

## 6 FE-Methoden für inkompressible Strömungen

In diesem Kapitel diskutieren wir die Anwendung der Methode der finiten Elemente zur numerischen Lösung der inkompressiblen Navier-Stokes-Gleichungen. Für weitere Details sei auf die Übersichtsartikel Glowinski [111], Becker et al. [99] und [131, 132, 133, 134, 135, 136], verwiesen.

### 6.1 FEM für die Stromfunktionsformulierung

In diesem Abschnitt diskutieren wir die Anwendung der Methode der finiten Elemente zur numerischen Lösung der stationären Stromfunktionsformulierung der inkompressiblen Navier-Stokes-Gleichungen. Die Diskretisierung dieses Randwertproblems vierter Ordnung folgt denselben Prinzipien wie die bei der schon betrachteten verwandten „Plattengleichung“. Wir betrachten zunächst wieder die klassischen „primalen“ Approximationen, als deren Verallgemeinerungen die „nicht-konformen“ Ansätze und dann durch Übergang zu einer anderen variationellen Formulierung auch „gemischte“ Verfahren. Um den exemplarischen Charakter dieser Diskussion zu wahren, werden wir uns auf die einfachste Randwertaufgabe der Stromfunktionsgleichung beschränken, d. h. werden nur homogene Dirichlet-Randbedingungen betrachten und uns meist sogar mit dem linearen Fall begnügen. Dasselbe gilt für die betrachteten Diskretisierungen. Im Interesse einer leichten Verständlichkeit werden jeweils wieder nur die einfachsten Vertreter der verschiedenen Verfahrensklassen beschrieben, auch wenn diese vielleicht nicht die „besten“ Verfahren für die allgemeine Praxis sind. Beweise werden nur gegeben, wenn diese von der entsprechenden Argumentation bei der Plattengleichung abweichen.

Wir formulieren nochmals die (nichtlineare) Stromfunktionsgleichung in variationeller Form mit homogenen Dirichlet-Randbedingungen:

(P) *Gesucht is ein  $\Psi \in H_0^2(\Omega)$  mit der Eigenschaft*

$$\nu(\Delta\Psi, \Delta\varphi) - (\partial_1\Psi\Delta\Psi, \partial_2\varphi) - (\partial_2\Delta\Psi, \partial_1\varphi) = (f, \text{rot}\varphi) \quad \forall\varphi \in H_0^2(\Omega). \quad (6.1.1)$$

Zur Abkürzung wird im Folgenden  $\nu = 1$  gesetzt und meist nur die lineare Gleichung betrachtet

$$(\Delta\Psi, \Delta\varphi) = (f, \text{rot}\varphi) \quad \forall\varphi \in H_0^2(\Omega). \quad (6.1.2)$$

Wenn in der ursprünglichen Navier-Stokes-Gleichungen die Volumenkraftdichte die „minimale“ Regularität  $f \in L^2(\Omega)$  besitzt, definiert  $F(\varphi) := (f, \text{rot}\varphi)$  offenbar ein stetiges lineares Funktional auf  $H_0^1(\Omega)$ , d. h.:  $F(\cdot) \in H^{-1}(\Omega)$ . Der Stromfunktionsoperator  $\Delta^2$  (Biharmonischer Operator) sollte Funktionale in  $H^{-1}(\Omega)$  in Lösungen in  $H^3(\Omega)$ , dem Sobolew-Raum der  $L^2$ -Funktionen mit verallgemeinerten Ableitungen bis zur Ordnung drei in  $L^2(\Omega)$ , abbildet. Auf Gebieten mit den bereits oben beschriebenen Eigenschaften,  $\Omega$  konvex oder  $\partial\Omega$  glatt, ist dies tatsächlich der Fall. Der Operator

$$\Delta^2 : H_0^2(\Omega) \cap H^3(\Omega) \rightarrow H^{-1}(\Omega)$$

ist dann ein Isomorphismus, und es gilt insbesondere die a priori Abschätzung

$$c\|\nabla^2\Psi\| \leq \|F\|_{-1} := \sup_{\varphi \in H_0^1(\Omega)} \frac{|(f, \operatorname{rot}\varphi)|}{\|\nabla\varphi\|} \leq c'\|f\|. \quad (6.1.3)$$

Diese Regularitätsstufe  $\Psi \in H_0^2(\Omega) \cap H^3(\Omega)$  erscheint also als generisch für die Stromfunktion und wird im Folgenden immer als gegeben angenommen. Zur einfacheren Konstruktion der FE-Approximation wird weiter das Gebiet  $\Omega$  als ein konvexes Polygonebiet angenommen. Die Lösung  $\Psi \in H_0^2(\Omega) \cap H^3(\Omega)$  erfüllt dann auch die Beziehung

$$-(\nabla\Delta\Psi, \nabla\varphi) = (f, \operatorname{rot}\varphi) \quad \forall \varphi \in H_0^1(\Omega),$$

was man durch partielle Integration in (6.1.2) und Ausnutzung der Dichtheit von  $H_0^2(\Omega)$  in  $H_0^1(\Omega)$  erschließt.

### 6.1.1 Konforme primale Ansätze

Der Vollständigkeit halber werden im Folgenden die wesentlichen Begriffe und Tatsachen aus dem Kapitel über die Numerik der Plattengleichung wiederholt. Wir betrachten zunächst das lineare Problem (6.1.2). Seien  $\mathbb{T}_h = \{K\}$  wieder Zerlegungen von  $\bar{\Omega}$  in abgeschlossene Zellen  $K$  (Dreiecke oder Vierecke) der maximalen Weite  $h := \max_{K \in \mathbb{T}_h} \operatorname{diam}(K)$ . Diese Zerlegungen sollen die folgenden, üblichen Regularitätseigenschaften haben, deren Formulierungen auf Dreiecks- und Viereckszerlegungen zugeschnitten sind:

1. *Strukturregularität:* Je zwei Zellen haben höchstens eine ganze Kante oder einen Eckpunkt gemeinsam.
2. *Formregularität:* Es bezeichnen  $h_K$  den Durchmesser und  $\rho_K$  den minimale Abstand eines Eckpunkts der Zelle  $K$  zu der gegenüberliegenden Seite. Es gibt eine Konstante  $c_1 > 0$ , so dass

$$\sup_{h>0} \max_{K \in \mathbb{T}_h} \frac{h_K}{\rho_K} \leq c_1.$$

3. *Größenregularität:* Es gibt eine  $h$ -unabhängige Konstante  $c_2 > 0$ , so dass

$$\max_{K \in \mathbb{T}_h} h_K \leq c_2 \min_{K \in \mathbb{T}_h} h_K.$$

Wir nehmen im Folgenden stets an, dass die verwendeten Gitter  $\mathbb{T}_h$  struktur- und formregulär sind. Größenregularität ist für die meisten Resultate nicht notwendig, was z. B. lokale Gitterverfeinerung im Rahmen adaptiver Techniken erlaubt. Für spätere Zwecke definieren wir noch die Menge  $\partial\mathbb{T}_h = \{\Gamma\}$  aller Kanten von Zellen in  $\mathbb{T}_h$ .

Auf den Gittern  $\mathbb{T}_h$  werden Finite-Elemente-Ansatzräume definiert:

$$V_h := \{\varphi_h \in L^2(\Omega) \mid \varphi_h|_K \in P(K), K \in \mathbb{T}_h, \varphi_h \text{ stetig und } \varphi_h = 0 \text{ auf } \partial\Omega \\ \text{bzgl. gewisser Knotenfunktionale } \chi(\cdot)\}.$$

Dabei sind  $P(K)$  gewisse Vektorräume von Polynomen auf der Zelle  $K$ . Meist ist  $P(K) = P_k(K)$  ein voller Polynomraum (Polynome vom Grad  $\leq k$ ), oder  $P(K) = P_k(K) \oplus \text{span}(p_1, \dots, p_r)$  mit gewissen Polynomen  $p_i$  höherer als  $k$ -ter Ordnung. In diesen Räumen werden Approximationen  $\Psi_h \in V_h$  zu  $\Psi \in V := H_0^2(\Omega)$  gesucht. Die konkreten FE-Ansätze auf den einzelnen Zellen  $K \in \mathbb{T}_h$  ergeben sich dabei mit Hilfe einer linearen bzw. bilinearen Transformation  $\varphi_K : \hat{K} \rightarrow K$  von einer „Referenzzelle“  $\hat{K}$  (z. B. Einheitsdreieck oder Einheitsviereck im  $\mathbb{R}^2$ ) auf  $K$  (sog. „isoparametrischer“ Ansatz):

$$v_{h|K}(x) := \hat{v}_h(\varphi_K^{-1}(x)), \quad x \in K.$$

**Definition 6.1:** Der Finite-Elemente-Ansatzraum  $V_h$  wird „konform (bzgl.  $V = H_0^2(\Omega)$ )“ genannt, wenn  $V_h \subset V$  ist, andernfalls „nicht-konform“.

Konformität bzgl.  $H_0^2(\Omega)$  bedingt bei einer stückweise polynomialen Ansatzfunktion  $\Psi_h$  die Stetigkeit von  $\Psi_h$  und  $\nabla\Psi_h$  über die Zellgrenzen hinweg sowie das Verschwinden von  $\Psi_h$  und  $\nabla\Psi_h$  auf dem Polygonzug  $\partial\Omega$ . Gängige Beispiele  $H_0^2$ -konformer Ansätze sind im Kapitel zur Numerik der Plattengleichung angegeben und werden deswegen nicht noch einmal wiederholt.

Für einen *konformen* Ansatz lautet die Gleichung zur Bestimmung von  $\Psi_h \in V_h$ :

$$(\Delta\Psi_h, \Delta\varphi_h) = (f, \text{rot}\varphi_h) \quad \forall \varphi_h \in V_h. \quad (6.1.4)$$

Die eindeutige Lösbarkeit folgt mit demselben Argument wie für das zugehörige kontinuierliche Problem (Darstellungssatz von Riesz). Analog zur FE-Approximation der Poisson-Gleichung gilt wieder die „Galerkin-Orthogonalität“ für den Fehler  $e_h := \Psi - \Psi_h$ :

$$(\Delta e_h, \Delta\varphi_h) = 0, \quad \varphi_h \in V_h. \quad (6.1.5)$$

Die Zuordnung  $\Psi_h := R_h\Psi$  ist daher eine orthogonale Projektion bzgl. des Skalarprodukts  $(\Delta\cdot, \Delta\cdot)$ . Daher gilt wieder die Bestapproximationsbeziehung

$$\|\Delta e_h\| = \min_{\varphi_h \in V_h} \|\Delta(\Psi - \varphi_h)\|, \quad (6.1.6)$$

womit die Frage nach der Konvergenz der Approximation,  $\Psi_h \rightarrow \Psi$  ( $h \rightarrow 0$ ) zurückgeführt ist auf ein lokales Approximationsproblem. Für die üblichen FE-Räume  $V_h$  gibt es einen zellweise definierten „Interpolationsoperator“  $I_h : V \cap H^3(\Omega) \rightarrow V_h$ , so dass

$$\|\nabla^2(\Psi - I_h\Psi)\|_K \leq c_i h_K \|\nabla^3\Psi\|_K, \quad K \in \mathbb{T}_h. \quad (6.1.7)$$

Mit Hilfe dieser Abschätzung erschließt man direkt die erste Fehlerabschätzung für die FE-Approximation:

$$\|\nabla^2(\Psi - \Psi_h)\| \leq ch\|f\|. \quad (6.1.8)$$

Die *konformen* FE-Ansätze haben Systemmatrizen

$$A = ((\Delta\varphi_h^{(j)}, \Delta\varphi_h^{(i)}))_{i,j=1}^N, \quad N := \dim V_h,$$

mit einer großen Bandbreite wegen der Kopplung der vielen lokalen Freiheitsgrade. Ferner bedingt die Verwendung höherer Ableitungen als Knotenwerte unterschiedliche Skalierung (mit  $h$ -Potenzen) der Matrixelemente. Die Verwendung niedriger Polynomgrade wäre daher wünschenswert. Die Konditionierung der Systemmatrizen ist in Abhängigkeit von der Ordnung des zugrunde liegenden Differentialoperators  $\Delta^2$  für alle Ansätze sehr schlecht, nämlich

$$\text{cond}_2(A) = \mathcal{O}(h^{-4}).$$

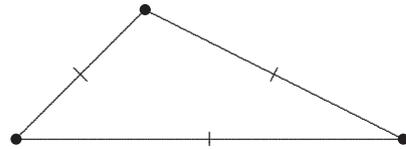
### 6.1.2 Nicht-konforme primale Ansätze

Die Schwierigkeit bei der Konstruktion *konformer* Finite-Elemente-Ansätze legt es nahe, die Stetigkeitsforderungen abzuschwächen, um mit niedrigeren Polynomgraden auszukommen. Wir betrachten wieder als einfachstes Beispiel das quadratische „Morleysche Dreieckselement“ mit den Ansatzräumen

$$V_h := \{\varphi \in L^2(\Omega) \mid \varphi|_K \in P_2(K), K \in \mathbb{T}_h, \chi_a(\varphi), \chi_b(\varphi) \text{ stetig und gleich null für Knotenpunkte auf } \partial\Omega\},$$

mit den Knotenfunktionalen ( $a$  Eckpunkte,  $b$  Seitenmitten):

$$\begin{aligned} P(K) &= P_2(K), \quad \dim P_2(K) = 6, \\ \chi_\Omega(\varphi) &\in \{\varphi(a), \partial_n \varphi(b)\}, \\ \chi_{\partial\Omega}(\varphi) &\in \{\varphi(a), \partial_n \varphi(b)\}. \end{aligned}$$



Alternativ kann man als Knotenfunkionale für das Morley-Element statt der Funktionswerte  $\partial_n \varphi(b)$  auch die Mittelwerte

$$\chi_\Gamma(\varphi) := |\Gamma|^{-1} \int_\Gamma \partial_n \varphi \, ds$$

über Zellkanten  $\Gamma$  verwenden. Diese Wahl hat den Vorteil, dass die zugehörige natürliche Knoteninterpolation dann für alle Funktionen aus  $H_0^2(\Omega)$  definiert ist. Weitere Beispiele *nichtkonformer* finiter Elemente finden sich im Kapitel über die Numerik der Plattengleichung.

Wir wollen nun Modifikationen der variationellen Formulierung (6.1.2) der Stromfunktionsgleichung für nichtkonforme Ansätze diskutieren. Dabei wird sich erweisen, dass ein zu „naiver“ Diskretisierungsansatz leicht in die Irre führen kann. Wegen der Nichtkonformität des Ansatzes ist die Bilinearform  $(\Delta \cdot, \Delta \cdot)$  nicht für Funktionen aus  $V_h$  definiert. Es ist naheliegend, stattdessen die „stückweise“ definierte Form

$$(\Delta \varphi, \Delta \psi)_h := \sum_{K \in \mathbb{T}_h} (\Delta \varphi, \Delta \psi)_K$$

zu verwenden, wobei der der Index  $K$  wieder Integration über die Zelle  $K$  bedeutet., Analog wird die rechte Seite modifiziert zu

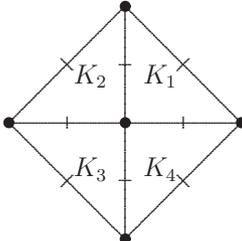
$$(f, \text{rot}(\psi))_h := \sum_{K \in \mathbb{T}_h} (f, \text{rot}(\psi))_K.$$

Als diskretes Problem ist dann ein  $\varphi_h \in V_h$  zu bestimmen mit

$$(\Delta\varphi_h, \Delta\psi_h)_h = (f, \text{rot}(\psi_h))_h \quad \forall \psi_h \in V_h. \quad (6.1.9)$$

Dieses Problem ist aber nicht wohl gestellt, da die Bilinearform  $(\Delta\cdot, \Delta\cdot)_h$  nicht notwendig definit ist. Dies wird anhand des folgenden, einfachen Beispiels belegt.

**Beispiel 6.1:** Auf einem Gitter bestehend aus vier Zellen betrachten wir die Funktion

$$\varphi(x, y) = \begin{cases} x(1-x) - y(1-y) & \text{auf } K_1, \\ x(1+x) + y(1-y) & \text{auf } K_2, \\ -x(1+x) + y(1+y), & \text{auf } K_3, \\ -x(1-x) - y(1+y), & \text{auf } K_4. \end{cases}$$


Diese Funktion hat die Eigenschaften  $\varphi(a_i) = 0$  ( $i = 1, \dots, 5$ ), und  $\partial_n \varphi(b_j) = 0$  in Seitenmitten auf dem Rand. Ferner ist  $\partial_n \varphi(b_j)$  stetig in inneren seitenmitten. Es gilt jedoch  $\|\Delta\varphi\|_h = 0$ .

Wir diskutieren jetzt die Morley-Approximation in etwas mehr Detail. Der zugehörige Ansatzraum sei mit  $V_h^M$  bezeichnet. Aufgrund der vorausgehenden Überlegung muss das variationelle Problem zur Verwendung des Morley-Elements angepasst werden. Eine Stabilisierung der Bilinearform  $(\Delta\cdot, \Delta\cdot)$  ist erforderlich. In Anlehnung an die Situation bei der biharmonischen Gleichung in der Plattenstatik (Kirchhoffsches Plattenmodell) führen wir den folgenden Ausdruck ein:

$$A[\nabla^2\varphi, \nabla^2\psi] := \Delta\varphi\Delta\psi + \sigma\{2\partial_1\partial_2\varphi\partial_1\partial_2\psi - \partial_1^2\varphi\partial_2^2\psi - \partial_2^2\varphi\partial_1^2\psi\}$$

mit einer Konstanter  $\sigma \in (0, 1]$ . Offenbar gilt  $A[\nabla^2\varphi, \nabla^2\psi] = A[\nabla^2\psi, \nabla^2\varphi]$  und wegen

$$A[\nabla^2\varphi, \nabla^2\varphi] = (\partial_1^2\varphi)^2 + (\partial_2^2\varphi)^2 + 2(1-\sigma)\partial_1^2\varphi\partial_2^2\varphi + 2\sigma(\partial_1\partial_2\varphi)^2$$

auch

$$\sigma|\nabla^2\varphi|^2 \leq A[\nabla^2\varphi, \nabla^2\varphi] \leq 2|\nabla^2\varphi|^2.$$

Es wird dann gesetzt:

$$A_h(\varphi, \psi) := \sum_{K \in \mathbb{T}_h} \int_K A[\nabla^2\varphi, \nabla^2\psi] dx.$$

Die modifizierte Bilinearform  $A_h(\cdot, \cdot)$  ist dann für  $0 < \sigma \leq 1$  bzgl. des diskreten Skalarprodukts

$$(\nabla^2\varphi, \nabla^2\psi)_h := \sum_{K \in \mathbb{T}_h} (\nabla^2\varphi, \nabla^2\psi)_K$$

koerzitiv, d. h.:

$$A_h(\varphi, \varphi) \geq \sigma \|\nabla^2 \varphi\|_h^2, \quad \varphi \in V \oplus V_h^M. \quad (6.1.10)$$

Dass hierdurch tatsächlich eine *definite* symmetrische Bilinearform und damit ein Skalarprodukt definiert ist, erschließt man leicht wie folgt: Ist  $\|\nabla^2 \varphi_h\|_h = 0$ , so folgt  $\nabla^2 \varphi_h|_K \equiv 0$  auf jeder Zelle  $K \in \mathbb{T}_h$ . Also ist  $\varphi_h$  linear. Wegen der geforderten Stetigkeit in den Eckpunkten ist dann  $\varphi_h \in H_0^1(\Omega)$ . In allen Zellen am Rand ist dann notwendig  $\varphi_h|_K \equiv 0$  wegen der Randbedingung  $\partial_n \varphi_h(b) = 0$ . Dies überträgt sich dann induktiv auf alle Zellen, so dass sich  $\varphi_h \equiv 0$  auf ganz  $\Omega$  ergibt. Folglich definiert  $\|\nabla^2 \cdot\|_h$  eine Norm auf dem „Morley-Raum“  $V_h^M$ .

Das zugehörige diskrete Problem lautet dann: *Finde  $\Psi_h \in V_h^M$ , so dass*

$$A_h(\Psi_h, \varphi_h) = (f, \text{rot} \varphi_h)_h \quad \forall \varphi_h \in V_h^M. \quad (6.1.11)$$

Wegen der Koerzitivitätsbeziehung (6.1.10) ist diese Aufgabe stets eindeutig lösbar.

### 6.1.3 Konvergenzanalyse für nichtkonforme Ansätze

Wir wollen einen allgemeinen Konvergenzsatz herleiten, welcher auf verschiedene nichtkonforme Ansätze anwendbar ist. Mit  $I_h \psi \in V_h$  wird die natürliche „Knoteninterpolierende“ von  $\psi \in V \cap H^3(\Omega)$  bezeichnet. Für diese gilt allgemein

$$\|\psi - I_h \psi\| + h \|\nabla(\psi - I_h \psi)\|_h + h^2 \|\nabla^2(\psi - I_h \psi)\|_h \leq ch^3 \|\nabla^3 \psi\|. \quad (6.1.12)$$

Für die Lösung  $\psi \in V \cap H^3(\Omega)$  der Randwertaufgabe

$$(\Delta \psi, \Delta \varphi) = (g, \text{rot} \varphi) \quad \forall \varphi \in V, \quad (6.1.13)$$

definieren wir auf  $V_h \oplus V$  die Form

$$N_h(\psi, \varphi_h) := A_h(\psi, \varphi_h) - (g, \text{rot} \varphi_h), \quad \varphi_h \in V_h \oplus V.$$

Speziell für die Lösung  $\psi = \Psi$  (zur rechten Seite  $g = f$ ) und  $\varphi_h \in V_h$  ist dann

$$N_h(\psi, \varphi) = A(\psi, \varphi) - (g, \text{rot} \varphi) = 0, \quad \varphi \in V. \quad (6.1.14)$$

Wir definieren nun die Größen (Maß für die Approximationsgüte)

$$\tau_h^{(k)} := \sup_{\psi \in V \cap H^k(\Omega)} \inf_{\psi_h \in V_h} \frac{\|\nabla^2(\psi - \psi_h)\|_h}{\|\nabla^k \psi\|}, \quad k \geq 2,$$

sowie (Maß für die Nichtkonformität)

$$\delta_h^{(k)} := \sup_{\psi \in V \cap H^k(\Omega)} \sup_{\varphi \in V \oplus V_h} \frac{|N_h(\psi, \varphi)|}{\|\nabla^k \psi\| \|\nabla^2 \varphi\|_h}, \quad k \geq 2.$$

**Hilfssatz 6.1 (Stromfunktionsapproximation):** Für die Stromfunktion  $\Psi \in V \cap H^3(\Omega)$  sei  $\Psi_h \in V_h^M$  die durch (6.1.11) definierte, nicht-konforme Approximierende. Dann gelten für den Fehler  $e_h := \Psi - \Psi_h$  die Abschätzungen

$$\|\nabla^2 e_h\|_h \leq c\{\tau_h^{(3)} + \delta_h^{(3)}\} \|f\|, \quad (6.1.15)$$

$$\|\nabla e_h\|_h \leq c\{\tau_h^{(3)} + \delta_h^{(3)}\}^2 \|f\|. \quad (6.1.16)$$

**Beweis:** (i) Sei  $\psi_h \in V_h$  eine beliebige Approximierende der Lösung  $\Psi$ . Wir schätzen zunächst die Differenz  $\tilde{e}_h := \psi_h - \Psi_h \in V_h$  ab. Es gilt

$$\begin{aligned} \sigma \|\nabla^2 \tilde{e}_h\|_h^2 &\leq A_h(\tilde{e}_h, \tilde{e}_h) \\ &= A_h(\psi_h - \Psi, \tilde{e}_h) + A_h(e_h, \tilde{e}_h) = A_h(\psi_h - \Psi, \tilde{e}_h) + N_h(\Psi, \tilde{e}_h), \end{aligned}$$

und folglich

$$\|\nabla^2 e_h\|_h^2 \leq \frac{2 + \sigma}{\sigma} \inf_{\psi_h \in V_h} \|\nabla^2(\Psi - \psi_h)\|_h + \frac{1}{\sigma} \sup_{\varphi_h \in V_h} \frac{|N_h(\Psi, \varphi_h)|}{\|\nabla^2 \varphi_h\|_h}. \quad (6.1.17)$$

Diese Abschätzung ist eine Verallgemeinerung der üblichen „Galerkin-Bestapproximation“ für nicht-konforme Ansätze (sog. „2. Strangsche Lemma“). Mit den oben definierten Größen erhalten wir also

$$\|\nabla^2 e_h\|_h \leq c\{\tau_h^{(3)} + \delta_h^{(3)}\} \|\nabla^3 \Psi\|.$$

Dies impliziert die Abschätzung (6.1.15).

(ii) Zum Nachweis von (6.1.16) verwenden wir ein Dualitätsargument. Sei  $z \in V$  Lösung des Hilfsproblems

$$(\Delta \varphi, \Delta z) = (\text{rot} \varphi, \text{rote}_h)_h \quad \forall \varphi \in V.$$

Die rechte Seite definiert offenbar ein stetiges lineares Funktional auf  $H_0^1(\Omega)$ . Folglich ist  $z \in H^3(\Omega)$ , und es gilt die a priori Abschätzung

$$\|\nabla^3 z\| \leq c \|\text{rote}_h\|_h. \quad (6.1.18)$$

Aufgrund der Definition der Form  $N_h(z, e_h)$  gilt dann

$$\|\text{rote}_h\|_h^2 = A_h(e_h, z) - N_h(z, e_h).$$

Mit der Knoteninterpolierenden  $I_h z \in V_h$  folgt weiter

$$\begin{aligned} \|\text{rote}_h\|_h^2 &= A_h(e_h, z - I_h z) + A_h(e_h, I_h z) - N_h(z, e_h) \\ &= A_h(e_h, z - I_h z) + N_h(\Psi, I_h z) - N_h(z, e_h) \\ &= A_h(e_h, z - I_h z) + N_h(\Psi, I_h z - z) - N_h(z, e_h). \end{aligned}$$

Wir können also wie folgt abschätzen:

$$\begin{aligned} \|\text{rote}_h\|_h^2 &\leq 2\|\nabla^2 e_h\|_h \|\nabla^2(z - I_h z)\|_h + \delta_h^{(3)} \|\nabla^3 \Psi\| \|\nabla^2(z - I_h z)\|_h \\ &\quad + \delta_h^{(3)} \|\nabla^3 z\| \|\nabla^2 e_h\|_h \\ &\leq \{2\tau_h^{(3)} \|\nabla^2 e_h\|_h + \delta_h^{(3)} \|\nabla^3 \Psi\| \tau_h^{(3)} + \delta_h^{(3)} \|\nabla^2 e_h\|_h\} \|\nabla^3 z\|. \end{aligned}$$

Mit Hilfe der a priori Abschätzung (6.1.18) erschließen wir weiter

$$\|\operatorname{rote}_h\|_h \leq 2\tau_h^{(3)} \|\nabla^2 e_h\|_h + \delta_h^{(3)} \|\nabla^3 \Psi\|_{\tau_h^{(3)}} + \delta_h^{(3)} \|\nabla^2 e_h\|_h,$$

und unter Berücksichtigung der bereits gezeigten Abschätzung (6.1.15):

$$\|\operatorname{rote}_h\|_h \leq c(\tau_h^{(3)} + \delta_h^{(3)})^2 \|\nabla^3 \Psi\| + \tau_h^{(3)} \delta_h^{(3)} \|\nabla^3 \Psi\|.$$

Hieraus ergibt sich die Abschätzung (6.1.16), was den Beweis vervollständigt. Q.E.D.

Die Anwendung dieses Hilfssatzes auf konkrete nicht-konforme Approximationen erfordert die Abschätzung der Größen  $\tau_h^{(k)}$  und  $\delta_h^{(k)}$ . Dies erfolgt unter Ausnutzung der jeweiligen Struktur der verwendeten Ansatzfunktionen.

**Satz 6.1 (Morley-Approximation):** *Die Lösung der Stromfunktionsgleichung erfülle  $\Psi \in V \cap H^3(\Omega)$ . Dann gilt für den Fehler  $e_h := \Psi - \Psi_h$  der Morley-Diskretisierung die a priori Fehlerabschätzung*

$$\|\nabla e_h\|_h + h \|\nabla^2 e_h\|_h \leq ch^2 \|f\|. \quad (6.1.19)$$

**Beweis:** Die Argumentation ist analog zu der bei dem entsprechenden Resultat für das Plattenproblem. Q.E.D.

Als Folgerung aus Satz 6.1 erhalten wir für das zur Stromfunktion  $\Psi$  gehörende Geschwindigkeitsfeld  $v = \operatorname{rot}\Psi$  sowie die Wirbelstärke  $\omega = -\Delta\Psi$  und die zugehörigen Approximierenden  $v_{h|K} := \operatorname{rot}\Psi_K$  bzw.  $\omega_{h|K} := -\Delta\Psi_{h|K}$  die Abschätzung

$$\|v - v_h\| + h \|\omega - \omega_h\| \leq ch^2 \|f\|. \quad (6.1.20)$$

Zum Abschluß formulieren wir noch ein Satz 6.1 entsprechendes Resultat für das nicht-konforme Adini-Element.

**Satz 6.2 (Adini-Approximation):** *Die Lösung der Stromfunktionsgleichung sei  $\Psi \in V \cap H^3(\Omega)$ . Dann gilt für den Fehler  $e_h := \Psi - \Psi_h$  der Adini-Diskretisierung die a priori Fehlerabschätzung*

$$\|\nabla e_h\| + h \|\nabla^2 e_h\|_h \leq ch^2 \|f\|. \quad (6.1.21)$$

*Wenn alle Innenwinkel von  $\Omega$  kleiner als  $\alpha_* \approx 126^\circ$  sind und  $\operatorname{rot}f \in L^2(\Omega)$  ist, gilt die verbesserte Fehlerabschätzung*

$$\|e_h\| + h \|\nabla e_h\| + h^2 \|\nabla^2 e_h\|_h \leq ch^4 \|\operatorname{rot}f\|. \quad (6.1.22)$$

**Beweis:** Wir haben wieder die Größen  $\tau_h^{(k)}$  und  $\delta_h^{(k)}$  für das Adini-Element abzuschätzen. Dabei verfahren wir ähnlich wie für das Morley-Element. Die Ausführung des Beweises wird wieder als Übungsaufgabe gestellt. Q.E.D.

### 6.1.4 Gemischte Ansätze

Wir haben gesehen, dass die Konstruktion konformer Ansätze für das Stromfunktionsproblem nicht einfach ist und gewisse praktische Nachteile besitzt. Ein alternativer Weg ist wieder die Verwendung sog. „gemischter“ Methoden; dazu wird von einer anderen, „gemischten“ variationellen Formulierung des Problems ausgegangen. Wir schreiben die Differentialgleichung vierter Ordnung wieder als ein System von zwei Gleichungen zweiter Ordnung für die Streamfunktion  $\Psi$  und die Wirbelstärke  $\omega$ :

$$-\Delta\Psi = \omega, \quad -\nu\Delta\omega + \partial_2(\partial_1\Psi\omega) + \partial_1(\partial_2\Psi\omega) = \operatorname{rot}f. \quad (6.1.23)$$

Gesucht ist also nun ein Paar  $\{\omega, \Psi\} \in L^2(\Omega) \times H_0^2(\Omega)$ , so dass

$$(\omega, \psi) + (\Delta\Psi, \psi) = 0 \quad \forall \psi \in L^2(\Omega), \quad (6.1.24)$$

$$\nu(\nabla\omega, \nabla\varphi) - (\partial_1\Psi\omega, \partial_2\varphi) - (\partial_2\Psi\omega, \partial_1\varphi) = (f, \operatorname{rot}\varphi) \quad \forall \varphi \in H_0^1(\Omega). \quad (6.1.25)$$

Dies ist keine für unsere Zwecke brauchbare variationelle Formulierung, da sie nach wie vor  $H_0^2$ -konforme Ansätze erfordert. Um dies zu umgehen, wird wieder in den obigen Gleichungen partiell integriert und wir erhalten die folgende „gemischte“ variationelle Formulierung der Stromfunktionsgleichung: *Finde*  $\{\omega, \Psi\} \in H^1(\Omega) \times H_0^1(\Omega)$ , *so dass*

$$(\omega, \psi) - (\nabla\Psi, \nabla\psi) = 0 \quad \forall \psi \in H^1(\Omega), \quad (6.1.26)$$

$$\nu(\nabla\omega, \nabla\varphi) - (\partial_1\Psi\omega, \partial_2\varphi) - (\partial_2\Psi\omega, \partial_1\varphi) = (f, \operatorname{rot}\varphi) \quad \forall \varphi \in H_0^1(\Omega). \quad (6.1.27)$$

In dieser Formulierung tritt die Randbedingung  $\partial_n\Psi|_{\partial\Omega} = 0$  nicht mehr explizit auf. Sie ist zu einer „natürlichen“ Randbedingung geworden, die von jeder hinreichend glatten Lösung des Problems (6.1.26, 6.1.27) automatisch erfüllt wird. Eine Komplikation stellen die nichtlinearen Terme dar, die für den gewählten Lösungsraum  $H^1(\Omega) \times H_0^1(\Omega)$  gar nicht definiert sind. Sie erfordern etwas mehr Regularität für die Stromfunktionskomponente als im Ansatz  $\Psi \in H_0^1(\Omega)$  verlangt wird. Die Formulierung (6.1.26, 6.1.27) kann aber trotzdem als Basis für eine Galerkin-Diskretisierung dienen. Wir konzentrieren uns im Folgenden wieder auf den linearen Fall und setzen  $\nu = 1$ . Die gemischte variationelle Formulierung erhält dann die Gestalt:

$$(\omega, \psi) - (\nabla\Psi, \nabla\psi) = 0 \quad \forall \psi \in H^1(\Omega), \quad (6.1.28)$$

$$(\nabla\omega, \nabla\varphi) = (f, \operatorname{rot}\varphi) \quad \forall \varphi \in H_0^1(\Omega). \quad (6.1.29)$$

Wir halten fest, dass für jede Lösung  $\Psi \in V \cap H^3(\Omega)$  der „primalen“ Formulierung der Stromfunktionsgleichung mit  $\omega := -\Delta\Psi$  eine Lösung  $\{\omega, \Psi\} \in H^1(\Omega) \times H_0^1(\Omega)$  der „gemischten“ variationellen Formulierung (6.1.28, 6.1.29) gegeben ist.

**Satz 6.3:** *Im linearen Fall hat die gemischte Formulierung (6.1.28, 6.1.29) höchstens eine Lösung  $\{\omega, \Psi\} \in H^1(\Omega) \times H_0^1(\Omega)$ , und für jede solche Lösung gilt die a priori Abschätzung*

$$\|\omega\| + \|\nabla\Psi\| \leq c\|f\|. \quad (6.1.30)$$

**Beweis:** Wir beweisen zunächst die Eindeutigkeitsaussage. Seien  $\{\omega^1, \Psi^1\}$  und  $\{\omega^2, \Psi^2\}$  zwei Lösungen. Für die Differenz  $\{\omega, \Psi\} := \{\omega^1 - \omega^2, \Psi^1 - \Psi^2\}$  gilt dann

$$\begin{aligned}(\omega, \psi) - (\nabla \Psi, \nabla \psi) &= 0 \quad \forall \psi \in H^1(\Omega), \\ \nu(\nabla \omega, \nabla \varphi) &= 0 \quad \forall \varphi \in H_0^1(\Omega).\end{aligned}$$

Wir setzen  $\psi := \omega$ ,  $\varphi := \Psi$  und erhalten

$$\|\omega\|^2 = (\nabla \Psi, \nabla \omega) = 0,$$

bzw.  $\omega = 0$ . Mit  $\psi := \Psi$  folgt dann auch

$$\|\nabla \Psi\|^2 = (\omega, \Psi) = 0, \quad (6.1.31)$$

bzw.  $\Psi = 0$ . Weiter erhält man bei der Setzung  $\Psi := \omega$  in (6.1.28) und  $\varphi := \Psi$  in (6.1.29):

$$\|\omega\|^2 = (\nabla \Psi, \nabla \omega) = (f, \text{rot} \Psi) \leq \|f\| \|\nabla \Psi\|, \quad (6.1.32)$$

und bei Setzung von  $\psi := \Psi$  in (6.1.28):

$$\|\Psi\|^2 = (\omega, \Psi) \leq c \|\omega\| \|\nabla \Psi\|. \quad (6.1.33)$$

Kombination dieser Abschätzungen ergibt die a priori Schranke (6.1.30). Q.E.D.

**Bemerkung 6.1:** Da man in (6.1.29) nicht mit  $\varphi := \omega \notin H_0^1(\Omega)$  testen darf, gibt die Formulierung (6.1.28, 6.1.29) direkt keine Kontrolle über  $\|\nabla \omega\|$ . Letztere gewinnt man erst auf dem Umweg über die primale Formulierung aus der (nichttrivialen) a priori Abschätzung

$$\|\nabla \omega\| \leq \|\nabla^3 \Psi\| \leq c \|f\|. \quad (6.1.34)$$

Zur Diskretisierung der gemischten Formulierung (6.1.28, 6.1.29) verwenden wir einen  $H^1$ -konformen Ansatz:

$$\begin{aligned}V_h &:= \{\psi_h \in H^1(\Omega) \mid \psi_h|_K \in P(K), K \in \mathbb{T}_h\}, \\ V_h^0 &:= \{\varphi_h \in H_0^1(\Omega) \mid \varphi_h|_K \in P^0(K), K \in \mathbb{T}_h\}.\end{aligned}$$

Die Polynomräume  $P(K)$ ,  $P^0(K)$  sind im einfachsten Fall gerade  $P(K) = P^0(K) = P_1(K)$  (lineare Dreieckselemente) oder  $P(K) = P^0(K) = \tilde{Q}_1(K)$ , (isoparametrische bilineare Viereckselemente). Die gemischte Finite-Elemente-Diskretisierung lautet dann wie folgt: Finde  $\{\omega_h, \Psi_h\} \in V_h \times V_h^0$ , so dass

$$(\omega_h, \psi_h) - (\nabla \Psi_h, \nabla \psi_h) = 0 \quad \forall \psi_h \in V_h, \quad (6.1.35)$$

$$(\nabla \omega_h, \nabla \varphi_h) = (f, \text{rot} \varphi_h) \quad \forall \varphi_h \in V_h^0. \quad (6.1.36)$$

Im Hinblick auf die vorausgehende Bemerkung betrachten wir im Folgenden nur den Standardfall, dass die Ansätze für  $\{\omega_h, \Psi_h\} \in V_h \times V_h^0$  die Eigenschaft  $V_h^0 \subset V_h$  besitzen und insbesondere vom gleichen Polynomgrad sind, d. h.:

$$V_h^0 = \{\varphi_h \in V_h \mid \varphi_h|_{\partial\Omega} = 0\}.$$

Dabei konzentrieren wir uns auf den einfachsten Fall linearer oder bilinearer Ansätze. Grundlage für die Fehleranalyse des gemischten Verfahrens ist wieder seine Galerkin-Orthogonalität, welche hier die folgende Form besitzt:

$$(\omega - \omega_h, \psi_h) - (\nabla(\Psi - \Psi_h), \nabla\psi_h) = 0, \quad \psi_h \in V_h, \quad (6.1.37)$$

$$(\nabla(\omega - \omega_h), \nabla\varphi_h) = 0, \quad \varphi_h \in V_h^0. \quad (6.1.38)$$

**Satz 6.4 (Gemischte Approximation):** Für die Diskretisierung (6.1.35, 6.1.36) mit linearen (oder bilinearen) Ansätzen gelten für die Fehler  $e_h^\omega := \omega - \omega_h$  sowie  $e_h^\Psi := \Psi - \Psi_h$  die a priori Abschätzungen

$$\|e_h^\omega\| + \|\nabla e_h^\Psi\| \leq ch^{1/2}L(h)^2\|f\|, \quad (6.1.39)$$

wobei  $L(h) := |\ln(h)| + 1$ . Für quadratische (oder bi-quadratische) Ansätze gilt

$$\|e_h^\omega\| + h\|\nabla e_h^\Psi\| \leq ch^2\|f\|. \quad (6.1.40)$$

**Beweis:** Im Folgenden bezeichnen  $P_h^0\Psi \in V_h^0$  sowie  $P_h\omega \in V_h$  Approximierende von  $\Psi$  bzw.  $\omega$  (Knoteninterpolierende,  $L^2$ -Projektion, etc.) mit den Eigenschaften

$$\|\Psi - P_h^0\Psi\| + h\|\nabla(\Psi - P_h^0\Psi)\| \leq ch^2\|\nabla^2\Psi\|, \quad (6.1.41)$$

$$\|\omega - P_h\omega\| + h\|\nabla(\omega - P_h\omega)\| \leq ch\|\nabla\omega\|. \quad (6.1.42)$$

i) Wir beginnen mit der Abschätzung von  $\|\nabla e_h^\Psi\|$ . Unter Verwendung der Orthogonalitätsbeziehung (6.1.37) folgt

$$\begin{aligned} \|\nabla e_h^\Psi\| &= (\nabla e_h^\Psi, \nabla(\Psi - P_h^0\Psi)) + (\nabla e_h^\Psi, \nabla(P_h^0\Psi - \Psi_h)) \\ &= (\nabla e_h^\Psi, \nabla(\Psi - P_h^0\Psi)) + (e_h^\omega, P_h^0\Psi - \Psi_h) \\ &= (\nabla e_h^\Psi, \nabla(\Psi - P_h^0\Psi)) + (e_h^\omega, P_h^0\Psi - \Psi) + (e_h^\omega, e_h^\Psi) \\ &\leq \|\nabla e_h^\Psi\| \|\nabla(\Psi - P_h^0\Psi)\| + \|e_h^\omega\| \{\|\Psi - P_h^0\Psi\| + \|e_h^\Psi\|\}. \end{aligned}$$

Mit der Poincaréschen Ungleichung,  $\|\varphi\| \leq c\|\nabla\varphi\|$ ,  $\varphi \in H_0^1(\Omega)$ , folgern wir, dass

$$\|\nabla e_h^\Psi\| \leq ch\|\nabla^2\Psi\| + c\|e_h^\omega\|. \quad (6.1.43)$$

ii) Als nächstes wird  $\|e_h^\omega\|$  abgeschätzt. Dazu verwenden wir eine Hilfskonstruktion. Sei  $R_h : H^1(\Omega) \rightarrow V_h$  die orthogonale Projektion bzgl. des Skalarprodukts  $(\nabla\cdot, \nabla\cdot) + (\cdot, \cdot)$ :

$$(\nabla(\varphi - R_h\varphi), \nabla\psi_h) + (\varphi - R_h\varphi, \psi_h) = 0 \quad \forall \psi_h \in V_h. \quad (6.1.44)$$

Für diese gilt die übliche  $L^2$ -Fehlerabschätzung

$$\|\varphi - R_h\varphi\| + h\|\nabla(\varphi - R_h\varphi)\| \leq h^r\|\nabla^r\varphi\|, \quad r \in \{1, 2\}. \quad (6.1.45)$$

sowie die Maximumnorm-Fehlerabschätzung (s. Ciarlet [80])

$$\|\varphi - R_h\varphi\|_\infty \leq ch^2L(h)^2\|\nabla^3\varphi\|. \quad (6.1.46)$$

Unter Verwendung der Orthogonalitätsbeziehungen (6.1.37) und (6.1.38) erschließen wir

$$\begin{aligned} \|e_h^\omega\|^2 &= (e_h^\omega, \omega - R_h\omega) + (e_h^\omega, R_h\omega - \omega_h) \\ &= (e_h^\omega, \omega - R_h\omega) - (\nabla e_h^\Psi, \nabla(R_h\omega - \omega_h)) \\ &= (e_h^\omega, \omega - R_h\omega) - (\nabla e_h^\Psi, \nabla(R_h\omega - \omega)) + (\nabla e_h^\Psi, \nabla e_h^\omega) \\ &\quad + (e_h^\Psi, R_h\omega - \omega) - (e_h^\Psi, R_h\omega - \omega) \\ &= (e_h^\omega, \omega - R_h\omega) - (\nabla(\Psi - R_h\Psi), \nabla(R_h\omega - \omega)) + (\Psi - R_h\Psi, R_h\omega - \omega) \\ &\quad - (e_h^\Psi, R_h\omega - \omega) - (\nabla e_h^\Psi, \nabla e_h^\omega), \end{aligned}$$

und weiter

$$\begin{aligned} \|e_h^\omega\|^2 &\leq \|e_h^\omega\| \|\omega - R_h\omega\| + \|\nabla(\Psi - R_h\Psi)\| \|\nabla(\omega - R_h\omega)\| \\ &\quad + \|\Psi - R_h\Psi\| \|\omega - R_h\omega\| + \|e_h^\Psi\| \|\omega - R_h\omega\| + |(\nabla e_h^\Psi, \nabla e_h^\omega)| \\ &\leq ch\{ \|e_h^\omega\| + (1+h)\|\nabla^2\Psi\| \} \|\nabla\omega\| + ch\|e_h^\Psi\| \|\nabla\omega\| + |(\nabla e_h^\Psi, \nabla e_h^\omega)|. \end{aligned}$$

Kombination der bisherigen Abschätzungen ergibt dann

$$\|e_h^\omega\|^2 \leq ch\|\nabla^3\Psi\|^2 + |(\nabla e_h^\Psi, \nabla e_h^\omega)|. \quad (6.1.47)$$

Für den letzten Term rechts erhalten wir mit Hilfe der Orthogonalitätseigenschaften, dass

$$(\nabla e_h^\Psi, \nabla e_h^\omega) = (\nabla\Psi, \nabla e_h^\omega) = (\nabla(\Psi - R_h\Psi), \nabla e_h^\omega) + (\nabla R_h\Psi, \nabla e_h^\omega),$$

Man beachte, dass zwar  $\Psi \in H_0^1(\Omega)$  ist, aber im Allgemeinen  $R_h\Psi \notin V_h^0$ , so dass der letzte Term rechts nicht verschwindet. Damit folgt weiter

$$\begin{aligned} (\nabla e_h^\Psi, \nabla e_h^\omega) &= (\nabla(\Psi - R_h\Psi), \nabla e_h^\omega) + (\nabla R_h\Psi, \nabla e_h^\omega) \\ &= (\nabla(\Psi - R_h\Psi), \nabla(\omega - R_h\omega)) + (\Psi - R_h\Psi, \omega - R_h\omega) \\ &\quad - (\Psi - R_h\Psi, e_h^\omega) + (\nabla R_h\Psi, \nabla e_h^\omega) \\ &\leq \|\nabla(\Psi - R_h\Psi)\| \|\nabla(\omega - R_h\omega)\| + \|\Psi - R_h\Psi\| \|\omega - R_h\omega\| \\ &\quad + \|\Psi - R_h\Psi\| \|e_h^\omega\| + |(\nabla R_h\Psi, \nabla e_h^\omega)| \\ &\leq c(h+h^3)\|\nabla^2\Psi\| \|\nabla\omega\| + ch^2\|\nabla^2\Psi\| \|e_h^\omega\| + |(\nabla R_h\Psi, \nabla e_h^\omega)|. \end{aligned}$$

Dies liefert zusammen mit (6.1.47), dass

$$\|e_h^\omega\|^2 \leq ch\|\nabla^3\Psi\|^2 + |(\nabla R_h\Psi, \nabla e_h^\omega)|. \quad (6.1.48)$$

iii) Der letzte Term rechts erfordert eine spezielle Behandlung. Seien mit  $a$  die Knotenpunkte der Zerlegung  $\mathbb{T}_h$ , und mit  $\psi_h \in V_h$  die zugehörigen Knotenbasisfunktionen bezeichnet. Wegen der Orthogonalitätsbeziehung (6.1.37) gilt dann

$$(\nabla R_h \Psi, \nabla e_h^\omega) = \sum_{a \in \partial\Omega} R_h \Psi(a) (\nabla \psi_h, \nabla e_h^\omega).$$

Mit Hilfe von

$$|R_h \Psi(a)| = |(R_h \Psi - \Psi)(a)| \leq \|R_h \Psi - \Psi\|_\infty \leq ch^2 L(h)^2 \|\nabla^3 \Psi\|$$

und  $|\nabla \psi_h| \leq ch^{-1}$  sowie  $\|1\|_{S_h} \leq ch^{1/2}$  folgt

$$|(R_h \Psi, \nabla e_h^\omega)| \leq ch^{3/2} L(h)^2 \|\nabla^3 \Psi\| \|\nabla e_h^\omega\|_{S_h}$$

wobei  $S_h := \cup_{a \in \partial\Omega} \text{Träger}(\psi_h)$ . Die inverse Beziehung ergibt wieder

$$\|\nabla e_h^\omega\| \leq ch^{-1} \|e_h^\omega\| + c \|\nabla \omega\|,$$

und folglich

$$\begin{aligned} |(\nabla R_h \Psi, \nabla e_h^\omega)| &\leq ch^{3/2} L(h)^2 \|\nabla^3 \Psi\| \{h^{-1} \|e_h^\omega\| + \|\nabla \omega\|\} \\ &\leq \frac{1}{2} \|e_h^\omega\|^2 + chL(h)^4 \|\nabla^3 \Psi\|^2. \end{aligned}$$

Einsetzen dieser Abschätzung in (6.1.48) ergibt schließlich

$$\|e_h^\omega\|^2 \leq chL(h)^4 \|\nabla^3 \Psi\|^2.$$

Zusammen mit (6.1.43) und der a priori Abschätzung (6.1.34) liefert dies

$$\|e_h^\omega\| + \|\nabla e_h^\Psi\| \leq ch^{1/2} L(h)^2 \|f\|,$$

was den Beweis von (6.1.39) vervollständigt. Der Beweis der Abschätzung (6.1.40) verwendet wieder Galerkin-Orthogonalität und ein Dualitätsargument analog zur Argumentation im Fall der Plattengleichung und sei als (nicht ganz einfache) Übungsaufgabe gestellt. Q.E.D.

**Bemerkung 6.2:** Für die gemischte Approximation der Stromfunktionsgleichung kann auch für *lineare* Ansätze auf allgemeinen Gittern Konvergenz garantiert werden, was an der Verwendung einer skalaren Variablen  $\omega = -\Delta \Psi$  liegt. Dies macht den „Trick“ in o. a. Beweisteil (iii) möglich. Im Fall der Plattengleichung ist die entsprechende Variable ein Tensor  $M = \nabla^2 u$ , für den diese Argumentation nicht funktioniert. Hier wird für Konvergenz eine gewisse Gleichmäßigkeit des Gitters gefordert, um Superapproximationseffekte ausnutzen zu können. Ob das wirklich notwendig ist, bleibt eine noch offene Frage für weitere Forschung.

## 6.2 FE-Diskretisierung des Stokes-Problems

Wir betrachten nun die direkte Diskretisierung des linearen Stokes-Problems. Die dabei verwendeten Finite-Elemente-Ansätze werden „Stokes-Elemente“ genannt. Sei  $\Psi \in W := H_0^2(\Omega)$  Lösung der (linearen) Stromfunktionsgleichung

$$(\Delta\Psi, \Delta\Phi) = (f, \text{rot}\Phi) \quad \forall \Phi \in W. \quad (6.2.49)$$

Für einen Moment werde  $\Psi \in H^3(\Omega)$  angenommen. Mit der Zuordnung  $v := \text{rot}\Psi$  und  $\varphi := \text{rot}\Phi$  folgt wegen  $\text{rot rot} = -\Delta$  durch partielle Integration:

$$\begin{aligned} (\Delta\Psi, \Delta\Phi) &= -(\Delta\Psi, \text{rot rot}(\Phi)) = -(\text{rot}(\Delta\Psi), \text{rot}(\Phi)) = -(\Delta\text{rot}(\Psi), \text{rot}(\Phi)) \\ &= -(\Delta v, \varphi) = (\nabla v, \nabla\varphi). \end{aligned}$$

Die Randintegrale sind wegen  $\partial_n \Psi|_{\partial\Omega} = 0$  gleich Null. Die Annahme  $\Psi \in H^3(\Omega)$  kann wegen der Dichtheit von  $H^3(\Omega) \subset H^2(\Omega)$  durch das übliche Approximationsargument zu  $\Psi \in W$  abgeschwächt werden. Es gilt also allgemein für  $v := \text{rot}\Psi$  und  $\varphi := \text{rot}\Phi$ :

$$(\nabla v, \nabla\varphi) = (\Delta\Psi, \Delta\Phi) = (f, \varphi) \quad \forall \varphi \in V, \quad (6.2.50)$$

wobei  $V := \text{rot}(W)$  gesetzt wird. Wenn über den Stromfunktionsansatz ein divergenz-freies Geschwindigkeitsfeld  $v$  gewonnen worden ist, muss noch ein zugehöriger Druck bestimmt werden. Dies geschieht über die variationelle Gleichung

$$(p, \nabla \cdot \varphi) = (\nabla v, \nabla\varphi) - (f, \varphi) \quad \forall \varphi \in H := H_0^1(\Omega). \quad (6.2.51)$$

Wie wir im Theoriekapitel gesehen haben, besitzt dieses Problem eine Lösung  $p \in L^2(\Omega)$ , welche bis auf eine Konstante eindeutig bestimmt ist. Für diese gilt die sog. „inf-sup“-Stabilitätsungleichung

$$\inf_{\chi \in L_0^2(\Omega)} \left\{ \sup_{\varphi \in H_0^1(\Omega)^2} \frac{(\chi, \nabla \cdot \varphi)}{\|\chi\| \|\nabla\varphi\|} \right\} \geq \beta > 0. \quad (6.2.52)$$

Diese gibt Kontrolle über den Druck durch die Daten des Problems. Unter Verwendung der a priori Abschätzung für das Geschwindigkeitsfeld

$$\|\nabla v\| \leq \|f\|_{-1},$$

mit der dualen Sobolew-Norm

$$\|f\|_{-1} := \sup_{\varphi \in H_0^1(\Omega)^d} \frac{(f, \varphi)}{\|\nabla\varphi\|}$$

folgt

$$\|p\| \leq \beta^{-1} \sup_{\varphi \in H_0^1(\Omega)^d} \frac{(\nabla v, \nabla\varphi) - (f, \varphi)}{\|\nabla\varphi\|} \leq 2\beta^{-1} \|f\|_{-1}. \quad (6.2.53)$$

### 6.2.1 „Exakt“ divergenzfreie Stokes-Elemente

Wir betrachten zunächst einen *konformen* Finite-Elemente-Ansatz für die Stromfunktions-Gleichung: *Finde*  $\Psi_h \in W_h \subset H_0^2(\Omega)$ , *so dass*

$$(\Delta \Psi_h, \Delta \Phi_h) = (f, \text{rot}(\Phi_h)) \quad \forall \Phi_h \in W_h. \quad (6.2.54)$$

Dieses Problem ist im Hinblick auf die Abschätzung

$$\|\nabla \Psi\| \leq c \|\nabla^2 \Psi\| \leq c \|\Delta \Psi\|, \quad \Psi \in H_0^2(\Omega),$$

stets eindeutig lösbar. Analog zu der Beziehung  $v = \text{rot} \Psi$  ordnen wir der diskreten Stromfunktion einen „diskreten“ Geschwindigkeitsvektor zu durch  $v_h := \text{rot} \Psi_h$ . Der zugehörige diskrete „Geschwindigkeitsraum“ ist

$$V_h := \text{rot}(W_h) \subset J_1(\Omega) \subset H_0^1(\Omega)^2.$$

Mit der Zuordnung  $\varphi_h := \text{rot} \Phi_h$  für  $\Phi \in W_h$  erhalten wir unter Beachtung von (6.2.50):

$$(\nabla v_h, \nabla \varphi_h) = (\Delta \Psi_h, \Delta \Phi_h) = (f, \text{rot} \Phi_h) = (f, \varphi_h) \quad \forall \varphi_h \in V_h. \quad (6.2.55)$$

Die so gewonnene Geschwindigkeitsnäherung  $v_h \in V_h$  ist offensichtlich „exakt“ divergenzfrei. Man mache sich diesen Zusammenhang zwischen konformen Ansätzen für die Stromfunktionsgleichung und entsprechenden, divergenzfreien Stokes-Elementen anhand des Argyris-Elements klar. Hier ist

$$V_h = \text{rot}(W_h^{\text{Argyris}}) \subset H_h = \{\varphi_h \in H_0^1(\Omega)^2 \mid \varphi_h|_K \in P_4(K), K \in \mathbb{T}_h\}.$$

Um direkt mit dem zugehörigen Ansatzraum  $V_h \subset J_1(\Omega)$  arbeiten zu können, benötigen wir eine Knotenbasis von  $V_h$ . Da der Rotationsoperator  $\text{rot} = (-\partial_2, \partial_1)^T$ , auf  $W_h^{\text{Argyris}}$  injektiv ist, erhalten wir eine solche Basis direkt durch Anwendung von  $\text{rot}$  auf die Elemente der natürlichen Knotenbasis von  $W_h^{\text{Argyris}}$ .

*Bestimmung des „diskreten“ Drucks:* Es bleibt, einen zugehörigen „diskreten“ Druck zu bestimmen. Offenbar ist  $V_h = \{\varphi_h \in H_h \mid \nabla \cdot \varphi_h = 0\} \subset H_h$ . Wir wählen einen „Druckansatzraum“  $L_h \subset L_0^2(\Omega)$ . Der zu  $v_h \in V_h$  gehörende diskrete Druck  $p_h \in L_h$  wird dann durch die folgende variationelle Gleichung bestimmt:

$$(p_h, \nabla \cdot \varphi_h) = (\nabla v_h, \nabla \varphi_h) - (f, \varphi_h) \quad \forall \varphi_h \in H_h. \quad (6.2.56)$$

**Lemma 6.1 (Druck):** *Für jede Wahl von Ansatzräumen  $V_h := \text{rot}(W_h)$  und  $L_h \subset L$  existiert ein Lösungspaar  $\{v_h, p_h\} \in V_h \times L_h$  der Gleichungen (6.2.55) und (6.2.56).*

**Beweis:** Die Existenz einer (eindeutigen) Lösung  $v_h \in V_h$  von (6.2.55) haben wir bereits oben begründet. Die Existenz dieses zugehörigen Drucks gemäß (6.2.56) diskutieren wir auf dem Algebra-Level. Seien  $\{\varphi_h^i, i = 1, \dots, N_H := \dim H_h\}$  sowie  $\{\chi_h^j, j = 1, \dots, N_L := \dim L_h\}$  Knotenbasen der Finite-Element-Räume  $H_h$  bzw.  $L_h$  und  $x \in \mathbb{R}^{N_H}$  sowie  $y \in \mathbb{R}^{N_L}$  die zugehörigen Knotenvektoren:

$$v_h = \sum_{i=1}^{N_H} x_i \varphi_h^i, \quad p_h = \sum_{i=1}^{N_L} y_i \chi_h^i.$$

Mit den Matrizen und dem Vektor

$$A := ((\nabla \varphi_h^i, \nabla \varphi_h^j))_{i,j=1}^{N_H}, \quad B := ((\chi_h^i, \nabla \cdot \varphi_h^j))_{i,j=1}^{N_L, N_H}, \quad b := (f, \varphi_h^j)_{j=1}^{N_H},$$

ist Problem (6.2.56) äquivalent zu dem folgenden linearen Gleichungssystem

$$By = Ax - b. \quad (6.2.57)$$

Wegen  $\text{Bild}(B) = \text{Kern}(B^T)^\perp$  ist dieses System genau dann lösbar, wenn die rechte Seite  $Ax - b$  orthogonal zu  $\text{Kern}(B^T)$  ist, d. h. wenn

$$\langle Ax - b, z \rangle = 0 \quad \forall z \in \text{Kern}(B^T).$$

Der Kern von  $B^T$  ist charakterisiert durch

$$\langle B^T x, y \rangle = (\nabla \varphi_h, \chi_h) = 0 \quad \forall \chi_h \in L_h,$$

d. h.: Alle Funktionen  $\varphi_h \in V_h$  haben Knotenvektoren  $x \in \text{Kern}(B^T)$ . Dies ist aber wegen

$$(\nabla v_h, \nabla \varphi_h) - (f, \varphi_h) = 0, \quad \varphi_h \in V_h,$$

automatisch der Fall.

Q.E.D.

Die Existenz eines diskreten Drucks  $p_h \in L_h$  besagt aber noch nicht, dass dieser auch eine brauchbare Approximation zum Druck  $p$  ist. Dazu sollte er eindeutig (bis auf Konstanten) bestimmt sein, stetig von den Problemdata abhängen und einer Fehlerabschätzung genügen. Der Schlüssel zu all diesen Forderungen ist das diskrete Analogon der „inf-sup“-Ungleichung (6.2.52), nämlich

$$\min_{\chi_h \in L_h} \left\{ \max_{\varphi_h \in H_h} \frac{(\chi_h, \nabla \cdot \varphi_h)}{\|\chi_h\| \|\nabla \varphi_h\|} \right\} \geq \beta_h \geq \beta_* > 0, \quad (6.2.58)$$

möglichst gleichmäßig für  $h \rightarrow 0$  nach unten beschränkten Stabilitätskonstanten  $\beta_h$ . Diese Bedingung wird in der Literatur aus historischen Gründen auch „Babuska-Brezzi“ oder kurz „BB-Bedingung“ genannt. Aus (6.2.58) folgt unmittelbar die Eindeutigkeit des diskreten Drucks  $p_h$  bis auf Konstanten. Weiter erhalten wir analog wie im kontinuierlichen Fall für  $p_h \in L_h$  die a priori Abschätzungen

$$\|\nabla v_h\| \leq \|f\|_{-1}, \quad \|p_h\| \leq 2\beta_*^{-1} \|f\|_{-1}.$$

Die Gültigkeit der Stabilitätsabschätzung hängt von der „richtigen“ Wahl des Druckansatzraumes  $L_h$  ab; dieser darf offenbar nicht zu „groß“ sein verglichen mit dem Geschwindigkeitsansatzraum  $H_h$ . Wir werden weiter unten sehen, dass z.B.

$$L_h^{(2)} := \{\chi_h \in L^2(\Omega) \mid \chi_h|_K \in P_2(K), K \in \mathbb{T}_h\}$$

zusammen mit dem Argyris-Element gewonnenen Geschwindigkeitsansatz zulässig ist. Die zugehörige Fehlerabschätzung formulieren wir nun als Satz.

**Satz 6.5 (Druckapproximation):** *i) Es gelte die gleichmäßige „inf-sup“-Stabilitätsungleichung (6.2.58). Dann gilt für den durch (6.2.56) eindeutig definierten diskreten Druck  $p_h \in L_h$  die allgemeine Fehlerabschätzung:*

$$\|p - p_h\| \leq (1 + c\beta_h^{-1}) \min_{\chi_h \in L_h} \|p - \chi_h\| + \beta_h^{-1} \min_{\Phi_h \in W_h} \|\nabla^2(\Psi - \Phi_h)\|. \quad (6.2.59)$$

*ii) Im Fall des Argyris-Elements gilt die „inf-sup“-Stabilitätsungleichung (6.2.58) gleichmäßig bzgl.  $h$  und es gilt mit  $p_h \in L_h^{(2)}$ :*

$$\|p - p_h\| \leq c\{h^3\|p\|_{H^3} + h^4\|v\|_{H^5}\}, \quad (6.2.60)$$

*vorausgesetzt, dass  $v$  und  $p$  diese Regularität besitzen.*

**Beweis:** i) Durch Vergleich der kontinuierlichen Gleichung (6.2.51) mit ihrem diskreten Analogon (6.2.56) folgt die Beziehung

$$(p - p_h, \nabla \cdot \varphi_h) = (\nabla(v - v_h), \nabla \varphi_h), \quad \varphi_h \in H_h. \quad (6.2.61)$$

Mit Hilfe der inf-sup-Stabilitätsabschätzung (6.2.58) und der Beziehung (6.2.61) schätzen wir wie dann folgt ab:

$$\begin{aligned} \|p - p_h\| &\leq \|p - \chi_h\| + \|\chi_h - p_h\| \\ &\leq \|p - \chi_h\| + \beta_h^{-1} \max_{\varphi_h \in H_h} \frac{(\chi_h - p_h, \nabla \cdot \varphi_h)}{\|\nabla \varphi_h\|} \\ &\leq \|p - \chi_h\| + \beta_h^{-1} \max_{\varphi_h \in H_h} \frac{(\chi_h - p, \nabla \cdot \varphi_h)}{\|\nabla \varphi_h\|} + \beta_h^{-1} \max_{\varphi_h \in H_h} \frac{(p - p_h, \nabla \cdot \varphi_h)}{\|\nabla \varphi_h\|} \\ &\leq (1 + c\beta_h^{-1})\|p - \chi_h\| + \beta_h^{-1} \max_{\varphi_h \in H_h} \frac{(p - p_h, \nabla \cdot \varphi_h)}{\|\nabla \varphi_h\|} \\ &= (1 + c\beta_h^{-1})\|p - \chi_h\| + \beta_h^{-1} \max_{\varphi_h \in H_h} \frac{(\nabla(v - v_h), \nabla \varphi_h)}{\|\nabla \varphi_h\|} \\ &\leq (1 + c\beta_h^{-1})\|p - \chi_h\| + \beta_h^{-1} \|\nabla(v - v_h)\| \\ &\leq (1 + c\beta_h^{-1})\|p - \chi_h\| + \beta_h^{-1} \|\nabla^2(\Psi - \Psi_h)\| \\ &\leq (1 + c\beta_h^{-1})\|p - \chi_h\| + \beta_h^{-1} \|\Delta(\Psi - \Psi_h)\| \end{aligned}$$

mit einem beliebigen  $\chi_h \in L_h$ . Dies impliziert die erste behauptete Ungleichung:

$$\|p - p_h\| \leq (1 + c\beta_h^{-1}) \min_{\chi_h \in L_h} \|p - \chi_h\| + \beta_h^{-1} \min_{\Phi_h \in W_h} \|\Delta(\Psi - \Phi_h)\|. \quad (6.2.62)$$

ii) Als nächstes bemerken, dass für das oben definierte Raumpaar  $\{H_h, L_h\}$  mit  $H_h = H_h^{(4)} := \{\varphi_h \in H_0^1(\Omega)^2 \mid \varphi_{|K} \in P_4(K), K \in \mathbb{T}_h\}$  und  $L_h = L_h^{(2)}$  die gleichmäßige „inf-sup“-Stabilitätsbedingung (6.2.58) erfüllt ist (Übungsaufgabe).

Aus der allgemeinen Fehlerabschätzung (6.2.59) folgt damit für das Argyris-Element:

$$\begin{aligned} \|p - p_h\| &\leq (1 + c\beta_*^{-1}) \min_{\chi_h \in L_h} \|p - \chi_h\| + \beta_*^{-1} \min_{\Phi_h \in W_h} \|\nabla^2(\Psi - \Phi_h)\| \\ &\leq c\{h^2\|p\|_{H^2} + h^4\|\Psi\|_{H^6}\}, \end{aligned}$$

was den Beweis vervollständigt.

Q.E.D.

Als zweites Beispiel betrachten wir das nicht-konforme Morley-Element. In diesem Fall ergibt sich durch Rotationsbildung  $H_h := \text{rot}(V_h^M)$  ein wieder nicht-konformer „linearer“ Geschwindigkeitsansatz, der zusammen mit stückweise konstanten Druckansätzen der inf-sup-Bedingung genügt. Das resultierende ‘Stokes-Element’ wird unten noch genauer diskutiert werden. Für die Knotenbasis  $\{\Psi_h^i\}$  des Morley-Ansatzes erhält man dann durch Setzung  $\psi_h^i := \text{rot}\Psi_h^i$  eine lokale Basis des Teilraums der diskret divergenz-freien Geschwindigkeitsansätze  $V_h := \{v_h \in H_h \mid (\nabla \cdot v_h, \chi_h)_h = 0 \ \forall \chi_h \in L_h\}$ .

### 6.2.2 Allgemeine „Stokes-Elemente“

Wir betrachten nun die Diskretisierung des allgemeinen, stationären Stokes-Problems losgelöst von der Stromfunktionsformulierung. Dies eröffnet insbesondere den Weg zur Behandlung auch des 3-dimensionalen Falls. Der Einfachheit halber wird die Stokes-Gleichung wieder auf einem polyedrischen Gebiet  $\Omega \subset \mathbb{R}^d$  ( $d = 2$  oder  $d = 3$ ) betrachtet, und es werden homogene Dirichlet-Randbedingungen  $v|_{\partial\Omega} = 0$  gestellt. Mit dem entsprechenden technischen Mehraufwand ist auch die Behandlung von Gebieten mit krummen Rändern sowie inhomogene Einströmdaten und freie Ausströmränder möglich.

Ausgangspunkt ist die variationelle Formulierung in den Räumen  $H := H_0^1(\Omega)^d$  und  $L := L_0^2(\Omega)$ : Finde  $\{v, p\} \in H \times L$ , so dass

$$(\nabla v, \nabla \varphi) - (p, \nabla \cdot \varphi) = (f, \varphi) \quad \forall \varphi \in H, \quad (6.2.63)$$

$$(\chi, \nabla \cdot v) = 0 \quad \forall \chi \in L. \quad (6.2.64)$$

Auf konvexen Polyedern oder allgemeinen glatt-berandeten Gebieten hat die (eindeutige) Lösung dieses Problems wieder die Regularität  $\{v, p\} \in H^2(\Omega)^d \times H^1(\Omega)$  und es gilt die a priori Abschätzung

$$\|v\|_{H^2} + \|p\|_{H^1} \leq c_\Omega \|f\|. \quad (6.2.65)$$

Seien nun  $H_h \subset H$  und  $L_h \subset L$  FE-Teilraum, welche „diskret“ inf-sup-stabil sind:

$$\min_{\chi_h \in L_h} \left\{ \max_{\varphi_h \in H_h} \frac{(\chi_h, \nabla \cdot \varphi_h)}{\|\chi_h\| \|\nabla \varphi_h\|} \right\} \geq \beta_h \quad (6.2.66)$$

mit möglicherweise von  $h$  abhängigen Konstanten  $\beta_h > 0$ . Die approximierenden Probleme lauten dann: Finde  $\{v_h, p_h\} \in H_h \times L_h$ , so dass

$$(\nabla v_h, \nabla \varphi_h) - (p_h, \nabla \cdot \varphi_h) = (f, \varphi_h) \quad \forall \varphi_h \in H_h, \quad (6.2.67)$$

$$(\chi_h, \nabla \cdot v_h) = 0 \quad \forall \chi_h \in L_h. \quad (6.2.68)$$

Die Ansatzräume  $H_h \times L_h$  wird wegen  $H_h \subset H$  (und  $L_h \subset L$ ) als „konform“ bezeichnet. Wir werden später auch noch „nicht-konforme“ Ansatzräume kennenlernen. Wir bezeichnen wieder mit

$$V_h := \{v_h \in H_h \mid (\nabla \cdot v_h, \chi_h) = 0 \ \forall \chi_h \in L_h\}.$$

den Teilraum der diskret divergenz-freien Geschwindigkeitsansätze.

Aufgrund der Diskussion im vorigen Abschnitt besitzen die diskreten Stokes-Probleme (6.2.67) - (6.2.68) eindeutig bestimmte Lösungspaare  $\{v_h, p_h\} \in H_h \times L_h$ . Für die Konvergenz  $\{v_h, p_h\} \rightarrow \{v, p\}$  für  $h \rightarrow 0$  ist erforderlich, dass die Räume  $V_h$  ausreichend „groß“ sind, um damit Funktionen aus  $V$  approximieren zu können. Dazu fordern wir die folgende Approximationseigenschaft:

$$\min_{\varphi_h \in H_h} \|\nabla(v - \varphi_h)\| + \min_{\chi_h \in L_h} \|p - \chi_h\| \leq ch\{\|v\|_{H^2} + \|p\|_{H^1}\} \leq ch\|f\|. \quad (6.2.69)$$

**Lemma 6.2 (Quasi-Bestapproximationseigenschaft):** *Die konforme Approximation (6.2.67) - (6.2.68) des Stokes-Problems (6.2.63) - (6.2.64) besitzt die „(quasi)-Bestapproximationseigenschaft“ bzgl. der Ansatzräume  $V_h$ :*

$$\|\nabla(v - v_h)\| \leq c\left\{ \min_{\chi_h \in V_h} \|\nabla(v - \varphi_h)\| + \min_{\varphi_h \in L_h} \|p - \chi_h\| \right\}. \quad (6.2.70)$$

**Beweis:** Die Subtraktion der diskreten Stokes-Gleichungen von den kontinuierlichen ergibt die „Galerkin-Orthogonalitätsgleichungen“

$$(\nabla(v - v_h), \nabla\varphi_h) - (p - p_h, \nabla \cdot \varphi_h) = 0, \quad \varphi_h \in H_h. \quad (6.2.71)$$

$$(\chi_h, \nabla \cdot (v - v_h)) = 0, \quad \chi_h \in L_h. \quad (6.2.72)$$

Wegen  $v_h \in V_h$  folgt für beliebiges  $\varphi_h \in V_h$  und  $\chi_h \in L_h$  wegen (6.2.71) und (6.2.72):

$$\begin{aligned} \|\nabla(v - v_h)\|^2 &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (\nabla(v - v_h), \nabla(\varphi_h - v_h))) \\ &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - p_h, \nabla \cdot (\varphi_h - v_h))) \\ &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - \chi_h, \nabla \cdot (\varphi_h - v)) + (p - \chi_h, \nabla \cdot (v - v_h))). \end{aligned}$$

Daraus folgt mit Hilfe der Youngschen Ungleichung  $ab \leq a^2 + \frac{1}{4}b^2$ :

$$\|\nabla(v - v_h)\|^2 \leq c\{\|\nabla(v - \varphi_h)\|^2 + \|p - \chi_h\|^2\}.$$

Da  $\varphi_h \in V_h$  und  $\chi_h \in L_h$  beliebig gewählt waren, ergibt sich durch Infimumbildung die behauptete Abschätzung. Q.E.D.

Die obige Fehlerabschätzung (6.2.70) ist unbefriedigend, da es nicht offensichtlich ist, wie man eine geeignete Approximierende  $i_h v \in V_h$  zu  $v \in V$  mit bekanntem Fehlerverhalten konstruieren kann. Dies ändert sich, wenn die Raumpaare  $H_h \times L_h \subset H \times L$  *gleichmäßig* inf-sup-stabil sind:

$$\min_{\chi_h \in L_h} \left\{ \max_{\varphi_h \in H_h} \frac{(\chi_h, \nabla \cdot \varphi_h)}{\|\chi_h\| \|\nabla \varphi_h\|} \right\} \geq \beta_* \quad (6.2.73)$$

mit einer festen Konstante  $\beta_* > 0$ .

**Lemma 6.3:** *Erfüllt das Teilraumpaar  $H_h \times L_h \subset H \times L$  die gleichmäßige inf-sup-Stabilitätsungleichung (6.2.73), so gilt auch die dazu „adjungierte“ Abschätzung*

$$\min_{\varphi_h \in H_h} \left\{ \max_{\chi_h \in L_h} \frac{(\chi_h, \nabla \cdot \varphi_h)}{\|\chi_h\| \|\nabla \varphi_h\|} \right\} \geq \beta_*. \quad (6.2.74)$$

Ferner gilt für  $v \in V$ :

$$\min_{\varphi_h \in \tilde{V}_h} \|\nabla(v - \varphi_h)\| \leq (1 + c\beta_*^{-1}) \min_{\varphi_h \in H_h} \|\nabla(v - \varphi_h)\|. \quad (6.2.75)$$

**Beweis:** i) Wir beweisen zunächst (6.2.74). Durch

$$\langle B_h \varphi_h, \chi_h \rangle := (\chi_h, \nabla \cdot \varphi_h) \quad \forall \chi_h \in L_h,$$

wird ein linearer Operator  $B_h : H_h \rightarrow L_h^*$  definiert. Der zugehörige adjungierte Operator  $B_h^* : L_h \rightarrow H_h^*$  ist definiert durch

$$\langle B_h^* \chi_h, \varphi_h \rangle := \langle B_h \varphi_h, \chi_h \rangle = (\chi_h, \nabla \cdot \varphi_h) \quad \forall \varphi_h \in H.$$

Da die Räume  $H_h$  und  $L_h$  endlich dimensional sind, gelten die Beziehungen

$$\begin{aligned} \text{Bild}(B_h) &= \text{Kern}(B_h^*)^\perp, & \text{Bild}(B_h)^\perp &= \text{Kern}(B_h^*), \\ \text{Bild}(B_h^*) &= \text{Kern}(B_h)^\perp, & \text{Bild}(B_h^*)^\perp &= \text{Kern}(B_h). \end{aligned}$$

Die Stabilitätsungleichung (6.2.73) impliziert dann

$$\|B_h^* \chi_h\|_{H^*} = \sup_{\varphi_h \in H_h} \frac{\langle B_h^* \chi_h, \varphi_h \rangle}{\|\nabla \varphi_h\|} = \sup_{\varphi_h \in H_h} \frac{(\chi_h, \nabla \cdot \varphi_h)}{\|\nabla \varphi_h\|} \geq \beta_* \|\chi_h\|, \quad \chi_h \in L_h.$$

Der Operator  $B_h^*$  ist also injektiv, d. h.  $\text{Bild}(B_h)^\perp = \text{Kern}(B_h^*) = \{0\}$  mit beschränkter Inverser  $B_h^{*-1} : \text{Kern}(B_h)^\perp \subset H^* \rightarrow L_h$ :

$$\|B_h^{*-1}\| \leq \beta_*^{-1}.$$

Dann ist auch  $B_h : H \rightarrow \text{Kern}(B_h^*)^\perp \subset L_h^*$  invertierbar mit derselben Schranke

$$\|B_h^{-1}\| \leq \beta_*^{-1}.$$

Dies wiederum impliziert

$$\beta_* \|\nabla \varphi_h\| \leq \|B_h \varphi_h\|_{L_h^*} = \sup_{\chi_h \in L_h} \frac{\langle B_h \varphi_h, \chi_h \rangle}{\|\chi_h\|} = \sup_{\chi_h \in L_h} \frac{(\chi_h, \nabla \cdot \varphi_h)}{\|\chi_h\|},$$

was zu beweisen war.

ii) Es bleibt, (6.2.75) zu zeigen. Seien  $v \in V$ ,  $v_h \in V_h$  und  $\varphi_h \in H_h$  beliebig. Damit gilt dann unter Verwendung von (6.2.74):

$$\begin{aligned} \|\nabla(v - v_h)\| &\leq \|\nabla(v - \varphi_h)\| + \|\nabla(\varphi_h - v_h)\| \\ &\leq \|\nabla(v - \varphi_h)\| + \beta_*^{-1} \sup_{\chi_h \in L_h} \frac{(\chi_h, \nabla \cdot (\varphi_h - v_h))}{\|\chi_h\|} \\ &= \|\nabla(v - \varphi_h)\| + \beta_*^{-1} \sup_{\chi_h \in L_h} \frac{(\chi_h, \nabla \cdot (\varphi_h - v))}{\|\chi_h\|} \\ &\leq (1 + c\beta_*^{-1}) \|\nabla(v - \varphi_h)\| \end{aligned}$$

Wegen der beliebigen Wahl von  $v_h \in V_h$  und  $\varphi_h \in H_h$  folgt hieraus (6.2.75). Q.E.D.

**Satz 6.6 (Bestapproximationseigenschaft):** *Die konforme Approximation (6.2.67) - (6.2.68) des Stokes-Problems (6.2.63) - (6.2.64) besitzt im Falle gleichmäßiger inf-sup-Stabilität die Bestapproximationseigenschaft:*

$$\|\nabla(v - v_h)\| + \|p - p_h\| \leq c \left\{ \min_{\varphi_h \in H_h} \|\nabla(v - \varphi_h)\| + \min_{\chi_h \in L_h} \|p - \chi_h\| \right\}, \quad (6.2.76)$$

$$\|v - v_h\| \leq ch \left\{ \min_{\varphi_h \in H_h} \|\nabla(v - \varphi_h)\| + \min_{\chi_h \in L_h} \|p - \chi_h\| \right\}. \quad (6.2.77)$$

Mit der Approximationseigenschaft (6.2.69) ergibt dies die Fehlerabschätzungen:

$$\|\nabla(v - v_h)\| + \|p - p_h\| \leq ch\|f\|, \quad (6.2.78)$$

$$\|v - v_h\| \leq ch^2\|f\|. \quad (6.2.79)$$

**Beweis:** Zunächst ergibt die Subtraktion der diskreten Stokes-Gleichungen von den Kontinuierlichen die Galerkin-Orthogonalitätsgleichungen

$$(\nabla(v - v_h), \nabla\varphi_h) - (p - p_h, \nabla \cdot \varphi_h) = 0, \quad \varphi_h \in H_h. \quad (6.2.80)$$

$$(\chi_h, \nabla \cdot (v - v_h)) = 0, \quad \chi_h \in L_h. \quad (6.2.81)$$

i) Wir beginnen mit der Abschätzung von  $\|\nabla(v - v_h)\|$ . Wir werden hierfür nicht das Resultat von Lemma 6.3 heranziehen, sondern stattdessen eine davon unabhängige, alternative Argumentation verwenden. Damit bereiten wir schon den später im Zusammenhang mit sog. „stabilisierten“ Stokes-Elementen verwendeten Beweisgang vor. Für beliebiges  $\varphi_h \in H_h$  folgt mit Hilfe von (6.2.80):

$$\begin{aligned} \|\nabla(v - v_h)\|^2 &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (\nabla(v - v_h), \nabla(\varphi_h - v_h)) \\ &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - p_h, \nabla \cdot (\varphi_h - v_h)) \\ &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - p_h, \nabla \cdot (\varphi_h - v)) + (p - p_h, \nabla \cdot (v - v_h)), \end{aligned}$$

und weiter mit Hilfe von (6.2.81) mit beliebigem  $\chi_h \in L_h$ :

$$\|\nabla(v - v_h)\|^2 = (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - p_h, \nabla \cdot (\varphi_h - v)) + (p - \chi_h, \nabla \cdot (v - v_h))).$$

Mit Hilfe der Youngschen Ungleichung  $ab \leq \varepsilon^2 a^2 + (4\varepsilon^2)^{-1} b^2$  folgern wir weiter

$$\|\nabla(v - v_h)\| \leq c \left\{ (1 + \varepsilon^{-1}) \|\nabla(v - \varphi_h)\| + \|p - \chi_h\| \right\} + \varepsilon \|p - p_h\|. \quad (6.2.82)$$

ii) Als nächstes wird  $\|p - p_h\|$  abgeschätzt. Mit Hilfe der inf-sup-Ungleichung (6.2.66) folgt

$$\begin{aligned} \|p - p_h\| &\leq \|p - \chi_h\| + \|\chi_h - p_h\| \\ &\leq \|p - \chi_h\| + \beta^{-1} \sup_{\psi_h \in H_h} \frac{(\chi_h - p_h, \nabla \cdot \psi_h)}{\|\nabla\psi_h\|} \\ &\leq \|p - \chi_h\| + \beta^{-1} \sup_{\psi_h \in H_h} \frac{(\chi_h - p, \nabla \cdot \psi_h)}{\|\nabla\psi_h\|} + \beta^{-1} \sup_{\psi_h \in H_h} \frac{(p - p_h, \nabla \cdot \psi_h)}{\|\nabla\psi_h\|} \\ &\leq c(1 + \beta^{-1}) \|p - \chi_h\| + \beta^{-1} \sup_{\psi_h \in H_h} \frac{(p - p_h, \nabla \cdot \psi_h)}{\|\nabla\psi_h\|}. \end{aligned}$$

Weiter impliziert die Galerkin-Orthogonalität (6.2.81)

$$\begin{aligned} \|p - p_h\| &\leq c(1 + \beta^{-1})\|p - \chi_h\| + \beta^{-1} \sup_{\psi_h \in H_h} \frac{(\nabla(v - v_h), \nabla\psi_h)}{\|\nabla\psi_h\|} \\ &\leq c(1 + \beta^{-1})\|p - \chi_h\| + c\|\nabla(v - v_h)\|. \end{aligned}$$

Durch Kombination dieser Abschätzung mit (6.2.82) erhalten wir

$$\begin{aligned} \|\nabla(v - v_h)\| &\leq c\{\|\nabla(v - \varphi_h)\| + \|p - \chi_h\|\} \\ &\quad + \varepsilon\{c(1 + \beta^{-1})\|p - \chi_h\| + c\|\nabla(v - v_h)\|\}. \end{aligned}$$

Bei hinreichend kleiner aber fester Wahl von  $\varepsilon > 0$  folgt

$$\|\nabla(v - v_h)\| \leq c\{\|\nabla(v - \varphi_h)\| + (1 + \beta^{-1})\|p - \chi_h\|\}.$$

Da  $\varphi_h \in H_h$  und  $\chi_h \in L_h$  beliebig gewählt waren, ergibt sich durch Infimumbildung die erste Abschätzung (6.2.76).

iii) Zur Abschätzung von  $\|v - v_h\|$  verwenden wir wieder ein Dualitätsargument. Sei  $\{z, q\} \in H \times L$  die Lösung des „dualen Stokes-Problems“

$$(\nabla\varphi, \nabla z) - (q, \nabla \cdot \varphi) = (\varphi, v - v_h)\|v - v_h\|^{-1} \quad \forall \varphi \in H, \quad (6.2.83)$$

$$(\chi, \nabla \cdot z) = 0 \quad \forall \chi \in L. \quad (6.2.84)$$

Diese ist in  $H^2(\Omega)^d \times H^1(\Omega)$  und es gilt die a priori Abschätzung

$$\|z\|_{H^2} + \|q\|_{H^1} \leq c\|v - v_h\|\|v - v_h\|^{-1} = c. \quad (6.2.85)$$

Mit der Testfunction  $\varphi := v - v_h$  ergibt sich dann unter Verwendung der Galerkin-Orthogonalität mit den natürlichen Knoteninterpolierenden  $i_h z \in H_h$  und  $j_h q \in L_h$ :

$$\begin{aligned} \|v - v_h\| &= (\nabla(v - v_h), \nabla z) - (q, \nabla \cdot (v - v_h)) \\ &= (\nabla(v - v_h), \nabla(z - i_h z)) + (\nabla(v - v_h), \nabla i_h z) - (q - j_h q, \nabla \cdot (v - v_h)) \\ &= (\nabla(v - v_h), \nabla(z - i_h z)) + (p - p_h, \nabla \cdot i_h z) - (q - j_h q, \nabla \cdot (v - v_h)) \\ &= (\nabla(v - v_h), \nabla(z - i_h z)) + (p - p_h, \nabla \cdot (i_h z - z)) - (q - j_h q, \nabla \cdot (v - v_h)). \end{aligned}$$

Mit Hilfe der Fehlerabschätzungen für die Knoteninterpolierenden schätzen wir weiter ab:

$$\begin{aligned} \|v - v_h\| &\leq \|\nabla(v - v_h)\|\|\nabla(z - i_h z)\| + \|p - p_h\|\|\nabla \cdot (i_h z - z)\| + \|q - j_h q\|\|\nabla \cdot (v - v_h)\| \\ &\leq ch\|\nabla(v - v_h)\|\|z\|_{H^2} + ch\|p - p_h\|\|z\|_{H^2} + ch\|q\|_{H^1}\|\nabla(v - v_h)\| \\ &\leq ch\{\|\nabla(v - v_h)\| + \|p - p_h\|\}. \end{aligned}$$

Dies vervollständigt den Beweis.

Q.E.D.

### I) Beispiele von „konformen“ Stokes-Elementen

a) *Elemente mit unstetigem Druck*: Diese Ansätze zeichnen sich dadurch aus, dass sie lokale Masseerhaltung garantieren

$$\int_{\partial K} n \cdot v_h \, do = \int_K \nabla \cdot v_h \, dx = 0, \quad K \in \mathbb{T}_h. \quad (6.2.86)$$

Ferner ist wegen der fehlenden Kopplung zwischen den Druckfreiheitsgraden zu verschiedenen Zellen die Massematrix des Druckansatzes block-diagonal bzw. im einfachsten Fall sogar diagonal. Die verschiedenen Stokes-Elemente werden im Folgenden wieder durch Angabe der Knotenwerte für Geschwindigkeit- und Druckansatz beschrieben. Wir beschränken uns auf Beispiele in 2 Dimensionen; in den meisten Fällen existieren natürliche Analoga in 3 Dimensionen.

i) Das  $P_1/P_0^{\text{dc}}$ -Element (a) und das  $Q_1^c/P_0^{\text{dc}}$ -Element (b):

$$a) \quad P_H(K) := P_1^c(K), \quad P_L(K) := P_0^{\text{dc}}(K);$$

$$b) \quad P_H(K) := Q_1^c(K), \quad P_L(K) := P_0^{\text{dc}}(K);$$

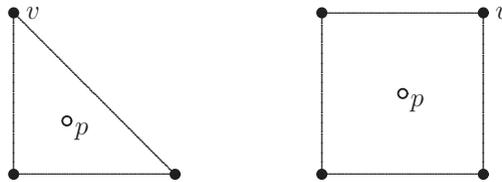


Abbildung 6.1: Die konformen Stokes-Elemente vom Typ  $P_1^c/P_0^{\text{dc}}$  (links) und  $Q_1^c/P_0^{\text{dc}}$  (rechts).

Das einfachste  $P_1/P_0$ -Element ist unbrauchbar, da hier der Teilraum  $V_h$  der „diskret divergenzfreien“ Geschwindigkeiten fast nur die Null enthält. Das  $Q_1/P_0$ -Element ist wegen seiner Einfachheit bei Ingenieuren sehr beliebt, erweist sich aber häufig als instabil, was sich in unphysikalischen Oszillationen im Druck zeigt. Die inf-sup-Stabilitätsbedingung ist hier nicht (oder nur mit  $h$ -abhängiger Konstante  $\beta_h$ ) erfüllt.

ii) Das  $P_2^c/P_0^{\text{dc}}$ -Element (a), das erweiterte  $P_2^b/P_1^{\text{dc}}$ -Element (b), und das  $Q_2^c/P_1^{\text{dc}}$ -Element (c):

$$a) \quad P_H(K) := P_2^c(K), \quad P_L(K) := P_0^{\text{dc}}(K);$$

$$b) \quad P_H(K) := P_2^c(K) \oplus \text{span}\{b_K^1, b_K^2\}, \quad P_L(K) := P_1^{\text{dc}}(K);$$

$$c) \quad P_H(K) := Q_2^c(K), \quad P_L(K) := P_1^{\text{dc}}(K).$$

Dabei ist  $b_K^1 = (b_K, 0)^T$  und  $b_K^2 = (0, b_K)^T$ , mit der (eindeutig bestimmten) kubischen „bulb-Funktion“

$$b_K \in P_3(K) : \quad b_K|_{\partial K} = 0, \quad |K|^{-1} \int_K b_K \, dx = 1.$$

Wir werden unten sehen, dass alle diese Stokes-Elemente gleichmäßig „inf-sup“-stabil sind. Dabei wird die Konstruktion des erweiterten Ansatzes  $P_2^b$  durch den Beweis der Stabilitätsabschätzung motiviert. Diese Konstruktion lässt sich auch für höhere Polynomgrade  $r \geq 3$  verallgemeinern. Das  $Q_2^c/P_1^{\text{dc}}$ -Element erweist sich in der Praxis als besonders gut geeignet hinsichtlich Komplexität und Genauigkeit.

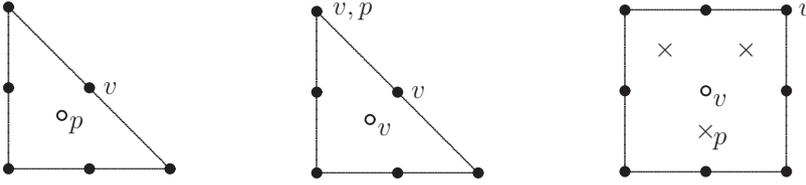


Abbildung 6.2: Die konformen Stokes-Elemente vom Typ  $P_2^c/P_0^{\text{dc}}$  (links),  $P_2^c \oplus \text{span}\{b_K^1, b_K^2\}/P_1^{\text{dc}}$  (mittig) und  $Q_2^c/P_1^{\text{dc}}$  (rechts).

**Lemma 6.4 („inf-sup“-Bedingung):** *Die konformen „Stokes-Elemente“ vom Typ  $P_2/P_0^{\text{dc}}$ ,  $P_2^b/P_1^{\text{dc}}$  und  $Q_2/P_1^{\text{dc}}$  genügen der gleichmäßigen „in-sup“-Bedingung (6.2.58).*

**Beweis:** i) Der Beweis basiert auf der kontinuierlichen inf-sup-Abschätzung (6.2.52) und den lokalen Eigenschaften der verwendeten Ansatzfunktionen. Die einzelnen Beweisschritte können dabei Hinweise zur Konstruktion von stabilen Stokes-Elementen geben. Sei  $q_h \in L_h$  beliebig gegeben. Aufgrund der kontinuierlichen inf-sup-Abschätzung gilt dann:

$$\beta \|q_h\| \leq \sup_{\varphi \in H} \frac{(q_h, \nabla \cdot \varphi)}{\|\nabla \varphi\|}. \quad (6.2.87)$$

Hiermit erhalten wir die Beziehung

$$\beta \|q_h\| \leq \sup_{\varphi \in H} \left\{ \frac{(q_h, \nabla \cdot \varphi_h)}{\|\nabla \varphi_h\|} \frac{\|\nabla \varphi_h\|}{\|\nabla \varphi\|} \right\} + \sup_{\varphi \in H} \frac{(q_h, \nabla \cdot (\varphi - \varphi_h))}{\|\nabla \varphi\|}, \quad (6.2.88)$$

für beliebiges  $\varphi_h \in H_h$ . Ein Weg zum Beweis der Behauptung ist nun die Konstruktion eines Interpolationsoperators  $r_h : H \rightarrow H_h$  mit den folgenden Eigenschaften:

$$(a) \quad (\chi_h, \nabla \cdot r_h v) = (\chi_h, \nabla \cdot v), \quad \chi_h \in L_h, \quad (6.2.89)$$

$$(b) \quad \|\nabla r_h v\| \leq c_1 \|\nabla v\|. \quad (6.2.90)$$

Damit folgt dann über (6.2.88):

$$\beta \|q_h\| \leq \sup_{\varphi \in H} \left\{ \frac{(q_h, \nabla \cdot r_h \varphi)}{\|\nabla r_h \varphi\|} \frac{\|\nabla r_h \varphi\|}{\|\nabla \varphi\|} \right\} \leq c_1 \sup_{\varphi_h \in H_h} \frac{(q_h, \nabla \cdot \varphi_h)}{\|\nabla \varphi_h\|},$$

Die diskrete inf-sup-Bedingung ist also erfüllt mit der Konstante  $\beta_* := \beta/c_1$ .

ii) Die Konstruktion des Interpolationsoperators  $r_h$  ist orientiert an den kritischen Eigenschaften (a) und (b) und natürlich an der Forderung  $r_h v \in H_h$ . Mit Hilfe von partieller Integration erhalten wir

$$(q_h, \nabla \cdot v) = \sum_{K \in \mathcal{T}_h} (q_h, \nabla \cdot v)_K = \sum_{K \in \mathcal{T}_h} \{ (q_h, n \cdot v)_{\partial K} - (\nabla q_h, v)_K \}.$$

Eine lokale Konstruktion des Operators  $r_h$  muss also die folgenden Bedingungen erfüllen:

$$(q_h, n \cdot v)_\Gamma = (q_h, n \cdot r_h v)_\Gamma, \quad \Gamma \in \partial \mathbb{T}_h, \quad (6.2.91)$$

$$(\nabla q_h, v)_K = (\nabla q_h, r_h v)_K, \quad K \in \mathbb{T}_h. \quad (6.2.92)$$

Diese Bedingungen sind für Funktionen  $v \in H$  stetig definiert. Im jeweiligen konkreten Fall müssen sie durch weitere Bedingungen ergänzt werden, um einen geeigneten Operator  $r_h$  zu erzeugen. Dazu gehört dann noch der Nachweis der Stabilitätseigenschaft (b). Wir wollen diese Konstruktion für die betrachteten Stokes-Elemente kurz skizzieren. Der detaillierte Nachweis der Unisolvenz, der Konformität sowie Stabilität sei als Übungsaufgabe gestellt. Für ein  $v \in H$  sei  $v_h \in H_h$  die globale Ritz-Projektion mit den Eigenschaften

$$(\nabla v_h, \nabla \varphi_h) = (\nabla v, \nabla \varphi_h) \quad \forall \varphi_h \in H_h, \quad \|\nabla v_h\| \leq \|\nabla v\|. \quad (6.2.93)$$

iii) Das  $P_2/P_0^{\text{dc}}$ -Element: Der zugehörige Interpolationsoperator  $r_h$  wird durch folgende lokale Bedingungen definiert (beachte  $\nabla q_h|_K \equiv 0$ ):

$$r_h v(a) = v_h(a), \quad a \in \partial^2 \mathbb{T}_h, \quad (\chi, r_h v)_\Gamma = (\chi, v)_\Gamma, \quad /; \chi \in P_0(\Gamma), \quad \Gamma \in \partial \mathbb{T}_h.$$

Für den Beweis für das  $P_2^b/P_1^{\text{dc}}$ -Element und das  $Q_2^c/P_1^{\text{dc}}$ -Element verweisen wir auf die einschlägige Literatur, z. B.: Girault/Raviart [110] und Matthies/Tobiska [122]. Q.E.D.

**Bemerkung 6.3:** Die Existenz eines (linearen) Operators  $r_h : H \rightarrow H_h$  mit den Eigenschaften (6.2.89)-(6.2.90) ist sogar notwendig für die Gültigkeit der *gleichmäßigen* diskreten inf-sup-Stabilität. Sei diese gegeben. Dann gilt auch (Hinweis: Man betrachte den durch  $(\chi_h, B_h v_h) := (\chi_h, \nabla \cdot v_h)$ , definierten Operator  $B_h : H_h \rightarrow L_h^*$  und seine Adjungierte  $B_h^* : L_h \rightarrow H_h^*$ .)

$$\inf_{v_h \in H_h} \sup_{\chi_h \in L_h} \frac{(\chi_h, \nabla \cdot v_h)}{\|\nabla v_h\| \|\chi_h\|} \geq \beta_*.$$

Für jedes  $v \in H$  existiert dann ein eindeutiges Element  $r_h v \in V_h^T$ , so dass

$$(\chi_h, \nabla \cdot r_h v) = (\chi_h, \nabla \cdot v) \quad \forall \chi_h \in L_h,$$

und

$$\beta_* \|\nabla r_h v\| \leq \sup_{\chi_h \in L_h} \frac{(\chi_h, \nabla \cdot r_h v)}{\|\chi_h\|} \leq \sup_{\chi_h \in L_h} \frac{(\chi_h, \nabla \cdot v)}{\|\chi_h\|} \leq c \|\nabla v\|.$$

Dadurch ist offensichtlich ein linearer Operator  $r_h : H \rightarrow H_h$  mit den gewünschten Eigenschaften definiert.

b) *Elemente mit stetigem Druck:*

Diese Elemente zeichnen sich durch eine geringe Zahl von Knotenfreiheitsgraden aus, besitzen aber nicht die lokale Masseerhaltungseigenschaft.

i) Das  $P_1^b/P_1^c$ - (MINI<sup>1</sup>)-Element (a), das  $P_2^c/P_1^c$ - (Taylor-Hood)-Element (b) und das entsprechende  $Q_2^c/Q_1^c$ -Element (c):

$$a) P_H(K) := P_1(K) \oplus \text{span}\{b_K^1, b_K^2\}, \quad P_L(K) := P_1^c(K);$$

$$b) P_H(K) := P_2^c(K), \quad P_L := P_1^c(K);$$

$$c) P_H(K) := Q_2^c(K), \quad P_L(K) := Q_1^c(K).$$

Dabei ist wieder  $b_K^1 = (b_K, 0)^T$  bzw.  $b_K^2 = (0, b_K)^T$  mit der oben definierten kubischen „bulb-Funktion“  $b_K$ .

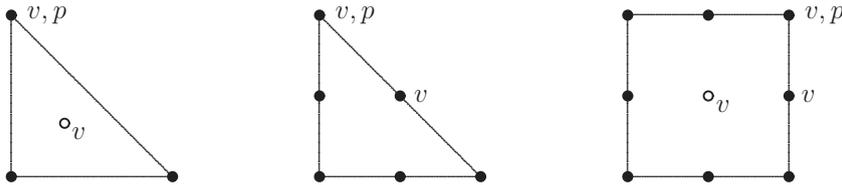


Abbildung 6.3: Die konformen Stokes-Elemente vom Typ  $P_1 \oplus \text{span}\{b_K^1, b_K^2\}/P_1^c$  (links),  $P_2^c/P_1^c$  (Mitte) und  $Q_2^c/Q_1^c$  (rechts).

Alle drei Stokes-Elemente sind inf-sup-stabil. Allgemein ist der stetige  $P_r^c/P_{r-2}^c$ -Ansatz (für  $r \geq 3$ ) inf-sup-stabil. Die Kombination  $P_r^c/P_{r-1}^c$  (verallgemeinerte Taylor-Hood-Elemente) ist nicht stabil und erfordert eine Anreicherung des Geschwindigkeitsansatzes (s. Brenner/Scott [78]).

**Lemma 6.5 („inf-sup“-Bedingung):** *Das „MINI-Element“ vom Typ  $P_1^b/P_1^c$  sowie die „Taylor-Hood-Elemente“ vom Typ  $P_2^c/P_1^c$  und  $Q_2^c/Q_1^c$  genügen der gleichmäßigen „inf-sup“-Bedingung (6.2.58).*

**Beweis:** Der Beweis verwendet ähnliche Argumente wie der von Lemma 6.2.2. Ausgangspunkt ist wieder die Beziehung

$$\beta \|q_h\| \leq \sup_{\varphi \in H} \left\{ \frac{(q_h, \nabla \cdot \varphi_h) \|\nabla \varphi_h\|}{\|\nabla \varphi_h\| \|\nabla \varphi\|} \right\} + \sup_{\varphi \in H} \frac{(q_h, \nabla \cdot (\varphi - \varphi_h))}{\|\nabla \varphi\|}, \quad (6.2.94)$$

für beliebige gewählte  $q_h \in L_h$  und  $\varphi_h \in H_h$ . Ziel ist dann wieder die Konstruktion eines Interpolationsoperators  $r_h : H \rightarrow H_h$  mit den Eigenschaften (a) und (b). Dies kann beim MINI-Element analog wie oben erfolgen. Wegen der Stetigkeit der Druckansätze ist

$$(q_h, \nabla \cdot \varphi) = -(\nabla q_h, \varphi),$$

<sup>1</sup>F. Brezzi, M. Fortin: *A minimal stabilization procedure for mixed finite element methods*, Numer. Math. 89, 457–491 (2001).

so dass für den Operator  $r_h$  neben der  $H$ -Stabilität nur die Eigenschaft

$$(\nabla q_h, \varphi) = (\nabla q_h, r_h \varphi)_K = 0, \quad K \in \mathbb{T}_h,$$

zu fordern ist. Hierfür stehen beim MINI-Element die zusätzlichen zwei Bulb-Freiheitsgrade pro Zelle  $K$  im Geschwindigkeitsansatz zur Verfügung. Der Stabilitätsbeweis für die Taylor-Hood-Elemente erfordert eine modifizierte Argumentation und ist wesentlich aufwendiger. Hierfür wird auf die o. a. Literatur verwiesen. Q.E.D.

## II) Beispiele von „nicht-konformen“ Stokes-Elementen

Wir haben gesehen, dass die Stabilitätsbedingung (6.2.58) die Konstruktion von Stokes-Elementen niedrigsten Polynomgrads verbietet; die  $P_1^c/P_0^{\text{dc}}$ - und  $Q_1^c/P_0^{\text{dc}}$ -Elemente sind nicht brauchbar. Die Erzeugung stabiler Ansätze erfordert hier die Verwendung von *nicht-konformen*, d. h. nicht vollständig stetigen, Geschwindigkeitsansätzen. Diese Elemente verwenden unstetige Druckansätze zur Gewährleistung der lokalen Masseerhaltung und Geschwindigkeitsansätze niedriger Ordnung.

Bei Verwendung von über die Zellkanten unstetigen Geschwindigkeitsansätzen müssen die Ableitungen  $\partial_i$  bzw. der Operator  $\nabla$  zellweise definiert werden:

$$(\nabla_h v_h)|_K := \nabla(v_h|_K), \quad K \in \mathbb{T}_h.$$

Mit dieser Notation ist dann ein Paar  $\{v_h, p_h\} \in H_h \times L_h$  gesucht mit

$$(\nabla_h v_h, \nabla_h \varphi_h) - (p_h, \nabla_h \cdot \varphi_h) = (f, \varphi_h) \quad \forall \varphi_h \in H_h, \quad (6.2.95)$$

$$(\nabla_h \cdot v_h, \chi_h) = 0 \quad \forall \chi_h \in L_h. \quad (6.2.96)$$

**Beispiele:** Das  $P_1^{\text{nc}}/P_0^{\text{dc}}$ - (Crouzeix<sup>2</sup>/Raviart<sup>3</sup>)- (a) und das  $Q_1^{\text{rot}}/P_0^{\text{dc}}$ -Element (b):

$$a) \quad P_H(K) := P_1^{\text{nc}}(K), \quad P_L(K) := P_0^{\text{dc}}(K);$$

$$b) \quad P_H(K) := \tilde{Q}_1^{\text{nc}}(K), \quad P_L := P_0^{\text{dc}}(K).$$

mit dem „rotiert-bilinearen“ Ansatz (auf dem Referenzelement)

$$\tilde{Q}_1(K) := \text{span}\{1, x_1, x_2, x_1^2 - x_2^2\}$$

---

<sup>2</sup>Michel Crouzeix (1928–): Französischer Mathematiker; Studium in Paris, Promotion 1975 bei Pierre-Arnaud Raviart; Prof. an der Univ. de Rennes 1; Arbeitsgebiet Numerik, speziell Matrix-Analysis, Numerische Lineare Algebra und Finite-Elemente-Methode.

<sup>3</sup>Pierre-Arnaud Raviart (1939–): Französischer Mathematiker; Prof. an der Univ. Paris 6; fundamentale Beiträge zur Methode der Finiten-Elemente und ihrer Anwendung (u. a. Interpolationsabschätzungen, FEM für Navier-Stokes-Gleichungen, „unstetige“ Galerkin-Verfahren - DG-Methode - für hyperbolische Erhaltungsgleichungen).

Diese Modifikation gegenüber dem üblichen *bilinearen* Ansatz  $\text{span}\{1, x_1, x_2, x_1 x_2\}$  ist erforderlich, da letzterer bzgl. der Kantenmittenwerte oder der Kantenmittelwerte nicht *unisolvant* ist. Dies erkennt man anhand des Polynoms  $p(x) := x_1 x_2$ , welches in den 4 Interpolationspunkten  $(\pm 1, \pm 1)$  Null ist. Der modifizierte Ansatz  $\tilde{Q}_1(K)$  geht aus  $Q_1(K)$  durch Koordinatenrotation um 90 Grad hervor (daher die Bezeichnung „rotiert-bilinear“ Ansatz) und ist folglich unisolvant bzgl. dieser Knotenwertsätze. Auf allgemeinen, nicht-kartesischen Gittern könnte man, dem allgemeinen Konstruktionsprinzip folgend, die lokalen Formfunktionen durch Transformation vom Referenzelement (Einheitsquadrat) erzeugen. In diesem Fall ist die Transformation im allgemeinen aber *bilinear*, und damit der kombinierte Ansatz nicht im strengen Sinne *isoparametrisch*. Dies führt auf stark verzerrten Gittern zu Konsistenzdefekten, welche durch einen *nicht parametrischen* Ansatz vermieden werden. Dabei wird die Formfunktion separat auf jeder Zelle bzgl. eines lokalen (fast) kartesischen Koordinatensystems  $(\xi, \eta)$  durch  $\text{span}\{1, \xi, \eta, \xi^2 - \eta^2\}$  definiert. Diese Stokes-Elemente haben natürliche 3-dimensionale Analoga, z. B.:

$$\tilde{Q}_1(K) := \text{span}\{1, x_1, x_2, x_3, x_1^2 - x_2^2, x_2^2 - x_3^2\}.$$

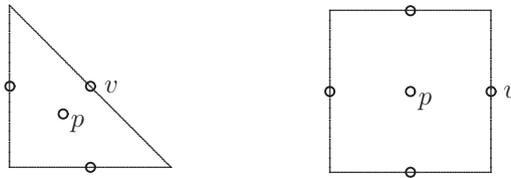


Abbildung 6.4: Die nichtkonformen Stokes-Elemente vom Typ  $P_1^{\text{nc}}/P_0$  (links) und  $\tilde{Q}_1^{\text{nc}}/P_0$  (rechts).

Zum Nachweis der Existenz von Lösungen des Systems (6.2.95), (6.2.96) im nicht-konformen Fall betrachten wir wieder eine Lösung  $\{v_h, p\} \in H_h \times L_h$  des zugehörigen homogenen Problems, d. h. für  $f \equiv 0$ . Für diese gilt dann notwendig

$$\|\nabla_h v_h\| = 0.$$

Dies impliziert  $\nabla v_{h|K} \equiv 0$  auf jeder einzelnen Zelle  $K \in \mathbb{T}_h$  und folglich aufgrund der Stetigkeitseigenschaften von  $v_h \in H_h$  und der Randbedingungen notwendig  $v_h \equiv 0$ .

**Lemma 6.6 („inf-sup“-Bedingung):** *Die nicht-konformen Stokes-Elemente vom Typ  $P_1^{\text{nc}}/P_0^{\text{nc}}$  und  $\tilde{Q}_1^{\text{nc}}/P_0^{\text{nc}}$  genügen der gleichmäßigen „inf-sup“-Bedingung:*

$$\sup_{\varphi_h \in H_h} \frac{(q_h, \nabla_h \cdot \varphi_h)}{\|\nabla_h \varphi_h\|} \geq \beta_* \|q_h\|, \quad q_h \in L_h. \quad (6.2.97)$$

**Beweis:** Wir werden wieder einen Interpolationsoperator  $r_h : H \rightarrow H_h$  konstruieren mit den folgenden Eigenschaften:

$$\|\nabla_h r_h v\| \leq c_1 \|\nabla v\|, \quad v \in H, \quad (6.2.98)$$

$$(q_h, \nabla_h \cdot r_h v) = (q_h, \nabla \cdot v), \quad v \in H, \quad q_h \in L_h. \quad (6.2.99)$$

Mit dieser Konstruktion können wir wieder wie folgt schließen:

$$\beta \|q_h\| \leq \sup_{\varphi \in H} \frac{(q_h, \nabla \cdot \varphi)}{\|\nabla \varphi\|} = \sup_{\varphi \in H} \left\{ \frac{(q_h, \nabla_h \cdot r_h \varphi)}{\|\nabla_h r_h \varphi\|} \frac{\|\nabla_h r_h \varphi\|}{\|\nabla \varphi\|} \right\} \leq c_1 \sup_{\varphi_h \in H_h} \frac{(q_h, \nabla_h \cdot \varphi_h)}{\|\nabla \varphi_h\|},$$

was die Stabilitätsabschätzung (6.2.97) impliziert mit der Konstante  $\beta_* := \beta c_1^{-1}$ . Wir kommen nun zur Konstruktion des Operators  $r_h$ . Durch die Vorschrift

$$\int_{\Gamma} r_h v \, ds = \int_{\Gamma} v \, ds, \quad \Gamma \in \partial \mathbb{T}_h,$$

wobei  $\partial \mathbb{T}_h$  wieder die Menge aller Zellkanten der Zerlegung  $\mathbb{T}_h$  bezeichnet, ist auf jeder Zelle  $K \in \mathbb{T}_h$  ein lokaler Interpolationsoperator  $r_h^K : H^1(K)^2 \rightarrow P_1(K)^2$  definiert. Für  $v \in H$  können die Stücke  $r_h^K v$  zu einer Funktion  $r_h v \in H_h$  zusammengefügt werden. Die (sogar zellweise)  $H^1$ -Stabilität dieses Interpolationsoperators ergibt sich unter Beachtung von  $\Delta r_h^K v|_K \equiv 0$  und  $\partial_n r_h^K v|_{\Gamma} \equiv \text{konst.}$  direkt aus seiner Konstruktion wie folgt:

$$\begin{aligned} \|\nabla r_h^K v\|_K^2 &= \int_{\partial K} \partial_n r_h^K v r_h^K v \, ds - (\Delta r_h^K v, r_h^K v)_K = \int_{\partial K} \partial_n r_h^K v v \, ds \\ &= (\nabla r_h^K v, \nabla v)_K + (\Delta r_h^K v, v)_K \leq \|\nabla r_h^K v\| \|\nabla v\|_K. \end{aligned}$$

Dies vervollständigt den Beweis. Q.E.D.

Der Interpolationsoperator  $r_h$  ist für den betrachteten, nicht-konformen Ansatzraum  $H_h$  gerade die natürliche Knoteninterpolierende  $r_h = i_h : H \rightarrow H_h$ . Für diese erhält man mit Hilfe des üblichen Transformationsarguments die Fehlerabschätzung

$$\|v - i_h v\| + h \|\nabla_h (v - i_h v)\| \leq c_i h^2 \|\nabla^2 v\|. \quad (6.2.100)$$

Unter der Annahme der Gültigkeit einer diskreten Variante der Poincaréschen Ungleichung (Übungsaufgabe: Der Beweis verwendet dieselben Argumente wie der des nachfolgenden Satzes 6.7.)

$$\|v_h\| \leq \gamma \|\nabla_h v_h\|, \quad v_h \in H_h, \quad (6.2.101)$$

folgt die Stabilität der diskreten Geschwindigkeiten. Dazu beachten wir, dass

$$\begin{aligned} \|\nabla_h v_h\|^2 - (p_h, \nabla_h \cdot v_h) &= (f, v_h), \\ (p_h, \nabla_h \cdot v_h) &= 0, \end{aligned}$$

und folglich

$$\|\nabla_h v_h\| \leq \|f\| \|v_h\| \|\nabla_h v_h\|^{-1} \leq \gamma \|f\|. \quad (6.2.102)$$

**Satz 6.7 (Fehlerabschätzungen):** Für die nicht-konformen Stokes-Elemente vom Typ  $P_1^{\text{nc}}/P_0^{\text{dc}}$  und  $\tilde{Q}_1^{\text{nc}}/P_0^{\text{dc}}$  (nicht-parametrisch) gelten die a priori Fehlerabschätzungen:

$$\|\nabla_h(v - v_h)\| + \|p - p_h\| \leq h\|f\|, \quad (6.2.103)$$

$$\|v - v_h\| \leq ch^2\|f\|. \quad (6.2.104)$$

**Beweis:** Wir führen den Beweis nur für das einfachere  $P_1^{\text{nc}}/P_0^{\text{dc}}$ -Element.

i) Sei wieder  $e_h := v - v_h$  und  $\eta_h := p - p_h$  gesetzt. Mit einem beliebigen  $\varphi_h \in H_h$  gilt

$$\|\nabla_h e_h\|^2 = (\nabla_h e_h, \nabla_h(v - \varphi_h)) + (\nabla_h e_h, \nabla_h(\varphi_h - v_h)),$$

sowie für beliebiges  $\psi_h \in H_h$ :

$$\begin{aligned} (\nabla_h e_h, \nabla_h \psi_h) &= (\nabla v, \nabla_h \psi_h) - (\nabla v_h, \nabla_h \psi_h) \\ &= (\nabla v, \nabla_h \psi_h) - (f, \psi_h) - (p_h, \nabla_h \cdot \psi_h) \\ &= (\nabla v, \nabla_h \psi_h) - (p, \nabla_h \cdot \psi_h) - (f, \psi_h) + (\eta_h, \nabla_h \cdot \psi_h). \end{aligned}$$

Ferner ist mit beliebigem  $\chi_h \in L_h$ :

$$\begin{aligned} (\eta_h, \nabla_h \cdot (\varphi_h - v_h)) &= (\eta_h, \nabla_h \cdot (\varphi_h - v)) + (\eta_h, \nabla_h \cdot e_h) \\ &= (\eta_h, \nabla_h \cdot (\varphi_h - v)) + (\eta_h, \nabla \cdot v) - (\eta_h, \nabla_h \cdot v_h) \\ &= (\eta_h, \nabla_h \cdot (\varphi_h - v)) + (p - \chi_h, \nabla_h \cdot e_h). \end{aligned}$$

Durch Kombination der vorausgehenden Beziehungen erhalten wir

$$\begin{aligned} \|\nabla_h e_h\|^2 &\leq \|\nabla_h e_h\| \|\nabla_h(v - \varphi_h)\| + \Delta_h(v, p) \{ \|\nabla_h(\varphi_h - v)\| + \|\nabla_h e_h\| \} \\ &\quad + \|\eta_h\| \|\nabla(v - \varphi_h)\| + \|p - \chi_h\| \|\nabla_h e_h\|, \end{aligned}$$

mit der Abkürzung

$$\Delta_h(v, p) := \max_{\psi_h \in H_h} \frac{(\nabla v, \nabla_h \psi_h) - (p, \nabla_h \cdot \psi_h) - (f, \psi_h)}{\|\nabla_h \psi_h\|}.$$

Unter Verwendung der Abschätzung  $ab \leq \frac{1}{4}\varepsilon a^2 + \varepsilon^{-1}b^2$ , für  $a, b \in \mathbb{R}_+$  und beliebiges  $\varepsilon > 0$ , erschließen wir weiter

$$\begin{aligned} \|\nabla_h e_h\|^2 &\leq \frac{3}{4}\varepsilon_1 \|\nabla_h e_h\|^2 + \varepsilon_1^{-1} \{ \|\nabla_h(v - \varphi_h)\|^2 + \|p - \chi_h\|^2 + \Delta_h(v, p)^2 \} \\ &\quad + \frac{1}{4}\varepsilon_2 \|\eta_h\|^2 + \{ \varepsilon_2^{-1} + \frac{1}{2} \} \Delta_h(v, p)^2 + \frac{1}{2} \|\nabla_h(v - \varphi_h)\|^2, \end{aligned}$$

und bei Setzung  $\varepsilon_1 = 1$ :

$$\begin{aligned} \|\nabla_h e_h\|^2 &\leq 6 \min_{\varphi_h \in H_h} \|\nabla_h(v - \varphi_h)\|^2 + 4 \min_{\chi_h \in L_h} \|p - \chi_h\|^2 \\ &\quad + \{ \varepsilon_2^{-1} + 6 \} \Delta_h(v, p)^2 + \varepsilon_2 \|\eta_h\|^2. \end{aligned} \quad (6.2.105)$$

Dieses Zwischenresultat wollen wir uns merken.

ii) Zur Abschätzung des Druckfehlers  $\|\eta_h\|$  bedienen wir uns der inf-sup-Ungleichung (6.2.97). Mit einem beliebigen  $\chi_h \in L_h$  ergibt sich dann

$$\begin{aligned} \|\eta_h\| &\leq \|p - \chi_h\| + \|\chi_h - p_h\| \\ &\leq \|p - \chi_h\| + \beta^{-1} \max_{\varphi_h \in H_h} \frac{(\chi_h - p_h, \nabla_h \cdot \varphi_h)}{\|\nabla_h \varphi_h\|} \\ &\leq \|p - \chi_h\| + \beta^{-1} \max_{\varphi_h \in H_h} \frac{(\chi_h - p, \nabla_h \cdot \varphi_h)}{\|\nabla_h \varphi_h\|} + \beta^{-1} \max_{\varphi_h \in H_h} \frac{(\eta_h, \nabla_h \cdot \varphi_h)}{\|\nabla_h \varphi_h\|} \\ &\leq (\beta^{-1} + 1) \|p - \chi_h\| + \beta^{-1} \max_{\varphi_h \in H_h} \frac{(\eta_h, \nabla_h \cdot \varphi_h)}{\|\nabla_h \varphi_h\|}. \end{aligned}$$

Weiter gilt

$$\begin{aligned} (\eta_h, \nabla_h \cdot \varphi_h) &= (p, \nabla_h \cdot \varphi_h) - (p_h, \nabla_h \cdot \varphi_h) \\ &= (p, \nabla_h \cdot \varphi_h) - (\nabla v_h, \nabla_h \varphi_h) + (f, \varphi_h) \\ &= (p, \nabla_h \cdot \varphi_h) - (\nabla v, \nabla_h \varphi_h) + (f, \varphi_h) - (\nabla e_h, \nabla_h \varphi_h) \end{aligned}$$

und folglich

$$\max_{\varphi_h \in H_h} \frac{(\eta_h, \nabla_h \cdot \varphi_h)}{\|\nabla_h \varphi_h\|} \leq \Delta_h(v, p) + \|\nabla_h e_h\|.$$

Dies ergibt

$$\|\eta_h\| \leq (\beta^{-1} + 1) \|p - \chi_h\| + \beta^{-1} \{ \Delta_h(v, p) + \|\nabla_h e_h\| \}. \quad (6.2.106)$$

Durch Kombination der Zwischenresultate (6.2.105) und (6.2.106) und Wahl von  $\varepsilon_2 := \frac{1}{2}\beta$  erhalten wir schließlich mit einer generischen Konstante  $c(\beta) \approx \beta^{-1} + 1 > 0$ :

$$\|\nabla_h e_h\| + \|\eta_h\| \leq c(\beta) \left\{ \min_{\varphi_h \in H_h} \|\nabla_h(v - \varphi_h)\| + \min_{\chi \in L_h} \|p - \chi_h\| + \Delta_h(v, p) \right\}. \quad (6.2.107)$$

iii) Es bleibt, den Nichtkonformitätsterm  $\Delta_h(v, p)$  abzuschätzen. Dazu formen wir unter Beachtung von  $-\Delta v + \nabla p = f$  wie folgt um:

$$\begin{aligned} (\nabla v, \nabla_h \psi_h) - (p, \nabla_h \cdot \psi_h) - (f, \psi_h) &= \sum_{K \in \mathbb{T}_h} \{ (-\Delta v + \nabla p - f, \psi_h)_K + (\partial_n v - pn, \psi_h)_{\partial K} \} \\ &= \sum_{K \in \mathbb{T}_h} (\partial_n v - pn, \psi_h)_{\partial K} = \sum_{\Gamma \in \partial \mathbb{T}_h} (\partial_n v - pn, \psi_h)_\Gamma, \end{aligned}$$

wobei  $\partial \mathbb{T}_h$  wieder die Menge aller Kanten  $\Gamma$  der Zellen  $K \in \mathbb{T}_h$  bezeichnet. Jede Kante  $\Gamma \in \partial \mathbb{T}_h$  ist entweder gemeinsame Kante zweier Zellen  $K, K' \in \mathbb{T}_h$  oder Teil von  $\partial \Omega$ . Die Funktion  $\partial_n v - pn$  auf  $\Gamma \subset K$  hat wegen der Stetigkeit von  $\nabla v$  und  $p$  entgegengesetztes Vorzeichen zu  $\partial_n v - pn$  auf  $\Gamma \subset K'$ . Daher können wir schreiben

$$\sum_{\Gamma \in \partial \mathbb{T}_h} (\partial_n v - pn, \psi_h)_\Gamma = \sum_{\Gamma \in \partial \mathbb{T}_h} (\partial_n v - pn, [\psi_h])_\Gamma$$

mit der Bezeichnung

$$[\psi_h]_\Gamma := \begin{cases} \frac{1}{2}(\psi_{h|\Gamma \cap K} - \psi_{h|\Gamma \cap K'}), & \Gamma = K \cap K', \\ \psi_{h|\Gamma}, & \Gamma \subset \partial\Omega. \end{cases}$$

Auf einer Kante  $\Gamma \in \partial\mathbb{T}_h$  hat  $[\psi_h]$  aufgrund der speziellen Eigenschaften von  $\psi_h \in H_h$  Mittelwert Null. Folglich ist

$$(\partial_n v - pn, [\psi_h])_\Gamma = (\partial_n v - pn, [\psi_h] - \overline{[\psi_h]})_\Gamma = (\partial_n v - pn - \overline{(\partial_n v - pn)}_\Gamma, [\psi_h] - \overline{[\psi_h]})_\Gamma$$

mit den Mittelwerten

$$\overline{(\partial_n v - pn)}_\Gamma := |\Gamma|^{-1} \int_\Gamma (\partial_n v - pn) ds, \quad \overline{[\psi_h]}_\Gamma := |\Gamma|^{-1} \int_\Gamma [\psi_h] ds.$$

Aufspaltung von  $[\psi_h]$  in die Beiträge der beteiligten Zellen  $K, K'$  ergibt dann

$$(\partial_n v - pn, [\psi_h])_\Gamma = (\partial_n v - pn - \overline{(\partial_n v - pn)}_\Gamma, \psi_{h|K} - \overline{\psi_{h\Gamma}})_\Gamma - \overline{(\partial_n v - pn)}_\Gamma (\psi_{h|K'} - \overline{\psi_{h\Gamma}})_\Gamma.$$

Mit Hilfe der üblichen Transformationstechnik zeigt man die folgenden Fehlerabschätzungen für die Mittelwertapproximation:

$$\begin{aligned} \|\partial_n v - pn - \overline{(\partial_n v - pn)}_\Gamma\|_\Gamma &\leq c_i h^{1/2} \{\|\nabla^2 v\|_K + \|\nabla p\|\}, \\ \|\psi_{h|K} - \overline{\psi_{h\Gamma}}\|_\Gamma &\leq c_i h^{1/2} \|\nabla \psi_h\|_K, \end{aligned}$$

und entsprechend für die Zelle  $K'$ . Unter Verwendung der vorausgehenden Ergebnisse erhalten wir nun für  $\Gamma = K \cap K'$ :

$$\begin{aligned} |(\partial_n v - pn, [\psi_h])_\Gamma| &\leq \|\partial_n v - pn - \overline{(\partial_n v - pn)}_\Gamma\|_\Gamma \{ \|\psi_{h|K} - \overline{\psi_{h\Gamma}}\|_\Gamma + \|\psi_{h|K'} - \overline{\psi_{h\Gamma}}\|_\Gamma \} \\ &\leq c^2 h \{ \|\nabla^2 v\|_K + \|\nabla p\|_K \} \|\nabla_h \psi_h\|_{K \cup K'}, \end{aligned}$$

und entsprechend für  $\Gamma \subset \partial\Omega$ . Hiermit erschließen wir

$$\left| \sum_{\Gamma \in \partial\mathbb{T}_h} (\partial_n v - pn, \psi_h)_\Gamma \right| \leq ch \{ \|\nabla^2 v\| + \|\nabla p\| \} \|\nabla_h \psi_h\|,$$

und folglich, mit einer von  $\beta$  unabhängigen Konstante  $c > 0$ ,

$$\Delta_h(v, p) \leq ch \{ \|\nabla^2 v\| + \|\nabla p\| \} \leq ch \|f\|. \quad (6.2.108)$$

Zusammen mit den bekannten Approximierbarkeitsabschätzungen

$$\min_{\varphi_h \in H_h} \|\nabla_h(v - \varphi_h)\| + \min_{\chi_h \in L_h} \|p - \chi_h\| \leq ch \{ \|\nabla^2 v\| + \|\nabla p\| \} \leq ch \|f\|$$

erhalten wir somit schließlich die Fehlerabschätzung (6.2.103).

iv) Zum Nachweis der  $L^2$ -Fehlerabschätzung (6.2.104) verwenden wir wieder ein Dualitätsargument. Die Details werden hier nicht ausgeführt (siehe die entsprechende Argumentation bei dem nichtkonformen Morley-Element für das Plattenproblem). Q.E.D.

### 6.2.3 Stabilisierte Stokes-Elemente

Es gibt Gründe, möglichst einfache, knoten-orientierte Elementansätze zur Diskretisierung der Stokes-Gleichungen zu verwenden, wobei die Freiheitsgrade für Geschwindigkeit und Druck in denselben Knotenpunkten „leben“. Einfachste Beispiel sind die konformen  $P_1^c/P_1^c$ - und  $Q_1^c/Q_1^c$ -Stokes-Element, welches leider nicht inf-sup-stabil sind. Wir wollen im folgenden eine Technik zur Stabilisierung solcher Ansätze diskutieren.

Wir rekapitulieren aus dem Beweis der Stabilitätseigenschaft für konforme Stokes-Elemente die folgende Beziehung für  $q_h \in L_h$ :

$$\beta \|q_h\| \leq \sup_{\varphi \in H} \left\{ \frac{(q_h, \nabla \cdot \varphi_h) \|\nabla \varphi_h\|}{\|\nabla \varphi_h\| \|\nabla \varphi\|} \right\} + \sup_{\varphi \in H} \frac{(q_h, \nabla \cdot (\varphi - \varphi_h))}{\|\nabla \varphi\|}. \quad (6.2.109)$$

Dabei ist  $\varphi_h := i_h^* \varphi \in H_h$  eine  $H^1$ -stabile (modifizierte) Knoteninterpolierende, für die auf jeder Zelle  $K \in \mathbb{T}_h$  gilt:

$$\|\varphi - i_h^* \varphi\|_K + h_K \|\nabla(\varphi - i_h^* \varphi)\|_K \leq \tilde{c}_i h_K \|\nabla \varphi\|_{\tilde{K}},$$

mit der Vereinigungsmenge  $\tilde{K}$  aller Zellen, die mit  $K$  einen nicht-leeren Schnitt haben. Ferner gilt wie für die übliche Knoteninterpolierende für  $\varphi \in H^2(K)$ :

$$\|\varphi - i_h^* \varphi\|_K + h_K \|\nabla(\varphi - i_h^* \varphi)\|_K \leq c_i h_K^2 \|\nabla^2 \varphi\|_K.$$

Für den zweiten Term rechts in (6.2.109) erhalten wir damit nach partieller Integration (beachte  $\varphi - i_h^* \varphi \in H$ ):

$$\begin{aligned} |(q_h, \nabla \cdot (\varphi - i_h^* \varphi))| &= |(\nabla q_h, \varphi - i_h^* \varphi)| \leq \sum_{K \in \mathbb{T}_h} \|\nabla q_h\|_K \|\varphi - i_h^* \varphi\|_K \\ &\leq c_i \sum_{K \in \mathbb{T}_h} h_K \|\nabla q_h\|_K \|\nabla \varphi\|_{\tilde{K}} \leq c_i \left( \sum_{K \in \mathbb{T}_h} h_K^2 \|\nabla q_h\|_K^2 \right)^{1/2} \left( \sum_{K \in \mathbb{T}_h} \|\nabla \varphi\|_{\tilde{K}}^2 \right)^{1/2} \\ &\leq c_* c_i \left( \sum_{K \in \mathbb{T}_h} h_K^2 \|\nabla q_h\|_K^2 \right)^{1/2} \|\nabla \varphi\|, \end{aligned}$$

mit einer für reguläre Triangulierungsfamilien  $(\mathbb{T}_h)_{h>0}$  fest wählbaren Konstante  $c_* \geq 1$ . Für den ersten Term in (6.2.109) haben wir wieder

$$\sup_{\varphi \in H} \left\{ \frac{(q_h, \nabla \cdot i_h^* \varphi) \|\nabla i_h^* \varphi\|}{\|\nabla i_h^* \varphi\| \|\nabla \varphi\|} \right\} \leq c_i \sup_{\varphi_h \in H_h} \frac{(q_h, \nabla \cdot \varphi_h)}{\|\nabla \varphi_h\|}.$$

Zusammengenommen ergeben die beiden letzten Abschätzungen

$$\beta_h^2 \|q_h\|^2 \leq \sup_{\varphi_h \in H_h} \frac{(q_h, \nabla \cdot \varphi_h)^2}{\|\nabla \varphi_h\|^2} + \sum_{K \in \mathbb{T}_h} h_K^2 \|\nabla q_h\|_K^2, \quad (6.2.110)$$

mit der Konstante  $\beta_h := (c_* c_i)^{-1} \beta$ . Dieses Resultat legt die folgende Modifikation des Approximationsschemas (6.2.67 - 6.2.68) nahe:

$$(\nabla v_h, \nabla \varphi_h) - (p_h, \nabla \cdot \varphi_h) = (f, \varphi_h) \quad \forall \varphi_h \in H_h, \quad (6.2.111)$$

$$(\chi_h, \nabla \cdot v_h) + s_h(\chi_h, p_h) = 0 \quad \forall \chi_h \in L_h, \quad (6.2.112)$$

mit der „Stabilisierungsform“

$$s_h(\chi_h, p_h) := \alpha \sum_{K \in \mathbb{T}_h} h_K^2 (\nabla \chi_h, \nabla p_h)_K,$$

wobei  $\alpha > 0$  eine (geeignet gewählte) Konstante ist.

**Lemma 6.7 (Stabilität):** *Das modifizierte Approximationsschema (6.2.111) - (6.2.112) besitzt eine eindeutige Lösung  $\{v_h, p_h\} \in H_h \times L_h$ , und es gilt die Stabilitätsabschätzung*

$$\|\nabla v_h\| + \|p_h\| + s_h(p_h, p_h)^{1/2} \leq c\|f\|. \quad (6.2.113)$$

**Beweis:** Zum Nachweis der Existenz von Lösungen genügt es wieder, die Eindeutigkeit derselben zu zeigen. Dazu setzen wir in (6.2.111)  $\varphi_h := v_h$  und in (6.2.112)  $\chi_h := p_h$  und erhalten

$$\begin{aligned} \|\nabla v_h\|^2 - (p_h, \nabla \cdot v_h) &= (f, v_h), \\ (p_h, \nabla \cdot v_h) + s_h(p_h, p_h) &= 0. \end{aligned}$$

Addition der Gleichungen ergibt

$$\|\nabla v_h\|^2 + s_h(p_h, p_h) = (f, v_h).$$

Im Fall homogener Daten, d. h.  $f \equiv 0$ , folgt hieraus  $v_h \equiv 0$  sowie  $\nabla p_h \equiv 0$ , d. h. die Eindeutigkeit einer Lösung (und damit auch deren Existenz). Weiter ergibt sich über die Beziehung

$$(f, v_h) \leq \|f\| \|v_h\| \leq c\|f\| \|\nabla v_h\|$$

die Stabilitätsabschätzung

$$\|\nabla v_h\|^2 + s_h(p_h, p_h) \leq c\|f\|^2.$$

Mit Hilfe der Beziehung (6.2.110) gewinnen wir zusätzlich die Stabilitätsabschätzung

$$\|p_h\| \leq c\|f\|,$$

was den Beweis vervollständigt.

Q.E.D.

**Satz 6.8 (Stabilisierte Stokes-Elemente):** *Für die Lösung  $\{v_h, p_h\} \in H_h \times L_h$  des modifizierten Approximationsschemas (6.2.111) - (6.2.112) gelten die a priori Fehlerabschätzungen*

$$\|\nabla(v - v_h)\| + \|p - p_h\| + s_h(p - p_h, p - p_h)^{1/2} \leq ch\|f\|, \quad (6.2.114)$$

$$\|v - v_h\| \leq ch^2\|f\|. \quad (6.2.115)$$

**Beweis:** Subtraktion der diskreten Stokes-Gleichungen von den kontinuierlichen ergibt

$$(\nabla(v - v_h), \nabla\varphi_h) - (p - p_h, \nabla \cdot \varphi_h) = 0, \quad \varphi_h \in H_h. \quad (6.2.116)$$

$$(\chi_h, \nabla \cdot (v - v_h)) + s_h(\chi_h, p - p_h) = s_h(\chi_h, p). \quad (6.2.117)$$

i) Wir beginnen mit der Abschätzung von  $\|\nabla(v - v_h)\|$  und  $s_h(p - p_h, p - p_h)$ . Mit beliebigem  $\varphi_h \in H_h$  folgt wegen (6.2.116)

$$\begin{aligned} \|\nabla(v - v_h)\|^2 &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (\nabla(v - v_h), \nabla(\varphi_h - v_h))) \\ &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - p_h, \nabla \cdot (\varphi_h - v_h))) \\ &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - p_h, \nabla \cdot (\varphi_h - v)) + (p - p_h, \nabla \cdot (v - v_h))), \end{aligned}$$

und weiter wegen (6.2.117) mit beliebigem  $\chi_h \in L_h$

$$\begin{aligned} \|\nabla(v - v_h)\|^2 &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - p_h, \nabla \cdot (\varphi_h - v)) + (p - \chi_h, \nabla \cdot (v - v_h))) \\ &\quad + (\chi_h - p_h, \nabla \cdot (v - v_h)) \\ &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - p_h, \nabla \cdot (\varphi_h - v)) + (p - \chi_h, \nabla \cdot (v - v_h))) \\ &\quad + (\chi_h - p_h, \nabla \cdot (v - v_h)) \\ &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - p_h, \nabla \cdot (\varphi_h - v)) + (p - \chi_h, \nabla \cdot (v - v_h))) \\ &\quad - s_h(\chi_h - p_h, p - p_h) + s_h(\chi_h - p_h, p) \\ &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - p_h, \nabla \cdot (\varphi_h - v)) + (p - \chi_h, \nabla \cdot (v - v_h))) \\ &\quad - s_h(\chi_h - p, p - p_h) - s_h(p - p_h, p - p_h) + s_h(\chi_h - p_h, p). \end{aligned}$$

Also gilt

$$\begin{aligned} \|\nabla(v - v_h)\|^2 + s_h(p - p_h, p - p_h) &= (\nabla(v - v_h), \nabla(v - \varphi_h) + (p - p_h, \nabla \cdot (\varphi_h - v))) \\ &\quad + (p - \chi_h, \nabla \cdot (v - v_h)) - s_h(\chi_h - p, p - p_h) + s_h(\chi_h - p_h, p) \\ &\leq \|\nabla(v - v_h)\| \|\nabla(v - \varphi_h)\| + \|p - p_h\| \|\nabla \cdot (\varphi_h - v)\| \\ &\quad + \|p - \chi_h\| \|\nabla \cdot (v - v_h)\| + s_h(\chi_h - p, \chi_h - p)^{1/2} s_h(p - p_h, p - p_h)^{1/2} \\ &\quad + s_h(\chi_h - p_h, \chi_h - p_h)^{1/2} s_h(p, p)^{1/2} \end{aligned}$$

Unter Beachtung von  $\|\nabla p\| \leq c\|f\|$  folgt

$$s_h(p, p)^{1/2} = \left( \alpha \sum_{K \in \mathbb{T}_h} h_K^2 \|\nabla p\|_K^2 \right)^{1/2} \leq ch \left( \sum_{K \in \mathbb{T}_h} \|\nabla p\|_K^2 \right)^{1/2} = ch \|\nabla p\| \leq ch \|f\|. \quad (6.2.118)$$

Ferner ergibt sich mit Hilfe der inversen Beziehung von finiten Elementen

$$\begin{aligned} s_h(\chi_h - p_h, \chi_h - p_h)^{1/2} &= \left( \alpha \sum_{K \in \mathbb{T}_h} h_K^2 \|\nabla(\chi_h - p_h)\|_K^2 \right)^{1/2} \leq c \left( \sum_{K \in \mathbb{T}_h} \|\chi_h - p_h\|_K^2 \right)^{1/2} \\ &\leq c\{\|p - p_h\| + \|p - \chi_h\|\}. \end{aligned}$$

Damit erschließen wir wieder mit Hilfe der Youngschen Ungleichung  $ab \leq \varepsilon^2 a^2 + (4\varepsilon^2)^{-1} b^2$  :

$$\begin{aligned} \|\nabla(v - v_h)\|^2 + s_h(p - p_h, p - p_h) &\leq c(1 + \varepsilon^{-1})\{\|\nabla(v - \varphi_h)\|^2 + \|p - \chi_h\|^2 \\ &\quad + s_h(p - \chi_h, p - \chi_h) + h^2\|f\|^2\} + \varepsilon\|p - p_h\|^2. \end{aligned} \quad (6.2.119)$$

ii) Als nächstes wird  $\|p - p_h\|$  abgeschätzt. Mit Hilfe der Stabilitätsungleichung (6.2.110) und der Galerkin-Orthogonalität (6.2.116) folgt

$$\begin{aligned} \|p - p_h\| &\leq \|p - \chi_h\| + \|\chi_h - p_h\| \\ &\leq \|p - \chi_h\| + \beta_*^{-1} \sup_{\psi_h \in H_h} \frac{(\chi_h - p_h, \nabla \cdot \psi_h)}{\|\nabla \psi_h\|} + \beta_*^{-1} s_h(\chi_h - p_h, \chi_h - p_h)^{1/2} \\ &\leq \|p - \chi_h\| + \beta^{-1} \sup_{\psi_h \in H_h} \frac{(\chi_h - p, \nabla \cdot \psi_h)}{\|\nabla \psi_h\|} + \beta^{-1} \sup_{\psi_h \in H_h} \frac{(p - p_h, \nabla \cdot \psi_h)}{\|\nabla \psi_h\|} \\ &\quad + \beta_*^{-1} s_h(\chi_h - p_h, \chi_h - p_h)^{1/2} \\ &= \|p - \chi_h\| + \beta_*^{-1} \sup_{\psi_h \in H_h} \frac{(\chi_h - p, \nabla \cdot \psi_h)}{\|\nabla \psi_h\|} + \beta_*^{-1} \sup_{\psi_h \in H_h} \frac{(\nabla(v - v_h), \nabla \psi_h)}{\|\nabla \psi_h\|} \\ &\quad + \beta_*^{-1} s_h(\chi_h - p_h, \chi_h - p_h)^{1/2} \end{aligned}$$

und damit

$$\begin{aligned} \|p - p_h\| &\leq (1 + \beta_*^{-1})\|p - \chi_h\| + \beta_*^{-1}\|\nabla(v - v_h)\| \\ &\quad + \beta_*^{-1} s_h(\chi_h - p_h, \chi_h - p_h)^{1/2}. \end{aligned} \quad (6.2.120)$$

Durch Kombination dieser Abschätzung mit (6.2.119) erhalten wir

$$\begin{aligned} \|\nabla(v - v_h)\|^2 + s_h(p - p_h, p - p_h) &\leq c(1 + \varepsilon^{-1})\{\|\nabla(v - \varphi_h)\|^2 + \|p - \chi_h\|^2 \\ &\quad + s_h(p - \chi_h, p - \chi_h) + h^2\|f\|^2\} \\ &\quad + c\varepsilon\{(1 + \beta_*^{-1})^2\|p - \chi_h\|^2 + \beta_*^{-2}\|\nabla(v - v_h)\|^2 + \beta_*^{-2}s_h(\chi_h - p_h, \chi_h - p_h)\}. \end{aligned}$$

Bei hinreichend kleiner aber fester Wahl von  $\varepsilon > 0$  folgt

$$\|\nabla(v - v_h)\| + s_h(p - p_h, p - p_h)^{1/2} \leq c\{\|\nabla(v - \varphi_h)\| + \|p - \chi_h\| + s_h(p - \chi_h, p - \chi_h)^{1/2} + h\|f\|\}.$$

Für  $\varphi_h = i_h^* v$  und  $\chi_h = i_h^* p$  ergibt sich Mit Hilfe der obigen Interpolationsabschätzungen

$$\|\nabla(v - v_h)\| + s_h(p - p_h, p - p_h)^{1/2} \leq ch\{\|\nabla^2 v\| + \|\nabla p\| + \|f\|\} \leq ch\|f\|.$$

Damit folgt aus (6.2.120) auch

$$\|p - p_h\| \leq ch\|f\|.$$

iii) Zur Abschätzung von  $\|v - v_h\|$  verwenden wir wieder ein Dualitätsargument. Sei  $\{z, q\} \in H \times L$  die Lösung des „dualen“ Stokes-Problems

$$(\nabla \varphi, \nabla z) - (q, \nabla \cdot \varphi) = (\varphi, v - v_h)\|v - v_h\|^{-1} \quad \forall \varphi \in H, \quad (6.2.121)$$

$$(\chi, \nabla \cdot z) = 0 \quad \forall \chi \in L. \quad (6.2.122)$$

Diese ist in  $H^2(\Omega)^d \times H^1(\Omega)$  und es gilt die a priori Abschätzung

$$\|z\|_{H^2} + \|q\|_{H^1} \leq c\|v - v_h\| \|v - v_h\|^{-1} = c. \quad (6.2.123)$$

Mit der Testfunction  $\varphi := v - v_h$  ergibt sich dann unter Verwendung der Galerkin-Orthogonalität mit der Knoteninterpolierenden  $i_h^* z \in H_h$  und  $i_h^* q \in L_h$ :

$$\begin{aligned} \|v - v_h\| &= (\nabla(v - v_h), \nabla z) - (q, \nabla \cdot (v - v_h)) \\ &= (\nabla(v - v_h), \nabla(z - i_h^* z)) + (\nabla(v - v_h), \nabla i_h^* z) - (q - i_h^* q, \nabla \cdot (v - v_h)) \\ &\quad - (i_h^* q, \nabla \cdot (v - v_h)) \\ &= (\nabla(v - v_h), \nabla(z - i_h^* z)) + (p - p_h, \nabla \cdot i_h^* z) - (q - i_h^* q, \nabla \cdot (v - v_h)) \\ &\quad + s_h(i_h^* q, p - p_h) - s_h(i_h^* q, p) \\ &= (\nabla(v - v_h), \nabla(z - i_h^* z)) + (p - p_h, \nabla \cdot (i_h^* z - z)) - (q - i_h^* q, \nabla \cdot (v - v_h)) \\ &\quad + s_h(i_h^* q, p - p_h) - s_h(i_h^* q, p). \end{aligned}$$

Mit Hilfe der Fehlerabschätzungen für die Interpolierenden schätzen wir dies weiter ab:

$$\begin{aligned} \|v - v_h\| &\leq \|\nabla(v - v_h)\| \|\nabla(z - i_h^* z)\| + \|p - p_h\| \|\nabla \cdot (i_h^* z - z)\| + \|q - i_h^* q\| \|\nabla \cdot (v - v_h)\| \\ &\quad + s_h(i_h^* q, i_h^* q)^{1/2} s_h(p - p_h, p - p_h)^{1/2} + s_h(i_h^* q, i_h^* q)^{1/2} s_h(p, p)^{1/2} \\ &\leq ch \|\nabla(v - v_h)\| \|z\|_{H^2} + ch \|p - p_h\| \|z\|_{H^2} + ch \|q\|_{H^1} \|\nabla(v - v_h)\| \\ &\quad + ch \|\nabla q\| s_h(p - p_h, p - p_h)^{1/2} + ch^2 \|\nabla q\| \|\nabla p\| \\ &\leq ch \{ \|\nabla(v - v_h)\| + \|p - p_h\| + s_h(p - p_h, p - p_h)^{1/2} \} + ch^2 \|f\|. \end{aligned}$$

Dies vervollständigt den Beweis.

Q.E.D.

Die Konvergenzabschätzungen (6.2.114) und (6.2.115) sind optimal bzgl. der  $h$ -Ordnung sowie der Regularitätsanforderungen an die Lösung  $\{v, p\}$ . Dasselbe Stabilisierungskonzept kann auch für die  $P_r^c/P_r^c$ - und  $Q_r^c/Q_r^c$ -Stokes-Elemente des Polynomgrades  $r \geq 2$  angewendet werden. In diesem Fall wird die Approximationsordnung aber durch die Stabilisierung in (6.2.112) zu  $\mathcal{O}(h^2)$  limitiert. Diese Ordnungsbarriere kann durch einen modifizierten, voll-konsistenten Stabilisierungsansatz überwunden werden:

$$(\nabla v_h, \nabla \varphi_h) - (p_h, \nabla \cdot \varphi_h) = (f, \varphi_h) \quad \forall \varphi_h \in H_h, \quad (6.2.124)$$

$$(\chi_h, \nabla \cdot v_h) + s_h(\chi_h, \nabla p_h) = g_h(\chi_h) \quad \forall \chi_h \in L_h, \quad (6.2.125)$$

wobei

$$g_h(\chi_h) := \alpha \sum_{K \in \mathbb{T}_h} h_K^2 \{ (\nabla \chi_h, f)_K + (\nabla \chi_h, \Delta v_h)_K \}.$$

In diesem Fall wird das diskrete System von der kontinuierlichen Lösung  $\{v, p\}$  exakt erfüllt, so dass volle „Galerkin-Orthogonalität“ besteht.

Das stabilisierte  $P_1^c/P_1^c$ -Stokes-Element hat eine enge Beziehung zum inf-sup-stabilen MINI-Element. Um diese zu verstehen, zerlegen wir den Geschwindigkeitsansatzraum des MINI-Elements in einen rein „bilinearen“ und einen „bulb“-Anteil:

$$v_h = v_h^{(1)} + v_h^b \in H_h^{(1)} \oplus H_h^b.$$

Dies ist wegen

$$(\nabla v_h^b, \nabla v_h^{(1)}) = \sum_{K \in \mathbb{T}_h} \{(v_h^b, \partial_n v_h^{(1)})_{\partial K} - (v_h^b, \Delta v_h^{(1)})_K\} = 0$$

tatsächlich eine  $H$ -orthogonale Aufspaltung. Das zugehörige diskrete Stokes-System lässt sich damit in folgender äquivalenten Form schreiben:

$$(\nabla v_h^{(1)}, \nabla \varphi_h) - (p_h, \nabla \cdot \varphi_h) = (f, \varphi_h) \quad \forall \varphi_h \in H_h^{(1)}, \quad (6.2.126)$$

$$(\nabla v_h^b, \nabla \varphi_h^b) - (p_h, \nabla \cdot \varphi_h^b) = (f, \varphi_h^b) \quad \forall \varphi_h^b \in H_h^b, \quad (6.2.127)$$

$$(\chi_h, \nabla \cdot v_h^{(1)}) - (\nabla \chi_h, v_h^b) = 0 \quad \forall \chi_h \in L_h. \quad (6.2.128)$$

Aus der zweiten Gleichung entnehmen wir die Beziehungen

$$\alpha_i^b \|\nabla \varphi_i^b\|_K^2 + \partial_i p_{h|K} (1, \varphi_i^b)_K = (f_i, \varphi_i^b)_K, \quad K \in \mathbb{T}_h,$$

mit den Koeffizienten  $\alpha_i^b$  von  $v_{h|K}$  bzgl. der Basis-Bulbs  $\varphi_i^b$ ,  $i = 1, 2$ . Dies ergibt

$$(\nabla \chi_h, v_h^b)_K = \alpha_i^b (\partial_i \chi_h, \varphi_i^b)_K = -\partial_i p_{h|K} \partial_i \chi_{h|K} \frac{(1, \varphi_i^b)_K^2}{\|\nabla \varphi_i^b\|_K^2} + (f_i, \varphi_i^b)_K \frac{(\partial_i \chi_h, \varphi_i^b)_K}{\|\nabla \varphi_i^b\|_K^2}.$$

Damit lässt sich die dritte Gleichung umformen zu

$$(\chi_h, \nabla \cdot v_h^{(1)}) + \sum_{K \in \mathbb{T}_h} \delta_K (\nabla \chi_h, \nabla p_h)_K = - \sum_{K \in \mathbb{T}_h} \frac{\delta_K |K|}{(1, \varphi_i^b)_K} (\partial_i \chi, (f_i, \varphi_i^b) \varphi_i^b)_K \quad \forall \chi_h \in L_h$$

mit den Parametern

$$\delta_K := \frac{(1, \varphi_K^b)_K^2}{|K| \|\nabla \varphi_K^b\|_K^2} \approx h_K^2.$$

Bis auf die rechte Seite entspricht dies gerade der modifizierten Gleichung (6.2.112) des druck-stabilisierten Verfahrens.

## 6.2.4 Lösung der diskreten Stokes-Probleme

Wir betrachten im Folgenden einheitlich sowohl inf-sup-stabile als auch stabilisierte Ansätze. Mit den natürlichen „Knotenwertbasen“ der Ansatzräume  $H_h$  und  $L_h$ .

$$\{\psi_h^i, i = 1, \dots, N_H := \dim H_h\}, \quad \{\chi_h^i, i = 1, \dots, N_L := \dim L_h\},$$

schreiben sich die Approximationsschemata (6.2.67) - (6.2.68), (6.2.95) - (6.2.96) bzw. (6.2.111) - (6.2.112) einheitlich in Form eines Blocksystems

$$\mathcal{A}\xi = \begin{bmatrix} A & B \\ -B^T & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix} =: \beta \quad (6.2.129)$$

für die Knotenwertvektoren  $\xi = \{x, y\}$  zu den Darstellungen

$$v_h = \sum_{i=1}^{N_H} x_i \psi_h^i, \quad p_h = \sum_{i=1}^{N_L} y_i \chi_h^i,$$

mit den zugehörigen Matrizen sowie den rechten Seiten

$$\begin{aligned} A &:= ((\nabla \psi_h^i, \nabla \psi_h^j))_{i,j}^{N_H}, & B &:= -((\chi_h^i, \nabla \cdot \psi_h^j))_{i,j=1}^{N_L, N_H}, & b &:= ((f, \psi_h^j))_{j=1}^{N_H}, \\ C &:= \left( \sum_{K \in \mathbb{T}_h} \delta_K (\nabla \chi_h^j, \nabla \chi_h^i)_K \right)_{i,j=1}^{N_L}, & c &:= \left( \sum_{K \in \mathbb{T}_h} \delta_K (\nabla \chi_h^j, f)_K \right)_{j=1}^{N_L}. \end{aligned}$$

### 6.2.5 Schur-Komplement-Verfahren

Die „Laplace-Matrix“  $A$  ist symmetrisch und positiv-definit und somit regulär. Folglich kann die Komponente  $x$  aus dem System eliminiert werden:

$$\Sigma y := (B^T A^{-1} B + C)y = B^T A^{-1} b + c, \quad x = A^{-1} b - A^{-1} B y. \quad (6.2.130)$$

Die Matrix  $\Sigma := B^T A^{-1} B + C$  wird das „Schur-Komplement“ von  $A$  in der Blockmatrix  $\mathcal{A}$  genannt. Mit ihr besteht die Block-Dreieckszerlegung

$$\mathcal{A} = \begin{bmatrix} A & B \\ -B^T & C \end{bmatrix} = \begin{bmatrix} A & 0 \\ -B^T & \Sigma \end{bmatrix} \begin{bmatrix} I & A^{-1} B \\ 0 & I \end{bmatrix}. \quad (6.2.131)$$

#### Uzawa-Algorithmus:

Die „klassische“ Methode zur Lösung des (quasi) Sattelpunktproblems (6.2.129) ist das „Uzawa<sup>4</sup>-Verfahren“. Ausgehend von einem Anfangswert  $y^0$  wird für  $l \in \mathbb{N}$  berechnet:

$$Ax^l = b - B y^{l-1}, \quad (6.2.132)$$

$$y^l = y^{l-1} + \theta (B^T x^l + C y^{l-1} - c), \quad (6.2.133)$$

wobei  $\theta > 0$  ein geeignet zu wählender Relaxationsparameter ist. Diese Iteration erfolgt im Raum der Druck-Knotenwerte. Um eventuelle Irregularitäten des Gitters  $\mathbb{T}_h$  (Zell-Anisotropien, heterogene Gitterweiten) auszugleichen, empfiehlt es sich, das System mit der Massematrix des Druckansatzraumes,

$$M = M_L = ((\chi_h^i, \chi_h^j))_{i,j=1}^{N_L},$$

---

<sup>4</sup>Hirofumi Uzawa (1928–2014): Japanischer Mathematischer Ökonom; Studium der Mathematik in Tokio 1951 und Ökonomie in Stanford 1956; Prof. für Ökonomie an verschiedenen Universitäten in den USA und in Japan; Mitinitiator der sog. „Mathematischen Ökonomie“; die nach ihm benannte „Uzawa-Iteration“ erschien ursprünglich im Rahmen der „Konkaven Optimierung“.

vorzukonditionieren. Der so modifizierte Uzawa-Algorithmus lautet dann:

$$Ax^i = b - By^{l-1}, \quad (6.2.134)$$

$$My^l = My^{l-1} + \theta(B^T x^l + Cy^{l-1} + c), \quad (6.2.135)$$

Jeder Iterationsschritt erfordert im wesentlichen die Invertierung der „Laplace-Matrix“  $A$  sowie der „Masse-Matrix“  $M$ , was etwa mit dem CG- oder einem Mehrgitter-Verfahren erfolgt. Durch Elimination von  $x^l$  erhalten wir den Uzawa-Algorithmus in Form einer Fixpunktiteration:

$$\begin{aligned} y^l &= y^{l-1} + \theta_l M^{-1}(B^T A^{-1}(b - By^{l-1}) + Cy^{l-1} + c) \\ &= (I - \theta M^{-1}\Sigma)y^{l-1} + \theta M^{-1}(B^T A^{-1}b + c). \end{aligned}$$

Der Uzawa-Algorithmus ist also eine gedämpfte Richardson-Iteration zur Lösung der Schur-Komplementgleichung. Für diese ergibt sich mit Hilfe des Banachschen Fixpunktsatzes das folgende Resultat.

**Satz 6.9 (Uzawa-Algorithmus):** *Der Uzawa-Algorithmus konvergiert unter der Bedingung  $\theta < \lambda_{\max}(M^{-1}\Sigma)^{-1}$  mit dem maximalen (reellen) Eigenwert von  $M^{-1}\Sigma$  gegen die Lösung  $\{x, y\}$  des Sattelpunktproblems (6.2.129). Dabei gilt mit*

$$0 < 1 - \lambda_{\min}(M^{-1}\Sigma)\lambda_{\max}(M^{-1}\Sigma)^{-1} = 1 - \text{cond}_2(M^{-1}\Sigma)^{-1} =: q < 1$$

die Fehlerabschätzung

$$|y^l - y| \leq \frac{q^l}{1 - q} |y^0 - y|. \quad (6.2.136)$$

**Beweis:** Die exakte Lösung  $y$  erfüllt die Fixpunktgleichung

$$y = (I - \theta M^{-1}\Sigma)y + \theta M^{-1}(B^T A^{-1}b + c).$$

Für den Iterationsfehler  $e^l := y - y^l$  gilt daher:

$$e^l = (I - \theta M^{-1}\Sigma)e^{l-1}.$$

Folglich liegt Konvergenz für beliebigen Startwert genau dann vor, wenn für den Spektralradius der Iterationsmatrix gilt  $\text{spr}(I - \theta M^{-1}\Sigma) < 1$ . Aus der Darstellung

$$M^{-1}\Sigma = M^{-1/2}(M^{-1/2}\Sigma M^{-1/2})M^{1/2}$$

folgt (Ähnlichkeitstransformation), dass die Eigenwerte der Matrix  $M^{-1}\Sigma$  gleich denen der symmetrischen und positiv definiten Matrix  $M^{-1/2}\Sigma M^{-1/2}$  sind. Daher sind alle Eigenwerte von  $M^{-1}\Sigma$  reell und positiv, und es gilt aufgrund der Voraussetzung an den Parameter  $\theta$ :

$$\text{spr}(I - \theta M^{-1}\Sigma) = \max\{1 - \theta\lambda(M^{-1}\Sigma)\} \leq 1 - \frac{\lambda_{\min}(M^{-1}\Sigma)}{\lambda_{\max}(M^{-1}\Sigma)}.$$

Die Behauptung folgt dann aus einem allgemeinen Resultat für die sukzessive Approximation (Fixpunktiteration). Q.E.D.

Wir werden weiter unten zeigen, dass  $\text{cond}_2(M^{-1}\Sigma) \leq c$  gleichmäßig für alle Gitterweiten  $h$ . Der Uzawa-Algorithmus konvergiert also mit einer von der Gitterfeinheit unabhängigen Rate, vorausgesetzt der Relaxationsparameter  $\theta$  ist ausreichend klein gewählt. Durch variable Wahl von  $\theta = \theta_l$  lässt sich die Konvergenzgeschwindigkeit des Verfahrens optimieren. Dann entspricht es gerade dem Gradientenverfahren zur Lösung der Schur-Komplement-Gleichung. Wir verzichten auf die zugehörige Analyse und betrachten lieber gleich das noch bessere CG-Verfahren.

### CG-Verfahren („Konjugierte-Gradienten-Verfahren“)

Ähnlich wie beim Uzawa-Algorithmus wird das System mit der Druck-Masse-Matrix  $M_L$  vorkonditioniert, d.h.: Das CG-Verfahren wird als „PCG-Verfahren“ auf das modifizierte System

$$M^{-1}\Sigma y = M^{-1}(B^T A^{-1}b + c) \quad (6.2.137)$$

angewendet. Dies ist gleichbedeutend mit der Anwendung des CG-Verfahrens auf die symmetrische und (modulo der Mittelwertbedingung) positiv-definite Matrix  $M^{-1/2}\Sigma M^{-1/2}$  die Spektralkondition

$$\kappa := \text{cond}_2(M^{-1}\Sigma) = \frac{\lambda_{\max}(M^{-1}\Sigma)}{\lambda_{\min}(M^{-1}\Sigma)}$$

bestimmt gemäß

$$|y^l - y| \leq \kappa \left( \frac{1 - \kappa^{-1/2}}{1 + \kappa^{-1/2}} \right)^l |y^0 - y|, \quad l \in \mathbb{N},$$

wobei  $y^0$  der Startwert der Iteration ist.

**Satz 6.10 (Schur-Komplement):** Für das Schur-Komplement  $\Sigma = B^T A^{-1}B$  gilt

$$\text{cond}_2(M^{-1}\Sigma) \leq \frac{c_0}{\beta_h^2}, \quad (6.2.138)$$

mit der Konstante  $\beta_h > 0$  in der „inf-sup“-Stabilitätsungleichung des jeweiligen Finite-Elemente-Ansatzes  $H_h/L_h$  zur Diskretisierung der Stokes-Gleichungen. Im konformen Fall  $H_h \subset H$  ist  $c_0 = 2$ , wogegen im allgemeinen Fall nur  $c_0 = 5$  gezeigt wird.

**Beweis:** Wir notieren zunächst die Beziehungen

$$\begin{aligned} \lambda_{\min}(M^{-1}\Sigma) &= \min_{y \in \mathbb{R}^{N_L}} \frac{\langle \Sigma y, y \rangle}{\langle M y, y \rangle} = \min_{y \in \mathbb{R}^{N_L}} \frac{\langle B^T A^{-1} B y + C y, y \rangle}{\langle M y, y \rangle} \\ &= \min_{y \in \mathbb{R}^{N_L}} \left\{ \frac{\langle A A^{-1} B y, A^{-1} B y \rangle}{\langle M y, y \rangle} + \frac{\langle C y, y \rangle}{\langle M y, y \rangle} \right\}. \end{aligned}$$

Für ein beliebiges Skalarprodukt etwa der Art  $\langle \cdot, \cdot \rangle_A := \langle A \cdot, \cdot \rangle$  gilt

$$|x|_A = \max_{z \in \mathbb{R}^{N_H}} \frac{\langle x, z \rangle_A}{|z|_A}, \quad z \in \mathbb{R}^{N_H}.$$

Unter Verwendung dieser Beziehung für  $x := A^{-1}By$  folgt

$$\begin{aligned} \lambda_{\min}(M^{-1}\Sigma) &= \min_{y \in \mathbb{R}^{N_L}} \left\{ \max_{z \in \mathbb{R}^{N_H}} \frac{\langle AA^{-1}By, z \rangle^2}{\langle Az, z \rangle \langle My, y \rangle} + \frac{\langle Cy, y \rangle}{\langle My, y \rangle} \right\} \\ &= \min_{y \in \mathbb{R}^{N_L}} \left\{ \max_{z \in \mathbb{R}^{N_H}} \frac{\langle By, z \rangle^2}{\langle Az, z \rangle \langle My, y \rangle} + \frac{\langle Cy, y \rangle}{\langle My, y \rangle} \right\}. \end{aligned}$$

Im Hinblick auf Definition der Systemmatrizen  $A$ ,  $B$ ,  $C$  und  $M$  ergibt sich bei Zuordnung  $y \in \mathbb{R}^{N_L} \leftrightarrow p_h \in L_h$  und  $z \in \mathbb{R}^{N_H} \leftrightarrow \psi_h \in H_h$  somit

$$\lambda_{\min}(M^{-1}\Sigma) = \min_{y \in \mathbb{R}^{N_L}} \left\{ \max_{z \in \mathbb{R}^{N_H}} \frac{(q_h, \nabla_h \cdot \psi_h)^2}{\|q_h\|^2 \|\nabla_h \psi_h\|^2} + \frac{1}{\|q_h\|^2} \sum_{K \in \mathbb{T}_h} \delta_K \|\nabla q_h\|_K^2} \right\} =: \beta_h^2.$$

Analog erschließen wir

$$\begin{aligned} \lambda_{\max}(M^{-1}\Sigma) &= \max_{y \in \mathbb{R}^{N_L}} \frac{\langle \Sigma y, y \rangle}{\langle My, y \rangle} \\ &= \max_{q_h \in L_h} \left\{ \max_{\psi_h \in H_h} \frac{(q_h, \nabla_h \cdot \psi_h)^2}{\|q_h\|^2 \|\nabla_h \psi_h\|^2} + \frac{1}{\|q_h\|^2} \sum_{K \in \mathbb{T}_h} \delta_K \|\nabla q_h\|_K^2} \right\} \leq 5. \end{aligned}$$

Dabei wurde die Abschätzung (gültig in  $d \leq 3$  Dimensionen)

$$\|\nabla_h \cdot \psi_h\|^2 \leq c_* \|\nabla_h \psi_h\|^2$$

mit  $c_* = 4$  verwendet. Im *konformen* Fall  $H_h \subset H$  kann man dies verschärfen zu  $c_* = 1$ . Damit ist die behauptete Schranke für  $\kappa$  bewiesen. Q.E.D.

Als Nebenprodukt des Beweises von Satz 6.2.5 erhalten wir auch die oben angekündigte Normschranke für das Schurkomplement

$$\|M^{-1}\Sigma\| \leq 4. \tag{6.2.139}$$

Bei der praktischen Durchführung des CG-Verfahrens für die Matrix  $M^{-1}\Sigma$  ist zu beachten, dass jeder Iterationsschritt im wesentlichen aus einer Matrix-Vektor-Multiplikation mit  $M^{-1}\Sigma$  besteht. Dies erfordert als „teuersten“ Schritt die Lösung eines Gleichungssystems mit der Laplace-Matrix  $A$ :

$$y \rightarrow M^{-1}\Sigma y \quad \Leftrightarrow \quad y \rightarrow By \rightarrow A^{-1}By \rightarrow (B^T A^{-1} B + C)y \rightarrow M^{-1}(B^T A^{-1} B + C)y.$$

Da  $A^{-1}$  i. Allg. nicht exakt zur Verfügung steht, muss die Auswertung von  $A^{-1}By$  iterativ erfolgen. Für die Matrix  $A$  (Diskretisierung des Vektor-Laplace-Operators) stehen auch auf „schlechten“ Gittern sehr leistungsfähige PCG- oder Mehrgitterverfahren zur Verfügung. Diese „inneren“ Iterationen im Rahmen der „äußeren“ CG-Iteration werden

durch ein Abbruchkriterium gesteuert, welches sich an der Entwicklung des jeweiligen Residuums orientiert. Dies bedeutet, dass die in der äußeren CG-Iteration zur Defektberechnung verwendete Matrix  $M^{-1}\tilde{A}^{-1}$  erstens fehlerhaft ist (z. B. mit elementweisem Fehler in der Größenordnung  $\mathcal{O}(10^{-8})$ ) und sich zweitens ständig ändert. Damit sind die Voraussetzungen für die Konvergenz des äußeren CG-Verfahrens nicht voll erfüllt, was sich in der Regel in einem „Festfressen“ des Residuums auf dem Genauigkeitslevel  $10^{-8}$  der inneren Iteration äußert. Diesen unangenehmen Defekt kann man dadurch beheben, dass die PCG-Iteration für die Schur-Komplementmatrix  $M^{-1}\Sigma$  in eine äußere Defektkorrekturiteration eingebettet wird. Die „ungenau“ PCG-Iteration wird dabei quasi als „Vorkonditionierung“  $S$  einer einfachen Richardson-Iteration verwendet:

$$y^l \rightarrow d^l := M^{-1}\Sigma y^l - M^{-1}(B^T A^{-1}b + c) \rightarrow r^l = S^{-1}d^l \rightarrow y^{l+1} := y^l + r^l.$$

Auf diese Weise erhält man eine sehr einfache, robuste und effiziente Lösungsmethode für die diskreten Stokes-Gleichungen. In neuerer Zeit sind aber auch konkurrierende Mehrgitterverfahren entwickelt worden, welche direkt an der Blockmatrix  $\mathcal{A}$  orientiert sind und die beschriebene „stabilisierte“ Schur-Komplement-CG-Iteration noch an Effizienz übertreffen (s. etwa Turek [145]). Dabei muss allerdings besondere Sorgfalt auf die Wahl des „Glätters“ verwendet werden. Wir werden dies später direkt für die allgemeinere Situation der linearisierten Navier-Stokes-Gleichungen diskutieren.

## 6.3 Lösung der stationären Navier-Stokes-Gleichungen

Die Lösung der *nichtlinearen* Navier-Stokes-Gleichungen erfolgt zwangsläufig iterativ. Dabei sind einige besonders im Fall dominanter Nichtlinearität (d. h. für höhere Reynolds-Zahl) kritische Punkte zu beachten.

### 6.3.1 Diskretisierung des Konvektionsterms

Ausgangspunkt sind die (stationären) Navier-Stokes-Gleichungen in dimensionsloser Form

$$-\nu\Delta v + v \cdot \nabla v + \nabla p = f, \quad \nabla \cdot v = 0 \quad \text{in } \Omega, \quad (6.3.140)$$

auf einem Gebiet  $\Omega \subset \mathbb{R}^d$  ( $d = 2, 3$ ) mit ausreichend regulärem Rand  $\partial\Omega$ . Hier steht  $\nu$  wieder für einen Parameter, welcher das Größenverhältnis zwischen Diffusion und Konvektion beschreibt. Im Fall einer charakteristischen Länge  $L \sim 1$  und Geschwindigkeit  $U \sim 1$  ist  $\nu \sim \text{Re}^{-1}$ . Der Einfachheit halber beschränken wir uns im Folgenden auf die Betrachtung reiner *homogener* Dirichlet-Randbedingungen,  $v|_{\partial\Omega} = 0$ . Im Fall inhomogener Randbedingungen (z. B. vorgegebene Einströmung im Kanal oder feste Randströmung in der Nische) oder Neumann-Randbedingungen („freie“ Ausströmung im Kanal oder im Außenraum) gelten analoge Aussagen.

Bei Verwendung eines der oben beschriebenen Stokes-Element-Raumpaare  $H_h/L_h \approx H/L$  ergibt sich aus der variationellen Formulierung die folgende Diskretisierung des Konvektionsterms:

$$(v \cdot \nabla v, \psi) \approx (v_h \cdot \nabla_h v_h, \psi_h),$$

wobei  $\nabla_h$  wieder die im *nicht-konformen* Fall notwendige „zellweise“ Interpretation des  $\nabla$ -Operators bezeichnet. Bei der üblichen Finite-Elemente-Approximation trägt dieser Term vor allem zu den Nebendiagonalen der Systemmatrix  $A$  bei, wodurch diese ihre Definitheitseigenschaften verliert. Dies wiederum bewirkt auf größeren Gittern unphysikalische Oszillationen in der Lösung sowie das Versagen der üblichen iterativen Lösungsverfahren.

### Stabilisierung im eindimensionalen Fall

Wir wollen diese Aussagen anhand einer einfachen Modellsituation illustrieren. Auf dem eindimensionalen Gebiet  $\Omega = I := (0, 1) \in \mathbb{R}^1$  wird die *singulär gestörte* Randwertaufgabe (Sturm-Liouville-Problem)

$$-\varepsilon u''(x) + q(x)u'(x) = 0, \quad x \in I, \quad u(0) = 1, \quad u(1) = 0, \quad (6.3.141)$$

betrachtet. Im Fall  $q \equiv 1$  hat die eindeutige Lösung die Gestalt (Abb. 6.5)

$$u^\varepsilon(x) = \frac{e^{1/\varepsilon} - e^{x/\varepsilon}}{e^{1/\varepsilon} - 1}.$$

Im Fall  $\varepsilon \ll 1$  ist für  $x = 1 - \delta$  und  $\delta > \varepsilon$ :

$$u^\varepsilon(1 - \delta) = \frac{e^{1/\varepsilon}}{e^{1/\varepsilon} - 1} (1 - e^{-\delta/\varepsilon}) \approx 1, \quad \sup_{x \in I} |u^{\varepsilon''}(x)| \approx \varepsilon^{-2},$$

was den Namen „Grenzschichtlösung“ rechtfertigt. Für  $\varepsilon = 0$  ergibt sich die Grenzlösung  $u^0 \equiv 1$ , welche die Randbedingung bei  $x = 1$  nicht erfüllt.

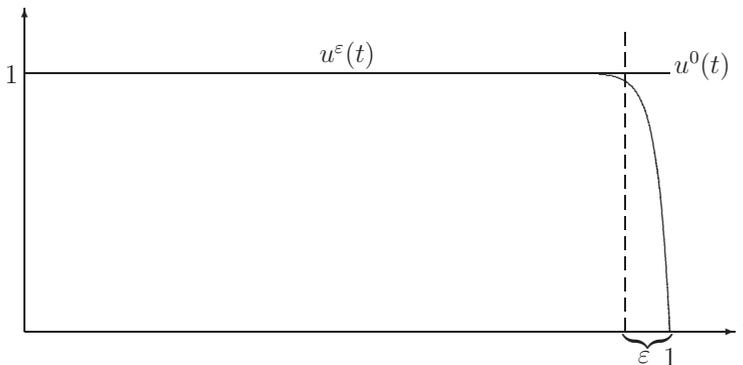


Abbildung 6.5: Lösung des singulär gestörten Sturm-Liouville-Problems für  $\varepsilon = 0.1$ .

Die Approximation dieses Problems mit dem üblichen (zentralen) Differenzenschema zur (äquidistanten) Schrittweite  $h = 1/(N + 1)$  ergibt

$$-(\varepsilon + \frac{1}{2}h)y_{n-1} + 2\varepsilon y_n - (\varepsilon - \frac{1}{2}h)y_{n+1} = 0, \quad 1 \leq n \leq N, \quad y_0 = 1, \quad y_{N+1} = 0.$$

Die zugehörige Koeffizientenmatrix ist offenbar nur unter der restriktiven Bedingung

$$h \leq 2\varepsilon \quad (6.3.142)$$

diagonal-dominant. Für  $h > 2\varepsilon$  zeigt die diskrete Lösung ein unphysikalisches Verhalten. Um dies zu sehen, machen wir wieder einen Lösungsansatz der Form  $y_n = \lambda^n$ . Die möglichen Werte für  $\lambda$  sind gerade die Wurzeln  $\lambda_{\pm}$  der quadratischen Gleichung

$$\lambda^2 + \frac{2\varepsilon}{\frac{1}{2}h - \varepsilon} \lambda + \frac{\frac{1}{2}h + \varepsilon}{\frac{1}{2}h - \varepsilon} = 0.$$

Berücksichtigung der Randbedingungen  $y_0 = 1$  und  $y_{N+1} = 0$  in dem Lösungsansatz

$$y_n = c_+ \lambda_+^n + c_- \lambda_-^n$$

ergibt für die Koeffizienten die Beziehungen  $c_+ + c_- = 1$ ,  $c_+ \lambda_+^{N+1} + c_- \lambda_-^{N+1} = 0$  und folglich

$$c_- = \frac{\lambda_+^{N+1}}{\lambda_+^{N+1} - \lambda_-^{N+1}}, \quad c_+ = 1 - \frac{\lambda_+^{N+1}}{\lambda_+^{N+1} - \lambda_-^{N+1}} = -\frac{\lambda_-^{N+1}}{\lambda_+^{N+1} - \lambda_-^{N+1}}.$$

Die Lösung hat also die Gestalt

$$y_n = \frac{\lambda_+^{N+1} \lambda_-^n - \lambda_-^{N+1} \lambda_+^n}{\lambda_+^{N+1} - \lambda_-^{N+1}}, \quad n = 0, \dots, N+1. \quad (6.3.143)$$

Im vorliegenden Fall sind die Wurzeln gegeben durch

$$\lambda_{+,-} = \frac{-\varepsilon \pm \sqrt{\varepsilon^2 + (\frac{1}{2}h + \varepsilon)(\frac{1}{2}h - \varepsilon)}}{\frac{1}{2}h - \varepsilon} = \frac{\varepsilon \mp \frac{1}{2}h}{\varepsilon - \frac{1}{2}h}, \quad \lambda_+ = 1, \quad \lambda_- = \frac{\varepsilon + \frac{1}{2}h}{\varepsilon - \frac{1}{2}h}.$$

Für  $\varepsilon \ll \frac{1}{2}h$  ist  $\lambda_- \approx -1$ . In diesem Fall wird also eine oszillierende Lösung erzeugt,

$$y_n = \frac{\lambda_-^n - \lambda_-^{N+1}}{1 - \lambda_-^{N+1}}, \quad n = 0, \dots, N+1.$$

welche qualitativ nicht den richtigen Lösungsverlauf wiedergibt. Zur Unterdrückung dieses Defekts gibt es verschiedene Strategien, die im Folgenden skizziert werden.

**i) Upwind-Diskretisierung:** Zunächst kann der Term erster Ordnung  $u'(x)$  in der Differentialgleichung statt mit dem zentralen mit einem der einseitigen Differenzenquotienten

$$\Delta_h^+ u(x) = \frac{u(x+h) - u(x)}{h}, \quad \Delta_h^- u(x) = \frac{u(x) - u(x-h)}{h}$$

approximiert werden. Bei der Wahl des *rückwärtigen* Differenzenquotienten  $\Delta_h^-$  wird dem physikalischen Vorgang eines Informationstransports in positive  $x$ -Richtung Rechnung getragen (vergl. die Form der Grenzlösung  $u^0(x)$ ). Dies führt auf die Differenzgleichungen

$$(-\varepsilon + h)y_{n-1} + (2\varepsilon + h)y_n - \varepsilon y_{n+1} = 0.$$

Die zugehörige Systemmatrix  $A_h$  ist für beliebiges  $h > 0$  wieder diagonal dominant; sie ist sogar eine  $M$ -Matrix. Der Lösungsansatz  $y_n \lambda^n$  führt in diesem Fall auf die Gleichung

$$\lambda^2 - \frac{2\varepsilon + h}{\varepsilon} \lambda + \frac{\varepsilon + h}{\varepsilon} = 0,$$

mit den Wurzeln

$$\lambda_{+,-} = \frac{2\varepsilon + h}{2\varepsilon} \pm \sqrt{(2\varepsilon + h)^2 - 4\varepsilon(\varepsilon + h)} = \frac{2\varepsilon + h \pm h}{2\varepsilon}, \quad \lambda_+ = \frac{\varepsilon + h}{\varepsilon}, \quad \lambda_- = 1.$$

Die kritische Wurzel  $\lambda_+$  ist hier stets positiv, so dass in der diskreten Lösung

$$y_n = \frac{\lambda_+^{N+1} - \lambda_+^n}{\lambda_+^{N+1} - 1},$$

keine ungewollten Oszillationen in der Näherungslösung entstehen. Diese spezielle Art der einseitigen Diskretisierung des Terms  $u'(x)$  nennt man „Rückwärtsdiskretisierung“ oder auch englisch „upwind discretization“. Da der verwendete einseitige Differenzenquotient aber nur die Approximationsordnung  $\mathcal{O}(h)$  hat, ist auch das Gesamtverfahren nur von erster Ordnung genau. Dies limitiert die Approximationsgenauigkeit in Bereichen, in denen die Lösung glatt ist, selbst wenn die Gitterweite in der Grenzschicht ausreichend fein gemäß  $h \approx \varepsilon$  gewählt wird.

**ii) Künstliche Diffusion:** Unter Beibehaltung der zentralen Diskretisierung des Terms  $u'(x)$  wird der *Diffusionskoeffizient*  $\varepsilon$  auf einen größeren Wert  $\varepsilon_h := \varepsilon + \delta h$  gesetzt. Dies führt auf die Differenzgleichungen

$$-(\varepsilon_h + \frac{1}{2}h)y_{n-1} + 2\varepsilon_h y_n - (\varepsilon_h - \frac{1}{2}h)y_{n+1} = 0, \quad 1 \leq n \leq N.$$

Für die zugehörige Lösung erhält man wieder durch einen Potenzansatz die Darstellung

$$y_n = \frac{\lambda_+^{N+1} - \lambda_+^n}{\lambda_+^{N+1} - 1}, \quad \lambda_+ = \frac{\varepsilon_h + \frac{1}{2}h}{\varepsilon_h - \frac{1}{2}h}.$$

Offenbar ist in diesem Fall  $\lambda_+ > 0$  für  $\varepsilon + \delta h > \frac{1}{2}h$ , d. h. für die Wahl  $\delta \geq \frac{1}{2}$ . Mit diesem Ansatz erhält man also ebenfalls wieder eine  $M$ -Matrix und somit eine stabile Diskretisierung. Allerdings wird nun die Grenzschicht stark verschmiert auf das Intervall  $[1 - \varepsilon_h, 1]$ , und die globale Approximationsgüte ist aufgrund der Störung des Differentialoperators ebenfalls lediglich  $\mathcal{O}(h)$ .

Für allgemeinen Transportkoeffizienten  $q(x)$  muss das „Upwinding“ abhängig vom Vorzeichen von  $q_n = q(x_n)$  angesetzt werden. Die einseitigen Differenzenquotienten werden gemäß der folgenden Schaltvorschrift angesetzt:

$$\text{sign}(q_n) = \begin{cases} +1 & : \quad \Delta_n^- \\ -1 & : \quad \Delta_n^+ \end{cases}.$$

Dies führt dann wieder auf eine für alle  $h > 0$  diagonal dominante Systemmatrix  $A_h$ .

Anhand des obigen einfachen Beispiels haben wir gesehen, dass bei singulär gestörten Problemen die einfachen Dämpfungsstrategien *Rückwärtsdiskretisierung* oder *künstliche Diffusion* zwar auf stabile Diskretisierungen führen, die Approximationsanordnung aber auf  $\mathcal{O}(h)$  reduzieren. Die Frage nach einer optimalen Dämpfungsstrategie höherer Ordnung zur Diskretisierung von *Transporttermen* ist noch nicht vollständig geklärt. Ansätze in diese Richtung bedienen sich z. B. einseitiger Differenzenquotienten höherer Ordnung (beim *Upwinding*) oder künstlicher Diffusionsterme der Form  $\delta h^2 u^{(iv)}$ . Allerdings kann die starke  $M$ -Matrizeigenschaft nur mit Diskretisierungen erster Ordnung erreicht werden. Diese Beschränkung kann durch eine Modifikation der „künstlichen Diffusion“ im Rahmen des Galerkin-Verfahrens mit finiten Elementen überwunden werden.

**iii) „Stromliniendiffusion“:** Wir betrachten wieder den Spezialfall  $q \equiv 1$ . Bei der sog. „Stromliniendiffusions-Methode“ wird die übliche variationelle Formulierung der Randwertaufgabe (6.3.141),

$$\varepsilon(u', \varphi') + (u' + \alpha u, \varphi) = (f, \varphi), \quad \forall \varphi \in H := H_0^1(I), \quad (6.3.144)$$

mit  $\alpha \geq 0$ , modifiziert zu

$$\varepsilon(u', \varphi') + (u' + \alpha u, \varphi + \delta \varphi') = (f, \varphi + \delta \varphi'), \quad \forall \varphi \in H. \quad (6.3.145)$$

mit einer Parameterfunktion  $\delta$ , die an die Gitterweite  $h$  gekoppelt wird. Die resultierende Bilinearform

$$a_\delta(u, v) := \varepsilon(u', v') + (u' + \alpha u, v + \delta v')$$

ist dann bzgl. der modifizierten Energie-Norm

$$\|v\|_\delta := (\varepsilon\|v'\|^2 + \|\delta^{1/2}v'\|^2 + \alpha\|v\|^2)^{1/2}$$

koerzitiv gemäß

$$a_\delta(v, v) \geq \|v\|_\delta^2, \quad v \in H. \quad (6.3.146)$$

Zum Nachweis dieser Beziehung nutzt man die Identität

$$\int_0^1 v'(x)v(x) dx = \frac{1}{2} \int_0^1 v^2(x) dx = \frac{1}{2} \{v^2(1) - v^2(0)\} = 0.$$

Es sei betont, dass der Parameter  $\delta$  i. Allg. eine Funktion von  $x$  ist (stückweise konstant auf der Zerlegung  $0 = x_0 < \dots < x_{N+1} = 1$ ) und folglich innerhalb der Norm  $\|\delta^{1/2}v'\|$  stehen muss. Analog verwenden wir im Folgenden auch das Symbol  $h = h(x)$  für eine stückweise konstante Gitterweitenfunktion mit  $h|_{I_n} \equiv h_n$  ( $n = 1, \dots, N+1$ ). Das zugehörige FE-Galerkin-Verfahren (mit linearen Ansatzfunktionen) im Ansatzraum  $H_h \subset H$  lautet nun: *Finde  $u_h \in H_h$ , so dass*

$$a_\delta(u_h, \varphi_h) = l_\delta(\varphi_h) \quad \forall \varphi_h \in H_h, \quad (6.3.147)$$

mit dem modifizierten Lastfunktional  $l_\delta(v) := (f, v + \delta v')$ . Durch Kombination der Variationsgleichungen für  $u$  und  $u_h$  erhalten wir die folgende gestörte Orthogonalitätsbeziehung für den Fehler  $e := u - u_h$ :

$$a_\delta(e, \varphi_h) = (u' + \alpha u - f, \delta \varphi_h') = \varepsilon(u'', \delta \varphi_h'). \quad (6.3.148)$$

Für die FE-Diskretisierung mit Stromliniendiffusionsstabilisierung (kurz „SD-FEM“) hat man dann das folgende Resultat:

**Satz 6.11 (Konvergenzsatz für SD-FEM):** *Es sei  $\varepsilon \ll h_{\min}$ , und der Stabilisierungsparameter in der Stromliniendiffusion sei auf jedem Teilintervall  $I_n$  wie  $\delta_n = h_n$  gewählt. Dann gilt für den Fehler  $e := u - u_h$  bzgl. der modifizierten Energienorm die a priori Abschätzung*

$$\|e\|_{\delta} \leq c \|h^{3/2} u''\|, \quad (6.3.149)$$

mit einer von  $h$  (und  $\delta$ ) unabhängigen Konstante  $c$ .

**Beweis:** Wir skizzieren den Beweis für den Fall  $\alpha = 1$ . Mit Hilfe der Koerzitivitätsbeziehung (6.3.146) und der Orthogonalitätsrelation (6.3.148) erhalten wir mit beliebigem  $\varphi_h \in H_h$ :

$$\|e\|_{\delta}^2 \leq a_{\delta}(e, u - \varphi_h) + \varepsilon(u'', \delta(\varphi_h - u_h)'). \quad (6.3.150)$$

Der erste Term rechts wird weiter abgeschätzt durch

$$\begin{aligned} |a_{\delta}(e, u - \varphi_h)| &\leq \varepsilon |(e', (u - \varphi_h)')| + |(e' + e, u - \varphi_h + \delta(u - \varphi_h)')| \\ &\leq \varepsilon \|e'\| \| (u - \varphi_h)' \| + \{ \|\delta^{1/2} e'\| + \|e\| \} \|\delta^{-1/2} (u - \varphi_h)\| \\ &\quad + \{ \|\delta^{1/2} e'\| + \|\delta^{1/2} e\| \} \|\delta^{1/2} (u - \varphi_h)'\|. \end{aligned}$$

Für die Wahl  $\varphi_h := i_h u$  folgt mit Hilfe der üblichen lokalen Interpolationsabschätzungen bei Beachtung von  $\delta \equiv h \leq 1$  und  $\varepsilon \leq h_{\min}$ :

$$\begin{aligned} |a_{\delta}(e, u - i_h u)| &\leq c \varepsilon \|e'\| \|h u''\| + c \{ \|\delta^{1/2} e'\| + \|e\| \} \|\delta^{-1/2} h^2 u''\| \\ &\quad + c \{ \|\delta^{1/2} e'\| + \|e\| \} \|\delta^{1/2} h u''\| \\ &\leq \frac{1}{4} \|e\|_{\delta}^2 + c \|h^{3/2} u''\|^2. \end{aligned}$$

Für den zweiten Term rechts in (6.3.150) folgt für  $\varphi_h := i_h u$  mit analogen Argumenten

$$\begin{aligned} \varepsilon |(u'', \delta(i_h u - u_h)')| &\leq \varepsilon \|\delta^{1/2} u''\| \{ \|\delta^{1/2} (i_h u - u_h)'\| + \|\delta^{1/2} e'\| \} \\ &\leq \frac{1}{4} \|e\|_{\delta}^2 + c \|h^{3/2} u''\|^2. \end{aligned}$$

Kombination der bisher gezeigten Abschätzungen ergibt das gewünschte Resultat. Q.E.D.

Die Fehlerabschätzung (6.3.149) besagt insbesondere, dass im Fall einer glatten Lösung  $u$  (ohne Grenzschicht), oder wenn die Gitterweite in der Grenzschicht hinreichend klein gewählt wird, die SD-FEM bzgl. der „Energie-Norm“ mit der Ordnung  $\mathcal{O}(h^{3/2})$  konvergiert. Damit ist die einfachste SD-FEM im transport-dominanten Fall von höherer Ordnung als das *Upwinding* stabilisierte Differenzenverfahren. Allerdings muss bemerkt werden, dass die zugehörige Systemmatrix  $A_h^{(\delta)}$  zwar definit ist, aber keine  $M$ -Matrix-Eigenschaft hat; insbesondere liegt in der Regel keine Diagonaldominanz vor.

### Stabilisierung der Navier-Stokes-Gleichungen

Die beschriebenen Methoden zur Transportstabilisierung im eindimensionalen Fall besitzen natürliche Verallgemeinerungen für die d-dimensionalen Navier-Stokes-Gleichungen.

**i) „upwind“-Stabilisierung:** Wir beschreiben diese Stabilisierungstechnik exemplarisch für das nichtkonforme „rotiert-bilineare“ Stokes-Element. Jedes Viereck  $K \in \mathbb{T}_h$  wird gemäß Abb. 6.6 zerlegt in acht Teildreiecke  $S_{ij}$ . Weiter wird für jede Kante  $\Gamma_l$  und deren Mittelpunkt  $m_l$  eine sog. „lumping region“  $R_l$  definiert durch  $R_l := \cup_{k \in \Lambda_l} S_{lk}$ , wobei  $\Lambda_l := \{k \mid m_l \text{ und } m_k \text{ gehören zur selben Zelle } K\}$ . Der Rand der „lumping region“  $R_l$  besteht aus den Kanten  $\Gamma_{lk} := \partial S_{lk} \cap \partial S_{kl}$ , d. h.:  $\partial R_l = \cup_{k \in \Lambda_l} \Gamma_{lk}$ . Auf diesem Wege erhalten wir eine kantenorientierte Zerlegung des Gitterbereichs  $\Omega_h = \cup_l R_l$  (s. Schieweck/Tobiska [140]).

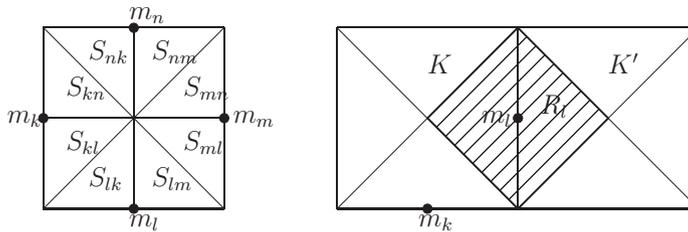


Abbildung 6.6: Zerlegung einer Rechteckzelle in „lumping regions“.

Mit diesen Bezeichnungen wird nun die folgende Modifikation der nichtlinearen Form  $n(\cdot, \cdot, \cdot)$  definiert:

$$\tilde{n}(u_h, v_h, w_h) := \sum_{K,l} (1 - \lambda_{lk}(u_h)) \{v_h(m_k) - v_h(m_l)\} w_h(m_l) \int_{\Gamma_l} u_h \cdot n_{lk} ds,$$

wobei die Parameter  $\lambda_{lk}$  von der lokalen „Flussrichtung“ abhängen. Mit der Setzung

$$x := \frac{1}{\nu} \int_{\Gamma_{lk}} u_h \cdot n_{lk} ds,$$

sind populäre Abhängigkeiten etwa das sog. „einfache upwinding“ oder das „Samarski<sup>5</sup> upwinding“:

$$\lambda_{lk} := \begin{cases} 1 & \text{für } x \geq 0, \\ 0 & \text{für } x < 0, \end{cases} \quad (\text{einfaches „upwinding“})$$

$$\lambda_{lk} := \begin{cases} (\frac{1}{2} + x)/(1 + x) & \text{für } x \geq 0, \\ 1/(2 - 2x) & \text{für } x < 0 \end{cases} \quad (,Samarski upwinding“).$$

<sup>5</sup>Alexander Andrejewitsch Samarski (1919–2008): Russischer Mathematiker; ab 1953 Prof. an der Lomonossow-Universität in Moskau; fundamentale Beiträge zu Differenzenverfahren für Probleme in der Mathematischen Physik.

Eine Fehleranalyse zeigt, dass diese Modifikation auf ein Diskretisierungsschema erster Ordnung in der Gitterweite führt. Die Stabilisierung drückt sich darin aus, dass die Hauptdiagonalblöcke der Matrix  $A + \tilde{N}(\cdot)$   $M$ -Matrizen werden. Dies ist die Schlüsseleigenschaft für die Konvergenz verschiedener iterativer Lösungsverfahren inkl. Mehrgitterverfahren. Die beschriebene „upwind“-Diskretisierung kann auf natürliche Weise auf drei Raumdimensionen übertragen werden. Ferner ist eine analoge Konstruktion auch für das konforme  $Q_1^c/Q_1^c$ -Element mit Druckstabilisierung möglich.

**ii) Stromliniendiffusion**<sup>6</sup> : Die Idee der „Stromliniendiffusion“ ist es, künstliche Diffusion nur in Transport- bzw. Stromlinienrichtung so einzuführen, dass die volle 2. Ordnung der Diskretisierung bewahrt bleibt. Dies kann man auf zwei (im Grunde äquivalenten) Wegen erreichen:

- durch Erweiterung der Testfunktionen um transport-orientierte Anteile, was auf ein sog. „Petrov<sup>7</sup>-Galerkin-Verfahren“ führt, oder
- durch Hinzufügung von gewissen „least-squares“ Termen zur variationellen Formulierung des Problems.

Wir beschreiben im Folgenden eine einfache Variante für die (stationären) Navier-Stokes-Gleichungen: *Bestimme  $v_h \in v_h^{\text{in}} + \mathbf{H}_h$  und  $p_h \in L_h$ , so dass*

$$\begin{aligned} a_h(v_h, \varphi_h) + n_h(v_h \cdot v_h, \varphi_h) + b_h(p_h, \varphi_h) + s_h(\{v_h, p_h\}, \{\varphi_h, \chi_h\}) \\ = (f, \varphi_h) + r_h(\{v_h, p_h\}, \{\varphi_h, \chi_h\}) \end{aligned} \quad (6.3.151)$$

für alle Paare  $\{\varphi_h, \chi_h\} \in \mathbf{H}_h \times L_h$ , wobei

$$\begin{aligned} s_h(\{v_h, p_h\}, \{\varphi_h, \chi_h\}) &= \sum_{K \in \mathbb{T}_h} \delta_K \{(\nabla p_h + v_h \cdot \nabla v_h, \nabla \chi_h + \bar{v}_h \nabla \varphi_h)_K + (\nabla \cdot v_h, \nabla \cdot \varphi_h)_K\}, \\ r_h(\{v_h, p_h\}, \{\varphi_h, \chi_h\}) &= \sum_{K \in \mathbb{T}_h} \delta_K (f + \nu \Delta v_h, \nabla \chi_h + \bar{v}_h \cdot \nabla \varphi_h)_K \end{aligned}$$

mit einer Referenzgeschwindigkeit  $\bar{v}_h$ . Die Stabilisierungsparameter  $\delta_K$  werden gemäß der folgenden Vorschrift gewählt:

$$\delta_K = \min \left\{ \frac{h_K^2}{\nu}, \frac{h_K}{|\bar{v}|_K} \right\}. \quad (6.3.152)$$

Diese Diskretisierung hat mehrere spezielle Eigenschaften. Der erste Term in der Summe

$$\sum_{K \in \mathbb{T}_h} \delta_K \{(\nabla p_h, \nabla \chi_h)_K + (v_h \cdot \nabla v_h, \bar{v}_h \cdot \nabla \varphi_h)_K + (\nabla \cdot v_h, \nabla \cdot \varphi_h)_K\}$$

stabilisiert den Druck für das konforme  $Q_1^c/Q_1^c$ -Element (inf-sup-Bedingung), der zweite Term stabilisiert den Transportoperator (Stromliniendiffusion), und der dritte Term verstärkt die Masseerhaltung (Inkompressibilität). Die anderen Terme dienen nur zur Korrektur, um die 2. Ordnung der Diskretisierung zu erhalten.

<sup>6</sup>T. J. R. Hughes, A. N. Brooks, Brooks: *Streamline upwind/PetrovGalerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equation*, Comput. Meth. Appl. Mech. Engrg. 32, 199–259 (1982).

<sup>7</sup>Georgi Iwanowitsch Petrov (1912–1987): Russischer Ingenieur; 1965–1973 Direktor des Instituts für Raumfahrtforschung; Publ.: „Application of the Galerkin method and the problem of flow stability of a viscous liquid“ (russ.), Prikl. Mat. Mekh. 4, 36–47 (1947)

**Bemerkung 6.4:** Eine theoretische Analyse (s. etwa Tobiska/Verfürth [144]) zeigt, dass die beschriebene Art der Stabilisierung im Rahmen des Galerkin-Verfahrens tatsächlich zu einer Verbesserung gegenüber dem standard „upwind“-Schema führt; genauer verhält sich der Fehler für das  $Q_1^c/Q_1^c$ -Element i. Allg. wie  $\mathcal{O}(h^{3/2})$ . Diese asymptotische Aussage wird allerdings dadurch relativiert, dass in die Fehlerkonstante Schranken für die  $H^2$ -Norm des Geschwindigkeitsfeldes eingehen, welche sich im Falle einer Grenzschichtlösung natürlich wie  $\mathcal{O}(\nu^{-1})$  verhalten. Um die höhere Approximationsordnung zu realisieren, muss die Grenzschicht also gegebenenfalls durch lokale (anisotrope) Gitterverfeinerung aufgelöst werden. Dies wäre natürlich auch bei jeder anderen Transportstabilisierung erforderlich.

**iii) Lokale Projektionsstabilisierung (LPS):** Vom praktischen Gesichtspunkt hat die oben beschriebene Stromlinien-Stabilisierung gravierende Nachteile. Ihre Verwendung in Verbindung mit der Neumann-Ausströmrandbedingung („do-nothing“ Bedingung) in Kanälen führt zu unphysikalischem Strömungsverhalten („numerische“ Grenzschicht) entlang des Ausströmrandes; Stromlinien erscheinen nach außen gebeugt. Ferner ist die Auswertung der vielen Zusatzterme in den Stabilisierungsformen  $c_h(\chi_h, p_h)$  und  $g_h(v_h; \chi_h)$ , insbesondere in 3D, sehr aufwendig.

Diese Probleme können behoben werden durch Verwendung der sog. „lokalen Projektionsstabilisierung“ (LPS) nach Becker/Braack [98]. Hier werden die Stabilisierungsformen  $r_h = 0$  und

$$s_h := (\nabla(p_h - \pi_{2h}p_h), \nabla(\chi_h - \pi_{2h}\chi_h)) + (\bar{v}_h \cdot \nabla(v_h - \pi_{2h}v_h), \delta_h \bar{v}_h \cdot \nabla(\varphi - \pi_{2h}\varphi)),$$

verwendet. Dabei ist  $\pi_{2h}$  eine Projektion oder Interpolation in den Ansatzraum  $L_{2h}$ , welcher auf einem größeren Gitter  $\mathbb{T}_{2h}$  definiert ist. Das resultierende Verfahren ist von 2-ter Ordnung genau, die Auswertung der zugehörigen Systemmatrizen ist vergleichsweise billig, und der Konsistenzfehler am Ausströmrand wird vermieden. Der Parameter  $\delta_h$  wird wieder zellweise in Abhängigkeit der lokalen Konvektion und Diffusion gewählt.

### 6.3.2 Linearisierung

Wir beschreiben zunächst die verschiedenen Linearisierungstechniken auf dem kontinuierlichen Level. Dabei wird ausgehend von einem Startwert  $v^0 \in H$  eine Folge von Iterierten  $\{v^l, p^l\} \in H \times L$  ( $l \in \mathbb{N}$ ) bestimmt, welche gegen die exakte Lösung  $\{v, p\}$ , konvergiert.

i) *Stokes-Linearisierung:*

Im Falle kleiner Reynolds-Zahl  $Re = 1/\nu \ll 1$  (zähe Flüssigkeit, kleine Geschwindigkeiten, kleine Abmessungen) genügt oft die einfache „Stokes-Linearisierung“

$$-\nu \Delta v^l + \nabla p^l = f - v^{l-1} \cdot \nabla v^{l-1}, \quad (6.3.153)$$

$$\nabla \cdot v^l = 0. \quad (6.3.154)$$

Die Nichtlinearität  $v \cdot \nabla v$  wird hier voll explizit behandelt. In jedem Iterationsschritt ist der symmetrische und positiv-definite Stokes-Operator zu invertieren.

ii) *Oseen-Linearisierung*:

Bei größeren Reynolds-Zahlen wird die sog. „Oseen<sup>8</sup>-Linearisierung“ verwendet:

$$-\nu\Delta v^l + \bar{v} \cdot \nabla v^l + \nabla p^l = f, \quad (6.3.155)$$

$$\nabla \cdot v^l = 0. \quad (6.3.156)$$

Hier ist  $\bar{v} \approx v^l$  eine geeignete Approximation, z. B. im einfachsten Fall  $\bar{v} := v^{l-1}$  (*konstante* Extrapolation). Die resultierende Iteration wird auch „Funktionaliteration“ genannt. Im Rahmen eines Pseudo-Zeitschrittverfahrens kann zur Erzielung höherer Genauigkeit auch  $\bar{v} := 2v^{l-1} - v^{l-2}$  (*lineare* Extrapolation) verwendet werden. Der in jedem Schritt zu invertierende Operator ist ein *unsymmetrischer* Diffusions-Transport-Operator. Zur Verbesserung der Konvergenz sollte diese Funktionaliteration in Form einer Defektkorrektur-Iteration durchgeführt werden. Dazu wird mit dem Defekt

$$d^{l-1} := f - \nu\Delta v^{l-1} - v^{l-1} \cdot \nabla v^{l-1} - \nabla p^{l-1}$$

eine Korrektur berechnet aus der Gleichung

$$-\nu\Delta w^l + \bar{v} \cdot \nabla w^l + \nabla q^l = d^{l-1}, \quad (6.3.157)$$

$$\nabla \cdot w^l = 0, \quad (6.3.158)$$

und anschließend gesetzt:  $v^l := v^{l-1} + \lambda_l w^l$ ,  $p^l := p^{l-1} + \lambda_l q^l$ . Der Parameter  $\lambda_l$  wird zur Dämpfung der Iteration verwendet und muss adaptiv angepasst werden. Zur Kosteneinsparung wird der unsymmetrische Term  $\bar{v} \cdot \nabla w^l$  meist nicht in jedem Iterationsschritt neu aufgebaut, sondern es wird im Extremfall mit einer festen Näherung  $\bar{v} \approx v^0$  gearbeitet.

iii) *Newton-Linearisierung*:

Durch Umschreiben erhält die Iteration (6.3.157), (6.3.158) die Gestalt

$$-\nu\Delta v^l + v^{l-1} \cdot \nabla v^l + v^l \cdot \nabla v^{l-1} + \nabla p^l = f + v^{l-1} \cdot \nabla v^{l-1}, \quad (6.3.159)$$

$$\nabla \cdot v^l = 0, \quad (6.3.160)$$

Dies entspricht gerade dem klassischen „Newton-Verfahren“ zur Berechnung einer Nullstelle des Gleichungssystems

$$-\nu\Delta v + v \cdot \nabla v + \nabla p - f = 0, \quad (6.3.161)$$

$$\nabla \cdot v = 0. \quad (6.3.162)$$

### 6.3.3 Algebraische Lösung der linearisierten Probleme

Wir betrachten wieder stabile Stokes-Elemente gegebenenfalls mit Druckstabilisierung. Mit den natürlichen „Knotenwertbasen“ der Ansatzräume  $H_h$  und  $L_h$  schreiben sich z. B. die diskreten Oseen-Gleichungen (6.3.157), (6.3.158) in Form eines Blocksystems

<sup>8</sup>Carl Wilhelm Oseen (1879–1944): Schwedischer theoretischer Physiker; 1909–1939 Prof. für Mechanik und Mathematische Physik in Uppsala, 1933 Vorstand des Nobelinstitutes in Stockholm; formulierte die Grundzüge der Elastizitätstheorie flüssiger Kristalle und führte 1904 die sog. „Oseen-Gleichungen“ als Linearisierung der nichtlinearen Navier-Stokes-Gleichungen ein.

$$\mathcal{A}(\bar{v})\xi = \begin{bmatrix} A(\bar{v}) & B \\ -B^T & \varepsilon C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix}. \quad (6.3.163)$$

für die Geschwindigkeits- und Druckknotenwertvektoren  $\xi = \{x, y\}$  mit den zugehörigen Matrizen sowie den rechten Seiten

$$A(\bar{v}) = ((\nu(\nabla_h \psi_h^i, \nabla_h \psi_h^j) + \bar{v} \cdot \nabla_h \psi_h^i, \psi_h^j) + \dots)_{i,j=1}^{N_H}, \quad B = ((\chi_h g^i, \nabla_h \cdot \psi_h^j)_{i,j=1}^{N_L, N_H}, \\ C = ((\delta \nabla \chi_h^i, \nabla \chi_h^j) + \dots)_{i=1,2}^{N_L}, \quad b = ((f, \psi_h^j)_{j=1}^{N_H}, \quad c = ((g, \chi_h^j)_{j=1}^{N_L}.$$

Dabei steht  $\varepsilon C$  gegebenenfalls für die durch Druckstabilisierung entstehende Matrix. Die Transportstabilisierung ist in der Matrix  $A(\bar{v})$  enthalten, wird aber zur Vereinfachung der Notation nicht explizit dargestellt. Die Matrix  $A(\bar{v})$  ist zwar unsymmetrisch, aber (unter natürlichen Bedingungen) regulär. Folglich kann die Komponente  $x$  wieder aus dem System eliminiert werden:

$$B^T A(\bar{v})^{-1} B y = B^T A(\bar{v})^{-1} b - c, \quad x = A(\bar{v})^{-1} (B y + b). \quad (6.3.164)$$

Mit dem „Schur-Komplement“  $\Sigma := B^T A(\bar{v})^{-1} B$  besteht die Block-Dreieckszerlegung

$$\mathcal{A}(\bar{v}) = \begin{bmatrix} A(\bar{v}) & B \\ -B^T & 0 \end{bmatrix} = \begin{bmatrix} A(\bar{v}) & 0 \\ -B^T & \Sigma \end{bmatrix} \begin{bmatrix} I & A(\bar{v})^{-1} B \\ 0 & I \end{bmatrix}. \quad (6.3.165)$$

Die modifizierte Uzawa-Iteration lautet dann entsprechend:

$$A(\bar{v})x^l = b - B y^{l-1}, \quad (6.3.166)$$

$$M y^l = M y^{l-1} + \theta_l (B^T x^l - c). \quad (6.3.167)$$

Jeder Iterationsschritt erfordert im Wesentlichen die Intervenierung der „Systemmatrix“  $A(\bar{v})$  sowie der „Masse-Matrix“  $M$ , was etwa mit dem GMRES- oder einem Mehrgitter-Verfahren erfolgt. Für die Uzawa-Iteration lässt sich wieder eine gitter-unabhängige Konvergenz beweisen. Wegen der durch den Transportterm verursachten Instabilität im System (6.3.163) ist diese aber mit zunehmender Reynolds-Zahl immer langsamer, so dass das Uzawa-Verfahren schnell zu ineffizient wird. Eine Alternative bietet das Mehrgitter-Verfahren, welches im Folgenden beschrieben wird.

### 6.3.4 Mehrgitter-Verfahren

Die dem „Mehrgitter-Verfahren“ zugrunde liegende Hauptidee ist die schnelle Reduzierung „hoch-frequenter“ Fehleranteile („smoothing“) durch „billige“ Relaxationsverfahren (z. B.: Punkt-Jacobi- oder Gauß-Seidel-Iteration) auf dem feinen Gitter und die Reduzierung des verbleibenden „glatten“, nieder frequenten Fehleranteils durch Defektkorrektur auf größeren Gittern („Grob-gitterkorrektur“). Wir wollen diesen Ansatz kurz beschreiben.

Die Mehrgitteriteration verwendet die Hierarchie von Finite-Elemente-Teilräumen

$$V_0 \subset V_1 \subset \dots \subset V_L,$$

die etwa im Zuge eines systematischen Gitterverfeinerungsprozesses gewonnen wird. Die Verbindung zwischen diesen Räumen wird durch „Fortsetzungsoperationen“ („prolongation“)  $P_{l-1}^l : V_{l-1} \rightarrow V_l$  und „Restriktionsoperationen“ („restriction“)  $R_l^{l-1} : V_l \rightarrow V_{l-1}$  hergestellt. Im FE-Kontext wird hierfür die folgende natürliche Wahl getroffen

$$P_{l-1}^l \quad \text{natürliche Einbettung,} \quad R_l^{l-1} \quad L^2\text{-Projektion.}$$

Der Hauptbestandteil eines Mehrgitteralgorithmus sind die Glättungsiterationen  $S_l : V_l \rightarrow V_l$  auf den verschiedenen Gitterleveln  $0 \leq l \leq L$  ( $l = 0$  korrespondiert zum größten Gitter und  $l = L$  zum feinsten Gitter.). Die Mehrgitteriteration

$$\mathcal{M}\xi = \mathcal{M}(l, z^0, \xi) \tag{6.3.168}$$

auf Level  $l$  zum Anfangswert  $z^0$  und mit  $m_1$  Vor- und  $m_2$  Nachglättungsschritten ist rekursiv definiert wie folgt:

**Mehrgitteralgorithmus**  $\mathcal{M}(l, z^0, \xi)$  für  $l \geq 0$ :

Für  $l = 0$  ist der Mehrgitteralgorithmus gegeben durch einen exakten Löser, d. h.:  $\mathcal{M}(0, z^0, \xi) := \mathcal{A}(\bar{v})^{-1}\xi$ . Für  $l \geq 1$  wird die folgende Iteration durchgeführt:

1. Vorglättung  $m_1$ -mal:  $z^1 := S_l^{m_1} z^0$ .
2. Residuum auf Level  $l$ :  $r^l := \xi - \mathcal{A}_l(\bar{v})z^0$ .
3. Restriktion zu Level  $l - 1$ :  $\tilde{r}^{l-1} := R_l^{l-1} r^l$ .
4. Grobgitterkorrektur beginnend mit  $q^0 := 0$ :  $q := \mathcal{M}(l - 1, q^0, \tilde{r}^{l-1})$ .
5. Prolongation zu Level  $l$ :  $z^2 := z^1 + P_{l-1}^l q$ .
6. Nachglättung  $m_2$ -mal:  $\mathcal{M}(l, z^0, \xi) := S_l^{m_2} z^2$ .

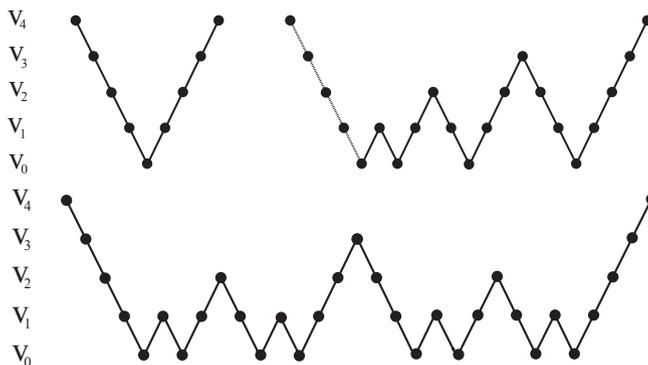


Abbildung 6.7: Schema des Mehrgitter- $V$ -Zyklus (oben links), des  $F$ -Zyklus (oben rechts) und des  $W$ -Zyklus (unten)

Wird der Mehrgitteralgorithmus  $\gamma$ -mal angewendet auf jedem Gitterlevel, so spricht man im Fall  $\gamma = 1$  vom „ $V$ -Zyklus“ und im Fall  $\gamma = 2$  vom „ $W$ -Zyklus“ (Abb. 6.7). Die Varianten mit  $\gamma \geq 3$  sind zu aufwendig und kommen in der Praxis nicht vor. Wird der Mehrgitteralgorithmus nur zur Vorkonditionierung einer robusten äußeren Iteration (z. B. dem GMRES-Verfahren) verwendet, so genügt in der Regel der  $V$ -Zyklus. Wenn der Mehrgitteralgorithmus der primäre Löser ist, erweist sich insbesondere bei unsymmetrischen Problemen der  $W$ -Zyklus als robuster und wird daher bevorzugt. Der sog. „ $F$ -Zyklus“ ist ein attraktiver Kompromiss zwischen  $V$ - und  $W$ -Zyklus.

Das Design eines Mehrgitterverfahrens zur Lösung des Sattelpunkt-Systems (6.3.163) erfordert Sorgfalt. Insbesondere die Wahl der Glättungsiteration ist delikat, da die üblichen Fixpunktiterationen wie „Punkt-Jacobi-“ oder „Gauß-Seidel-Verfahren“ hier nicht funktionieren. Dieses Problem kann auf verschiedene Weise angegangen werden:

1) *Gedämpfter Jacobi-Glätter*: Im Fall  $\varepsilon > 0$  ist die Matrix  $\mathcal{A}(\bar{v})$  schwach-definit, wodurch sogar die Anwendung von standard Methoden wie die gedämpfte Jacobi-Iteration möglich wird. Aber die resultierenden Algorithmen sind wenig robust und die Wahl der Iterationsparameter erfordert große Sorgfalt. Für größere Reynolds-Zahlen wird diese Methode schnell ineffizient und Konvergenz geht verloren.

2) *Block-Gauß-Seidel-Glätter* („*Vanka*<sup>9</sup>-Glätter“): Einen einfachen, aber sehr erfolgreichen Glätter für die Matrix  $\mathcal{A}(\bar{v})$  erhält man durch zellweise Blockung der Freiheitsgrade aller physikalischen Variablen innerhalb einer globalen Gauß-Seidel-Iteration. Dies war ursprünglich für eine Finite-Differenzen-Diskretisierung vorgeschlagen worden. Wir wollen kurz die Realisierung dieser Idee für das nicht-konforme „rotierte“  $Q_1^{\text{rot}}/P_0^{\text{dc}}$  Stokes-Element beschreiben. Die Geschwindigkeits- und Druckfreiheitsgrade zu einer Zelle  $K$  oder einer Gruppe von Zellen werden zusammengefasst, d.h. fortlaufend nummeriert, und die entsprechenden Element-Systemmatrizen durch den Index „loc“ gekennzeichnet. Diese lokalen Freiheitsgrade werden nun im Rahmen einer Gauß-Seidel-Iteration simultan iteriert:

$$S_{\text{loc}}(\bar{v})v_{\text{loc}}^k + B_{\text{loc}}p_{\text{loc}}^k = \text{„bekannt“}, \quad B_{\text{loc}}^T v_{\text{loc}}^k = \text{„bekannt“},$$

wobei  $S_{\text{loc}} := A_{\text{loc}}(\bar{v})$ . Diese Iteration läuft über alle Zellblöcke. Die lokalen Stokes-Probleme haben die Dimension  $d_{\text{loc}} = 9$  (in 2D) bzw.  $d_{\text{loc}} = 19$  (in 3D). Die zugehörigen Matrizen (in 2D) haben die folgende Gestalt:

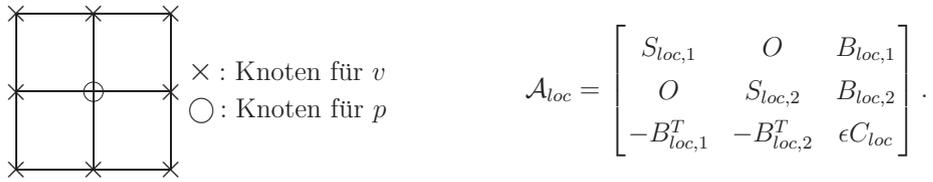
a) Das nichtkonforme  $Q_1^{\text{rot}}/P_0^{\text{nc}}$ -Stokes-Element:

× : Knoten für  $v$   
○ : Knoten für  $p$

$$\mathcal{A}_{\text{loc}} = \begin{bmatrix} S_{\text{loc},1} & O & B_{\text{loc},1} \\ O & S_{\text{loc},2} & B_{\text{loc},2} \\ -B_{\text{loc},1}^T & -B_{\text{loc},2}^T & 0 \end{bmatrix}.$$

<sup>9</sup>Surya Pratap Vanka (????-): US-Amerikanischer Ingenieurwissenschaftler indischer Herkunft; Prof. em. Department of Mechanical Science and Engineering, Univ. of Illinois, Urbana; Beiträge zur numerischen Strömungsmechanik: *Block implicit multigrid solution of Navier-Stokes equations in primitive variables*, J. Comp. Phys. 65, 138–158 (1986).

b) Das konforme  $Q_1^c/Q_1^c$ -Stokes-Element:



Zur Kostenreduktion können die Hauptdiagonalblöcke  $S_{loc,i}$  noch „gelumped“ werden,  $S_{loc,i} \approx D_{loc,i}$ . Weiter wird die Iteration zur Erhöhung der Robustheit gedämpft,  $v_h^{k+1} = v_h^k + \omega(\tilde{v}_h^{k+1} - v_h^k)$  mit  $\omega \in (0, 1)$ .

Wir illustrieren die Eigenschaften dieses Mehrgitteralgorithmus für das  $Q_1^{rot}/P_0^{dc}$ -Stokes-Element anhand von Resultaten für die sog. „lid-driven cavity“ (s. Turek [145]).

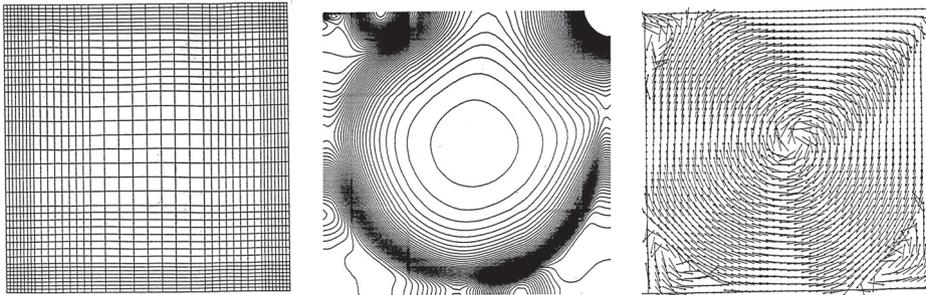


Abbildung 6.8: Resultate für das „driven cavity“-Testproblem: Rechengitter (links), Druckisolines (Mitte) und Geschwindigkeitsvektoren (rechts).

Tabelle 6.1: Mehrgitterkonvergenzraten (2 Vor- und 1 Nachglättungsschritte mit dem „Vanka-Glätter“) und Anzahl der äußeren Fixpunktiterationen auf gleichförmig verfeinerten Gittern.

#Zellen	1600	6400	25600	#Iterationen
Re = 1	0.081	0.096	0.121	4
Re = 100	0.098	0.099	0.130	6
Re = 1000	0.227	0.245	0.168	9
Re = 5000	0.285	0.368	0.370	18

Da der Vanka-Glätter wie ein „Punkt“-Gauß-Seidel-Verfahren auf den Geschwindigkeitsunbekannten wirkt, erwarten wir Probleme im Fall stark gestreckter Gitter. Diese Schwierigkeit kann aber überwunden werden. Die Zellstreckungen sind i. Allg. groß nur in einem kleinen Teil des Rechengebiets, so dass man einen „adaptiven“ Glätter verwenden sollte. Wir werden dabei den konventionellen „Punkt-Glätter“ mit einem robusteren

Glätter kombinieren, nur wo es erforderlich ist, z. B. auf Zellen mit großer Streckung. In diesem Fall werden die Knoten in der Richtung der anisotropen Gitterverfeinerung zu Gruppen zusammengefasst, was im Rahmen der globalen Gauß-Seidel-Glättung zur Bezeichnung „stringwise“ Block-Gauß-Seidel-Glättung führt.

**Bemerkung 6.5 (Mehrgitteriteration als Vorkonditionierer):** Sei  $\mathcal{A}$  die Finite-Elemente-Systemmatrix der linearisierten Navier-Stokes-Gleichungen oder der Approximation derselben. Während die Theorie von Mehrgitterverfahren für skalare Probleme sehr gut entwickelt ist, erscheint die Situation von Systemen weniger klar. Seitens der theoretischen Analyse wissen wir, dass die Verwendung von Mehrgitteriterationen als Vorkonditionierer innerhalb einer äußeren Iteration (z. B. Krylov-Raum-Methode wie GMRES) weniger einschränkende Voraussetzungen erfordert als ihre direkte Verwendung als Löser. Bezeichnet  $\mathcal{M}$  den Operator eines Mehrgitterschritts, so ist es ausreichend, eine obere Schranke für die Kondition des Produkts  $\mathcal{M}\mathcal{A}(\bar{v})$  zu haben, wogegen sonst die Eigenwerte der Iterationsmatrix  $\mathcal{B} := \mathcal{I} - \mathcal{M}\mathcal{A}(\bar{v})$  alle gleichmäßig von eins weg beschränkt werden müssen. Daher verwenden wir meist bei der Lösung von Strömungsproblemen das Mehrgitterverfahren nur als Vorkonditionierer innerhalb einer robusten, äußeren GMRES-Iteration. Hier wird der Mehrgitteroperator  $\mathcal{M}$  als eine approximative Inverse interpretiert,  $\mathcal{M} \approx \mathcal{A}(\bar{v})$ . Dabei ist es natürlich nicht notwendig, diese Matrix explizit zu berechnen; vielmehr genügt es, Matrix-Vektor-Produkte  $\mathcal{M}\xi$  auszuwerten, d. h. den Mehrgitterprozess für eine feste rechte Seite durchzuführen.

## 6.4 Lösung der instationären Navier-Stokes-Gleichungen

Wir haben oben gesehen, dass die *stationären* Navier-Stokes-Gleichungen im Fall von Dirichlet-Randbedingungen für beliebig große Reynolds-Zahl eine Lösung besitzen. Diese Lösung ist aber i. Allg. nicht eindeutig und auch nicht stabil, d. h. physikalisch nicht realisiert. Für höhere Reynolds-Zahlen beobachtet man nur noch instationäre Strömungen, die natürlich auch aus den *instationären* Navier-Stokes-Gleichungen berechnet werden müssen. Hierfür gebräuchliche Diskretisierungsmethoden werden im Folgenden diskutiert.

Der Übergang von *stationärer* zu *instationärer* Lösung erfolgt je nach Strömungskonfiguration bei sehr unterschiedlichen Reynolds-Zahlen. Z. B. geschieht dies in 2D bei der Nischenströmung („lid-driven cavity“; Fig. 6.9) bei  $Re \sim 10^4$ , dagegen bei der Kanalsströmung um einen Zylinder bereits bei  $Re \sim 80$  (Fig. 6.10). Die Größe der Reynolds-Zahl allein sagt also noch nicht notwendig etwas über die Komplexität einer Strömung aus.

Ein weiterer, für die Schwierigkeit der numerischen Simulation instationärer Strömungsvorgänge wichtiger Aspekt ist die Quelle der Instationarität. Wir unterscheiden:

- „erzwungene“ *Instationarität*: Die zeitliche Varianz des Strömungsfelds wird aufgeprägt durch eine gegebene Veränderung der treibenden Daten (z. B. Einströmung, Volumenkraft, u.s.w.).

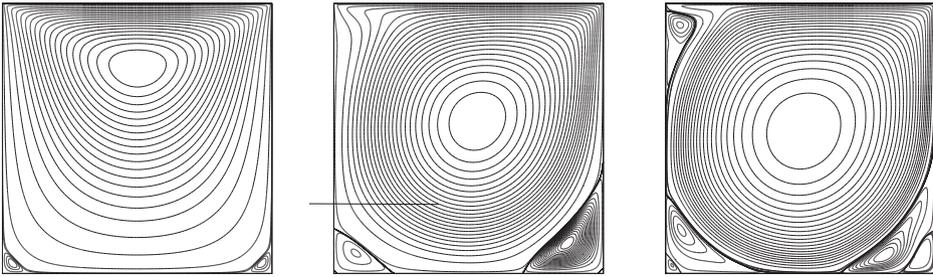


Abbildung 6.9: Stationäre Nischenströmung (getrieben durch konstante Überströmung) für  $Re = 1, 1000, 9000$  (von links nach rechts); s. Turek [145].

- „inhärente“ *Instationarität*: Trotz stationärer Daten zeigt die Strömung aufgrund einer sog. „Hopf-Instabilität“ ein instationäres, oft periodisches Verhalten. Ein typisches Beispiel ist die „von Kármánsche Wirbelstraße“ in Abb. 6.10.

Die numerische Berechnung einer erzwungenen instationären Strömung entspricht im wesentlichen der mehrfachen Berechnung einer stationären Strömung, während die Berechnung einer inhärent instationären Strömung sehr schwierig und aufwendig sein kann.



Abbildung 6.10: Instationäre „von Kármánsche Wirbelstraße“ bei  $Re = 100$  (Experiment im Strömungskanal).

#### 6.4.1 Zeitschrittschemata

Das System der instationären Navier-Stokes-Gleichungen

$$\partial_t v - \nu \Delta v + v \cdot \nabla v + \nabla p = f, \quad \nabla \cdot v = 0, \quad (6.4.169)$$

bildet in der Sprache der Theorie gewöhnlicher Differentialgleichungen ein „differentiell-algebraisches System“ („DAE“). Die „algebraische“ Nebenbedingung  $\nabla \cdot v = 0$  bestimmt eine lineare Mannigfaltigkeit im sog. „Phasenraum“  $H = H_0^1(\Omega)^d$ , auf der sich die Dynamik abspielt.

Zur numerischen Lösung dieses Problems betrachten wir eine Diskretisierung nach dem „Rothe<sup>10</sup>-Verfahren“, d. h.: Es wird zunächst in der Zeit und dann in jedem einzelnen Zeitschritt das auftretende quasi-stationäre Problem im Ort diskretisiert. Seien

$$0 = t_0 < t_1 < \dots < t_m < \dots < t_M = T$$

diskrete Zeitpunkte,  $k_m = t_m - t_{m-1}$  die einzelnen Zeitschrittweiten und  $k := \sup_m k_m$ . Zur Bestimmung von Näherungen  $\{v^m, p^m\} \approx \{v(\cdot, t_m), p(\cdot, t_m)\}$  betrachten wir zunächst das sog. „Einschritt- $\theta$ -Schema“ mit dem Parameter  $\theta \in [0, 1]$ :

$$k_m^{-1}\{v^m - v^{m-1}\} - \nu\Delta\{\theta v^m + (1-\theta)v^{m-1}\} + \theta v^m \cdot \nabla v^m + (1-\theta)v^{m-1} \cdot \nabla v^{m-1} \\ + \theta \nabla p^m + (1-\theta)\nabla p^{m-1} = \theta f^m + (1-\theta)f^{m-1},$$

mit  $f^m := f(t_m)$ , wobei noch die Nebenbedingung

$$\nabla \cdot v^m = 0 \tag{6.4.170}$$

einzuhalten ist. Dieses Schema ist generisch „implizit“. Selbst wenn, etwa im Fall  $\theta = 0$  („explizites Euler-Verfahren“), das Geschwindigkeitsfeld rein explizit fortgepflanzt wird, erfordert die Nebenbedingung (6.4.170) die Lösung eines impliziten Gleichungssystems. Wir werden diesen Schritt weiter unten noch im Zusammenhang mit sog. „Projektionsverfahren“ diskutieren.

Da das *explizite* Euler-Verfahren (Fall  $\theta = 0$ ) nur von erster Ordnung ist und außerdem einer Schrittweitenbedingung der Form

$$k \leq c\nu^{-1}h_{\min}^2 \tag{6.4.171}$$

unterliegt, wird es hier nicht weiter diskutiert. Wir konzentrieren uns im Folgenden auf voll *implizite* Verfahren, d. h. auf den Parameterbereich  $\frac{1}{2} < \theta \leq 1$ . Von besonderem Interesse sind die Spezialfälle  $\theta = 1$  („implizites Euler-Schema“ erster Ordnung) und  $\theta = \frac{1}{2}$  („Crank<sup>11</sup>-Nicolson<sup>12</sup>-Verfahren (Trapezregel) zweiter Ordnung). In jedem Zeitschritt z. B. des impliziten Euler-Schemas ist ein quasi-stationäres Problem der Form

$$k_m^{-1}v^m - \nu\Delta v^m + v^m \cdot \nabla v^m + \nabla p^m = \text{„bekannt“} \tag{6.4.172}$$

$$\nabla \cdot v^m = 0, \tag{6.4.173}$$

zu lösen. Mit derselben Argumentation wie im Fall der einfachen stationären Navier-Stokes-Gleichungen zeigt man auch hier die Existenz (für hinreichend kleine Schrittweite  $k$  eindeutiger) Lösungen  $\{v^m, p^m\}$ .

<sup>10</sup>Erich Rothe (1895–1988): Deutscher Mathematiker; Promotion und Habilitation in Berlin (1928), danach Assistent in Breslau, wegen seiner jüdischen Herkunft Entlassung aus dem Staatsdienst und 1937 Emigration in die USA, ab 1955 Prof. an der University of Michigan, Ann Arbor; Arbeitsgebiete Analysis und Nichtlineare Funktionalanalysis (Abbildungsgrad).

<sup>11</sup>John Crank (1916–2006): Englischer Mathematiker; Prof. an der Brunel University, Uxbridge; Beiträge zur Numerik partieller Differentialgleichungen, bekannt durch das „Crank-Nicolson-Verfahren“.

<sup>12</sup>Phyllis L. Nicolson (1917–1968): Englische Physikerin; Lecturer an der Univ. Leeds und an der Univ. Manchester; bekannt durch das „Crank-Nicolson-Verfahren“.

Das implizite Euler-Schema ist „stark A-stabil“, während das Crank-Nicolson-Schema nur einfach „A-stabil“ ist. Diese Eigenschaften werden sichtbar, wenn man die Schemata auf die skalare „Testgleichung“

$$y'(t) = qy(t), \quad t \geq 0, \quad q \in \mathbb{C}, \quad y(t) = e^{qt}y(0), \quad (6.4.174)$$

anwendet. Ein Zeitschritt  $y^{m-1} \rightarrow y^m$  mit Schrittweite  $k$  ist dabei gegeben in der Form

$$y^m = R(kq)y^{m-1} \quad (6.4.175)$$

mit einem „Verstärkungsfaktor“  $R(qk)$ . Für diesen gilt

$$R(qk) = e^{kq} + \mathcal{O}(k^{r+1}). \quad (6.4.176)$$

mit der „Ordnung“  $r \geq 1$ , des Verfahrens. Im Folgenden setzen wir  $\lambda := qk$ . Für das Einschritt- $\theta$ -Schema gilt:

$$R_\theta(\lambda) = \frac{1 + (\theta - \frac{1}{2})\lambda}{1 - \theta\lambda} = e^\lambda + (\theta - \frac{1}{2})|\lambda|^2 + \mathcal{O}(|\lambda|^3), \quad |\lambda| \leq 1.$$

Das robuste implizite Euler-Schema ( $\theta = 1$ ) ist sehr „dissipativ“, d. h. dämpft besonders zeitlich hoch-frequente Anteile in der Lösung sehr stark, und ist daher wenig geeignet für die Berechnung inhärent instationärer Strömungen. Im Gegensatz dazu ist das Crank-Nicolson-Schema nur sehr gering dissipativ, leidet aber unter einer geringen Stabilität gegenüber hochfrequenten Störungen (fehlende *starke* A-Stabilität). Diese Effekte werden wir unten anhand von Beispielen illustrieren.

Ein „gutes“ Zeitschrittverfahren charakterisiert durch seine Approximation  $R(\lambda)$  der Exponentialfunktion sollte die folgenden Eigenschaften haben:

- *A-Stabilität* ( $\Rightarrow$  lokale Konvergenz):  $|R(\lambda)| \leq 1$ .
- *Globale Stabilität* ( $\Rightarrow$  globale Konvergenz):  $\overline{\lim}_{\text{Re}\lambda \rightarrow \infty} |R(\lambda)| \leq 1 - \mathcal{O}(k)$ .
- *Starke A-Stabilität* ( $\Rightarrow$  Glättungseigenschaft):  $\overline{\lim}_{\text{Re}\lambda \rightarrow \infty} |R(\lambda)| \leq 1 - \delta < 1$ .
- *Geringe Dissipation* ( $\Rightarrow$  Energie-Erhaltung):  $|R(\lambda)| = 1 - \mathcal{O}(|\text{Im}(\lambda)|)$ ,  $\text{Re}(\lambda) \rightarrow 0$ .

Alternative Verfahren höherer Ordnung basieren auf sog. „diagonal impliziten“ Runge-Kutta-Formeln oder auf sog. „Rückwärtsdifferenzenformeln“. In der Praxis werden aber hauptsächlich das Crank-Nicolson-Verfahren und Varianten davon verwendet. Eine sehr erfolgreiche Variante ist das sog. „Teilschritt- $\theta$ -Schema“ (Fractional Step- $\theta$  Scheme“) nach Glowinski/Periaux [112, 111]. Zu seiner kompakten Beschreibung führen wir die folgende an der algebraischen Schreibweise orientierte Notation ein:

$$A := -\nu\Delta, \quad N^m := v^m \cdot \nabla, \quad A^m := A + N^m, \quad B := \nabla, \quad B^T := \nabla \cdot .$$

**Teilschritt- $\theta$ -Verfahren:** (drei Teilschritte:  $t_{m-1} \rightarrow t_{m-1+\theta} \rightarrow t_{m-\theta} \rightarrow t_m$ )

$$\begin{aligned} (1) & \begin{cases} [M + \alpha\theta k A^{m-1+\theta}]v^{m-1+\theta} + \theta k B p^{m-1+\theta} = [M - \beta\theta k A^{m-1}]v^{m-1} + \theta k f^{m-1}, \\ -B^T v^{m-1+\theta} = 0, \end{cases} \\ (2) & \begin{cases} [M + \beta\theta' k A^{m-\theta}]v^{m-\theta} + \theta' k B p^{m-\theta} = [M - \alpha\theta' k A^{m-1+\theta}]v^{m-1+\theta} + \theta' k f^{m-\theta}, \\ -B^T v^{m-\theta} = 0, \end{cases} \\ (3) & \begin{cases} [M + \alpha\theta k A^m]v^m + \theta k B p^m = [M - \beta\theta k A^{m-\theta}]v^{m-\theta} + \theta k f^{m-\theta}, \\ -B^T v^m = 0, \end{cases} \end{aligned}$$

wobei  $\theta \in (0, \frac{1}{2})$ ,  $\theta' = 1 - 2\theta$ ,  $\alpha \in [0, 1]$ , und  $\beta = 1 - \alpha$ . Dieses Verfahren entspricht einer rationalen Approximation der Exponentialfunktion der Form

$$R_\theta(\lambda) = \frac{(1 + \alpha\theta'\lambda)(1 + \beta\theta\lambda)^2}{(1 - \alpha\theta\lambda)^2(1 - \beta\theta'\lambda)} = e^\lambda + \mathcal{O}(|\lambda|^3), \quad |\lambda| \leq 1.$$

Für  $\theta = 1 - 1/\sqrt{2} = 0.292893\dots$  ist das Verfahren von zweiter Ordnung und für  $\alpha > \frac{1}{2}$  stark A-stabil:

$$\overline{\lim}_{\operatorname{Re}\lambda \rightarrow \infty} |R_\theta(-\lambda)| = \frac{\beta}{\alpha} < 1. \quad (6.4.177)$$

Für die spezielle Wahl  $\alpha = (1 - 2\theta)/(1 - \theta) = 0.585786\dots$  gilt  $\alpha\theta = \beta\theta'$ , so dass die Hauptteile der in jedem der drei Teilschritte zu bildenden Systemmatrizen gleich sind.

Dieses Verfahren war ursprünglich in Form eines „Operatorsplitting-Schemas“ vorgeschlagen worden, um die Komplikationen „Nichtlinearität“ und „Inkompressibilität“ innerhalb des Zyklus  $t_{m-1} \rightarrow t_{m-1+\theta} \rightarrow t_{m-\theta} \rightarrow t_m$  zu separieren:

$$\begin{aligned} (1) & \begin{cases} [M + \alpha\theta k A]v^{m-1+\theta} + \theta k B p^{m-1+\theta} = [M - \beta\theta k A^{m-1}]v^{m-1} + \theta k f^{m-1} - \theta k N^{m-1}v^{m-1}, \\ -B^T v^{m-1+\theta} = 0, \end{cases} \\ (2) & \begin{cases} [M + \beta\theta' k A^{m-\theta}]v^{m-\theta} + \theta' k B p^{m-\theta} = [M - \alpha\theta' k A^{m-1+\theta}]v^{m-1+\theta} + \theta' k f^{m-\theta}, \\ \dots, \end{cases} \\ (3) & \begin{cases} [M + \alpha\theta k A]v^m + \theta k B p^m = [M - \beta\theta k A^{m-\theta}]v^{m-\theta} + \theta k f^{m-\theta} - \theta k N^{m-\theta}v^{m-\theta}, \\ -B^T v^m = 0. \end{cases} \end{aligned}$$

Das Teilschritt- $\theta$ -Verfahren besitzt aber auch wesentliche Vorzüge als ein einfaches Zeitschritt-schema. Es ist von zweiter Ordnung und stark „A-stabil“ für jede Wahl von  $\alpha \in (\frac{1}{2}, 1]$  (im Gegensatz zum Crank-Nicolson-Verfahren), und sein Verstärkungsfaktor hat Betrag  $|R(-\lambda)| \approx 1$  für  $\lambda$  nahe bei der imaginären Achse (z. B.:  $|R(-0.8i)| = 0.9998\dots$ ), was eine sehr geringe Dissipation bedeutet. Hinsichtlich Genauigkeit entspricht ein Zyklus des Teilschritt- $\theta$ -Verfahrens in etwa drei Crank-Nicolson-Schritten.

## Numerischer Test

Wir präsentieren einige Resultate numerischer Tests aus Müller et al. [125] für das implizite Euler-, das Crank-Nicolson- und das Teilschritt- $\theta$ -Verfahren. Die Strömungskonfiguration ist in Abb. 6.11 gezeigt: Strömung um eine schräg angestellte Platte in einem Kanal für

$Re = 500$ . Die Ortsdiskretisierung erfolgt mit dem oben beschriebenen, nicht konformen, „rotiert-bilinearen“  $Q_1^{\text{rot}}/P_0^{\text{dc}}$ -Stokes-Element auf einem gleichförmig verfeinerten Gitter mit ca. 13.000 Zellen. Die Rechnung erstreckt sich über das Zeitintervall  $I = [0, 60]$ .

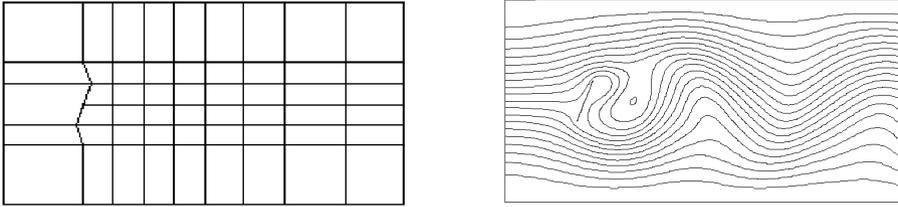


Abbildung 6.11: Konfiguration des Plattenumströmungstests: grobes Ausgangsgitter (links) und Stromlinien (rechts).

Der erste Test betrifft die Genauigkeit. Abb. 6.12 zeigt, dass das implizite Euler-Schema (BE) nicht geeignet zur Berechnung freier, zeitperiodischer Strömungen ist, während das Crank-Nicolson-Schema (CN) und das Teilschritt- $\theta$ -Schema (FS) etwa gleich gut sind. Diese gleiche Approximationsqualität wird auch bestätigt durch den Vergleich anhand einer sensitiven Fehlerquantität, der mittleren Druckdifferenz zwischen Vorder- und Rückseite der Platte in Abb. 6.13:

$$\Delta P := \int_{\Gamma_{\text{vorn}}} p \, ds - \int_{\Gamma_{\text{hinten}}} p \, ds.$$

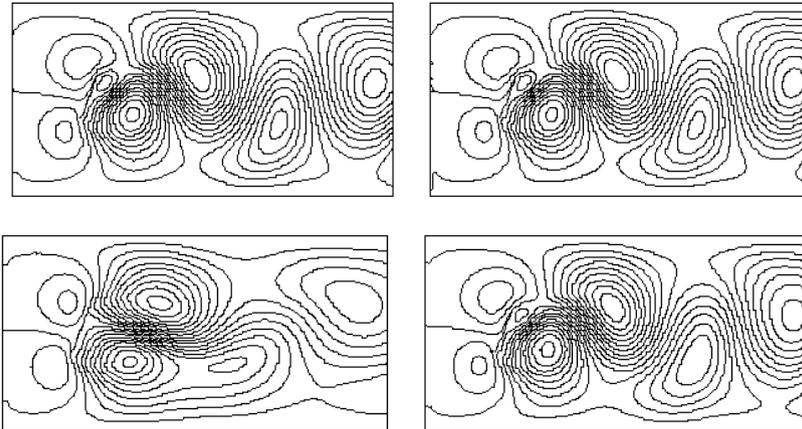


Abbildung 6.12: Druck-Isolinien: BE-Schema mit  $3k = 1$  (oben links), BE-Schema mit  $3k = 0.1$  (oben rechts), CN-Schema mit  $3k = 1$  (unten links), FS-Schema mit  $k = 1$  (unten rechts).

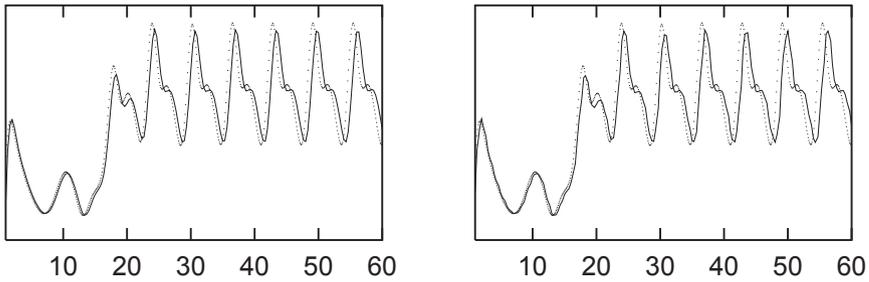


Abbildung 6.13: Mittlere Druckdifferenz mit voll-impliziter Behandlung der Nichtlinearität: CN-Schema mit  $3k = 0.33$  (links), FS-Schema mit  $k = 0.33$  (rechts), beide verglichen mit einer Referenzlösung (gestrichelte Linie)

Schließlich vergleichen wir die Robustheit der Verfahren; Abb. 6.14 zeigt den Mangel an Robustheit des Crank-Nicolson-Schemas (in Verbindung mit linearer Extrapolation zur Approximation der Nichtlinearität) für größere Zeitschritte.

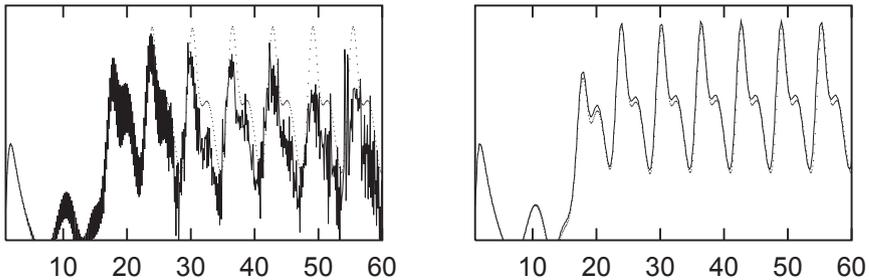


Abbildung 6.14: Mittlere Druckdifferenz mit linearer Zeitextrapolation der Nichtlinearität (ohne Defektkorrektur): CN-Schema mit  $3k = 0.11$  (links), FS-Schema mit  $k = 0.11$  (rechts), beide verglichen mit einer Referenzlösung (gestrichelte Linie)

## 6.4.2 Projektionsverfahren

Das klassische „Chorinsche Projektionsverfahren“ nach A. Chorin<sup>13</sup> (1967) ist physikalisch motiviert und umgeht das Problem der „teuren“ Lösung eines Sattelpunktsystems in jedem Zeitschritt durch die Verwendung eines sog. „Operator-Splittings“. Verwandte Vorgehensweisen findet man auch bei der Behandlung von DAEs in der Numerik von

<sup>13</sup>Alexandre Joel Chorin (1938–): US-Amerikanischer Mathematiker (geb. in Warschau); Studium zunächst in Lausanne (Diplomingenieur), dann in New York, Promotion bei P. Lax (1966), Prof. für Mathematik an der Univ. of California, Berkeley, seit 1989 auch Prof. an der Univ. Tel Aviv; fundamentale Arbeiten zur numerischen Lösung der Navier-Stokes-Gleichungen: Projektionsmethode (parallel zu R. Temam, 1968), Methode der künstlichen Kompressibilität, Vortex-Method.

gewöhnlichen Differentialgleichungen. Dabei wird in jedem Zeitschritt zunächst einmal die Lösung gemäß des ODE-Anteils, unter Vernachlässigung der algebraischen Nebenbedingung fortgepflanzt und dann auf die Mannigfaltigkeit zurückprojiziert. Die sachgerechte Wahl dieser Projektion ist im Rahmen von PDEs wesentlich kritischer als bei ODEs.

Das ursprüngliche Chorin-Verfahren basiert auf dem impliziten Euler-Schema:

$$k^{-1}\{v^m - v^{m-1}\} - \nu \Delta v^m + v^m \cdot \nabla v^m + \nabla p^m = f^m, \quad \nabla \cdot v^m = 0. \quad (6.4.178)$$

Ausgehend vom Lösungspaar  $\{v^{m-1}, p^{m-1}\}$  zum vorausgehenden Zeitpunkt  $t_{m-1}$  wird zunächst das Geschwindigkeitsfeld fortgepflanzt (bei Vernachlässigung des Drucks):

$$k^{-1}\{\tilde{v}^m - v^{m-1}\} - \nu \Delta \tilde{v}^m + \tilde{v}^m \cdot \nabla \tilde{v}^m = f^m. \quad (6.4.179)$$

Die so gewonnene Zwischenlösung  $\tilde{v}^m \in H$  ist i. Allg. nicht divergenzfrei. Dies wird nun korrigiert, indem man  $\tilde{v}^m$  orthogonal auf die Mannigfaltigkeit der divergenz-freien Vektorfelder projiziert. Diese Projektion erfolgt aus Kostengründen im  $L^2$ -Sinne:

$$v^m = P\tilde{v}^m \in J_0(\Omega) : \quad (v^m, \varphi) = (\tilde{v}^m, \varphi) \quad \forall \varphi \in J_0(\Omega).$$

Dabei ist  $J_0(\Omega)$  der schon oben eingeführte Teilraum der im „schwachen“ Sinne divergenz-freien Vektorfunktionen:

$$J_0(\Omega) := \{v \in L^2(\Omega)^d \mid (v, \nabla \chi) = 0 \quad \forall \chi \in C^1(\overline{\Omega})\}.$$

Die Funktionen in  $J_0(\Omega)$  erfüllen also im „schwachen“ (bzw. distributionellen) Sinne  $\nabla \cdot v = 0$  und  $n \cdot v|_{\partial\Omega} = 0$ . Man erhält  $J_0(\Omega)$  auch als Abschluss des Raumes  $\Phi(\Omega) = \{\varphi \in C_0^\infty(\Omega)^d \mid \nabla \cdot \varphi = 0\}$  bzgl. der  $L^2$ -Norm.

Zur Berechnung dieser Projektion betrachten wir die Neumannsche Randwertaufgabe

$$\Delta q^m = k^{-1} \nabla \cdot \tilde{v}^m \text{ in } \Omega, \quad \partial_n q^m|_{\partial\Omega} = 0. \quad (6.4.180)$$

Wegen

$$\int_{\Omega} \nabla \cdot \tilde{v}^m dx = \int_{\partial\Omega} n \cdot \tilde{v}^m ds = 0$$

existiert eine (schwache) Lösung  $q \in H^1(\Omega)$ , welche durch die Zusatzbedingung  $q^m \in L_0^2(\Omega)$  eindeutig bestimmt ist. Die Funktion

$$v^m := \tilde{v}^m - k \nabla q^m \in H^1(\Omega)^d$$

ist dann wegen

$$(v^m, \varphi) = (\tilde{v}^m, \varphi) - k(\nabla q^m, \varphi) = (\tilde{v}^m, \varphi) + k(q^m, \nabla \cdot \varphi) = (\tilde{v}^m, \varphi), \quad \varphi \in J_0(\Omega),$$

die gesuchte  $L^2$ -Projektion von  $\tilde{v}^m$  in  $J_0(\Omega)$ . Sie erfüllt

$$\nabla \cdot v^m = \nabla \cdot \tilde{v}^m - k \nabla \cdot \nabla q^m = \nabla \cdot \tilde{v}^m - k \Delta q^m = 0 \quad (6.4.181)$$

und

$$n \cdot v^m|_{\partial\Omega} = -\partial_n q^m|_{\partial\Omega} = 0.$$

I. Allg. ist aber  $\tau \cdot v^m|_{\partial\Omega} = \partial_\tau q^m|_{\partial\Omega} \neq 0$ , d. h.: Die projizierte Näherung  $v^m$  ist zwar divergenzfrei, erfüllt aber nicht vollständig die Haftbedingung am Gebietsrand. Trotz dieses Mankos wird  $v^m$  als Approximation zu  $v(\cdot, t_m)$  akzeptiert. Weiter wird die Hilfsfunktion  $q^m$  als neuer Druck verwendet

$$p^m := q^m.$$

Im nächsten Zeitschritt wird dann analog verfahren. Ein Zeitschritt  $t_{m-1} \rightarrow t_m$  zerfällt also im wesentlichen in die Lösung eines d-dimensionalen (nichtlinearen) Konvektions-Diffusions-Systems („Impulstransport-Gleichung“) für die Geschwindigkeit  $\tilde{v}^m$  und eines skalaren Neumann-Problems („Druckkorrekturgleichung“) für den Druckterm  $q^m$ . Durch die Modifikation

$$k^{-1}\{\tilde{v}^m - v^{m-1}\} - \nu\Delta\tilde{v}^m + v^{m-1} \cdot \nabla\tilde{v}^m = f^m \quad (6.4.182)$$

wird die Impulstransport-Gleichung linearisiert. Wir fassen das „klassische“ Chorinsche Projektionsverfahren wie folgt zusammen:

### Chorinsches Projektionsverfahren:

i) Impliziter „Transport-Diffusions-Schritt“:

$$k^{-1}\{\tilde{v}^m - v^{m-1}\} - \nu\Delta\tilde{v}^m + \tilde{v}^m \cdot \nabla\tilde{v}^m = f^m \text{ in } \Omega, \quad \tilde{v}^m|_{\partial\Omega} = 0.$$

ii) Druck-Poisson-Gleichung:

$$\Delta q^m = k^{-1}\nabla \cdot \tilde{v}^m \text{ in } \Omega, \quad \partial_n q^m|_{\partial\Omega} = 0.$$

iii) Druck- und Geschwindigkeitskorrektur:

$$v^m := \tilde{v}^m - k\nabla q^m, \quad p^m := q^m.$$

Dieses Verfahren wurde in Verbindung mit einer Finite-Differenzen-Diskretisierung im Ort erfolgreich zur (qualitativen) Lösung der instationären Navier-Stokes-Gleichungen eingesetzt. Es bestehen aber mehrere zu klärende Fragen:

1. Die physikalische Rolle der Zwischenlösung  $\tilde{v}^m$  sowie des Druckterms  $q^m$  ist unklar.
2. Es zeigt sich ein besonders großer Fehler in den Spannungen  $\sigma_h^m := 2\mu\varepsilon(v_h^m) - p_h^m I$  am Gebietsrand.
3. Das Verfahren ist ungeeignet zur Berechnung „stationärer Limiten“ für  $t_m \rightarrow \infty$ .
4. Das Verfahren besitzt nur eine geringe Zeitgenauigkeit  $\mathcal{O}(k)$ .
5. Das Verfahren funktioniert auch in Verbindung mit „instabiler“ Ortsdiskretisierung, z. B. mit den  $Q_1^c/P_0^{\text{dc}}$ - oder  $Q_1^c/Q_1^c$ -Stokes-Elementen ohne explizite Druckstabilisierung.

Wir wollen die Quellen und Lösungen dieser Probleme im Einzelnen diskutieren. Zunächst kann aufgrund seiner Konstruktion das Verfahren höchstens von erster Ordnung  $\mathcal{O}(k)$  sein; tatsächlich ergibt eine globale Fehleranalyse sogar nur:

$$\|\nabla(v(t_m) - v^m)\| + \|p(t_m) - p^m\| = \mathcal{O}(\sqrt{k}).$$

Ferner wird die Folge der Iterierten  $(v^m)_{m \in \mathbb{N}}$  auch im Falle der Konvergenz der exakten Lösung gegen eine stationäre Lösung,  $v(t) \rightarrow v^\infty$  ( $t \rightarrow \infty$ ), wegen des zwischengeschalteten Projektionsschrittes i. Allg. ebenfalls nicht gegen diesen Limes konvergieren. Selbst wenn mit der exakten (stationären) Lösung  $v^0 := v^\infty$ ,  $p^0 := p^\infty$ , gestartet wird, erzeugt das Verfahren für  $\nabla p^\infty \neq 0$  im ersten Schritt eine leicht gestörte Näherung

$$\begin{aligned} \tilde{v}^1 &= kf + v^\infty + k\nu\Delta\tilde{v}^1 - kv^\infty \cdot \nabla\tilde{v}^1 \neq v^\infty, \\ -\Delta q^1 &= \nabla \cdot \tilde{v}^1 \neq 0, \\ v^1 &= \tilde{v}^1 - k\nabla q^1 \neq v^\infty. \end{aligned}$$

Die Frage nach der Relevanz der Druckapproximation war lange Zeit ungeklärt. Gelegentlich wurde sogar in Zweifel gezogen, dass  $p^n$  überhaupt eine brauchbare Näherung zu  $p(\cdot, t_m)$  darstellt. In der Tat ist in der Regel die Geschwindigkeitsapproximation recht gut, die Druckapproximation aber besonders entlang des Randes fast unbrauchbar schlecht. Diese Einschätzung wurde erst korrigiert durch die Beobachtung, dass der große Druckfehler (sowie die damit zusammenhängende große Divergenz  $\nabla \cdot \tilde{v}^m$ ) auf eine sehr kleine Grenzschicht entlang des festen Randes beschränkt ist und ins Gebietsinnere exponentiell abfällt. Genauer gilt für die Chorin-Näherung:

$$|p^m(x) - p(x, t_m)| \approx k + \exp\left(\frac{d(x)}{\sqrt{\nu k}}\right)\sqrt{k}, \quad (6.4.183)$$

wobei  $d(x) := \min\{|x - y| : y \in \partial\Omega\}$ . Die Breite der „Druckgrenzschicht“ verhält sich demnach wie  $\mathcal{O}(\sqrt{\nu k})$ . Auf jedem Teilgebiet  $\Omega' \subset \Omega$  mit positivem Abstand zum Rand  $\partial\Omega$  hat man sogar die optimale Approximationsordnung

$$\|p^m - p(t_m)\|_{\Omega'} = \mathcal{O}(k). \quad (6.4.184)$$

Dieses Resultat ist aber nur von zweifelhaftem Wert, da der Druck als Bestandteil der Spannungskräfte in der Praxis gerade entlang fester Ränder zu bestimmen ist.

Wir wollen die beim Projektionsverfahren wirksame Quelle des großen Fehlers am Rand durch eine Modellbetrachtung zu verstehen versuchen. Einsetzen von  $\tilde{v}^{m-1} = v^{m-1} + k\nabla q^{m-1}$  und  $p^{m-1} = q^{m-1}$  in die Impulstransportgleichung (6.4.179) ergibt in diesem Fall zusammen mit der Beziehung (6.4.181):

$$k^{-1}\{\tilde{v}^m - \tilde{v}^{m-1}\} - \nu\Delta\tilde{v}^m + \tilde{v}^m \cdot \nabla\tilde{v}^m + \nabla p^{m-1} = f^m, \quad (6.4.185)$$

$$\nabla \cdot \tilde{v}^m - k\Delta p^m = 0. \quad (6.4.186)$$

Dies kann wie folgt interpretiert werden: Das im Projektionsverfahren erzeugte Paar  $\{\tilde{v}^m, p^m\} \in H \times L$  entspricht einer aus der impliziten Euler-Approximation des Navier-Stokes-Problems mit *Druckstabilisierung* (Stabilisierungsparameter  $k$ ) bei expliziter Ankopplung des Drucks gewonnenen Näherung. Dies legt es nahe, zunächst unter Vernachlässigung der Zeitabhängigkeit sowie Nichtlinearität den Effekt des Störungsterms

$-\varepsilon\Delta p$  im *linearen* Modell

$$-\nu\Delta v_\varepsilon + \nabla p_\varepsilon = f, \quad (6.4.187)$$

$$\nabla \cdot v_\varepsilon - \varepsilon\Delta p_\varepsilon = 0, \quad (6.4.188)$$

zu untersuchen. Dieses Problem besitzt natürlich für jedes  $\varepsilon > 0$  eine eindeutige Lösung  $\{v_\varepsilon, p_\varepsilon\} \in H \times L$ .

**Satz 6.12 (Druckstabilisierung):** Für den durch die Druckstabilisierung in (6.4.188) hervorgerufenen Fehler gilt die globale Abschätzung

$$\nu\|\nabla(v - v_\varepsilon)\| + \|p - p_\varepsilon\| \leq c\sqrt{\nu\varepsilon}\|\nabla p\|, \quad (6.4.189)$$

$$\nu\|v - v_\varepsilon\| + \|p - p_\varepsilon\|_{-1} \leq c\nu\varepsilon\|\nabla p\|, \quad (6.4.190)$$

mit der „dualen“ Norm  $\|p\|_{-1} := \sup_{\chi \in H_0^1(\Omega)} \{(p, \chi), \|\nabla \chi\| = 1\}$ , sowie auf Teilgebieten  $\Omega' \subset \Omega$  mit positivem Abstand zu Rand  $\partial\Omega$  die „innere“ Abschätzung

$$\|p - p_\varepsilon\|_{\Omega'} \leq c(\Omega')\nu\varepsilon\{\|\nabla p\| + \|\Delta p\|\}. \quad (6.4.191)$$

**Beweis:** i) Die Fehler  $e^v := v_\varepsilon - v$  und  $e^p := p_\varepsilon - p$  erfüllen die Beziehungen

$$-\nu\Delta e^v + \nabla e^p = 0 \text{ in } \Omega, \quad e^v|_{\partial\Omega} = 0, \quad (6.4.192)$$

$$\nabla \cdot e^v - \varepsilon\Delta e^p = \varepsilon\Delta p \text{ in } \Omega, \quad \partial_n e^p|_{\partial\Omega} = -\partial_n p|_{\partial\Omega}. \quad (6.4.193)$$

bzw. in variationeller Form:

$$\begin{aligned} \nu(\nabla e^v, \nabla \varphi) - (e^p, \nabla \cdot \varphi) &= 0 \quad \forall \varphi \in H, \\ (\nabla e^v, \chi) + \varepsilon(\nabla e^p, \nabla \chi) &= \varepsilon(\nabla p, \nabla \chi) \quad \forall \chi \in H^1(\Omega). \end{aligned}$$

Wir testen hier mit  $\varphi := e^v$  bzw. mit  $\chi := e^p$  und kombinieren die Resultate zu

$$\nu\|\nabla e^v\|^2 + \varepsilon\|\nabla e^p\|^2 = -\varepsilon(\nabla p, \nabla e^p).$$

Dies impliziert die erste Fehlerabschätzung

$$\nu\|\nabla e^v\| + \sqrt{\nu\varepsilon}\|\nabla e^p\| \leq c\sqrt{\nu\varepsilon}\|\nabla p\|. \quad (6.4.194)$$

ii) Zur Abschätzung des Druck- $L^2$ -Fehlers verwenden wir die Stabilität des div-Operators:

$$\beta\|e^p\| \leq \sup_{\varphi \in H} \frac{(e^p, \nabla \cdot \varphi)}{\|\nabla \varphi\|} = \sup_{\varphi \in H} \frac{\nu(\nabla e^v, \nabla \varphi)}{\|\nabla \varphi\|} \leq \nu\|\nabla e^v\|.$$

Dies vervollständigt den Beweis von (6.4.189).

iii) Als nächstes bedienen wir uns eines Dualitätsarguments. Sei  $\{z, r\} \in H \times L$  die (eindeutige) Lösung des Stokes-Problems

$$-\nu\Delta z + \nabla r = e^v, \quad \nabla \cdot z = 0 \text{ in } \Omega, \quad z|_{\partial\Omega} = 0.$$

Hierfür gilt die a priori Abschätzung  $\|\nabla r\| \leq c\|e^v\|$ . Damit erhalten wir ( $\nabla \cdot z = 0$ )

$$\|e^v\|^2 = \nu(\nabla e^v, \nabla z) - (\nabla \cdot e^v, r) = \varepsilon(\nabla(e^p + p), \nabla r) \leq \varepsilon\|\nabla p_\varepsilon\|\|\nabla r\|,$$

und folglich unter Beachtung von (6.4.194):

$$\|e^p\| \leq c\varepsilon\|\nabla p\|. \quad (6.4.195)$$

iv) Zur Abschätzung des Druck- $H^{-1}$ -Fehlers verwenden wir wieder die Stabilität des div-Operators in einer modifizierten Form wie folgt:

$$\beta\|e^p\|_{-1} \leq \sup_{\varphi \in H \cap H^2(\Omega)^d} \frac{(e^p, \nabla \cdot \varphi)}{\|\nabla^2 \varphi\|} = \sup_{\varphi \in H \cap H^2(\Omega)^d} \frac{\nu(\nabla e^v, \nabla \varphi)}{\|\nabla^2 \varphi\|} \leq c\nu\|e^v\|.$$

Dies vervollständigt den Beweis von (6.4.190).

v) Zum Beweis der „inneren“ Abschätzung (2.5.172) wählen wir eine „Abschneidefunktion“  $\sigma \in C_0^\infty(\Omega)$  mit der Eigenschaft  $0 \leq \sigma \leq 1$  und  $\sigma_{\Omega'} \equiv 1$ . Hierfür gilt

$$|\nabla \sigma^2| \leq c_\delta \sigma, \quad \delta := \text{dist}\{\Omega', \partial\Omega\},$$

sowie

$$\nu\|\nabla e^v\|_{\Omega'}^2 + \varepsilon\|\nabla e^p\|_{\Omega'}^2 \leq \nu\|\sigma \nabla e^v\|^2 + \varepsilon\|\sigma \nabla e^p\|^2.$$

Wir wollen nun die „gewichteten“  $L^2$ -Fehlernormen abschätzen. Mit Hilfe der Fehlergleichungen (6.4.192), (6.4.193) und Beachtung von  $\sigma|_{\partial\Omega} \equiv 0$  erhalten wir

$$\begin{aligned} \nu\|\sigma \nabla e^v\|^2 + \varepsilon\|\sigma \nabla e^p\|^2 &= \nu(\nabla e^v, \nabla(\sigma^2 e^v) - \nu(\nabla e^v, e^v \nabla \sigma^2) \\ &\quad + \varepsilon(\nabla e^p, \nabla(\sigma^2 e^p)) - \varepsilon(\nabla e^p, e^p \nabla \sigma^2) \\ &= (e^p, \nabla \cdot (\sigma^2 e^v)) - \nu(\nabla e^v, e^v \nabla \sigma^2) \\ &\quad - (\nabla \cdot e^v, \sigma^2 e^p) + \varepsilon(\nabla p, \nabla(\sigma^2 e^p)) - \varepsilon(\nabla e^p, e^p \nabla \sigma^2) \\ &= (e^p, \sigma^2 \nabla \cdot e^v) + (e^p, e^v \nabla \sigma^2) - \nu(\nabla e^v, e^v \nabla \sigma^2) \\ &\quad - (\nabla \cdot e^v, \sigma^2 e^p) + \varepsilon(\nabla p, \nabla(\sigma^2 e^p)) - \nu(\nabla e^v, e^v \nabla \sigma^2) \\ &= (e^p, e^v \nabla \sigma^2) - \nu(e^v \nabla \sigma^2, \nabla e^v) - \varepsilon(\Delta p, \sigma^2 e^p) - \varepsilon(\nabla e^p, e^p \nabla \sigma^2). \end{aligned}$$

Die vier Ausdrücke auf der rechten Seite werden wie folgt abgeschätzt:

$$\begin{aligned} (e^p, e^v \nabla \sigma^2) &\leq \frac{\alpha}{2\nu}\|\sigma e^p\|^2 + c_\alpha \nu\|e^v\|^2, \\ -\nu(e^v \nabla \sigma^2, \nabla e^v) &\leq c\nu\|e^v\|^2, \\ -\varepsilon(\Delta p, \sigma^2 e^p) &\leq \frac{\alpha}{2\nu}\|\sigma e^p\|^2 + c_\alpha \nu \varepsilon^2 \|\Delta p\|^2, \\ -\varepsilon(\nabla e^p, e^p \nabla \sigma^2) &\leq c\varepsilon\|e^p\|^2. \end{aligned}$$

Kombination dieser Abschätzungen ergibt dann

$$\nu\|\sigma \nabla e^v\| \leq \alpha\|\sigma e^p\| + c_\alpha \varepsilon \{\nu\|e^v\| + \sqrt{\nu\varepsilon}\|e^p\|\} + c_\alpha \nu \varepsilon \|\Delta p\|. \quad (6.4.196)$$

Jetzt verwenden wir wieder die Stabilität des div-Operators:

$$\hat{\beta} \|\sigma e^p - \overline{\sigma e^p}\| \leq \sup_{\varphi \in H} \frac{(\sigma e^p - \overline{\sigma e^p}, \nabla \cdot \varphi)}{\|\nabla \varphi\|},$$

wobei  $\overline{\sigma e^p}$  der Mittelwert von  $\sigma e^p$  über  $\Omega$  ist. Wegen

$$|\overline{\sigma e^p}| \leq c \|\nabla \sigma\| \|e^p\|_{-1}$$

erhalten wir folglich

$$\|\sigma e^p\| \leq c \sup_{\varphi \in H} \frac{(\sigma e^p, \nabla \cdot \varphi)}{\|\nabla \varphi\|} + c \|e^p\|_{-1}.$$

Weiter gilt unter Verwendung von (6.4.196):

$$\begin{aligned} (\sigma e^p, \nabla \cdot \varphi) &= (e^p, \nabla \cdot (\sigma \varphi)) - (e^p, \varphi \cdot \nabla \sigma) \\ &= \nu (\sigma \nabla e^v, \nabla \varphi) - \nu (\nabla e^v, \varphi \cdot \nabla \sigma) - (e^p, \varphi \cdot \nabla \sigma). \end{aligned}$$

Hiermit erschließen wir, dass

$$\|\sigma e^p\| \leq c \{ \nu \|\sigma \nabla e^v\| + \nu \|e^v\| + \|e^p\|_{-1} \}.$$

Dieses Resultat wird in (6.4.196) eingesetzt und bei geeigneter Wahl von  $\alpha$  ergibt sich

$$\nu \|\sigma e^v\| + \|\sigma e^p\| \leq \{ \nu \|e^v\| + \sqrt{\nu \varepsilon} \|e^p\| + \|e^p\|_{-1} \} + c \nu \varepsilon \|\Delta p\|.$$

Die Resultate von Schritt (1) liefern dann schließlich die Abschätzung

$$\nu \|\sigma \nabla e^v\| + \|\sigma e^p\| \leq c \nu \varepsilon \{ \|\nabla p\| + \|\Delta p\| \},$$

womit die gewünschte lokale Abschätzung auf  $\Omega'$  folgt.

Q.E.D.

**Satz 6.13 (Druckgrenzschicht):** Für den durch die Druckstabilisierung in (6.4.188) hervorgerufenen Fehler gilt auf Teilgebieten  $\Omega_\delta := \{x \in \Omega \mid \text{dist}(x, \partial\Omega) > \delta\}$ :

$$\|p - p_\varepsilon\|_{\Omega_\delta} \leq \left\{ \exp\left(\frac{-\delta}{2\sqrt{\nu\varepsilon}}\right) \sqrt{\nu\varepsilon} \|\nabla p\| + \nu\varepsilon \|\Delta p\| \right\}. \quad (6.4.197)$$

**Beweis:** Wir skizzieren einen Beweisansatz, in welchem aber eine Lücke besteht, die leider als Problem offen bleibt. Ausgangspunkt sind wieder die Fehlergleichungen (6.4.192) und (6.4.193):

$$-\nu \Delta e^v + \nabla e^p = 0 \quad \text{in } \Omega, \quad e^v|_{\partial\Omega} = 0, \quad (6.4.198)$$

$$\nabla \cdot e^v - \varepsilon \Delta e^p = \varepsilon \Delta p \quad \text{in } \Omega, \quad \partial_n e^p|_{\partial\Omega} = -\partial_n p|_{\partial\Omega}. \quad (6.4.199)$$

Sei  $\Delta_D : H_0^1(\Omega)^d \rightarrow L^2(\Omega)^d$  und  $\Delta_N : H^1(\Omega) \rightarrow L^2(\Omega)$  der Laplace-Operator auf  $\Omega$  zu homogenen Dirichlet- bzw. Neumann-Randbedingungen. Mit diesen Bezeichnungen kann das System der Fehlergleichungen umgeformt werden zu

$$(\nu^{-1} \nabla \cdot \Delta_D^{-1} \nabla - \varepsilon \Delta_N) e^p = \varepsilon \Delta p \quad \text{in } \Omega, \quad \partial_n e^p|_{\partial\Omega} = -\partial_n p|_{\partial\Omega}. \quad (6.4.200)$$

Der Operator  $\nu^{-1}\nabla\cdot\Delta_D^{-1}\nabla$  ist von 0-ter Ordnung, auf ganz  $L^2(\Omega)$  definiert, symmetrisch und positiv-definit. An dieser Stelle entsteht eine Lücke im Beweis, da wir im folgenden den vereinfachten („lokalen“) Operator  $\nu^{-1}I - \varepsilon\Delta_N$  betrachten werden. Wir setzen  $d(x) := \text{dist}(x, \partial\Omega)$  und  $\varepsilon' := \nu\varepsilon$  und definieren die Gewichtsfunktion

$$\sigma(x) := \min \left\{ e^{d(x)/\sqrt{\varepsilon'}}, e^{\delta/\sqrt{\varepsilon'}} \right\}.$$

Damit gilt

$$\|e^p\|_{\Omega_\delta} + \varepsilon'\|\nabla e^p\|_{\Omega_\delta} \leq e^{-\delta/\sqrt{\varepsilon'}} \left\{ (\sigma e^p, e^p) + \varepsilon'(\sigma\nabla e^p, \nabla e^p) \right\} =: e^{-\delta/\sqrt{\varepsilon'}} A.$$

Weiter ist

$$A = (\sigma e^p, e^p) + \varepsilon'(\nabla(\sigma e^p), \nabla e^p) - \varepsilon'(e^p \nabla \sigma, \nabla e^p)$$

und, wegen  $|\nabla \sigma| \leq \sigma/\sqrt{\varepsilon'}$ ,

$$\varepsilon'|e^p, \nabla \sigma, \nabla e^p| \leq \frac{1}{2} \left\{ (\sigma e^p, e^p) + \varepsilon'(\sigma\nabla e^p, \nabla e^p) \right\}.$$

Hiermit erschließen wir

$$A \leq \frac{1}{2} \left\{ (\sigma e^p, e^p) + \varepsilon'(\sigma\nabla e^p, \nabla e^p) \right\} = 2\varepsilon'(\Delta p, \sigma e^p) + 2\varepsilon'(\sigma e^p, \partial_n p)_{\partial\Omega}.$$

Unter Beachtung von  $\sigma|_{\partial\Omega} \equiv 1$  folgt

$$\varepsilon'(\sigma e^p, \partial_n p)_{\partial\Omega} = \varepsilon'(\nabla e^p, \nabla p) + \varepsilon'(e^p, \Delta p) = \varepsilon'(\nabla e^p, \nabla p) + \varepsilon'(e^p, \Delta p).$$

Hieraus folgern wir

$$(\sigma e^p, e^p) + \varepsilon'(\nabla(\sigma e^p), \nabla e^p) \leq 2A^{1/2} \left\{ \varepsilon'^2 e^{\delta/\sqrt{\varepsilon'}} \|\Delta p\|^2 + \varepsilon' \|\nabla p\|^2 \right\}^{1/2}$$

und schließlich

$$A \leq c \left\{ \varepsilon' e^{\delta/\sqrt{\varepsilon'}} \|\Delta p\|^2 + \varepsilon' \|\nabla p\|^2 \right\}.$$

Dies vervollständigt den Beweis. Q.E.D.

Diese Argumentation kann auf die Situation des Chorin-Verfahrens zur Approximation der instationären Navier-Stokes-Gleichungen übertragen werden. Die Beweise werden allerdings sehr technisch und können hier nicht präsentiert werden. Die Bedeutung des Resultats von Satz 6.4.2, d. h. der exponentielle Charakter der Grenzschicht des Druckfehlers, wird durch Abb. 6.15 illustriert.

Das Resultat von Satz 6.13 legt es nahe, dass durch Extrapolation der „guten“ Druckwerte aus dem Innern von  $\Omega$  zum Rand auch dort ausreichend genaue Druckapproximationen gewonnen werden können. Dies wird bestätigt durch das in Abb. 6.16 dargestellte Resultat eine Testrechnung.

Ein wichtiger Schritt in Richtung auf ein randfehler-freies Projektionsverfahren ist das von A. Prohl [129] vorgeschlagene „Chorin-Uzawa-Verfahren“, das wie folgt lautet:

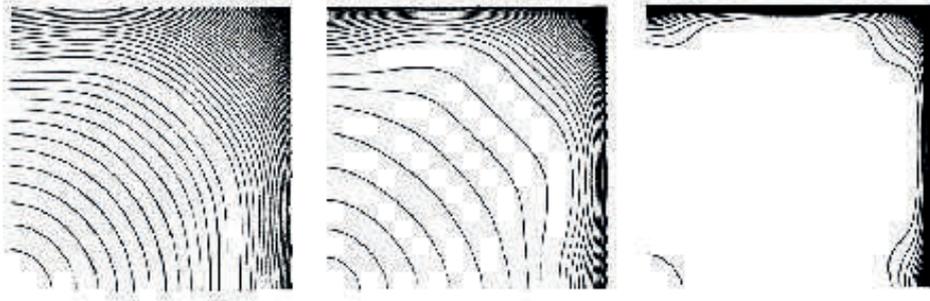


Abbildung 6.15: Folge von Druck-Isolinien berechnet mit dem Chorin-Verfahren mit Zeitschrittweiten  $k = 2.5 \cdot 10^{-2}$ ,  $6.25 \cdot 10^{-3}$ ,  $1.56 \cdot 10^{-3}$  (Modellproblem auf dem Einheitsquadrat mit  $\nu = 1$  mit polynomialer Lösung).

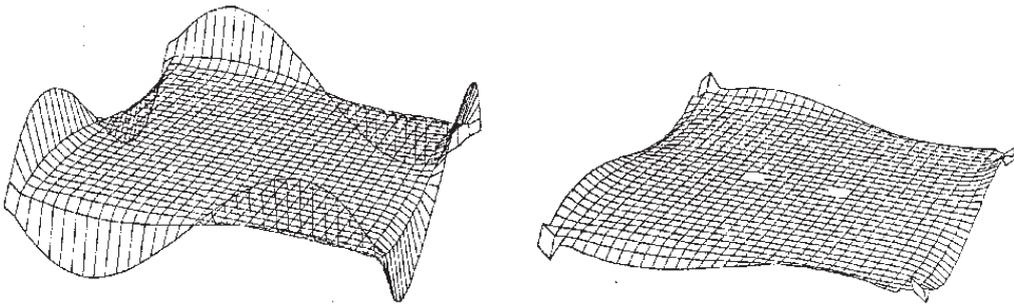


Abbildung 6.16: Druckfehler für eine polynomialer Lösung vor (links) und nach (rechts) Korrektur durch Extrapolation zum Rand.

### Chorin-Uzawa-Verfahren:

i) Impliziter „Konvektions-Diffusions-Schritt“:

$$k^{-1}\{\tilde{v}^m - v^{m-1}\} - \nu \Delta \tilde{v}^m + \tilde{v}^m \cdot \nabla \tilde{v}^m + \nabla(q^{m-1} - p^{m-1}) = f^m \quad \text{in } \Omega, \quad \tilde{v}^m|_{\partial\Omega} = 0.$$

ii) Druck-Poisson-Gleichung:

$$\Delta q^m = k^{-1} \nabla \cdot \tilde{v}^m \quad \text{in } \Omega, \quad \partial_n q^m|_{\partial\Omega} = 0.$$

iii) Druck- und Geschwindigkeitskorrektur ( $\alpha \in (0, 1)$ ):

$$v^m := \tilde{v}^m - k \nabla q^m, \quad p^m := p^{m-1} - \alpha \nabla \cdot \tilde{v}^m.$$

Der Namensbestandteil „Uzawa“ deutet an, dass dieser Algorithmus von seiner Struktur her zum Uzawa-Algorithmus zur Lösung der stationären Navier-Stokes-Gleichungen

korrespondiert. Mit derselben Argumentation wie oben sehen wir, dass das Chorin-Uzawa-Verfahren einer Druckstabilisierung der Form

$$\nabla \cdot v + \alpha^{-1} k \partial_t p = 0. \quad (6.4.201)$$

entspricht. Diese Stabilisierung bewirkt keine „singuläre“ Störung in der Druckgleichung, so dass keine Grenzschicht im Fehler auftreten sollte. Diese Erwartung wird durch die Ergebnisse einer Testrechnung mit polynomialer Lösung bestätigt (siehe Abb. 6.16).

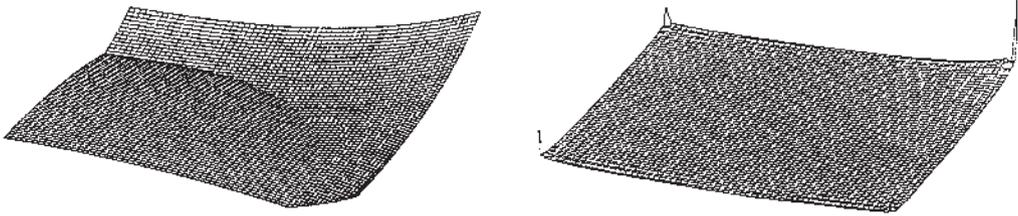


Abbildung 6.17: Druckfehler für eine polynomialer Lösung mit dem „klassischen“ Chorin-Verfahren (links) und dem Chorin-Uzawa-Verfahren (rechts); Rechnung für  $\nu = 1$  mit dem  $Q_1^c/Q_1^c$ -Stokes-Element auf einem äquidistanten Gitter mit  $h = 1/64$  und  $k = 1/100$ .

Die geringe Approximationsgenauigkeit des Chorin-Verfahrens kann formal behoben werden, indem das Verfahren zu einer „Defektkorrektur-Iteration“ aufbauend auf dem Crank-Nicolson-Schema erweitert wird zum sog. „Projektionsverfahren von van Kan<sup>14</sup>“. Hier erfolgt der Schritt  $t_{m-1} \rightarrow t_m$  gemäß

#### van Kan-Verfahren:

i) Impliziter „Konvektions-Diffusions-Schritt“:

$$\begin{aligned} k^{-1} \{ \tilde{v}^m - v^{m-1} \} - \frac{1}{2} k \Delta \{ \tilde{v}^m + v^{m-1} \} + \frac{1}{2} \bar{v}^m \cdot \nabla \tilde{v}^m + \frac{1}{2} v^{m-1} \cdot \nabla v^{m-1} \\ = \frac{1}{2} \{ f^m + f^{m-1} \} - \nabla p^{m-1} \quad \text{in } \Omega, \quad \tilde{v}^m|_{\partial\Omega} = 0. \end{aligned}$$

ii) Druck-Poisson-Gleichung:

$$\Delta q^m = k^{-1} \nabla \cdot \tilde{v}^m \quad \text{in } \Omega, \quad \partial_n q^m|_{\partial\Omega} = 0.$$

iii) Druck- und Geschwindigkeitskorrektur:

$$v^m := \tilde{v}^m + k \nabla q^m, \quad p^m := p^{m-1} + q^m.$$

Dabei ist  $\bar{v}^m := \tilde{v}^m$  (nicht-lineare Formulierung) oder  $\bar{v}^m := 2v^{m-1} - v^{m-2}$  (Linearisierung durch Zeitextrapolation zweiter Ordnung). Dieses Verfahren ist formal von zweiter

<sup>14</sup>Jos van Kan (1944–): Niederländischer Mathematiker; Prof. em. für Mathematik an der Delft University of Technology; Beiträge zur Numerischen Strömungsmechanik (Druckkorrekturverfahren).

Ordnung,  $\mathcal{O}(k^2)$ , und erzeugt im Fall der Konvergenz  $\tilde{v}^m \rightarrow v^\infty$ ,  $q^m \rightarrow 0$  ( $m \rightarrow \infty$ ) Limiten, welche Lösungen der stationären Navier-Stokes-Gleichungen sind:

$$\begin{array}{ccccccc} -\nu\Delta\tilde{v}^m + \tilde{v}^m \cdot \nabla\tilde{v}^m + \nabla p^{m-1} - f^m = k^{-1}(\tilde{v}^m - v^{m-1}) & & & & & & \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & (m \rightarrow \infty) \\ -\nu\Delta v^\infty + v^\infty \cdot \nabla v^\infty + \nabla p - f = & & & & & & 0. \end{array}$$

Das van Kan-Verfahren erlaubt eine Fehlerabschätzung der Form

$$|p^m(x) - p(x, t_m)| \approx k^2 + \exp\left(\frac{d(x)}{\sqrt{\nu k}}\right)k \quad (6.4.202)$$

und korrespondiert zu einer Druckstabilisierung der Form

$$\nabla \cdot v + \alpha^{-1}k\partial_t\Delta p = 0. \quad (6.4.203)$$

Wir fassen zusammen, dass alle bisher betrachteten Projektionsverfahren (und einige Vorläufer) zu Spezialfällen aus einer Familie von Druckstabilisierungsansätzen wachsender Komplexität korrespondieren:

- $\nabla \cdot v + \varepsilon p = 0$  in  $\Omega$ , („Penalty-Methode“).
- $\nabla \cdot v + \varepsilon\partial_t p = 0$  in  $\Omega$ ,  $p|_{t=0} = p^0$  („Chorin-Uzawa-Verfahren“).
- $\nabla \cdot v + \varepsilon\Delta p = 0$  in  $\Omega$ ,  $\partial_n p|_{\partial\Omega} = 0$  („Chorin-Verfahren“).
- $\nabla \cdot v + \varepsilon\partial_t\Delta p = 0$  in  $\Omega$ ,  $\partial_n p|_{\partial\Omega} = 0$ ,  $p|_{t=0} = p^0$  („van Kan-Verfahren“).

Zum Abschluss unserer Diskussion wollen wir klären, wieso beim Chorin-Algorithmus auch in Verbindung mit „instabilen“ Stokes-Elementen häufig keine Probleme mit unphysikalischen Oszillationen im Druck beobachtet werden. Wir haben gesehen, dass der Chorin-Algorithmus implizit eine Modifikation der Kontinuitätsgleichung der Form

$$\nabla \cdot v - k\Delta p^m = 0$$

enthält. Wenn die Zeitschrittweite nicht zu klein wird,  $k \geq h^2$ , wirkt dies in Verbindung mit einer Ortsdiskretisierung wie eine Druckstabilisierung. Im Gegensatz dazu korrespondiert das van Kan-Verfahren zu einer Stabilisierung der Form

$$\nabla \cdot v + \alpha^{-1}k\partial_t\Delta p = 0,$$

welche bei Konvergenz gegen eine stationäre Lösung,  $\partial_t p(t) \rightarrow 0$  ( $t \rightarrow \infty$ ) nur erhalten bleibt, wenn die Zeitschrittweite entsprechend vergrößert wird.

### 6.4.3 Lösung der algebraischen Teilprobleme

Bei den beschriebenen Zeitschrittverfahren müssen in jedem einzelnen Zeitschritt im Rahmen eines Linearisierungsprozesses mehrere lineare Sattelpunktprobleme der Form

$$\begin{bmatrix} M + \alpha k A(\bar{x}) & kB \\ -B^T & \varepsilon C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix} \quad (6.4.204)$$

gelöst werden. Dabei sind wieder  $\{x, y\}$  die Koeffizientenvektoren der Geschwindigkeits- und Druckvariablen bzgl. geeigneter Knotenbasen der Ansatzräume  $H_h$  und  $L_h$ , und die Systemmatrizen  $M, A, N(\cdot), B$  und  $A(\cdot) = \nu A + N(\cdot)$  sind gegeben durch

$$\begin{aligned} M &:= ((\psi_h^i, \psi_h^j))_{i,j=1}^{N_H}, & A &:= ((\nabla_h \psi_h^i, \nabla_h \psi_h^j))_{i,j=1}^{N_H}, & B &:= -((\chi_h^i, \nabla_h \cdot \psi_h^j))_{i,j=1}^{N_L, N_H}, \\ N(\bar{v}) &:= ((\bar{v} \cdot \nabla_h \psi_h^i, \psi_h^j))_{i,j=1}^{N_H}, & C &:= ((\nabla_h \chi_h^i, \nabla_h \chi_h^j))_{i,j=1}^{N_L}. \end{aligned}$$

wobei  $\bar{v}$  der zum (geschätzten) Vektor  $\bar{x}$  korrespondierende Geschwindigkeitsvektor ist. Die Matrix  $C$  repräsentiert die Druckstabilisierung (wenn notwendig) und der Vektor  $c$  die zugehörigen Korrekturterme. In diesem Fall ist der Hauptdiagonalblock

$$S(\bar{x}) := M + \alpha k A(\bar{x})$$

nicht symmetrisch, aber bei geeigneter Stabilisierung (z. B.: „upwinding“, Stromliniendiffusion, u.s.w.) regulär. Folglich kann die Komponente  $x$  wieder aus dem System eliminiert werden:

$$[kB^T S(\bar{x})^{-1} B + \varepsilon C] y = B^T S(\bar{x})^{-1} b + c, \quad x = S(\bar{x})^{-1} b - k S(\bar{x})^{-1} y. \quad (6.4.205)$$

Das resultierende Schur-Komplement

$$\Sigma(\bar{x}) := kB^T S(\bar{x})^{-1} C + \varepsilon C$$

kann dann etwa mit dem Uzawa-Algorithmus oder mit einer angepassten Variante des CG-Verfahrens für unsymmetrische Matrizen (GMRES- oder BiCGstab-Verfahren) invertiert werden. Auch hier ist wieder die Kondition bzw. die Norm von  $\Sigma(\bar{x})$  entscheidend für die Konvergenzgeschwindigkeit.

Nun sollte man annehmen, in Analogie zur Situation der Wärmeleitungsgleichung, dass für kleiner werdende Zeitschrittweite  $k$  die Lösung der linearen Gleichungssysteme immer leichter wird. Dies ist aber beim Schur-Komplement  $\Sigma(\bar{x})$  wegen des DAE-Charakters des Systems nicht der Fall, da hier für  $k \ll 1$  (und  $\varepsilon \ll 1$ )

$$\Sigma(\bar{x}) = kB^T S(\bar{x})^{-1} B + \varepsilon C \approx kB^T M^{-1} B.$$

Die Grenzmatrix  $B^T M^{-1} B$  ist eine Approximation des Laplace-Operators mit der Kondition (auf gleichförmigem Gitter)

$$\text{cond}_2(B^T M^{-1} B) = \mathcal{O}(h^{-2}).$$

Dies bedeutet, dass die bisher zur Invertierung von  $\Sigma(\bar{x})$  diskutierten Verfahren für kleine Zeitschrittweiten zunehmend langsam werden. Also ist „Vorkonditionierung“ notwendig. Diese wird aber schwierig, da die Matrix  $\Sigma(\bar{x})$  nicht elementweise zur Verfügung steht, so dass die üblichen punkt- oder blockweisen Iterationsschemata wie „Jacobi“- , „Gauß-Seidel“- , „SOR“- , „ADI“- und „ILR“-Iterationen nicht unmittelbar anwendbar sind. Dasselbe Problem ergibt sich bei der Anwendung von Mehrgitter-Algorithmen für das Schur-Komplement bei der Konstruktion von Glättern. In diesem Fall bietet sich der Uzawa-Algorithmus selber als „Glätter“ in einem äußeren Mehrgitterverfahren an.

### Mehrgitter-Algorithmus

Zum Abschluss wollen wir noch einen Mehrgitteralgorithmus präsentieren, welcher speziell auf die Lösung „instationärer“ Systeme zugeschnitten ist. Dieser Ansatz ist besonders effizient im Fall

$$\sigma = \frac{1}{k} \approx \frac{\nu}{h^2}.$$

Die Lösung der Schur-Komplement-Gleichungen

$$B^T S^{-1} B p^m = B^T S^{-1} g^m, \quad v^m = S^{-1}(g^m - B p^m), \quad (6.4.206)$$

durch CG-artige Verfahren erfordert die Auswertung von  $S^{-1}$ , was zunehmend ineffizient wird für kleine Zeitschrittweiten  $k$ , größere Reynolds-Zahlen und auf stark anisotropen Gittern. Dieses Problem kann gemeistert werden durch die Verwendung einfacher Richardson-Iterationen für die Schur-Komplement-Gleichung mit einem Vorkonditionierer der Art  $B^T C^{-1} B$ . Populäre Beispiele solcher Vorkonditionierer  $C$  sind:

- $C^{-1} := I$ ;
- $C^{-1} := \tilde{M}^{-1}$ ;
- $C^{-1} := \tilde{M}^{-1} + \alpha^{-1} B^T B$ .

Die resultierende Iterationsmethode, sog. „diskrete Projektionsverfahren“, lautet:

$$p^{m,l+1} = p^{m,l} - (B^T C^{-1} B)^{-1} (B^T S^{-1} B p^{m,l} - B^T S^{-1} g^m). \quad (6.4.207)$$

Nach  $L$  Iterationsschritten wird  $p^m := p^{m,L}$  gesetzt, und man berechnet die zugehörige Geschwindigkeitskomponente aus

$$S v^m = g^m - B p^m + \alpha^{-1} (\alpha I - S C^{-1}) B (p^{m,L} - p^{m,L-1}). \quad (6.4.208)$$

mit einem Relaxationsparameter  $\alpha \in (0, 1)$ . Diese Konstruktion von  $v^m$  garantiert, dass die resultierende Geschwindigkeit im diskreten Sinne „divergenz-frei“ ist,  $B^T v^m = 0$ , und motiviert damit den Namen „diskrete Projektionsmethode“ für das gesamte Verfahren. Dieses Iterationsverfahren wird dann als Glätter in einem äußeren Mehrgitterverfahren eingesetzt.

Im Spezialfall  $L = 1$  entspricht dieses Verfahren einer diskreten Version der Projektionsverfahren nach Chorn (für die Wahl  $p^{m,0} := 0$ ) und Van Kan (für die Wahl  $p^{m,0} := p^{m-1}$ ). Diese Operator-Splitting-Schemata haben die Form

1.  $S\tilde{v}^m = g^m - kBp^{m-1}$  (Transport-Diffusions-Schritt).
2.  $B^T\tilde{M}^{-1}Bq^m = k^{-1}B^T\tilde{v}^m$  (Druck-Poisson-Gleichung).
3.  $v^m = \tilde{v}^m - k\tilde{M}^{-1}Bq^m$  (Geschwindigkeitskorrektur).
4.  $p^m = p^{m-1} + \alpha q^m$  (Druckkorrektur).

Der Mehrgitter-Algorithmus mit Glättung durch die „diskrete“ Projektionsiteration hat sich in praktischen Tests als sehr effizient zur Lösung der voll-gekoppelten instationären Navier-Stokes-Gleichungen erwiesen. Er ist robust für alle relevanten Reynolds-Zahlen (*laminare* Strömung) und alle physikalisch realistischen Zeitschrittweiten. Er basiert auf optimalen „inneren“ Mehrgitterkomponenten für die Teilprobleme „Transport-Diffusions-Problem“ und „Druck-Poisson-Gleichung“. In Beispielen erfordert der Gesamtalgorithmus bei Verwendung z. B. des nicht konformen  $Q_1^{\text{rot}}/P_0^{\text{dc}}$ -Stokes-Elements in 3D pro Zelle etwa 1 KByte Speicher, so dass Simulationen mit mehr als  $10^7$  Unbekannten auf Workstations mit 2 GByte Arbeitsspeicherspeicher möglich sind; für Details verweisen wir auf die Monographie Turek [145]).