligned}$$

Dabei wird immer modulo drei gerechnet:

$$X_k^{i+3} \equiv X_k^i \quad Y_k^{i+3} \equiv Y_k^i$$

Es folgen Definitionen der Normalenvektoren:

$$\underline{N} := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Die i -te Kante:

$$e_k^i := \begin{pmatrix} X_k^{i+2} - X_k^{i+1} \\ Y_k^{i+2} - Y_k^{i+1} \end{pmatrix}$$

Die innere Normale zur i -ten Kante:

$$n_k^i := \underline{N}e_k^i \quad \underline{n}_k^i := \frac{n_k^i}{\|e_k^i\|} \quad (5.1)$$

$$n_{k,1}^i = Y_k^{i+1} - Y_k^{i+2} \quad n_{k,2}^i = X_k^{i+2} - X_k^{i+1}$$

Bezeichne mit F_k die Funktion, die das Referenzdreieck \hat{T} (= Dreieck $(0,0)(1,0)(0,1)$) auf das k -te Dreieck T_k (= Dreieck $(X_k^1, Y_k^1)(X_k^2, Y_k^2)(X_k^3, Y_k^3)$) der Diskretisierung abbildet:

$$\hat{T} := \{(x, y) \in \mathbb{R}^2 : x, y \geq 0 : x + y \leq 1\}$$

$$\begin{aligned} F_k : \hat{T} &\rightarrow T_k \\ \hat{x} &\mapsto F_k(\hat{x}) = B_k \hat{x} + \begin{pmatrix} X_k^1 \\ Y_k^1 \end{pmatrix} =: x \end{aligned}$$

mit

$$B_k = \begin{pmatrix} X_k^2 - X_k^1 & X_k^3 - X_k^1 \\ Y_k^2 - Y_k^1 & Y_k^3 - Y_k^1 \end{pmatrix} = (e_k^3, -e_k^2).$$

Daraus folgt

$$\begin{aligned} \det B_k &= \det DF_k \\ &= (X_k^2 - X_k^1)(Y_k^3 - Y_k^1) - (Y_k^2 - Y_k^1)(X_k^3 - X_k^1) \\ F_k^{-1}(x) &= \frac{1}{\det B_k} \begin{pmatrix} Y_k^3 - Y_k^1 & X_k^1 - X_k^3 \\ Y_k^1 - Y_k^2 & X_k^2 - X_k^1 \end{pmatrix} \cdot \left(x - \begin{pmatrix} X_k^1 \\ Y_k^1 \end{pmatrix}\right) \end{aligned}$$

Die Schwerpunktskoordinaten λ sind wie üblich definiert:

$$\begin{aligned} \hat{\lambda}^i &: \hat{T} \rightarrow \mathbb{R} \\ \lambda_k^i &: T_k \rightarrow \mathbb{R} \\ \hat{\lambda}^1(x, y) &:= 1 - x - y \\ \hat{\lambda}^2(x, y) &:= x \\ \hat{\lambda}^3(x, y) &:= y \\ \lambda_k^i &:= \hat{\lambda}^i \circ F_k^{-1} \end{aligned}$$

Das Kronecker Symbol ist wie üblich definiert:

$$\delta_j^m := \begin{cases} 0 & j \neq m \\ 1 & j = m. \end{cases}$$

5.2 Ableitungen und Integrale

Die Transformationsformel für lineare Abbildungen besagt insbesondere:

$$\int_{T_k} f(x)dx = \int_{\hat{T}} f(F_k(\hat{x})) \det B_k d\hat{x}$$

Ebenso:

$$\int_{T_k} \lambda_k^i(x)dx = \int_{\hat{T}} \hat{\lambda}^i(\hat{x}) \det B_k d\hat{x}.$$

Die partiellen Ableitungen der Schwerpunktskoordinaten sind:

$$\partial_1(\lambda_k^i(x)) = \frac{1}{\det B_k} (Y_k^{i+1} - Y_k^{i+2})$$

und

$$\partial_2(\lambda_k^i(x)) = \frac{1}{\det B_k} (X_k^{i+2} - X_k^{i+1}).$$

Beweis

Sei in diesem Beweis $e_j := (\delta_1^j, \delta_2^j)$.

$$\begin{aligned} \partial_j(\lambda_k^i(x)) &= \partial_j(\hat{\lambda}^i(\hat{x})) \\ &= \partial_j(\hat{\lambda}^i(F^{-1}x)) \\ &= D(\hat{\lambda}^i \circ F^{-1})(x) \cdot e_j \\ &= D\hat{\lambda}^i(F^{-1}(x)) \cdot DF^{-1}(x) \cdot e_j \\ &= \begin{cases} [-1, -1] & \text{für } i = 1 \\ [1, 0] & \text{für } i = 2 \\ [0, 1] & \text{für } i = 3 \end{cases} \frac{1}{\det B_k} \begin{bmatrix} Y_k^3 - Y_k^1 & X_k^1 - X_k^3 \\ Y_k^1 - Y_k^2 & X_k^2 - X_k^1 \end{bmatrix} \cdot e_j \\ &= \frac{1}{\det B_k} \begin{cases} (Y_k^{i+1} - Y_k^{i+2}) & \text{für } j = 1 \\ (X_k^{i+2} - X_k^{i+1}) & \text{für } j = 2 \end{cases} \\ &= \frac{1}{\det B_k} \cdot n_{k,j}^i. \end{aligned}$$

□

Integral über die Ableitung einer Schwerpunktskoordinate: $\forall i = 1, 2, 3 :$

$$\begin{aligned} \int_{T_k} \partial_1 \lambda_k^i(x) dx &= +\frac{1}{2} n_{k,1}^i \\ \int_{T_k} \partial_2 \lambda_k^i(x) dx &= +\frac{1}{2} n_{k,2}^i \end{aligned} \tag{5.2}$$

Für die Schwerpunktskoordinaten berechnet man leicht die Integrale auf dem Referenzdreieck:

$$\begin{aligned}\forall i : \int_{\hat{T}} \hat{\lambda}_k^i &= \frac{1}{6} \\ i \neq j : \int_{\hat{T}} \hat{\lambda}_k^i \hat{\lambda}_k^j &= \frac{1}{24} \\ \forall i : \int_{\hat{T}} (\hat{\lambda}_k^i)^2 &= \frac{1}{12}\end{aligned}$$

6 1.Programm

Im ersten Programm wird folgendes Problem implementiert:

Finde $u \in V_h$, $p \in P_h$ und $\sigma \in H_h$ sodaß:

$$\begin{aligned}\forall \tau \in H_h : \quad & (\sigma, \tau) + (\nabla \cdot \tau, u) = \langle u_0, \tau \cdot n \rangle \\ \forall v \in V_h : \quad & -(\nabla \cdot \sigma, v) + (\nabla p, v) = (f, v) \\ \forall q \in P_h : \quad & (u, \nabla q) = \langle u_0 \cdot n, q \rangle\end{aligned}\tag{6.1}$$

Es wird mit den im Theorieteil definierten Vektorräumen gearbeitet und als Polynomgrad wird

$$k = 1$$

gewählt. Das heißt: lineare Drücke p und stückweise konstante Geschwindigkeiten v .

6.1 Vektorräume

6.1.1 Basis und Funktionen in P_h

Basis in P_h : Die Basis in P_h bilden die Vektoren q_ν mit $\nu \in \{1, 2, \dots, NV\}$, wobei NV die Zahl der Eckpunkte in der Triangulierung angibt. Die q_ν seien wie folgt definiert:

$$q_\nu(x) := \begin{cases} \lambda_k^i(x) & \text{für } \nu = \text{ITNODE}(i, k) \\ 0 & \text{sonst} \end{cases}$$

wobei λ_k^i die i -te Schwerpunktskoordinate im Dreieck T_k bezeichnet.

Funktionen in P_h : Damit können wir $p \in P_h$ schreiben als:

$$p(x) = \sum_{\nu=1}^{NV} p_\nu q_\nu(x) = \sum_{k=1}^{NT} \sum_{i=1}^3 p_{\text{ITNODE}(i,k)} \lambda_k^i(x)$$

mit p_ν in \mathbb{R} .

Dimension von P_h : $\dim(P_h) = NV$

6.1.2 Basis und Funktionen in V_h

Basis:

$$v_k^j(x) := \begin{cases} \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \text{für } j = 1, x \in T_k \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \text{für } j = 2, x \in T_k \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \text{sonst} \end{cases}$$

Dabei bezeichnet j die Komponente der Geschwindigkeit. $j = 1$ für x -Komponente; $j = 2$ für y -Komponente. k läuft von 1 bis NT , der Zahl der Dreiecke.

Damit können wir jede Funktion u in V_h schreiben als

$$u(x) = \sum_{j=1}^2 \sum_{k=1}^{NT} u_k^j v_k^j(x)$$

mit $u_k^j \in \mathbb{R}$.

Dimension von V_h : $\dim(V_h) = 2 \times \text{Anzahl Dreiecke} = 2NT$

6.1.3 Basis und Funktionen in H_h

Sei

$$\gamma_k^i := \frac{\pm |e_k^i|}{\det B_k}$$

dabei wird das negative Vorzeichen genommen, wenn $k < \text{ITEDGE}(i, k)$ ist. Und im Fall $k > \text{ITEDGE}(i, k)$ wird das positive Vorzeichen verwendet. Das heißt, die Normale zu einer Kante zeigt immer vom Dreieck mit der größeren Nummer zum Dreieck mit der kleineren Nummer (vergleiche Abbildung 2 auf Seite 30).

$$\varphi_k^i \begin{pmatrix} x \\ y \end{pmatrix} := \gamma_k^i \begin{pmatrix} x - x_k^i \\ y - y_k^i \end{pmatrix}.$$

Mit diesen Definitionen gilt

$$\varphi_k^i \begin{pmatrix} x_k^{i+1} + t(x_k^{i+2} - x_k^{i+1}) \\ y_k^{i+1} + t(y_k^{i+2} - y_k^{i+1}) \end{pmatrix} \cdot \frac{-n_k^i}{|e_k^i|} = \pm 1.$$

Das heißt auf der i -ten Kante ($t \in [0, 1]$) ist $\varphi \cdot n \equiv 1$ für ein fest vorgegebenes n . Also ist φ in $H(\nabla \cdot, (\Omega))$.

Die Funktionen in H_h sind jetzt auf den Dreiecken T_k wie folgt definiert: ($e = \text{ITMDPT}(k, i)$)

Abbildung 2: Festlegung der Normalenrichtungen

$$\tau_e^{[1]} := \begin{bmatrix} (\varphi_k^i)_1 & (\varphi_k^i)_2 \\ 0 & 0 \end{bmatrix} = \gamma_k^i \begin{bmatrix} x - x_k^i & y - y_k^i \\ 0 & 0 \end{bmatrix}$$

$$\tau_e^{[2]} := \begin{bmatrix} 0 & 0 \\ (\varphi_k^i)_1 & (\varphi_k^i)_2 \end{bmatrix} = \gamma_k^i \begin{bmatrix} 0 & 0 \\ x - x_k^i & y - y_k^i \end{bmatrix}$$

Somit können wir jede Funktion σ aus H_h wie folgt schreiben:

$$\sigma \begin{pmatrix} x \\ y \end{pmatrix} = \sum_{e=1}^{NE} \sum_{m=1}^2 \sigma_e^{[m]} \cdot \tau_e^{[m]} \begin{pmatrix} x \\ y \end{pmatrix}$$

Dimension von H_h : $\dim(H_h) = 2NE$

6.2 Matrixschreibweise und Berechnung der Integrale

Sei

$$\begin{aligned} \Sigma &:= (\sigma_1^{[1]}, \sigma_2^{[1]}, \dots, \sigma_{NE}^{[2]}) \\ U &:= (u_1^1, u_2^1, u_3^1, \dots, u_{NT}^1, u_1^2, u_2^2, \dots, u_{NT}^2) \\ P &:= (p_1, p_2, \dots, p_{NV}) \end{aligned}$$

Damit ist das Gleichungssystem (6.1) äquivalent zum Gleichungssystem:

$$\begin{aligned}
A\Sigma + BU &= E \\
-B^t\Sigma + CP &= F \\
C^tU &= G
\end{aligned}$$

für geeignete A, B, C, E, F, G .

Es folgen jetzt die Berechnungen von A, B, C, E, F, G , sowie die Implementierung.

6.2.1 Berechnung von A

$$A = \begin{bmatrix} (\tau_1^{[1]}, \tau_1^{[1]}) & \dots & (\tau_1^{[1]}, \tau_{NE}^{[2]}) \\ \vdots & \ddots & \vdots \\ (\tau_{NE}^{[2]}, \tau_1^{[1]}) & \dots & (\tau_{NE}^{[2]}, \tau_{NE}^{[2]}) \end{bmatrix}$$

Wegen der Wahl von $\tau_e^{[m]}$ gilt:

$$(\tau_e^{[1]}, \tau_f^{[2]}) = 0 \quad \forall e, f = 1, \dots, NE$$

Und, weil $(\tau_e^{[1]}, \tau_f^{[1]}) = (\tau_e^{[2]}, \tau_f^{[2]})$, ist (τ_e, τ_f) wohldefiniert:

$$(\tau_e, \tau_f) := (\tau_e^{[1]}, \tau_e^{[1]}) := \int_{\Omega} \tau_e^{[1]} \cdot \tau_f^{[1]}$$

Für $e = f$ gilt:

$$\begin{aligned}
(\tau_e, \tau_e) = & \frac{|e_k^i|^2}{6 \det B_k} (((m_k^1)_1 - x_k^i)^2 + ((m_k^1)_2 - y_k^i)^2 \\
& + ((m_k^2)_1 - x_k^i)^2 + ((m_k^2)_2 - y_k^i)^2 \\
& + ((m_k^3)_1 - x_k^i)^2 + ((m_k^3)_2 - y_k^i)^2) \\
& + \frac{|e_k^j|^2}{6 \det B_{\tilde{k}}} (((m_{\tilde{k}}^1)_1 - x_{\tilde{k}}^j)^2 + ((m_{\tilde{k}}^1)_2 - y_{\tilde{k}}^j)^2 \\
& + ((m_{\tilde{k}}^2)_1 - x_{\tilde{k}}^j)^2 + ((m_{\tilde{k}}^2)_2 - y_{\tilde{k}}^j)^2 \\
& + ((m_{\tilde{k}}^3)_1 - x_{\tilde{k}}^j)^2 + ((m_{\tilde{k}}^3)_2 - y_{\tilde{k}}^j)^2)
\end{aligned}$$

mit $e = \text{ITMDPT}(i, k) = \text{ITMDPT}(j, \tilde{k})$ und den Seitenmittelpunkten

$$(m_k^i) := \frac{1}{2} \begin{pmatrix} x_k^{i+2} + x_k^{i+1} \\ y_k^{i+2} + y_k^{i+1} \end{pmatrix}$$

Dies folgt sofort mit der Quadraturformel, die die Summe der Werte in den Seitenmittelpunkten mit der Fläche und $\frac{1}{3}$ multipliziert und somit exakt ist für Polynome 2. Grades.

Für $e \neq f$ ist das Integral (τ_e, τ_f) Null, falls die beiden Kanten e und f nicht zum selben Dreieck gehören, da die Funktionen τ_e, τ_f nur auf den angrenzenden Dreiecken leben. Gehören aber e und f zu den selben Dreiecken so ist

$$\begin{aligned}
(\tau_e, \tau_f) = \frac{(\pm|e_k^i|) \cdot (\pm|e_k^j|)}{6 \det B_k} \cdot (& ((m_k^1)_1 - x_k^i)((m_k^1)_1 - x_k^j) \\
& + ((m_k^1)_2 - y_k^i)((m_k^1)_2 - y_k^j) \\
& + ((m_k^2)_1 - x_k^i)((m_k^2)_1 - x_k^j) \\
& + ((m_k^2)_2 - y_k^i)((m_k^2)_2 - y_k^j) \\
& + ((m_k^3)_1 - x_k^i)((m_k^3)_1 - x_k^j) \\
& + ((m_k^3)_2 - y_k^i)((m_k^3)_2 - y_k^j)
\end{aligned}$$

für $e = \text{ITMDPT}(i, k), f = \text{ITMDPT}(j, k)$. Dies folgt mit derselben Quadraturformel. Dabei wird wieder das positive Vorzeichen verwendet, falls die festgelegte Normale (vgl. Abbildung 2 auf Seite 30) vom Dreieck T_k weg zeigt; sonst das negative.

6.2.2 Berechnung von B

$$B = \begin{bmatrix} (\nabla \cdot \tau_1^{[1]}, v_1^1) & \dots & (\nabla \cdot \tau_1^{[1]}, v_{NT}^2) \\ \vdots & & \vdots \\ (\nabla \cdot \tau_{NE}^{[2]}, v_1^1) & \dots & (\nabla \cdot \tau_{NE}^{[2]}, v_{NT}^2) \end{bmatrix}$$

mit den Integralen:

$$(\nabla \cdot \tau_e^{[m]}, v_k^j)$$

Wobei $m = 1, 2$; $j = 1$ für x-Komponente $j = 2$ für y-Komponente; $k =$ Dreiecksnummer. Weil die Funktionen v_k^j nur Werte auf den Dreiecken annehmen ist sofort klar, daß:

$$(\nabla \cdot \tau_e^{[m]}, v_k^j) = 0$$

falls

$$\forall i \in \{1, 2, 3\} : e \neq \text{ITMDPT}(i, k).$$

Sei im folgenden e die i -te Kante im Dreieck T_k ; ($e = \text{ITMDPT}(i, k)$).

$$\begin{aligned}
(\nabla \cdot \tau_e^{[m]}, v_k^j) &= \int_{T_k} \nabla \cdot \tau_e^{[m]} \cdot v_k^j \\
&= \int_{T_k} [\partial_1(\tau_e^{[m]})_{11} + \partial_2(\tau_e^{[m]})_{12}, \partial_1(\tau_e^{[m]})_{21} + \partial_2(\tau_e^{[m]})_{22}] v_k^j \\
&= \int_{T_k} \partial_1(\tau_e^{[m]})_{j1} + \partial_2(\tau_e^{[m]})_{j2} \\
&= \delta_j^m \int_{T_k} \partial_1(\varphi_k^i)_1 + \partial_2(\varphi_k^i)_2
\end{aligned}$$

$$\begin{aligned}
&= \delta_j^m \int_{T_k} \partial_1 \gamma_k^i (x - x_k^i) + \partial_2 \gamma_k^i (y - y_k^i) \\
&= \delta_j^m \int_{T_k} \gamma_k^i + \gamma_k^i \\
&= 2\delta_j^m \gamma_k^i \int_{T_k} dx \\
&= \pm \delta_j^m \frac{2|e_k^i|}{\det B_k} \frac{1}{2} \det B_k \\
&= \pm \delta_j^m |e_k^i|
\end{aligned}$$

6.2.3 Berechnung von C

$$C = \begin{bmatrix} (\nabla q_1, v_1^1) & \dots & (\nabla q_{NV}, v_1^1) \\ \vdots & & \vdots \\ (\nabla q_1, v_{NT}^2) & \dots & (\nabla q_{NV}, v_{NT}^2) \end{bmatrix}$$

Nun zur Berechnung von $(\nabla q_\nu, v_k^j)$:

Wieder (wie in der Berechnung von B) ist klar, daß das Integral verschwindet, wenn es kein i gibt mit $\nu = \text{ITNODE}(i, k)$. Sei wieder i , sodaß $\nu = \text{ITNODE}(i, k)$:

$$\begin{aligned}
(\nabla q_\nu, v_k^j) &= \int_{\Omega} \nabla q_\nu \cdot v_k^j \\
&= \int_{T_k} \nabla q_\nu \cdot v_k^j \\
&= \int_{T_k} [\partial_1 q_\nu, \partial_2 q_\nu] \begin{pmatrix} \delta_1^j \\ \delta_2^j \end{pmatrix} \\
&= \int_{T_k} \partial_j q_\nu \\
&= \int_{T_k} \partial_j (\lambda_k^i) \\
&= +\frac{1}{2} (n_{k,j}^i).
\end{aligned}$$

Dabei gilt die letzte Gleichung wegen den Gleichungen 5.2 auf Seite 27.

6.2.4 Berechnung von E

$$E = \begin{pmatrix} \langle u_0, \tau_1^{[1]} \cdot n \rangle \\ \langle u_0, \tau_2^{[1]} \cdot n \rangle \\ \vdots \\ \langle u_0, \tau_{NE}^{[2]} \cdot n \rangle \end{pmatrix}$$

Dabei wird das Integral

$$\langle u_0, \tau_e^{[j]} \cdot n \rangle$$

approximiert durch:

$$\langle u_0, \tau_e^{[j]} \cdot n \rangle \approx |e_k^i| \cdot (u_0(m_k^i))$$

Dabei ist $e = \text{ITMDPT}(i, k)$. Falls e keine Randkante ist wird das Integral Null gesetzt.

(Die Bezeichnung m_k^i wurde schon erwähnt und steht einfach für den Mittelpunkt der Kante e .)

6.2.5 Berechnung von F

Um das Integral (f, v_k^j) zu berechnen reicht es, die gegebene Funktion f auf den Dreiecken konstant zu interpolieren. Es macht keinen Sinn, f genauer zu interpolieren als die Basisfunktionen in V_h .

Zuerst sei der Schwerpunkt im Dreieck T_k durch folgende Schreibweise definiert:

$$\begin{aligned} X S_k &:= \frac{X_k^1 + X_k^2 + X_k^3}{3} \\ Y S_k &:= \frac{Y_k^1 + Y_k^2 + Y_k^3}{3} \end{aligned}$$

damit erhalten wir:

$$\begin{aligned} (f, v_k^1) &= \int_{T_k} f \cdot v_k^1 \\ &\approx \int_{T_k} f_1(X S_k, Y S_k) dx \\ &= \frac{1}{2} \det B_k \cdot F U(X S_k, Y S_k) \end{aligned}$$

und

$$\begin{aligned} (f, v_k^2) &= \int_{T_k} f \cdot v_k^2 \\ &\approx \int_{T_k} f_2(X S_k, Y S_k) dx \\ &= \frac{1}{2} \det B_k \cdot F V(X S_k, Y S_k). \end{aligned}$$

6.2.6 Berechnung von G

$$G = \begin{pmatrix} \langle u_0 \cdot n, q_1 \rangle \\ \vdots \\ \langle u_0 \cdot n, q_{NV} \rangle \end{pmatrix}$$

Achtung: n bezeichnet hier die äußere Normale auf Ω , nicht wie n_k , was die innere Normale zum Dreieck T_k bezeichnet (vgl. 5.1).

Jetzt wird $\langle u_0 \cdot n, q_\nu \rangle$ als Summe geschrieben:

$$\begin{aligned} \langle u_0 \cdot n, q_\nu \rangle &= - \int_{e_k^i} u_0 \cdot \underline{n}_k^i \cdot q_\nu - \int_{e_{\tilde{k}}^{\tilde{i}}} u_0 \cdot \underline{n}_{\tilde{k}}^{\tilde{i}} \cdot q_\nu \\ &\approx -\frac{1}{2}(u_0(e_k^i)n_k^i + u_0(e_{\tilde{k}}^{\tilde{i}})n_{\tilde{k}}^{\tilde{i}}) \end{aligned}$$

mit $\nu = \text{ITNODE}(i, k) = \text{ITNODE}(\tilde{i}, \tilde{k})$. $u_0(e_k^i)$ soll einfach den Wert von u_0 im Mittelpunkt der Kante e_k^i bezeichnen. Hierbei wurde konstant interpoliert.

6.3 Implementierung

6.3.1 Konjugierte Residuen

Zum Lösen des linearen Gleichungssystems werden konjugierte Residuen mit Lanczos-Schritt verwendet (vgl. [6] Seiten 261,264,265). Konjugierte Gradienten dürfen nicht verwendet werden, da das System nicht positiv-definit ist.

Algorithmus:

Wähle x_0 .

Setze $r_0 := b - \mathcal{A}x_0$, $p_{-1} := 0$ **und** $p_0 := r_0$.

FOR $k := 0, 1, 2, \dots$ **berechne**

$$\alpha_k := \frac{(r_k, \mathcal{A}p_k)}{(p_k, \mathcal{A}^2 p_k)}$$

$$x_{k+1} := x_k + \alpha_k p_k$$

$$r_{k+1} := r_k - \alpha_k \mathcal{A}p_k$$

$$\gamma_k := \frac{(\mathcal{A}p_k, \mathcal{A}^2 p_k)}{(p_k, \mathcal{A}^2 p_k)}$$

$$\beta_k := \begin{cases} 0 & \text{für } k = 0 \\ \frac{(p_k, \mathcal{A}^2 p_k)}{(p_{k-1}, \mathcal{A}^2 p_{k-1})} & \text{sonst} \end{cases}$$

$$p_{k+1} := \mathcal{A}p_k - \gamma_k p_k - \beta_k p_{k-1}$$

ENDFOR.

In unserem Fall ist \mathcal{A} die Stokes-Matrix:

$$\mathcal{A} := \begin{pmatrix} A & B & 0 \\ B^t & 0 & -C \\ 0 & -C^t & 0 \end{pmatrix}$$

und b die Rechte Seite:

$$b := \begin{pmatrix} E \\ -F \\ -G \end{pmatrix}.$$

Also ist (6.1) äquivalent zum Lösen von

$$\mathcal{A} \cdot \begin{pmatrix} \Sigma \\ U \\ P \end{pmatrix} = b$$

6.3.2 Vorkonditionierung

Auf eine eigentliche Vorkonditionierung für die konjugierten Residuen wird bewußt verzichtet. Bei einer echten Vorkonditionierung wäre es nicht mehr sinnvoll, die beiden Programme (vgl. Kapitel 8) miteinander zu vergleichen. Die gewählte Vorkonditionierung beschränkt sich auf eine Skalierung der Einträge der Blöcke A, B und C . Alle drei Blöcke sollen bei Verfeinerung von derselben Ordnung bezüglich h (der Gittergröße) sein. Skaliert man mit der Matrix

$$\begin{bmatrix} I & 0 & 0 \\ 0 & h & 0 \\ 0 & 0 & I \end{bmatrix},$$

so erhalten alle Einträge der drei Blöcke die Ordnung h^2 .

6.3.3 Implementierung von STOLIN

Das Unterprogramm (Funktion) STOLIN berechnet die Multiplikation der Stokes-Matrix mit einem gegebenen Vektor. Dieses Unterprogramm wird gebraucht, da das Gleichungssystem iterativ gelöst wird.

PARAMETERS

Input: SIG1,SIG2,U,V,P;

Output: RSIG1,RSIG2,RU,RV,RP;

BEGIN

RSIG1 := RSIG2 := RU := RV := RP := 0;

FOR $k := \text{NTL}$ TO NTU DO

FOR $i := 1$ TO 3 DO

E := ITMDPT(i, k);

RSIG1(E) := RSIG1(E) + $(\tau_E, \tau_E) \cdot \text{SIG1}(\mathbf{E})$;

RSIG2(E) := RSIG2(E) + $(\tau_E, \tau_E) \cdot \text{SIG2}(\mathbf{E})$;

IF $i < 3$ THEN

FOR $j := i + 1$ TO 3 DO

F := ITMDPT(j, k);

RSIG1(E) := RSIG1(E) + $(\tau_E, \tau_F) \cdot \text{SIG1}(\mathbf{F})$;

RSIG2(E) := RSIG2(E) + $(\tau_E, \tau_F) \cdot \text{SIG2}(\mathbf{F})$;

RSIG1(F) := RSIG1(F) + $(\tau_E, \tau_F) \cdot \text{SIG1}(\mathbf{E})$;

RSIG2(F) := RSIG2(F) + $(\tau_E, \tau_F) \cdot \text{SIG2}(\mathbf{E})$;

ENDDO

ENDIF;

RSIG1(E) := RSIG1(E) + $(\nabla \cdot \tau_E, v_k) \cdot \mathbf{U}(k)$;

RSIG2(E) := RSIG2(E) + $(\nabla \cdot \tau_E, v_k) \cdot \mathbf{V}(k)$;

RU(k) := RU(k) + $(\nabla \cdot \tau_E, v_k) \cdot \text{SIG1}(\mathbf{E})$;

RV(k) := RV(k) + $(\nabla \cdot \tau_E, v_k) \cdot \text{SIG2}(\mathbf{E})$;

$\nu := \text{ITNODE}(i, k)$;

RU(k) := RU(k) - $(q_\nu, v_k^1) \cdot \mathbf{P}(\nu)$;

RV(k) := RV(k) - $(q_\nu, v_k^2) \cdot \mathbf{P}(\nu)$;

RP(ν) := RP(ν) - $(q_\nu, v_k^1) \cdot \mathbf{U}(k)$ - $(q_\nu, v_k^2) \cdot \mathbf{V}(k)$;

ENDDO

ENDDO

END

7 2.Programm

Im 2. Program wollen wir auf die Bedingung $\sigma \in H(\nabla \cdot, (\Omega))$ verzichten. Wie in Lemma A.1.6 (Seite 68) bewiesen wird, ist $\sigma \in H(\nabla \cdot, (\Omega))$ äquivalent zu “ $\sigma \cdot n$ ist stetig entlang innerer Kanten”.

Sei zunächst

$$H_h^* := \{\tau \in L^2(\Omega)^{n \times n} : \tau|_T \in P_k(\Omega)^{n \times n}, T \in \mathcal{C}_h\}.$$

Aus der ursprünglichen Definition von $\sigma(= \nabla u)$ folgt für alle $\tau_h \in H_h^*$:

$$\begin{aligned} 0 &= (\sigma, \tau_h)_\Omega - (\nabla u, \tau_h)_\Omega \\ &= (\sigma, \tau_h)_\Omega - \sum_T (\nabla u, \tau_h)_T \\ &\stackrel{A.1.5}{=} (\sigma, \tau_h) + \sum_T \{(u, \nabla \cdot \tau_h)_T - \langle u, \tau \cdot n_T \rangle_{\partial T}\}. \end{aligned}$$

Daraus folgt

$$(\sigma, \tau_h) + \sum_T (\nabla \cdot \tau_h, u)_T - \sum_T \langle u, \tau \cdot n_T \rangle_{\partial T \setminus \partial \Omega} = \langle u_0, \tau \cdot n \rangle_{\partial \Omega}.$$

Mit dieser Hybridisierungstechnik von Fraijs de Veubeke [10] [11] erhalten wir eine neue Diskretisierung: Finde ein $(\sigma_h, u_h, m_h, p_h) \in H_h^* \times V_h \times M_h \times P_h$, sodaß

$$\begin{aligned} (\sigma_h, \tau) + \sum_{T \in \mathcal{C}_h} \{(\nabla \cdot \tau, u_h)_T - \langle \tau \cdot n_T, m_h \rangle_{\partial T}\} &= \langle \tau \cdot n, u_0 \rangle, & \tau \in H_h^*, \\ - \sum_{T \in \mathcal{C}_h} (\nabla \cdot \sigma_h, v)_T + (\nabla p_h, v) &= (f, v), & v \in V_h, \\ \sum_{T \in \mathcal{C}_h} \langle \sigma_h \cdot n_T, m \rangle_{\partial T} &= 0, & m \in M_h, \\ (u_h, \nabla q) &= \langle u_0 \cdot n, q \rangle & q \in P_h, \end{aligned} \tag{7.1}$$

mit

$$M_h := \{m : m|_e \in P_k(e)^2, e \in \Gamma_h; m|_e = 0, e \subset \partial \Omega\}.$$

Nebenbei ist m_h der Lagrangesche Multiplikator zum Sprung von $\tau \cdot n$ über innere Kanten.

7.1 Vektorräume

7.1.1 Basis und Funktionen in P_h

Die Funktionen sind genau dieselben wie im ersten Programm. Vergleiche dazu also 6.1.1 auf Seite 28.

7.1.2 Basis und Funktionen in V_h

Die Funktionen aus V_h sind genau dieselben wie im ersten Programm. Vergleiche dazu 6.1.2 auf Seite 29.

7.1.3 Basis und Funktionen in H_h^*

Basis in H_h^* :

Für $x \in T_k$ sei

$$\tau_k^{i[m]}(x) := \begin{pmatrix} \delta_{11}^m & \delta_{12}^m \\ \delta_{21}^m & \delta_{22}^m \end{pmatrix} \cdot (\lambda_k^{i+1} + \lambda_k^{i+2} - \lambda_k^i)(x)$$

und falls x nicht in T_k , so ist $\tau_k^{i[m]}$ Null.

Funktionen in H_h^* : Mit obigen Basisfunktionen können wir jede Funktion aus H_h^* schreiben als

$$\sigma(x) = \sum_{m \in \{11,12,21,22\}} \sum_{k=1}^{NT} \sum_{i=1}^3 \sigma_k^{i[m]} \cdot \tau_k^{i[m]}(x)$$

mit geeigneten reellen $\sigma_k^{i[m]}$.

Dimension von H_h^* : $\dim(H_h^*) = 12 \cdot NT$

7.1.4 Basis und Funktionen in M_h

Basis in M_h :

$$M_h = \{m : m|_e \in P_1(e)^2 : e \in \Gamma_h \text{ und } m|_e = 0 \text{ für } e \in \partial\Omega\}$$

Da die Funktionen aus M_h nur auf den (inneren) Kanten leben, müssen wir diese geeignet nummerieren. Weil **FEM-FLOW** [9] dreiecksorientiert arbeitet, hängt eine Kante aber von einem Dreieck \tilde{k} und einer lokalen Kantenummer \tilde{i} ab. Damit eine innere Kante nicht zweimal vorkommt soll für eine Kante $e_{\tilde{k}}^{\tilde{i}}$ immer erfüllt sein, daß \tilde{k} größer ist als die Dreiecksnummer des angrenzenden Dreiecks: $\tilde{k} > \text{ITNODE}(\tilde{i}, \tilde{k})$.

Da die Funktionen auf den Kanten linear interpoliert werden müssen, existieren pro Kante immer 4 Funktionen. Je eine für x - und y -Komponente ($j = 1$ für x und $j = 2$ für y) und je eine für den Wert an beiden Enden der Kante. $l = 1$ für den Wert vorne und $l = 2$ für den Wert hinten auf der Kante. Vergleiche dazu die Abbildung 3 auf Seite 41.

Sei nun $e_{\tilde{k}}^{\tilde{i}}$ keine Randkante und $x = (x_1, x_2) \in e_{\tilde{k}}^{\tilde{i}}$, so definieren wir

Abbildung 3: lineare Kantenfunktionen

$$m_{(e_{\tilde{k}}^i)}^{jl}(x) := \begin{pmatrix} \delta_1^j \\ \delta_2^j \end{pmatrix} \frac{1}{|e_{\tilde{k}}^i|^2} \begin{pmatrix} x_1 - X_{\tilde{k}}^{\tilde{i}+3-l} \\ x_2 - Y_{\tilde{k}}^{\tilde{i}+3-l} \end{pmatrix} \begin{pmatrix} X_{\tilde{k}}^{\tilde{i}+3-l} - X_{\tilde{k}}^{\tilde{i}+l} \\ Y_{\tilde{k}}^{\tilde{i}+3-l} - Y_{\tilde{k}}^{\tilde{i}+l} \end{pmatrix}$$

Funktionen in M_h : Damit läßt sich jede Funktion I aus M_h schreiben als

$$I(x) = \sum_{e_{\tilde{k}}^i=1}^{NE} \sum_{j=1}^2 \sum_{l=1}^2 I_{(e_{\tilde{k}}^i)}^{jl} \cdot m_{(e_{\tilde{k}}^i)}^{jl}(x)$$

für geeignete reelle $I_{(e_{\tilde{k}}^i)}^{jl}$.

Dimension von M_h : $\dim(M_h) = 4NE$

7.2 Matrixschreibweise und Berechnung der Integrale

Die Gleichungen (7.1) (Seite 39) lassen sich als lineares Gleichungssystem schreiben:

$$\begin{pmatrix} A & B & -C & 0 \\ -B^t & 0 & 0 & D \\ C^t & 0 & 0 & 0 \\ 0 & D^t & 0 & 0 \end{pmatrix} \begin{pmatrix} \Sigma \\ U \\ M \\ P \end{pmatrix} = \begin{pmatrix} E \\ F \\ 0 \\ L \end{pmatrix} \quad (7.2)$$

Mit den entsprechenden Integralen:

$$\begin{aligned}
A &\hat{=} (\sigma_h, \tau)_T \\
B &\hat{=} (\nabla \cdot \tau, u_h)_T \\
B^t &\hat{=} (\nabla \cdot \sigma_h, v)_T \\
C &\hat{=} \langle \tau \cdot n, m_h \rangle_{\partial T} \\
C^t &\hat{=} \langle \sigma_h \cdot n, I \rangle_{\partial T} \\
D &\hat{=} (\nabla p_h, v)_T \\
D^t &\hat{=} (u_h, \nabla q)_T \\
E &\hat{=} \langle \tau \cdot n, u_0 \rangle \\
F &\hat{=} (f, v) \\
L &\hat{=} \langle u_0 \cdot n, q \rangle
\end{aligned}$$

7.2.1 Elimination

Wie im nächsten Kapitel gezeigt wird, ist die Matrix A leicht zu invertieren (sie ist diagonal). Damit kann die erste Gleichung aus (7.2) umgeschrieben und Σ ganz eliminiert werden

$$\Sigma = A^{-1}E - A^{-1}BU + A^{-1}CM.$$

Setzen wir dieses Σ in obiges Gleichungssystem ein, so erhalten wir folgendes neue System:

$$\begin{pmatrix} B^t A^{-1} B & -B^t A^{-1} C & D \\ -C^t A^{-1} B & C^t A^{-1} C & 0 \\ D^t & 0 & 0 \end{pmatrix} \begin{pmatrix} U \\ M \\ P \end{pmatrix} = \begin{pmatrix} F + B^t A^{-1} E \\ -C^t A^{-1} E \\ L \end{pmatrix}$$

Wie gleich berechnet wird, ist auch $B^t A^{-1} B$ diagonal und U kann aus dieser Gleichung eliminiert werden.

$$U = (B^t A^{-1} B)^{-1} (F + B^t A^{-1} E + B^t A^{-1} CM - DP)$$

Dies eingesetzt ergibt die letzte eliminierte Gleichung:

$$\begin{pmatrix} C^t A^{-1} C - C^t A^{-1} B (B^t A^{-1} B)^{-1} B^t A^{-1} C & C^t A^{-1} B (B^t A^{-1} B)^{-1} D \\ D^t (B^t A^{-1} B)^{-1} B^t A^{-1} C & -D^t (B^t A^{-1} B)^{-1} D \end{pmatrix} \begin{pmatrix} M \\ P \end{pmatrix} = \begin{pmatrix} \mathcal{A} \\ \mathcal{B} \end{pmatrix} \quad (7.3)$$

mit

$$\begin{pmatrix} \mathcal{A} \\ \mathcal{B} \end{pmatrix} = \begin{pmatrix} -C^t A^{-1} E + C^t A^{-1} B (B^t A^{-1} B)^{-1} (F + B^t A^{-1} E) \\ L - D^t (B^t A^{-1} B)^{-1} (F + B^t A^{-1} E) \end{pmatrix}.$$

7.2.2 Berechnung von A

Behauptung:

$$(\tau_k^{i[m]}, \tau_{\tilde{k}}^{\tilde{i}[\tilde{m}]}) = \begin{cases} \frac{1}{12} |\det B_k| & \text{für } i = \tilde{i}, m = \tilde{m}, k = \tilde{k} \\ 0 & \text{sonst} \end{cases}$$

Beweis

Für $k \neq \tilde{k}$ ist die Behauptung trivial, da $\tau_k^{i[m]}$ und $\tau_{\tilde{k}}^{\tilde{i}[\tilde{m}]}$ unterschiedliche Träger haben.

Sei also $k = \tilde{k}$ und $\tilde{i} \neq i$ o.B.d.A $\tilde{i} = i + 1$:

$$\begin{aligned} (\tau_k^{i[m]}, \tau_{\tilde{k}}^{\tilde{i}[\tilde{m}]}) &= (\tau_k^{i[m]}, \tau_k^{i+1[\tilde{m}]}) \\ &= \delta_m^{\tilde{m}} \int_{T_k} (\lambda_k^{i+1} + \lambda_k^{i+2} - \lambda_k^i) (\lambda_k^{i+2} + \lambda_k^i - \lambda_k^{i+1}) \\ &= \delta_m^{\tilde{m}} \int_{T_k} \lambda_k^{i+1} \lambda_k^{i+2} + \lambda_k^{i+1} \lambda_k^i - \lambda_k^{i+1} \lambda_k^{i+1} \\ &\quad + \lambda_k^{i+2} \lambda_k^{i+2} + \lambda_k^{i+2} \lambda_k^i - \lambda_k^{i+2} \lambda_k^{i+1} \\ &\quad - \lambda_k^i \lambda_k^{i+2} - \lambda_k^i \lambda_k^i + \lambda_k^i \lambda_k^{i+1} \\ &= \delta_m^{\tilde{m}} |T_k| \left(2 \cdot \frac{1}{24} - \frac{1}{12} + \frac{1}{12} - \frac{1}{12} \right) \\ &= 0 \end{aligned}$$

Jetzt sei $i = \tilde{i}, k = \tilde{k}, m = \tilde{m}$:

$$\begin{aligned} (\tau_k^{i[m]}, \tau_k^{i[m]}) &= \int_{T_k} (\lambda_k^{i+1} + \lambda_k^{i+2} - \lambda_k^i) (\lambda_k^{i+1} + \lambda_k^{i+2} - \lambda_k^i) \\ &= \int_{T_k} \lambda_k^{i+1} \lambda_k^{i+1} + \lambda_k^{i+1} \lambda_k^{i+2} - \lambda_k^{i+1} \lambda_k^i \\ &\quad + \lambda_k^{i+2} \lambda_k^{i+1} + \lambda_k^{i+2} \lambda_k^{i+2} - \lambda_k^{i+2} \lambda_k^i \\ &\quad - \lambda_k^i \lambda_k^{i+1} - \lambda_k^i \lambda_k^{i+2} + \lambda_k^i \lambda_k^i \\ &= |T_k| \left(\frac{1}{12} + \frac{1}{24} - \frac{1}{24} + \frac{1}{24} + \frac{1}{12} - \frac{1}{24} - \frac{1}{24} - \frac{1}{24} + \frac{1}{12} \right) \\ &= \frac{1}{6} |T_k| \\ &= \frac{1}{12} |\det B_k| \end{aligned}$$

□

7.2.3 Berechnung von B

$$\begin{aligned}
 \sum_{T_k} (\nabla \cdot \tau_k^{i[m]}, u_k^j)_{T_k} &= \int_{T_k} \nabla \cdot \tau_k^{i[m]} \cdot u_k^j \\
 &= \int_{T_k} \nabla \cdot \left\{ \begin{pmatrix} \delta_m^{11} & \delta_m^{12} \\ \delta_m^{21} & \delta_m^{22} \end{pmatrix} (\lambda_k^{i+1} + \lambda_k^{i+2} - \lambda_k^i) \right\} \begin{pmatrix} \delta_1^j \\ \delta_2^j \end{pmatrix} \\
 &= \int_{T_k} (\delta_m^{j1} \partial_1 + \delta_m^{j2} \partial_2) (\lambda_k^{i+1} + \lambda_k^{i+2} - \lambda_k^i) \\
 &= \int_{T_k} \delta_{m_1}^j \partial_{m_2} (\lambda_k^{i+1} + \lambda_k^{i+2} - \lambda_k^i) \\
 &= \delta_{m_1}^j \cdot \left(+\frac{1}{2} n_{k,m_2}^{i+1} + \frac{1}{2} n_{k,m_2}^{i+2} - \frac{1}{2} n_{k,m_2}^i \right) \\
 &= -\delta_{m_1}^j n_{k,m_2}^i
 \end{aligned}$$

(Achtung: n_k^i bezeichnet die innere Normale im Dreieck T_k .)

7.2.4 Berechnung von $B^t A^{-1} B$

Die Matrix $B(k, i, m; \tilde{k}, j)$ hat nur Einträge, falls $k = \tilde{k}$. Die Einträge in diesem Fall betragen

$$B(k, i, m; k, j) = \delta_{m_1}^j \cdot n_{k,m_2}^i$$

Für festes k und j ist also

$$(B^t B)_{k,j} = \sum_{i=1}^3 \sum_{m_2=1}^2 (n_{k,m_2}^i)^2$$

Und weil $(A^{-1})_k = \frac{12}{|\det B_k|}$, ist

$$(B^t A^{-1} B)_{k,j} = \frac{12}{|\det B_k|} \sum_{m_2=1}^2 \sum_{i=1}^3 (n_{k,m_2}^i)^2$$

Berechnung von $(B^t A^{-1} B)^{-1}$ Weil $B^t A^{-1} B$ diagonal ist, so ist die Berechnung der Inversen ganz leicht.

$$[(B^t A^{-1} B)^{-1}]_{k,j} = \left(\frac{|\det B_k|}{12 \cdot \sum_{m_2=1}^2 \sum_{i=1}^3 (n_{k,m_2}^i)^2} \right)$$

7.2.5 Berechnung von C

Die Matrix C hat folgende Einträge:

$$C_{(k,i,m;e,j,l)} := \sum_{\hat{k}=1}^{NT} \int_{\partial T_{\hat{k}} \setminus \partial \Omega} m_e^{j_l} \cdot \tau_k^{i[m]}(-\underline{n}_{\hat{k}})$$

Die Berechnung dieser Integrale ist nicht weiter schwierig. Da sich diese Arbeit auf lineare Elemente beschränkt, kann das Produkt von zwei linearen Elementen höchstens quadratisch sein und wir können die Integrale mit der Simpson-Regel berechnen. Da $C_{(k,i,m;e,j,l)}$ von vielen Parametern abhängt wird C mit einer Fallunterscheidung angegeben:

1.Fall: e ist Randkante Liegt e auf dem Rand, so ist $m_e^{j_l}$ nach Definition Null und somit auch C :

$$C_{(k,i,m;e,j,l)} = 0$$

2.Fall: k grenzt nicht an die Kante e In diesem Fall sind die Träger von $\tau_k^{i[m]}$ und von $m_e^{j_l}$ disjunkt und auch hier gilt:

$$C_{(k,i,m;e,j,l)} = 0$$

3.Fall: $\text{ITMDPT}(\tilde{i}, k) = e = (e_{\tilde{k}}^{\tilde{i}})$ und $k > \text{ITEDGE}(\tilde{i}, k)$ Hier zeigt die Kante e im Gegenuhrzeigersinn im Dreieck T_k ; der vordere Punkt von e ($l = 1$) trägt also die höhere lokale Nummer i als der hintere ($l = 2$) Punkt der Kante e .

Hier kann die Basisfunktion $\tau_k^{i[m]}$ aber auf drei verschiedene Arten relativ zur Kante e liegen. Im ersten Fall ist einfach $\tilde{i} = i$:

$$C_{(k,i,m;e,j,k)} = -\delta_j^{m_1} \cdot \underline{n}_{k,m_2}^{\tilde{i}} \frac{|e|}{2}.$$

Oder es ist ($i + 1 = \tilde{i}$ und $l = 1$ (vorn)) oder ($i + 2 = \tilde{i}$ und $l = 2$ (hinten)):

$$C_{(k,i,m;e,j,k)} = +\delta_j^{m_1} \cdot \underline{n}_{k,m_2}^{\tilde{i}} \frac{|e|}{6}.$$

Oder es ist ($i + 1 = \tilde{i}$ und $l = 2$ (hinten)) oder ($i + 2 = \tilde{i}$ und $l = 2$ (vorn)):

$$C_{(k,i,m;e,j,k)} = -\delta_j^{m_1} \cdot \underline{n}_{k,m_2}^{\tilde{i}} \frac{|e|}{6}.$$

4.Fall: $\text{ITMDPT}(\tilde{i}, k) = e$ und $k < \text{ITEDGE}(\tilde{i}, k)$ Hier zeigt die Kante e im Uhrzeigersinn; der vordere Punkt von e ($l = 1$) trägt also die kleinere lokale Nummer i als der hintere ($l = 2$) Punkt der Kante e .

Im ersten Fall ist wieder $\tilde{i} = i$:

$$C_{(k,i,m;e,j,k)} = -\delta_j^{m_1} \cdot \underline{n}_{k,m_2}^{\tilde{i}} \frac{|e|}{2}.$$

Oder es ist $(i + 1 = \tilde{i}$ und $l = 1$ (vorn)) oder $(i + 2 = \tilde{i}$ und $l = 2$ (hinten)):

$$C_{(k,i,m;e,j,k)} = -\delta_j^{m_1} \cdot \underline{n}_{k,m_2}^{\tilde{i}} \frac{|e|}{6}.$$

Oder es ist $(i + 1 = \tilde{i}$ und $l = 2$ (hinten)) oder $(i + 2 = \tilde{i}$ und $l = 2$ (vorn)):

$$C_{(k,i,m;e,j,k)} = +\delta_j^{m_1} \cdot \underline{n}_{k,m_2}^{\tilde{i}} \frac{|e|}{6}.$$

7.2.6 Berechnung von D

Bei der Matrix D handelt es sich um genau dieselben Integrale wie im ersten Programm bei der dortigen Matrix C . Die Berechnung und die Resultate sind dieselben wie bei 6.2.3 auf Seite 33.

7.2.7 Berechnung von E

Der Vektor E enthält die Komponenten

$$\langle \tau_k^{i[m]} \cdot n, u_0 \rangle_{\partial\Omega}.$$

Auch hier sei wieder erwähnt, daß die Normale n die auf Länge 1 normierte **äußere** Normale an Ω ist, wohingegen n_k immer die **innere** Normale zum Dreieck T_k bezeichnet.

$$\begin{aligned} \langle \tau_k^{i[m]} \cdot n, u_0 \rangle_{\partial\Omega} &= -\int_{\partial T_k} (\tau_k^{i[m]} \cdot n_k) u_0 \\ &= -\int_{e_k^i} \begin{pmatrix} \delta_m^{11} & \delta_m^{12} \\ \delta_m^{21} & \delta_m^{22} \end{pmatrix} (\lambda_k^{i+1} + \lambda_k^{i+2} - \lambda_k^i) \cdot n_k \cdot u_0 \\ &= -\int_{e_k^i} (\lambda_k^{i+1} + \lambda_k^{i+2} - \lambda_k^i) \frac{n_{k,m_2}^i}{|e_k^i|} u_0^{m_1} \\ &\approx -n_{k,m_2}^i \cdot u_0^{m_1}(e_k^i) \end{aligned}$$

Hierbei wird $u_0^{m_1}(e_k^i)$ konstant interpoliert, d.h. $u_0(e_k^i)$ bezeichnet einfach den Wert der Funktion u_0 im Mittelpunkt der Kante e_k^i .

7.2.8 Berechnung von F

Die Berechnung von F geht wie im ersten Programm. (Vgl. 6.2.5 auf Seite 34.)

7.2.9 Berechnung von L

Die Integrale sind genau dieselben wie in 6.2.6 auf Seite 35.

7.3 Implementierung

7.3.1 Konjugierte Residuen und Skalierung

Wie im ersten Programm wird auch hier das Gleichungssystem mit den konjugierten Residuen mit Lanczos-Schritt gelöst. Auch wird lediglich eine Skalierung vorgenommen, damit alle Einträge der Stokes-Matrix wenigstens die gleiche Größenordnung haben. Dazu berechnet man zuerst die Ordnungen der einzelnen Teilmatrizen.

$$\begin{aligned}A &= \mathcal{O}(h^2) \\ B &= \mathcal{O}(h) \\ C &= \mathcal{O}(h) \\ D &= \mathcal{O}(h)\end{aligned}$$

Somit ergibt sich, daß die effektive Stokes-Matrix von links und von rechts mit

$$\begin{pmatrix} h & 0 \\ 0 & I \end{pmatrix}$$

multipliziert werden muß, um in allen Komponenten eine Ordnung von h^2 zu haben.

7.3.2 Implementierung von STOLIN

Da die Stokes-Matrix ((7.3) auf Seite 42) aus Multiplikationen von vielen einzelnen Matrizen besteht, wäre die Fehleranfälligkeit sehr groß, wenn die Matrizen gleich in einem Unterprogramm berechnet würden. Es wurden nur die Matrix \times Vektor-Multiplikationen

$$A^{-1}, B, B^t, C, C^t, D, D^t, (B^t A^{-1} B)^{-1}$$

implementiert. Die Fehleranfälligkeit ist viel geringer, zudem können die einzelnen Matrix-Vektor-Multiplikationen viel einfacher getestet werden. Leider geht der Speicherplatzgewinn der Elimination zum Teil wieder verloren, da zeitweilig (in Unterprogrammen) wieder Vektoren der Längen von $\Sigma = 12NT$ auftreten.

8 Vergleich der Programme

8.1 Speicherplatz

Zuerst eine Bezeichnung:

Sei NV die Anzahl der Eckpunkte, NT die Anzahl der Dreiecke und NE die Zahl der Kanten in einer Triangulierung. Wie in Lemma A.1.7 (s.70) bewiesen wird gelten die Annäherungen

$$NT \approx 2NV$$

und

$$NE \approx 3NV.$$

Wenn aber die Dreiecke der niedrigeren Levels auch immer noch in den Vektoren mitbehalten werden (z.B. für Mehrgitteralgorithmen) rechnet man (bei gleichmäßiger Unterteilung jedes Dreieckes in 2 Dreiecke) vorsichtiger mit $NT \approx 4NV$.

Dimensionen der Vektorräume:

	1.Programm	2.Programm
σ_h	2 $NE = 6NV$	12 $NT = 24NV$
u_h	2 $NT = 4NV$	2 $NT = 4NV$
p_h	NV	NV
m_h	–	4 $NE = 12NV$
Total	11 NV	41 NV
Nach Elimination	–	13 NV

Doch wie schon erwähnt, werden bei meiner Wahl der Implementierung auch beim 2.Programm die Vektoren σ_h und u_h in Unterprogrammen noch auftreten, sodaß der durch die Elimination gewonnene Speicher zum Teil trotzdem belegt wird.

8.2 Geschwindigkeit und Genauigkeit

Getestet wurden die beiden Programme mit folgenden fünf Funktionspaaren (u,p):

1. quad.f:

$$u(x, y) = 200 \cdot \begin{pmatrix} x^2(1-x)^2y(1-y)(1-2y) \\ -y^2(1-y)^2x(1-x)(1-2x) \end{pmatrix}$$

$$p(x, y) = 100x(1-x)y(1-y) - \frac{25}{9}$$

2. trig.f:

$$u(x, y) = \pi \cdot \begin{pmatrix} \sin(4\pi y)(1 - \cos(4\pi x)) \\ -\sin(4\pi x)(1 - \cos(4\pi y)) \end{pmatrix}$$

$$p(x, y) = 4\pi^2 \sin(4\pi x) \sin(4\pi y)$$

3. pl1.f:

$$u(x, y) = \begin{pmatrix} y(1-y) \\ 0 \end{pmatrix}$$
$$p(x, y) = 1 - 2x$$

4. pl2.f:

$$u(x, y) = \begin{pmatrix} y^2 \\ 0 \end{pmatrix}$$
$$p(x, y) = 2x - 2$$

5. csl.f: (in Polarkoordinaten)

$$u(r, \phi) = \frac{3}{4}\sqrt{r} \sin \frac{\phi}{2} \cdot \begin{pmatrix} \sin \phi \\ 1 - \cos \phi \end{pmatrix}$$
$$p(x, y) = \frac{-1.5}{\sqrt{r}} \cos \frac{\phi}{2}$$

Zu diesen fünf Funktionspaaren gehören sieben Triangulierungen: vergleiche dazu Abbildung 4 auf Seite 50.

Zu jedem Funktionspaar, zugehörigem Gebiet und aktuellem Level werden für beide Programme jeweils folgende Werte angegeben:

- Durchschnittliche Seitenlänge der Dreiecke: $h(\text{Mittel})$ oder \bar{h} .
- Anzahl der Unbekannten.
- Die Anzahl der Iterationen ($=: n$) bis das Problem ausiteriert war; oder das Maximum der Iterationen. Falls das Maximum der Iterationen vor einer genügende Genauigkeit erreicht wurde, so wird dieser Wert mit einem Stern versehen.
- Totale Zeit und die Zeit pro Iteration.
- Konvergenzraten: Die Konvergenzraten (CR) geben ein Maß für die Effizienz des Lösers (hier: konjugierte Residuen vgl. (6.3.1)). Für jeden Iterationsschritt sei die Konvergenzrate folgendermaßen gegeben:

$$\text{CR}_i := \left(\frac{\|Ax_i - b\|}{\|Ax_0 - b\|} \right)^{\frac{1}{i}}.$$

Damit ist die beste Konvergenzrate

$$\text{CR}(\text{best}) := \min_{1 \leq i \leq n} \text{CR}_i$$

die schlechteste Konvergenzrate

Abbildung 4: Triangulierungen zu den Funktionspaaren

$$\text{CR}(\text{worst}) := \max_{1 \leq i \leq n} \text{CR}_i$$

die letzte Konvergenzrate

$$\text{CR}(\text{last}) := \text{CR}_n$$

die durchschnittliche Konvergenzrate

$$\text{CR}(\text{mean}) := \frac{1}{n} \sum_{i=1}^n \text{CR}_i$$

und die Konvergenzrate pro Sekunde ($s = \text{Anzahl Sekunden}$)

$$\text{CR}(\text{sec}) := (\text{CR}_n)^{\frac{n}{s}}.$$

- Der relative Verfahrensfehler (Residuum):

$$\frac{\| Ax_n - b \|}{\| Ax_0 - b \|}$$

Jeweils in der Geschwindigkeitskomponente, in der Druckkomponente und über den ganzen Vektor.

- Der relative Fehler der Lösung (x^* bezeichnet die Interpolation der exakten Lösung):

$$\frac{\| x_n - x^* \|}{\| x^* \|}$$

Jeweils in der Geschwindigkeitskomponente, in der Druckkomponente und über den ganzen Vektor.

Die Abbruchbedingung für die konjugierten Residuen war einerseits eine maximal vorgegebene Anzahl Iterationen, und andererseits wurde jeweils auch abgebrochen, wenn die Druckkomponente im Vektor genügend genau war:

$$\| x_n - x^* \|_0(p) < \frac{\bar{h}}{10}$$

Dabei ist \bar{h} die durchschnittliche Seitenlänge der Elemente.

8.2.1 Die Funktionen quad.f auf q_0

	1.Programm	2.Programm	1.Programm	2.Programm
Level	2	2	3	3
h(Mittel)	0.1768	0.1768	8.839E-2	8.839E-2
Unbekannte	377	457	1457	1745
Iterationen	50	73	138	112
Zeit (total)	19 Sek.	1314 Sek.	297 Sek.	8391 Sek.
t /Iteration	0.38 Sek.	18 Sek.	2.15 Sek.	74.92 Sek.
CR(mean)	0.88	0.9314	0.9326	0.9554
CR(sec)	0.7431	0.9964	0.9811	0.9995
CR(best)	0.8259	0.8937	0.7939	0.9216
CR(worst)	0.9527	0.9623	0.9599	0.9631
CR(last)	0.8933	0.9375	0.9599	0.9631
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u)$	2.699E-8	4.951E-9	3.940E-8	3.998E-8
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(p)$	1.333E-5	3.614E-6	1.092E-6	1.820E-5
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u, p)$	1.977E-7	4.593E-8	1.589E-7	9.159E-8
$\frac{\ x_n - x^*\ }{\ x^*\ }(u)$	5.944E-4	4.412E-4	5.435E-4	2.428E-4
$\frac{\ x_n - x^*\ }{\ x^*\ }(p)$	9.440E-3	8.869E-3	4.810E-3	4.727E-3
$\frac{\ x_n - x^*\ }{\ x^*\ }(u, p)$	2.801E-3	4.116E-3	1.442E-3	1.678E-3

8.2.2 Die Funktionen quad.f auf q

	1.Programm	2.Programm	1.Programm	2.Programm
Level	2	2	3	3
h(Mittel)	0.125	0.125	6.25E-2	6.25E-2
Unbekannte	753	913	2913	3489
Iterationen	68	96	206	328
Zeit (total)	79 Sek.	3451 Sek.	880 Sek.	14735 Sek.
t /Iteration	1.162 Sek.	35.95 Sek.	4.272 Sek.	44.92 Sek.
CR(mean)	0.7790	0.9460	0.9515	0.9817
CR(sec)	0.9309	0.9987	0.9939	0.99987
CR(best)	0.8021	0.9176	0.8135	0.9355
CR(worst)	0.9398	0.9761	0.9744	0.9927
CR(last)	0.9202	0.9532	0.9744	0.9869
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u)$	9.204E-8	3.226E-9	2.016E-7	8.151E-9
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(p)$	4.657E-6	5.130E-6	7.712E-7	3.766E-6
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u, p)$	2.709E-7	5.022E-8	3.317E-7	2.524E-8
$\frac{\ x_n - x^*\ }{\ x^*\ }(u)$	8.066E-4	2.774E-4	1.044E-3	1.583E-4
$\frac{\ x_n - x^*\ }{\ x^*\ }(p)$	6.009E-3	6.876E-3	3.405E-3	3.376E-3
$\frac{\ x_n - x^*\ }{\ x^*\ }(u, p)$	2.185E-3	2.797E-3	1.517E-3	1.036E-3

8.2.3 Die Funktionen quad.f auf qs

	1.Programm	2.Programm	1.Programm	2.Programm
Level	2	2	3	3
h(Mittel)	0.1768	0.1768	8.839E-2	8.839E-2
Unbekannte	377	457	1457	1745
Iterationen	315	481	1000*	1000*
Zeit (total)	168 Sek.	8558 Sek.	2057 Sek.	73182 Sek.
t /Iteration	0.5333 Sek.	17.79 Sek.	2.057 Sek.	73.18 Sek.
CR(mean)	0.9737	0.9827	0.9877	0.9922
CR(sec)	0.9683	0.9994	0.9973	0.99996
CR(best)	0.8890	0.9297	0.8570	0.9535
CR(worst)	0.9830	0.9903	0.9945	0.9971
CR(last)	0.9830	0.9899	0.9945	0.9971
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u)$	1.604E-7	3.082E-8	7.476E-8	8.468E-6
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(p)$	1.674E-7	1.516E-5	9.856E-7	1.414E-4
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u, p)$	2.431E-7	6.233E-8	1.115E-7	8.655E-6
$\frac{\ x_n - x^*\ }{\ x^*\ }(u)$	7.036E-4	8.950E-4	3.469E-4	2.828E-3
$\frac{\ x_n - x^*\ }{\ x^*\ }(p)$	1.012E-2	9.993E-3	4.983E-3	3.189E-2
$\frac{\ x_n - x^*\ }{\ x^*\ }(u, p)$	3.161E-3	5.215E-3	1.467E-3	1.274E-2

Hier sieht man deutlich, daß die Programme auf Level 3 nicht ausiteriert waren: beim 2.Programm ist der absolute Fehler in der feineren Triangulierung schlechter approximiert worden.

8.2.4 Die Funktionen trig.f auf q_0

	1.Programm	2.Programm	1.Programm	2.Programm
Level	2	2	3	3
h(Mittel)	0.1768	0.1768	8.839E-2	8.839E-2
Unbekannte	377	457	1457	1745
Iterationen	92	80	118	177
Zeit (total)	49 Sek.	1421 Sek.	194 Sek.	13110 Sek.
t /Iteration	0.5326 Sek.	17.76 Sek.	1.644 Sek.	74.07 Sek.
CR(mean)	0.8874	0.9338	0.9051	0.9503
CR(sec)	0.8473	0.9946	0.9574	0.9994
CR(best)	0.7080	0.9011	0.7997	0.8799
CR(worst)	0.9989	0.9964	0.9915	0.9708
CR(last)	0.9155	0.9089	0.9310	0.9586
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u)$	2.552E-10	5.281E-10	3.386E-10	1.234E-10
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(p)$	6.155E-6	1.544E-6	4.221e-8	3.712E-8
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u, p)$	2.377E-9	5.541E-10	1.258E-9	3.616E-10
$\frac{\ x_n - x^*\ }{\ x^*\ }(u)$	1.834E-4	7.579E-5	1.076E-4	7.107E-5
$\frac{\ x_n - x^*\ }{\ x^*\ }(p)$	1.013E-3	6.311E-4	5.291E-4	5.450E-4
$\frac{\ x_n - x^*\ }{\ x^*\ }(u, p)$	2.978E-4	4.818E-4	1.832E-4	3.102E-4

8.2.5 Die Funktionen trig.f auf q

	1.Programm	2.Programm	1.Programm	2.Programm
Level	2	2	3	3
h(Mittel)	0.125	0.125	6.250E-2	6.250E-2
Unbekannte	753	913	2913	3489
Iterationen	135	112	324	613
Zeit (total)	107 Sek.	3896 Sek.	1039 Sek.	87713 Sek.
t /Iteration	0.7926 Sek.	34.79 Sek.	8.379 Sek.	143.1 Sek.
CR(mean)	0.9170	0.9402	0.9515	0.9806
CR(sec)	0.9244	0.9981	0.9918	0.999915
CR(best)	0.8284	0.9227	0.8149	0.9569
CR(worst)	0.9981	0.9863	0.9936	1*
CR(last)	0.9396	0.9345	0.9740	0.9880
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u)$	2.530E-10	2.553E-10	8.860E-11	9.896E-12
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(p)$	4.831E-6	7.649E-7	2.987E-9	6.227E-9
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u, p)$	1.171E-9	4.889E-10	4.243E-10	7.008E-11
$\frac{\ x_n - x^*\ }{\ x^*\ }(u)$	9.320E-5	9.579E-5	7.841E-5	2.443E-5
$\frac{\ x_n - x^*\ }{\ x^*\ }(p)$	8.779E-4	8.802E-4	3.485E-4	3.467E-4
$\frac{\ x_n - x^*\ }{\ x^*\ }(u, p)$	2.220E-4	5.080E-4	1.121E-4	1.856E-4

* Es ist möglich, daß die Geschwindigkeitskomponente divergiert, während die Druckkomponente konvergiert und vice versa, obschon das gesamte Gleichungssystem konvergiert. Somit kann es passieren, daß die Konvergenzrate 1 übersteigt, wie das hier der Fall war. Das muß aber nicht heißen, daß diese Iteration schlecht war.

8.2.6 Die Funktionen trig.f auf qs

	1.Programm	2.Programm	1.Programm	2.Programm
Level	2	2	3	3
h(Mittel)	0.1768	0.1768	8.839E-2	8.839E-2
Unbekannte	377	457	1457	1745
Iterationen	748	775	1000*	1000*
Zeit (total)	330 Sek.	13446 Sek.	2050 Sek.	71640 Sek.
t /Iteration	0.4412 Sek.	17.34 Sek.	2.050 Sek.	71.64 Sek.
CR(mean)	0.9789	0.9910	0.9874	0.9942
CR(sec)	0.9756	0.9995	0.9972	0.999965
CR(best)	0.7791	0.9759	0.8577	0.9655
CR(worst)	0.9890	0.9960	0.9942	9975
CR(last)	0.9890	0.9911	0.9942	0.9975
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u)$	6.50E-11	1.29E-5	4.676E-8	2.23E-1
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(p)$	7.98E-8	2.61E-3	1.194E-6	8.56E-4
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u, p)$	3.16E-10	1.42E-5	5.759E-8	2.31E-1
$\frac{\ x_n - x^*\ }{\ x^*\ }(u)$	1.27E-4	2.52E-4	1.024E-3	1.86E-2
$\frac{\ x_n - x^*\ }{\ x^*\ }(p)$	1.76E-3	1.69E-3	7.268E-3	1.12E-1
$\frac{\ x_n - x^*\ }{\ x^*\ }(u, p)$	3.90E-4	1.09E-3	2.345E-3	6.38E-2

Es sind beide Programme nicht ausiteriert. Bei gleich vielen Iterationen (1000) ist das Resultat des 1.Programmes genauer.

8.2.7 Die Funktionen `trig.f` auf `slit`

	1.Programm	2.Programm	1.Programm	2.Programm
Level	2	2	3	3
h(Mittel)	0.1768	0.1768	8.839E-2	8.893E-2
Unbekannte	1469	1765	5753	6857
Iterationen	250	244	735	1000*
Zeit (total)	406 Sek.	2170 Sek.	4784 Sek.	365000 Sek.
t /Iteration	1.624 Sek.	8.893 Sek.	6.509 Sek.	365 Sek.
CR(mean)	0.9336	0.9540	0.9668	0.9818
CR(sec)	0.9780	0.9996	0.9980	0.99995
CR(best)	0.6940	0.9000	0.7967	0.8748
CR(worst)	0.9995	0.9975	0.9953	0.9924
CR(last)	0.9645	0.9686	0.9872	0.9924
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u)$	4.311E-12	3.04E-11	3.174E-13	2.09E-12
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(p)$	1.185E-7	1.16E-6	2.895E-10	1.64E-10
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u, p)$	4.448E-11	1.62E-11	9.095E-12	2.45E-12
$\frac{\ x_n - x^*\ }{\ x^*\ }(u)$	1.519E-4	1.10E-4	7.972E-5	2.60E-5
$\frac{\ x_n - x^*\ }{\ x^*\ }(p)$	5.426E-4	5.34E-4	2.688E-4	4.72E-4
$\frac{\ x_n - x^*\ }{\ x^*\ }(u, p)$	1.196E-4	4.12E-4	6.335E-5	2.65E-4

Obschon nicht ausiteriert, ist wie üblich die Geschwindigkeitskomponente im 2.Programm auf Level 3 besser approximiert worden als im 1.Programm!

8.2.8 Die Funktionen `cs1.f` auf `cs1`

	1.Programm	2.Programm	1.Programm	2.Programm
Level	2	2	3	3
h(Mittel)	0.3093	0.3093	0.1562	0.1562
Unbekannte	389	477	1481	1785
Iterationen	127	143	293	414
Zeit (total)	53 Sek.	2954 Sek.	476 Sek.	3650 Sek.
$t/\text{Iteration}$	0.417 Sek.	20.66 Sek.	1.624 Sek.	88.16 Sek.
CR(mean)	0.9656	0.9650	0.9799	0.9804
CR(sec)	0.9306	0.9986	0.9911	0.99975
CR(best)	0.9230	0.8014	0.9310	0.8013
CR(worst)	0.9951	0.9761	0.9985	0.9886
CR(last)	0.9704	0.9714	0.9857	0.9886
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u)$	2.05E-7	1.20E-9	2.86E-8	3.76E-11
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(p)$	1.09E-4	9.89E-6	5.73E-6	6.21E-7
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u, p)$	2.34E-6	1.26E-8	7.24E-7	8.71E-10
$\frac{\ x_n - x^*\ }{\ x^*\ }(u)$	1.59E-3	7.63E-4	8.03E-4	3.84E-4
$\frac{\ x_n - x^*\ }{\ x^*\ }(p)$	1.36E-2	1.39E-2	6.26E-3	6.47E-3
$\frac{\ x_n - x^*\ }{\ x^*\ }(u, p)$	8.69E-3	6.77E-3	4.11E-3	2.52E-3

8.2.9 Die Funktionen pl1.f auf pl

	1.Programm	2.Programm	1.Programm	2.Programm
Level	2	2	3	3
h(Mittel)	0.125	0.125	6.25E-2	6.25E-2
Unbekannte	1481	1785	5777	6897
Iterationen	96	127	212	638
Zeit (total)	154 Sek.	11200 Sek.	1361 Sek.	231400 Sek.
t /Iteration	1.604 Sek.	88.19 Sek.	6.42 Sek.	362.7 Sek.
CR(mean)	0.9443	0.9606	0.9698	0.9840
CR(sec)	0.9807	0.9997	0.9978	0.999985
CR(best)	0.8502	0.7451	0.8689	0.6949
CR(worst)	0.9694	0.9837	1*	0.9948
CR(last)	0.9692	0.9779	0.9859	0.9947
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u)$	1.69E-6	2.68E-8	6.02E-7	3.70E-10
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(p)$	6.63E-5	4.02E-5	1.29E-5	3.16E-6
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u, p)$	2.86E-5	2.23E-7	1.19E-5	1.56E-8
$\frac{\ x_n - x^*\ }{\ x^*\ }(u)$	3.04E-3	2.66E-3	2.37E-3	1.30E-3
$\frac{\ x_n - x^*\ }{\ x^*\ }(p)$	7.45E-3	7.46E-3	3.82E-3	3.79E-3
$\frac{\ x_n - x^*\ }{\ x^*\ }(u, p)$	7.71E-3	6.01E-3	3.83E-3	2.66E-3

Für * gilt das gleiche wie in `trig.f` auf q .

8.2.10 Die Funktionen pl2.f auf pl

	1.Programm	2.Programm	1.Programm	2.Programm
Level	2	2	3	3
h(Mittel)	0.125	0.125	6.25E-2	6.25E-2
Unbekannte	1481	1785	5777	6897
Iterationen	107	179	272	626
Zeit (total)	175 Sek.	15700 Sek.	1763 Sek.	226400 Sek.
t /Iteration	1.64 Sek.	87.7 Sek.	6.48 Sek.	362 Sek.
CR(mean)	0.9533	0.9391	0.9790	0.9736
CR(sec)	0.9829	0.9997	0.9983	0.999977
CR(best)	0.8686	0.6750	0.9298	0.6788
CR(worst)	1*	0.9750	1*	0.9918
CR(last)	0.9721	0.9749	0.9888	0.9917
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u)$	1.71E-6	3.04E-10	1.65E-6	1.58E-11
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(p)$	1.36E-5	1.59E-6	2.08E-6	2.96E-7
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u, p)$	1.24E-5	4.15E-9	2.08E-6	6.80E-10
$\frac{\ x_n - x^*\ }{\ x^*\ }(u)$	1.92E-3	9.70E-4	9.88E-4	5.32E-4
$\frac{\ x_n - x^*\ }{\ x^*\ }(p)$	7.54E-3	7.52E-3	3.80E-3	3.79E-3
$\frac{\ x_n - x^*\ }{\ x^*\ }(u, p)$	5.51E-3	3.56E-3	2.65E-3	1.37E-3

Für * gilt das gleiche wie in `trig.f` auf q .

8.2.11 Die Funktionen `trig.f` auf `lshape`

	1.Programm	2.Programm	1.Programm	2.Programm
Level	2	2	3	3
h(Mittel)	0.1768	0.1768	8.839E-2	8.839E-2
Unbekannte	1105	1329	4321	5153
Iterationen	167	217	409	643
Zeit (total)	207 Sek.	14400 Sek.	1979 Sek.	175737 Sek.
t /Iteration	1.24 Sek.	66.4 Sek.	4.84 Sek.	273 Sek.
CR(mean)	0.9186	0.9539	0.9499	0.9751
CR(sec)	0.9588	0.9995	0.9953	0.999954
CR(best)	0.6934	0.9001	0.7970	0.8754
CR(worst)	0.9993	0.9977	0.9949	0.9876
CR(last)	0.9492	0.9653	0.9775	0.9875
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u)$	2.21E-10	5.87E-11	1.36E-12	2.03E-12
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(p)$	1.78E-6	1.79E-6	1.06E-9	2.22E-10
$\frac{\ Ax_n - b\ }{\ Ax_0 - b\ }(u, p)$	6.86E-10	1.87E-10	3.30E-11	2.69E-12
$\frac{\ x_n - x^*\ }{\ x^*\ }(u)$	2.42E-4	8.49E-5	1.19E-4	2.38E-5
$\frac{\ x_n - x^*\ }{\ x^*\ }(p)$	5.90E-4	6.18E-4	3.08E-4	3.10E-4
$\frac{\ x_n - x^*\ }{\ x^*\ }(u, p)$	1.65E-4	4.73E-4	7.56E-5	1.74E-4

8.3 Auswertung

Die beiden Gleichungssysteme (vom 1. Programm und vom 2. Programm) unterscheiden sich nicht stark in der Zahl der Unbekannten. Diese verhalten sich vom 1. zum 2. Programm wie 11 zu 13. Im 2. Programm treten in Unterprogrammen wieder Vektoren der Länge $24NV$ auf. Somit verhält sich der Speicherplatz effektiv wie 11 zu 37. Doch schon das Verhältnis 11 zu 13 läßt erwarten, daß das 2. Programm mehr Iterationen braucht, um das Gleichungssystem mit gleicher Genauigkeit zu lösen. Das wird auch mit den meisten Resultaten bestätigt.

Das 2. Programm braucht auf Level 3 immer (außer mit `quad.f` auf q_0) mehr Iterationen als das 1. Programm. Falls die Zahl der Iterationen mit einem Stern (*) versehen sind, darf man die beiden Programme nicht mehr mit der Zahl der Iterationen vergleichen. Es wurde nämlich immer abgebrochen, sobald mehr als 1000 Iterationen gerechnet waren. Die einzigen Vergleichswerte sind dann noch die Zeit pro Iteration und die Konvergenzraten.

Die absoluten Fehler

$$\frac{\|x_n - x^*\|}{\|x^*\|}$$

sind nicht von großer Bedeutung. Sie zeigen lediglich an, daß die Druckkomponente wirklich auf ein gewünschtes Maß reduziert wurde. Sind die Zahlen der absoluten Fehler für beide Programme etwa gleich groß, so kann die Effizienz einfach an der Anzahl der Iterationen oder den Konvergenzraten abgelesen werden. Gerade in der Fehlerreduktion hat das 2. Programm aber einen Vorteil gegenüber dem 1. Programm. Auf Level 3 wurde die Geschwindigkeitskomponente vom 2. Programm immer besser approximiert. Dies wiederum darf natürlich nur ausgesagt werden, falls beide Programme ihren Fehler in der Druckkomponente ganz ausiteriert hatten. Dazu nehme man den absoluten Fehler in p auf Level 3 als 100% an und vergleiche die Fehler in den Geschwindigkeitskomponenten.

Zur Fehlerreduktion sei noch erwähnt, daß diejenigen Funktionen, die nicht auf konvexen Gebieten leben, ebensogut approximiert wurden. Das ist aufgrund der Fehlerabschätzung (Kapitel 4.1) leicht erklärbar. Für den Polynomgrad $k = 1$ hat sie die Ordnung h^2 und zwar unabhängig davon, ob nun die Regularitätsannahme (2.3) erfüllt ist oder nicht: ob nun Ω konvex ist oder nicht!

Die Fehlerreduktion kann am einfachsten an der Konvergenzrate pro Sekunde abgelesen werden. Je näher die Zahl bei 1 ist, umso schlechter die Konvergenzgeschwindigkeit. Doch hier sei nochmals gesagt, daß mit einem Abspeichern der Matrizen das 2. Programm einiges besser abgeschnitten hätte.

Zusammenfassend kann das 2. Programm für $k = 1$ nicht empfohlen werden. Es benötigt mehr Speicher und liefert in unseren Beispielen schlechtere Konvergenz.

A Anhang

A.1 Lemmata

A.1.1 Äquivalenz zur Inf-Sup-Bedingung

Behauptung:

$$\exists C > 0 \in \mathbb{R} : \forall v \in V_h : \exists \gamma \in H_h : \\ (\nabla \cdot \gamma, v) \geq C \|v\|_{1,h}^2$$

und

$$\|\gamma\|_{0,h} \leq \|v\|_{1,h}$$

ist äquivalent zu

$$\exists C > 0 : \inf_{\substack{v \in V_h \\ v \neq 0}} \sup_{\substack{\gamma \in H_h \\ \gamma \neq 0}} \frac{(\nabla \cdot \gamma, v)}{\|v\|_{1,h} \|\gamma\|_{0,h}} \geq C$$

Beweis

“ \implies ”: Sei $v \in V_h$ beliebig, dann existiert nach Voraussetzung ein γ , sodaß

$$(\nabla \cdot \gamma, v) \geq C \|v\|_{1,h}^2$$

und

$$\|\gamma\|_{0,h} \leq \|v\|_{1,h}.$$

Hieraus folgt

$$\sup_{\sigma \in H_h} \frac{(\nabla \cdot \sigma, v)}{\|v\|_{1,h}} \geq \frac{(\nabla \cdot \gamma, v)}{\|v\|_{1,h}} \geq C \|v\|_{1,h}$$

“ \impliedby ”: Sei $v \in V_h$ beliebig. Dann folgt $\exists \gamma$ mit

$$\frac{(\nabla \cdot \gamma, v)}{\|\gamma\|_{0,h} \|v\|_{1,h}} \geq \frac{1}{2} \sup_{\sigma} \frac{(\nabla \cdot \sigma, v)}{\|\sigma\|_{0,h} \|v\|_{1,h}} \geq \frac{1}{2} C$$

Setze $\tilde{\gamma} := \frac{\|v\|_{1,h}}{\|\gamma\|_{0,h}} \gamma$. Dann folgt:

$$\|\tilde{\gamma}\|_{0,h} = \|v\|_{1,h}$$

und

$$(\nabla \cdot \tilde{\gamma}, v) = \frac{\|v\|_{1,h} (\nabla \cdot \gamma, v)}{\|\gamma\|_{0,h} \|v\|_{1,h}} \geq \frac{1}{2} C \|v\|_{1,h}.$$

□

A.1.2 Inverse Abschätzung

Behauptung: Für $u \in V_{h|T}$ ($T \in \mathcal{C}_h$) gilt $\|\nabla u\|_{0,T} \leq |u|_{1,T} \leq Ch^{-1} \|u\|_{0,T}$.

Beweis

$$\hat{u} := u \circ F_T \quad \hat{u} \in V_{h|\hat{T}} \quad (F_T \hat{x} = B_T \hat{x} + b)$$

Die Menge $\{\hat{u} \in V_{h|\hat{T}} : \|\hat{u}\|_{0,\hat{T}} = 1\}$ ist kompakt (, da $V_{h|\hat{T}}$ endlichdimensional). Daher nimmt die stetige Abbildung

$$\hat{u} \longmapsto |\hat{u}|_{1,\hat{T}}$$

auf dieser kompakten Menge ihr Maximum $=: C$ an. (Stetig: $|\hat{u}|_{1,\hat{T}} = \|\nabla \hat{u}\|_{0,\hat{T}}$)

$$\implies \exists C > 0 : \forall \hat{u} \in V_{h|\hat{T}} : \|\hat{u}\|_{0,\hat{T}} = 1 : |\hat{u}|_{1,\hat{T}} \leq C$$

$$\implies \exists C > 0 : \forall \hat{u} \in V_{h|\hat{T}} : |\hat{u}|_{1,\hat{T}} \leq C \cdot \|\hat{u}\|_{0,\hat{T}} \quad (\text{A.1})$$

Also gilt für alle $u \in V_{h|T} : u \in H^1(T)$ (für * siehe [5] Theorem 3.1.2) :

$$\begin{aligned} \|\nabla u\|_{0,T} &= |u|_{1,T} \\ &\stackrel{*}{\leq} C \|B_T^{-1}\|^1 \cdot |\det B_T|^{\frac{1}{2}} |\hat{u}|_{1,\hat{T}} \\ &\stackrel{(\text{A.1})}{\leq} C \|B_T^{-1}\| \cdot |\det B_T|^{\frac{1}{2}} \|\hat{u}\|_{0,\hat{T}} \\ &\stackrel{*}{\leq} C \|B_T^{-1}\| \cdot |\det B_T|^{\frac{1}{2}} \|B_T\|^0 |\det B_T|^{-\frac{1}{2}} \|u\|_{0,T} \\ &= C \|B_T^{-1}\| \cdot \|u\|_{0,T} \\ &\leq C \frac{\hat{h}}{\rho_T} \|u\|_{0,T} \\ &\leq C \cdot h_T^{-1} \|u\|_{0,T}. \end{aligned} \quad (\text{A.2})$$

Dabei gilt die zweitletzte Ungleichung für reguläre Triangulierungen. Vergleiche dazu (2.1) auf Seite 2.

Die letzte Ungleichung gilt wegen der Beziehung:

$$\|B_T^{-1}\| \leq \frac{\hat{h}}{\rho_T}$$

Vergleiche dazu auch [5] Theorem 3.1.3 und Remark 3.1.3.

□

A.1.3 Lemma für den Beweis der Fehlerabschätzung

In der Fehlerabschätzung (Kapitel. 4) wird für $q \in C(\Omega) \cap L_0^2(\Omega)$ und $\tilde{q} \in P_h(\Omega)$ folgende Abschätzung gebraucht:

$$\left(\sum_{e \in \Gamma_h} h_e \int_e |q - \tilde{q}|^2 \right)^{\frac{1}{2}} \leq Ch|q|_1 \quad (\text{A.3})$$

mit $h = \max_{e \in \Gamma_h} h_e$. Dabei ist h_e einfach die Länge der Kante e .

Beweis

Nach Definition ist:

$$\|q - \tilde{q}\|_{0,e}^2 = \int_e |q - \tilde{q}|^2$$

Der Transformationssatz liefert:

$$\int_e |q - \tilde{q}|^2 \leq h_e \int_{\hat{e}} |\hat{q} - \hat{\tilde{q}}|^2 = h_e \|\hat{q} - \hat{\tilde{q}}\|_{0,\hat{e}}^2$$

$$\begin{aligned} \Rightarrow \sum_{e \in \partial T} h_e \|q - \tilde{q}\|_{0,e}^2 &\leq \sum_{\hat{e} \in \partial \hat{T}} h_e^2 \|\hat{q} - \hat{\tilde{q}}\|_{0,\hat{e}}^2 \\ &\leq Ch_T^2 |q|_{1,\hat{T}}^2 \\ &\stackrel{[5]}{\leq} Ch_T^2 |q|_{1,T}^2 \end{aligned}$$

$$\begin{aligned} \Rightarrow \sum_{T \in \mathcal{C}_h} \left(\sum_{e \in \partial T} h_e \|q - \tilde{q}\|_{0,e}^2 \right) &\leq C \sum_{T \in \mathcal{C}_h} h_T^2 |q|_{1,T}^2 \\ &\leq Ch^2 \sum_{T \in \mathcal{C}_h} |q|_{1,T}^2 \\ &\leq Ch^2 |q|_1^2 \end{aligned}$$

$$\begin{aligned} \Rightarrow \sum_{e \in \Gamma_h} h_e \int_e |q - \tilde{q}|^2 &= \sum_{e \in \Gamma_h} h_e \|q - \tilde{q}\|_{0,e}^2 \\ &\leq \sum_{T \in \mathcal{C}_h} \sum_{e \in \partial T} h_e \|q - \tilde{q}\|_{0,e}^2 \\ &\leq Ch^2 |q|_1^2. \end{aligned}$$

□

A.1.4 Abschätzung für die $(1, h)$ -Norm

Im Beweis von Lemma (3.4.2) brauchen wir die Abschätzung:

$$\| u \|_{1,h}^2 \leq C \sum_{T \in \mathcal{C}_h} h_T^{-2} \| u \|_{0,T}^2$$

Beweis

Nach Definition ist

$$\| u \|_{1,h}^2 = \sum_{T \in \mathcal{C}_h} \| \nabla u \|_{0,T}^2 + \sum_{e \in \Gamma_h} h_e^{-1} \int_e |[u]_e|^2$$

Wir schätzen zuerst die ersten Summanden ab:

Dreiecksweise gilt wegen der inversen Abschätzung (A.1.2):

$$\exists C > 0 : \forall u : \| \nabla u \|_{0,T} \leq C h_T^{-1} \| u \|_{0,T}$$

und daraus erhalten wir:

$$\sum_{T \in \mathcal{C}_h} \| \nabla u \|_{0,T}^2 \leq C \sum_{T \in \mathcal{C}_h} h_T^{-2} \| u \|_{0,T}^2 .$$

Jetzt zur Abschätzung des 2. Summanden: Sei $e \in \Gamma_h$ die innere Kante zwischen den Dreiecken T_ℓ und T_r aus \mathcal{C}_h . Wegen $(a-b)^2 \leq 2a^2 + 2b^2$ folgt für diese Kante e :

$$\begin{aligned} \int_e |[u]_e|^2 &\leq 2(\int_e |u_\ell|^2 + \int_e |u_r|^2) \\ &\leq C(h^{-1} \| u \|_{0,T_\ell}^2 + h^{-1} \| u \|_{0,T_r}^2) \end{aligned}$$

\implies

$$\sum_{e \in \Gamma_h} h_e^{-1} \int_e |[u]_e|^2 \leq C \sum_{T \in \mathcal{C}_h} h^{-2} \| u \|_{0,T}^2 .$$

□

A.1.5 Satz von Gauß

Für $\tau : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ und $u : \Omega \rightarrow \mathbb{R}^n$ definieren wir :

$$(\nabla \cdot \tau)_i := \sum_{j=1}^n \partial_j \tau_{ij}$$

$$\nabla u := (\partial_j u_i)_{1 \leq i, j \leq n}$$

Daraus folgt mit dem Satz von Gauß:

$$\boxed{\int_{\Omega} \tau \cdot \nabla u = - \int_{\Omega} u \cdot (\nabla \cdot \tau) + \int_{\partial\Omega} u \cdot (\tau \cdot n)}$$

Bemerkung:

$$u \cdot (\tau \cdot n) = (u \cdot n^t) \tau$$

Beweis

$$\begin{aligned} \int_{\Omega} \tau \cdot \nabla u &= \sum_{i,j=1}^n \int_{\Omega} \tau_{ij} \partial_j u_i \\ &= \sum_{i,j=1}^n (- \int_{\Omega} u_i \partial_j \tau_{ij} + \int_{\partial\Omega} u_i n_j \tau_{ij}) \\ &= \sum_{i=1}^n (- \int_{\Omega} u_i (\nabla \cdot \tau)_i + \int_{\partial\Omega} u_i (\tau \cdot n)_i) \\ &= - \int_{\Omega} u \cdot (\nabla \cdot \tau) + \int_{\partial\Omega} u \cdot (\tau \cdot n) \end{aligned}$$

Dabei gilt die zweite Gleichung wegen folgender Anwendung des Satzes von Gauß: Sei $v : \Omega \rightarrow \mathbb{R}^n$ mit $v = [0 \dots 0 v_k 0 \dots 0]$. Der Satz von Gauß

$$\int_{\Omega} \nabla \cdot v = \int_{\partial\Omega} v \cdot n$$

angewandt auf dieses spezielle v liefert nun:

$$\int_{\Omega} \partial_k v_k = \int_{\partial\Omega} v_k \cdot n_k.$$

Sei nun $v_k = \phi \cdot \psi$ mit $\phi, \psi : \Omega \rightarrow \mathbb{R}$ zwei stetig differenzierbaren Funktionen auf Ω . Mit der Kettenregel für Produkte erhalten wir:

$$\int_{\Omega} \partial_k (\phi \cdot \psi) = \int_{\partial\Omega} (\phi \cdot \psi) \cdot n_k$$

\implies

$$\int_{\Omega} (\partial_k \phi) \psi = - \int_{\Omega} \phi \cdot \partial_k \psi + \int_{\partial\Omega} \phi \cdot n_k \cdot \psi.$$

A.1.6 Lemma: Äquivalenz zu $H(\nabla \cdot, (\Omega))$

Seien Ω_1, Ω_2 offen in \mathbb{R}^2 . Sei

$$\Omega := (\overline{\Omega_1} \cup \overline{\Omega_2})^\circ$$

Für $v \in L^2(\Omega)$ mit $v|_{\Omega_1} \in C^1(\Omega_1)$ und $v|_{\Omega_2} \in C^1(\Omega_2)$ sind äquivalent:

$v \in H(\nabla \cdot, (\Omega))$ und $v \cdot n$ ist stetig entlang innerer Kanten.

Beweis

Sei

$$\gamma := \partial\Omega_1 \cap \partial\Omega_2$$

$$v \in H(\nabla \cdot, (\Omega)) : \iff v \in L^2(\Omega) \text{ und } \nabla \cdot v \in L^2(\Omega)$$

Sei jetzt $\phi_i := \nabla \cdot (v|_{\Omega_i})$ und $\phi := \phi_i$ auf Ω_i . Das macht Sinn, da $v|_{\Omega_i} \in C^1(\Omega_i)$.

Falls jetzt ϕ in L^2 ist, so ist $v \in H(\nabla \cdot, (\Omega))$. Wann ist aber $\phi = \nabla \cdot v$ im Distributionssinn?

$\phi = \nabla \cdot v$ im Distributionssinn

$$\iff \int_{\Omega} \phi \psi = - \int_{\Omega} v \nabla \psi \quad \forall \psi \in C_0^\infty(\Omega)$$

Für ψ sind jetzt zwei Fälle möglich:

a) $\text{supp } \psi$ ist kompakte Teilmenge von Ω_i :

$$\begin{aligned} \int_{\Omega} \phi \psi &= \int_{\Omega_i} \phi \psi \\ &= \int_{\Omega_i} \nabla \cdot (v|_{\Omega_i}) \psi \\ &= - \int_{\Omega_i} v|_{\Omega_i} \nabla \psi \quad \psi = 0 \text{ auf } \partial\Omega_i \\ &= - \int_{\Omega_i} v \nabla \psi \end{aligned}$$

b) sei $\gamma \cap \text{supp } \psi \neq \{\}$.

$$\begin{aligned} \int_{\Omega} \phi \psi &= \int_{\Omega_1} \phi \psi + \int_{\Omega_2} \phi \psi \\ &= - \int_{\Omega_1} v \nabla \psi - \int_{\Omega_2} v \nabla \psi \\ &\quad + \int_{\partial\Omega_1} n_{\Omega_1} v \psi - \int_{\partial\Omega_2} n_{\Omega_2} v \psi \\ &= - \int_{\Omega} v \nabla \psi \pm \int_{\gamma \cap \text{supp } (\psi)} [n \cdot v] \psi \end{aligned}$$

Dabei bezeichnet $[\cdot]$ den Sprung über γ , also die Differenz zwischen links- und rechtsseitigem Grenzwert der Funktion. ϕ ist folglich Divergenz von v genau wenn

$$\forall \psi \in C_0^\infty(\Omega) : \text{supp } (\psi) \cap \gamma \neq \{\} : \int_{\gamma \cap \text{supp } (\psi)} [n \cdot v] \psi = 0.$$

Wir definieren nun eine Folge ψ_h so, daß

$$\int_{\gamma \cap \text{supp}(\psi)} [n \cdot v] \psi_h \xrightarrow{k \rightarrow \infty} [n \cdot v](x)$$

für h gegen 0. Genauer sei der Träger von ψ_h ein Ball mit Radius h um $x \in \gamma$. Somit muß der Sprung $[n \cdot v] = 0$ sein, damit ϕ distributionelle Divergenz von v ist. Der Sprung ist aber genau dann $= 0$ wenn $n \cdot v$ stetig ist auf γ .

□

A.1.7 Lemma von Euler (NT, NE, NV)

Für die Triangulierung bezeichnen:

NT =Anzahl der Dreiecke

NE =Anzahl aller Kanten

NE_{Γ} =Anzahl der Randkanten

$NE - NE_{\Gamma}$ =Anzahl innere Kanten

NV =Anzahl aller Dreieckseckpunkte

NV_{Γ} =Anzahl der Dreieckspunkte auf dem Rand von Ω

$NV - NV_{\Gamma}$ =Anzahl der inneren Punkte

k =Anzahl der Ränder von Ω

Es gilt jetzt

$$NV = 2NT - NE + NV_{\Gamma} - k + 2$$

$$NT = 2(NV - NV_{\Gamma}) + NE_{\Gamma} + 2(k - 2)$$

Daraus erhalten wir die Näherungen

$$NT \sim 2NV$$

$$NE \sim 3NV$$

$$NE \sim \frac{3}{2}NT$$

Beweis

Wir können jede Triangulierung mit den fünf folgenden Konstruktionsschritten erzeugen:

- Start: Mit Start bezeichnen wir den Fall, daß die Triangulierung nur aus genau einem Dreieck besteht.
- a) Wird in einer Triangulierung ein Dreieck von einem inneren Punkt aus unterteilt in drei Dreiecke, so kommen zur Triangulierung: +2 Dreiecke; +1 innerer Punkt; +3 innere Kanten.
- b) Wird in einer Triangulierung ein Dreieck an einer Randkante angefügt, so kommen zur Triangulierung: +1 innere Kante; +1 Randkante; +1 Dreieck.
- c) Wird bei einer Triangulierung mit einer einspringenden Ecke, diese Ecke eliminiert, indem außen ein Dreieck angehängt wird, so kommen zur Triangulierung: +2 innere Kanten; -1 Randkante; +1 innerer Punkt; +1 Dreieck.
- d) wird bei einer Triangulierung ein inneres Dreieck entfernt, so kommen zur Triangulierung: -3 innere Kanten; +3 Randkanten; -3 innere Punkte; -1 Dreieck; +1 Rand.

Das ganze in Tabellenform:

Fall	$NE - NE_{\Gamma}$	NE_{Γ}	$NV - NV_{\Gamma}$	NT	k
Start	0	3	0	1	1
a)	+3		+1	+2	
b)	+1	+1		+1	
c)	+2	-1	+1	+1	
d)	-3	+3	-3	-1	+1

Daraus ergibt sich das Gleichungssystem:

$$NE - NE_{\Gamma} = 3a + b + 2c - 3d$$

$$NE_{\Gamma} = 3 + b - c + 3d$$

$$NV - NV_{\Gamma} = a + c - 3d$$

$$NT = 1 + 2a + b + c - d$$

$$k = 1 + d$$

Daraus folgt:

$$NE = NV - NV_\Gamma + NT + k + NE_\Gamma - 2$$

Weil stets

$$3NT + NE_\Gamma = 2NE$$

gilt, so folgt die Behauptung.

□

A.2 Notationen

$(\cdot, \cdot) :=$ Skalarprodukt in $L^2(\Omega)$

$\langle \cdot, \cdot \rangle :=$ Skalarprodukt auf $L^2(\partial\Omega)$

$[u]_e :=$ Sprung von u über die Kante e .

A.2.1 Räume

$$L^2(\Omega) := \{\phi : \Omega \rightarrow \mathbb{R} : \int_\Omega \phi^2 \text{ existiert.}\}$$

$$H^k(\Omega) := \{\phi \in L^2(\Omega) : \text{alle Distributionsableitungen} \\ \text{bis zum Grad } k \text{ existieren und sind in } L^2\}$$

$$H_0^k(\Omega) := \{\phi \in H^k(\Omega) : \frac{\partial^\ell \phi}{\partial n^\ell} = 0 \text{ auf } \partial\Omega : 0 \leq \ell \leq k-1\} \\ = \text{Vervollständigung von } C_0^\infty(\Omega) \text{ bezüglich } \|\cdot\|_k$$

$$H^{-1}(\Omega) := \mathcal{L}(H_0^1(\Omega), \mathbb{R})$$

$$M_h^{k,-1} := \{\phi : \Omega \rightarrow \mathbb{R} : \forall T \in \mathcal{C}_h : \phi|_T = \hat{\phi} \circ F_T^{-1} : \hat{\phi} \in \hat{P}_k\}$$

$$M_h^{k,r} := M_h^{k,-1} \cap C^r(\Omega)$$

$$\hat{P}_k := \text{span}\{\prod_{\mu=1}^d x_\mu^{j_\mu} : j_\mu \in \mathbf{N} : j_\mu \geq 0 : \sum_{\mu=1}^d j_\mu \leq k\}$$

$$L_0^2(\Omega) := \{\phi \in L^2(\Omega) : \int_\Omega \phi = 0\}$$

$$H(\nabla \cdot, (\Omega)) := \{\tau \in L^2(\Omega)^{2 \times 2} : \nabla \cdot \tau \in L^2(\Omega)\}$$

$$P_k(T) := \{\phi : T \rightarrow \mathbb{R} : \phi \text{ ist Polynom vom Grad } \leq k\}$$

$$H_h := \{\tau \in H(\nabla \cdot, (\Omega)) : \tau|_T \in P_k(T), T \in \mathcal{C}_h\}$$

$$V_h := \{v \in L^2(\Omega) : v|_T \in P_{k-1}(T) : T \in \mathcal{C}_h\}$$

$$V_h^* := \{v \in L^2(\Omega) : v|_T \in P_{k+1}(T) : T \in \mathcal{C}_h\}$$

$$P_h := \{p \in C(\Omega) \cup L_0^2(\Omega) : p|_T \in P_k(T) : T \in \mathcal{C}_h\}$$

A.2.2 Normen und Halbnormen

$\| \cdot \|$:= Norm auf \mathbb{R}^n

$\| \cdot \|$:= zugehörige Matrixnorm

$\| \phi \|_0 := (\int_{\Omega} \phi^2)^{\frac{1}{2}} \quad \phi \in L^2(\Omega)$

$\| \phi \|_k := (\sum_{|\alpha| \leq k} \| D^{\alpha} \phi \|_0^2)^{1/2} \quad \phi \in H^k(\Omega)$

$|\phi|_k := (\sum_{|\alpha|=k} \| D^{\alpha} \phi \|_0^2)^{1/2} \quad \phi \in H^k(\Omega)$

$\| \phi \|_{-1} := \| \phi \|_{\mathcal{L}(H_0^1(\Omega), \mathbb{R})}$

$$= \sup_{\substack{\psi \in H_0^1(\Omega) \\ \psi \neq 0}} \frac{|\phi(\psi)|}{\| \psi \|_1}$$

$\| \sigma \|_{0,h} := (\| \sigma \|_0^2 + \sum_{e \in \Gamma_h} h_e \int_e |\sigma \cdot n|^2)^{1/2}$

$\| u \|_{1,h} := (\sum_{T \in \mathcal{C}_h} \| \nabla u \|_{0,T}^2 + \sum_{e \in \Gamma_h} h_e^{-1} \int_T |[u]|^2)^{1/2}$

$\| \sigma, u \|_h := (\| \sigma \|_{0,h}^2 + \| u \|_{1,h}^2)^{1/2}$

Literatur

- [1] Babuška, I., Osborn, J.E., Pitkäranta, J. ; Analysis of mixed methods using mesh dependent norms, *Math. Comput.* 35 (1980) (s. 1039-1062)
- [2] Brezzi F.; On the existence, uniqueness and approximation of saddle-point problems arising from Lagrange multipliers, *RAIRO Ser. Rouge* 8 (1974) (s. 129-151)
- [3] Brezzi, F., Douglas, J., Marini, L.D.; Two Families of Mixed Finite Elements for Second Order Elliptic Problems, *Numer. Math.* 47 (1985) (s.217-235)
- [4] Brezzi, F., Douglas, J., Duran, R., Fortin, M.; Mixed Finite Elements for Second Order Elliptic Problems in Three Variables, *Numer. Math.* 51 (1987) (s. 237-250)
- [5] Ciarlet, P.G.; *The Finite Element Method for Elliptic Problems* , NORTH-HOLLAND (1978)
- [6] Deuffhard P., Hohmann A.; *Numerische Mathematik*, de Gruyter (1991)
- [7] Girault, V., Raviart P.A.; *Finite Element Methods for Navier-Stokes Equations*, Springer series in computational mathematics, Springer-Verlag (1986)
- [8] Stenberg R.; Some new Families of Finite Elements for the Stokes Equations, *Numer. Math.* 56 (1990) (s.827-838)
- [9] Verfürth, R.; *FEMFLOW-User Guide*, Institut für Angewandte Mathematik Universität Zürich, Rämistr. 74 (1989)
- [10] Fraijs de Veubeke, B.X.: Displacement and equilibrium models in the finite element method. In: *Stress analysis* (O.C. Zienkiewicz, G.Holister, eds.) New York: John Wiley (1965)
- [11] Fraijs de Veubeke, B.X.: Stress function approach. *World Congress on the Finite Element Method in Structural Mechanics*. Bournemouth (1965)