Untersuchung von iterativen Lösungsverfahren und Vorkonditionierungsstrategien für die Oseen-Gleichungen

Diplomarbeit

vorgelegt von
Susann Müller
aus
Wippra

angefertigt am
Institut für
Numerische und Angewandte Mathematik
der
Georg–August–Universität zu Göttingen
1997

Inhaltsverzeichnis

Einleitung

Die Navier-Stokes-Gleichungen für inkompressible Flüssigkeiten zählen zu den wichtigsten Gleichungen der Strömungsphysik. Gesucht wird beim Navier-Stokes-Problem das Geschwindigkeitsfeld u und der Druck p mit

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

Dieses nichtlineare Problem kann durch Verwendung der Iterationsvorschrift

$$\begin{aligned} -\nu\triangle u^{n+1} + (u^n\cdot\nabla)u^{n+1} + \nabla p^{n+1} &= f, \\ \nabla\cdot u^{n+1} &= 0 \end{aligned}$$

linearisiert werden. Die in jedem Iterationsschritt entstehenden linearisierten Gleichungen werden auch als Oseen-Gleichungen bezeichnet. Betrachtet man die durch die Linearisierung entstehenden Probleme als eigenständige Gleichungen, so haben sie die folgende Form:

Gesucht sind u und p mit

$$-\nu\triangle u + (a\cdot\nabla)u + \nabla p = f, \tag{1}$$

$$\nabla\cdot u = 0, \tag{2}$$

wobei a ein divergenzfreies Strömungsfeld ist.

Die vorliegende Arbeit beschäftigt sich mit der numerischen Behandlung von Oseen-Gleichungen, besonders mit dem Problem der schnellen Lösung des bei der Finiten-Elemente-Methode entstehenden linearen Gleichungssystems durch iterative Lösungsverfahren.

Eine schwache Formulierung von (1) und (2) führt auf gemischte Variationsgleichungen (vgl. Kapitel 3). Die theoretischen Grundlagen zu gemischten Variationsproblemen werden in Kapitel 2 dargelegt.

Bei der Diskretisierung von Oseen-Gleichungen mit finiten Elementen entstehen große schwachbesetzte lineare Gleichungssysteme. Die Effektivität von iterativen Verfahren zur Lösung solcher Systeme hängt von der Kondition der Matrix ab, die durch geeignete Vorkonditionierung verbessert werden kann. In Kapitel 4 wird auf Vorkonditionierer

eingegangen, die den Oseen-Problemen angepaßt sind. Diese zeichnen sich besonders dadurch aus, daß die Eigenwerte des vorkonditionierten Systems unabhängig von der Diskretisierungsschrittweite h beschränkt sind.

Am Institut für Numerische und Angewandte Mathematik der Universität Göttingen wird das Programm ParallelNS (Parallelized solution of Navier-Stokes equations) zur Lösung partieller Differentialgleichungen entwickelt. In ParallelNS ist das Programmpaket BLANC eingebunden, um die bei der Diskretisierung entstehenden linearen Gleichungssysteme zu lösen. In diesem Paket sind verschiedene iterative Lösungsverfahren und Vorkonditionierer implementiert. In Kapitel 5 sind numerische Experimente dokumentiert, die das Ziel hatten, für verschiedene Oseen-Probleme effiziente Löser-Vorkonditionierer-Kombinationen zu ermitteln. Dabei sind die in BLANC zur Verfügung stehenden Löser und Präkonditionierer getestet worden.

Kapitel 1

Angepaßte Funktionenräume

Eines der einfachsten Beispiele partieller Differentialgleichungen ist das sogenannte Poissonproblem mit homogenen Dirichlet-Randbedingungen:

Sei $\Omega \subset \mathbb{R}^n$ ein beschränktes Gebiet. Gesucht ist $u : \bar{\Omega} \longrightarrow \mathbb{R}$ mit

$$\begin{aligned} -\triangle u &= f \quad \text{in } \Omega, && (1.1)\\ u &= 0 \quad \text{auf } \Gamma := \partial\Omega, && (1.2) \end{aligned}$$

wobei $\triangle u := \sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2}$.
Eine Funktion $u \in C^2(\Omega) \cap C(\bar{\Omega})$, die die Gleichungen (1.1) und (1.2) punktweise erfüllt, heißt klassische Lösung des Poisson-Problems.
Ein Problem der klassischen Lösungstheorie ist, daß die Anforderungen an die Daten des Problems zu hoch sind. Ein Ausweg führt bei partiellen Differentialgleichungen auf den Begriff der schwachen Lösung. Dazu ist es erforderlich, entsprechende Räume einzuführen. Zuerst wird auf den Begriff der Lebesque-Räume (L^p-Räume) eingegangen.

Definition 1.1 *Man bezeichnet mit $L^p(\Omega)$, $1 \leq p < \infty$, die Menge aller Äquivalenzklassen meßbarer Funktionen $u : \Omega \longrightarrow \mathbb{R}$ mit*

$$\|u\|_{L^p(\Omega)} := \left(\int_\Omega |u(x)|^p \, dx \right)^{1/p} < \infty. \qquad (1.3)$$

Bemerkung 1.2
(i) Die Menge $L^p(\Omega)$, $1 \leq p < \infty$, ist ein Banach-Raum bezüglich der Norm (1.3).
(ii) Die Menge $L^2(\Omega)$ ist mit dem Skalarprodukt

$$(u, v)_{L^2(\Omega)} := \int_\Omega u(x)v(x) \, dx \quad \textit{für alle } u, v \in L^2(\Omega)$$

ein Hilbert-Raum.

Beweis: vgl. [Alt92] S.27ff

Eine Erweiterung auf den Fall $p = \infty$ kann auf folgende Weise vorgenommen werden.

Definition 1.3 *Die Menge der Äquivalenzklassen der auf Ω wesentlich beschränkten Funktionen ist*

$$L^\infty(\Omega) := \{u : \Omega \longrightarrow \mathbb{R} \text{ meßbar} : \exists M < \infty \text{ mit } |u(x)| \leq M \text{ f.ü. in } \Omega\}$$

mit der Norm

$$\|u\|_{L^\infty(\Omega)} := ess \sup_{x\in\Omega} |u(x)| := \inf M.$$

Eine wichtige Rolle für die Einführung der Sobolev-Räume spielen auch die Räume der unendlich oft differenzierbaren Funktionen mit kompaktem Träger und der lokal integrierbaren Funktionen.

Definition 1.4
(i) $C^\infty(\Omega) := \{u : \Omega \longrightarrow \mathbb{R} : u$ ist unendlich oft differenzierbar $\}$ ist die Menge der unendlich oft differenzierbaren Funktionen.
(ii) Die Menge der unendlich oft differenzierbaren Funktionen mit kompaktem Träger ist definiert durch

$$C_0^\infty(\Omega) := \{u \in C^\infty(\Omega) : \ supp\,(u) \text{ ist kompakte Teilmenge von } \Omega\},$$

wobei $supp\,(u) := cl\{x \in \Omega : u(x) \neq 0\}$.

Definition 1.5 *Eine meßbare Funktion $u : \Omega \longrightarrow \mathbb{R}$ gehört zu der Menge $L^1_{loc}(\Omega)$, wenn für alle abgeschlossenen beschränkten Teilmengen $A \subset \Omega$ gilt:*

$$\int_A |u(x)|\, dx < \infty.$$

Als nächstes ist die Einführung der Multiindexschreibweise zweckmäßig.

Definition 1.6 *Ein Vektor $\alpha := (\alpha_1, \ldots, \alpha_n)$ mit nichtnegativen ganzen Zahlen α_i heißt Multiindex der Länge*

$$|\alpha| := \sum_{i=1}^{n} \alpha_i.$$

Die partielle Ableitung der Ordnung α einer hinreichend oft differenzierbaren Funktion $u : \Omega \longrightarrow \mathbb{R}$ im Punkt $x \in \Omega$ schreibt man in folgender Form:

$$\begin{aligned} D^\alpha u(x) &:= \frac{\partial^{|\alpha|} u}{\partial^{\alpha_1} x_1 \cdots \partial^{\alpha_n} x_n}(x), \ (|\alpha| \geq 1) \\ D^{(0,\ldots,0)} u(x) &= u(x). \end{aligned}$$

Nun ist es möglich, den Begriff der verallgemeinerten Ableitung zu definieren.

Definition 1.7 *Eine Funktion $\omega_\alpha \in L^1_{loc}(\Omega)$ heißt verallgemeinerte Ableitung $D^\alpha u$ von $u \in L^1_{loc}(\Omega)$, falls gilt*

$$\int_\Omega \omega_\alpha v \, dx = (-1)^{|\alpha|} \int_\Omega u D^\alpha v \, dx \quad \text{für alle } v \in C_0^\infty(\Omega).$$

Definition 1.8 *Für $1 \leq p < \infty$ heißt die Menge*

$$W^{k,p}(\Omega) := \{u \in L^p(\Omega) : \exists D^\alpha u \in L^p(\Omega), \forall \alpha : |\alpha| \leq k\}$$

Sobolev-Raum der Funktionen mit verallgemeinerten und zur p-ten Potenz auf Ω integrierbaren Ableitungen bis zur Ordnung k. Auf dem $W^{k,p}$ läßt sich eine Norm definieren durch

$$\|v\|_{W^{k,p}(\Omega)} := \left(\sum_{|\alpha| \leq k} \|D^\alpha u\|^p_{L^p(\Omega)} \right)^{1/p}. \tag{1.4}$$

Bemerkung 1.9
Der $W^{k,p}(\Omega)$ ist mit der Norm (1.4) ein Banach-Raum.

Beweis: vgl. [Alt92] Abschnitt 1.15

Definition 1.10 *Der $W_0^{k,p}(\Omega)$ ist der Abschluß des $C_0^\infty(\Omega)$ in der Norm $\|\cdot\|_{W^{k,p}(\Omega)}$.*

Für die Anwendungen auf partielle Differentialgleichungen zweiter Ordnung hat der Spezialfall $k = 1$ und $p = 2$ eine besondere Bedeutung. In diesem Fall gilt

Bemerkung 1.11
Für $u \in W_0^{1,2}(\Omega)$ ist

$$|u|_{W^{1,2}(\Omega)} := \left(\sum_{|\alpha|=1} \|D^\alpha u\|^2_{L^2(\Omega)} \right)^{1/2}$$

eine zu $\|\cdot\|_{W^{1,2}(\Omega)}$ äquivalente Norm.

Beweis: vgl. [GR94] S.84

Bemerkung 1.12
Die Räume $W^{1,2}(\Omega)$ und $W_0^{1,2}(\Omega)$ sind Hilbert-Räume mit dem Skalarprodukt

$$(u, v)_{W^{1,2}(\Omega)} := \sum_{0 \leq |\alpha| \leq 1} \int_\Omega D^\alpha u \, D^\alpha v \, dx.$$

Beweis: vgl. [Zei90] Pro. 18.23

In dieser Arbeit werden folgende Schreibweisen verwendet:

$$\begin{aligned} H^1(\Omega) &:= W^{1,2}(\Omega),\ H_0^1(\Omega) := W_0^{1,2}(\Omega), \\ \|\cdot\|_1 &:= \|\cdot\|_{W^{1,2}(\Omega)},\ |\cdot|_1 := |\cdot|_{W^{1,2}(\Omega)}, \\ \|\cdot\|_\infty &:= \|\cdot\|_{L^\infty(\Omega)},\ \|\cdot\|_0 := \|\cdot\|_{L^2(\Omega)}. \end{aligned}$$

Das Poisson-Gleichung (1.1), (1.2) ist ein skalares Problem, d.h. die Bilder der gesuchten Funktionen liegen im $\mathbb{R}$. Die in Kapitel 3 betrachteten Oseen-Gleichungen stellen ein vektorwertiges Problem dar. Aus diesem Grund ist es erforderlich, Normen auf den Räumen $H_0^1(\Omega)^n$, $L^2(\Omega)^n$ und $L^\infty(\Omega)^n$ zu definieren.

Definition 1.13 *Sei $u = (u_1, \ldots, u_n) \in H_0^1(\Omega)^n$ und $a = (a_1, \ldots, a_n) \in L^\infty(\Omega)^n$. Dann sind die Normen definiert durch*

$$\begin{aligned} \|u\|_1 &:= \left(\sum_{i=1}^n \|u_i\|_1^2 \right)^{1/2}, \\ |u|_1 &:= \left(\sum_{i=1}^n |u_i|_1^2 \right)^{1/2}, \\ \|u\|_0 &:= \left(\sum_{i=1}^n \|u_i\|_0^2 \right)^{1/2}, \\ \|a\|_\infty &:= \max_{j=1,\ldots,n} \|a_j\|_\infty. \end{aligned}$$

Kapitel 2

Gemischte Variationsgleichungen

In diesem Kapitel werden einige Aussagen über die Existenz und die Eindeutigkeit der Lösung gemischter Variationsgleichungen gemacht. Das Ziel ist die Anwendung dieser Theorie auf das Oseen-Problem, welches in Kapitel 3 eingeführt wird. Die schwache Formulierung der Oseen-Gleichungen führt auf ein gemischtes Variationsproblem.
Außerdem wird auf die Diskretisierung von gemischten Variationsgleichungen eingegangen, sowie eine kurze Darstellung über die Stabilität und die Konvergenz des diskreten Problems gegeben.

2.1 Existenz und Eindeutigkeit der Lösung gemischter Variationsgleichungen

Seien X und M (reelle) Hilbert-Räume mit den Normen $\|\cdot\|_X$ und $\|\cdot\|_M$. Weiterhin seien X' und M' die zu X und M korrespondierenden Dualräume mit den Dualnormen $\|\cdot\|_{X'}$ und$\|\cdot\|_{M'}$.
Mit $\langle\cdot,\cdot\rangle$ werde das Dualitätsprodukt bezeichnet, welches für $l \in X'$ definiert ist durch

$$\langle l, x\rangle := l(x) \quad \text{für alle } x \in X.$$

Außerdem seien die stetigen Bilinearformen

$$\begin{aligned} a(\cdot,\cdot) &: X \times X \to I\!R, \\ b(\cdot,\cdot) &: X \times M \to I\!R \end{aligned}$$

mit den Normen

$$\begin{aligned} \|a\| &= \sup_{u,v\in X\setminus\{0\}} \frac{a(u,v)}{\|u\|_X\|v\|_X}, \\ \|b\| &= \sup_{u\in X\setminus\{0\},\mu\in M\setminus\{0\}} \frac{b(v,\mu)}{\|v\|_X\|\mu\|_M} \end{aligned}$$

gegeben.

Betrachtet wird nun folgendes Variationsproblem:
Gegeben $l \in X'$, $\chi \in M'$.

$$(Q)\begin{cases} \text{Gesucht } (u,\lambda) \in X \times M \text{ mit} & & & \\ \qquad a(u,v) + b(v,\lambda) & = & \langle l, v\rangle & \text{für alle } v \in X, \\ \qquad b(u,\mu) & = & \langle \chi, \mu\rangle & \text{für alle } \mu \in M. \end{cases}$$

Der Bilinearform $a(\cdot,\cdot)$ wird der Operator A, definiert durch

$$\begin{aligned} A : X &\rightarrow X' \\ \langle Au, v\rangle &= a(u,v) \quad \text{für alle } u, v \in X, \end{aligned}$$

zugeordnet.

Ebenso werden der Bilinearform $b(\cdot,\cdot)$ die Operatoren B und B' zugeordnet

$$\begin{aligned} B : X &\rightarrow M' \\ \langle Bv, \mu\rangle &= b(v,\mu) \quad \text{für alle } v \in X, \mu \in M, \end{aligned}$$

$$\begin{aligned} B' : M &\rightarrow X' \\ \langle B'\mu, v\rangle &= b(v,\mu) \quad \text{für alle } v \in X, \mu \in M. \end{aligned}$$

Mit den Operatoren A, B und B' ergibt sich für das Problem (Q) folgende äquivalente Formulierung:

Gesucht $(u,\lambda) \in X \times M$ mit

$$\begin{aligned} Au + B'\lambda &= l \quad \text{in } X', \\ Bu &= \chi \quad \text{in } M'. \end{aligned}$$

Für die weiteren Betrachtungen sind noch einige Bezeichnungen erforderlich. Die Räume $V(\chi)$ und V seien definiert durch

$$\begin{aligned} V(\chi) &:= \{v \in X : b(v,\mu) = \langle \chi, \mu\rangle \text{ für alle } \mu \in M\}, \\ V &:= V(0) = ker B \\ &= \{v \in X : b(v,\mu) = 0 \text{ für alle } \mu \in M\}. \end{aligned}$$

Wegen der Stetigkeit von $b(\cdot,\cdot)$ ist V ein abgeschlossener Unterraum von X.

In der folgenden Definition werden die Begriffe Polare und orthogonales Komplement von V erläutert.

Definition 2.1
(i) Die Menge $V^0 := \{l \in X' : \langle l, v\rangle = 0$ *für alle* $v \in V\}$ *heißt Polare von* V.
(ii) Das orthogonale Komplement von V *ist definiert durch*

$$V^{\perp} := \{x \in X : (x,v) = 0 \text{ } \textit{für alle}\text{ } v \in V\}.$$

Betrachte nun das restringierte Variationsproblem:

$$(P)\begin{cases} \text{Gesucht } u \in V(\chi) \text{ mit} \\ \qquad\qquad a(u,v) \;=\; \langle l, v\rangle \quad \text{für alle } v \in V. \end{cases}$$

Um Aussagen über die Äquivalenz der Probleme (Q) und (P) machen zu können, sind einige Vorbetrachtungen notwendig. Zuerst soll ein Satz ohne Beweis formuliert werden. Den Nachweis findet man zum Beispiel in [Bra92].

Satz 2.2 (Closed Range Theorem) *Seien Y, Z Banach-Räume. Für eine beschränkte lineare Abbildung $L : Y \to Z$ sind folgende Aussagen äquivalent:*
(i) Das Bild $L(Y)$ ist abgeschlossen in Z.
(ii) Es gilt $L(Y) = (ker L')^0$, wobei $L' : Z' \to Y'$ der zu L adjugierte Operator ist.

Definition 2.3 *Seien U,W normierte Räume. Eine lineare bijektive Abbildung $L : U \to W$ heißt Isomorphismus, wenn L und L^{-1} stetig sind.*

Als weiteres Hilfsmittel ist das folgende Lemma erforderlich.

Lemma 2.4 *Seien U,W Hilbert-Räume. Der Bilinearform $s : U \times W \to \mathbb{R}$ wird der lineare Operator $S : U \to W'$ definiert durch $\langle Su, v\rangle = s(u,v)$ für alle $u \in U, v \in W$ zugeordnet. Die lineare Abbildung $S : U \to W'$ ist genau dann ein Isomorphismus, wenn die zugehörige Form $s : U \times W \to \mathbb{R}$ folgende Bedingungen erfüllt:*
(i) (Stetigkeit)
Es gilt mit $C \geq 0$

$$| s(u,v) | \leq C\|u\|_U \|v\|_W \quad \textit{für alle } u \in U, v \in W. \tag{2.1}$$

(ii) Es gilt mit einem $\alpha > 0$

$$\sup_{v \in W\setminus\{0\}} \frac{s(u,v)}{\|v\|_W} \geq \alpha \|u\|_U \quad \textit{für alle } u \in U. \tag{2.2}$$

(iii) Zu jedem $v \in W\setminus\{0\}$ gibt es ein $u \in U$ mit

$$s(u,v) \neq 0. \tag{2.3}$$

Zusatz:
Werden die Bedingungen (i) und (ii) vorausgesetzt, ist

$$S : U \to \{v \in W : s(u,v) = 0 \textit{ für alle } u \in U\}^0 \subset W'$$

ein Isomorphismus.
Ferner ist (2.2) äquivalent zu

$$\|Su\|_{W'} \geq \alpha \|u\|_U \quad \textit{für alle } u \in U. \tag{2.4}$$

Beweis:
(a) Es seien zunächst die Bedingungen (i) bis (iii) erfüllt. Der Beweis, daß die lineare Abbildung $S : U \to W'$ ein Isomorphismus, sowie der Nachweis des Zusatzes erfolgt nun in mehreren Schritten.

(1) Aus (2.1) folgt für $u \in U$, $v \in W$

$$| \langle Su, v \rangle | \leq C \|u\|_U \|v\|_W. \tag{2.5}$$

Weiterhin ist

$$\begin{aligned} \|Su\|_{W'} &= \sup_{v \in W \setminus \{0\}} \frac{| \langle Su, v \rangle |}{\|v\|_W} \\ &\leq \sup_{v \in W \setminus \{0\}} \frac{C \|u\|_U \|v\|_W}{\|v\|_W} \qquad \text{(wegen (2.5))} \\ &= C \|u\|_U. \end{aligned}$$

Daraus folgt also, daß S stetig ist.

(2) Sei $u \in kerS$. Dann gilt $\langle Su, v \rangle = 0$ für alle $v \in W$.
Also folgt daraus

$$\sup_{v \in W \setminus \{0\}} \frac{s(u,v)}{\|v\|_W} = 0.$$

Mit (2.2) gilt dann

$$\begin{aligned} 0 &= \sup_{v \in W \setminus \{0\}} \frac{s(u,v)}{\|v\|_W} \\ &\overset{(2.2)}{\geq} \underbrace{\alpha}_{>0} \underbrace{\|u\|_U}_{\geq 0} \geq 0. \end{aligned}$$

Somit erhält man $u = 0$. Damit ist die Injektivität von S gezeigt.

(3) Es gilt für alle $u \in U$

$$\begin{aligned} \|Su\|_{W'} &= \sup_{v \in W \setminus \{0\}} \frac{\langle Su, v \rangle}{\|v\|_W} = \sup_{v \in W \setminus \{0\}} \frac{s(u,v)}{\|v\|_W} \\ &\geq \alpha \|u\|_U. \end{aligned} \tag{2.6}$$

Damit ist die Äquivalenz zwischen (2.2) und (2.4) gezeigt.
Sei nun $f \in S(U)$. Wegen der in (2) gezeigten Injektivität gibt es ein eindeutiges Inverses $u = S^{-1} f$. Offensichtlich ist S^{-1} linear. Aus (2.6) folgt dann

$$\|S^{-1} f\|_U \leq \frac{1}{\alpha} \|f\|_{W'}.$$

Also ist S^{-1} auf dem Bild von S beschränkt. Wegen der Linearität von S^{-1} folgt daraus dann, daß S^{-1} auf dem Bild von S stetig ist.

(4) Sei $\{f_n\}$ eine Folge in $S(U)$ mit $f_n \to f$, also gilt $\|f_n - f\| \to 0$. Dann existiert eine Folge $\{u_n\}$ in U mit $Su_n = f_n$. Weiterhin gilt

$$\begin{aligned} \|u_n - u_m\| &= \|S^{-1}f_n - S^{-1}f_m\| \\ &= \|S^{-1}(f_n - f_m)\| \quad (\text{da } S^{-1} \text{ linear}) \\ &\leq \|S^{-1}\| \underbrace{\|f_n - f_m\|}_{\to 0} \quad (\text{da } S^{-1} \text{ auf dem Bild von } S \text{ stetig }). \end{aligned}$$

Dann gilt $\|u_n - u_m\| \to 0$, also ist $\{u_n\}$ eine Cauchyfolge. Da U vollständig ist, konvergiert u_n gegen $u \in U$.
Nun bleibt zu zeigen, daß $Su = f$ ist.
Es gilt

$$\begin{aligned} \|Su - f\| &\leq \|Su - Su_n\| + \|Su_n - f\| \\ &\leq \|S\| \underbrace{\|u - u_n\|}_{\to 0} + \underbrace{\|f_n - f\|}_{\to 0}. \end{aligned}$$

Folglich ergibt sich $Su = f$, also ist $S(U)$ abgeschlossen.
Die Abbildung $S' : W \longrightarrow U'$ ist definiert durch $\langle S'v, u\rangle := s(u, v)$ für alle $v \in W$, $u \in U$.
Aus der Abgeschlossenheit von $S(U)$ folgt dann mit dem Satz 2.2, daß $S(U) = (kerS')^0$ ist. Nun gilt

$$\begin{aligned} kerS' &= \{v \in W : \langle S'v, u\rangle = 0 \text{ für alle } u \in U\} \\ &= \{v \in W : s(v, u) = 0 \text{ für alle } u \in U\}. \end{aligned}$$

Also ist $(kerS')^0 = \{v \in W : s(v, u) = 0 \text{ für alle } u \in U\}^0 \subset W'$.

In den Beweisschritten (3) und (4) ist der Zusatz des Satzes bewiesen worden.

(5) Sei $M = \{0\} \subset W$. Dann ist

$$\begin{aligned} M^0 &= \{l \in W' : \langle l, m\rangle = 0 \text{ für alle } m \in M\} \\ &= \{l \in W' : \langle l, 0\rangle = 0\} \\ &= W'. \end{aligned} \tag{2.7}$$

Außerdem gilt aber wegen Bedingung (iii), daß $\{v \in W : s(u, v) = 0 \text{ für alle } u \in U\} = \{0\}$ ist. Aus (2.7) folgt dann $\{v \in W : s(u, v) = 0 \text{ für alle } u \in U\}^0 = \{0\}^0 = W'$. Folglich ist $S(U) = W'$, somit ist S surjektiv.

Im Teil (a) des Beweises ist also gezeigt worden, daß die Abbildung S ein Isomorphismus ist, wenn die Bedingungen (i) bis (iii) erfüllt sind.

(b) Umgekehrt ist nun zu zeigen, daß die Bedingungen (i) bis (iii) gelten, wenn

$S : U \to W'$ ein Isomorphismus ist.
Die Abbildung S sei also ein Isomorphismus. Dann gilt:
(i)

$$\begin{aligned} | \, s(u,v) \, | &= | \, \langle \underbrace{Su}_{\in W'}, v \rangle \, | \\ &\leq \|Su\|_{W'} \|v\|_W \\ &\leq \|S\| \, \|u\|_U \, \|v\|_W \quad \text{(da } S \text{ stetig)} \end{aligned}$$

(ii) Nach Teil (3) gilt

$$\sup_{v \in W \setminus \{0\}} \frac{s(u,v)}{\|v\|_W} = \|Su\|_{W'}.$$

Da $\|u\|_U = \|S^{-1}Su\|_U \leq \|S^{-1}\| \, \|Su\|_{W'}$, folgt

$$\sup_{v \in W \setminus \{0\}} \frac{s(u,v)}{\|v\|_W} \geq \underbrace{\frac{1}{\|S^{-1}\|}}_{=:\alpha} \|u\|_U.$$

(iii) Sei $v \in W \setminus \{0\}$, weiterhin sei $R : W' \to W$ der Rieszsche Darstellungsoperator. Folglich ist $R^{-1}v \in W'$. Da $S : U \to W'$ ein Isomorphismus ist, existiert ein $u \in U$ mit $Su = R^{-1}v$. Dann ist

$$\begin{aligned} s(u,v) = \langle Su, v \rangle &= \langle R^{-1}v, v \rangle \\ &= (v,v) \neq 0. \end{aligned}$$

Also gibt es zu jedem $v \in W \setminus \{0\}$ ein $u \in U$ mit $s(u,v) \neq 0$.

□

Mit Hilfe dieses Lemmas ist es nun möglich, einen weiteren Satz zu beweisen, der erforderlich ist, um Aussagen über die Existenz und Eindeutigkeit der Lösung von gemischten Variationsgleichungen machen zu können.

Satz 2.5 *Folgende Aussagen sind äquivalent.*
(i) Es existiert ein $\beta > 0$ mit

$$\inf_{\mu \in M \setminus \{0\}} \sup_{v \in X \setminus \{0\}} \frac{b(v,\mu)}{\|v\|_X \|\mu\|_M} \geq \beta. \tag{2.8}$$

Diese Bedingung wird im folgenden als inf-sup-Bedingung bezeichnet.
(ii) Sei $V^\perp$ das orthogonale Komplement von V. Die Abbildung $B : V^\perp \to M'$ ist ein Isomorphismus, und es gilt

$$\|Bv\|_{M'} \geq \beta \|v\|_X \quad \textit{für alle } v \in V^\perp. \tag{2.9}$$

(iii) Sei V^0 die Polare von V. Die Abbildung $B' : M \to V^0$ ist ein Isomorphismus, und es gilt

$$\|B'\mu\|_{X'} \geq \beta \|\mu\|_M \quad \textit{für alle } \mu \in M. \tag{2.10}$$

Beweis:
(a) In diesem Schritt wird die Äquivalenz der Aussagen (i) und (iii) gezeigt.
Es existiert ein $\beta > 0$ mit

$$\inf_{\mu \in M\setminus\{0\}} \sup_{v \in X\setminus\{0\}} \frac{b(v,\mu)}{\|v\|_X \|\mu\|_M} \geq \beta.$$

Äquivalent dazu ist

$$\sup_{v \in X\setminus\{0\}} \frac{b(v,\mu)}{\|v\|_X} \geq \beta\|\mu\|_M \quad \text{für alle } \mu \in M. \tag{2.11}$$

Nach dem Zusatz zu Lemma 2.4 gilt, daß (2.11) äquivalent zu $\|B'\mu\|_{X'} \geq \beta\|\mu\|_M$ für alle $\mu \in M$ ist. Somit ist gezeigt, daß (2.10) äquivalent zur inf-sup-Bedingung (2.8) ist. Da $b(\cdot,\cdot)$ eine stetige Bilinearform ist und (2.11) gilt, folgt ebenfalls mit dem Zusatz zum Lemma 2.4, daß $B' : M \to \{v \in X : b(v,\mu) = 0 \text{ für alle } \mu \in M\}^0 = V^0 \subset X'$ ein Isomorphismus ist.

(b) In diesem Beweisschritt wird gezeigt, daß aus Aussage (iii) die Eigenschaft (ii) folgt.
Sei $V = \{v \in X : b(v,\mu) = 0 \text{ für alle } \mu \in M\}$ und
$V^\perp := \{x \in X : (x,v) = 0 \text{ für alle } v \in V\}$. Weiterhin sei $v \in V^\perp$ beliebig. Dann ist durch $g(w) = (v,w)$ ein Funktional $g \in X'$ mit $\|g\|_{X'} = \|v\|_X$ definiert. Da $v \in V^\perp$ gilt $g(w) = (v,w) = 0$ für alle $w \in V$. Daraus folgt dann, daß
$g \in V^0 := \{l \in X' : \langle l, v\rangle = 0 \text{ für alle } v \in V\}$.
Da $B' : M \to V^0 \subset X'$ ein Isomorphismus ist, gibt es ein $\lambda \in M$ mit

$$\langle B'\lambda, w\rangle = b(w,\lambda) = g(w) = (v,w) \quad \text{für alle } w \in X. \tag{2.12}$$

Weiterhin gilt

$$\begin{aligned} \|v\|_X = \|g\|_{X'} &= \sup_{w \in X\setminus\{0\}} \frac{\langle g, w\rangle}{\|w\|_X} \\ &= \sup_{w \in X\setminus\{0\}} \frac{\langle B'\lambda, w\rangle}{\|w\|_X} \\ &= \|B'\lambda\|_{X'} \overset{(2.10)}{\geq} \beta\|\lambda\|_M. \end{aligned}$$

Also ist

$$\|v\|_X \geq \beta\|\lambda\|_M. \tag{2.13}$$

Setze in (2.12) $w = v$. Dann gilt

$$\langle B'\lambda, v\rangle = b(v,\lambda) = (v,v). \tag{2.14}$$

Somit erhält man für $\lambda \in M$

$$\begin{aligned} \|Bv\|_{M'} &= \sup_{\mu \in M\setminus\{0\}} \frac{\langle Bv, \mu\rangle}{\|\mu\|_M} = \sup_{\mu \in M\setminus\{0\}} \frac{b(v,\mu)}{\|\mu\|_M} \\ &\geq \frac{b(v,\lambda)}{\|\lambda\|_M} \overset{(2.14)}{=} \frac{(v,v)}{\|\lambda\|_M} \\ &= \frac{\|v\|_X^2}{\|\lambda\|_M} \overset{(2.13)}{\geq} \|v\|_X \beta \frac{\|\lambda\|_M}{\|\lambda\|_M} = \beta \|v\|_X. \end{aligned}$$

Damit ist (2.9) gezeigt.
Es gelten die drei Bedingungen des Lemma 2.4, nämlich:
(1) $b(\cdot,\cdot)$ ist stetig.
(2) Es gilt mit einem $\beta > 0$

$$\sup_{\mu \in M\setminus\{0\}} \frac{b(v,\mu)}{\|\mu\|_M} \geq \beta \|v\|_X \quad \text{für alle } v \in V^{\perp}.$$

(3) Zu jedem $\mu \in M\setminus\{0\}$ gibt es ein $v \in V^{\perp}$ mit $b(v,\mu) \neq 0$, denn:
Sei $\mu \in M$ mit $b(v,\mu) = 0$ für alle $v \in V^{\perp}$. Weiterhin gilt

$$\langle B'\mu, v_0\rangle = b(v_0,\mu) = 0 \quad \text{für alle } v_0 \in V.$$

Außerdem ist $X = V \bigoplus V^{\perp}$ (vgl. [Gri81] S.193), also läßt sich jedes $x \in X$ darstellen als $x = v_0 + v$ mit $v \in V^{\perp}$ und $v_0 \in V$. Damit gilt dann für alle $x \in X$

$$\begin{aligned} \langle B'\mu, x\rangle &= \langle B'\mu, v_0 + v\rangle \\ &= \underbrace{\langle B'\mu, v_0\rangle}_{=0} + \underbrace{\langle B'\mu, v\rangle}_{0} \\ &= 0. \end{aligned}$$

Daraus folgt dann, daß $B'\mu = 0$ ist. Also ist $\mu \in kerB'$. Da aber $B' : M \to V^0$ ein Isomorphismus ist, ist $\mu = 0$. Somit gibt es zu jedem $\mu \in M\setminus\{0\}$ ein $v \in V^{\perp}$ mit $b(v^{\perp},\mu) \neq 0$.
Somit erhält man unter Verwendung von Lemma 2.4, daß die Abbildung $B : V^{\perp} \longrightarrow M'$ ein Isomorphismus ist.

(c) Nun wird nachgewiesen, daß die Aussage (i) aus (ii) folgt.
Sei also $B : V^{\perp} \to M'$ ein Isomorphismus. Außerdem sei $\mu \in M$ gegeben, weiterhin sei $R : M' \to M$ der Rieszoperator. Dann gilt

$$\begin{aligned} \|\mu\|_M &= \|R^{-1}\mu\|_{M'} = \sup_{\alpha \in M} \frac{\langle R^{-1}\mu, \alpha\rangle}{\|\alpha\|_M} \\ &= \sup_{\alpha \in M} \frac{(\mu,\alpha)}{\|\alpha\|_M} = \sup_{g \in M'} \frac{(\mu, Rg)}{\|Rg\|_M} \end{aligned}$$

$$
\begin{aligned}
&= \sup_{g\in M'} \frac{(Rg,\mu)}{\|g\|_{M'}} = \sup_{g\in M'} \frac{\langle R^{-1}Rg,\mu\rangle}{\|g\|_{M'}} \\
&= \sup_{v\in V^\perp} \frac{\langle Bv,\mu\rangle}{\|Bv\|_{M'}} \quad \text{(da } B \text{ Isomorphismus)} \\
&= \sup_{v\in V^\perp} \frac{b(v,\mu)}{\|Bv\|} \le \sup_{v\in V^\perp} \frac{b(v,\mu)}{\beta\|v\|_X}.
\end{aligned}
$$

Damit gilt dann für alle $\mu \in M$

$$
\|\mu\|_M \le \sup_{v\in V^\perp} \frac{b(v,\mu)}{\beta\|v\|_X} \le \sup_{v\in X\setminus\{0\}} \frac{b(v,\mu)}{\beta\|v\|_X}.
$$

Folglich ist

$$
\beta \le \inf_{\mu\in M\setminus\{0\}} \sup_{v\in X\setminus\{0\}} \frac{b(v,\mu)}{\|v\|_X\|\mu\|_M}.
$$

Damit ist (i) gezeigt.

Insgesamt folgen aus (a), (b) und (c) sämtliche Äquivalenzen.

□

Im folgenden Satz wird gezeigt unter welchen Voraussetzungen die Probleme (P) und (Q) äquivalent sind.

Satz 2.6
(i) Ist $(u,\lambda) \in X \times M$ Lösung von (Q), dann löst $u \in X$ das Problem (P).
(ii) Es existiert ein $\beta > 0$ mit

$$
\inf_{\mu\in M\setminus\{0\}} \sup_{v\in X\setminus\{0\}} \frac{b(v,\mu)}{\|v\|_X\|\mu\|_M} \ge \beta. \tag{2.15}
$$

Dann gibt es zu jeder Lösung $u \in X$ von (P) ein eindeutig bestimmtes $\lambda \in M$, so daß (u,λ) Lösung von (Q) ist.

Beweis:
(i) Sei $(u,\lambda) \in X \times M$ Lösung von (Q). Dann gilt also

$$
b(u,\mu) = \langle \chi,\mu\rangle \quad \text{für alle } \mu \in M.
$$

Also ist $u \in V(\chi)$. Wähle nun ein $v \in V \subset X$. Da $\lambda \in M$ ist, gilt dann $b(v,\lambda) = 0$. Mit der ersten Gleichung aus (Q) folgt

$$
a(u,v) = \langle l,v\rangle \quad \text{für alle } v \in V.
$$

Also löst u das Problem (P).

(ii) Sei $u \in V(\chi)$ Lösung von (P). Daraus folgt dann

$$a(u,v) = \langle l, v\rangle \quad \text{für alle } v \in V$$

und

$$b(u,\mu) = \langle \chi, \mu\rangle \quad \text{für alle } \mu \in M.$$

Nun bleibt zu zeigen, daß es ein $\lambda \in M$ mit $B'\lambda = l - Au$ gibt.
Es ist für $v \in V$

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

Daraus folgt dann $l - Au \in V^0$.
Nach dem Satz 2.5 ist die Bedingung (2.15) äquivalent zu der Aussage, daß die Abbildung $B' : M \to V^0$ ein Isomorphismus ist. Also existiert eindeutig ein $\lambda \in M$ mit $B'\lambda = l - Au$. Daraus folgt dann (u,λ) löst das Problem (Q).

□

Eine wichtige Rolle in der Lösbarkeitstheorie gemischter Variationsgleichungen spielt die inf-sup-Bedingung. Dies wird im folgenden Satz, der Ausssagen über die Existenz und Eindeutigkeit der Lösung gemischter Variationsgleichungen enthält, deutlich. Ein weiteres Hilfsmittel zum Beweis des Satzes ist das Lemma von Lax-Milgram.

Lemma 2.7 (Lax-Milgram) *Sei V ein Hilbert-Raum und V' der zugehörige Dualraum. Weiterhin sei $a(\cdot,\cdot) : V \times V \longrightarrow \mathbb{R}$ eine stetige, V-elliptische Bilinearform. Dann besitzt für jedes $f \in V'$ die Variationsgleichung*

$$a(u,v) = f(v) \quad \textit{für alle } v \in V$$

eine eindeutige Lösung $u \in V$.

Beweis: vgl. [GR94] Lemma 3.6

Mit dem Lemma 2.7 und den Sätzes 2.5 sowie 2.6 kann nun ein Satz über die Lösbarkeit gemischter Variationsgleichungen bewiesen werden.

Satz 2.8 *Sei $a(\cdot,\cdot)$ eine V-elliptische Bilinearform, d.h. es existiert ein $\alpha > 0$ mit*

$$a(v,v) \geq \alpha \|v\|_V^2 \quad \textit{für alle } v \in V.$$

Dann sind folgende Aussagen äquivalent:
(i)

$$\begin{aligned} L : X \times M &\to X' \times M' \\ (u,\lambda) &\mapsto (Au + B'\lambda, Bu) \end{aligned}$$

ist ein Isomorphismus.

(ii) $b(\cdot,\cdot)$ erfüllt die inf-sup-Bedingung, d.h. es gibt ein $\beta > 0$ mit

$$\inf_{\mu \in M\setminus\{0\}} \sup_{v \in X\setminus\{0\}} \frac{b(v,\mu)}{\|v\|_X \|\mu\|_M} \geq \beta.$$

Beweis:
(a) Sei $b(\cdot,\cdot)$ eine Bilinearform, die die inf-sup-Bedingung erfüllt. Zuerst wird gezeigt, daß zu jedem Paar von Funktionalen $(l,\chi) \in X' \times M'$ genau eine Lösung $(u,\lambda) \in X \times M$ des Problems (Q) mit folgenden Eigenschaften existiert:

$$\|u\|_X \leq \alpha^{-1}\|l\|_{X'} + \beta^{-1}\left(1+\frac{C}{\alpha}\right)\|\chi\|_{M'}, \tag{2.16}$$

$$\|\lambda\|_M \leq \beta^{-1}\left(1+\frac{C}{\alpha}\right)\|l\|_{X'} + \beta^{-1}\left(1+\frac{C}{\alpha}\right)\frac{C}{\beta}\|\chi\|_{M'}. \tag{2.17}$$

Sei $\chi \in M'$. Wegen Satz 2.5 folgt aus (ii), daß die Abbildung $B : V^\perp \to M'$ ein Isomorphismus ist. Also existiert ein $u_0 \in V^\perp$ mit $Bu_0 = \chi$. Folglich gilt dann für alle $\mu \in M$

$$b(u_0,\mu) = \langle Bu_0,\mu\rangle = \langle\chi,\mu\rangle. \tag{2.18}$$

Da $b(u_0,\mu) = \langle\chi,\mu\rangle$ für alle $\mu \in M$ gilt, ist $u_0 \in V(\chi)$. Somit ist $V(\chi) \neq \emptyset$.

Weiterhin gilt wegen (2.9)

$$\|Bu_0\|_{M'} \geq \beta\|u_0\|_X.$$

Daraus folgt dann

$$\beta^{-1}\|\chi\|_{M'} \geq \|u_0\|_X. \tag{2.19}$$

Betrachte nun die Gleichung $a(u,v) + b(v,\lambda) = \langle l,v\rangle$ für alle $v \in X$. Durch Nullergänzung erhält man dann

$$a(u,v) + a(u_0,v) - a(u_0,v) + b(v,\lambda) = \langle l,v\rangle \quad \text{für alle } v \in X.$$

Äquivalent dazu ist

$$a(\underbrace{u-u_0}_{=:w},v) + b(v,\lambda) = \langle l,v\rangle - a(u_0,v) \quad \text{für alle } v \in X. \tag{2.20}$$

Betrachte außerdem die Gleichung $b(u,\mu) = \langle\chi,\mu\rangle$ für alle $\mu \in M$. Wiederum durch Nullergänzung erhält man

$$b(u,\mu) + b(u_0,\mu) - b(u_0,\mu) = \langle\chi,\mu\rangle \quad \text{für alle } \mu \in M.$$

Äquivalent dazu ist

$$\begin{aligned} b(\underbrace{u-u_0}_{=:w},\mu) &= \langle\chi,\mu\rangle - \underbrace{b(u_0,\mu)}_{=\langle\chi,\mu\rangle} \quad \text{(wegen (2.18))} \\ \Leftrightarrow b(w,\mu) &= 0 \quad \text{für alle } \mu \in M. \end{aligned} \tag{2.21}$$

Aus (2.20) und (2.21) folgt nun, daß (Q) äquivalent zu folgendem Problem ist:

$$(Q') \begin{cases} \text{Gesucht } (w,\lambda) \text{ mit} \\ \quad a(w,v) + b(v,\lambda) &= \langle l, v\rangle - a(u_0,v) \quad \text{für alle } v \in X, \\ \quad b(w,\mu) &= 0 \quad \text{für alle } \mu \in M. \end{cases}$$

Betrachte nun das Problem (P'):

$$(P') \begin{cases} \text{Gesucht } w \in V \text{ mit} \\ \quad a(w,v) &= \langle l, v\rangle - a(u_0,v) \quad \text{für alle } v \in V. \end{cases}$$

Da $a(\cdot,\cdot)$ V-elliptisch ist, hat das Problem (P') nach Lemma 2.7 eine eindeutige Lösung $w \in V$. Wie im Satz 2.6 gezeigt, existiert ein eindeutig bestimmtes $\lambda \in M$, so daß (w,λ) das Problem (Q') löst. Für eine Lösung (w,λ) von (Q') gilt aber, daß w das Problem (P') löst. Also ist die Lösung von (Q') eindeutig. Damit hat (Q) eine eindeutige Lösung (u,λ). Für die Lösung $w \in V$ des Problems (P') gilt weiterhin

$$\begin{aligned} \|w\|_X^2 &\le \alpha^{-1} a(w,w) \\ &= \alpha^{-1}(\langle l, w\rangle - a(u_0,w)) \\ &\le \alpha^{-1}(\|l\|_{X'}\|w\|_X + C\|u_0\|_X\|w\|_X) \\ &= \alpha^{-1}(\|l\|_{X'} + C\|u_0\|_X)\|w\|_X. \end{aligned}$$

Folglich ist

$$\|w\|_X \le \alpha^{-1}(\|l\|_{X'} + C\|u_0\|_X). \tag{2.22}$$

Sei nun (u,λ) Lösung von (Q). Da die inf-sup-Bedingung gilt, folgt aus Satz 2.5 $\|B'\lambda\|_{X'} \ge \beta\|\lambda\|_M$, also $\beta^{-1}\|B'\lambda\|_{X'} \ge \|\lambda\|_M$.

Für $\|B'\lambda\|_{X'}$ erhält man

$$\begin{aligned} \|B'\lambda\|_{X'} &= \sup_{v\in X\setminus\{0\}} \frac{\langle B'\lambda, v\rangle}{\|v\|_X} \\ &= \sup_{v\in X\setminus\{0\}} \frac{b(v,\lambda)}{\|v\|_X} \\ &= \sup_{v\in X\setminus\{0\}} \frac{\langle l, v\rangle - a(u,v)}{\|v\|_X} \quad (\text{da } w := u - u_0) \\ &\le \sup_{v\in X\setminus\{0\}} \frac{\|l\|_{X'}\|v\|_X + C\|u\|_X\|v\|_X}{\|v\|_X} \\ &= \|l\|_{X'} + C\|u\|_X. \end{aligned}$$

Damit folgt

$$\|\lambda\|_M \le \beta^{-1}(\|l\|_{X'} + C\|u\|_X). \tag{2.23}$$

Weiterhin gilt für $u = u_0 + w$

$$\|u\|_X = \|u_0 + w\|_X \leq \|u_0\|_X + \|w\|_X.$$

Unter Benutzung von (2.19) und (2.22) folgt dann

$$\begin{aligned}
\|u\|_X &\leq \|u_0\|_X + \|w\|_X \\
&\leq \beta^{-1}\|\chi\|_{M'} + \alpha^{-1}(\|l\|_{X'} + C\|u_0\|_X) \\
&\leq \beta^{-1}\|\chi\|_{M'} + \alpha^{-1}(\|l\|_{X'} + C\beta^{-1}\|\chi\|_{M'}) \\
&= \alpha^{-1}\|l\|_{X'} + \beta^{-1}\left(1 + \frac{C}{\alpha}\right)\|\chi\|_{M'}.
\end{aligned}$$

Unter Verwendung von (2.23) bekommt man

$$\begin{aligned}
\|\lambda\|_M &\leq \beta^{-1}(\|l\|_{X'} + C\|u\|_X) \\
&\leq \beta^{-1}\left(\|l\|_{X'} + C\left(\alpha^{-1}\|l\|_{X'} + \beta^{-1}\left(1 + \frac{C}{\alpha}\right)\|\chi\|_{M'}\right)\right) \\
&= \beta^{-1}\|l\|_{X'} + \beta^{-1}\frac{C}{\alpha}\|l\|_{X'} + \beta^{-1}\left(1 + \frac{C}{\alpha}\right)\frac{C}{\beta}\|\chi\|_{M'} \\
&= \beta^{-1}\left(1 + \frac{C}{\alpha}\right)\|l\|_{X'} + \beta^{-1}\left(1 + \frac{C}{\alpha}\right)\frac{C}{\beta}\|\chi\|_{M'}.
\end{aligned}$$

Die Abbildung L ist also bijektiv und erfüllt die Eigenschaften (2.16) und (2.17). Weiterhin ist L^{-1} stetig, denn:

Definiere folgende Norm auf $X \times M$: $\|(u, \lambda)\|_{X\times M} = \|u\|_X + \|\lambda\|_M$.
Dann gilt

$$\begin{aligned}
\|L^{-1}(l, \chi)\|_{X\times M} &= \|(u, \lambda)\|_{X\times M} = \|u\|_X + \|\lambda\|_M \\
&\leq \alpha^{-1}\|l\|_{X'} + \beta^{-1}\left(1 + \frac{C}{\alpha}\right)\|\chi\|_{M'} \\
&\quad + \beta^{-1}\left(1 + \frac{C}{\alpha}\right)\|l\|_{X'} + \beta^{-1}\left(1 + \frac{C}{\alpha}\right)\frac{C}{\beta}\|\chi\|_{M'} \\
&= \left(\alpha^{-1} + \beta^{-1}\left(1 + \frac{C}{\alpha}\right)\right)\|l\|_{X'} \\
&\quad + \left(\beta^{-1}\left(1 + \frac{C}{\alpha}\right) + \beta^{-1}\left(1 + \frac{C}{\alpha}\right)\frac{C}{\beta}\right)\|\chi\|_{M'}.
\end{aligned}$$

Da L^{-1} bijektiv und stetig ist, ist auch L stetig. Also ist L ein Isomorphismus.

(b) Der Beweis, daß die inf-sup-Bedingung aus der Isomorphismuseigenschaft von L folgt, findet sich in [Bra92].

□

2.2 Diskretisierung gemischter Variationsgleichungen

In diesem Abschnitt wird das diskrete Problem eingeführt, außerdem werden einige Aussagen über Stabilität und Konvergenz dieses Problems gemacht. Es erfolgt meistens nur eine Darstellung der Fakten. Auf Beweise wird weitestgehend verzichtet, jedoch werden die entsprechenden Referenzen angegeben.

2.2.1 Das diskrete Problem

Zur Diskretisierung von gemischten Variationsgleichungen werden endlich dimensionale Teilräume $X_h \subset X,\ M_h \subset M$ eingeführt. Das Problem (Q) aus Kapitel 2.1 wird dann approximiert durch:

$$(\mathrm{Q}_h)\begin{cases} \text{Gesucht } (u_h, \lambda_h) \in X_h \times M_h \text{ mit} & & & \\ \qquad a(u_h, v_h) + b(v_h, \lambda_h) & = & \langle l, v_h \rangle & \text{für alle } v_h \in X_h, \\ \qquad b(u_h, \mu_h) & = & \langle \chi_h, \mu_h \rangle & \text{für alle } \mu_h \in M_h. \end{cases}$$

Analog zu $V(\chi)$ definiert man

$$\begin{aligned} V_h(\chi) &:= \{v_h \in X_h : b(v_h, \mu_h) =< \chi, \mu_h > \text{ für alle } \mu_h \in M_h\} \\ \text{und} & \\ V_h &:= V_h(0). \end{aligned}$$

Ebenso wird das Problem (P) aus Kapitel 2.1 approximiert durch:

$$(\mathrm{P}_h)\begin{cases} \text{Gesucht } u_h \in V_h(\chi) \text{ mit} & & \\ \qquad a(u_h, v_h) & = & \langle l, v_h \rangle \text{ für alle } v_h \in V_h. \end{cases}$$

Bemerkung 2.9
Im allgemeinen gilt nicht $V_h(\chi) \subset V(\chi)$, da M_h ein echter Unterraum von M ist!

Das diskrete Problem (Q_h) kann in ein lineares Gleichungssystem überführt werden.

Sei $\{\phi_1, \ldots, \phi_l\}$ eine Basis von X_h und $\{\psi_1, \ldots, \psi_k\}$ eine Basis von M_h. Dann ergibt sich für $u_h \in X_h$ bzw. $\lambda_h \in M_h$ die Basisdarstellung

$$u_h = \sum_{i=1}^{l} U_i \phi_i,$$

$$\lambda_h = \sum_{j=1}^{k} \Lambda_j \psi_j.$$

Setze nun $U := (U_1, \ldots, U_l)$ bzw. $\Lambda := (\Lambda_1, \ldots, \Lambda_k)$.
Definiere weiterhin

$$A \in \mathbb{R}^{l \times l} : A_{ij} := a(\phi_j, \phi_i),$$

$$B \in \mathbb{R}^{k\times l} : B_{ij} := b(\phi_j, \psi_i),$$
$$L \in \mathbb{R}^{l} : L_i := \langle l, \phi_i \rangle,$$
$$N \in \mathbb{R}^{k} : N_i := \langle \chi, \psi_i \rangle.$$

Man kann damit nachrechen, daß

$$\begin{aligned} (AU, V) &= a(u_h, v_h) \text{ für alle } u_h, v_h \in X_h, \\ (BU, \Lambda) &= b(u_h, \lambda_h) \text{ für alle } u_h \in X_h,\ \lambda_h \in M_h. \end{aligned}$$

Daher ist das zu (Q_h) äquivalente Problem :

Finde $(U, \Lambda) \in \mathbb{R}^l \times \mathbb{R}^k$ mit

$$\begin{aligned} AU + B^T\Lambda &= L, \\ BU &= N \end{aligned}$$

oder anders geschrieben

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} U \\ \Lambda \end{pmatrix} = \begin{pmatrix} L \\ N \end{pmatrix}.$$

Damit hat man ein zu dem diskretisierten Problem (Q_h) äquivalentes lineares Gleichungssystem.
In Abschnitt 3.3 wird dann speziell das zum Oseen-Problem äquivalente Gleichungssystem vorgestellt.

2.2.2 Stabilität und Konvergenz des diskreten Problems

Aus dem Satz 2.8 über die eindeutige Lösbarkeit des kontinuierlichen Problems erhält man bei Anwendung auf die diskreten Räume und unter Beachtung von $\|l\|_{X_h'} \le \|l\|_{X'}$ und $\|\chi\|_{M_h'} \le \|\chi\|_{M'}$

Satz 2.10 *Sei $a(\cdot,\cdot)$ V_h-elliptisch, d.h.*

$$a(v_h, v_h) \ge \alpha_h \|v_h\|_X^2 \text{ für alle } v_h \in V_h,$$

und $b(\cdot,\cdot)$ erfülle die Babuška-Brezzi-Bedingung

$$\sup_{v_h \in X_h} \frac{b(v_h, \mu_h)}{\|v_h\|_X} \ge \beta_h \|\mu_h\|_M \quad \text{für alle } \mu_h \in M_h.$$

Dann ist das diskrete Problem (Q_h) eindeutig lösbar.

Für die Lösung (u_h, λ_h) des Problems (Q_h) gelten die Abschätzungen

$$\begin{aligned} \|u_h\|_X &\leq \alpha_h^{-1}\|l\|_{X'} + \beta_h^{-1}\left(1+\frac{C}{\alpha_h}\right)\|\chi\|_{M'}, \\ \|\lambda_h\|_M &\leq \beta_h^{-1}\left(1+\frac{C}{\alpha_h}\right)\|l\|_{X'} + \beta_h^{-1}\left(1+\frac{C}{\alpha}\right)\frac{C}{\beta}\|\chi\|_{M'}. \end{aligned}$$

Bemerkung 2.11
Sind α_h und β_h unabhängig von h, so hat man ein Stabilitätsresultat. Man bezeichnet die Diskretisierung in diesem Fall als stabil.

Zum Abschluß dieses Abschnitts soll noch ein Konvergenzresultat formuliert werden. Auf den Beweis des Satzes wird an dieser Stelle nicht eingegangen, man findet ihn zum Beispiel in [GR86] Theorem 1.1.

Satz 2.12 *Die Voraussetzungen des letzten Satzes seien erfüllt. Weiterhin sei $a(\cdot,\cdot)$ V-elliptisch und $b(\cdot,\cdot)$ erfülle die inf-sup-Bedingung.*
Dann gilt für die Lösung (u,λ) von (Q) bzw. (u_h,λ_h) von (Q_h) :

(i) Es existiert eine Konstante C_1 (abhängig von $\alpha_h, \|a\|, \|b\|$) mit

$$\|u-u_h\|_X \leq C_1\left\{\inf_{v_h\in V_h(\chi)}\|u-v_h\|_X + \inf_{\mu_h\in M_h}\|\lambda-\mu_h\|_M\right\}.$$

(ii) Es gibt eine Konstante C_2 (abhängig von $\alpha_h, \beta_h, \|a\|, \|b\|$) mit

$$\|u-u_h\|_X + \|\lambda-\lambda_h\|_M \leq C_2\left\{\inf_{v_h\in X_h}\|u-v_h\|_X + \inf_{\mu_h\in M_h}\|\lambda-\mu_h\|_M\right\}.$$

Kapitel 3

Die Oseen-Gleichungen

Die Navier-Stokes-Gleichungen für inkompressible Flüssigkeiten zählen zu den wichtigsten Gleichungen der Strömungsmechanik. Sei u das Geschwindigkeitsfeld und p der Druck, dann wird

$$-\nu \triangle u + (u \cdot \nabla)u + \nabla p = f \quad \text{in } \Omega$$

als Navier-Stokes-Problem bezeichnet. Die Inkompressibilitätsbedingung ist

$$div\, u = 0 \quad \text{in } \Omega.$$

Der Term $(u \cdot \nabla)u$ macht deutlich, daß es sich um ein nichtlineares Problem handelt. Durch Verwendung der Iterationsvorschrift

$$-\nu \triangle u^{n+1} + (u^n \cdot \nabla)u^{n+1} + \nabla p^{n+1} = f$$

entstehen in jedem Iterationsschritt linearisierte Gleichungen, die auch als Oseen-Gleichungen bezeichnet werden. In die folgenden Betrachtungen wird zusätzlich ein ableitungsfreier Term für die Geschwindigkeit einbezogen, der bei der Zeitdiskretisierung zeitabhängiger Probleme entsteht.

Betrachtet wird also das folgende Problem:

- $\Omega \subset \mathbb{R}^n$ sei beschränktes Gebiet, $\Gamma := \partial\Omega$ sei Lipschitz-stetig
- $f : \Omega \longrightarrow \mathbb{R}^n$
- $g : \Omega \longrightarrow \mathbb{R}^n$ mit $\int\limits_\Gamma g \cdot n\, ds = 0$, wobei $n = n(x)$ den äußeren Normaleneinheitsvektor im Punkt $x \in \Gamma$ bezeichnet
- $a \in H^1(\Omega)^n \cap L^\infty(\Omega)^n$ mit $\nabla \cdot a = 0$ fast überall
- $\theta \geq 0$

Gesucht sind Funktionen

$$\begin{aligned} u : \Omega &\longrightarrow \mathbb{R}^n, \\ p : \Omega &\longrightarrow \mathbb{R} \end{aligned}$$

mit

$$\begin{aligned} -\nu \triangle u + (a \cdot \nabla) u + \theta u + \nabla p &= f \quad \text{in } \Omega && (3.1) \\ div\ u &= 0 \quad \text{in } \Omega && (3.2) \\ u\,|_\Gamma &= g. && (3.3) \end{aligned}$$

Die Gleichung (3.1) ist dabei komponentenweise zu verstehen. Der Operator *div* ist definiert durch

$$div\ u := \sum_{i=1}^{n} \frac{\partial u_i}{\partial x_i}.$$

Bemerkung 3.1
Man erkennt, daß der Druck p durch Gleichung (3.1) nur bis auf eine Konstante bestimmt ist. Daher normiert man p durch die Forderung

$$\int_\Omega p(x)\, dx = 0.$$

3.1 Schwache Formulierung der Oseen-Gleichungen

Multiplikation der i-ten Gleichung in (3.1) mit einer beliebigen Testfunktion $v_i \in C_0^\infty(\Omega)$ und anschließende Integration über Ω liefern

$$\begin{aligned} \int_\Omega \left(-\nu \triangle u_i + (a \cdot \nabla) u_i + \theta u_i + \frac{\partial p}{\partial x_i} \right) v_i \, dx &= \int_\Omega f_i v_i \, dx \\ \iff \int_\Omega \left(-\nu \sum_{j=1}^{n} \frac{\partial^2 u_i}{\partial x_j^2} + (a \cdot \nabla) u_i + \theta u_i + \frac{\partial p}{\partial x_i} \right) v_i \, dx &= \int_\Omega f_i v_i \, dx. \end{aligned}$$

Mit partieller Integration und unter Berücksichtigung von $v_i\,|_\Gamma = 0$ erhält man

$$\begin{aligned} \nu \int_\Omega \sum_{j=1}^{n} \frac{\partial u_i}{\partial x_j} \frac{\partial v_j}{\partial x_j} \, dx + \int_\Omega \{(a \cdot \nabla) u_i + \theta u_i\} v_i \, dx - \int_\Omega p \frac{\partial v_i}{\partial x_i} \, dx &= \int_\Omega f_i v_i \, dx \\ \iff \nu \int_\Omega \nabla u_i \nabla v_i \, dx + \int_\Omega \{(a \cdot \nabla) u_i + \theta u_i\} v_i \, dx - \int_\Omega p \frac{\partial v_i}{\partial x_i} \, dx &= \int_\Omega f_i v_i \, dx. \quad (3.4) \end{aligned}$$

Definiert man $(u,v) := \sum_{i=1}^{n}(u_i, v_i)$, $(u_i, v_i) := \int_\Omega u_i\, v_i\, dx$ und summiert über alle i in Gleichung (3.4), so ergibt sich

$$\nu(\nabla u, \nabla v) + ((a \cdot \nabla)u + \theta u, v) - (p, \nabla v) = (f, v). \tag{3.5}$$

Analog verfährt man mit Gleichung (3.2).
Multiplikation mit einer Testfunktion $q \in L^2(\Omega)$ und Integration liefern

$$\int_\Omega div\ u \cdot q\, dx = 0. \tag{3.6}$$

Definiere nun $L^2_0(\Omega) := \{p \in L^2(\Omega) : \int_\Omega p\, dx = 0\}$.
Somit ergibt sich mit (3.5) und (3.6) als schwache Formulierung des Oseen-Problems:
Gesucht ist $(u,p) \in H^1(\Omega)^n \times L^2_0(\Omega)$ mit

$$\begin{aligned} \nu(\nabla u, \nabla v) + (\theta u, v) + ((a \cdot \nabla)u, v) - (\nabla \cdot v, p) &= (f, v) \quad \forall v \in H^1_0(\Omega)^n, & (3.7)\\ (\nabla \cdot u, q) &= 0 \quad \forall q \in L^2_0(\Omega), & (3.8)\\ \gamma u &= g. & (3.9) \end{aligned}$$

Dabei ist $\gamma : H^1(\Omega) \longrightarrow L^2(\Gamma)$ die Spurabbildung (vgl. [GR94] S.83).
Nun definiert man folgende drei Bilinearformen

$$\begin{aligned} a_1 : H^1(\Omega)^n \times H^1(\Omega)^n &\longrightarrow \mathbb{R}\\ a_1(u,v) := (\nabla u, \nabla v) + \frac{1}{\nu}(\theta u, v), & \end{aligned}$$

$$\begin{aligned} a_2 : H^1(\Omega)^n \times H^1(\Omega)^n &\longrightarrow \mathbb{R}\\ a_2(u,v) := ((a \cdot \nabla)u, v), & \end{aligned}$$

$$\begin{aligned} b : H^1(\Omega)^n \times L^2_0(\Omega) &\longrightarrow \mathbb{R}\\ b(v,q) := -(\nabla \cdot v, q) & \end{aligned}$$

und die Linearform

$$\begin{aligned} f : H^1(\Omega)^n &\longrightarrow \mathbb{R}\\ \langle f, v\rangle := (f, v). & \end{aligned}$$

Bemerkung 3.2
(i) Die Bilinearform $a_1(u,v) = (\nabla u, \nabla v) + \frac{\theta}{\nu}(u,v)$ ist symmetrisch.
(ii) Aufgrund der Divergenzfreiheit des Vektorfeldes a gilt

$$a_2(u,v) = -a_2(v,u) \quad \textit{für alle } u, v \in H^1_0(\Omega)^n,$$

denn

$$\begin{aligned}
a_2(u,v) &= ((a\cdot\nabla)u, v) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\int_\Omega a_j v_i \frac{\partial u_i}{\partial x_j}\,dx \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n}\int_\Omega u_i \frac{\partial}{\partial x_j}(a_j v_i)\,dx \quad (\textit{partielle Integration}) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n}\int_\Omega u_i \left\{\frac{\partial a_j}{\partial x_j}v_i + \frac{\partial v_i}{\partial x_j}a_j\right\} dx \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n}\int_\Omega u_i \frac{\partial v_i}{\partial x_j}a_j\,dx \quad (da\ \nabla\cdot a = 0) \\
&= -\sum_{i=1}^{n}\int_\Omega (a\cdot\nabla)v_i \cdot u_i\,dx \\
&= -((a\cdot\nabla)v, u) \\
&= -a_2(v,u).
\end{aligned}$$

Man definiere nun die Bilinearform

$$\begin{aligned}
\tilde{a} : H^1(\Omega)^n \times H^1(\Omega)^n &\longrightarrow \mathbb{R} \\
\tilde{a}(u,v) &:= \nu a_1(u,v) + a_2(u,v) \\
&= \nu(\nabla u, \nabla v) + ((a\cdot\nabla)u + \theta u, v).
\end{aligned}$$

Aufgrund der in Bemerkung 3.2 (ii) dargestellten Eigenschaft der Bilinearform $a_2(\cdot,\cdot)$ erhält man dann für $u, v \in H_0^1(\Omega)^n$

$$\tilde{a}(u,v) = \nu(\nabla u, \nabla v) + \frac{1}{2}\left\{((a\cdot\nabla)u, v) - ((a\cdot\nabla)v, u)\right\} + \theta(u,v). \tag{3.10}$$

3.2 Existenz und Eindeutigkeit der Lösung der Oseen-Gleichungen

In diesem Abschnitt wird die Existenz und Eindeutigkeit der Lösung der verallgemeinerten Aufgabenstellung bewiesen. Dies geschieht durch Anwendung der Theorie aus Kapitel 2.
Da nicht von homogenen Dirichlet-Randbedingungen ausgegangen wird, sind für den Beweis noch zwei Hilfssätze erforderlich, die an dieser Stelle jedoch nicht bewiesen werden. Ein Beweis findet man bei [GR86] (Abschnitt I, Corollar 2.4 und Lemma 2.2).

Lemma 3.3 *Sei* $V = \{v \in H_0^1(\Omega)^n : \nabla \cdot v = 0\}$. *Dann ist der Operator* $\nabla\cdot$ *ein Isomorphismus von* $V^\perp \longrightarrow L_0^2(\Omega)$. *Dabei ist* $V^\perp$ *das orthogonale Komplement in* $H_0^1(\Omega)^n$ *bezüglich des Skalarprodukts* $(u,v) := \int\limits_\Omega \nabla u \cdot \nabla v \, dx$.

Lemma 3.4 *Sei* $g \in H^{1/2}(\Gamma)^n := \{w \in L^2(\Gamma) : \exists v \in H^1(\Omega) \text{ mit } \gamma v = w\}^n$. *Es gelte* $\int\limits_\Gamma g \cdot n \, ds = 0$. *Dann gibt es ein* $u \in H^1(\Omega)^n$ *mit*

$$\begin{aligned} \nabla \cdot u &= 0 \quad \text{in } \Omega, \\ \gamma u &= g. \end{aligned}$$

Mit diesen beiden Lemmata stehen alle Hilfsmittel für den Beweis des wesentlichen Satzes dieses Abschnitts zur Verfügung.

Satz 3.5 *Sei* Ω *ein beschränktes zusammenhängendes Gebiet.* $\Gamma = \partial\Omega$ *sei Lipschitzstetig. Sei* $f \in L^2(\Omega)^n$ *und* $g \in H^{1/2}(\Gamma)^n$ *mit*

$$\int\limits_\Omega g \cdot n \, ds = 0.$$

Dann existiert eindeutig $(u,p) \in H^1(\Omega)^n \times L_0^2(\Omega)^n$ *als Lösung von (3.7)-(3.9).*

Beweis:
(i) Nach Lemma 3.4 gibt es ein $u_0 \in H^1(\Omega)^n$ mit

$$\begin{aligned} \nabla \cdot u_0 &= 0 \quad \text{in } \Omega, \\ \gamma u_0 &= g. \end{aligned}$$

Definiere eine Linearform $l \in (H_0^1(\Omega)^n)'$ durch

$$\langle l, v\rangle := \langle f, v\rangle - \tilde{a}(u_0, v) \quad \text{für alle } v \in H_0^1(\Omega)^n.$$

Betrachte das Problem:

Finde $w \in H_0^1(\Omega)^n,\ p \in L_0^2(\Omega)$ mit

$$\tilde{a}(w,v) + b(v,p) = \langle l, v\rangle \quad \text{für alle } v \in H_0^1(\Omega)^n, \tag{3.11}$$

$$b(w,q) = 0 \quad \text{für alle } q \in L_0^2(\Omega). \tag{3.12}$$

Mit $X := H_0^1(\Omega)^n$, $M := L_0^2(\Omega)$ und $\chi = 0$ erhält man das Oseen-Problem als gemischte Variationsgleichung.

(ii) $\tilde{a}(\cdot,\cdot)$ ist stetig, denn für $u, v \in H_0^1(\Omega)^n$ ist

$$|\nu(\nabla u, \nabla v)| = \left| \nu \sum_{i,j=1}^n \int\limits_\Omega \frac{\partial u_i}{\partial x_j} \cdot \frac{\partial v_i}{\partial x_j} \, dx \right|$$

$$
\begin{aligned}
&\leq \nu \sum_{i,j=1}^{n} \int_\Omega \left| \frac{\partial u_i}{\partial x_j} \cdot \frac{\partial v_i}{\partial x_j} \right| dx \\
&\leq \nu \sum_{i,j=1}^{n} \left\| \frac{\partial u_i}{\partial x_j} \right\|_0 \cdot \left\| \frac{\partial v_i}{\partial x_j} \right\|_0 \\
&= \nu \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \left\| \frac{\partial u_i}{\partial x_j} \right\|_0 \cdot \left\| \frac{\partial v_i}{\partial x_j} \right\|_0 \right) \\
&\leq \nu \sum_{i=1}^{n} \left(\left(\sum_{j=1}^{n} \left\| \frac{\partial u_i}{\partial x_j} \right\|_0^2 \right)^{1/2} \left(\sum_{i=1}^{n} \left\| \frac{\partial v_i}{\partial x_j} \right\|_0^2 \right)^{1/2} \right) \\
&\leq \nu \left(\sum_{i=1}^{n} \left(\sum_{j=1}^{n} \left\| \frac{\partial u_i}{\partial x_j} \right\|_0^2 \right) \right)^{1/2} \left(\sum_{i=1}^{n} \left(\sum_{i=1}^{n} \left\| \frac{\partial v_i}{\partial x_j} \right\|_0^2 \right) \right)^{1/2} \\
&= \nu \left(\sum_{i=1}^{n} |u_i|_1^2 \right)^{1/2} \left(\sum_{i=1}^{n} |v_i|_1^2 \right)^{1/2} \\
&= \nu |u|_1 |v|_1,
\end{aligned}
$$

$$
\begin{aligned}
|((a \cdot \nabla)u, v)| &= \left| \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_j \frac{\partial u_i}{\partial x_j}, v_i \right) \right| \\
&\leq \sum_{i=1}^{n} \sum_{j=1}^{n} \left| \int_\Omega a_j \frac{\partial u_i}{\partial x_j} v_i \, dx \right| \\
&\leq \sum_{i=1}^{n} \sum_{j=1}^{n} \|a_j\|_\infty \left\| \frac{\partial u_i}{\partial x_j} \right\|_0 \|v_i\|_0 \\
&\leq \|a\|_\infty \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\left\| \frac{\partial u_i}{\partial x_j} \right\|_0 \|v_i\|_0 \right) \\
&\leq \|a\|_\infty \sum_{j=1}^{n} \left(\left(\sum_{i=1}^{n} \|v_i\|_0^2 \right)^{1/2} \left(\sum_{i=1}^{n} \left\| \frac{\partial u_i}{\partial x_j} \right\|_0^2 \right)^{1/2} \right) \\
&= \|a\|_\infty \|v\|_0 \sum_{j=1}^{n} 1 \cdot \left(\sum_{i=1}^{n} \left\| \frac{\partial u_i}{\partial x_j} \right\|_0^2 \right)^{1/2} \\
&\leq C \|a\|_\infty |v|_1 \sqrt{n} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} \left\| \frac{\partial u_i}{\partial x_j} \right\|_0^2 \right)^{1/2} \quad \text{(Friedrichsche Ungleichung)} \\
&= C_1 \|a\|_\infty |v|_1 |u|_1,
\end{aligned}
$$

Wiederum unter Verwendung der Friedrichschen Ungleichung erhält man

$$\theta(u,v) \le \theta\|u\|_0\|v\|_0 \le \theta C_2|u|_1|v|_1.$$

Damit ergibt sich

$$|\tilde{a}(u,v)| \le (\nu + C_1\|a\|_\infty + \theta C_2)|u|_1|v|_1.$$

Die Bilinearform $b(\cdot,\cdot)$ ist ebenfalls stetig, denn für $v \in H_0^1(\Omega)^n$, $q \in L_0^2(\Omega)$ gilt

$$\begin{aligned}
|(\nabla\cdot v, q)| &= \left|\int_\Omega (\nabla\cdot v) q\,dx\right| \\
&= \left|\int_\Omega \sum_{i=1}^n \frac{\partial v_i}{\partial x_i}\cdot q\,dx\right| \\
&\le \left(\sum_{i=1}^n 1\cdot\left\|\frac{\partial v_i}{\partial x_i}\right\|_0\right)\|q\|_0 \\
&\le \sqrt{n}\left(\sum_{i=1}^n\left\|\frac{\partial u_i}{\partial x_i}\right\|_0^2\right)^{1/2}\|q\|_0 \\
&\le \sqrt{n}\left(\sum_{j=1}^n\sum_{i=1}^n\left\|\frac{\partial v_j}{\partial x_i}\right\|_0^2\right)^{1/2}\|q\|_0 \\
&\le C_3|v|_1\|q\|_0.
\end{aligned}$$

Auf ähnliche Weise läßt sich nachrechnen, daß $\langle l,\cdot\rangle$ stetig ist.

(iii) $\tilde{a}(\cdot,\cdot)$ ist X-elliptisch, denn für $v \in H_0^1(\Omega)^n$ ist mit (3.10)

$$\begin{aligned}
\tilde{a}(v,v) &= \nu\int_\Omega \nabla v\cdot\nabla v\,dx + \underbrace{\theta\|v\|_0^2}_{\ge 0} \\
&\ge \nu\sum_{i=1}^n\int_\Omega \nabla v_i\cdot\nabla v_i\,dx \\
&= \nu\sum_{i,j=1}^n\int_\Omega\left(\frac{\partial v_i}{\partial x_j}\right)^2 dx \\
&= \nu\sum_{i,j=1}^n\left\|\frac{\partial v_i}{\partial x_j}\right\|_0^2 \\
&= \nu|v|_1^2.
\end{aligned}$$

(iv) Nun ist die inf-sup-Bedingung zu zeigen, d.h.

$$\sup_{v\in H_0^1(\Omega)^n\setminus\{0\}}\frac{(q,\nabla\cdot v)}{|v|_1} \ge \beta\|q\|_0 \quad \text{für alle } q \in L_0^2(\Omega).$$

Sei nun $q \in L_0^2(\Omega)$ beliebig. Nach Lemma 3.3 gibt es ein eindeutig bestimmtes $v \in V^\perp$ mit $\nabla \cdot v = q$. Da $\nabla\cdot : V^\perp \longrightarrow L_0^2(\Omega)$ ein Isomorphismus, also auch stetig ist, gibt es eine Konstante $\hat{C}$ mit

$$|v|_1 \leq \hat{C}\|q\|_0.$$

Damit erhält man dann

$$\begin{aligned}\frac{(q, \nabla \cdot v)}{|v|_1} &= \frac{(q,q)}{|v|_1} = \frac{\|q\|_0}{|v|_1} \underbrace{\geq \frac{1}{\hat{C}}}_{=:\beta} \|q\|_0 \\ &= \beta\|q\|_0.\end{aligned}$$

Da β unabhängig von q ist, gilt also für alle $q \in L_0^2(\Omega)$, daß ein $v \in H_0^1(\Omega)^n$ mit der Eigenschaft

$$\frac{(q, \nabla \cdot v)}{|v|_1} \geq \beta\|q\|_0$$

existiert. Daraus ergibt sich die inf-sup-Bedingung.

(v) Mit Satz 2.8 folgt aus dem in (i)-(iv) gezeigten, daß es ein eindeutiges $(w, p) \in H_0^1(\Omega)^n \times L_0^2(\Omega)$ gib, welches die Gleichungen (3.11) und (3.12) erfüllt. Dann ist $(u = u_0 + w, p) \in [u_0 + H_0^1(\Omega)^n] \times L_0^2(\Omega)$ Lösung des Problems (3.7)-(3.9).
Ist umgekehrt (u, p) Lösung von (3.7)-(3.9), so löst $w := u - u_0$ die Gleichungen (3.11) und (3.12).
Insgesamt sind damit die Oseen-Gleichungen eindeutig lösbar.

□

3.3 Das diskrete Oseen-Problem

In diesem Abschnitt werden die Oseen-Gleichungen mit homogenen Randbedingungen betrachtet. Das Problem ist also folgendermaßen formuliert:

Gesucht (u, p) mit

$$\begin{aligned}-\nu\triangle u + (a \cdot \nabla)u + \theta u + \nabla p &= f \text{ in } \Omega, \\ \nabla \cdot u &= 0 \text{ in } \Omega, \\ u|_\Gamma &= 0.\end{aligned}$$

Daraus ergibt sich, wie in Abschnitt 3.1 dargestellt, folgende Variationsfomulierung:

Gesucht ist $(u, p) \in X \times M := H_0^1(\Omega)^n \times L_0^2(\Omega)$ mit

$$\begin{aligned}\nu a_1(u, v) + a_2(u, v) + b(v, p) &= \langle f, v\rangle \quad \text{für alle } v \in H_0^1(\Omega)^n, \\ b(u, q) &= 0 \qquad\quad \text{für alle } q \in L_0^2(\Omega).\end{aligned}$$

Die Bilinearformen $a_1(\cdot,\cdot)$, $a_2(\cdot,\cdot)$ und $b(\cdot,\cdot)$, sowie die Linearform $\langle f, \cdot\rangle$ seien definiert wie in Abschnitt 3.1. Durch Einführung von diskreten Räumen $X_h \subset X$ und $M_h \subset M$

(vgl. Abschnitt 2.2.1) kann das obige Problem approximiert werden durch:

Gesucht $(u_h, p_h) \in X_h \times M_h$ mit

$$\begin{aligned} \nu a_1(u_h, v_h) + a_2(u_h, v_h) + b(v_h, p_h) &= \langle f, v_h \rangle \quad \text{für alle } v_h \in X_h, \qquad (3.13) \\ b(u_h, q_h) &= 0 \quad \text{für alle } q_h \in M_h. \qquad (3.14) \end{aligned}$$

Sei nun $\{\phi_1, \ldots, \phi_l\}$ eine Basis von X_h und $\{\psi_1, \ldots, \psi_k\}$ eine Basis von M_h. Wie in Abschnitt 2.2.1 gezeigt, ist das diskrete Oseen-Problem (3.13), (3.14) äquivalent zu dem linearen Gleichungssystem

$$\begin{pmatrix} \nu A + N & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix},$$

wobei

$$U := (U_1, \ldots, U_l) \text{ mit } u_h = \sum_{i=1}^{l} U_i \phi_i,$$

$$P := (P_1, \ldots, P_k) \text{ mit } p_h = \sum_{j=1}^{k} P_j \psi_j$$

und

$$\begin{aligned} A \in \mathbb{R}^{l \times l} &: A_{ij} := a_1(\phi_j, \phi_i), \\ N \in \mathbb{R}^{l \times l} &: N_{ij} := a_2(\phi_j, \phi_i), \\ B \in \mathbb{R}^{k \times l} &: B_{ij} := b(\phi_j, \psi_i), \\ F \in \mathbb{R}^{l} &: F_i := \langle f, \phi_i \rangle. \end{aligned}$$

Aufgrund der Bemerkung 3.2 (i) ist die Matrix A symmetrisch. Aus Teil (ii) dieser Bemerkung folgt die Schiefsymmetrie von N, d.h. $N^T = -N$.

Da die Koeffizientenmatrix im allgemeinen schlecht konditioniert ist, empfiehlt sich der Einsatz von Vorkonditionierern. Mit dem Problem der Auswahl geeigneter Vorkonditionierer beschäftigt sich Kapitel 4. Dabei spielt das folgende Lemma eine wichtige Rolle.

Lemma 3.6 *Die Diskretisierung sei stabil, d.h. es existiert ein $\beta > 0$, das unabhängig von der Diskretisierungsschrittweite h ist, so daß gilt*

$$\sup_{u_h \in X_h \setminus \{0\}} \frac{b(u_h, p_h)}{\|u_h\|_X} \geq \beta \|p_h\|_P \quad \textit{für alle } p_h \in M_h. \qquad (3.15)$$

Dann gibt es Konstanten C_F und κ, die unabhängig von h sind, so daß für die Matrizen des Oseen-Problems folgende Abschätzung gilt

$$\frac{\beta^2}{1 + \frac{\theta}{\nu} C_F^2} \leq \frac{(P, BA^{-1}B^T P)}{(P, QP)} \leq \kappa^2 \quad \textit{für alle } P \in \mathbb{R}^k,$$

hierbei ist $Q_{ij} := (\psi_i, \psi_j)$ für $i, j = 1, \ldots, k$ die Massematrix des Druckes.

Beweis:
Die Normen auf X bzw. M seien definiert durch

$$\|u_h\|_X := |u_h|_1,$$
$$\|p_h\|_M := \|p_h\|_0.$$

Definiere nun

$$|||u_h|||_X^2 := |u_h|_1^2 + \frac{\theta}{\nu}\|u_h\|_0.$$

Für beliebiges $u_h \in X$ gilt unter Verwendung der Friedrichschen Ungleichung (vgl. [GR94])

$$\begin{aligned} |||u_h|||_X^2 &= |u_h|_1^2 + \frac{\theta}{\nu}\|u_h\|_0^2 \\ &\leq |u_h|_1^2 + \frac{\theta}{\nu}C_F^2|u_h|_1^2 \\ &= \left(1 + \frac{\theta}{\nu}C_F^2\right)|u_h|_1^2. \end{aligned} \tag{3.16}$$

Weiterhin gilt für beliebiges $u_h \in X$ aber auch

$$|u_h|_1^2 = |||u_h|||_X^2 - \underbrace{\frac{\theta}{\nu}\|u_h\|_0^2}_{\geq 0} \tag{3.17}$$

$$\leq |||u_h|||_X^2. \tag{3.18}$$

Daraus folgt dann, daß $|||\cdot|||_X$ und $|\cdot|_1$ äquivalente Normen auf X sind.

Unter Verwendung von (3.16) erhält man mit (3.15)

$$\begin{aligned} \beta\|p_h\|_P &\leq \sup_{u_h \in X_h \setminus \{0\}} \frac{b(u_h, p_h)}{|u_h|_1} \\ &\leq \left(1 + \frac{\theta}{\nu}C_F^2\right)^{1/2} \sup_{u_h \in X_h \setminus \{0\}} \frac{b(u_h, p_h)}{|||u_h|||_X}. \end{aligned}$$

Dies ist äquivalent zu

$$\frac{\beta}{\left(1 + \frac{\theta}{\nu}C_F^2\right)^{1/2}}\|p_h\|_P \leq \sup_{u_h \in X_h \setminus \{0\}} \frac{b(u_h, p_h)}{|||u_h|||_X}. \tag{3.19}$$

Weiterhin gilt

$$\begin{aligned} |||u_h|||_X^2 &= (\nabla u_h, \nabla u_h) + \frac{\theta}{\nu}(u_h, u_h) \\ &= a_1(u_h, u_h) = (AU, U), \end{aligned} \tag{3.20}$$

sowie

$$b(u_h, p_h) = (BU, P). \tag{3.21}$$

Außerdem ist

$$\begin{aligned} \|p_h\|_M = (p_h, p_h) &= (\sum_{i=1}^{n} P_i \psi_i, \sum_{j=1}^{n} P_j \psi_j) \\ &= \sum_{i=1}^{n} P_i \sum_{j=1}^{n} P_j \underbrace{(\psi_i, \psi_j)}_{=Q_{ij}} \\ &= (P, QP). \end{aligned} \tag{3.22}$$

Unter Verwendung von (3.20), (3.21) und (3.22) folgt aus (3.19)

$$\frac{\beta}{\left(1 + \frac{\theta}{\nu} C_F^2\right)^{1/2}} (P^T Q P)^{1/2} \leq \sup_{U \in \mathbb{R}^l \setminus \{0\}} \frac{P^T B U}{(U^T A U)^{1/2}}.$$

Daraus ergibt sich dann

$$\begin{aligned} (P, BA^{-1}B^T P) &= (AA^{-1}B^T P, A^{-1}B^T P) \\ &= \sup_{U \in \mathbb{R}^l \setminus \{0\}} \frac{(AA^{-1}B^T P, U)^2}{(AU, U)} \\ &= \sup_{U \in \mathbb{R}^l \setminus \{0\}} \frac{(P, BU)(P, BU)}{(AU, U)^{1/2}(AU, U)^{1/2}} \\ &\geq \left(\frac{\beta}{\left(1 + \frac{\theta}{\nu} C_F^2\right)^{1/2}} \right)^2 (P, QP). \end{aligned}$$

Folglich gilt

$$\frac{(P, BA^{-1}B^T P)}{(P, QP)} \geq \frac{\beta^2}{1 + \frac{\theta}{\nu} C_F^2}.$$

Die Bilinearform $b(u, p) := -(\nabla \cdot u, p)$ ist stetig (vgl. Beweis von Satz 3.5), also gilt für beliebiges $u_h \in X_h$, $p_h \in P_h$

$$\begin{aligned} (P, BU) = b(u_h, p_h) &\leq \kappa |u_h|_1 \|p_h\|_0 \\ &\leq \kappa |||u_h|||_X \|p_h\|_0 \\ &= \kappa (U^T A U)^{1/2} (P^T Q P)^{1/2}. \end{aligned}$$

Daraus folgt dann

$$\begin{aligned} (P, BA^{-1}B^T P) &\leq \kappa \left((A^{-1}B^T P)^T A A^{-1} B^T P) \right)^{1/2} (P^T Q P)^{1/2} \\ &= (P^T B A^{-1} B^T P)^{1/2} (P^T Q P)^{1/2}. \end{aligned}$$

Damit erhält man dann

$$\frac{(P, BA^{-1}B^TP)}{(P, QP)} \leq \kappa^2.$$

□

Bemerkung 3.7
Für den Fall $\theta = 0$ hat die Abschätzung aus Lemma 3.6 folgende Form

$$\beta^2 \leq \frac{(P, B\hat{A}^{-1}B^TP)}{(P, QP)} \leq \kappa^2 \quad \textit{für alle } P \in \mathbb{R}^k,$$

wobei $\hat{A}_{ij} := (\nabla\phi_j, \nabla\phi_j)$ für alle $i, j = 1, \ldots, l$ der diskrete Laplace-Operator ist.

Die Wahl der Räume X_h und M_h erfolgt bei konkreten Problemen, wie den Oseen-Gleichungen, mit sogenannten finiten Elementen. Das Gebiet $\Omega \subset \mathbb{R}^n$ wird in geometrisch einfachere Teilgebiete zerlegt. Zum Beispiel sind dies im zweidimensionalen Fall Dreiecke und Vierecke, im dreidimensionalen Raum können Tetraeder und Quader verwendet werden. Das Gebiet Ω wird also in endlich viele disjunkte Teilgebiete Ω_i, $i = 1, \ldots, m$, zerlegt, d.h.

$$\bar{\Omega} = \bigcup_{i=1}^{m} \bar{\Omega}_i \quad \text{und} \quad int\,\Omega_i \cap int\,\Omega_j = \emptyset, \quad \text{falls } i \neq j.$$

$\mathcal{Z}_h := \{\Omega_i\}_{i=1}^m$ bezeichnet man als Zerlegung des Gebietes Ω. Nun ist es möglich, auf jedem dieser Teilgebiete Ansatzfunktionen zu definieren. Zum Beispiel kann man stetige stückweise polynomiale Funktionen auf jedem Element verwenden. Als diskrete Räume erhält man dann

$$\begin{aligned} X_h &:= \{v_h \in C(\bar{\Omega}) : v_h|_{\Omega_i} \in \mathcal{P}_k \quad \text{für alle } \Omega_i \in \mathcal{Z}_h\}, \\ M_h &:= \{q_h \in C(\bar{\Omega}) : q_h|_{\Omega_i} \in \mathcal{P}_l \quad \text{für alle } \Omega_i \in \mathcal{Z}_h\}. \end{aligned}$$

Bemerkung 3.8
Finite Elemente, welche die Babuška-Brezzi-Bedingung

$$\sup_{u_h \in X_h \setminus \{0\}} \frac{b(u_h, p_h)}{\|u_h\|_X} \geq \beta \|p_h\|_P \quad \textit{für alle } p_h \in M_h$$

mit einer von der Diskretisierungsschrittweite h unabhängigen Konstante β erfüllen, heißen BB-stabil.

3.4 Stabilisierte Verfahren

Viele für praktische Implementierungen interessante finite Elemente erfüllen die Babuška-Brezzi-Bedingung mit einer von h unabhängigen Konstante nicht. Ein Ausweg ist die

Stabilisierung durch geeignete Abänderung des diskreten Problems, zum Beispiel durch die Verwendung der Streamline-Diffusion-FEM. Diese Methode besteht nun darin, einen least-squares-Term für die Divergenzgleichung einzuführen und die Geschwindigkeitsgleichung elementweise mit Funktionen der Form $(a\cdot\nabla)v+\nabla q$ zu testen. Dann erhält man als diskretes Problem:

Finde $\hat{u}_h=(u_h,p_h)\in X_h\times M_h$ mit

$$B_{SD}(\hat{u}_h,\hat{v}_h)=L_{SD}(\hat{v}_h)\quad\text{für alle }\hat{v}_h=(v_h,q_h)\in X_h\times P_h,$$

wobei

$$\begin{aligned}
B_{SD}(\hat{u}_h,\hat{v}_h) &:= \nu(\nabla u_h,\nabla v_h)+((a\cdot\nabla)u_h+\theta u_h,v_h)-(p_h,\nabla\cdot v_h)+(q_h,\nabla\cdot u_h)\\
&+ \sum_K \alpha_K(\nabla\cdot u_h,\nabla\cdot v_h)_K\\
&+ \sum_K \delta_K(-\nu\triangle u_h+(a\cdot\nabla)u_h+\theta u_h+\nabla p_h,(a\cdot\nabla)v_h+\nabla q_h)_K,\\
L_{SD}(\hat{v}_h) &:= (f,v_h)+\sum_K \delta_K(f,(a\cdot\nabla)v_h+\nabla q_h)_K.
\end{aligned}$$

Die zur Diskretisierung möglichen Ansatzräume und Aussagen zur Analysis der Streamline-Diffusion-FEM findet man zum Beispiel in [Mü97].

Eine weitere Stabilisierungsmethode ist das Galerkin/least-squares-Verfahren (GLS). Bei dieser Methode wird mit dem Differentialoperator $L\hat{u}:=\nu\triangle u+(a\cdot\nabla)u+\theta u+\nabla p$ elementweise getestet. Man erhält folgendes Problem:

Finde $\hat{u}_h=(u_h,p_h)\in X_h\times M_h$ mit

$$B_{GLS}(\hat{u}_h,\hat{v}_h)=L_{GLS}(\hat{v}_h)\quad\text{für alle }\hat{v}_h=(v_h,q_h)\in X_h\times P_h,$$

wobei

$$\begin{aligned}
B_{GLS}(\hat{u}_h,\hat{v}_h) &:= \nu(\nabla u_h,\nabla v_h)+((a\cdot\nabla)u_h+\theta u_h,v_h)-(p_h,\nabla\cdot v_h)+(q_h,\nabla\cdot u_h)\\
&+ \sum_K \alpha_K(\nabla\cdot u_h,\nabla\cdot v_h)_K+\sum_K \delta_K(L\hat{u}_h,L\hat{v}_h)_K,\\
L_{GLS}(\hat{v}_h) &:= (f,v_h)+\sum_K \delta_K(f,L\hat{v}_h)_K.
\end{aligned}$$

Dieses Verfahren wird für den Fall $\theta=0$ in [Lub94] genauer analysiert. Das Programm PNS, welches für die numerischen Experimente in Kapitel 5 genutzt wird, arbeitet mit linearen Elementen in Druck und Geschwindigkeit und verwendet zur Stabilisierung das GLS-Verfahren.

Bei linearen Elementen stimmen das Galerkin/least-squares-Verfahren und die Streamline-Diffusion-FEM überein.

Kapitel 4

Vorkonditionierung bei stabilen finiten Elementen

Das diskrete Oseen-Problem ist wie in Abschnitt 3.3 gezeigt äquivalent zu einem linearen Gleichungssystem. Dieses System kann nun mit verschiedenen Iterationsverfahren gelöst werden. Von einigen iterativen Verfahren kann gezeigt werden, daß die Konvergenzeigenschaften von der Kondition der Matrix abhängen (vgl. [Saa96]). Bei Verfahren, für die solche Aussagen nicht bewiesen sind, wird aufgrund von experimentellen Untersuchungen angenommen, daß auch in diesen Fällen ein Zusammenhang zwischen Konvergenz der Verfahren und Kondition der Matrix besteht. Die Konditionszahl ist nun vom betragsmäßig größten und kleinsten Eigenwert abhängig. Um die Kondition der Matrix zu verbessern, ist der Einsatz von Vorkonditionierern empfehlenswert. Für die bei der Diskretisierung von partiellen Differentialgleichungen entstehenden Gleichungssysteme ist es außerdem noch wünschenswert, daß die Eigenwerte der vorkonditionierten Systeme unabhängig von der Diskretisierungsschrittweite h beschränkt sind, also folglich ist die Kondition unabhängig von h.
In diesem Kapitel wird nun auf zwei Vorkonditionierungsstrategien für das Oseen-Problem bei BB-stabilen Elementen (vgl. Bemerkung 3.8) eingegangen, nämlich einen Blockdreiecks- und einen Blockdiagonalvorkonditionierer. In der Literatur sind sehr wenig Ausarbeitungen zum Thema der Vorkonditionierung der Oseen-Gleichung für BB-stabile Elemente zu finden. In [ES96] wird dieses Thema behandelt. Diese Ausarbeitung ist auch Grundlage des folgenden Kapitel. Abweichend von dem genannten Artikel wird jedoch noch der Term θu, der durch die Zeitdiskretisierung zeitabhängiger Probleme entsteht, berücksichtigt. Jedoch ist es mir nicht möglich zu zeigen, daß für den Fall $\theta > 0$ die Eigenwerte des vorkonditionierten Systems unabhängig von der Diskretisierungsschrittweite h beschränkt sind. Auf die kritischen Stellen wird entsprechend hingewiesen. Für den Fall $\theta = 0$ wird in Abschnitt 4 nachgewiesen, daß die Eigenwerte des vorkonditionierten Systems unabhängig von h beschränkt sind, dies ist auch das Hauptziel dieses Kapitels.

4.1 Die Vorkonditionierungsstrategien

In diesem Abschnitt werden nun die beiden Vorkonditionierungsstrategien vorgestellt. Außerdem wird gezeigt, daß die Eigenwerte des vorkonditionierten Systems für den Fall $\theta = 0$ unabhängig von der Diskretisierungsschrittweite h beschränkt sind.

Das zum diskreten Oseen-Problem äquivalente Gleichungssystem ist

$$\begin{pmatrix} \nu A + N & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix},$$

vgl. Abschnitt 3.3.
Für dieses Gleichungssystem kann man spezielle Vorkonditionierer definieren, die der Struktur des Problems angepaßt sind. Der Blockdiagonalvorkonditionierer ist definiert durch

$$M_D := \begin{pmatrix} \nu A + N & 0 \\ 0 & \frac{1}{\nu} Q \end{pmatrix},$$

wobei Q die Massematrix des Drucks ist.

Eine weitere Möglichkeit ist der Blockdreiecksvorkonditionierer, definiert durch

$$M_T := \begin{pmatrix} \nu A + N & B^T \\ 0 & \frac{1}{\nu} Q \end{pmatrix}.$$

Sei nun

$$D := \begin{pmatrix} \nu A + N & B^T \\ B & 0 \end{pmatrix}$$

die Koeffizientenmatrix des diskreten Oseen-Problems. M^{-1} sei die Vorkonditionierungsmatrix. Ziel ist es nun, Aussagen über die Kondition von DM^{-1} zu machen. Dazu ist die Abschätzung der Eigenwerte des vorkonditionierten Systems notwendig. Also wird das verallgemeinertes Eigenwertproblem

$$\begin{aligned} DM^{-1}w &= \lambda w \\ \Leftrightarrow Dv &= \lambda M v \end{aligned} \tag{4.1}$$

mit $v := M^{-1}w$ betrachtet.

Zuerst wird auf den Blockdiagonalvorkonditionierer

$$M_D := \begin{pmatrix} \nu A + N & 0 \\ 0 & \frac{1}{\nu} Q \end{pmatrix}$$

eingegangen.
Man erhält für das verallgemeinerte Eigenwertproblem (4.1)

$$\begin{pmatrix} \nu A + N & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \lambda \begin{pmatrix} \nu A + N & 0 \\ 0 & \frac{1}{\nu} Q \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix}. \tag{4.2}$$

Äquivalent dazu ist

$$\begin{aligned} Fu + B^T p &= \lambda F u, && (4.3)\\ Bu &= \lambda \frac{1}{\nu} Q p, && (4.4) \end{aligned}$$

mit $F := \nu A + N$.
Eine Lösung des Eigenwertproblems (4.2) ist $\lambda = 1$ mit dem zugehörigen Eigenvektor $(u, 0)^T$, für den gilt $Bu = 0$.
Die Matrix $B \in \mathbb{R}^{k \times l}$ hat den Rang k, da die Zeilen linear unabhängig sind. Der Lösungsraum des homogenen Gleichungssystems $Bu = 0$ hat dann die Dimension l-k. Also gibt es l-k linear unabhängige u, die das System $Bu = 0$ lösen. Daraus folgt, daß der Eigenwert $\lambda = 1$ die Vielfachheit l-k hat. Um die weiteren Eigenwerte zu ermitteln ist es erforderlich, die Gleichungen (4.3) und (4.4) in eine Gleichung für p umzuformen. Sei nun $\lambda \neq 1$. Dann erhält man durch Umformung von (4.3) folgende Gleichung für u

$$\begin{aligned} B^T p &= (\lambda - 1) F u \\ \Leftrightarrow F^{-1} B^T p &= (\lambda - 1) u \\ \Leftrightarrow \frac{1}{\lambda - 1} F^{-1} B^T p &= u. && (4.5) \end{aligned}$$

Einsetzen von (4.5) in (4.4) liefert

$$\begin{aligned} \frac{1}{\lambda - 1} B F^{-1} B^T p &= \lambda \frac{1}{\nu} Q p \\ \Leftrightarrow B F^{-1} B^T p &= \lambda(\lambda - 1) \frac{1}{\nu} Q p. \end{aligned}$$

Diese Gleichung wird als Schur-Komplement-Gleichung bezeichnet. Definiere nun $S := BF^{-1}B^T$ und $\mu := \lambda(\lambda - 1)$. Die weiteren Eigenwerte des Problems (4.2) erhält man also aus der quadratischen Gleichung

$$\lambda(\lambda - 1) = \mu,$$

wobei μ ein Eigenwert von

$$S p = \mu \frac{1}{\nu} Q p \qquad (4.6)$$

ist.
Äquivalent zu $\lambda(\lambda - 1) = \mu$ ist

$$\lambda = \frac{1 \pm \sqrt{1 + 4\mu}}{2}.$$

Im folgenden soll gezeigt werden, daß die Eigenwerte des Problems (4.6) unabhängig von der Diskretisierungsschrittweite h beschränkt sind. Dazu sind jedoch einige Lemmata erforderlich, die nun formuliert und teilweise auch bewiesen werden.

Lemma 4.1 *Sei* $S := B(\nu A + N)^{-1}B^T = BF^{-1}B^T$.
Für jeden Eigenwerte $\mu \in \mathbb{C}$ *des Schur-Komplement Problems*

$$Sp = \mu \frac{1}{\nu} Qp$$

gelten folgende Abschätzungen:

(i)

$$\min_{p \neq 0} \frac{(p, Cp)}{(p, \frac{1}{\nu}Qp)} \leq Re\mu \leq \max_{p \neq 0} \frac{(p, Cp)}{(p, \frac{1}{\nu}Qp)},$$

mit $C := \frac{1}{2}(S + S^T)$,

(ii)

$$|Im\mu| \leq \max_{p \neq 0} \frac{|(p, Rp)|}{(p, \frac{1}{\nu}Qp)},$$

mit $R := \frac{1}{2}(S - S^T)$.

Beweis:
(1) S ist eine quadratische Matrix. Diese läßt sich zerlegen in

$$S = C + R,$$

wobei

$$C = \frac{1}{2}(S + S^T)$$

der symmetrische Teil und

$$R = \frac{1}{2}(S - S^T)$$

der schiefsymmetrische Teil ist, d.h. es gilt $R^T = -R$.

Sei nun μ eine Lösung der verallgemeinerten Eigenwertproblems

$$Sp = \mu \frac{1}{\nu} Qp.$$

Daraus folgt dann

$$p^H Sp \;\; = \;\; \mu p^H \frac{1}{\nu} Qp.$$

Da Q positiv definit ist, also $p^H \frac{1}{\nu} Qp \neq 0$, gilt für μ dann

$$\mu = \frac{p^H Sp}{p^H \frac{1}{\nu} Qp}. \tag{4.7}$$

(2) Unter Verwendung von (4.7) und der Zerlegung von S erhält man für den Realteil von μ folgende Abschätzung

$$\begin{aligned}
Re\mu &\leq \max_{p\neq 0} Re\frac{p^H Sp}{p^H \frac{1}{\nu}Qp} \\
&= \max_{p\neq 0} \frac{1}{p^H \frac{1}{\nu}Qp} Re(p^H Sp) \quad (\text{da } p^H \tfrac{1}{\nu}Qp \in \mathbb{R}) \\
&= \max_{p\neq 0} \frac{1}{p^H \frac{1}{\nu}Qp}\frac{1}{2}\left(p^H Sp + p^H S^T p\right) \\
&= \max_{p\neq 0} \frac{1}{p^H \frac{1}{\nu}Qp}\frac{1}{2}\left(p^H (C+R)p + p^H (C+R)^T p\right) \\
&= \max_{p\neq 0} \frac{1}{p^H \frac{1}{\nu}Qp}\frac{1}{2}\left(p^H Cp + p^H Rp + p^H C^T p + p^H R^T p\right) \\
&= \max_{p\neq 0} \frac{1}{p^H \frac{1}{\nu}Qp} p^H Cp \quad (C^T = C \text{ und } R^T = -R) \\
&= \max_{p\neq 0} \frac{(p,Cp)}{(p,\frac{1}{\nu}Qp)}.
\end{aligned}$$

Analog zu dieser Abschätzung bekommt man auch

$$Re\mu \geq \min_{p\neq 0} \frac{(p,Cp)}{(p,\frac{1}{\nu}Qp)}.$$

Insgesamt gilt also

$$\min_{p\neq 0} \frac{(p,Cp)}{(p,\frac{1}{\nu}Qp)} \leq Re\mu \leq \max_{p\neq 0} \frac{(p,Cp)}{(p,\frac{1}{\nu}Qp)}.$$

(3) Für den Betrag des Imaginärteil von μ gilt

$$\begin{aligned}
|Im\mu| &\leq \left|\max_{p\neq 0} Im\frac{p^H Sp}{p^H \frac{1}{\nu}Qp}\right| \\
&\leq \max_{p\neq 0}\left|\frac{1}{p^H \frac{1}{\nu}Qp}\frac{1}{2i}\left(p^H Sp - p^H S^T p\right)\right| \\
&= \max_{p\neq 0}\left|\frac{1}{p^H \frac{1}{\nu}Qp}\right|\left|\frac{1}{i}p^H Rp\right| \\
&= \max_{p\neq 0}\frac{1}{p^H \frac{1}{\nu}Qp}\left|\frac{1}{i}\right||p^H Rp| \\
&= \max_{p\neq 0}\frac{|(p,Rp)|}{(p,\frac{1}{\nu}Qp)}.
\end{aligned}$$

Damit ist Abschätzung (ii) gezeigt.

□

Lemma 4.2 *Seien* $T := BA^{-1}B^T$ *und* $C := B\left(\frac{F^{-1}+F^{-T}}{2}\right)B^T$.
Dann gilt für alle $p \in \mathbb{C}^k \backslash \{0\}$

$$\frac{(p,Cp)}{(p,\frac{1}{\nu}Tp)} = \frac{\left(v,\left(I-\frac{1}{\nu^2}\tilde{N}^2\right)^{-1}v\right)}{(v,v)}, \tag{4.8}$$

wobei $\tilde{N} := A^{-1/2}NA^{-1/2}$ *und* $v := A^{-1/2}B^Tp$.

Beweis:
Sei $F := \nu A + N$. Man erhält unter Ausnutzung der Symmetrie von A und der Schiefsymmetrie von N

$$\begin{aligned}
\frac{F^{-1}+F^{-T}}{2} &= F^{-1}\left(\frac{F+F^T}{2}\right)F^{-T} \\
&= (\nu A+N)^{-1}\left(\frac{\nu A+N+(\nu A+N)^T}{2}\right)(\nu A+N)^{-T} \\
&= (\nu A+N)^{-1}\left(\frac{2\nu A+N-N}{2}\right)(\nu A-N)^{-1} \\
&= \left((\nu A-N)\left(\frac{1}{\nu}A^{-1}\right)(\nu A+N)\right)^{-1} \\
&= \left(\nu A-\frac{1}{\nu}NA^{-1}N\right)^{-1} \\
&= \left(\nu A^{1/2}A^{1/2}-\frac{1}{\nu}A^{1/2}A^{-1/2}NA^{-1/2}A^{-1/2}NA^{-1/2}A^{1/2}\right)^{-1} \\
&= \left(A^{1/2}(\nu I-\frac{1}{\nu}A^{-1/2}NA^{-1/2}A^{-1/2}NA^{-1/2})A^{1/2}\right)^{-1} \\
&= A^{-1/2}\left(\nu I-\frac{1}{\nu}\underbrace{A^{-1/2}NA^{-1/2}}_{=\tilde{N}}\underbrace{A^{-1/2}NA^{-1/2}}_{=\tilde{N}}\right)^{-1}A^{-1/2} \\
&= A^{-1/2}\left(\nu I-\frac{1}{\nu}\tilde{N}^2\right)^{-1}A^{-1/2}.
\end{aligned}$$

Also ist

$$C := B\left(\frac{F^{-1}+F^{-T}}{2}\right)B^T = BA^{-1/2}\left(\nu I-\frac{1}{\nu}\tilde{N}^2\right)^{-1}A^{-1/2}B^T.$$

Damit bekommt man

$$\frac{(p,Cp)}{(p,\frac{1}{\nu}Tp)} = \frac{\left(p,BA^{-1/2}(\nu I-\frac{1}{\nu}\tilde{N}^2)^{-1}A^{-1/2}B^Tp\right)}{(p,\frac{1}{\nu}BA^{-1}B^Tp)}$$

$$
\begin{aligned}
&= \frac{\left(A^{-1/2}B^Tp, (\nu I - \frac{1}{\nu}\tilde{N}^2)^{-1}A^{-1/2}B^Tp\right)}{(A^{-1/2}B^Tp, \frac{1}{\nu}A^{-1/2}B^Tp)} \\
&= \frac{\left(v, (\nu I - \frac{1}{\nu}\tilde{N}^2)^{-1}v\right)}{(v, \frac{1}{\nu}v)} \\
&= \frac{\left(v, (I - \frac{1}{\nu^2}\tilde{N}^2)^{-1}v\right)}{(v, v)},
\end{aligned}
$$

wobei $v := A^{-1/2}B^Tp$ ist.

□

Lemma 4.3 *Seien* $T := BA^{-1}B^T$ *und* $R := B\left(\frac{F^{-1}-F^{-T}}{2}\right)B^T$.
Dann gilt für alle $p \in \mathbb{C}^k \backslash \{0\}$

$$\frac{(p, Rp)}{(p, \frac{1}{\nu}Tp)} = -\frac{\nu(v, \tilde{N}v)}{(v, (\nu^2 I - \tilde{N}^2)v)},$$

mit $\tilde{N} := A^{-1/2}NA^{-1/2}$ *und* $v := (\nu I + \tilde{N})^{-1}A^{-1/2}B^Tp$.

Beweis:
Mit $F := \nu A + N$ erhält man

$$
\begin{aligned}
\frac{F^{-1} - F^{-T}}{2} &= F^{-T}\left(\frac{F^T - F}{2}\right)F^{-1} \\
&= (\nu A + N)^{-T}\left(\frac{(\nu A + N)^T - (\nu A + N)}{2}\right)(\nu A + N)^{-1} \\
&= (\nu A - N)^{-1}(-N)(\nu A + N)^{-1} \quad (\text{da } A = A^T,\ N^T = -N) \\
&= -\left(\nu A^{1/2}A^{1/2} - A^{1/2}A^{-1/2}NA^{-1/2}A^{1/2}\right)^{-1} N \\
&\quad \left(\nu A^{1/2}A^{1/2} + A^{1/2}A^{-1/2}NA^{-1/2}A^{1/2}\right)^{-1} \\
&= -(A^{1/2}(\nu I - \underbrace{A^{-1/2}NA^{-1/2}}_{=\tilde{N}})A^{1/2})^{-1}N \\
&\quad (A^{1/2}(\nu I + \underbrace{A^{-1/2}NA^{-1/2}}_{=\tilde{N}})A^{1/2})^{-1} \\
&= -A^{-1/2}(\nu I - \tilde{N})^{-1}\underbrace{A^{-1/2}NA^{-1/2}}_{=\tilde{N}}(\nu I + \tilde{N})^{-1}A^{1/2} \\
&= -A^{-1/2}(\nu I - \tilde{N})^{-1}\tilde{N}(\nu I + \tilde{N})^{-1}A^{-1/2}.
\end{aligned}
$$

Folglich ist

$$R := B\left(\frac{F^{-1} - F^{-T}}{2}\right)B^T = B\left(-A^{-1/2}(\nu I - \tilde{N})^{-1}\tilde{N}(\nu I + \tilde{N})^{-1}A^{-1/2}\right)B^T.$$

Für $\tilde{N} := A^{-1/2} N A^{-1/2}$ gilt

$$\begin{aligned}\tilde{N}^T &= (A^{-1/2} N A^{-1/2})^T \\ &= A^{-1/2}(-N)A^{-1/2} \quad (\text{da } (A^{-1/2})^T = A^{-1/2},\ N^T = -N) \\ &= -\tilde{N}.\end{aligned}$$

Also ist $\tilde{N}$ schiefsymmetrisch. Unter Verwendung dieser Eigenschaft bekommt man dann

$$\begin{aligned}\frac{(p, Rp)}{(p, \frac{1}{\nu} Tp)} &= \frac{\left(p, B\left(-A^{-1/2}(\nu I - \tilde{N})^{-1}\tilde{N}(\nu I + \tilde{N})^{-1}A^{-1/2}\right)B^T p\right)}{(p, \frac{1}{\nu} B A^{-1} B^T p)} \\ &= -\frac{((\nu I - \tilde{N})^{-T} A^{-1/2} B^T p, \tilde{N} \overbrace{(\nu I + \tilde{N})^{-1} A^{-1/2} B^T p}^{=v})}{(A^{-1/2} B^T p, \frac{1}{\nu} A^{-1/2} B^T p)} \\ &= -\frac{(\overbrace{(\nu I + \tilde{N})^{-1} A^{-1/2} B^T p}^{=v}, \tilde{N} v)}{(A^{-1/2} B^T p, \frac{1}{\nu}(\nu I + \tilde{N}) \underbrace{(\nu I + \tilde{N})^{-1} A^{-1/2} B^T p}_{=v})} \\ &= -\frac{(v, \tilde{N} v)}{\left(A^{-1/2} B^T p, \frac{1}{\nu}(\nu I - \tilde{N})^{-1}(\nu I - \tilde{N})(\nu I + \tilde{N}) v\right)} \\ &= -\frac{(v, \tilde{N} v)}{(\underbrace{(\nu I + \tilde{N})^{-1} A^{-1/2} B^T p}_{=v}, \frac{1}{\nu}(\nu^2 I - \tilde{N}^2) v)} \\ &= -\frac{\nu (v, \tilde{N} v)}{(v, (\nu^2 I - \tilde{N}^2) v)}. \qquad (4.9)\end{aligned}$$

□

Lemma 4.4 *Sei $\hat{A}$ der diskrete Laplace-Operator und N ein Operator erster Ordnung. Dann existiert eine Konstante $\delta \geq 0$, die unabhängig von der Gitterweite h ist, so daß gilt*

$$\rho(\hat{A}^{-1/2} N \hat{A}^{-1/2}) \leq \delta,$$

wobei ρ der Spektralradius ist.

Beweis: vgl. [ES86]

Nun sind alle Hilfsmittel bereitgestellt, um Aussagen über die Eigenwerte des Schur-Komplement Problems (4.6) machen zu können.

Satz 4.5 *Sei $S := B(\nu A + N)^{-1}B^T$. Für den Fall $\theta = 0$ gilt:*
Die Eigenwerte des Schur-Komplement Problems

$$Sp = \mu \frac{1}{\nu} Qp \tag{4.10}$$

für den diskreten Oseen-Operator sind enthalten in einem Reckteck in der rechten Halbebene, dessen Grenzen unabhängig von h sind.

Bemerkung 4.6
Es ist mir nicht gelungen, die Ausssage des Satzes 4.5 auf den Fall $\theta > 0$ zu übertragen. Da Lemma 4.4 nicht gilt, treten im Beweis Probleme auf. Trotzdem wird der Beweis, soweit dies möglich ist, allgemeiner formuliert. An den entsprechenden Stellen erfolgt ein expliziter Hinweis auf die Probleme im Fall $\theta > 0$. Dort wird auch eine Fallunterscheidung vorgenommen.

Beweis von Satz 4.5:
Sei $\mu \in \mathbb{C}$ eine Lösung der Eigenwertproblems $Sp = \mu \frac{1}{\nu} Qp$.

(1) Die Matrix $S := B(\nu A + N)^{-1}B^T$ läßt sich zerlegen in $S = C + R$, wobei C bzw. R definiert sind durch $C := \frac{1}{2}(S + S^T)$ bzw. $R := \frac{1}{2}(S - S^T)$. Unter Benutzung der Definition von S erhält man

$$\begin{aligned} C &= \frac{1}{2}\left(BF^{-1}B^T + (BF^{-1}B^T)^T\right) \\ &= \frac{1}{2}\left(BF^{-1}B^T + BF^{-T}B^T\right) \\ &= B\left(\frac{F^{-1} + F^{-T}}{2}\right)B^T. \end{aligned}$$

Durch analoge Rechnung bekommt man

$$R = B\left(\frac{F^{-1} - F^{-T}}{2}\right)B^T.$$

Nach Lemma 4.1 (i) gilt für den Realteil von μ die Abschätzung

$$\min_{p \neq 0} \frac{(p, Cp)}{(p, \frac{1}{\nu}Qp)} \leq Re\mu \leq \max_{p \neq 0} \frac{(p, Cp)}{(p, \frac{1}{\nu}Qp)}.$$

Im folgenden werden Grenzen für den Rayleigh-Quotient $\frac{(p,Cp)}{(p,\frac{1}{\nu}Qp)}$ konstruiert. Dazu definiere man $T := BA^{-1}B^T$. Für den Rayleigh-Quotienten gilt

$$\begin{aligned} \frac{(p, Cp)}{(p, \frac{1}{\nu}Qp)} &= \frac{(p, Cp)}{(p, \frac{1}{\nu}Qp)} \frac{(p, Tp)}{(p, Tp)} \\ &= \frac{(p, Cp)}{(p, \frac{1}{\nu}Tp)} \frac{(p, Tp)}{(p, Qp)}. \end{aligned} \tag{4.11}$$

Nach Lemma 3.6 gilt

$$\frac{\beta^2}{1+\frac{\theta}{\nu}C_F^2} \leq \frac{(p,Tp)}{(p,Qp)} \leq \kappa^2. \tag{4.12}$$

Die Konstanten β, C_F und κ sind unabhängig von der Diskretisierungsschrittweite h. Weiterhin ist nach Lemma 4.2

$$\frac{(p,Cp)}{(p,\frac{1}{\nu}Tp)} = \frac{\left(v,(I-\frac{1}{\nu^2}\tilde{N}^2)^{-1}v\right)}{(v,v)}, \tag{4.13}$$

mit $\tilde{N} := A^{-1/2}NA^{-1/2}$ und $v := A^{-1/2}B^Tp$. Da $\tilde{N}$ schiefsymmetrisch ist, folgt dann

$$-\tilde{N}^2 = -\tilde{N}\tilde{N} = \tilde{N}^T\tilde{N}. \tag{4.14}$$

Da $\tilde{N}^T\tilde{N}$ symmetrisch ist, sind die Eigenwerte von $-\tilde{N}^2$ reell.

Sei nun λ ein beliebiger Eigenwert von $-\tilde{N}^2$. Mit (4.14) folgt dann

$$\begin{aligned} x^T\tilde{N}^T\tilde{N}x &= \lambda\|x\|_2^2 \\ \Leftrightarrow \|\tilde{N}x\|_2^2 &= \lambda\|x\|_2^2 \\ \Leftrightarrow \lambda &= \frac{\|\tilde{N}x\|_2^2}{\|x\|_2^2} \geq 0. \end{aligned}$$

Also ist λ nichtnegativ. Folglich besitzt $-\tilde{N}^2$ nur nichtnegative Eigenwerte.

An dieser Stelle muß zwischen den Fällen $\theta = 0$ und $\theta > 0$ unterschieden werden.
Fall 1: $\theta = 0$
In diesem Fall ist A nur noch der diskrete Laplace-Operator. Somit ist Lemma 4.4 auf $\tilde{N}$ anwendbar. Die Eigenwerte von $\tilde{N}$ sind betragsmäßig durch eine Konstante δ, die unabhängig von h ist, beschränkt. Deshalb liegt das Spektrum von $I-\frac{1}{\nu^2}\tilde{N}^2$ im Intervall $\left[1, 1+\frac{\delta^2}{\nu^2}\right]$. Daraus folgt dann, daß das Spektrum von $\left(I-\frac{1}{\nu^2}\tilde{N}^2\right)^{-1}$ im Intervall $\left[\frac{\nu^2}{\nu^2+\delta^2}, 1\right]$ liegt. Zusammen mit (4.13) ergibt sich folgende Abschätzung

$$\frac{\nu^2}{\delta^2+\nu^2} \leq \frac{(p,Cp)}{(p,\frac{1}{\nu}Tp)} \leq 1. \tag{4.15}$$

Diese Abschätzung kombiniert mit (4.11), (4.12) und unter Berücksichtigung von $\theta = 0$ ergibt

$$\frac{\beta^2\nu^2}{\delta^2+\nu^2} \leq \frac{(p,Cp)}{(p,\frac{1}{\nu}Qp)} \leq \kappa^2.$$

Damit sind die Grenzen für den Realteil der Eigenwertes μ im Fall $\theta = 0$ konstruiert.

Fall 2: $\theta > 0$
In diesem Fall ist Lemma 4.4 nicht anwendbar. Es ist mir auch nicht gelungen, eine ähnliche Abschätzung zu zeigen. Aus diesem Grund bleibt die Abschätzung der Eigenwerte für $\tilde{N}$ im Fall $\theta > 0$ ein offenes Problem.

(2) Nach Lemma 4.1 (ii) gilt

$$|Im\mu| \leq \max_{p \neq 0} \frac{|(p, Rp)|}{(p, \frac{1}{\nu} Qp)}.$$

Mit $T := BA^{-1}B^T$ erhält man für den Quotienten $\frac{(p,Rp)}{(p,\frac{1}{\nu}Qp)}$

$$\frac{(p, Rp)}{(p, \frac{1}{\nu} Qp)} = \frac{(p, Rp)}{(p, \frac{1}{\nu} Tp)} \frac{(p, Tp)}{(p, Qp)}.$$

Nach Lemma 4.3 ist

$$\frac{(p, Rp)}{(p, \frac{1}{\nu} Tp)} = -\frac{\nu(v, \tilde{N}v)}{(v, (\nu^2 I - \tilde{N}^2)v)}, \tag{4.16}$$

mit $v := (\nu I + \tilde{N})^{-1} A^{1/2} B^T p$.
Für die schiefsymmetrische Matrix $\tilde{N} := A^{-1/2} N A^{-1/2}$ gilt

$$\tilde{N}^T \tilde{N} = -\tilde{N}(-\tilde{N}^T) = \tilde{N}\tilde{N}^T,$$

also ist $\tilde{N}$ normal. Deshalb kann man $\tilde{N}$ unitär auf Diagonalgestalt transformieren. Folglich existiert eine unitäre Matrix U, so daß gilt

$$\begin{aligned} U^H \tilde{N} U &= \tilde{\Lambda} \\ \Leftrightarrow \tilde{N} &= U \tilde{\Lambda} U^H, \end{aligned} \tag{4.17}$$

wobei $\tilde{\Lambda} = diag(\lambda_j)$ mit λ_j Eigenwerte von $\tilde{N}$.
Bei schiefsymmetrischen Matrizen treten die Eigenwerte stets in der Form $\lambda = \pm i\eta$ auf. Dann ergibt sich mit (4.17)

$$\tilde{N} = iU\Lambda U^H,$$

mit $\tilde{\Lambda} = i\Lambda$, $\Lambda \in \mathbb{R}^{l \times l}$. Daraus folgt dann

$$\begin{aligned} \tilde{N}^2 &= \tilde{N}\tilde{N} \\ &= i^2 U \Lambda \underbrace{U^H U}_{=I} \Lambda U^H \\ &= -U\Lambda^2 U^H. \end{aligned}$$

Mit (4.16) ergibt sich damit

$$\begin{aligned}
-\frac{\nu(v,\tilde{N}v)}{(v,(\nu^2 I-\tilde{N}^2)v)} &= -\frac{\nu(v,iU\Lambda U^H v)}{(v,(\nu^2 UU^H+U\Lambda^2 U^H)v)}\\
&= -\frac{\nu(v,iU\Lambda U^H v)}{(v,U(\nu^2 I+\Lambda^2)U^H v)}\\
&= -\frac{\nu(\overbrace{U^H v}^{=:w},i\Lambda\overbrace{U^H v}^{=:w})}{(\underbrace{U^H v}_{=:w},(\nu^2 I+\Lambda^2)\underbrace{U^H v}_{=:w})}\\
&= -\frac{\nu(w,i\Lambda w)}{(w,(\nu^2 I+\Lambda^2)w)}.
\end{aligned}$$

Weiterhin gilt

$$\begin{aligned}
\frac{|-\nu(w,i\Lambda w)|}{(w,(\nu^2 I+\Lambda^2)w)} &= \frac{|-i\nu||(w,\Lambda w)|}{(w,(\nu^2 I+\Lambda^2)w)}\\
&= \frac{\nu|(w,\Lambda w)|}{(w,(\nu^2 I+\Lambda^2)w)}. \qquad (4.18)
\end{aligned}$$

An dieser Stelle ist wieder eine Fallunterscheidung notwendig.

Fall 1: $\theta = 0$
In diesem Fall ist der Betrag der Eigenwerte von $\tilde{N}$ nach Lemma 4.4 durch eine Konstante δ beschränkt. Folglich ist (4.18) durch

$$\max_{-\delta\le\eta\le\delta}\frac{\nu|\eta|}{\nu^2+\eta^2}=\max_{0\le\eta\le\delta}\frac{\nu\eta}{\nu^2+\eta^2}$$

beschränkt. Dieses Maximum ist $\frac{1}{2}$ und wird angenommen, wenn $\eta=\nu$ ist. Insgesamt erhält man dann

$$\frac{|(p,Rp)|}{(p,\frac{1}{\nu}Qp)}\le\frac{\kappa^2}{2}.$$

Der Betrag des Imaginärteils von μ ist damit unabhängig von der Diskretisierungsschrittweite h beschränkt.

Fall 2: $\theta > 0$
Bei diesem Fall tritt das gleiche Problem, wie in Beweisteil (1) auf, nämlich die Abschätzung der Eigenwerte von $\tilde{N}$. Auch an dieser Stelle bleibt ein offenes Problem.

□

Im Satz 4.5 sind die Grenzen für den Real- bzw. Imaginärteil der Eigenwerte des Schur-Komplement Problems ermittelt worden.

Ziel ist es aber, die Eigenwerte von

$$\begin{pmatrix} \nu A + N & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \lambda \begin{pmatrix} \nu A + N & 0 \\ 0 & \frac{1}{\nu} Q \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} \tag{4.19}$$

abzuschätzen. Da aber $\mu := \lambda(\lambda - 1)$ erhält man für die Eigenwerte des Problems (4.19) folgende Aussage.

Folgerung 4.7
Für den Fall $\theta = 0$ gilt:
$\lambda = 1$ ist Eigenwert mit der Vielfachheit $l - k$ des mit

$$M_D^{-1} := \begin{pmatrix} \nu A + N & 0 \\ 0 & \frac{1}{\nu} Q \end{pmatrix}^{-1}$$

vorkonditionierten diskreten Oseen-Operators. Die weiteren Eigenwerte des vorkonditionierten Systems sind vier Mengen bestehend aus Punkten der Form $1 + (a \pm bi)$ und $-a \pm bi$. Diese Mengen können in zwei rechteckige Gebiete eingeschlossen werden, wobei die Gebiete symmetrisch bezüglich $Re(\lambda) = \frac{1}{2}$ sind. Außerdem sind die Gebiete unabhängig von der Diskretisierungsschrittweite h beschränkt.

Beweis:
Die Eigenwerte λ von (4.19) ergeben sich aus μ durch

$$\begin{aligned} \lambda &= \frac{1 \pm \sqrt{1 + 4\mu}}{2} \\ &= \frac{1 \pm (x_1 + i x_2)}{2} \quad (\, x_1, x_2 \in \mathbb{R}) \\ &= \frac{1}{2} \pm \frac{x_1}{2} \pm i \frac{x_2}{2}. \end{aligned}$$

Definiere nun

$$\begin{aligned} a &:= \frac{x_1}{2} - \frac{1}{2}, \\ b &:= \frac{x_2}{2}. \end{aligned}$$

Dann ist

$$\begin{aligned} 1 + (a \pm ib) &= 1 + \left(\frac{x_1}{2} - \frac{1}{2} \right) \pm i \frac{x_2}{2} \\ &= \frac{1}{2} + \frac{x_1}{2} \pm i \frac{x_2}{2}, \\ -(a \pm ib) &= -\left(\frac{x_1}{2} - \frac{1}{2} \right) \pm i \frac{x_2}{2} \\ &= \frac{1}{2} - \frac{x_1}{2} \pm i \frac{x_2}{2}. \end{aligned}$$

Also lassen sich alle λ aus (4.19) in der Form $1 + (a \pm ib)$ und $-a \pm ib$ darstellen. Da der Realteil und der Imaginärteil von μ nach Satz 4.5 beschränkt sind, erkennt man an obiger Darstellung, daß die Eigenwerte von (4.19) in zwei rechteckigen Gebieten liegen, die bezüglich $Re(\lambda) = \frac{1}{2}$ symmetrisch sind. Die Rechtecke sind unabhängig von h beschränkt, da die Schranken von $Re(\mu)$ und $Im(\mu)$ unabhängig von h sind.

□

Die Eigenwerte des vorkonditionierten Systems liegen auf beiden Seiten der imaginären Achse. Aus diesem Grund kann es sein, daß der betragsmäßig kleinste Eigenwert nicht gegen Null beschränkt ist. Diese Tatsache kann ein möglicher Nachteil der Vorkonditionierung mit M_D sein. Eine Alternative, die dieses Problem vermeidet, ist ein Blockdreiecksvorkonditionierer, der definiert ist durch

$$M_T := \begin{pmatrix} F & B^T \\ 0 & -\frac{1}{\nu}Q \end{pmatrix},$$

wobei $F := A + \nu N$. Für diesen Fall lautet das verallgemeinerte Eigenwertproblem

$$\begin{pmatrix} F & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \lambda \begin{pmatrix} F & B^T \\ 0 & -\frac{1}{\nu}Q \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix}. \tag{4.20}$$

Äquivalent dazu ist

$$\begin{aligned} Fu + B^T p &= \lambda(Fu + B^T p), && (4.21) \\ Bu &= \lambda\left(-\frac{1}{\nu}Qp\right). && (4.22) \end{aligned}$$

Eine Lösung des Problems (4.20) ist $\lambda = 1$. Dann erhält man mit (4.22)

$$\begin{aligned} Bu &= -\frac{1}{\nu}Qp \\ \Leftrightarrow -\nu Bu &= Qp. \end{aligned} \tag{4.23}$$

Der Eigenwert $\lambda = 1$ hat die Vielfachheit l. Für $\lambda \neq 1$ erhält man durch Multiplikation von (4.21) mit BF^{-1}

$$\begin{aligned} BF^{-1}(Fu + B^T p) &= \lambda BF^{-1}(Fu + B^T p) \\ \Leftrightarrow Bu + \underbrace{BF^{-1}B^T}_{=S} p &= \lambda Bu + \lambda \underbrace{BF^{-1}B^T}_{=S} p. \end{aligned}$$

Unter Benutzung von (4.22) ergibt sich

$$\begin{aligned} \lambda\left(-\frac{1}{\nu}Qp\right) + Sp &= \lambda\left(\lambda\left(-\frac{1}{\nu}Qp\right)\right) + \lambda Sp \\ \Leftrightarrow (1-\lambda)Sp &= (\lambda - \lambda^2)\left(\frac{1}{\nu}Qp\right) \\ \Leftrightarrow Sp &= \lambda\left(\frac{1}{\nu}Qp\right). \end{aligned} \tag{4.24}$$

Die Eigenwerte dieses Problems sind für den Fall $\theta = 0$ im Satz 4.5 abgeschätzt worden. Also erhält man folgendes Resultat.

Satz 4.8 *Für den Fall $\theta = 0$ gilt:*
$\lambda = 1$ ist Eigenwert mit der Vielfachheit l des mit

$$M_T^{-1} := \begin{pmatrix} \nu A + N & B^T \\ 0 & \frac{1}{\nu} Q \end{pmatrix}^{-1}$$

vorkonditionierten diskreten Oseen-Operators. Die weiteren Eigenwerte des vorkonditionierten Systems erhält man aufgrund der obigen Herleitung aus $Sp = \lambda \frac{1}{\nu} Qp$. Folglich sind die Eigenwerte der vorkonditionierten Oseen-Problems unabhängig von der Diskretisierungsschrittweite h beschränkt.

Bemerkung 4.9
Sei $\lambda \in \mathbb{C}$ ein Eigenwert von (4.24). Dann gelten für $\theta = 0$ nach Satz 4.5 folgende Abschätzungen:

(i)

$$\frac{\beta^2 \nu^2}{\delta^2 + \nu^2} \leq Re(\lambda) \leq \kappa^2,$$

(ii)

$$|Im(\lambda)| \leq \frac{\kappa^2}{2}.$$

Folglich ist der betragsmäßig kleinste Eigenwert des mit M_T^{-1} vorkonditionierten Gleichungssystems unabhängig von h gegen Null beschränkt. Damit ist das bei der Vorkonditionierung mit M_D^{-1} auftretende Problem behoben.

Bemerkung 4.10
Bei Benutzung einer der Vorkonditionierer M_D oder M_T ist es erforderlich, in jedem Schritt eines iterativen Verfahrens die Wirkung von F^{-1} zu berechnen. Die Anwendung von F^{-1} auf einen Vektor mittels direkter Methoden ist sehr teuer.
Eine Alternative zu der direkten Berechnung wird im folgenden Abschnitt vorgestellt.

4.2 Praktische Aspekte der Berechnung

Die hauptsächlichen Kosten der Vorkonditionierer entstehen durch die Anwendung von F^{-1} bzw. F^{-T} auf einen Vektor v in jedem Schritt der Iteration. In diesem Abschnitt soll gezeigt werden, daß man diese Operation durch eine billigere ersetzen kann, zum Beispiel durch Berechnung der approximativen Lösung des Systems $Fw = v$ bzw.

$F^T w = v$ mit iterativen Verfahren . Die Idee ist, die Vorkonditionierungsoperatoren M_D und M_T durch

$$\hat{M}_D := \begin{pmatrix} \hat{F} & 0 \\ 0 & \frac{1}{\nu}Q \end{pmatrix} \qquad \text{und} \qquad \hat{M}_T := \begin{pmatrix} \hat{F} & B^T \\ 0 & -\frac{1}{\nu}Q \end{pmatrix} \tag{4.25}$$

zu ersetzen, wobei $\hat{F} \approx F$ ist. Im folgenden werden die Vorkonditionierer, die die exakte Wirkung von F^{-1} verwenden, als exakte Versionen und die auf der Approximation in (4.25) beruhenden als inexakte Versionen bezeichnet. Weiterhin sei Q_ν definiert durch $Q_\nu := \frac{1}{\nu}Q$. Sei $\mathcal{A}_D$ die vorkonditionierte Matrix bei Verwendung des exakten Blockdiagonalvorkonditionierers M_D. Es gilt also

$$\mathcal{A}_D := \begin{pmatrix} F & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} F & 0 \\ 0 & Q_\nu \end{pmatrix}^{-1} = \begin{pmatrix} I & B^T Q_\nu^{-1} \\ BF^{-1} & 0 \end{pmatrix}.$$

Sei $\hat{\mathcal{A}}_D$ die vorkonditionierte Matrix bei Verwendung der inexakten Version des Blockdiagonalvorkonditionierers. Man erhält unter Benutzung der Definition von $\mathcal{A}_D$ und von $E := \hat{F} - F$

$$\begin{aligned}
\hat{\mathcal{A}}_D &= \begin{pmatrix} F & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \hat{F} & 0 \\ 0 & Q_\nu \end{pmatrix}^{-1} \\
&= \begin{pmatrix} F\hat{F}^{-1} & B^T Q_\nu^{-1} \\ B\hat{F}^{-1} & 0 \end{pmatrix} \\
&= \begin{pmatrix} F\hat{F}^{-1} + I - I & B^T Q_\nu^{-1} \\ B\hat{F}^{-1} + BF^{-1} - BF^{-1} & 0 \end{pmatrix} \\
&= \begin{pmatrix} I & B^T Q_\nu^{-1} \\ BF^{-1} & 0 \end{pmatrix} - \begin{pmatrix} \hat{F}\hat{F}^{-1} - F\hat{F}^{-1} & 0 \\ BF^{-1} - B\hat{F}^{-1} & 0 \end{pmatrix} \\
&= \mathcal{A}_D - \begin{pmatrix} (\hat{F} - F)\hat{F}^{-1} & 0 \\ BF^{-1}(\hat{F} - F)\hat{F}^{-1} & 0 \end{pmatrix} \\
&= \mathcal{A}_D - \begin{pmatrix} E\hat{F}^{-1} & 0 \\ BF^{-1}E\hat{F}^{-1} & 0 \end{pmatrix}.
\end{aligned}$$

Definiere nun

$$\mathcal{E}_D = -\begin{pmatrix} E\hat{F}^{-1} & 0 \\ BF^{-1}E\hat{F}^{-1} & 0 \end{pmatrix}.$$

Damit erhält man $\hat{\mathcal{A}}_D = \mathcal{A}_D + \mathcal{E}_D$.

Sei $\mathcal{A}_T$ die vorkonditionierte Matrix bei Verwendung der exakten Version der Blockdreiecksvorkonditionierers. Es gilt

$$\begin{aligned}
\mathcal{A}_T &:= \begin{pmatrix} F & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} F & B^T \\ 0 & Q_\nu \end{pmatrix}^{-1} \\
&= \begin{pmatrix} I & 0 \\ BF^{-1} & BF^{-1}B^T Q_\nu^{-1} \end{pmatrix}
\end{aligned}$$

Weiterhin sei $\hat{\mathcal{A}}_T$ die vorkonditionierte Matrix bei Verwendung der inexakten Version des Blockdreiecksvorkonditionierers. Unter Verwendung der Definition von $\mathcal{A}_T$ und von $E := \hat{F} - F$ bekommt man

$$\begin{aligned}
\hat{\mathcal{A}}_T &= \begin{pmatrix} F & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \hat{F} & B^T \\ 0 & -Q_\nu \end{pmatrix}^{-1} \\
&= \begin{pmatrix} F\hat{F}^{-1} & F\hat{F}^{-1}B^TQ_\nu^{-1} - B^TQ_\nu^{-1} \\ B\hat{F}^{-1} & B\hat{F}^{-1}B^TQ_\nu^{-1} \end{pmatrix} \\
&= \begin{pmatrix} F\hat{F}^{-1} + I - \hat{F}\hat{F}^{-1} & F\hat{F}^{-1}B^TQ_\nu^{-1} - \hat{F}\hat{F}^{-1}B^TQ_\nu^{-1} \\ B\hat{F}^{-1} + BF^{-1} - BF^{-1} & (B\hat{F}^{-1} + BF^{-1} - BF^{-1})B^TQ_\nu^{-1} \end{pmatrix} \\
&= \begin{pmatrix} I & 0 \\ BF^{-1} & BF^{-1}B^TQ_\nu^{-1} \end{pmatrix} - \begin{pmatrix} E\hat{F}^{-1} & E\hat{F}^{-1}B^TQ_\nu^{-1} \\ BF^{-1}E\hat{F}^{-1} & BF^{-1}E\hat{F}^{-1}B^TQ_\nu^{-1} \end{pmatrix} \\
&= \mathcal{A}_T - \begin{pmatrix} E\hat{F}^{-1} & E\hat{F}^{-1}B^TQ_\nu^{-1} \\ BF^{-1}E\hat{F}^{-1} & BF^{-1}E\hat{F}^{-1}B^TQ_\nu^{-1} \end{pmatrix}.
\end{aligned}$$

Definiere nun

$$\mathcal{E}_T = -\begin{pmatrix} E\hat{F}^{-1} & E\hat{F}^{-1}B^TQ_\nu^{-1} \\ BF^{-1}E\hat{F}^{-1} & BF^{-1}E\hat{F}^{-1}B^TQ_\nu^{-1} \end{pmatrix}.$$

Dann erhält man für die vorkonditionierte Matrix $\hat{\mathcal{A}}_T = \mathcal{A}_T + \mathcal{E}_T$. Um Aussagen über die Grenzen der Eigenwerte des vorkonditionierten Systems bei Verwendung der inexakten Vorkonditionierer machen zu können, ist des Satz von Bauer-Fike als Hilfsmittel erforderlich. Auf den Beweis des Satzes wird an dieser Stelle verzichtet. Man kann in [GL89] S.342 den Nachweis des Satzes finden.

Satz 4.11 (Bauer-Fike) *Sei μ ein Eigenwert von $A+E \in \mathbb{C}^{n\times n}$ und $X^{-1}AX = D = diag(\lambda_1, \ldots, \lambda_n)$. Dann gilt*

$$\min_{i=1,\ldots,n} |\lambda_i - \mu| \leq \kappa_p(X)\|E\|_p,$$

wobei $\|\cdot\|_p$ die p-Norm ($p = 1, 2, \infty$) und $\kappa_p(X)$ die Konditionszahl der Matrix X bezüglich der p-Norm bezeichnet.

Für den Fall, daß die Matrizen $\mathcal{A}_D$ bzw. $\mathcal{A}_T$ diagonalisierbar sind, ergeben sich für die Eigenwerte des vorkonditionierten Systems bei Verwendung der inexakten Vorkonditionierer folgende Grenzen

Satz 4.12 *(i) Sei $\mathcal{A}_D = \mathcal{V}_D\Lambda_D\mathcal{V}_D^{-1}$ diagonalisierbar, dann gilt für alle Eigenwert $\mu \in \sigma(\hat{\mathcal{A}}_D)$*

$$\min_{\lambda\in\sigma(\mathcal{A}_D)} |\lambda - \mu| \leq \|E\hat{F}^{-1}\|_\infty \kappa_\infty(\mathcal{V}_D) max\left(1, \|BF^{-1}\|_\infty\right),$$

wobei $\kappa_\infty(\mathcal{V}_D)$ die Konditionszahl der Matrix $\mathcal{V}_D$ bezüglich der ∞-Norm ist.
(ii) Sei $\mathcal{A}_T = \mathcal{V}_T\Lambda_T\mathcal{V}_T^{-1}$ diagonalisierbar, dann gilt für alle Eigenwerte $\mu \in \sigma(\hat{\mathcal{A}}_T)$

$$\min_{\lambda\in\sigma(\mathcal{A}_T)} |\lambda - \mu| \leq \|E\hat{F}^{-1}\|_\infty \kappa_\infty(\mathcal{V}_T)(1 + \|B^TQ^{-1}\|_\infty max\left(1, \|BF^{-1}\|_\infty\right),$$

wobei $\kappa_\infty(\mathcal{V}_T)$ die Konditionszahl der Matrix $\mathcal{V}_T$ bezüglich der ∞-Norm ist.

Beweis: (i) Sei $\mu \in \sigma(\mathcal{A}_D + \mathcal{E}_D)$ und $\mathcal{A}_D = \mathcal{V}_D \Lambda_D \mathcal{V}_D^{-1}$, dann gilt mit Satz 4.11

$$\min_{\lambda \in \sigma(\mathcal{A}_D)} |\lambda - \mu| \leq \kappa_\infty(\mathcal{V}_D) \|\mathcal{E}_D\|_\infty. \tag{4.26}$$

Mit der Definition von $\mathcal{E}_D$ gilt dann

$$\begin{aligned}
\|\mathcal{E}_D\|_\infty &= \left\| \begin{pmatrix} E\hat{F}^{-1} & 0 \\ BF^{-1}E\hat{F}^{-1} & 0 \end{pmatrix} \right\|_\infty \\
&= \left\| \begin{pmatrix} I & 0 \\ BF^{-1} & 0 \end{pmatrix} \begin{pmatrix} E\hat{F}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \right\|_\infty \\
&\leq \left\| \begin{pmatrix} I & 0 \\ BF^{-1} & 0 \end{pmatrix} \right\|_\infty \left\| \begin{pmatrix} E\hat{F}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \right\|_\infty \\
&= max(1, \|BF^{-1}\|_\infty) \|E\hat{F}^{-1}\|_\infty.
\end{aligned}$$

Aus dieser Abschätzung erhält man mit (4.26)

$$\min_{\lambda \in \sigma(\mathcal{A}_D)} |\lambda - \mu| \leq \kappa_\infty(\mathcal{V}_D) max(1, \|BF^{-1}\|_\infty) \|E\hat{F}^{-1}\|_\infty.$$

(ii) Sei $\mu \in \sigma(\mathcal{A}_T + \mathcal{E}_T)$ und $\mathcal{A}_T = \mathcal{V}_T \Lambda_T \mathcal{V}_T^{-1}$, dann gilt mit Satz 4.11

$$\min_{\lambda \in \sigma(\mathcal{A}_T)} |\lambda - \mu| \leq \kappa_\infty(\mathcal{V}_T) \|\mathcal{E}_T\|_\infty. \tag{4.27}$$

Mit der Definition von $\mathcal{E}_T$ gilt nun

$$\begin{aligned}
\|\mathcal{E}_T\|_\infty &= \left\| \begin{pmatrix} E\hat{F}^{-1} & E\hat{F}^{-1}B^T Q_\nu^{-1} \\ BF^{-1}E\hat{F}^{-1} & BF^{-1}E\hat{F}B^T Q_\nu^{-1} \end{pmatrix} \right\|_\infty \\
&= \left\| \begin{pmatrix} I & 0 \\ BF^{-1} & 0 \end{pmatrix} \begin{pmatrix} E\hat{F}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} I & B^T Q_\nu^{-1} \\ 0 & 0 \end{pmatrix} \right\|_\infty \\
&\leq \left\| \begin{pmatrix} I & 0 \\ BF^{-1} & 0 \end{pmatrix} \right\|_\infty \left\| \begin{pmatrix} E\hat{F}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \right\|_\infty \left\| \begin{pmatrix} I & B^T Q_\nu^{-1} \\ 0 & 0 \end{pmatrix} \right\|_\infty \\
&= max(1, \|BF^{-1}\|_\infty) \|E\hat{F}^{-1}\|_\infty \left(1 + \|B^T Q_\nu^{-1}\|_\infty\right)
\end{aligned}$$

Aus dieser Abschätzung bekommt man mit (4.27)

$$\min_{\lambda \in \sigma(\mathcal{A}_T)} |\lambda - \mu| \leq \kappa_\infty(\mathcal{V}_T) max(1, \|BF^{-1}\|_\infty) \|E\hat{F}^{-1}\|_\infty \left(1 + \|B^T Q_\nu^{-1}\|_\infty\right).$$

□

4.3 Ausblick auf eine weitere Vorkonditionierungsstrategie

Ein weitere Möglichkeit der Blockdreiecksvorkonditionierung wird in [KS97] vorgestellt. Der Vorkonditionierer ist definiert durch

$$B^{-1} := \begin{pmatrix} \nu A + N & B^T \\ 0 & -B(\nu A + N)^{-1} B^T \end{pmatrix}^{-1}.$$

Die Konvergenzrate des iterativen Lösungsverfahren GMRES (vgl.[Saa96]) in Kombination mit den dem Blockdreiecksvorkonditionierer B^{-1} ist unabhängig von der Diskretisierungsschrittweite h. Die genaue Analyse findet man in [KS97]. Weiterhin wird auch noch eine Konvergenzanalyse durchgeführt, wenn man F^{-1} und $S^{-1} := (BF^{-1}B^T)^{-1}$ durch die approximativen Inversen $\hat{F}^{-1}$ und $\hat{S}^{-1}$ ersetzt.

Kapitel 5

Experimentelle Untersuchungen iterativer Löser für Stokes- und Oseen-Gleichungen

Am Institut für Numerische und Angewandte Mathematik der Universität Göttingen wird ein Programmmpaket namens ParallelNS (Parallelized solution of Navier-Stokes equations) zur Lösung von partiellen Differentialgleichungen entwickelt. Mit ParallelNS können verschiedene Probleme von skalaren Diffusions-Konvektions-Reaktions-Gleichungen bis hin zu Navier-Stokes-Gleichungen gelöst werden. Zur Lösung der bei der Diskretisierung dieser partiellen Differentialgleichungen entstehenden Gleichungssysteme wird das Programmpaket BLANC verwendet. Dieses Paket enthält verschiedene iterative Verfahren zur Lösung linearer Gleichungssysteme. Die Lösungsverfahren können nun mit unterschiedlichen Vorkonditionierern kombiniert werden. Ziel der in diesem Abschnitt dokumentierten Untersuchungen war es, für verschiedene Stokes- und Oseen-Probleme iterative Lösungsverfahren zu ermitteln, die die gestellten Probleme effizient lösen können. Dabei sind die Löser mit verschiedenen Vorkonditionierern getestet worden. Im Abschnitt 5.5 werden an einem lid-driven cavity Problem die im Kapitel 4 vorgestellten Vorkonditionierungsstrategien und die im Paket BLANC zur Verfügung gestellten Vorkonditionierer in Verbindung mit dem iterativen Löser QMR verglichen.

5.1 Durchführung der Experimente

Das Programmpaket BLANC ist linearen Gleichungssystemen angepaßt, die bei der Diskretisierung von vektorwertigen partiellen Differentialgleichungen entstehen.

Bei der Lösung einer vektorwertigen partiellen Differentialgleichung, z.B. eines zweidimensionalen Oseen-Problems, sind auf jedem Gitterpunkt drei Unbekannte (x-Komponente der Geschwindigkeit, y-Komponente der Geschwindigkeit und der Druck). Der klassische Weg ist die sequentielle Anordnung der Vektorkomponenten. Damit ist ge-

meint, daß zuerst alle Vektoreinträge, die zur ersten Unbekannten gehören, gespeichert werden. Dann werden alle Einträge gespeichert, die zur zweiten Unbekannten gehören usw..
In BLANC erfolgt eine punktweise Anordnung der Vektorkomponenten. Für jeden Gitterpunkt werden die Einträge der x-Komponente der Geschwindigkeit, der y-Komponente der Geschwindigkeit und des Drucks zusammengefaßt. Deshalb hat jeder Vektor in BLANC Blockstruktur, wobei jeder Block zu einem Gitterpunkt gehört. Die Dimension eines solchen Blocks entspricht der Anzahl der Unbekannten in diesem Gitterpunkt. Die Matrizen in BLANC sind ähnlich strukturiert. Jede Matrix besteht aus Blockmatrizen, deren Dimension der Anzahl der Unbekannten in einem Gitterpunkt entspricht. Aufgrund der dargestellten Blockstruktur ist es möglich, beim SSOR-Verfahren jede Komponente des Lösungsvektors für sich zu relaxieren. In den folgenden Untersuchungen wurden die einzelnen Elemente jedes Blocks verschieden relaxiert. Damit ist berücksichtigt worden, daß in jedem Block Geschwindigkeitskomponenten und eine Druckkomponente enthalten sind. In die experimentellen Untersuchungen sind aus dem Programmpaket BLANC folgende iterative Löser einbezogen worden:

- stationäre Verfahren:
  - SSOR (Symmetric Successive OverRelaxation)
- Petrov-Galerkin-Krylov-Verfahren (PGK-Verfahren):
  - Bi-CG (Bi-Conjugate Gradient-Verfahren)
  - CGS (Conjugate Gradient Stabilized-Verfahren)
  - Bi-CGSTAB (Bi-CG Stabilized)
  - QMR (Quasi Minimized Residual)
  - TFQMR (Transposed-Free QMR)
  - QMRCGSTAB
  - CGNR (Conjugate Gradient Normal Residual)

Die PGK-Verfahren sind noch mit unterschiedlichen Vorkonditionierungen kombiniert worden:

- ohne Vorkonditionierung
- Jacobi
- SSOR
- ILU(0)

Weitere Informationen zu diesem Programmpaket findet man in [Pri96]. Genauere Aussagen über die verwendeten iterativen Löser sowie Algorithmen kann man zum Beispiel in [Saa96] bzw. [Pri96] nachlesen.

Um die Qualität der Lösungsverfahren miteinander zu vergleichen, wurde bei den numerischen Experimenten der zeitliche Konvergenzverlauf betrachtet, also die Entwicklung der Konvergenz $\|r_k\|_2/\|r_0\|_2$ des Residuums in der Euklidischen Norm bezüglich der Rechenzeit. Die Diskretisierung der Probleme und somit die Generierung der Matrix wurde mit dem Programm ParallelNS vorgenommen. ParallelNS arbeitet mit linearen Elementen für Druck und Geschwindigkeit, welche die Babuška-Brezzi mit einer von h unabhängigen Konstante nicht erfüllen. Deshalb wird in diesem Programm das in Abschnitt 3.4 kurz vorgestellte GLS-Verfahren zur Stabilisierung verwendet. Die Zerlegung des Einheitsquadrats wurde durch ein regelmäßiges 16×16-, 32×32-, 64×64- oder 128×128-Schachbrettmuster realisiert, bei dem jedes Feld diagonal halbiert wurde, vgl. Abbildung 5.1.

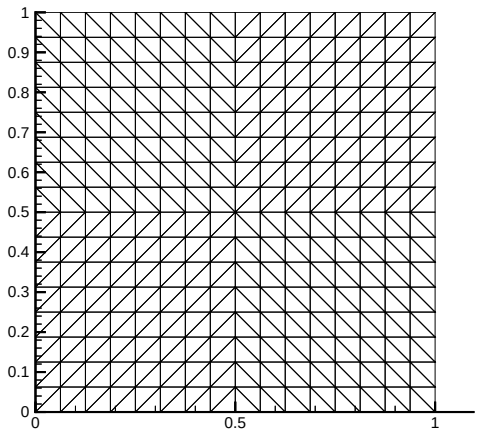


Abbildung 5.1: Beispiel einer verwendeten Zerlegung (17×17)

Es entstanden Gleichungssysteme mit $n = 3\cdot17\cdot17 = 867$, $n = 3\cdot33\cdot33 = 3267$, $n = 3\cdot65\cdot65 = 12675$, $n = 3\cdot129\cdot129 = 49923$ Unbekannten.
Zur Durchführung der Experimente wurden die Startlösung $x_0 = \frac{1}{\sqrt{n}}(1,...,1)^T$ und die rechte Seite $b = 0$ gesetzt. Verglichen wurde die Rechenzeit die erforderlich war, um das Abbruchkriterium $\|r_k\|_2/\|r_0\|_2 < 10^{-6}$ zu erreichen.
Beim SSOR-Verfahren wurden die Relaxationsparameter für die Geschwindigkeitskomponenten und den Druck verschieden gewählt. Im folgenden wird mit ω_v der Relaxationsparameter für die Geschwindigkeitskomponenten und mit ω_p der Relaxationsparameter für die Druckkomponente bezeichnet. Als Rechenumgebung für die numerischen Experimente sind DEC Alpha 3000/600 verwendet worden.

5.2 Stokes-Gleichung

Zuerst wurde eine Stokes-Gleichung, bei der das Strömungsfeld a verschwindet, untersucht. Da das Programm ParallelNS mit der GLS-Stabilisierungstechnik arbeitet, ist es erforderlich, bestimmte Stabilisierungsparameter zu wählen (vgl. Kapitel 3.4). Genauere Aussagen über die geeignete Wahl der Stabilisierungsparameter für verschiedene Probleme kann man in [Mü97] nachlesen. In diesem Beispiel wurde nur eine Druckstabilisierung verwendet. Die Rechnungen sind mit folgender Parameterwahl durchgeführt worden:

- Diffusionskoeffizient: $\nu = 1.0$,
- Stabilisierungsparameter für Geschwindigkeit: $c_1^u = 0.0$,
- Stabilisierungsparameter für Druck: $c_1^p = 1.0$,
- Stabilisierungsparameter für die Divergenzgleichung: $c_2 = 0.0$.

Außerdem wurde mit Konsistenzsicherung gerechnet. Diese benötigt man, denn bei der Verwendung von linearen Elementen für die Geschwindigkeit, wie in ParallelNS, ist $\triangle u_h|_K = 0$. Also verschwindet in der diskreten Formulierung der Laplace-Operator, deshalb erfüllt die kontinuierliche Lösung die diskreten Gleichungen nicht mehr. Aus diesem Grund ist dieses Verfahren inkonsistent und führt zu Fehlern in der Lösung. Die Darstellung einer Konsistenzsicherung findet man in [DH94]. Zuerst wurden für das SSOR-Verfahren die optimalen Relaxationsparameter ermittelt. In Tabelle 5.1 sind die optimalen Relaxationsparameter der Geschwindigkeitskomponenten und der Druckkompomente in Abhängigkeit von der Zerlegung angegeben. Außerdem enthält die Tabelle Informationen über die Rechenzeit, die erforderlich war, um das Abbruchkriteriums $\|r_k\|_2/\|r_0\|_2 < 10^{-6}$ zu erreichen. Zusätzlich ist noch die Zahl der dafür notwendigen Iterationen angegeben.

Zerlegung	ω_v	ω_p	CPU-sec	Iterationen
17×17	1.4	0.2	0.7666	120
33×33	1.8	0.2	4.233	160
65×65	1.9	0.1	32.899	305
129×129	1.95	0.05	313.704	606

Tabelle 5.1: Optimale Relaxationsparameter für SSOR beim Stokes-Problem

Die Abhängigkeit der optimalen Relaxationsparamter von der Feinheit der Zerlegung ist aus der Tabelle deutlich erkennbar. Mit kleiner werdender Diskretisierungsschrittweite h müssen die Geschwindigkeitskomponenten stärker überrelaxiert werden, bei der Druckkomponente ist eine stärkere Unterrelaxation erforderlich.
Die oben genannten PGK-Verfahren sind bei jeder Zerlegung mit den angegebenen Vorkonditionierern getestet worden, wobei für SSOR-Vorkonditionierung der vorher ermittelte optimale Relaxationsparameter verwendet wird. Außerdem sind noch weitere Relaxationsparameter für die SSOR-Vorkonditionierung getestet worden, um festzustellen, wie sich die nicht optimale Wahl des Relaxationsparameters für SSOR-Vorkonditionierung auf die Konvergenz des jeweiligen PGK-Verfahrens auswirkt. In der Tabelle 5.2 sind für jedes Verfahren die optimalen Vorkonditionierer, die Rechenzeit und die Iterationszahl angegeben. Der Eintrag ">sec" bedeutet, daß das PGK-Verfahren mit keinem der angegebenen Vorkonditionierer in der Zeit "sec" das Abbruchkriterium erfüllt. Weiterhin wird mit $\omega_{opt(A)}$ der optimale Relaxationsparameter der Zerlegung AxA bezeichnet (vgl. Tabelle 5.1).

Zerlegung	Verfahren	optimaler Vorkonditionierer	CPU-sec	Iterationen
17×17	BICG	ILU	0.9166	71
	CGNR		>20	
	QMR	ILU	0.9499	69
	BICGSTAB	SSOR($\omega_v = 1.6; \omega_p = 0.2$)	0.5166	34
	CGS	SSOR($\omega_v = 1.6; \omega_p = 0.2$)	0.6499	42
	TFQMR	SSOR($\omega_v = 1.6; \omega_p = 0.2$)	0.9166	86
	QMRCGSTAB	SSOR($\omega_v = 1.6; \omega_p = 0.2$)	0.5333	34
33×33	BICG	ILU	13.033	270
	CGNR		>50	
	QMR	ILU	15.316	273
	BICGSTAB	SSOR($\omega_{opt(33)}$)	2.567	44
	CGS	SSOR($\omega_{opt(33)}$)	3.383	58
	TFQMR	SSOR($\omega_{opt(33)}$)	5.183	124
	QMRCGSTAB	SSOR($\omega_{opt(33)}$)	2.817	47
65×65	BICG		>100	
	CGNR		>100	
	QMR		>100	
	BICGSTAB	SSOR($\omega_{opt(65)}$)	21.332	72
	CGS	SSOR($\omega_{opt(65)}$)	28.282	102
	TFQMR	SSOR($\omega_{opt(65)}$)	38.398	211
	QMRCGSTAB	SSOR($\omega_{opt(65)}$)	23.866	76
129×129	BICG		>400	
	CGNR		>400	
	QMR		>400	
	BICGSTAB	SSOR($\omega_{opt(129)}$)	146.961	119
	CGS	SSOR($\omega_{opt(129)}$)	321.637	261
	TFQMR		>400	
	QMRCGSTAB	SSOR($\omega_{opt(129)}$)	158.160	126

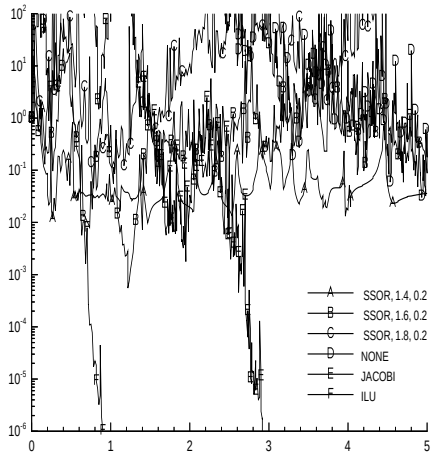
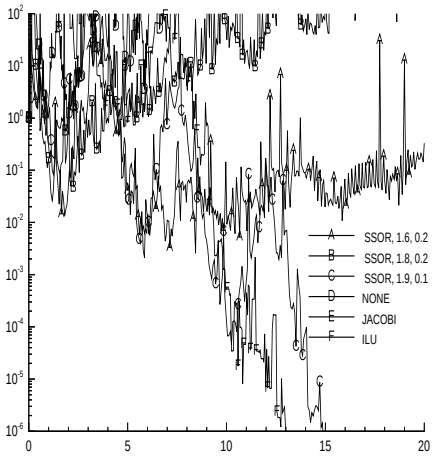
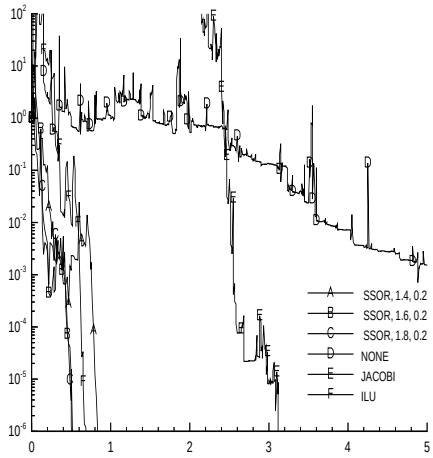
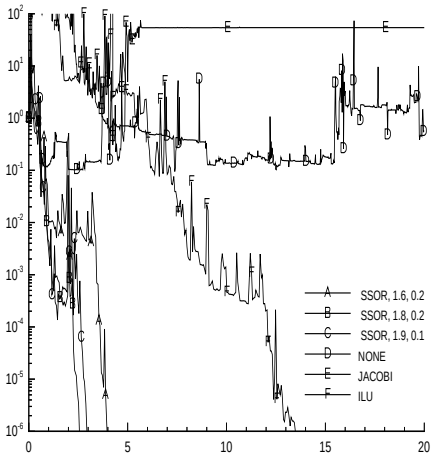
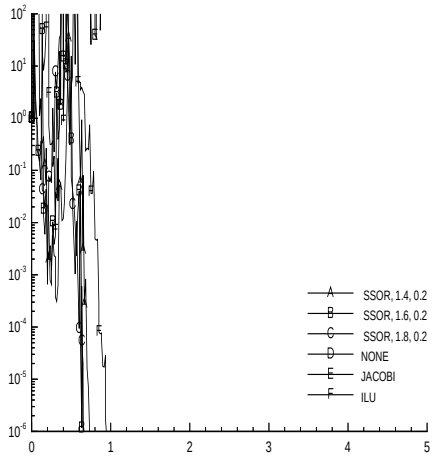
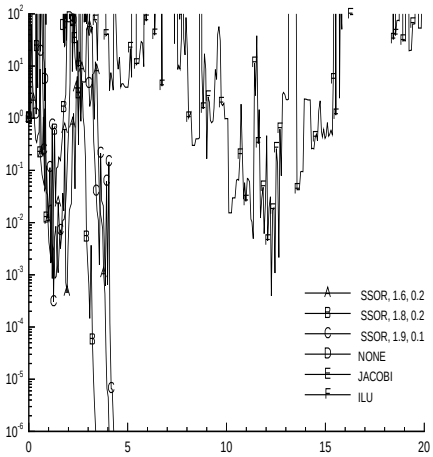
Tabelle 5.2: CPU-Zeit und Iterationen für PGK-Verfahren bei einem Stokes-Problem

Zur Berechnung des Stokes-Problems sind die PGK-Verfahren BICGSTAB und QMRCGSTAB mit SSOR-Vorkonditionierung sehr gut geeignet. Für die Vorkonditionierung muß allerdings der optimale Relaxationsparameter verwendet werden. Die Abbildungen auf den Seiten 62 bis 65 zeigen die Konvergenzverläufe der PGK-Verfahren BICG, BICGSTAB, CGS, QMRCGSTAB, TFQMR und QMR mit verschiedenen Vorkonditionierern bzw. ohne Vorkonditionierung. Auf eine Darstellung der Konvergenzverläufe des CGNR wird an dieser Stelle verzichtet, da der Löser keine guten Konvergenzeigenschaften besitzt. In der Legende sind die Vorkonditionierer angegeben, wobei

die verwendeten Abkürzungen folgende Bedeutung haben

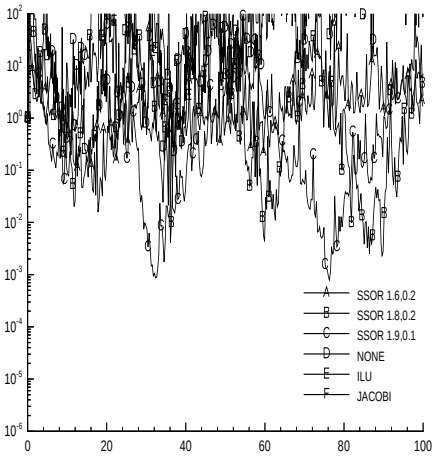
- NONE keine Vorkonditionierung
- JACOBI Verwendung der Jacobi-Vorkonditionierung
- ILU Verwendung der ILU(0)-Vorkonditionierung
- "SSOR, A, B" Verwendung der SSOR-Vorkonditionierung, wobei die Geschwindigkeitskomponenten mit A relaxiert werden und die Druckkomponente mit B relaxiert wird.

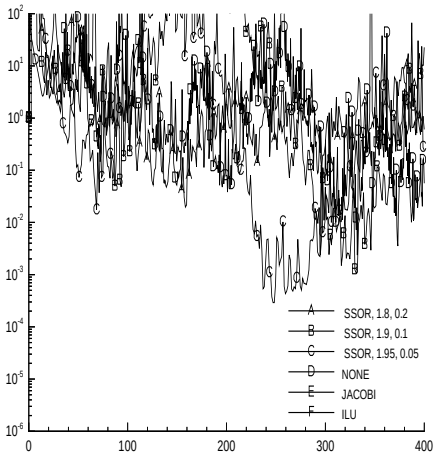
Gerade bei feinen Zerlegungen verschlechtert sich die Konvergenz der Verfahren dramatisch bei nicht optimaler Wahl des Relaxationsparameter. Die Abhängigkeit der SSOR-Vorkonditionierung von der optimalen Wahl des Relaxationsparameters ist auch ein entscheidender Nachteil dieser Vorkonditionierungstechnik. Es ist möglich, die Rechenzeiten mit geeigneten PGK-Verfahren je nach Zerlegung, um bis zu 50% gegenüber dem SSOR-Verfahren mit optimalen Relaxationsparametern als Lösungsverfahren zu senken.

	17×17	33×33
BICG		
BICGSTAB		
CGS		

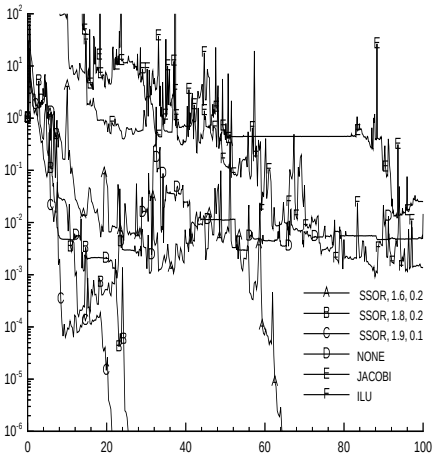
65×65 129×129

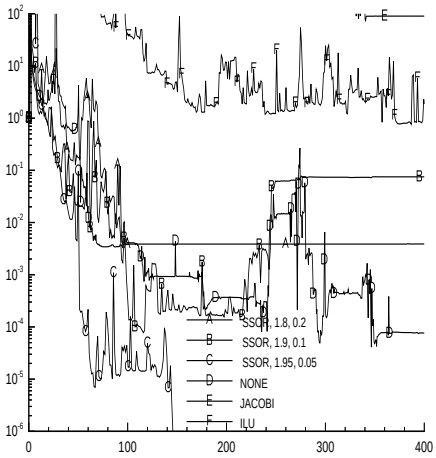
BICG



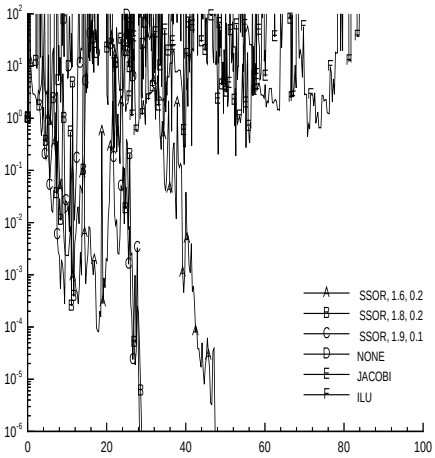


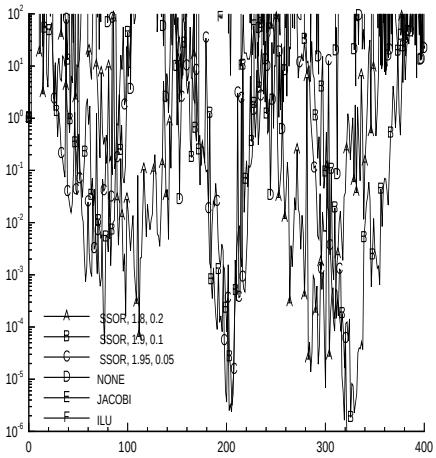
BICGSTAB





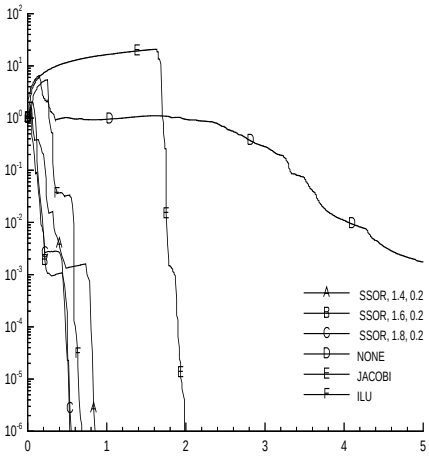
CGS

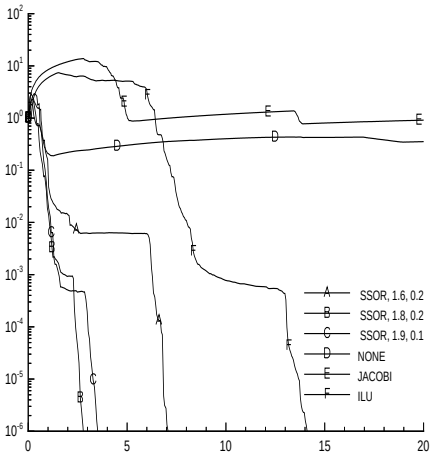




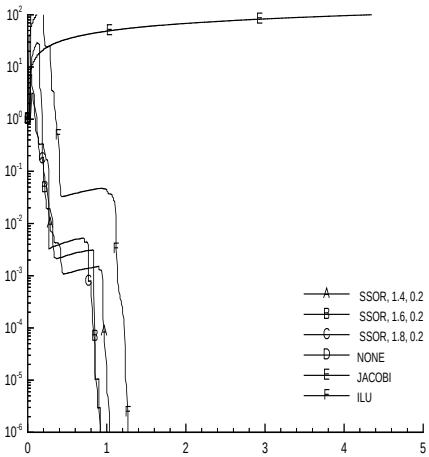
17×17 33×33

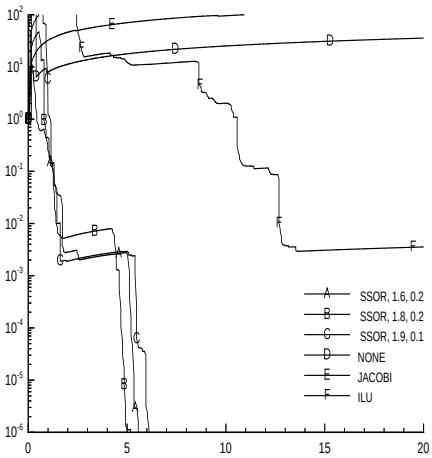
QMRCGSTAB



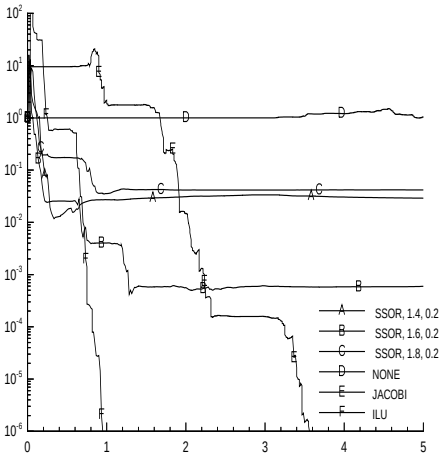


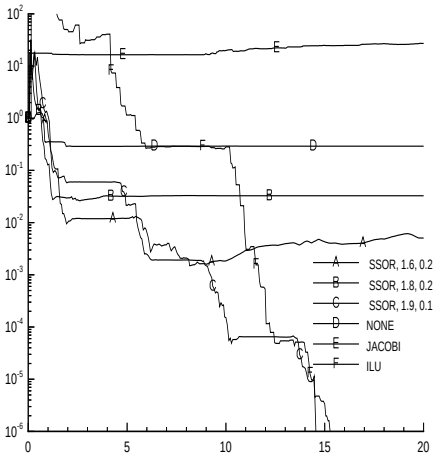
TFQMR





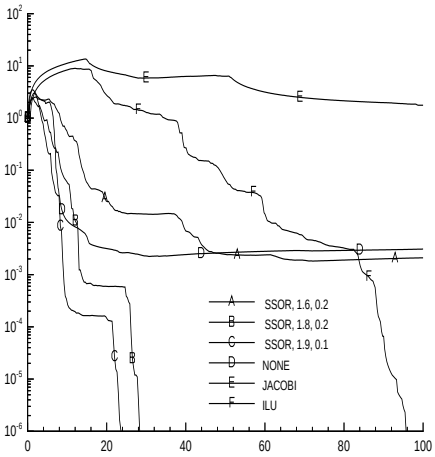
QMR

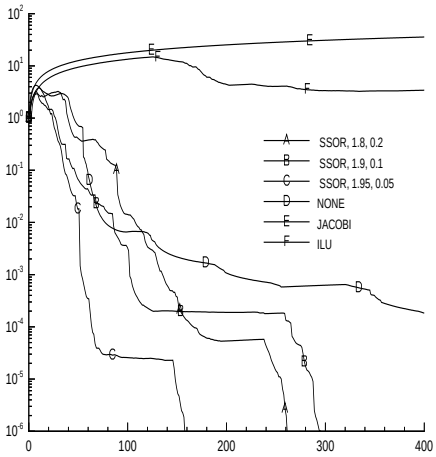




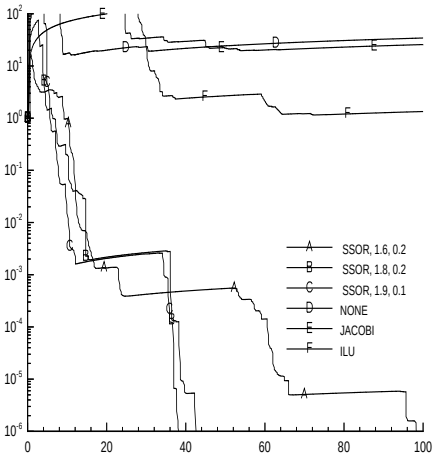
65×65 129×129

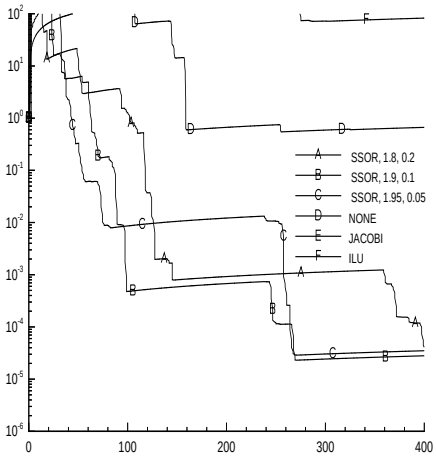
QMRCGSTAB



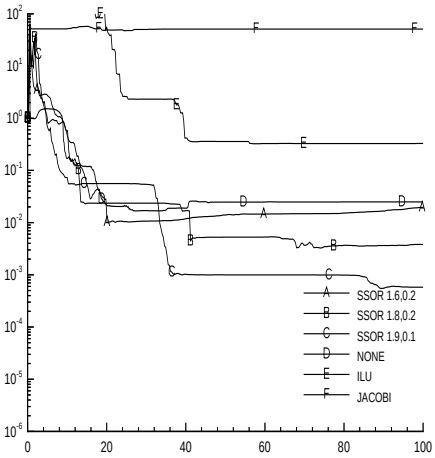


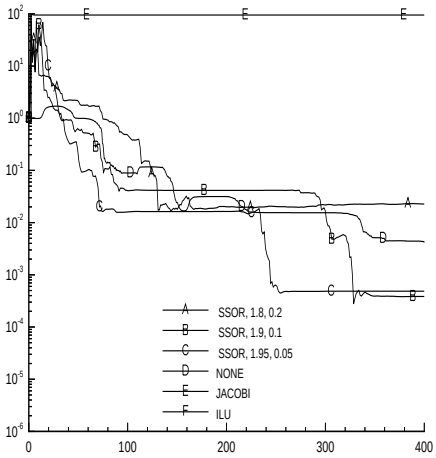
TFQMR





QMR





5.3 Oseen-Gleichung im diffusionsdominanten Fall

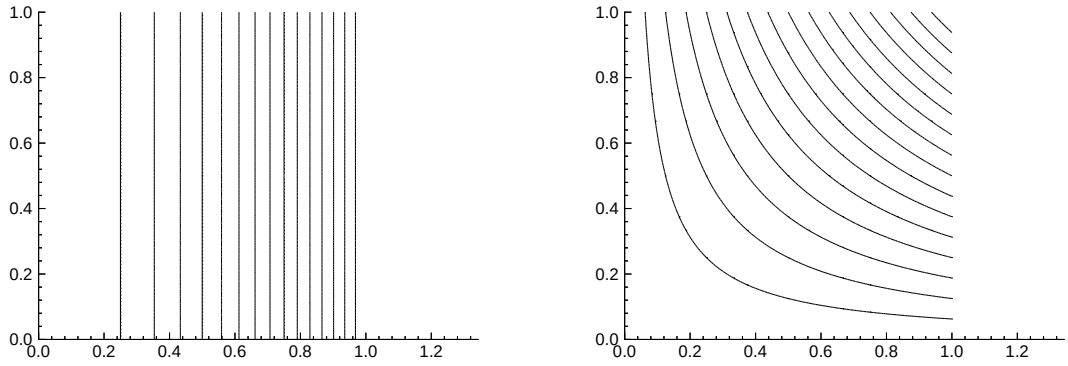
Als zweites Beispiel wurde eine diffusiondominante Oseen-Gleichung gerechnet. Als Strömungsfeld wurde

$$a = \begin{pmatrix} x^2 \\ -2xy \end{pmatrix}$$

verwendet, vgl. Abbildung 5.2. Die Rechnungen sind mit folgender Parameterwahl durchgeführt worden:

- Diffusionskoeffizient: $\nu = 0.5$,
- Stabilisierungsparameter für Geschwindigkeit: $c_1^u = 1.0$,
- Stabilisierungsparameter für Druck: $c_1^p = 1.0$,
- Stabilisierungsparameter für die Divergenzgleichung: $c_2 = 0.0$.

Außerdem wurde mit Konsistenzsicherung gerechnet.



$u_1 = const.$ $u_2 = const.$

Abbildung 5.2: Niveaulinien der Strömungsfeldes

Bei diesem Problem ergaben sich in Abhängigkeit von der Diskretisierungsschrittweite h die in Tabelle 5.3 angegeben optimalen Relaxationsparameter (ω_{opt}) für das SSOR-Verfahren, weiterhin sind die Rechenzeit und die Iterationszahl angegeben, die bis zum Erreichen des Abbruchkriteriums notwendig waren. Die für die Oseen-Gleichung im diffusionsdominanten Fall ermittelten optimalen Relaxationsparameter stimmen in wesentlichen mit den Ergebnissen des Stokes-Problem überein.

Zerlegung	ω_v	ω_p	CPU-sec	Iterationen
17×17	1.6	0.2	0.5666	86
33×33	1.8	0.2	3.5165	146
65×65	1.9	0.1	32.599	288
129×129	1.95	0.05	296.705	573

Tabelle 5.3: Optimale Relaxationsparameter für SSOR beim Oseen-Problem im diffusionsdominanten Fall

Auch bei diesem Problem sind die optimalen Relaxationsparameter von der Feinheit der Zerlegung abhängig. In der Tabelle 5.4 sind für die oben genannten PGK-Verfahren die optimalen Vorkonditionierer und die Rechenzeiten, sowie die Iterationszahlen, die erforderlich waren, damit das Abbruchkriterium $\|r_k\|_2/\|r_0\|_2 < 10^{-6}$ erfüllt wurde, angegeben.
Auch bei diesem Problem können gute Ergebnisse mit den PGK-Verfahren BICGSTAB und QMRCGSTAB mit SSOR-Vorkonditionierung erzielt werden. Dabei treten jedoch die gleichen Probleme wie bei der vorher betrachteten Stokes-Gleichung auf, nämlich die Abhängigkeit der Konvergenz der Verfahren von der Wahl der Relaxationsparameter. Die Konvergenzverläufe der PGK-Verfahren mit verschiedenen Vorkonditionierern sind in den Abbildungen auf den Seiten 70 bis 73 dargestellt. Weiterhin wird in diesen Diagrammen deutlich, daß andere Vorkonditionierer nicht so leistungsfähig wie die SSOR-Vorkonditionierung sind. Die Rechenzeiten von geeigneten PGK-Verfahren sinken um bis zu 2/3 gegenüber dem SSOR-Verfahren mit optimalen Relaxationsparametern als Lösungsverfahren. Bei der 129×129-Zerlegung sind die CPU-Zeiten für die Oseen-Gleichung im diffusionsdominanten Fall zum Teil deutlich besser als beim Stokes-Problem. Die Rechenzeiten verbessern sich teilweise um 2/3 bzw. um die Hälfte.

Zerlegung	Verfahren	optimaler Vorkonditionierer	CPU-sec	Iterationen
17×17	BICG	ILU	0.9832	72
	CGNR		>30	
	QMR	ILU	1.0999	73
	BICGSTAB	SSOR($\omega_{opt(17)}$)	0.3666	20
	CGS	SSOR($\omega_v = 0.6; \omega_p = 1.0$)	0.5999	38
	TFQMR	SSOR($\omega_v = 1.4; \omega_p = 0.4$)	0.8999	87
	QMRCGSTAB	SSOR($\omega_{opt(17)}$)	0.3833	22
33×33	BICG	ILU	10.599	204
	CGNR		>100	
	QMR	ILU	11.199	211
	BICGSTAB	SSOR($\omega_{opt(33)}$)	3.0665	44
	CGS	SSOR($\omega_{opt(33)}$)	4.0831	62
	TFQMR	SSOR($\omega_{opt(33)}$)	5.2497	128
	QMRCGSTAB	SSOR($\omega_{opt(33)}$)	2.9332	47
65×65	BICG		>200	
	CGNR		>200	
	QMR		>200	
	BICGSTAB	SSOR($\omega_{opt(65)}$)	11.549	39
	CGS	SSOR($\omega_{opt(65)}$)	12.016	40
	TFQMR	SSOR($\omega_{opt(65)}$)	41.048	213
	QMRCGSTAB	SSOR($\omega_{opt(65)}$)	11.982	40
129×129	BICG		>400	
	CGNR		>400	
	QMR		>400	
	BICGSTAB	SSOR($\omega_{opt(129)}$)	76.530	59
	CGS	SSOR($\omega_{opt(129)}$)	86.579	68
	TFQMR	SSOR($\omega_{opt(129)}$)	291.954	352
	QMRCGSTAB	SSOR($\omega_{opt(129)}$)	79.096	61

Tabelle 5.4: CPU-Zeit und Iterationen für PGK-Verfahren für Ossen-Gleichung im diffusionsdominanten Fall

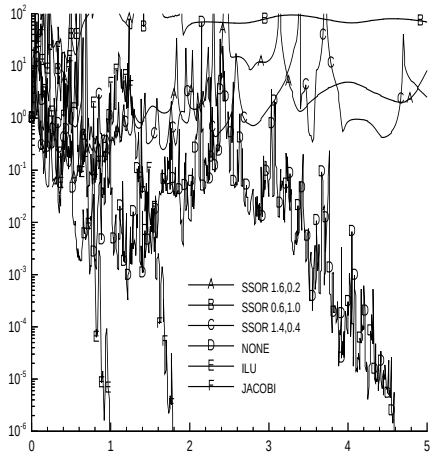
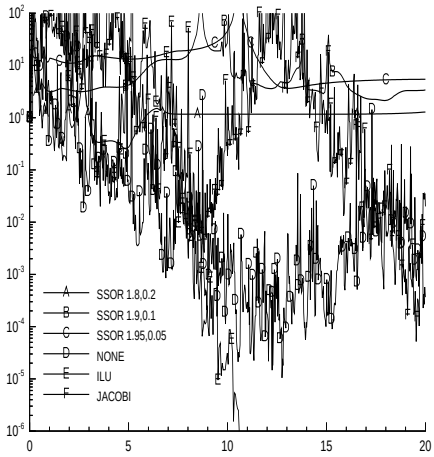
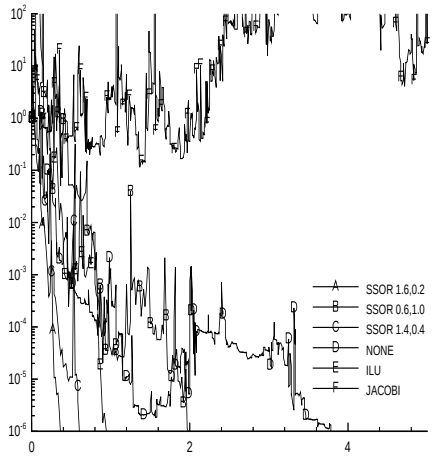
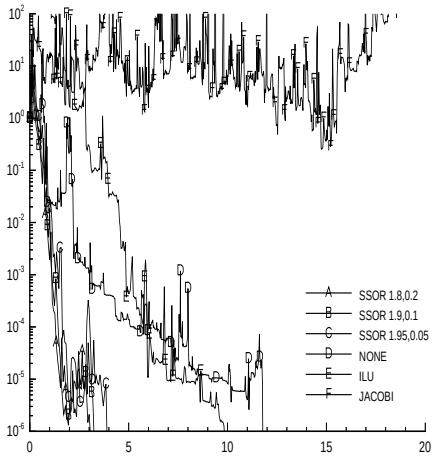
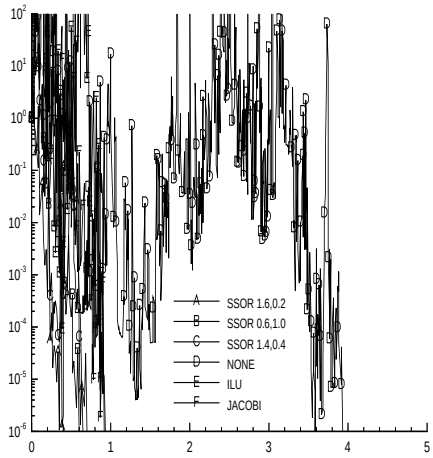
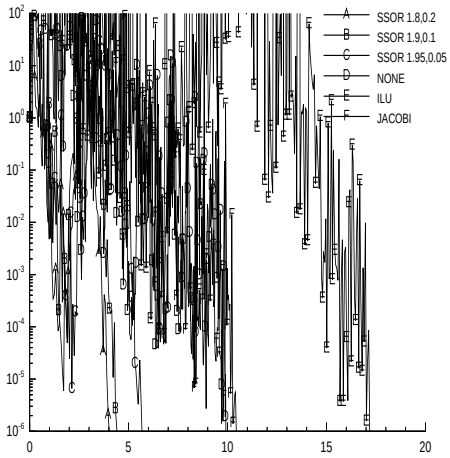
Weiterhin wurde untersucht, wie sich die Rechenzeiten verschlechtern, wenn die optimalen Relaxationsparameter der 129×129-Zerlegung auch für die anderen Zerlegungen benutzt werden. Mit den Relaxationsparametern $\omega_v = 1.95$ und $\omega_p = 0.05$ für die SSOR-Vorkonditionierung von BICGSTAB, CGS, TFQMR und QMRCGSTAB erhält man bei unterschiedlichen Zerlegungen die in Tabelle 5.5 dargestellten Ergebnisse.

Zerlegung	Verfahren	CPU-sec	Iterationen
17×17	BICGSTAB	0.79996	46
	CGS	1.08329	65
	TFQMR	1.48327	131
	QMRCGSTAB	0.8333	48
33×33	BICGSTAB	3.8998	65
	CGS	5.6831	93
	TFQMR	7.6163	192
	QMRCGSTAB	4.9831	72
65×65	BICGSTAB	13.799	45
	CGS	38.498	127
	TFQMR	50.881	262
	QMRCGSTAB	14.699	47

Tabelle 5.5: Rechenzeiten bei SSOR-Vorkonditionierung mit $\omega_v = 1.95$ und $\omega_p = 0.05$

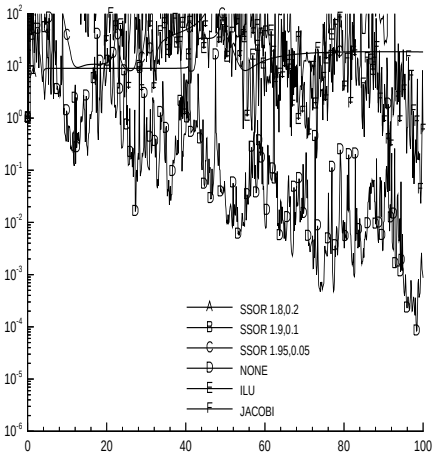
Bei der 17×17-Zerlegung und bei der 33×33-Zerlegung wird die Rechenzeit für die einzelnen Verfahren fast verdoppelt, dies stellt jedoch bei kleinen Rechenzeiten kein Problem dar. Die Rechenzeiten für BICGSTAB und QMRCGSTAB bei einer 65×65-Zerlegung verschlechtern sich nur um 1/6.

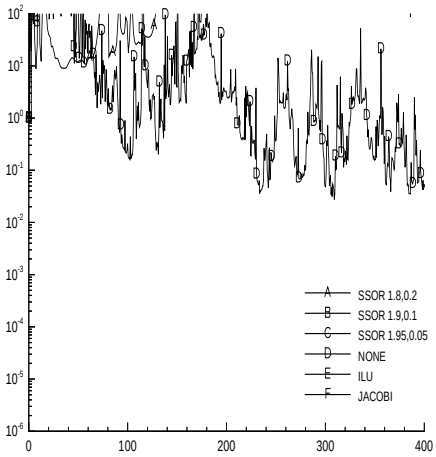
Es ist also möglich, für die Löser BICGSTAB und QMRCGSTAB, die optimalen Relaxationsparameter der 129×129-Zerlegung auch für die Zerlegungen mit geringerer Feinheit zu benutzen. Als Empfehlung für das in diesem Abschnitt betrachtete Problem ergibt sich, die PGK-Verfahren BICGSTAB oder QMRCGSTAB mit SSOR-Vorkonditionierung und den Relaxationsparametern $\omega_v = 1.95$ und $\omega_p = 0.05$ zu verwenden.

	17×17	33×33
BICG		
BICGSTAB		
CGS		

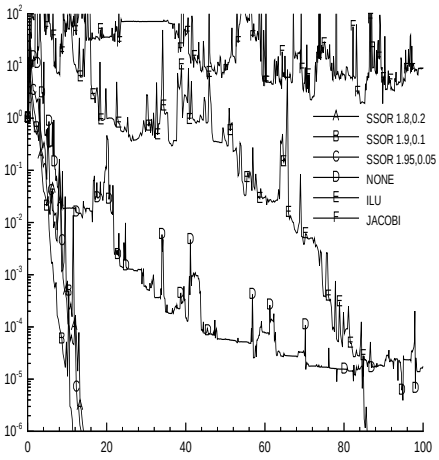
65×65 129×129

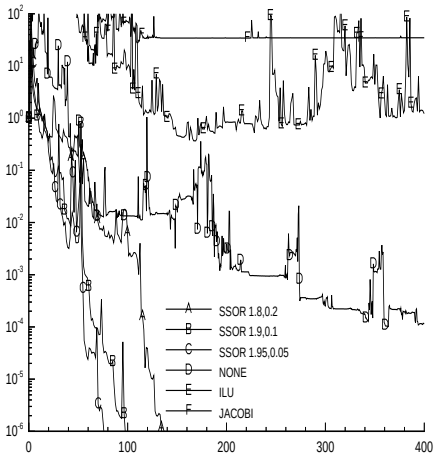
BICG



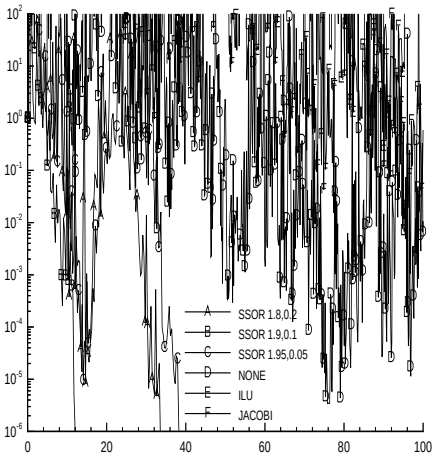


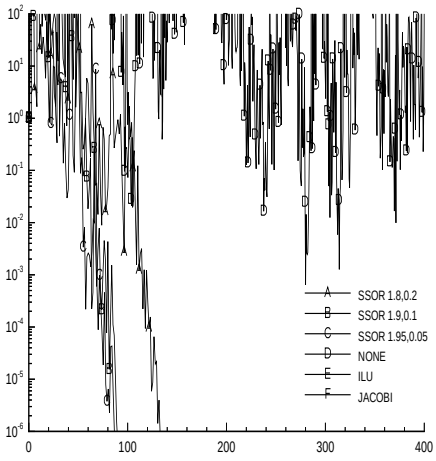
BICGSTAB





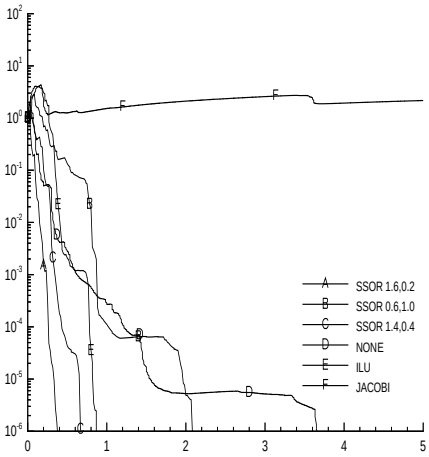
CGS

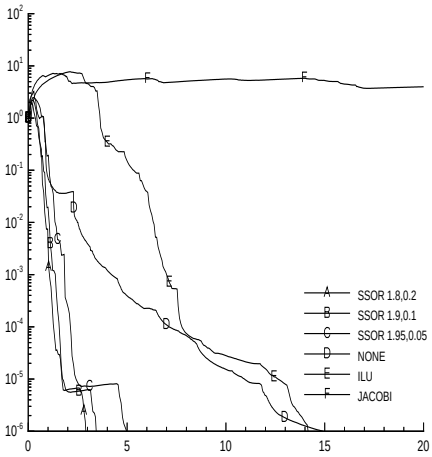




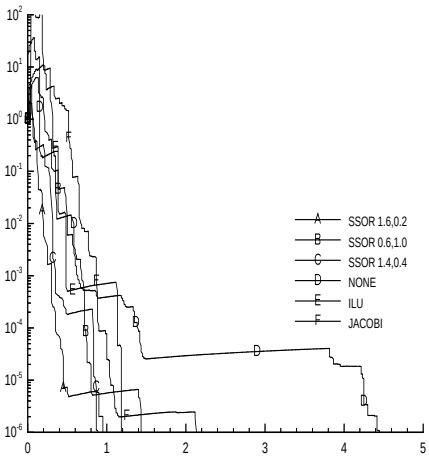
17×17 33×33

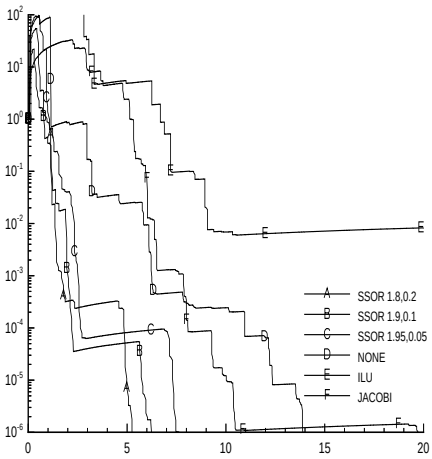
QMRCGSTAB



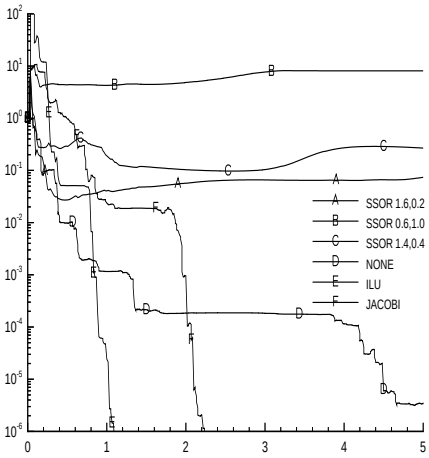


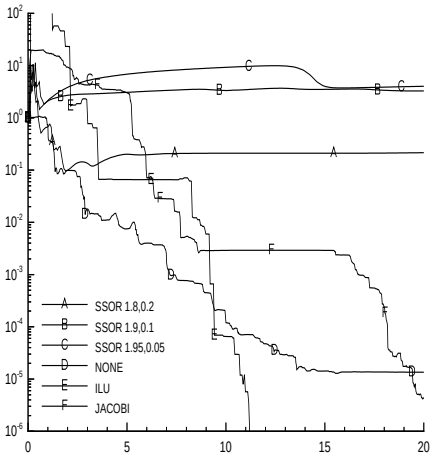
TFQMR





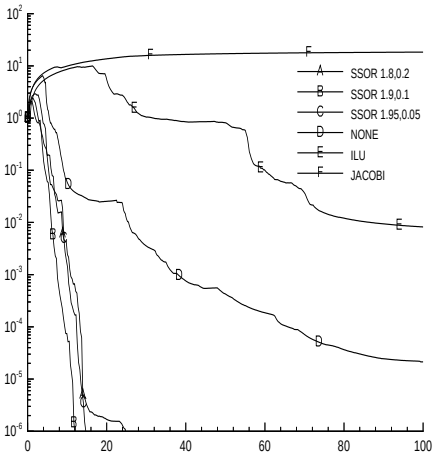
QMR

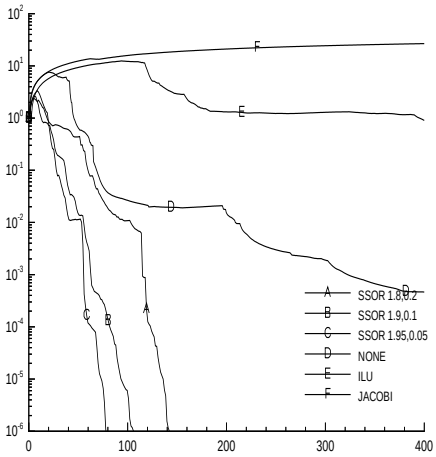




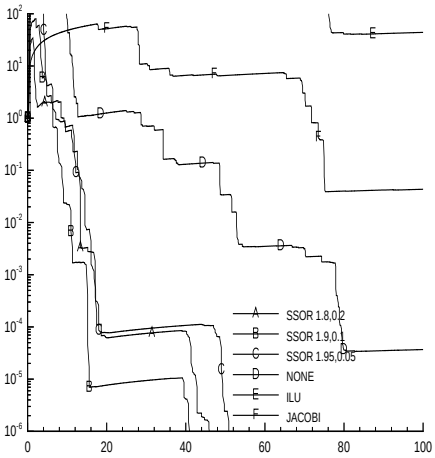
65×65 129×129

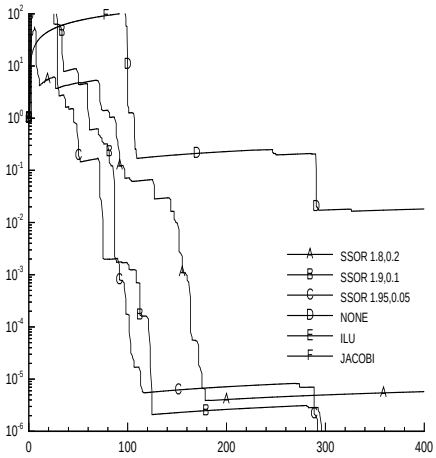
QMRCGSTAB



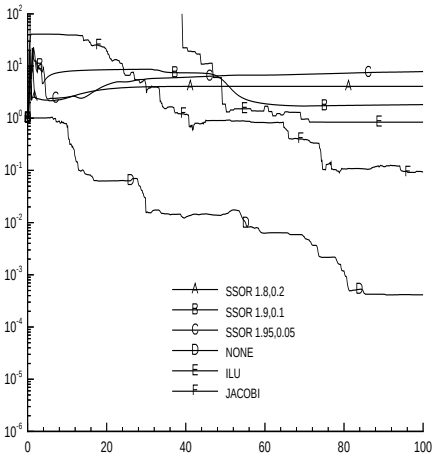


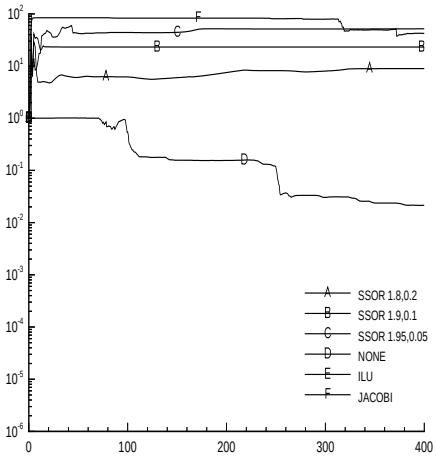
TFQMR





QMR





5.4 Oseen-Gleichung im konvektionsdominanten Fall

Als weiteres Beispiel wurde eine Oseen-Gleichung im konvektionsdominanten Fallmit dem in Abschnitt 5.3 verwendeten Strömungsfeld betrachtet, also

$$a = \begin{pmatrix} x^2 \\ -2xy \end{pmatrix}.$$

Für die Berechnungen sind folgende Parameter verwendet worden:

- Diffusionskoeffizient: $\nu = 0.01$,
- Stabilisierungsparameter für Geschwindigkeit: $c_1^u = 0.1$,
- Stabilisierungsparameter für Druck: $c_1^p = 1.0$,
- Stabilisierungsparameter für die Divergenzgleichung: $c_2 = 1.0$.

Die Berechnungen wurden ohne Konsistenzsicherung durchgeführt.

In der Tabelle 5.6 sind die optimalen Relaxationsparameter des SSOR-Verfahrens in Abhängigkeit von der Diskretisierungsschrittweite für dieses Beispiel angegeben. Bei der Oseen-Gleichung im konvektionsdominanten Fall sind die Rechenzeiten und die Iterationszahlen des SSOR-Verfahrens mit optimalen Relaxationsparametern für alle Zerlegungen besser als im diffusionsdominanten Fall. Dabei ist aber zu berücksichtigen, daß im konvektionsdominanten Fall ohne Konsistenzsicherung gerechnet worden ist.

Zerlegung	ω_v	ω_p	CPU-sec	Iterationen
17×17	1.4	1.6	0.3499	54
33×33	1.6	1.2	1.816	70
65×65	1.8	0.6	15.199	128
129×129	1.9	0.2	222.207	440

Tabelle 5.6: Optimale Relaxationsparameter für SSOR bei Oseen-Problem im konvektionsdominanten Fall

Auch in diesem Fall sind die optimalen Relaxationsparameter von der Feinheit der Zerlegung abhängig. Mit kleiner werdender Diskretisierungsschrittweite h müssen die Geschwindigkeitskomponeten stärker überrelaxiert werden. Mit zunehmender Feinheit der Zerlegung verschiebt sich die Relaxation der Druckkomponente von Überrelaxation zu Unterrelaxation. Die Relaxationsparameter für die Druckkomponente müssen im konvektionsdominanten Fall völlig anders gewählt werden als bei Diffusionsdominanz. Dieser Unterschied ist damit zu erklären, daß die Verwendung der Konsistenzsicherung einen Einfluß auf die Wahl der Relaxationsparameter hat.

In Tabelle 5.7 sind für die getesteten PGK-Verfahren die optimalen Vorkonditionierer angegeben.

Zerlegung	Verfahren	optimaler Vorkonditionierer	CPU-sec	Iterationen
17×17	BICG	ILU	0.5999	45
	CGNR	ILU	1.0166	62
	QMR	ILU	0.6330	44
	BICGSTAB	ILU	0.2333	15
	CGS	ILU	0.2499	18
	TFQMR	ILU	0.6333	76
	QMRCGSTAB	ILU	0.3166	22
33×33	BICG	ILU	4.5331	81
	CGNR	ILU	7.6663	115
	QMR	ILU	4.2664	81
	BICGSTAB	SSOR($\omega_{opt(33)}$)	1.3999	22
	CGS	SSOR($\omega_{opt(33)}$)	1.3666	23
	TFQMR	SSOR($\omega_{opt(33)}$)	2.0832	52
	QMRCGSTAB	SSOR($\omega_{opt(33)}$)	1.5499	26
65×65	BICG	ILU	43.614	177
	CGNR	ILU	95.829	328
	QMR	ILU	48.114	185
	BICGSTAB	SSOR($\omega_{opt(65)}$)	8.9163	32
	CGS	SSOR($\omega_{opt(65)}$)	9.7996	35
	TFQMR	SSOR($\omega_{opt(65)}$)	13.799	74
	QMRCGSTAB	SSOR($\omega_{opt(65)}$)	9.6996	34
129×129	BICG		>400	
	CGNR		>400	
	QMR		>400	
	BICGSTAB	SSOR($\omega_{opt(129)}$)	74.630	60
	CGS	SSOR($\omega_v = 1.8$,$\omega_p = 0.6$)	74.597	60
	TFQMR	SSOR($\omega_v = 1.8$,$\omega_p = 0.6$)	104.579	128
	QMRCGSTAB	SSOR($\omega_{opt(129)}$)	90.879	70

Tabelle 5.7: CPU-Zeit und Iterationszahl für PGK-Verfahren am Beispiel einer Oseen-Gleichung im konvektionsdominanten Fall

Bei diesem Problem sind die PGK-Verfahren BICGSTAB, QMRCGSTAB und CGS mit SSOR-Vorkonditionierung gute Löser. Es tritt aber wiederum das Problem der optimalen Wahl der Relaxationsparameter für die SSOR-Vorkonditionierung auf. Die Löser CGS und TFQMR konvergieren bei einer 129×129-Zerlegung schneller, wenn für die SSOR-Vorkonditionierung nicht die optimalen Relaxationsparameter, sondern

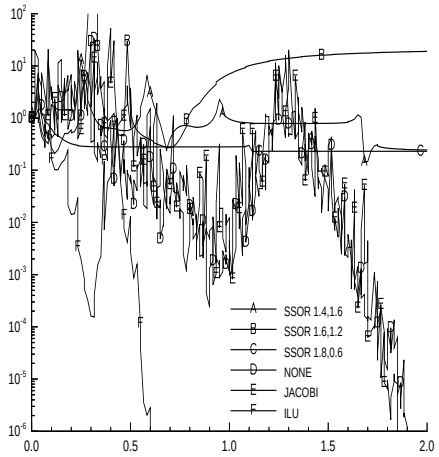
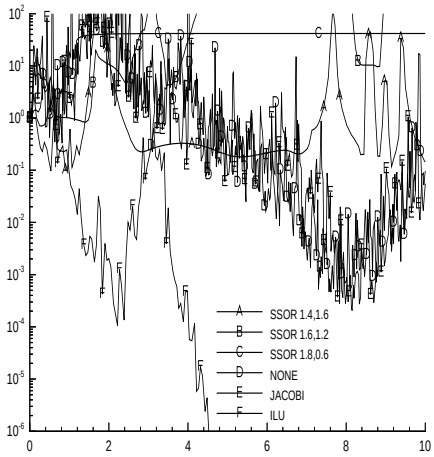
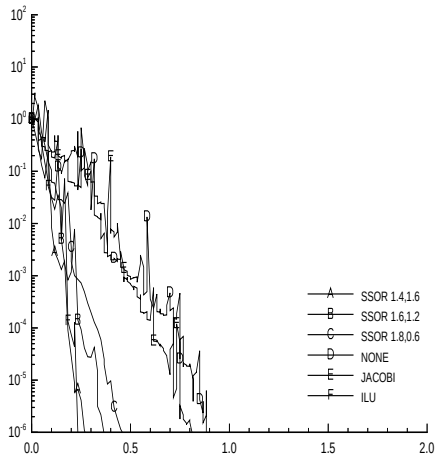
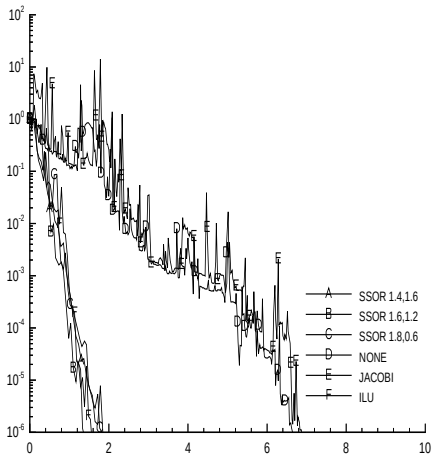
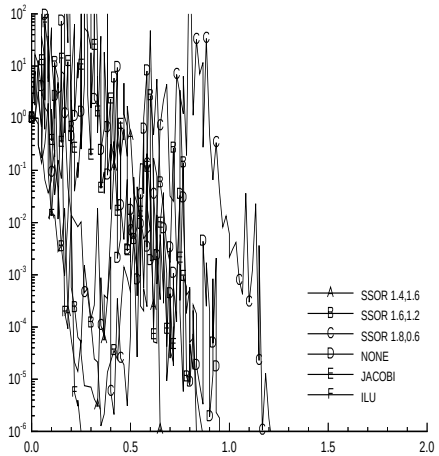
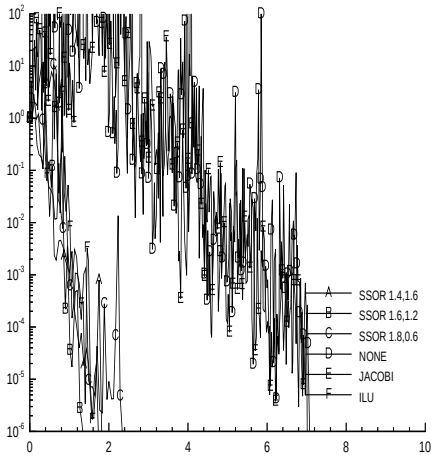
$\omega_v = 1.8$ und $\omega_p = 0.6$ verwendet werden. Bei der 17×17-Zerlegung konvergieren alle betrachteten Verfahren mit der ILU-Vorkonditionierung am schnellsten. Die dargestellten Zusammenhänge sind in den Konvergenzdiagrammen für die betrachteten Löser auf den Seiten 78 bis 81 erkennbar.
Weiterhin wurde untersucht, wie sich die Rechenzeiten der Verfahren verschlechtern, wenn für die SSOR-Vorkonditionierung die Relaxationsparameter der feinsten Zerlegung benutzt werden. Für die 129×129-Zerlegung sind außerdem die Rechenzeiten für die Löser CGS und TFQMR bei Verwendung der SSOR-Vorkonditionierung mit $\omega_v = 1.9$ und $\omega_p = 0.2$ ermittelt worden. Die Ergebnisse sind in Tabelle 5.8 zusammengestellt.

Zerlegung	Verfahren	CPU-sec	Iterationen
17×17	BICGSTAB	0.6666	41
	CGS	0.7166	44
	TFQMR	2.4499	231
	QMRCGSTAB	0.8832	54
33×33	BICGSTAB	3.3831	50
	CGS	3.8831	54
	TFQMR	5.3664	118
	QMRCGSTAB	3.6831	54
65×65	BICGSTAB	17.3826	57
	CGS	17.6492	63
	TFQMR	23.8323	128
	QMRCGSTAB	17.2159	59
129×129	CGS	93.1129	75
	TFQMR	131.8947	162

Tabelle 5.8: Rechenzeit für PGK-Verfahren mit SSOR-Vorkonditionierung mit Relaxationsparametern $\omega_v = 1.9$ und $\omega_p = 0.2$

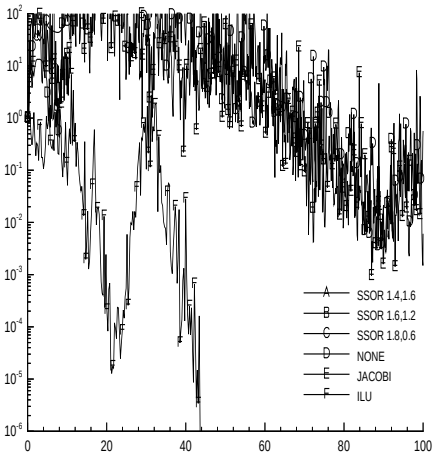
Die Rechenzeiten verschlechtern sich um ca. 50 Prozent bei der 33×33- und der 65×65-Zerlegung. Bei diesem Problem ist es nicht empfehlenswert, für alle Zerlegungen die SSOR-Vorkonditionierung mit den Relaxationsparametern $\omega_v = 1.9$ und $\omega_p = 0.2$ durchzuführen. Es werden aber bei allen Zerlegungen gute Ergebnisse mit dem CGS-Verfahren mit SSOR-Vorkonditionierung unter Verwendung der Relaxationsparameter $\omega_v = 1.8$ und $\omega_p = 0.6$ erzielt. Für die einzelnen Zerlegungen wird dies in den Konvergenzdiagrammen deutlich. Zusammenfassend kann man feststellen, daß bei allen betrachteten Gleichungen das Problem der optimalen Wahl der Relaxationsparameter bei SSOR- Vorkonditionierung auftritt. Die Kombination von PGK-Verfahren mit anderen Vorkonditionierern als SSOR liefert in den meisten Fällen keine guten Ergebnisse, insbesondere ist BICGSTAB in Kombination mit ILU-Vorkonditionierung bei feinen

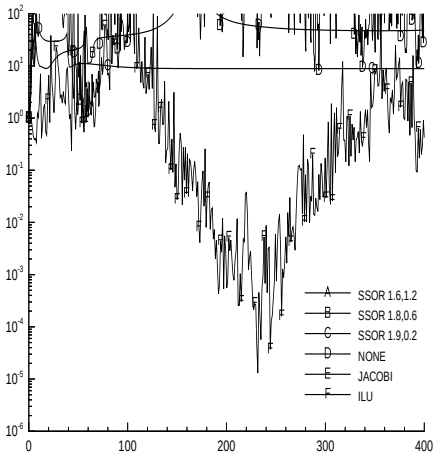
Zerlegungen nicht empfehlenswert. Aus den Konvergenzdiagrammen für die 65×65- bzw. 129×129-Zerlegung wird deutlich, daß sich bei Verwendung der Lösers BICG-STAB mit ILU als Vorkonditionierer die Rechenzeiten verfünffachen. Folglich gibt es bei den betrachteten Lösern und Vorkonditionierern keine brauchbare Alternative zur SSOR-Vorkonditionierung, gerade bei feinen Zerlegungen.

	17×17	33×33
BICG		
BICGSTAB		
CGS		

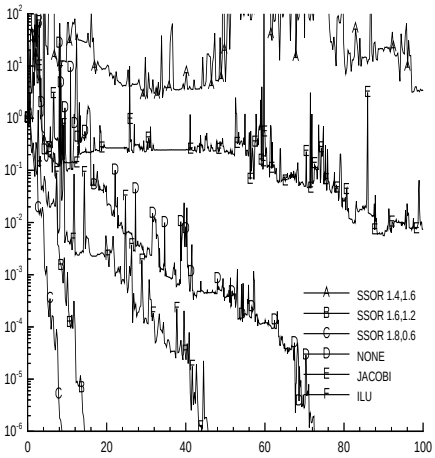
65×65 129×129

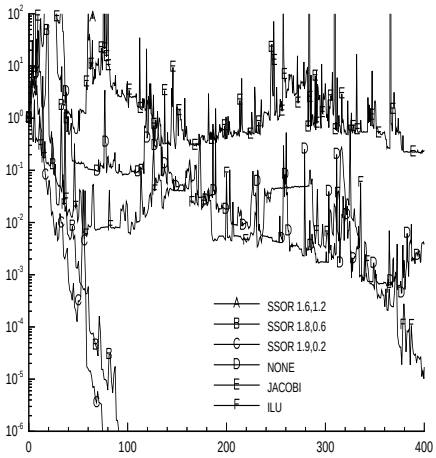
BICG



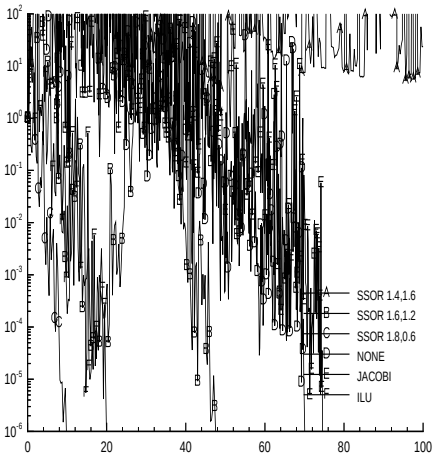


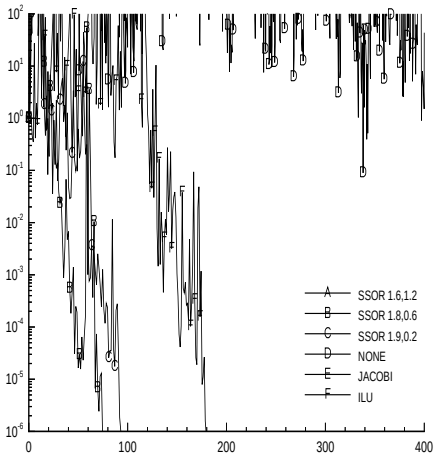
BICGSTAB





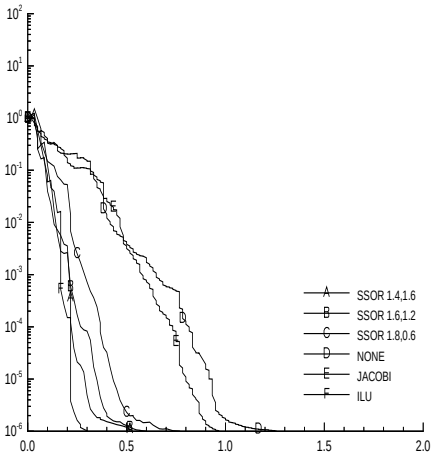
CGS

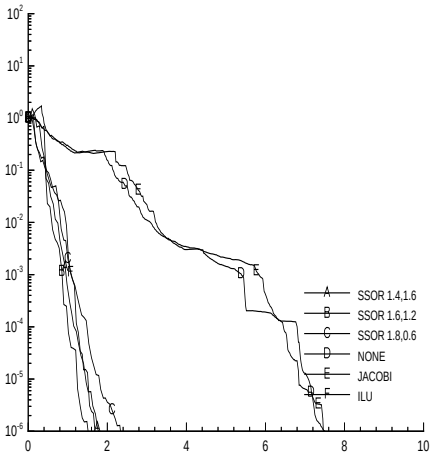




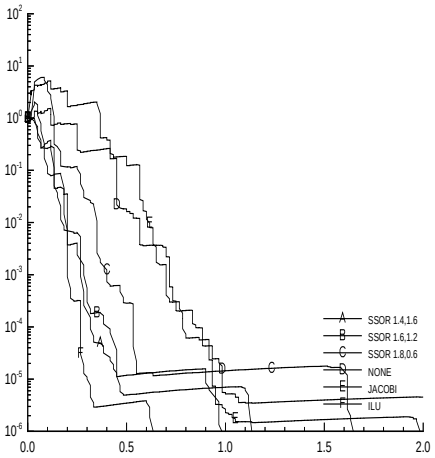
17×17 33×33

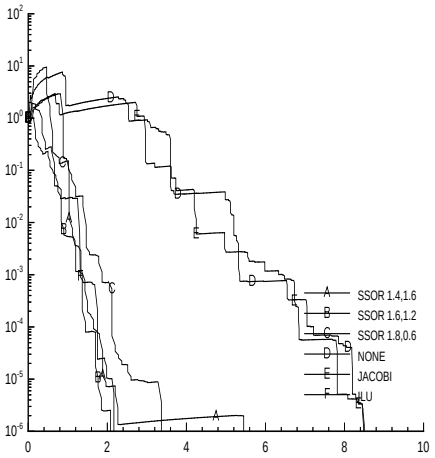
QMRCGSTAB



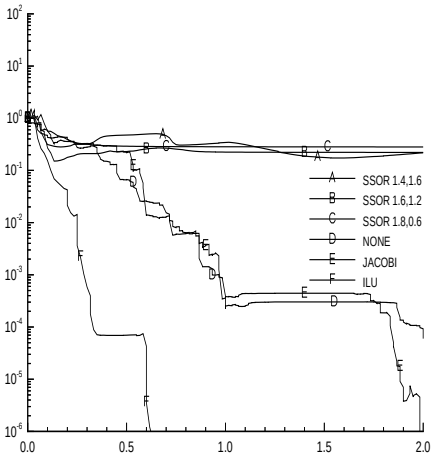


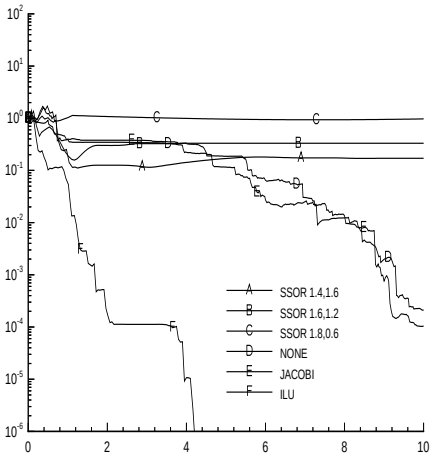
TFQMR





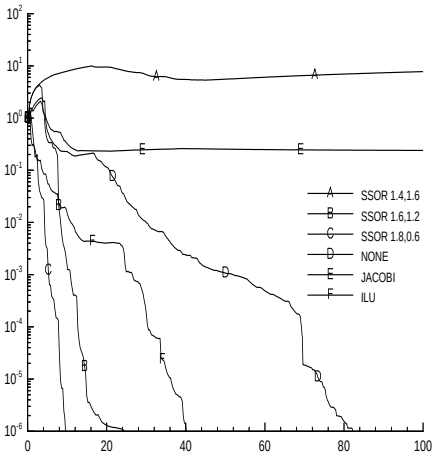
QMR

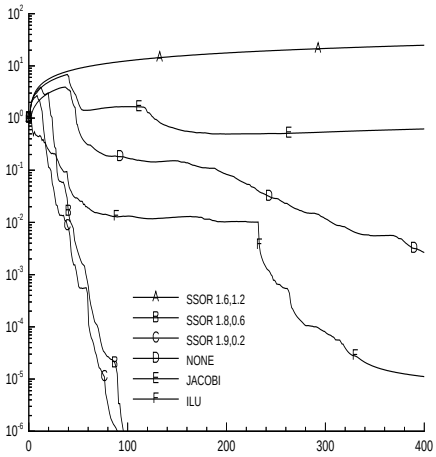




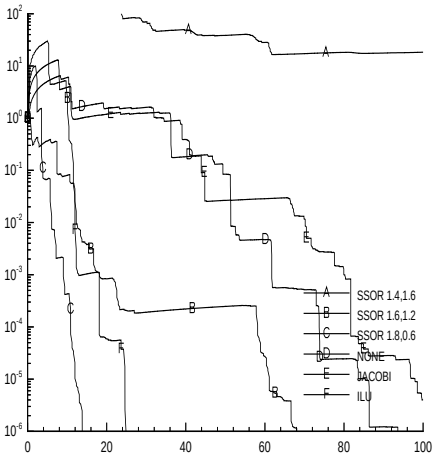
65×65 129×129

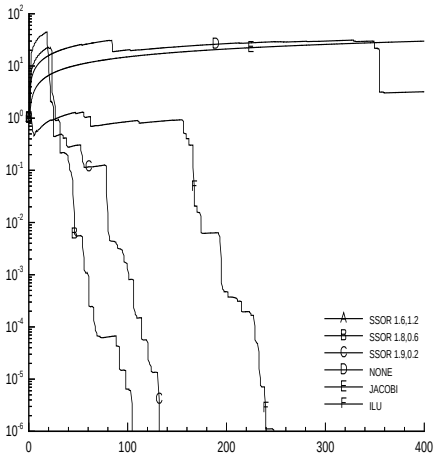
QMRCGSTAB



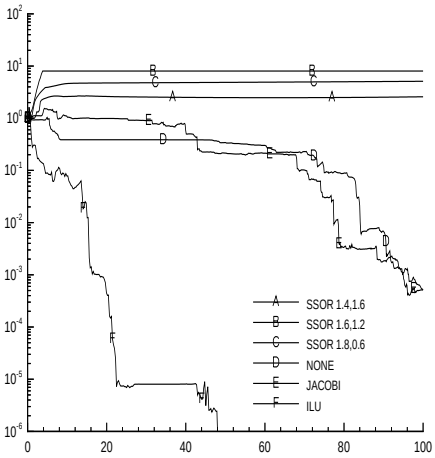


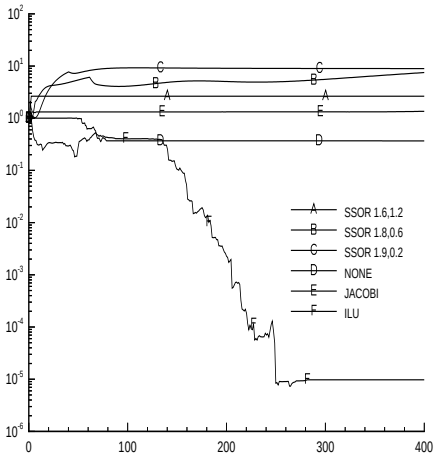
TFQMR





QMR





5.5 Vergleich der Leistungsfähigkeit von problembezogenen Vorkonditionierern und Vorkonditionierern aus BLANC

Für die in Kapitel 4 vorgestellten Vorkonditionierer bei stabilen Elementen wurden in [ES96] einige numerische Ergebnisse vorgestellt. Um im Vergleich dazu die Leistungsfähigkeit der in BLANC zur Verfügung gestellten Vorkonditionierer zu testen, wurde das in diesem Artikel betrachtete Problem gerechnet. Die Generierung der Matrix erfolgte mit ParallelNS, d.h. im Unterschied zu den Rechnungen [ES96] sind stabilisierte Verfahren verwendet worden. Es wurde ein zweidimensionales lid-driven cavity Problem auf dem Einheitsquadrat $\Omega = (0,1) \times (0,1)$ mit den in Abbildung 5.3 gegebenen Randbedingungen betrachtet.

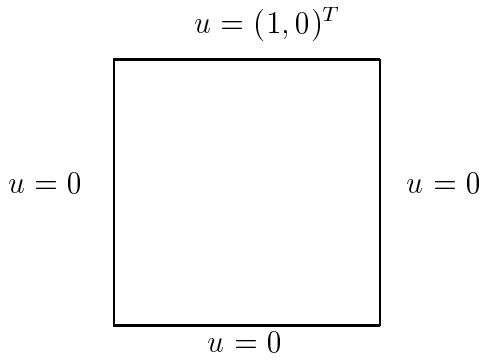


Abbildung 5.3: Vorgabe der Randbedingungen

Das verwendete Strömungsfeld

$$a = \begin{pmatrix} 8(2y-1)(-x^2+x) \\ -8(2x-1)(-y^2+y) \end{pmatrix}$$

ist in Abbildung 5.4 dargestellt.

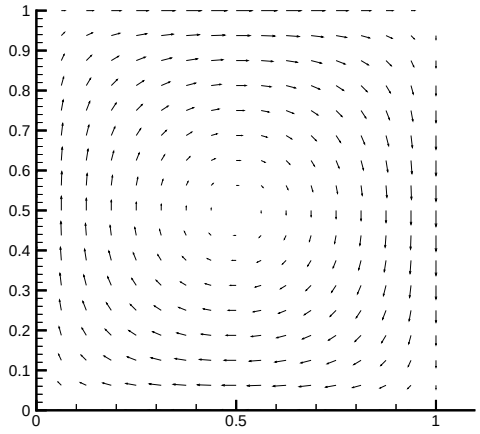


Abbildung 5.4: Strömungsfeld

In die Untersuchungen wurden ein diffusionsdominanter Fall ($\nu = 1$) und ein konvektiondominater Fall ($\nu = 1/100$) einbezogen. Die Stabilisierungsparameter für ParallelNS sind im Fall $\nu = 1$ wie in Kapitel 5.3 und im Fall $\nu = 1/100$ wie im Kapitel 5.4 gewählt worden.

In [ES96] betrachtete man einerseits eine Diskretisierung mittels der Galerkin Methode und andererseits die streamline-upwind Diskretisierung. In allen Fällen wurden die Iterationszahlen verglichen, die erforderlich waren, um das Residuum um 10^{-6} zu reduzieren. In diesem Artikel sind bei Verwendung des Blockdreiecksvorkonditionierers bzw. des Blockdiagonalvorkonditioniers in Kombination mit dem iterativen Löser QMR die in den Tabellen 5.9 und 5.10 enthaltenen Iterationszahlen angegeben worden. Dabei beziehen sich die in Klammern angegebenen Iterationszahlen auf die Blockdiagonalvorkonditionierung.

Zerlegung	17×17	33×33	65×65	129×129
$\nu = 1$	22(43)	22(43)	22(41)	16
$\nu = 1/100$	73(143)	126(246)	189(375)	253

Tabelle 5.9: Iterationszahlen für QMR bei Galerkin-Diskretisierung

Zerlegung	17×17	33×33	65×65	129×129
$\nu = 1$	25(49)	27(51)	25(47)	22
$\nu = 1/100$	76(157)	133(249)	190(382)	229

Tabelle 5.10: Iterationszahlen für QMR bei Streamline-upwind-Diskretisierung

Das obige Beispiel wurde mit den iterativen Löser QMR in Kombination mit den in BLANC zur Verfügung stehenden Vorkonditionierern Jacobi, SSOR und ILU(0) gerechnet. Die besten Ergebnisse hinsichtlich der benötigten Iterationszahlen erzielt man bei Verwendung der ILU(0)-Vorkonditionierung. Die Resultate sind in Tabelle 5.11 zusammengestellt.

Zerlegung	17×17	33×33	65×65	129×129
$\nu = 1$	69	386	>1000	>1000
$\nu = 1/100$	50	105	206	>2000

Tabelle 5.11: Iterationszahlen für QMR mit ILU-Vorkonditionierung

Im diffusionsdominanten Fall ($\nu = 1$) sind die Ergebnisse ab der 65×65-Zerlegung bei Verwendung des QMR-Verfahren mit den in BLANC zür Verfügung gestellten Vorkonditionierern völlig unbefriedigend. Selbst mit 1000 Iterationen kann das Residuum nicht um 10^{-6} reduziert werden. Im konvektionsdominanten Fall sind die Ergebnisse der ILU(0)-Vorkonditionierung im Vergleich zu den Ergebnissen in der Tabellen 5.9 und 5.10 bis zur 65×65-Zerlegung konkurrenzfähig. Problematisch wird auch bei diesem Fall die Zerlegung 129×129. Selbst mit 2000 Iterationen ist keine Reduktion des Residuums um 10^{-6} möglich. Sowohl im diffusionsdominanten als auch im konvektionsdominanten Fall wird deutlich, daß sich die Iterationszahl mindestens verdoppelt, wenn die Diskretisierungsschrittweite halbiert wird. Bei den in [ES96] vorgestellten numerischen Ergebnissen tritt dieses Probelm zumindest im diffusionsdominanten Fall nicht auf.

In den Abschnitten 5.3 und 5.4 wird deutlich, daß mit anderen in BLANC zur Verfügung stehenden Iterationsverfahrender bessere Ergebnisse als mit dem iterativen Löser QMR erzielt werden können. Aus diesem Grund wurde das obige Beispiel mit den Lösern QMRCGSTAB, CGS, TFQMR und BICGSTAB mit verschiedenen Vorkonditionierern berechnet. Dazu war es zuerst erforderlich, die optimalen Relaxationsparameter für das SSOR-Verfahren zu ermitteln. Für den diffusionsdominanten Fall ($\nu = 1$) und den konvektionsdominanten Fall ($\nu = 1/100$) ergaben sich in Abhängigkeit von der Feinheit der Zerlegung die in Tabelle 5.12 angebenen optimalen Relaxationsparameter.

	$\nu = 1$		$\nu = 1/100$	
Zerlegung	ω_v	ω_p	ω_v	ω_p
17×17	1.4	0.2	1.4	1.6
33×33	1.8	0.2	1.4	1.4
65×65	1.9	0.1	1.4	1.2
129×129	1.95	0.05	1.6	0.8

Tabelle 5.12: Optimale Relaxationsparameter für SSOR

Für den diffusionsdominanten Fall sind die besten Ergebnisse mit dem iterativen Löser BICGSTAB mit SSOR-Vorkonditionierung erzielt worden, vgl. Tabelle 5.13.

Zerlegung	Verfahren	optimaler Vorkonditionierer	CPU-sec	Iterationen
17×17	BICGSTAB	SSOR($\omega_v = 1.8; \omega_p = 0.2$)	0.5000	33
33×33	BICGSTAB	SSOR($\omega_{opt(33)}$)	2.6998	45
65×65	BICGSTAB	SSOR($\omega_{opt(65)}$)	19.8825	70
129×129	BICGSTAB	SSOR($\omega_{opt(129)}$)	142.6776	115

Tabelle 5.13: Iterationszahlen und CPU-Zeit für BICGSTAB im diffusionsdominanten Fall

Selbst bei Verwendung von sehr guten Lösern und Vorkonditionierern ist es im diffusionsdominanten Fall nicht möglich, Iterationszahlen zu erreichen, die den Ergebnissen bei Verwendung von Blockdiagonal- bzw. Blockdreiecksvorkonditionierung entsprechen. Aus der obigen Tabelle wird deutlich, daß das größte Problem die Abhängigkeit der Iterationszahl von der Diskretisierungsschrittweite h ist.
Auch im konvektionsdominaten Fall ($\nu = 1/100$) erhielt man mit dem iterative Löser BICGSTAB in Kombination mit der SSOR-Vorkonditionierung die besten Ergebnisse, die in Tabelle 5.14 zusammengestellt sind.

Zerlegung	Verfahren	optimaler Vorkonditionierer	CPU-sec	Iterationen
17×17	BICGSTAB	SSOR($\omega_{opt(17)}$)	0.3666	23
33×33	BICGSTAB	SSOR($\omega_{opt(33)}$)	2.3499	42
65×65	BICGSTAB	SSOR($\omega_{opt(65)}$)	27.2822	99
129×129	BICGSTAB	SSOR($\omega_{opt(129)}$)	175.193	144

Tabelle 5.14: Iterationszahlen und CPU-Zeit für BICGSTAB im konvektionsdominanten Fall

Die mit dieser Löser-Vorkonditionierer-Kombination im konvektionsdominanten Fall benötigten Iterationszahlen sind 50 Prozent besser als die Ergebnisse, die mit den Blockdreiecks- bzw. Blockdiagonalvorkonditionierern erzielt worden sind. Abschließend kann man feststellen, daß im Fall von Konvektionsdominanz die in BLANC zur Verfügung gestellten Löser und Vorkonditionierer im Vergleich zu [ES96] konkurrenzfähige Ergebnisse liefern.

5.6 Fazit der numerischen Experimente

In den Abschnitten 5.3 bis 5.5 sind verschiedene Beispiele betrachtet worden. Dabei wurde deutlich, daß die Löser BICGSTAB, CGS und QMRCGSTAB in Kombination mit der SSOR-Vorkonditionierung bei geeigneter Wahl der Relaxationsparameter gute Ergebnisse hinsichtlich der benötigten Rechenzeit liefern. Da bei Verwendung der ILU-Vorkonditionierung in den meisten Fällen keine akzeptablen Ergebnisse erzielt werden konnten, eignet sich die Verwendung dieser Vorkonditionierungsstrategie nicht. Ein Problem der SSOR-Vorkonditionierung stellt allerdings die Wahl der Relaxationsparameter dar. An den Beispielen, besonders im konvektionsdominanten Fall, sieht man deutlich, daß die optimalen Parameter vom Problem abhängig sind. Außerdem kann man an den Experimenten erkennen, daß die Rechenzeiten bzw. die Iterationszahlen von h abhängig sind. Wünschenswert wären Ergebnisse, wie sie im diffusionsdominanten Fall in [ES96] erzielt worden sind.

Danksagung

Zum Schluß dieser Arbeit bedanke ich mich bei allen, die mich während ihrer Fertigstellung unterschützt haben.
Besonders bedanke ich mich bei Herrn Prof. G. Lube für die intensive und motivierende Betreuung meiner Arbeit.
Mein Dank gilt auch Frank-Christian Otto, der mir bei Fragen und Problemen stets hilfreich zur Seite stand.
Für die tatkräftige Unterstützung bei allen Computerproblemen bedanke ich mich bei Ralph Hangleiter.
Ganz herzlich danke ich meinen Eltern, die mir das Studium ermöglicht haben und mich während der gesamten Studienzeit unterstützt haben.
Mein ganz besonderer Dank gilt meinem Lars, der während meines Studiums viele Höhen aber auch Tiefen hautnah miterlebt hat. Eine große Hilfe waren auch die fachlichen Diskussionen, die wir häufig geführt haben.

Literaturverzeichnis

[Alt92] Alt, H.W.: *Lineare Funktionalanalysis*. 3. Auflage. Berlin : Springer, 1992

[Bra92] Braess, D.: *Finite Elemente*. Berlin : Springer, 1992

[DH94] Droux, L.P. ; Hughes, T.J. R.: A boundary integral modification of the Galerkin least squares formulation of the Stokes problem. **In:** *Comp. methods in appl. mechanics and enginering* 113 (1994), S. 173–182

[ES86] Elman, H. C. ; Schultz, M. H.: Preconditioning by fast direct methods for nonselfadjoint nonseparable elliptic problems. **In:** *SIAM J. Numer. Anal* (1986), Nr. 23, S. 44–57

[ES96] Elman, H. ; Silvester, D.: Fast nonsymmetric iterations and preconditioning for Navier-Stokes equations. **In:** *SIAM J. Sci. Comput.* 17 (1996), S. 33–46

[GL89] Golub, G.H. ; Loan, C.F. van: *Matrix Computations*. second. Baltimore : The John Hopkins University Press, 1989

[GR86] Girault, V. ; Raviart, P.-A.: *Finite Elemente for Navier-Stokes Equations*. Berlin : Springer, 1986

[GR94] Grossmann, Ch. ; Roos, H.-G.: *Numerik partieller Differentialgleichungen*. 2. Stuttgart : Teubner, 1994

[Gri81] Griffel, D.H.: *Applied functional analysis*. Chicester : Ellis Horwood Limited, 1981

[KS97] Klawonn, A. ; Starke, G.: Block triangular preconditioners for Nonsymmetric Saddle Point Problems: Field-of-Values Analysis / Universität Münster. 1997 (4/97-N). – Forschungsbericht

[Lub94] Lube, G.: Stabilized galerkin finite element methods for convection dominated and incompressible flow problems. **In:** *Numerical analysis and mathematical modelling* 29 (1994), S. 85–104

[Mü97] Müller, L.: *Untersuchung einer stabilisierten Finite-Elemente-Methode für die Oseen-Gleichungen*, Universität Göttingen, Diplomarbeit, 1997

[Pri96] PRIESNITZ, A.P.: *Untersuchung iterativer Lösungsverfahren am Beispiel diskretisierter Konvektions-Diffusions-Reaktions-Gleichungen*, Universität Göttingen, Diplomarbeit, 1996

[Saa96] SAAD, Y.: *Iterative methods for sparse linear systems*. Boston : PWS Publishing Company, 1996

[Zei90] ZEIDLER, E.: *Nonlinear functional analysis and its applications II/A (Linear monotone operators)*. Springer, 1990