WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER
Mathematisches Institut

Numerische Analyse und Simulation des Blutflusses durch künstliche Herzklappen

Sabine Aust

Münster im März 2000

Inhaltsverzeichnis

Abbildungsverzeichnis

Einleitung

In den letzten Jahrzehnten haben sich sowohl Hardware als auch numerische Algorithmen so stark weiterentwickelt, daß es Wissenschaftlern und Ingenieuren verschiedenster Fachrichtungen möglich ist, komplexe Vorgänge am Computer zu simulieren. Selbst in den klassischen Naturwissenschaften wie Biologie und Medizin sind numerische Simulationen zu einem essentiellen Werkzeug der Forschung und Weiterentwicklung geworden.

In dieser Arbeit soll nun speziell das medizintechnische Problem der künstlichen Herzklappen betrachtet werden. Zur genauen Analyse und Simulation des Blutflusses durch eine künstliche Herzklappe bedarf es einiger biologischer, physikalischer und mathematischer Erläuterungen:

Das 1. Kapitel befaßt sich ausschließlich mit den biologischen Gegebenheiten des menschlichen Herzens, speziell der Aortenklappe, während das 2. Kapitel dann die unterschiedlichen Formen möglicher Klappenprothesen beschreibt und deren Vor- und Nachteile vergleicht. Die medizinischen bzw. biologischen Fachbegriffe, die hierzu nötig sind, werden im Text durch *kursive* Schreibweise gekennzeichnet und im Anhang genauer erläutert.

Kapitel 3 und 4 schaffen im Anschluß daran die physikalischen Grundlagen, die zum Verständnis der Klappenbewegungen notwendig sind. Es werden die Ursachen der wirbeligen oder sogar turbulenten Blutströmung skizziert und die beschreibenden mathematischen Gleichungen, die **Navier-Stokes Gleichungen** hergeleitet.

Die komplexe mathematische Theorie der Variationsgleichungen und Funktionenräume, die allgemein hinter partiellen Differentialgleichungen dieser Gestalt steht, wird in Kapitel 5 vorbereitet, so daß in Kapitel 6 auf Lösungsansätze in geeigneten Funktionenräumen eingegangen werden kann. Im Hinblick auf die spätere Implementierung wird auf das Galerkin-Verfahren, die Methode der finiten Element und auf iterative Methoden zur Lösung der resultierenden Gleichungssysteme besonders eingegangen.

Das Ziel des theoretischen Teils dieser Arbeit wird in Kapitel 7 erreicht. Hier werden zunächst zwei Beweise der Existenz schwacher Lösungen des linearen und nichtlinearen Navier-Stokes-Problemes geliefert, um anschließend einen Beweis der Existenz starker Lösungen der vollständigen Navier-Stokes Gleichungen in der Dimension $n \geq 3$ zu erläutern. Diesers letzte Existenzbeweis liefert schon anhand von genügend kleinen Anfangswerten eine globale, eindeutige Lösung. Die Berechtigung der gängigen numerischen Verfahren ist hiermit gesichert.

Die Umsetzung dieser biologischen, physikalischen und mathematischen Grundlagen und Erkenntnisse erfolgt in Kapitel 8, welches die Implementierung eines vereinfachten 2-dimensionalen, stationären Modelles zur Simulation des Blutflusses durch eine künstliche Aortenklappe beinhaltet. Unter Verwendung der C++ Bibliothek "Diffpack 3.0" wird der Aufbau eines klassischen Finite-Elemente-Lösers für die Navier-Stokes Gleichungen beschrieben und auf das stationäre Problem mit unterschiedlichen Positionen der Klappenflügel angewendet. Besondere Aufmerksamkeit soll dabei den Auswirkungen verschiedener Öffnungswinkel gewidmet werden. Zum Vergleich der Ergebnisse dienen sowohl die simulierten, als auch die experimentellen Ergebnisse von Mary J. King aus ihrer Doktorarbeit "Computational and experimental studies of flow through a bileaflet mechanical heartvalve" aus dem Jahre 1994 (siehe [23]).

Der Vergleich zeigt deutlich, daß ein 2-dimensionales zeitunabhängiges Modell nicht exakt die Bildung von Turbulenzen und Wirbeln lokalisieren kann, aber dennoch durchaus in der Lage ist, Tendenzen der Blutströmung in der unmittelbaren Umgebung der mechanischen Herzklappe vorherzusagen.

Danksagung

Die ersten, denen ich danken möchte, sind Prof. Colin Cryer und Prof. Michael Wiegner. Prof. Cryer war derjenige, der meine Begeisterung für das interdisziplinäre Arbeiten geweckt und durch das vergangene Jahr begleitet und gefördert hat. Prof. Wiegner brachte mir durch geduldige und wiederholte Hilfestellungen seinen Beweis zu Theorem 7.1 näher, so daß Kapitel 7 in greifbare Nähe rückte. An dieser Stelle sei für das mir und meiner Arbeit entgegengebrachte Interesse gedankt.

Doch ohne die technische Hilfe vieler gutmütiger Menschen wäre diese Diplomarbeit nicht so bald fertiggestellt worden. Allen voran sind da Dr. Frank Wübbeling und Axel Feldmann zu nennen, die mich immer wieder milde lächelnd mit einem Knopfdruck von mir unlösbar erscheinenden Computer-Problemen befreiten, dicht gefaolg von Jutta Lücking, die sich gemeinsam mit mir durch Diffpack kämpfte und über jedes bunte Bild freute. Insgesamt war die Atmosphäre in dem Institut für numerische Mathematik durch den allgemein humorvollen Umgangston mehr als angenehm und ich bin froh, in diese Etage gelandet zu sein.

Begleitet, unterstützt und motiviert haben mich in den letzten Monaten viele gute Freunde, unter denen Michael G. W. Jacob ein besonderer Dank gebührt. Seine allumfassenden Ratschläge und sein tapferes Korrekturlesen, das von Axel Naumann noch unterstützt wurde, waren mir eine große Hilfe. Dann waren da noch viele wichtige Menschen in meinem Umfeld wie Kerstin Ronneberger, Helke Kläning, Sandra Denningmann und die unverwüstliche und doch so charmante Mensa-Runde samt allen gelegentlichen Gästen, um nur einige von denen zu nennen, die ich vermissen werde.

Zu guter Letzt noch ein Danke, das mit Worten kaum noch auszudrücken ist, an Markus Kirchler, der in den vergangenen Monaten liebevoll über manche Stresserscheinung hinweggesehen hat, und an meine Familie, ohne die ich jetzt nicht da wäre, wo ich bin. Ohne Euch hätte ich selbst nicht geglaubt, daß es zu schaffen ist.

1 Das menschliche Herz

Im folgenden sollen Aufbau, Funktion und krankhafte Veränderungen des menschlichen Herzens in ihren Grundzügen beschrieben werden; vergleiche dazu [12], [33] und [29].

1.1 Aufbau und Funktion

Das menschliche Herz ist ein pulsierender Hohlmuskel, der die Bewegung der Blutflüssigkeit in unserem geschlossenen Blutsystem bewirkt. Form und Größe des Herzens entsprechen in etwa der Faustgröße der jeweiligen Person. Seine Oberseite, die Herzbasis, an welcher die großen Gefäße münden bzw. entspringen, hat einen deulich größeren Umfang als die nach unten links gerichtete Herzspitze. Die Herzbasis liegt unmittelbar hinter dem Brustbein, während die Herzspitze nach links von der Mittellinie abweicht.

Der größte Teil des Herzens ist vom Perikard, dem Herzbeutel, eingehüllt. Das Perikard bildet eine allseits geschlossene *seröse Höhle* mit normalerweise nicht mehr als 20 ml Flüssigkeit im Binnenraum. Wie andere seröse Höhlen auch, dient der Herzbeutel dazu, die Reibung zwischen dem beweglichen Herzen und den mehr oder weniger fixierten Nachbarorganen herabzusetzen (ähnlich zweier Glasplatten mit einem Tropfen Wasser dazwischen). Außerdem überträgt dieser Herzbeutel den Sog des Lungengewebes auf das Herz und sorgt somit für einen Druckausgleich im gesamten Thoraxraum.

Das Herz wird von einer muskulösen Scheidewand, dem Septum, in eine linke und eine rechte Herzhälfte getrennt, welche sich jeweils in Vorhof (*Atrium*) und Kammer (*Ventrikel*) unterteilen. Das Septum ist im Vorhofbereich dünn und sehnig, zwischen den Kammern aber außerordentlich muskelstark, da die mechanische Beanspruchung hier deutlich höher ist.

Der rechte Vorhof erhält durch die Hohlvenen sauerstoffarmes Blut aus dem Körperkreislauf und befördert es in die rechte Kammerhälfte. Von hier aus wird es zum Gasaustausch über die Lungenarterie in die Lunge gepumpt, gelangt nach deren Passage als sauerstoffreiches Blut in die Lungenvenen und dann in den linken Vorhof („kleiner“ Blutkreislauf). Aus dem linken Vorhof wird es in die linke Kammer, von dieser mit hohem Druck in die *Aorta* und weiter in den Körperkreislauf gepumpt („großer“ Blutkreislauf).

Wie das gesamte Gefäßsystem, so hat auch das Herz einen dreischichtigen Aufbau:

Abbildung 1: Längsschnitt durch das Herz (aus [33], S.127)

- Das **Endokard**, die glatte Innenschicht des Herzens, kleidet den Blutraum vollständig aus und ist somit entscheidend für eine gute Gleitfähigkeit des Blutes. Schon kleine Unebenheiten, z.B. aufgrund einer *Endokarditis*, erhöhen den Strömungswiderstand relevant. Das Blut fließt an diesen Stellen dann langsamer, es können sich *Thrombozyten* ablagern und ineinander verkeilen - es enstehen Blutgerinsel, bzw. *Thromben*. Wenn sich ein Thrombus ablöst und in den allgemeinen Kreislauf gelangt, kann er eine wichtige Arterie, z.B. im Gehirn, verstopfen und damit einen Schlaganfall verursachen (*Embolie*).

  Unterhalb dieses einschichtigen Gewebes finden sich reichlich elastische Fasern. Aus einer besonders kollagenfaserreichen Falte des *Endokards* bestehen auch die Herzklappen und Sehnenfäden.

- Das **Myokard**, die Herzmuskelschicht, schließt sich nach außen an das *Endokard* an. Da das *Myokard* die eigentliche Herzarbeit verrichten muß, ist es die dickste Schicht der Herzwandung. Es besteht aus einkernigen Herzmuskelzellen, die mechanisch und elektrisch miteinander verbunden sind. Im Ventrikelbereich schrauben sich die Muskelfasern in den äußeren *Myokardschichten* spiraliförmig von der Herzklappenebene zur Herzspitze hinab, biegen dort um und winden sich innen wieder in Richtung Herzbasis zurück. Dadurch stehen die Muskelfasern der äußeren und inneren Myokardschicht häufig senkrecht aufeinander. Die Aktin- und Myosinfilamente der Herzmuskulatur überlappen sich also in Ruhe so stark, daß sie sich tendenziell gegenseitig behindern. Bei körperlicher Belastung erhöht sich das Volumen der *Ventrikel*, die Muskulatur wird besser vorgedehnt und arbeitet effektiver, wodurch sich Kraft und Schlagvolumen des Herzens erhöhen.

  Diese autonome Anpassung der Muskulatur besitzt eine besonder große Bedeutung bei der *Herzinsuffizienz*: Ein geschwächter Herzmuskel neigt stets dazu, ein größeres Volumen anzunehmen („ausgelatschtes Herz“), was den oben beschriebenen Mechanismus aktiviert und die Herzschwäche über lange Zeit hinweg ausgleichen kann.

- Das **Epikard** sorgt für eine gute Gleitfähigkeit und Beweglichkeit des Herzens in seiner Umgebung. Es stellt gleichzeitig die äußerste Schicht des Herzens und das innere Blatt des Herzbeutels dar.

Wegen der extremen Dicke der Herzwand kann das Herz nicht direkt aus dem durchgepumpten Blut versorgt werden — die Diffusionsstrecken wären viel zu lang. Es ist ein eigenes Blutversorgungssystem nötig, welches von den Herzkranzgefäße gebildet wird. Bei Verschluß einer Herzkranzarterie oder eines größeren Arterienastes wird die von diesem Gefäß versorgte Muskulatur nicht mehr mit Sauerstoff und Nährstoffen beschickt und stirbt ab. Auf diese Weise ensteht ein Herzinfarkt.

1.2 Diastole und Systole

Um eine Flüssigkeit durch eine Leitung zu pumpen, gibt es zwei sich grundlegend unterscheidende Möglichkeiten: die Kreiselpumpe und die Ventilpumpe. Da die biologische Evolution das Rad und damit auch drehbare Systeme aller Art niemals entwickelt hat, muß das Herz eine Ventilpumpe sein, die ähnlich wie ein Blasebalg funktioniert: Ein- und Auslaßventil entsprechen den Herzklappen, der eigentliche Blasebalg ist die Herzkammer.

Bei der rhythmischen Tätigkeit des Herzens unterscheidet man in einem Zyklus zwischen der Kontraktions- und Austreibungssphase (**Systole**) sowie der Erschlaffungs- und Füllungsphase (**Diastole**).

1. **Systole**

    - **Kontraktionsphase:** Die Systole beginnt mit einer Anspannungsphase von ca. 60 ms, in der alle Herzklappen geschlossen sind. Zunächst spannt sich die Ventrikelmuskulatur an und übt einen Druck auf das Blutvolumen in den *Ventrikeln* aus. Das Volumen der *Ventrikel* ändert sich dabei nicht, da das Blut wie jede wässrige Flüssigkeit praktisch inkompressibel ist. Daher wird die Anspannungsphase auch isovolumetrische Kontraktion genannt. Der Ventrikeldruck nimmt so lange zu, bis er den Druck in der *Aorta* bzw. in dem *Truncus pulmonalis* übersteigt. Der rechte *Ventrikel* hat es dabei deutlich leichter, da er nur den diastolischen Pulmonalarteriendruck von 10 mmHg (Aorta 80 mmHg) überwinden muß. Die Anspannungsphase endet daher rechts früher als links.

      Zu Beginn der Anspannungsphase machen beide *Ventrikel* eine Formveränderung durch: Aus der länglichen Gestalt während der Diastole wird eine annähernde Kugelform. Diese Formveränderung teilt sich der Brustwand als dumpfer Schlag mit und ist als erster Herzton zu hören und über der Herzspitze zu tasten.

    - **Austreibungsphase:** Nachdem der Ventrikulardruck die diastolischen Werte in *Truncus pulmonalis* und *Aorta* überschritten hat, öffnen sich die Klappen zu den beiden großen Arterien und es beginnt die ca. 200 ms lange Austreibungsphase, also die eigentliche Bewegung der Blutsäule in den großen Arterien. Die Einlaßventile zwischen *Ventrikel* und *Atrium* bleiben wegen des hohen Drucks im *Ventrikel* weiterhin geschlossen. Trotz des ständigen Auswurfes von Blutvolumen steigt der Ventrikeldruck während der ersten zwei Drittel der Austreibungsphase weiterhin an und erreicht Spitzenwerte von ca. 130 mmHg (links) bzw. 25 mmHg (rechts).

Normalerweise treiben beide *Ventrikel* nur knapp die Hälfte ihres Inhaltes in die großen Arterien aus; das Schlagvolumen beträgt durchschnittlich 70 ml, kann bei körperlicher Belastung jedoch deutlich gesteigert werden. Durch den Blutauswurf verkleinert sich natürlich das Ventrikelvolumen. Dieses verringerte Volumen zieht einerseits die Herzspitze geringfügig in Richtung Herzbasis, andererseits senkt sich die Klappenebene in Richtung Herzspitze. Die Bewegung dehnt und öffnet die Vorhöfe; venöses Blut wird in das rechte und linke *Atrium* „eingesogen" (Ventilmechanismus).

2. **Diastole**

   - **Erschlaffungsphase:** Die Diastole beginnt mit einer Erschlaffungsphase von ca. 40 ms. Die Muskulatur der Ventrikelwand verliert ihre Spannung, der Ventrikeldruck sinkt unter den Blutdruck in der *Aorta* bzw. den *Pulmonalarterien* ab. Nun schließen sich auch die Ausflußklappen wieder; man spricht von isovolumetrischer Entspannung, wobei sich das eigentliche Zuschlagen durch einen hell klingenden zweiten Herzton verrät. Die Ventilebene wandert geringfügig in Richtung Herzbasis, wird aber durch das große Atriumvolumen an einer vollständigen Hebung gehindert.
   - **Füllungsphase:** Ist die Herzmuskulatur vollständig erschlafft, fällt der Druck im *Ventrikel* unter den im *Atrium*. Die Einlaßklappen öffnen sich und es beginnt eine Füllungsphase von sehr variabler Dauer, die mit einer erneuten Systole endet. Da die *Atrien* aufgrund des Ventilebenenmechanismus prall mit Blut gefüllt sind, findet am Beginn der Füllungsphase ein schneller Einstrom in die *Ventrikel* statt, der mit verschwindendem Druckunterschied allmählich nachläßt. Am Ende der Füllungsphase kontrahiert sich das *Atrium* und trägt so aktiv zur Vergrößerung des Blutvolumens in der Herzkammer bei.

Die rhythmische Fortbewegung der Blutflüssigkeit bewirkt in den nachfolgenden zentralen Gefäßabschnitten pulsierende Schwankungen des Wand- bzw. Blutdruckes. Deshalb werden allgemein für den Blutdruck zwei Werte, der systolische und der diastolische Druck angegeben. Um die starken Blutdruckschwankungen zwischen Diastole und Systole zu mildern, sind die großen Arterien mit sehr elastischen Wänden ausgestattet, durch die sie als sogenannte Windkessel fungieren:

Der Begriff „Windkessel" stammt aus der Dampfmaschinentechnik und bezeichnet eine technische Lösung des Problems, den stoßweisen Ausstrom von Flüssigkeit aus einer Pumpe in eine gleichmäßige Strömung umzuwandeln. Der physiologische Windkessel besteht im Ausdehnen bzw. Zusammenziehen der Wandung von *Aorta* und großen Arterien. Das Herz wirft in der Systole Blut aus, während der Blutstrom in der Diastole zum Erliegen kommt. Die Windkesselarterien dehnen sich während der Systole aus und nehmen so eine Teil des Schlagvolumens zusätzlich auf, der in der Diastole passiv wieder ausgepresst wird. Je elastischer die Arterien sind und je größer ihr Binnenvolumen im Verhältnis zum Schlagvolumen ist, desto gleichmäßiger strömt das Blut in die Peripherie ab.

Abbildung 2: Phasen der Herztätigkeit (aus [33], S. 131)

In der oberen Hälfte von Abb. 2 erkennt man die Bewegung der Ventilebene in der Austreibungs- und Füllungsphase. Die untere Hälfte zeigt die Druckverhältnisse in der linken Herzkammer (rot) und in der *Aorta* (grün) während der Herzaktion.

1.3 Herzklappen

Wie zuvor beschrieben, verhält sich das Herz wie eine Ventilpumpe, wobei die Funktion der Ein- und Auslaßventile von der Herzklappen übernommen wird. Je nach

Abbildung 3: Querschnitt durch das Herz (aus [33], S. 128)

Aufbau und Funktionalität unterscheidet man grundsätzlich zwischen **Segelklappen** und **Taschenklappen**.

Zwischen Vorhöfen und Kammern liegen die beiden großen Segelklappen (siehe Abb. 3): Die Mitral- und die Trikuspidalklappe.

- Die zweizipfelige **Mitralklappe** liegt zwischen dem linken *Atrium* und dem linken *Ventrikel*. Sie besteht aus zwei dünnen, von glattem Endothel überzogenen Einzelsegeln: einem größeren Aortensegel an der Vorderwand und einem kleineren hinteren Wandsegel.
- Ihr Gegenstück im rechten Teil des Herzens ist die dreizipfelige **Trikuspidalklappe** mit je einem vorderen, hinteren und Septumsegel.

Die Segelklappen sind Einwegventile, die einen Blutstrom nur vom *Atrium* in den *Ventrikel*, nicht aber in umgekehrter Richtung zulassen. Zwei weitere Klappen, sogenannte **Taschenklappen**(sieh Abb. 3), trennen die linke bzw. rechte Herzkammer von den großen Arterien (*Aorta* und *Truncus pulmonalis*). Sie sorgen dafür, daß zwischen den einzelnen Herzschlägen kein Blut aus den Arterien in die *Ventrikel* zurückfließen kann.

- Je nach der verschlossenen Abgangsarterie heißt die rechte Taschenklappe **Pulmonalklappe**,
- die linke dementsprechend **Aortenklappe**.

Die Segelklappen sind großflächiger als die Taschenklappen und benötigen daher Sehnenfäden, die von ihrem freien Rand in die Herzkammer ragen und ein „Durchschlagen“ der Klappen während der Systole in die Vorhöfe verhindern. Sie sind an der Ventrikelinnenwand über Papillarmuskeln befestigt, welche die Länge der Sehnenfäden dem Kontraktionszustand der Kammern anpassen können. Die Taschenklappen bestehen jeweils aus drei identischen Teilen, die sich wie Schöpfkellen in Richtung der Herzkammern vorwölben. Die Klappen sind zwischen zwei Herzschlägen, in der Diastole, durch den Druckunterschied zwischen Arterien und Herzkammern geschlossen; sie stoßen in der Form eines „Mercedessterns“ aufeinander. Spannt sich der Herzmuskel an (Systole), so drückt das aus dem Herz ausströmende Blut die Taschenklappen an die Wand von *Aorta* bzw. *Truncus pulmonalis*; es entsteht eine dreieckige Öffnung. Da die Taschenklappen nur eine relativ kleine Fläche zu verschließen haben, benötigen sie keine Sehnenfäden. Im weiteren wird nur noch die Aortenklappe, also die Taschenklappe zwischen *Aorta* und linkem *Ventrikel* betrachtet, da diese das wesentliche Ventil zwischen Herz und Körperkreislauf darstellt.

2 Künstliche Herzklappen

Herzklappen können von Geburt an falsch angelegt oder durch Krankheiten (z.B. Infektionen) geschädigt sein (vergl. [33], S. 129). Weil dadurch der reguläre Blutfluß gestört ist, kann das Blut unter Umständen in die falsche Richtung strömen. Um den Kreislauf aufrecht zu erhalten, muß das Herz deshalb stärker arbeiten.

Diese Situation kann zum Herzversagen führen. Die geschädigte Herzklappe muß daher entweder operativ behandelt oder durch eine neue ersetzt werden. Ein chirurgischer Eingriff am Herzen löst bei den meisten Patienten allerdings große Ängste und Unsicherheit aus. Die veränderten Lebensbedingungen nach der Operation können diese Empfindungen noch verstärken.

So wird von den Betroffenen — besonders in der ersten Zeit nach der Operation — das „Klicken" der künstlichen Herzklappen sehr bewußt wahrgenommen. Je nach Herzklappe (Kunststoffklappe oder biologische Klappe) kann es zudem erforderlich sein, daß die Patienten in ihrem weiteren Leben regelmäßig Medikamente einnehmen.

2.1 Angeborene und erworbene Klappenfehler

Zu den seltenen, angeborenen Klappenfehlern (vergl. [21], S. 127-156) zählen die sogenannten *Stenosen*, wobei am häufigsten noch die angeborenen Aortenklappenstenosen vorkommen. Es handelt sich hierbei um eine Entwicklungsstörung der Klappenflügel. Statt mit drei Flügeln ist die Aortenklappe bei diesen Patienten durch einen genetischen Fehler nur *bikuspidal* ausgebildet.

Neben den angeborenen *Stenosen* ist als Hauptursache von Herzklappenerkrankungen das *rheumatische Fieber* zu nennen. Die rheumatische *Endokarditis* führt zu einer Verklebung der Klappenflügelränder, woraus nicht selten eine *bikuspidale* Klappe entstehen kann. Daraufhin kommt es in der Regel zu einer Kalkeinlagerung durch Wirbelbildungen und zu einer hohen Blut-Flußgeschwindigkeit an der Klappe. Die Öffnung der Aortenklappe verengt sich noch mehr. Die analoge Krankheitsentwicklung kann man bei angeborenen Aortenstenosen beobachten.

Je weiter die Klappenstenose ausgebildet ist, desto größer wird das Druckgefälle zwischen linkem *Ventrikel* und *Aorta* während der Systole. Der erschwerte Ausfluß des Blutes aus dem *Ventrikel* in die *Aorta* führt also zu erhöhtem Druck und somit zu einer starken *Hypertrophie* der Ventrikelmuskulatur, die dadurch einen erhöhten Sauerstoffbedarf hat.

Im weiteren Verlauf der Erkrankung sinken Herzminutenvolumen und Schlagvolumen in Ruhe ab, wodurch der Weg zur *Herzinsuffizienz* vorgezeichnet ist. Um einen frühzeitigen Tod durch Myocardversagen, Herzrhythmusstörungen oder Herzinfarkt zu vermeiden, läßt sich die Implantation eines Herzklappenersatzes kaum umgehen.

2.2 Die wichtigsten künstlichen Herzklappentypen und ihre charakteristischen Eigenschaften

Sowohl Aortenklappe als auch Mitralklappe können durch eine künstliche Prothese operativ ersetzt werden, was die Abbildung 4 zeigt:

Abbildung 4: Doppelklappenersatz (aus [6])

Allgemein unterscheidet man zwischen mechanischen und biologischen Prothesen (vergl. [25], S. 25ff und [6]). Jede dieser Prothesen zeigt charakteristische Vor- und Nachteile. Zu den mechanischen Prothesen gehören Ball-, Hubscheiben-, Kippscheiben- und Doppelflügelprothesen.

- Bei den **Ball- und Hubscheibenprothesen** (siehe Abb. 5, a und b) bewegt sich eine Kugel bzw. eine Scheibe in der Öffnungsphase in der Richtung des

Abbildung 5: Mechanische Prothesen und Bioprothesen (aus [25], S.26)

Blutflusses und wird von einem Käfig festgehalten. In der Schlußphase liegt der Schließkörper auf dem Prothesenring auf. Das Flußprofil der Kugel-Käfig-Prothesen weist einen exzentrischen Verlauf auf, da der Blutstrom von der sich verlagernden Kugel seitlich zu den Wänden abgedrängt wird. Das diastolische Flußmuster einer Hubscheibenklappe zeigt ein ähnliches Flußprofil. Der diastolische Einstrom wird an der Scheibe abgelenkt, seitlich der Scheibe fließt das Blut entlang der Wände nach apikal. Die kurze Flußumkehr in Richtig des Herzens zum Schließen der Klappe erfolgt dann überwiegend zentral auf die Scheibe zu. (siehe Abb. 6, b und c)

- Die **Kippscheibenprothesen** (siehe Abb. 5, c) besitzen eine bewegliche Scheibe auf einem exzentrisch gelegenen Drehschanier. Durch diese Konstruktion können die meisten Kippscheibenklappen nur bis ca. 70° geöffnet werden, so daß der Ein- und Ausstrom abhängig von der Einbaurichtung der Prothese unterschliedliche Flußmuster aufweist.

  Man unterscheidet einen stärkeren Hauptstrom zur einen und einen schwächeren Nebenstrom zur anderen Wand. (siehe Abb. 6, d und e)

- Ein anderes mechanisches Modell ist die **Doppelflügelprothese** (siehe Abb. 5, d). Hierbei öffnen sich zwei Klappendeckel wie Türflügel mit einem fast rechten Winkel, ca. 85°, so daß das Strombahnhindernis im inneren Prothesenbereich nicht sehr ausgeprägt ist. Wenn bei maximaler Öffnung die Flügel beinahe senkrecht stehen, kann das Blut fast ungehindert durch zwei breitere seitliche und einen schmalere zentrale Öffnung ausströmen. Die Flußumkehr erfolgt entlang der Seitenwände. (siehe Abb. 6, f und g)

- Die implantierten **Bioprothesen** (siehe Abb. 5, e und f) bestehen meist aus Schweineaortenklappen oder Kalbsperikard und werden in der Regel auf flexible Bügel aufgenäht, wobei die Basis der Klappe von einem Metallring umgeben wird. Das biologische Gewebe ist an einem Nahtring mit Haltestreben befestigt und zeigt im intakten Zustand fast denselben Bewegungsablauf wie eine normale, menschliche Herzklappe. (siehe Abb. 6, h bis m)

Die weiteren Ausführungen werden sich von jetzt an nur noch auf Doppelflügelprothesen beziehen, da diese in der modernen Praxis am häufigsten eingesetzt werden. Obwohl die Bioprothesen haemodynamisch weitaus günstiger sind, haben sie doch im Vergleich zu den Doppelflügelprothesen einen entscheidenden Nachteil: Wie jedes biologisches Gewebe sind sie nicht unbegrenzt haltbar. Nach ca. 10-15 Jahren zeigen Bioprothesen die ersten Verkalkungs- und Ermüdungserscheinungen, so daß sie in der Regel nur bei Patienten über 60 Jahre eingesetzt werden. Bei jüngeren Patienten wäre das Risiko von weiteren Operationen zu hoch. Da es das Ziel ist, das natürliche

Abbildung 6: Strömungsprofile verschiedener Klappenprothesen (aus [25], S.28)

Flußprofil so gut wie möglich zu immitieren, erscheinen alle Varianten bis auf die Doppelflügelprothesen und Bioprothesen als nicht besonders geeignet. Weicht das Flußprofil zu sehr vom natürlichen Fall ab, so sind die Belastungen für manche Teile der Aortenwand zu hoch und es entstehen Stellen zu hoher und Stellen zu geringer Flußgeschwindigkeit.

Die gängigsten Modelle der Doppelflügelprothesen sind die **St. Jude Medical** (kurz: **SJM**) und die **Carbomedics** Herzklappen, die sich im Wesentlichen nur durch leicht verschiedene Öffnungswinkel unterscheiden:

Die Carbomedics-Klappen zeigen in maximaler Position einen Öffnungswinkel von 78°, die St. Jude-Klappen sogar 84°. Wie sich später heraus stellen wird, rufen diese auf den ersten Blick geringfügigen Unterschiede deutliche Abweichungen in den Fluß- und Geschwindigkeitsprofilen hervor.

2.3 Schwachstellen bzw. krankhafte Veränderungen des Herzens durch Klappenprothesen

Die häufigsten und wohl auch schwerwiegendsten Krankheitsbilder aufgrund von Herzklappenoperationen sind Hirnblutungen, Magen-Darm-Blutungen, *Thromboembolien*, *haemolytische Anaemien* und eine evtl. noch verbleibende *Herzinsuffizienz*.

- **Hirn-** und **Magen-Darm-Blutungen** können als Nebenwirkung von starken gerinnungshemmenden Medikamenten enstehen, die aber meist verabreicht werden müssen, um die erhöhte Trombosegefahr in der Umgebung der Klappenprothese zu kontrollieren.

- Das andere Extrem findet man bei Patienten mit **Thromboembolien**:

  In der Umgebung der Klappenprothese bilden sich Blutgerinsel, die auseinanderbrechen und vom Blutstrom weitergetragen werden. Diese „Bruchstücke“ können sich festsetzen und z.B. eines der lebensnotwendigen Herzgefäße verstopfen.

- Durch verstärkte **Haemolyse** an den Herzklappenflügeln kann der Anteil der roten Blutkörperchen stark abfallen. Man spricht von einer *haemolytischen Anaemie*, der Blutarmut, die eine Störung des Sauerstoff-Transportes durch das Blut verursacht. Alle sauerstoffabhängigen Leistungen, z.B. auch alle Muskel-Leistungen werden auf diese Weise erschwert.

- Sowohl durch die verminderte Sauerstoffversorgung, als auch durch die Thromboseneigung ist das Herz ungewöhnlich großen Belastungen ausgesetzt, so daß es nicht mehr in der Lage ist, das notwendige Blutvolumen in die weiterführenden Gefäße zu pumpen, was zu einer **Herzinsuffizienz** führen kann.

Abbildung 7: SJM-Herzklappenpatienten **ohne** spätere Komplikationen, in Prozent; **AVR** bzw. **MVR** entsprechen **aortic** bzw. **mitral valve replacement** (aus [19], S. 10)

Abbildung 8: links: Carbomedics-Herzklappenpatienten **ohne** spätere Thromboembolien und **ohne** spätere Blutungen durch gerinnungshemmende Medikamente (aus [1], S. 631)

Wie man in Langzeitstudien an den St. Jude Medical (siehe Abb. 7 und 10) und den Carbomedics Herzklappen (siehe Abb. 8 und 9) feststellen konnte, findet man in allen Testgruppen nach einigen Jahren Patienten mit den oben beschriebenen

Abbildung 9: Überlebensraten der Carbomedics-Herzklappenpatienten nach **Aortenklappen-Austausch** und nach **Mitralklappen-Austausch** (aus [1], S. 630)

Abbildung 10: Überlebensraten der SJM-Herzklappenpatienten; **AVR** bzw. **MVR** entsprechen **aortic** bzw. **mitral valve replacement** (aus [19], S. 10)

Kranheitsbildern, die in vielen Fällen zum Tod führen. Obwohl diese Ereignisse zeitlich stark verzögert auftreten, sind sie immer noch auf die Klappenoperation bzw. auf die klassischen Schwierigkeiten mit der künstlichen Prothese zurückführbar.

Die Mortalitäts-Kurven der beiden betrachteten Klappentypen ähneln sich sehr stark, wobei die Kurve bzgl. der Carbomedics-Prothesen die frühen, postoperativen Todesfälle nicht miteinbezieht, was in der SJM-Graphik der Fall ist. In allen betrachteten Testgruppen fallen diese postoperativen, frühen Fälle mit ca. 7% ins Gewicht.

Der Überblick über diese statistischen Gegebenheiten zeigt, wie häufig doch die unterschiedlichen Krankheitsbilder in der heutigen Zeit noch auftreten. Diese ließen sich theoretisch vermeiden, wenn es möglich wäre, ein naturgetreues Modell einer Herzklappe zu konstruieren, welches keine Unregelmäßigkeiten im Flußprofil mehr aufzeigt. Ein zentrales Kriterium dafür sind die exakte Modellierung und die anschließende Simulation des Blutflusses durch eine mechanische Herzklappe.

3 Physikalische Grundlagen

3.1 Der Idealfall

Betrachtet man die in Abschnitt 1 und 2 beschrieben biologischen und medizinischen Gegebenheiten, so zeichnet sich ab, welche Eigenschaften eine **ideale künstliche Herzklappe** haben sollte, um dem natürlichen Fall möglichst nahe zu kommen und möglichst wenige Komplikationen zu verursachen (vergl. [23], S. 5):

- Reibungsfreies Öffnen und Schließen der Klappe bereits bei minimaler Druckänderung.
- Keine Druckunterschiede an der Klappenoberfläche während des Blutausstromes.
- Kein Blutrückfluß, während die Klappe sich schließt bzw. ganz geschlossen ist.
- Lebenslange Haltbarkeit ohne Errosionserscheinungen.
- Keine durch die Herzklappe verursachten krankhaften Veränderungen des umliegenden Gewebes.
- Keine Verursachung von Haemolysen und Thrombosen.

Die letzten beiden Bedingungen zielen direkt auf das Flußverhalten des Blutes beim Strom durch die Klappe. Wird die Strömung beim Ausfluß zu turbulent, bilden sich hinter der Klappe, wie wir später noch genauer erkennen werden, starke Wirbel. Diese lassen in den Klappentaschen und in der *Aorta* einerseits Stellen mit zu hoher und andererseits Stellen mit zu geringer bzw. gar keiner Flußgeschwindigkeit entstehen. Bei zu geringer Flußgeschwindigkeit besteht eine extrem hohe Gerinnungsgefahr, da das Blut fast steht. Zu hohe Geschwindigkeiten lösen meist vermehrte Haemolyse (durch Zerquetschen und Reißen von Blutzellen an den starren Klappenflügeln) und unkontrolliertes Wuchern der arteriellen Wand (durch zu starken Aufprall des Blutstromes) aus.

Um eine ideale Klappenprothese zu konstruieren, braucht man also detaillierte Kenntnisse über das Flußverhalten des Blutes, bzw. Druck und Geschwindigkeit innerhalb des Blutflusses.

3.2 Historische Ansätze

Der Bewegung von Herzklappen widmen sich schon seit Jahrzehnten Physiologen, Physiker und nicht zuletzt Mathematiker. Im Jahre 1912 erkannten **Yandell Henderson** und **F. Elmer Johnson** (siehe [18]) als erste die fluiddynamischen Vehältnisse, die zum Öffnen und Schließen der Herzklappen führen und machten dies an einigen erstaunlich einfachen in vitro Experimenten deutlich.

- Die erste Versuchsanordnung besteht aus einem Glasröhrchen, das die Verbindung zwischen einem erhöhten Tintenreservoir und einem darunter liegenden Wasserbehälter bildet. Wird der Tintenzulauf von oben abrupt unterbrochen, kann man beobachten, daß die Tintensäule innerhalb des Röhrchens sofort stillsteht, während der sich schon im Wasser befindende Reststrahl seine Vorwärtsbewegung beibehält.

  Dieses produziert in einem kleinen Bereich an der unteren Rohröffnung kurzzeitig einen Unterdruck, so daß klares Wasser von den Seiten in diesen Bereich einströmt.

- Im zweiten Versuch wird zusätzlich noch ein D-förmiges Teilstück in die Mitte des Rohres eingefügt und am unteren Übergang mit einer freischwingenden Membranklappe versehen. Läuft nun zuerst der Tintenfluss gerade das Rohr hinunter in den Wasserbehälter, so verursacht er fast keine Bewegung im gebogenen Teilstück.

  Wird dann aber der Fluss z.B. durch kurzes Verschließen des Röhrchens wiederum unterbrochen, so beginnt die Tinte in dem D-Stück zu zirkulieren und veranlaßt die Membranklappe, sich entegegen dem Vorwärtsstrom nach oben zu schließen.

- Der dritte Versuch simuliert nun mit Hilfe einer angepaßten, flexiblen Gummimanschette an der unteren Öffnung des Röhrchnes das Verhalten der Klappenflügel. Die Flüssigkeitssäule innerhalb des Rohres wird deutlich über den Wasserstand des Behälters angehoben und mit einer Deckplatte oben gehalten. Entfernt man die Deckplatte und läßt die Säule nach unten fallen bzw. fließen, so kann man folgendes beobachten:

  Fällt der Level innerhalb des Röhrchens unter den des Wasserbehälters, wird der Druck von außen zu groß, so daß die Manschette dann zusammen fällt und das Röhrchen so verschließt.

Mit Hilfe von angefärbter Flüssigkeit und photographischen Momentaufnahmen zeigten Henderson und Johnson recht deutlich, daß die Klappenflügel sich in dem

Moment, in dem der Blutstrom unterbrochen wird, nicht wie „schwingende Türen" (vergl. [18], S. 81) schließen, sondern sich durch deutliche Wirbel in den Klappenhöhlen von der Basis her „einrollen" (vergl. [18], S. 81), während sich das Blut noch im Vorwärtsfluß befindet. Durch diesen Mechanismus wird ein Blutrückstrom in die linke Herzkammer also fast vollständig vermieden. Leider wurden die Schlußfolgerungen von Henderson und Johnson viele Jahre nicht beachtet oder fehlinterpretiert.

Erst im Jahre 1969 veröffentlichten **B. J. Bellhouse** und **F. H. Bellhouse** (siehe [2] und [3]) neue fluiddynamische Erkentnisse zur Klappenbewegung, deren Blickwinkel sich etwas von dem von Henderson und Johnson unterschied. Statt, wie diese, die Verantwortung für den Schluß der Klappen ausschließlich dem Effekt eines abbrechenden Flüssigkeitsstrahles zuzuschreiben („Breaking-of-a-jet-Theorie"), sahen sie die Ursache in der Bildung von Wirbeln an den Klappenflügeln.

- Der Versuchsaufbau sicherte einen laminaren, pulsierenden Fluß durch eine Modellklappe, in dem Wirbelbildung, Geschwindigkeit und Klappenöffnungsgrad zu unterschiedlichen Stadien gemessen und als zeitliche Funktionen aufgezeichnet wurden.

- Die gemessenen Werte lieferten folgende Ergebnisse: Die Stromlinien folgen dem Verlauf der Klappenflügel bis zum distalen Ende des Sinus und teilen sich dann in Sinus-Wirbel (rückwärts) und Hauptstrom (vorwärts). Die Wirbelbildung beginnt in früher Systole, erreicht ihr Maximum am Scheitelpunkt des Herzzyklus und hält bis weit in die Diastole an. Das deutliche Druckgefälle während der ausklingenden Diastole führt zu einer Verlangsamung der Flußgeschwindigkeit in der Aorta, die eine Ausweitung der Sinuswirbel erlaubt, so daß die Klappe sich schließt.

- Erstmals wurden die gemessenen Werte nun auch mit berechneten, also einem mathematischen Modell, verglichen. Auf der Grundlage der physikalischen Gegebenheiten wurden Gleichungen für den Druck aufgestellt, die folgende Werte miteinbezogen: Druckunterschied $p_1 - p_2$, Geschwindigkeit u, Beschleunigung $\frac{\partial u}{\partial t}$ (d. h. zeitliche Veränderung der Geschwindigkeit), Dichte ρ, Viskosität γ, Aortenradius a, Radius der Klappenöffnung r und Klappenflügellänge l. Diese Werte lieferten die Differentialgleichung

  $$p_1 - p_2 = -\frac{\rho l a}{r} \cdot \frac{\partial u}{\partial t} - \frac{1}{2} \cdot \rho u^2 (\frac{a^4}{r^4} - 1)\,.$$

  Die berechneten Werte stimmten in erstaunlich hohem Maße mit den gemessenen überein.

Obwohl dem Konzept von Henderson und Johnson keine Schwachstellen nachgewiesen werden konnten, bevorzugten die meisten die „Wirbel-Theorie“ von Bellhouse und Bellhouse, die der „Breaking-of-a-jet-Theorie“ nicht grundlegend widerspricht. Es ist vielmehr eine andere Sichtweise bzw. ein anderes Modell für dieselben fluiddynamischen Gegebenheiten.

Ebenfalls im Jahre 1969 wagten **B. J. Bellhouse** und **L. Talbot** (siehe [4]) den ersten Versuch, die Funktion von Herzklappen mathematisch zu modellieren. Sie erreichten aber nur eine grobe Annäherung an die gemessenen Werte. Bessere Ergebnisse erzielte **C. S. Peskin** (siehe [30]) im Jahre 1977, indem er vollständige Bewegungsgleichungen aufstellte und diese analytisch löste. Um die Lösbarkeit zu gewährleisten, mußte er sich auf ein 2-dimensionales Modell beschränken und, um den Rechenaufwand in einem realisierbaren Rahmen zu halten, auf eine recht kleine Reynoldszahl (Maß für die Viskosität der betrachteten Flüssigkeit), die den physiologischen Gegebenheiten nicht gerecht wurde.

Die ersten akzeptablen Werte durch Lösen von fluiddynamischen Gleichungen erhielten **C. S. F. Lee** und **L. Talbot** 1979 (siehe [28] und [15]), indem sie die Gleichungen von 1969 noch weiter verfeinerten. Sie versuchten, den Effekt der Blutviskosität als Randbedingungen für den Prozeß der Klappenöffnung einzubringen, d.h. die Geschwindigkeit an verschiedenen Stellen der Klappenflügel sollte gleich sein. Zusammen mit den Gleichgewichtsgleichungen für Masse, Impuls und Energie erhielten sie eine übersichtliche aber doch genaue Beschreibung des Öffnugsprozesses. Für das Schließen der Klappen wurden die Gleichungen von 1969 übernommen, da dieser Vorgang nur unwesentlich durch Viskosität beeinflußt wird.

3.3 Ein moderner Ansatz: Doktorarbeit von Mary J. King, 1994

Im Jahre 1994 veröffentlichte Mary J. King an der Universität Leeds ihre Arbeit mit dem Titel „Numerische und experimentelle Studien zum Fluß durch zweiflügelige mechanische Herzklappen“ und bewirkte damit einen großen Schritt in der detaillierten Berchnung der Herzklappenfunktion. Bis zu diesem Zeitpunkt existierte nur ein 2-dimensionales, laminares, zeitunabhängiges Modell bezüglich einer physiologisch nicht angemessenen Reynoldszahl, das den Verlauf der Stromlinien zwar grob annähern konnte, aber nicht in der Lage war, die Bildung von Wirbeln und Turbulenzen exakt vorher zu sagen. Das Ziel der Doktorarbeit von King war also ein 3-dimensionales, zeitabhängiges Modell einer zweiflügeligen Klappe in aortischer Position bezüglich einer physiologisch akzeptablen Reynoldszahl.

Als Grundlage, bzw. als experimentelle Vergleichswerte, wurden die mechanischen Daten der Carbomedics und der St. Jude Doppelflügelklappen benutzt und mit zwei leicht veränderten Prototypen verglichen. Der wesentliche Unterschied der vier Klappentypen liegt in den verschiedenen Öffnungswinkeln der Klappenflügel zwischen 78° und 85°.

Wie in der Doktorarbeit (vergl. [23], S. 6ff) ausführlich beschrieben, wird zum vereinfachten Verständnis der fluiddynamischen Verhältnisse zunächst ein Fluß um einen 2-D Zylinder herum angenommen. Bei einer sehr niederigen Reynolds-Zahl bewegt sich das Fluid wie in Abb. 11, A, um den Zylinder herum. Wird die Reynolds-Zahl erhöht, lösen sich die Randschichten ab und es entstehen zwei zusammenhängende Wirbel, wie man in Abb. 11, B, sehen kann. Diese Wirbel dehnen sich aus und werden instabil, sobald die Reynolds-Zahl noch weiter ansteigt (siehe Abb. 11, C), bis sie schließlich alternierend von gegenüberliegenden Seiten des Zyliders strömen. Werden Winkelgeschwindigkeit und Scherkräfte der Wirbel zu groß, so brechen sie zusammen und werden zu Turbulenzen (siehe Abb. 11, D). Als nächstes wird ei-

Abbildung 11: Entwicklung des Flußverhaltens um einen 2-D zylinder bei wachsender Reynolds-Zahl (aus [23], S.7)

ne Doppelflügel-Klappenprothese betrachtet. Die Klappenflügel bestehen aus semielliptischen flachen Platten, die, wenn die Klappe vollständig geöffnet ist, in einem bestimmten Winkel (siehe Abb. 12) zur Strömung geneigt sind. Da die Klappenflügel sich über die Klappenöffnung ausdehnen, kann man weitere Modelle auf eine zweidimensionale Platte in der Mitte der Flügel beschränken. Diese dünne Platte würde sich innerhalb des Flüssigkeitsstromes ähnlich verhalten wie der oben beschriebene Zylinder.

Der Herz-Zyklus wird des weiteren mit folgenden Annäherungs- bzw. Durchschnittswerten beschrieben:

- Die Flußrate Q wird durch eine Sinus-Kurve $Q = Q_m \sin \omega t$ approximiert, wobei Q_m die maximale Flußrate ist. Durch ω können unterschiedliche Herzfrequenzen eingehen.

- Das ausgeworfene Blutvolumen pro Zyklus wird auf einen Durchschnittswert von 70 ml bei einer Herzfrequenz von 72 bpm (Schlägen pro Minute) festgelegt. Integriert man über den Abschnitt einer Herz-Zyklus-Kurve von 0 bis π, so läßt sich die maximale Flußrate und somit auch die Flußgeschwindigkeit berechnen.

- Man nimmt eine dynamische Blutviskosität von $\eta = 4 \cdot 10^{-3}\, Pa \cdot s$, die Dichte von Blut mit $\rho = 1000\, kg \cdot m^{-3}$, eine menschliche Durchschnitts-Aorta (achsensymmetrisch bzw. gerades Rohr) mit $r = 0,0145\, m$ Radius und eine Durchschnittsgeschwindigkeit des Blutes in der Aorta von $u = 0,3995\, m \cdot s^{-1}$ an und erhält so bzgl. des Durchmessers eine Reynoldszahl von $Re_D = 3000$, bzw. $Re_R = 1500$ bzgl. des Radius.

  $$Re_R = \frac{\rho u r}{\eta} = 1000 \cdot 0,3995 \cdot 0,0145 \cdot 250 \approx 1500\,.$$

  Für Rohrströmung liegt diese Reynoldszahl zwischen den Werten für laminaren und turbulenten Fluß. Daher würde man für das betrachtete 2D Herzklappenmodell ein Flußverhalten erwarten, das zwischen den Ansätzen 'C' und 'D' in Abb. 11 liegt. Die genaue Bedeutung und Berechnung der Reynoldszahl Re wird in Abschnitt 4.3 noch detailliert beschrieben. Obwohl man Blut eigentlich nicht als Newtonsches Fluid bezeichnen kann, wird zur weiteren Modellierung von laminarer Strömung ausgegangen. Da sich das Blut nur im Zentrum der großen Arterien nicht wie ein Newtonsches Fluid verhält, ist diese Vereinfachung im Bereich direkt hinter einer Herzklappe, wo ungewöhnlich große Schubkräfte herrschen, durchaus gerechtfertigt.

Eine flache, leicht geneigt Fläche, wie in diesem Modell, wird sich im Blutstrom wie eine Flugzeugtragfläche im Luftstrom verhalten, d.h. sie wird Auftrieb erzeugen. Zur Modellierung von Doppelflügelprothesen können die Flügel also wie Tragflächen behandelt werden, die sich in einem Winkel zwischen 5° und 12° bewegen, wie in Abb. 12 angedeutet.

Betrachtet man ein ideales Fluid im Fluß über eine Tragfläche, erkennt man, daß der Oberflächendruck an der Unterseite der Tragfläche größer ist als der an der Oberseite, so daß eine Auftriebskraft entsteht. Dieser Druckunterschied kann z.B. auf eine Zirkulation um den Flügel herum zurückgeführt werden.

Die Theorie zu solchen Problemen liefert nun die Vorhersage von Wirbelbildungen an der Tragfläche, was auf die Klappenflügel übertragbar sein sollte. Um diese Wirbelbildungen genau erfassen zu können, muß besonderer Augenmerk auf die Viskosität des Fluids und somit auch auf die Trägheit der Grenzschichten gerichtet werden, die sich in der Flußgeschwindigkeit bemerkbar macht: Im Hauptteil des Fluids findet man eine Geschwindigkeit $u \neq 0$, während man an der Oberfläche der Tragfläche und an der Außenwand des Rohres aufgrund von Reibung als Randbedingung eine Geschwindigkeit von $u = 0$ voraussetzt.

Abbildung 12: schematische Darstellung einer Doppelflügelklappe, Modell vergleichbar mit einer Tragfläche (aus [23], S.8)

Aus diesen Voraussetzungen erhält man eine speziell angepaßte Variante der Navier-Stokes-Gleichungen mit dazugehörigen Randbedingungen. Die Software, die zur Lösung benutzt wurde, basiert auf der Methode finiter Elemente, benutzt gewichtete Residuen und die Galerkin-Methode.

Das heißt, der zu betrachtende Raum wird zerteilt, z.B. trianguliert, die entstandenen Knoten werden bezeichnet und das Problem mit den dazugehörigen Randbedingungen für die neuen Elemente spezialisiert. Die daraus resultierende globale

Matrix bzw. das Gleichungssystem wurde iterativ gelöst, wobei bei den Toleranzgrenzen zwischen linearen und nicht-linearen Lösungen unterschieden wurde.

Die rein mathematischen und numerischen Fragestellungen, die diese Problematik der Navier-Stokes-Gleichungen hervorruft, sollen in den folgenden Abschnitten detailliert erläutert werden.

4 Navier-Stokes-Gleichungen und Reynoldszahl

Um die Navier-Stokes-Gleichungen herleiten und verstehen zu können (vergl. [7] und [38]), macht man häufig einen Umweg und betrachtet zuerst ein ideales Fluid, d. h. eine Flüssigkeit ohne innere Reibung bzw. ohne Zähigkeit, und ergänzt anschließend in der beschreibenden Gleichung Ausdrücke, die die Reibung beschreiben. Zur genauen Markierung von verschieden Orten innerhalb des Fluids zu verschiedenen Zeiten versieht man das Modell mit einem Koordinatensystem, in dem die Bahnlinie einzelner Punkte in Raum und Zeit verfolgt werden kann.

Im weiteren werden folgende Bezeichnungen verwendet:

- D sei ein Gebiet in $\mathbb{R}^2$ oder $\mathbb{R}^3$ gefüllt mit Flüssigkeit.
- $\underline{x} \in D$ sei ein Punkt in D; betrachtet werden die Flüssigkeits-Teilchen, die zum Zeitpunkt t den Punkt $\underline{x}$ passieren.
- $\underline{u}(\underline{x}, t)$ sei die Geschwindigkeit der Teilchen, die $\underline{x}$ zum Zeitpunkt t passieren.
- Man nimmt an, daß das Fluid eine zu jedem Zeitpunkt t und an jedem Ort $\underline{x}$ wohldefinierte Dichte $\rho(\underline{x}, t) > 0$ hat.
- $p(\underline{x}, t)$ beschreibt den orts- und zeitabhängigen Druck, $\underline{b}(\underline{x}, t)$ die auf einen Punkt wirkende Schwerkraft.
- $W \subset D$ Teilmenge mit glattem Rand, so ist die Masse $m(W, t)$ der Flüssigkeit in W zum Zeitpunkt t gegeben durch

  $$m(W, t) = \int_W \rho(\underline{x}, t)\, dV\,,$$

  wobei dV ein Volumen- bzw. Flächenelement beschreibt.

4.1 Herleitung der Euler-Gleichungen

Die Herleitung der Euler-Gleichungen basiert auf drei elementaren, physikalischen Gesetzen:

1. **Massenerhaltung**
2. **Impulserhaltung**

3. **Energieerhaltung**

Zu 1.: In einem beliebig abgegrenzten Volumen $W \subset D$, welches fest gewählt und unabhängig von der Zeit sei, muß die zeitliche Abnahme der darin enthaltenen Masse gleich der Differenz von ein- und ausströmender Masse sein. Diese Massenveränderung wird beschrieben durch:

$$\frac{\partial}{\partial t} m(W,t) = \frac{\partial}{\partial t} \int_W \rho(\underline{x},t)\, dV = \int_W \frac{\partial \rho}{\partial t}(\underline{x},t)\, dV\,.$$

Die Differenz des Massenflusses durch ein Oberflächenelement dA des Volumens W in Richtung der äußeren Normalen $\underline{n}$ von A erhält man über das Oberflächenintegral $\oint \rho\underline{u} \cdot \underline{n}\, dA$. Das Gesetzt der Massenerhaltung lautet also

$$\int_W \frac{\partial \rho}{\partial t}(\underline{x},t)\, dV \;=\; -\oint_{\partial W} \rho\underline{u} \cdot \underline{n}\, dA\,.$$

Die Anwendung des **Gaußschen Integralsatzes** liefert

$$\int_W \frac{\partial \rho}{\partial t}(\underline{x},t)\, dV = -\oint_{\partial W} \rho\underline{u} \cdot \underline{n}\, dA \overset{Gauß}{=} -\int_W \operatorname{div}(\rho\underline{u})\, dV\,.$$

Da $W \subset D$ beliebig gewählt war, führt dies zu

$$\int_W \{\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho\underline{u})\}\, dV = 0 \iff \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho\underline{u}) = 0\,. \tag{1}$$

(Differentialform der Massenerhaltung bzw. Kontinuitätsgleichung)

Zu 2.: Man unterscheidet stets zwischen zwei Arten von Kräften, die auf ein fest gewähltes Volumenteil W des Fluids wirken können:

Einerseits Druck auf die Oberfläche, also direkte Krafteinwirkung durch benachbarte Materie, und andererseits externe Kräfte wie Schwerkraft oder durch magnetische Felder verursachte Kräfte. Da wir zunächst nur eine ideale Flüssigkeit, d. h. eine Flüssigkeit ohne innere Reibung, betrachten, findet man an der Oberfläche ∂W keine Tangentialkräfte, sondern ausschließlich orthogonale Kräfte in Richtung der äußeren Normalen. Wie unter 1. ist der Druck p bzw. die Kraft $\underline{S}_{\partial W}$, die in Richtung eines äußeren Normalenvektors auf den Rand von W wirkt, mittels Gaußschem Integralsatz gegeben durch

$$\underline{S}_{\partial W} = -\oint_{\partial W} p \cdot \underline{n}\, dA = -\int_W \operatorname{grad} p\, dV\,. \tag{2}$$

Wenn $\underline{b}(\underline{x},t)$ die gegebene Anziehungskraft pro Masseneinheit bezeichnet, so ist analog die gesamte Schwerkraft

$$\underline{B} = \int_W \rho \underline{b} \, dV ,$$

so daß insgesamt auf jede Volumeneinheit folgende Kräfte einwirken:

$$\text{Kraft pro Volumeneinheit} \;=\; -\text{grad}p \;+\; \rho\underline{b} .$$

Bezeichnet man mit $\underline{x}(t) = (x(t), y(t), z(t))$ die Kurve, die ein Fluidpartikel beschreibt, so läßt sich dessen Geschwindigkeit $\underline{u}$ schreiben als $\underline{u}(\underline{x}(t),t) = \frac{d\underline{x}}{dt}(t)$, so daß man für die Beschleunigung folgende Schreibweise erhält:

$$\begin{aligned}
\frac{d\underline{u}}{dt}(\underline{x}(t)\,,\,t) \;&=\; \frac{d}{dt}\underline{u}(x(t), y(t), z(t), t) \\
&=\; \frac{\partial \underline{u}}{\partial x}\cdot \dot{x} \;+\; \frac{\partial \underline{u}}{\partial y}\cdot \dot{y} \;+\; \frac{\partial \underline{u}}{\partial z}\cdot \dot{z} \;+\; \frac{\partial \underline{u}}{\partial t} \\
&=:\; \text{grad}\underline{u}\cdot \underline{u} \;+\; \partial_t \underline{u} \\
&=:\; \frac{D\underline{u}}{Dt}, \qquad \text{wobei} \quad \frac{D}{Dt} := \underline{u}\cdot\underline{\nabla} + \partial_t \,.
\end{aligned} \tag{3}$$

Mit dem 2. Gestz von Newton (Kraft = Masse × Beschleunigung) gelangt man von hier zur Differentialform der zweiten Gleichung:

$$\rho \cdot \frac{D\underline{u}}{Dt} \;=\; -\text{grad}p \;+\; \rho\underline{b} \tag{4}$$

(Impulserhaltung)

Ein Fluid heißt inkompressibel, wenn für eine beliebige Teilmenge W des Fluids gilt, daß das Volumen $\int_W \, dV$ von W konstant in t ist, woraus man die Bedingung $\text{div}\,\underline{u} \;=\; 0$ herleiten kann, falls das Fluid von einer festen Wand umgeben ist.

Aus der Kontinuitätsgleichung

$$\begin{aligned}
\frac{\partial \rho}{\partial t} + \text{div}(\rho\underline{u}) \;&=\; \partial_t\rho + \text{grad}\rho\cdot\underline{u} + \rho\text{div}\underline{u} \\
&=:\; \frac{D\rho}{Dt} + \rho\text{div}\underline{u} \\
&=\; 0
\end{aligned} \tag{5}$$

und der Tatsache, daß stets $\rho > 0$ ist, folgt, daß die Kontinuitätsgleichung für ein inkompressibles Fluid

$$\frac{D\rho}{Dt} \;=\; 0 \tag{6}$$

lautet.

Man kann also sagen, daß ein Fluid inkompressibel ist, wenn die Massen-Dichte ρ konstant den Bewegungen des Fluids folgt. Ist ρ nicht nur unabhängig von der Zeit, sondern zusätzlich noch unabhängig vom Ort, so spricht man von einem homogenen Fluid. Eine solche Flüssigkeit, die zum Zeitpunkt $t = 0$ homogen ist, muß nicht zwingend homogen bleiben; dieses ist nur dann der Fall, wenn sie auch inkompressibel ist.

Zu 3.: Betrachtet man den Flüssigkeitsstrom in drei Dimensionen, so erhält man mit $\underline{u}(\underline{x}(t),t) = \frac{dx}{dt}(t) = (\dot{x},\dot{y},\dot{z})\,(t)$ aus den bisherigen Gleichungen ein Gleichungssystem mit 4 Gleichungen, aber 5 Unbekannten:

1. $\frac{D\rho}{Dt} + \rho \mathrm{div}\underline{u} = 0$ (Masse)
2. $\rho \cdot \frac{D\underline{u}}{Dt} = -\mathrm{grad}p + \rho \underline{b}$ (Impuls)

Die noch fehlende Gleichung gewinnt man aus dem Gesetz der Energieerhaltung.

Man nimmt an, daß die Gesamtenergie des Fluids geschrieben werden kann als:

$$E_{total} = E_{kin} + E_{int} + E_{ext}\,,$$

wobei die kinetische Energie in einem Volumen D gegeben ist durch

$$E_{kin} = \frac{1}{2}\int_D \rho\|\underline{u}\|^2\,dV\,.$$

E_{int} bezeichnet die innere Energie bzw. die Wärme, die sich aus intermolekularen Potentialen, molekularen Schwingungen u. ä. zusammensetzt. Die aus externen Kräften, wie etwa der Schwerkraft, resultierende Energie $E_{ext} = \underline{b}$ wird meist vernachlässigt, da sie nur minimal ins Gewicht fällt. Aus diesen Voraussetzungen kann man verschiedene Varianten von Energiegleichungen ableiten. Nimmt man z. B. E_{int} als konstant an, so muß E_{kin} unabhängig von Bewegung bzw. Zeit sein, um das Gleichgewicht zu erhalten, d. h. es handelt sich um einen inkompressibeln Fluß. Es gilt also

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\int_D \rho\|\underline{u}\|^2\,dV\right) = 0\,. \tag{7}$$

Für jede stetig differenzierbare Funktion $f(\underline{x},t)$ gilt

$$\frac{\partial}{\partial t}\int_D f\,dV = \int_D \frac{Df}{Dt}\,dV\,.$$

Dies liefert zusammen mit der Kettenregel und der Kontinuitätsgleichung für inkompressiblen Fluß (6):

$$\begin{aligned} 0 &= \frac{\partial}{\partial t}\left(\frac{1}{2}\int_D \rho\|\underline{u}\|^2\,dV\right) \\ &= \frac{1}{2}\int_D \left(\underbrace{\frac{D\rho}{Dt}}_{=0}\cdot\|\underline{u}\|^2 \;+\; 2\rho\underline{u}\frac{D\underline{u}}{Dt}\right) dV \\ &= \int_D \rho\cdot\underline{u}\cdot\frac{D\underline{u}}{Dt}\,dV \;=\; 0\,. \end{aligned} \tag{8}$$

Das Gleichgewicht des Momentums liefert mit $\underline{b} = 0$ weiter

$$-\int_D \underline{u}\cdot\mathrm{grad}p\,dV \;=\; 0\,. \tag{9}$$

Nimmt man sinnvollerweise an, daß das Fluid den äußeren Rand von D nicht überschreitet, also $\underline{u}\cdot\underline{n} = 0$ auf ∂D, und integriert partiell, so folgt

$$\begin{aligned} -\int_D \underline{u}\cdot\mathrm{grad}p\,dV &= \underbrace{-\oint_{\partial D}\underline{u}\cdot p\cdot\underline{n}\,dA}_{=0} + \int_D \mathrm{div}(\underline{u})p\,dV \\ &= \int_D \mathrm{div}(\underline{u})p\,dV \;=\; 0 \qquad (10) \\ &\Longrightarrow\; \mathrm{div}\,\underline{u} \;=\; 0\,, \qquad (11) \end{aligned}$$

da der Fall $p = 0$, d.h. nicht vorhandener Druck, für die hier betrachtete Fragestellung physikalisch nicht sinnvoll ist.

Also haben die **Euler-Gleichungen für ein inkompressibles Fluid** insgesamt folgende Gestalt:

$$\left.\begin{array}{rcll} \rho\cdot\frac{D\underline{u}}{Dt} &=& -\mathrm{grad}p & \text{(Impuls)} \\ \frac{D\rho}{Dt} &=& 0 & \text{(Masse)} \\ \mathrm{div}\,\underline{u} &=& 0 & \text{(Energie)} \end{array}\right\}\text{auf } D$$

$$\text{mit}\quad \underline{u}\cdot\underline{n} \;=\; 0 \quad\text{auf}\quad \partial D\,. \tag{12}$$

4.1.1 Beispiel

Schon an einem einfachen Beispiel kann man erkennen, daß die Euler-Gleichungen wegen der fehlenden Reibungskomponente nicht in der Lage sind, das Flußverhalten eines realen Fluids vollständig zu beschreiben.

Betrachtet man ein **inkompressibles**, **homogenes Fluid** beim Fluß durch eine „zweidimensionale“ Röhre in x-Richtung, so ergibt sich ein deutlicher Widerspruch. Gegebene Voraussetzungen sind also:

- $\rho = \rho_0$ konstant, da das Fluid inkompressibel und homogen ist; die Kontinuitätsgleichung entfällt also.
- Da eine Lösung der Form

$$\begin{aligned} \underline{u}(x,y,t) &= (\underline{u}(x,t),0) \\ \text{und}\, p(x,y,t) &= p(x) \end{aligned}$$

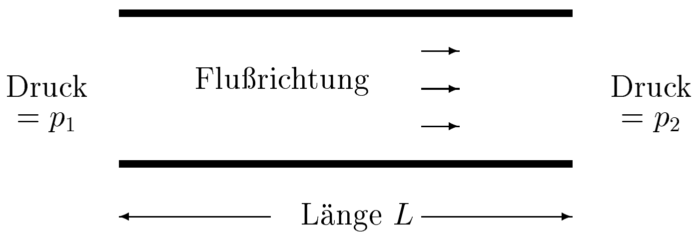
  gesucht ist, liefert die Energie-Gleichung

$$\begin{aligned} \operatorname{div}\underline{u} = 0 \;\Rightarrow\; & \partial_x \underline{u} = 0 \\ \Rightarrow\; & \underline{u}\cdot \operatorname{grad}\underline{u} = 0\,. \end{aligned}$$

- Dieses vereinfacht auch die Gleichung zur Impulserhaltung zu

$$\rho_0(\partial_t \underline{u} + \underline{u}\cdot\operatorname{grad}\underline{u}) = \rho_0\cdot\partial_t\underline{u} = -\operatorname{grad} p = \partial_x p\,. \tag{13}$$

Man beschränkt sich nun auf ein Rohstück in x-Richtung von $0 - L$ und bezeichnet die Druckverhältnisse am Rand mit $p(0) = p_1$ und $p(L) = p_2$, wobei natürlich $p_1 > p_2$ sein muß.



Für p erhalten wir also aus Gleichung (13):

$$\begin{aligned}
\rho_0 \partial_t \underline{u} &= -\partial_x p \\
\Rightarrow \quad \partial_x^2 p &= 0 \text{ mit } p(0) = p_1 \ , \ p(L) = p_2 \\
\Rightarrow \quad p(x) &= p_1 - \left(\frac{p_1 - p_2}{L} \right) \cdot x \\
\Rightarrow \quad -\partial_x p &= \frac{p_1 - p_2}{L} = \rho_0 \partial_t \underline{u} \\
\Rightarrow \quad \underline{u} &= \frac{p_1 - p_2}{\rho_0 \cdot L} \cdot t + \text{konst.}
\end{aligned}$$

D. h. bei konstantem Druckgradienten $\partial_x p$ bzw. konstantem Druckunterschied $p_1 - p_2$ wächst die Geschwindigkeit $\underline{u}$ unendlich an. Dieser Widerspruch kann nur entstehen, weil die Reibung nicht berücksichtigt wurde.

4.2 Herleitung der Navier-Stokes-Gleichungen

Im folgenden werden einige Ausdrücke zur besseren Übersichtlichkeit mit Nabla- bzw. Laplace-Operatoren bezeichnet:

Den Laplace-Operator Δ eines **skalaren Feldes** $u(D)$ erklärt man für kartesische Koordinaten durch den skalaren Wert

$$\Delta\, u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}$$

Ist $v(D) = (v_x, v_y, v_z)(D)$ ein **Vektorfeld**, so definiert man ebenfalls für kartesische Koordinaten Δv durch den Vektor

$$\Delta\, v = [(\Delta\, v_x), (\Delta\, v_y), (\Delta\, v_z)]^t$$

Der Nabla-Operator $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$ hingegen bezeichnet für ein **skalares Feld** $u(D)$ den Vektor $\nabla\, u = \operatorname{grad} u$ und für ein **Vektorfeld** $v(D) = (v_x, v_y, v_z)(D)$ andererseits entweder den skalaren Wert $\nabla\, v = \operatorname{div} v$, oder die Jacobi-Matrix von $\underline{u}$ mit der Spur $\operatorname{div} \underline{u}$.

Betrachtet man eine Ebene S innerhalb des Fluids, so wurde bis jetzt vorausgesetzt, daß alle Kräfte die auf S wirken, orthogonal zu dieser sind. Dieser Gedanke sollte nun um Tangential- bzw. Reibungskräfte erweitert werden.

Statt wie bisher anzunehmen, daß gilt

$$\text{Kraft auf } S \text{ pro Flächeneinheit } = p(\underline{x}, t) \cdot \underline{n}$$

verändert man diese Definition zu

$$\text{Kraft auf } S \text{ pro Flächeneinheit } = p(\underline{x}, t) \cdot \underline{n} + \underline{\underline{\sigma}}(\underline{x}, t) \cdot \underline{n}\,,$$

wobei die Reibungskomponente oder der Tensor $\underline{\underline{\sigma}}$ eine Matrix mit 4 wichtigen Eigenschaften ist:

1. $\underline{\underline{\sigma}} \cdot \underline{n}$ muß nicht zwingend parallel zu $\underline{n}$ sein.
2. $\underline{\underline{\sigma}}$ hängt ausschließlich von der Jacobi-Matrix $\nabla \underline{u}$ ab und ist proportional zu diesem Gradienten.
3. $\underline{\underline{\sigma}}$ ist symmetrisch.
4. $\underline{\underline{\sigma}}$ ist invariant unter Körper-Rotation und Translation, d. h. für jede orthogonale Matrix $\underline{\underline{U}}$ gilt

$$\underline{\underline{\sigma}}(\underline{\underline{U}} \cdot \nabla \underline{u} \cdot \underline{\underline{U}}^{-1}) = \underline{\underline{U}} \cdot \underline{\underline{\sigma}}(\nabla \underline{u}) \cdot \underline{\underline{U}}^{-1}\,.$$

Aus 2. und 3. folgt:

$\underline{\underline{\sigma}}$ hängt nur von dem symmetrischen Teil von $\nabla \underline{u}$ ab. Mit $\nabla \underline{u} = \underline{\underline{D}} + \underline{\underline{S}}$ und

$$\underbrace{\underline{\underline{D}} =: \frac{1}{2}[\nabla \underline{u} + (\nabla \underline{u})^t]}_{Diagonale}\,,\ \underbrace{\underline{\underline{S}} =: \frac{1}{2}[\nabla \underline{u} - (\nabla \underline{u})^t]}_{Rest}$$

folgert man nun weiter:

$\quad\ \underline{\underline{\sigma}}$ ist linear abhängig von $\underline{\underline{D}}$.

$\Rightarrow\ \underline{\underline{\sigma}}$ und $\underline{\underline{D}}$ kommutieren, d.h. $\underline{\underline{\sigma}} \cdot \underline{\underline{D}} = \underline{\underline{D}} \cdot \underline{\underline{\sigma}}$.

$\Rightarrow\ \underline{\underline{\sigma}}$ und $\underline{\underline{D}}$ können simultan diagonalisiert werden.

$\Rightarrow$ Die Eigenwerte von $\underline{\underline{\sigma}}$ sind Linearkombinationen der Eigenwerte von $\underline{\underline{D}}$.

$\overset{4.}{\Rightarrow}$ Die Eigenwerte von $\underline{\underline{\sigma}}$ müssen „symmetrisch“ sein, da man $\underline{\underline{U}}$ so wählen kann, daß zwei Eigenwerte von $\underline{\underline{D}}$ permutiert werden.

Die einzigen in diesem Sinne symmetrischen Linearkombinationen sind

$$\sigma_i \;=\; \lambda(d_1 \;+\; d_2 \;+\; d_3) \;+\; 2\mu d_i \text{ mit } i = 1,2,3\,,$$

wobei σ_i die Eigenwerte von $\underline{\underline{\sigma}}$ und d_i die Eigenwerte von $\underline{\underline{D}}$ bezeichnen. Der Parameter μ wird als 1. Viskositätskoeffizient bezeichnet.

Da die Spur der Jacobi-Matrix von $\underline{u} = (u_1, u_2, u_3)$ gegeben ist durch

$$d_1 \;+\; d_2 \;+\; d_3 \;=\; \partial_x u_1 \;+\; \partial_y u_2 \;+\; \partial_z u_3 \;=\; \mathrm{div}\underline{u}$$

läßt sich $\underline{\underline{\sigma}}$ also als Linearkombination aus $\underline{\underline{D}}$ und $(\mathrm{div}\,\underline{u})\underline{\underline{I}}$ schreiben, wobei $\underline{\underline{I}}$ die Einheitsmatrix bezeichnet.

$$\begin{aligned} \underline{\underline{\sigma}} &= \lambda(\mathrm{div}\underline{u})\underline{\underline{I}} \;+\; 2\mu\underline{\underline{D}} \\ &=: 2\mu[\,\underline{\underline{D}} \;-\; \frac{1}{3}(\mathrm{div}\underline{u})\underline{\underline{I}}\,] \;+\; \xi(\mathrm{div}\underline{u})\underline{\underline{I}} \end{aligned} \tag{14}$$

$$\begin{aligned} \text{mit }\; \mu \quad &\hat{=} \quad \text{1. Viskositätskoeffizient} \\ \text{und }\; \xi \;=\; (\lambda + \frac{2}{3}\mu) \quad &\hat{=} \quad \text{2. Viskositätskoeffizient} \end{aligned}$$

Aus der ursprünglichen Gleichung zur Impulserhaltung folgt also mit dem 2. Gesetz von Newton (Masse $\times$ Beschleunigung = Kraft)

$$\begin{aligned} \int_W \rho \cdot \frac{D\underline{u}}{Dt}\, dV &= \int_{\partial W} (-p \cdot \underline{n} \;+\; \underline{\underline{\sigma}} \cdot \underline{n})\, dA \\ \text{(Gauß)} \quad &= \int_W (-\nabla p \;+\; \nabla\underline{\underline{\sigma}})\, dV \\ \text{wobei} \quad \nabla\underline{\underline{\sigma}} &= \nabla(2\mu[\,\underline{\underline{D}} \;-\; \frac{1}{3}(\mathrm{div}\underline{u})\underline{\underline{I}}\,] \;+\; \xi(\mathrm{div}\underline{u})\underline{\underline{I}}) \\ &= 2\mu[\,\nabla\underline{\underline{D}} \;-\; \frac{1}{3}\nabla(\mathrm{div}\underline{u})\,] \;+\; \xi\nabla(\mathrm{div}\underline{u}) \\ &= 2\mu(\frac{1}{2}\Delta\underline{u} + \frac{1}{2}\nabla(\mathrm{div}\underline{u}) - \frac{1}{3}\nabla(\mathrm{div}\underline{u})) \;+\; \xi\nabla(\mathrm{div}\underline{u}) \\ &= \mu\Delta\underline{u} \;+\; (\lambda \;+\; \mu)\nabla(\mathrm{div}\underline{u}) \end{aligned}$$

$$\begin{aligned} \Rightarrow \int_W \rho\frac{D\underline{u}}{Dt}\, dV &= \int_W (-\nabla p \;+\; \mu\Delta\underline{u} \;+\; (\lambda \;+\; \mu)\nabla(\mathrm{div}\underline{u}))\, dV \\ \Rightarrow \rho\frac{D\underline{u}}{Dt} &= -\nabla p \;+\; \mu\Delta\underline{u} \;+\; (\lambda \;+\; \mu)\nabla(\mathrm{div}\underline{u})\,. \end{aligned} \tag{15}$$

Zusammen mit der Kontinuitätsgleichung und einer Ernergieerhaltungsgleichung beschreibt Gleichung (15) vollständig die Flußbewegung eines viskosen Fluids. Im Falle eines inkompressiblen, homogenen Flusses mit $\rho \equiv \rho_0 =$ konst. reduziert sich das komplette Gleichungssystem auf folgendes:

$$\begin{aligned} \frac{D\underline{u}}{Dt} &= -\mathrm{grad}\tilde{p} + \nu\Delta\underline{u} \\ & \quad \text{mit} \quad \tilde{p} = p/\rho_0 \,,\; \nu = \mu/\rho_0 \\ \mathrm{div}\,\underline{u} &= 0 \end{aligned} \tag{16}$$

Navier-Stokes-Gleichungen für inkompressiblen, homogenen Fluß

Diese Gleichungen müssen noch durch Randbedingungen ergänzt werden. Die passende Randbedingung zu den Euler-Gleichungen war $\underline{u} \cdot \underline{n} = 0$, d.h. der Fluß kann den Rand nicht kreuzen, sich aber tangential dazu bewegen. Die Navier-Stokes-Gleichungen enthalten nun den zusätzlichen Term $\nu\Delta\underline{u}$, durch den neben den 1. Ableitungen von $\underline{u}$ nun auch die 2. Ableitungen eingebracht werden. Sowohl aus experimentellen, als auch aus mathematischen Gründen fordert diese Tatsache direkt eine 2. Randbedingung, da das Ziel eine eindeutige, stetig von den Anfangswerten abhängende Lösung ist.

Beobachtet man die Grenzschichten einer fließenden viskosen Flüssigkeit in einem Rohr, so kann man erkennen, daß auch die Tangentialgeschwindigkeit gegen 0 geht, was also insgesamt die **Randbedingung**

$$\underline{u} = 0 \quad \text{auf } \partial D \,,$$

natürlich nur bei fester, unbeweglicher Wand, ergibt.

4.3 Die Reynolds-Zahl

Die oben hergeleiteten Gleichungen reichen zwar zum vollständigen Beschreiben aller Flüssigkeitsbewegungen aus, sind jedoch häufig nicht analytisch zu lösen. Die wenigen berechenbaren Strömungen haben die Eigenschaft, daß alle Strömungsgrößen wie Geschwindigkeit, Druck usw. in jedem Raum-Zeit-Punkt des Strömungsfeldes definiert und berechenbar sind. Bahn- und Stromlinien bilden glatte Raumkurven, d. h. die Strömung besteht in einem Vorübergleiten benachbarter Flüssigkeitsschichten. Man spricht deshalb hier von **laminarer Strömung**.

In den meisten für die Praxis interessanten Fällen erweist sich diese Laminarströmung aber als instabil; es ensteht eine völlig andersartige **turbulente Strömungsform**.

Hier sind dann selbst bei unveränderlichen, stationären Randbedingungen Größen wie Geschwindigkeit und Druck im festgehaltenen Raumpunkt nicht mehr zeitlich konstant, sondern schwanken schnell und völlig unregelmäßig um einen zeitlichen Mittelwert, der sich bei instationärer turbulenter Strömung sogar noch ändern kann.

Da die Gleichung zur Impulserhaltung für turbulente Strömung aus den obigen Gründen differiert, ist es notwendig, durch einfache Modellversuche festzustellen, ob es sich in dem zu untersuchenden Fall um eine laminare oder turbulente Strömung handelt. Ein Maß für die Turbulenz ist die Reynolds-Zahl Re, die den Effekt von Viskosität auf das Flußverhalten beschreibt.

Für ein gegebenes Problem werden folgende charakterisierenden Größen eingeführt:

$$\begin{aligned} L &\hat{=} \quad \textit{charakteristische Länge} \\ U &\hat{=} \quad \textit{charakteristische Geschwindigkeit} \\ T = L/U &\hat{=} \quad \textit{betrachtete Zeit} \end{aligned}$$

Dies ist eine äußerst ungenaue Definition, da z. B. bei einem Fluß durch einen Zylinder sowohl der Radius, als auch der Durchmesser als charakteristische Länge betrachtet werden könnte. Mittels dieser für das Problem typischen Größen führt man nun dimensionslose oder auch „normierte“ Variablen ein:

$$\underline{\tilde{u}} := \underline{u}/U \ , \ \underline{\tilde{x}} := \underline{x}/L \ , \ \tilde{t} := t/T$$

Die x-Komponente der Navier-Stokes-Gleichung für inkompressiblen Fluß lautet für $\underline{u} = (u, v, w)$ und $\underline{\tilde{u}} = (\tilde{u}, \tilde{v}, \tilde{w})$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} = -\frac{1}{\rho_0}\frac{\partial p}{\partial x} + \nu \left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right] . \tag{17}$$

Ein Tausch der Variablen führt zu

$$\frac{\partial \tilde{u}}{\partial \tilde{t}} + \tilde{u}\frac{\partial \tilde{u}}{\partial \tilde{x}} + \tilde{v}\frac{\partial \tilde{u}}{\partial \tilde{y}} + \tilde{w}\frac{\partial \tilde{u}}{\partial \tilde{z}} = -\frac{1}{\rho_0 U^2}\frac{\partial p}{\partial \tilde{x}} + \frac{\nu}{L \cdot U} \left[\frac{\partial^2 \tilde{u}}{\partial \tilde{x}^2} + \frac{\partial^2 \tilde{u}}{\partial \tilde{y}^2} + \frac{\partial^2 \tilde{u}}{\partial \tilde{z}^2}\right] . \tag{18}$$

Mit $\tilde{p} := p/\rho_0 U^2$, der kinetischen Viskosität $\nu := \mu/\rho_0$ und den analogen y- und z-Komponenten haben die **dimensionslosen Navier-Stokes-Gleichungen** folgende Gestalt:

$$
\begin{aligned}
\frac{\partial \underline{\tilde{u}}}{\partial \tilde{t}} + (\underline{\tilde{u}} \cdot \nabla)\, \underline{\tilde{u}} &= -\operatorname{grad} \tilde{p} + \frac{\mu}{\rho_0\, L\, U}\, \Delta\, \underline{\tilde{u}} \\
\operatorname{div} \underline{\tilde{u}} &= 0 \quad \text{auf } D \\
\underline{\tilde{u}} &= 0 \quad \text{auf } \partial D\,.
\end{aligned}
\tag{19}
$$

Man definiert die **Reynolds-Zahl** Re nun durch

$$
\begin{aligned}
Re &:= \frac{\text{charakteristische Geschwindigkeit} \times \text{Länge}}{\text{kinematische Viskosität}} \\
&= \frac{L \cdot U}{\nu} \\
&= \frac{\rho_0 \cdot L \cdot U}{\mu}\,.
\end{aligned}
$$

Sie kennzeichnet die Größenordnung des Verhältnisses von Trägheits- zu Zähigkeitskräften. Eine kleine Re-Zahl entspricht geringer Geschwindigkeit, sehr zäher Flüssigkeit oder kleinen räumlichen Ausmaßen und steht damit für ein Überwiegen der Zähigkeitskräfte. Bei großer Re-Zahl spielt die Zähigkeit eine kleinere Rolle und es kommt vor allem auf das Gleichgewicht zwischen Trägheits-, Druck- und äußeren Kräften an. Die in der Praxis interessanteren Fälle sind also diejenigen, die durch eine hohe Reynolds-Zahl ausgezeichnet sind.

Setzt man die Kraftdichte bzw. die externen Kräfte nicht vereinfacht auf 0, so müssen die obigen Gleichungen (19) noch um einen Summanden $f = f(x,t)$ ergänzt werden, der eben diese Kräfte beschreibt. Auf diese Weise erhält man die **allgemeinen Navier-Stokes-Gleichungen**, wie sie in den meisten Standardwerken zur Strömungslehre zu finden ist:

$$
\boxed{
\begin{aligned}
\tfrac{\partial \underline{u}}{\partial t} + (\underline{u} \cdot \nabla)\, \underline{u} &= -\operatorname{grad} p + \tfrac{1}{Re}\, \Delta\, \underline{u} + f \\
\operatorname{div} \underline{u} &= 0 \quad \text{auf } D \\
\underline{u} &= 0 \quad \text{auf } \partial D
\end{aligned}
}
\tag{20}
$$

5 Theoretische Vorbereitungen

Differentialgleichungen können auf verschieden Arten numerisch oder analytisch behandelt bzw. gelöst werden (vergl. [36],[11], [17], [8], [31], [9] und [10]). Eine häufig benutzte Technik sind die **Differenzenverfahren**, deren Grundidee darin besteht, auftretende Ableitungen der gesuchten Funktion durch Informationen an diskreten Stellen mittels Gitterfunktionen zu approximieren, also Ableitungen durch geeignete Differenzenquotienten näherungsweise zu ersetzen. Die Methode, die im folgenden näher betrachtet werden soll, ist die **Methode der finiten Elemente**, die sich von den Differenzenverfahren grundlegend unterscheidet:

Anstelle der punktweisen Gültigkeit wird eine unter gewissen Zusatzbedingungen äquivalente Integral- oder Variationsbedingung zugrunde gelegt, die durch die Aproximationen u_n erfüllt werden muß. Bei diesen Approximationen handelt es sich nun nicht mehr um Gitterfunktionen, sondern um Summen von „echten" Funktionen, die nur begrenzt oft differenzierbar sind, dafür aber kleine Träger besitzen. Um die Lösbarkeit der Variationaufgabe zu sichern, müssen angepaßte Funktionenräume verwendet werden, die genauere Aussagen z.B. zum Definitionsbereich der Lösung und ihren Approximationen zulassen.

5.1 Funktionenräume

Die Navier-Stokes-Gleichungen machen nur dann mathematisch und physikalisch Sinn, wenn der betrachtete fluidgefüllte Raum Ω eine offene, aber beschränkte Teilmenge des $\mathbb{R}^n$ ist mit $n = 2$ oder $n = 3$. Der Rand von Ω sei von nun an bezeichnet mit Γ, bzw. mit der Funktion $\Gamma : \mathbb{R}^n \to \mathbb{R}^{n-1}$. Im weiteren wird immer wieder eine bestimmte „Glattheitseigenschaft" von Γ verlangt, die sich folgendermaßen darstellt (vgl.: [14], S. 102):

Definition 5.1 $\Omega \subset \mathbb{R} \times \mathbb{R}^{n-1}$. Sei $\Gamma : \Omega \to \mathbb{R}^{n-1}$ eine Funktion. Man sagt, Γ genüge in $\mathbb{R}^n$ lokal einer Lipschitz-Bedingung mit der Lipschitz-Konstanten $L \geq 0$, falls jeder Punkt $(x, y) \in \Omega$ eine Umgebung U und eine davon abhängige Konstante $L \geq 0$ besitzt, so daß für alle $(x.y), (x, \tilde{y}) \in \Omega$ gilt:

$$\|\Gamma(x, y) - \Gamma(x, \tilde{y})\| \leq L\|y - \tilde{y}\| \,.$$

Im folgenden sei immer vorausgesetzt, daß $\Omega \subset \mathbb{R}^n$ offen und beschränkt mit Lipschitz-stetigem Rand ist.

Bezeichnungen Sei $U \subset \mathbb{R}^n$ eine offene Menge. Wir bezeichnen mit $\mathcal{C}(U)$ den Vektorraum aller stetigen Funktionen $f : U \to \mathbb{R}$ und mit $\mathcal{C}_0(U)$ den Untervektorraum aller Funktionen $f \in \mathcal{C}(U)$, die kompakten Träger in U haben. Für eine natürliche Zahl k sei $\mathcal{C}^k(U)$ der Vektorraum der k-mal stetig partiell differenzierbaren Funktionen $f : U \to \mathbb{R}$ sowie

$$\mathcal{C}^\infty(U) := \bigcap_{k=0}^{\infty} \mathcal{C}^k(U)\,,$$
$$\mathcal{C}_0^k(U) := \mathcal{C}_0(U) \cap \mathcal{C}^k(U)\,.$$

Für die Beweise einiger Existenzsätze sind geeignete Hilberträume von Nöten, z.B. die Sobolevräume, deren Verständnis nur durch den Weg über die Lebesguesche Theorie gewährleistet werden kann. Im Unterschied zur Riemannschen Theorie werden Ober- und Unterintegrale hier mit Hilfe der halbstetigen Funktionen aus $\mathcal{H}^\uparrow(\mathbb{R}^n)$ bzw. $\mathcal{H}^\downarrow(\mathbb{R}^n)$ anstelle der Treppenfunktionen definiert.

Definition 5.2 Die Menge aller Funktionen $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, die sich als Limiten monoton wachsender (bzw. fallender) Funktionenfolgen f_ν aus $\mathcal{C}_0(\mathbb{R}^n)$ darstellen lassen, d.h.

$$f(x) := \lim_{\nu\to\infty} f_\nu(x) \quad \text{für alle} \quad x \in \mathbb{R}^n\,,$$

bezeichnet man mit $\mathcal{H}^\uparrow(\mathbb{R}^n)$ (bzw. $\mathcal{H}^\downarrow(\mathbb{R}^n)$) oder kurz $\mathcal{H}^\uparrow$ (bzw. $\mathcal{H}^\downarrow$).

Definition 5.3 (Oberintegral, Unterintegral) Sei $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ eine beliebige Funktion. Dann setzt man

$$\int^* f(x)dx \;:=\; \inf\left\{\int_{\mathbb{R}^n} \varphi(x)dx \;:\; \varphi \in \mathcal{H}^\uparrow,\, \varphi \geq f\right\}\,,$$
$$\int_* f(x)dx \;:=\; \sup\left\{\int_{\mathbb{R}^n} \psi(x)dx \;:\; \psi \in \mathcal{H}^\downarrow,\, \psi \leq f\right\}\,.$$

Wie in der Riemannschen Theorie heißt eine Funktion nun integrierbar, falls Ober- und Unterintegral übereinstimmen.

Definition 5.4 Eine Funktion $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ heißt Lebesgue-integrierbar, falls

$$-\infty \;<\; \int_* f(x)dx \;=\; \int^* f(x)dx \;<\; +\infty\,.$$

Der gemeinsame Wert des Ober- und Unterintegrals heißt dann das Lebesgue-Integral von f und wird mit $\int f(x)dx$ bezeichnet.

Bezeichnung Wir bezeichnen mit $\mathcal{L}_1(\mathbb{R}^n)$ die Menge aller Lebesgue-integrierbaren Funktionen $f : \mathbb{R}^n \to \mathbb{R}$, ohne die Werte $\pm\infty$. Mit den Rechenregeln der Integralrechnung wird $\mathcal{L}_1(\mathbb{R}^n)$ zu einem Vektorraum, auf dem das Integral ein lineares, monotones Funktional darstellt.

Der letzte Schritt auf dem Weg zu dem ersten Ziel, den Lebesgue-Räumen, ist die Definition meßbarer Mengen, bzw. meßbarer Funktionen:

Definition 5.5 Sei $M \subset \mathbb{R}^n$ eine beliebige Teilmenge. Eine Funktion $f : M \to \mathbb{R} \cup \{\pm\infty\}$ heißt Lebesgue-meßbar auf M, falls die trivial fortgesetzte Funktion $\tilde{f}$ mit

$$\tilde{f}(x) := \begin{cases} f(x) & \text{für} \quad x \in M \\ 0 & \text{für} \quad x \in \mathbb{R}^n \backslash M \end{cases}$$

über $\mathbb{R}^n$ Lebesgue-integrierbar ist. Man setzt dann

$$\int_M f(x)dx := \int_{\mathbb{R}^n} \tilde{f}(x)dx \,.$$

Definition 5.6 (Lebesgue-Räume) Sei Ω eine offene Teilmenge des $\mathbb{R}^n$, $n \in \mathbb{N}$, $p \in \mathbb{R}$, $m \geq 0$, $p \geq 1$. Man bezeichnet mit $L^p(\Omega)$ den Lebesgueschen Raum. Die Elemente dieses Raumes sind Klassen äquivalenter Funktionen $f(x)$, die auf Ω Lebesgue-meßbar sind und für die $|f(x)|^p$ Lebesgue-summierbar ist, d.h.

$$\|f\|_p = \left[\int_\Omega |f(x)|^p dx\right]^{1/p} \neq 0 \ \text{ für } 1 \leq p < \infty$$

$$\text{und} \quad \|f\|_\infty = \inf\{K \ : \ |f(x)| \leq K \quad \text{fast überall}\} \neq 0 \,.$$

Um letztendlich Sobolevräume definieren zu können, benötigen wir noch die Definition einer weiteren Norm. Mit $\alpha = (\alpha_1, \ldots, \alpha_n)$, $\alpha_i \in \mathbb{N}$, $|\alpha| = \alpha_1 + \ldots + \alpha_n$ und

$$D^\alpha = D_1^{\alpha_1} \cdots D_n^{\alpha_n} = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$$

hat diese Norm folgende Gestalt:

$$\|u\|_{m,p} = \left[\sum_{0 \leq |\alpha| \leq m} (\|D^\alpha u\|_p)^p\right]^{1/p} , \ 1 \leq p < \infty$$

$$\text{und} \quad \|u\|_{m,\infty} = \max_{|\alpha| \leq m} \|D^\alpha u\|_\infty \,.$$

Definition 5.7 (Sobolev-Räume durch Vervollständigung) Unter der Vervollständigung eines metrischen Raumes X versteht man einen über Cauchy-Folgen (wie bei der Vervollständigung von $\mathbb{Q}$ zu $\mathbb{R}$) gewonnenen vollständigen Raum $\tilde{X}$, der zu X isometrisch und isomorph ist. Auf diese Weise erhält man die erste Definition von Sobolev-Räumen.

$$\begin{array}{lll} H^{m,p}(\Omega) & := & \text{Vervollständigung von } \{u \in \mathcal{C}^\infty(\Omega) : \|u\|_{m,p} < \infty\} \text{ bzgl. } \|\cdot\|_{m,p} \\ H_0^{m,p}(\Omega) & := & \text{Vervollständigung von } \{u \in \mathcal{C}_0^\infty(\Omega) : \|u\|_{m,p} < \infty\} \text{ bzgl. } \|\cdot\|_{m,p} \end{array}$$

Obwohl diese Definition von $H^{m,p}(\Omega)$ und $H_0^{m,p}(\Omega$ sehr einfach und häufig nützlich ist, hat sie den Nachteil, daß die funktionstheoretischen Eigenschaften der Elemente von $H^{m,p}(\Omega)$ nicht geklärt sind. Deshalb wird zusätzlich ein anderer Weg der Definition über schwache Ableitungen eingeschlagen.

Sei Ω offene Teilmenge des $\mathbb{R}^n$. Sei $X := \mathcal{C}_0^\infty(\Omega)$, also die Menge aller unendlich oft differenzierbaren Funktionen Φ auf Ω mit kompaktem Träger $\text{supp}(\Phi) \subset \Omega$.

Ist $K \subset \Omega$ kompakt, so definiert man

$$\mathcal{D}_K(\Omega) := \{\Phi \in X \ : \ \text{supp}(\Phi) \subset K\}$$

Mit der Seminorm

$$p_{K,m}(\Phi) := \sup_{|\alpha| \le m_{m \in K}} |D^\alpha \Phi|$$

und der Nullumgebungsbasis

$$U(\epsilon_1, \ldots, \epsilon_n \,;\, m_1, \ldots, m_n \,;\, K) = \{\Phi \in \mathcal{D}_K(\Omega) : p_{K,m_j}(\Phi) \le \epsilon_j \ , \ 1 \le j \le n\}$$

wird $\mathcal{D}_K(\Omega) := X_K$ zum lokalkonvexen topologischen Vektorraum.

Definition 5.8 Mit der induktiven Topologie τ bzgl. $\{X_K\}$ ist (X, τ) eine lokalkonvexer topologischer Raum, der mit $\mathcal{D}(\Omega)$ bezeichnet wird. Der duale Raum $\mathcal{D}'(\Omega)$ zu $\mathcal{D}(\Omega)$ besteht aus allen stetigen linearen Funktionalen auf $\mathcal{D}(\Omega)$. Ein Element $T \in \mathcal{D}'(\Omega)$ heißt Distribution, eine Element $\phi \in \mathcal{D}(\Omega)$ heißt Testfunktion.

Definition 5.9 Sei $T \in \mathcal{D}'(\Omega)$, α ein Multiindex. Die Ableitung $D^\alpha T$ ist die Distribution

$$(D^\alpha T)(\phi) = (-1)^{|\alpha|} T(D^\alpha \phi) \,.$$

Sei nun $u \in L_1^{loc}(\Omega)$, d.h. $u|_A \in L_1(A)$ für jede meßbare Teilmenge $A \subset \Omega$, für die $\bar{A} \subset \Omega$ und $\bar{A}$ kompakt ist. Sei außerdem

$$T_u \phi := \int_\Omega u(x)\phi(x)dx \ .$$

Das Funktional T_u ist nach Definition linear und sogar stetig, d.h. T_u ist Distribution bzw. $T_u \in \mathcal{D}'(\Omega)$.

Definition 5.10 Sei $u \in L_1^{loc}(\Omega)$. Die Funktion $v \in L_1^{loc}(\Omega)$ heißt schwache Ableitung $D^\alpha u$ von u, falls

$$T_v = D^\alpha(T_u)\,,$$

d.h.

$$\int_\Omega v(x)\phi(x)dx \;=\; (-1)^{|\alpha|}\int u(x)D^\alpha\phi(x)dx \quad \text{für alle } \phi \in \mathcal{D}(\Omega).$$

Dieses führt nun zur gesuchten Definition von Sobolev-Räumen:

Definition 5.11 (Sobolev-Räume über schwache Ableitungen) Sei $m \in \mathbb{N}_0$, $p \in \mathbb{R}$, $p \geq 1$, $\Omega \subset \mathbb{R}^n$. $W^{m,p}(\Omega)$ sei die Menge aller Funktionen u, die schwache Ableitungen $D^\alpha u \in L^p(\Omega)$ besitzen. Man setzt

$$\|u\|_{m,p} \;=\; \left[\sum_{|\alpha|\leq m}(\|D^\alpha u\|_p)^p\right]^{1/p},$$

womit $W^{m,p}(\Omega)$ zum Banachraum wird.

Um sich das Verhältnis zwischen $W^{m,p}(\Omega)$ und $H^{m,p}(\Omega)$ verständlich zumachen, betrachtet man den Raum $X := \mathcal{C}^\infty(\Omega) \cap H^{m,p}(\Omega)$. X ist dicht in $H^{m,p}(\Omega)$. Jedes Element $x \in X$ liegt auch in $W^{m,p}(\Omega)$. Es folgt:

$$H^{m,p}(\Omega) \subset W^{m,p}(\Omega)$$

Für $1 \leq p < \infty$ gilt sogar

$$H^{m,p}(\Omega) \;=\; W^{m,p}(\Omega).$$

Für die Konvergenzuntersuchungen bei Randwertaufgaben zweiter bzw. vierter Ordnung sind vor allem die Sobolev-Räume $H^{1,2}(\Omega) \;=\; W^{1,2}(\Omega)$ und $H^{2,2}(\Omega) \;=\; W^{2,2}(\Omega)$ von Bedeutung. Der Raum $H^{m,2}(\Omega)$ bildet mit dem Skalarprodukt

$$(u,v) \;:=\; \int_\Omega(\sum_{|\alpha|\leq m} D^\alpha u D^\alpha v)dx$$

einen Hilbert-Raum, d.h. einen vollständigen normierten Raum mit Skalarprodukt, der alle im weiteren verlangten Eigenschaften bieten wird.

5.2 Variationsgleichungen

5.2.1 Beispiel 1: Lineares Randwertproblem

Als erstes und einfachstes Beispiel einer Variationsgleichung betrachten wir zunächst das lineare Randwertproblem

$$-u''(x) + d(x)u'(x) + c(x)u(x) = f(x) \quad \text{in } \Omega := (0,1) \qquad (21)$$

$$\text{mit } u(0) = u(1) = 0\,.$$

Dabei seien d, c und f gegebene stetige Funktionen und das Problem besitze eine Lösung, d.h. es existiere eine Funktion $u \in \mathcal{C}^2(\Omega)$, die der Differentialgleichung genügt. Damit gilt für diese Lösung u trivialerweise auch

$$\int_\Omega (-u'' + du' + cu)\varphi\,dx \;=\; \int_\Omega f\varphi\,dx$$

für beliebige Funktionen $\varphi \in \mathcal{C}_0^\infty(\Omega)$. Genügt umgekehrt eine Funktion $u \in \mathcal{C}^2(\Omega)$ der Integralgleichung, so genügt u auch der zugrunde gelegten Differentialgleichung. Integriert man nun partiell, so erhält man zuerst

$$-u'\varphi|_{x=0}^{1} \;+\; \int_\Omega u'\varphi'\,dx \;+\; \int_\Omega (du' + cu)\varphi\,dx \;=\; \int_\Omega f\varphi\,dx$$

und unter Verwendung der Randwerte dann schließlich

$$\int_\Omega u'\varphi'\,dx \;+\; \int_\Omega (du' + cu)\varphi\,dx \;=\; \int_\Omega f\varphi\,dx.$$

Also hat die zugehörige **Variationsgleichung** die Form

$$a(u,\varphi) \;=\; F(\varphi) \quad \text{für alle } \varphi \in \mathcal{C}_0^\infty(\Omega)\,, \tag{22}$$

wobei

$$\begin{aligned} a(u,\varphi) &:= \int_\Omega [u'\varphi' + (du' + cu)\varphi]\,dx\,, \\ F(\varphi) &:= \int_\Omega f\varphi\,dx\,. \end{aligned}$$

5.2.2 Beispiel 2: Poissonsche Differentialgleichung

Als nächstes Beispiel einer einfachen elliptischen Randwertaufgabe bietet sich die *Poissonsche Differentialgleichung* mit homogenen Dirichlet-Randwerten an:

$$-\Delta\, u(x_1,x_2) \;=\; f(x_1,x_2) \quad \text{in } \Omega \subset \mathbb{R}^2 \tag{23}$$

$$\text{mit } u|_\Gamma \;=\; 0\,, \quad \text{wobei } \Gamma =: \partial\Omega \quad \text{stückweise glatt ist.}$$

Mit entsprechenden Funktionen $\varphi \in \mathcal{C}_0^\infty(\Omega)$ erhält man analog zum ersten Beispiel mittels partieller Integration eine zugehörige schwache Variante:

$$\begin{aligned}
& -\int_\Omega \Delta u \varphi \, dx \;=\; \int_\Omega f \varphi \, dx \\
\Rightarrow \quad & -\left(\frac{\partial u}{\partial x_1} + \frac{\partial u}{\partial x_2}\right)\varphi\Big|_\Gamma + \int_\Omega \left(\frac{\partial u}{\partial x_1}\frac{\partial \varphi}{\partial x_1} + \frac{\partial u}{\partial x_2}\frac{\partial \varphi}{\partial x_2}\right) dx \;=\; \int_\Omega f \varphi \, dx \\
\Rightarrow \quad & \int_\Omega \left(\frac{\partial u}{\partial x_1}\frac{\partial \varphi}{\partial x_1} + \frac{\partial u}{\partial x_2}\frac{\partial \varphi}{\partial x_2}\right) dx \;=\; \int_\Omega f \varphi \, dx \\
\Rightarrow \quad & a(u,\varphi) \;=\; F(\varphi) \quad \text{für alle } \varphi \in \mathcal{C}_0^\infty(\Omega) \text{ mit } \varphi|_\Gamma = 0 \qquad (24)
\end{aligned}$$

wobei

$$\begin{aligned}
a(u,\varphi) &:= \int_\Omega \left(\frac{\partial u}{\partial x_1}\frac{\partial \varphi}{\partial x_1} + \frac{\partial u}{\partial x_2}\frac{\partial \varphi}{\partial x_2}\right) dx \,, \\
F(\varphi) &:= \int_\Omega f \varphi \, dx \,.
\end{aligned}$$

5.2.3 Ausblick auf Lösungsmethoden

Erweitert man nun die Definitionsbereiche der Abbildungen a und F auf einen geeigneten Hilbert-Raum H, so lassen sich mit der zugehörigen Norm Aussagen formulieren, die die Existenz und Eindeutigkeit einer Lösung $u \in H$ mit

$$a(u,\varphi) \;=\; F(\varphi) \quad \text{für alle } \varphi \in H$$

sichern. Zur Lösung eines Variationsproblems kann man das klassische Galerkin-Verfahren anwenden, dessen Grundidee sich folgendermaßen darstellt (vgl.: [10], S. 116):

1. Wähle $H_n = span\{h_1, \ldots, h_n\} \subset H$.
2. Bestimme $u_n \in H_n$ mit $a(u_n, \varphi_n) = F(\varphi_n)$ für alle $\varphi_n \in H_n$.

Es kommt also darauf an, endlichdimensionale Untervektorräume H_n von H zu finden, die einerseits geometrisch einfach sind und andererseits gute Approximationseigenschaften besitzen. Diese Eigenschaften werden besonders gut von den Räumen finiter Elemente erfüllt. Hierbei wird das Gebiet Ω in möglichst gleichmäßige Polyeder, z.B. Dreiecke oder Quadrate im zweidimensionalen Fall, zerlegt (bzw. durch eine solche Zerlegung approximiert, falls der Rand von Ω gekrümmt ist). Finite Elemente sind dann stückweise polynomiale Funktionen von vorgegebenem Grad, im einfachsten Fall also stückweise lineare Funktionen auf Dreiecken. (vgl.: [20], S. 171)

Die weitere Vorgehensweise besteht dann darin, die Basisfunktionen h_k so zu wählen, daß jedes h_k

1. einen kleinen kompakten Träger besitzt,
2. stückweise glatt ist und
3. durch einfache Funktionen, z.B. Polynome, definiert ist.

Die Approximationsgüte hängt dann einerseits von der Polynomordnung und andererseits von der Feinheit der gewählten Gebietszerlegung ab. Meist ist es sinnvoll, die Polynomordnung festzuhalten, also z. B. nur stückweise lineare oder quadratische Elemente zuzulassen, und dafür die Zerlegung genügend zu verfeinern.

5.3 Das Variationsproblem im Hilbert-Raum

Die Existenz und Eindeutigkeit der Lösung u einer Variationsgleichung wird in einem geeigneten Hilbert-Raum, z. B. in dem Sobolev-Raum $H^{m,p}(\Omega)$, durch den Satz von Lax-Milgram gesichert, der einige Bedingungen an die Abbildungen a und F stellt. Diese sollen zuvor definiert werden:

Definition 5.12 Sei H ein Hilbert-Raum mit zugehörigem Skalarprodukt $(\cdot,\cdot)$ und zugehöriger Norm $\|\cdot\|$, a eine Bilinearform $H \times H \to \mathbb{R}$ und F ein lineares Funktional $H \to \mathbb{R}$. F heißt **beschränkt**, wenn es eine Konstante C gibt, so daß

$$\|Fx\| \leq C\|x\| \quad \text{für alle } \; x \in H\,,$$

und die Norm von F ist definiert durch

$$\|F\| := \sup_{x \in H, x \neq 0} \frac{\|Fx\|}{\|x\|}\,.$$

a heißt **beschränkt**, wenn es eine Konstante K gibt, so daß

$$a(u,v) \leq K\,\|u\|\,\|v\| \quad \text{für alle } \; u,v \in H\,,$$

und die Norm von a ist definiert durch

$$\|a\| := \sup \frac{a(u,v)}{\|u\| \cdot \|v\|}\,.$$

a heißt **symmetrisch**, falls für alle $u, v \in H$ $a(u,v) = a(v,u)$ gilt und a heißt **koerziv**, falls es eine Konstante $\alpha > 0$ gibt, so daß

$$a(u,u) \geq \alpha \|u\|^2 \quad \text{für alle } \; u \in H\,.$$

Von zentraler Bedeutung für den Beweis den Satzes von Lax-Milgram ist der Rieszsche Darstellungssatz:

Satz 5.1 (Rieszscher Darstellungssatz) Jedes lineare beschränkte Funktional F auf einem Hilbert-Raum H ist von der Form

$$F(\varphi) = (\varphi, b)\,,$$

wobei $b \in H$ durch F eindeutig bestimmt ist. Umgekehrt definiert (φ, b) für jedes $b \in H$ ein beschränktes, lineares Funktional auf H. Es gilt $\|F\| = \|b\|$.

(zum Beweis siehe [10], S. 120)

5.3.1 Satz von Lax-Milgram

Satz 5.2 (Lax-Milgram) Sei a eine beschränkte, koerzive Bilinearform auf dem Hilbert-Raum H und F ein beschränktes lineares Funktional auf H. Dann gibt es ein eindeutiges Element $u \in H$, so daß

$$a(\varphi, u) \;=\; F(\varphi) \quad \text{für alle } \; \varphi \in H\,.$$

Ferner gilt die Abschätzung $\|u\| \leq \frac{1}{\alpha}\|F\|$ für ein $\alpha > 0$.

Beweis: Sei $v \in H$. Die Abbildung

$$\varphi \longmapsto a(\varphi, v)$$

ist ein beschränktes lineares Funktional auf H. Aufgrund des Rieszschen Satzes existiert ein eindeutiges $u_v \in H$ mit

$$a(\varphi, v) = (\varphi, u_v) \quad \text{für alle } \;\; \varphi \in H\,.$$

Die Abbildung $T : H \to H$ mit $v \mapsto u_v$ ist beschränkt und linear:

- **Beschränktheit:** Nach Definition gilt

  $$\|u_v\|^2 = (u_v, u_v) = a(u_v, v) \leq \|a\| \cdot \|u_v\| \cdot \|v\|\,,$$

  so daß mit der Beschränktheit von a folgt:

  $$\|Tv\| = \|u_v\| \leq \|a\| \cdot \|v\| \leq K \cdot \|v\|$$

  für eine Konstante $K \in \mathbb{R}$.

- **Linearität:** Für $v_1, v_2 \in H$ gilt

$$a(\varphi, v_1) = (\varphi, Tv_1) \quad \text{und} \quad a(\varphi, v_2) = (\varphi, Tv_2)\,.$$

Dann gilt aufgrund der Bilinearität von a

$$a(\varphi, v_1 + v_2) = (\varphi, Tv_1 + Tv_2)\,,$$

$$\text{d.h.} \quad T(v_1 + v_2) = Tv_1 + Tv_2\,.$$

Durch die Koerzvität von a folgt schließlich noch, daß

$$\alpha\|v\|^2 \;\leq\; a(v,v) \;=\; (v, Tv) \;\leq\; \|v\| \cdot \|Tv\|\,,$$

$$\Rightarrow \;\|Tv\| \;\geq\; \alpha\|v\|\,.$$

T ist also injektiv und T^{-1} ist beschränkt; der Wertebereich $R(T)$ ist abgeschlossen. Damit folgt bereits $R(T) = H$, d.h. T ist surjektiv. Sonst gäbe es aufgrund des Projektionssatzes (siehe z.B. [10], S. 119) mit $M = R(T)$ ein Element

$$z \;\in\; M^{\perp} \subset H \;\;,\;\; z \neq 0\,.$$

D.h. für alle $v \in H$ würde $a(z, v) = (z, Tv)$ gelten, insbesondere für $z = v$:

$$a(z, z) = (z, Tz) = 0 \;\Rightarrow\; z = 0\,,$$

was den erwünschten Widerspruch zur Voraussetzung liefert. Die inverse Abbildung T^{-1} ist deshalb eine lineare, beschränkte Abbildung von H auf H. Also gilt:

$$a(\varphi, T^{-1}w) = (\varphi, w) \quad \text{für alle} \quad \varphi, w \in H\,.$$

Die Anwendung des Rieszschen Satzes auf die Abbildung F liefert ein **eindeutiges** Element $\tilde{u} \in H$, für das gilt:

$$F(\varphi) \;=\; (\varphi, \tilde{u}) \quad \text{für alle} \quad \varphi \in H\,.$$

Es folgt für eben dieses $\tilde{u} \in H$ und für alle $\varphi \in H$

$$F(\varphi) \;=\; (\varphi, \tilde{u}) \;=\; a(\varphi, T^{-1}\tilde{u})\,.$$

Setzt man nun $u \;:=\; T^{-1}\tilde{u}$, so hat man eine eindeutige Lösung u der Variationsgleichung erhalten. (q.e.d.)

5.3.2 Anwendung auf ein Modellproblem

Wir betrachten erneut die Poissongleichung mit homogenen Dirichlet Randbedingungen bzw. die zugehörige Variationsgleichung (siehe Beispiel 2): $a(u,\varphi) = F(\varphi)$, wobei

$$\begin{aligned} a(u,\varphi) &:= \int_\Omega \left(\frac{\partial u}{\partial x_1}\frac{\partial \varphi}{\partial x_1} + \frac{\partial u}{\partial x_2}\frac{\partial \varphi}{\partial x_2} \right) dx\,, \\ F(\varphi) &:= \int_\Omega f\varphi\, dx\,, \end{aligned}$$

In diesem Beispiel ist der geeignete Hilbert-Raum $H = H_0^{1,2}(\Omega)$ mit der Norm

$$\|u\|_{1,2} := \left[\int_\Omega \sum_{|\alpha|\le 1} |D^\alpha u(t)|^2\, dt \right]^{1/2}\,,$$

welche äquivalent ist zu der Halbnorm

$$|u|_{1,2} := \left[\int_\Omega \sum_{|\alpha|=1} |D^\alpha u(t)|^2\, dt \right]^{1/2}\,,$$

wie die Poincarésche Ungleichung zeigt:

Satz 5.3 (Poincarésche Ungleichung) Sei $\Omega \subset \mathbb{R}^n$, $d > 0$. Ω liege zwischen den Ebenen $x_n = 0$ und $x_n = d$ (d.h. Ω ist Lebesgue-meßbar). Dann gibt es eine Konstante K mit

$$|u|_{m,p} \le \|u\|_{m,p} \le K \cdot |u|_{m,p} \quad \text{für alle} \quad u \in H_0^{m,p}(\Omega)$$

(Zum Beweis siehe z.B. [20], S. 159.)

Für das betrachtete Modellproblem der Poissongleichung liefert uns das

$$\begin{aligned} a(u,u) &= \int_\Omega \left(\frac{\partial u}{\partial t_1}\right)^2 + \left(\frac{\partial u}{\partial t_2}\right)^2 dt \\ &= \int_\Omega \sum_{|\alpha|=1} |D^\alpha u|^2\, dt \\ &= (|u|_{1,2})^2 \end{aligned}$$

und damit die geforderte **Koerzivität** der Abbildung a:

$$a(u,u) = (|u|_{1,2})^2 \ge \lambda \cdot (\|u\|_{1,2})^2$$

für ein $\lambda > 0$.

Die nächste geforderte Eigenschaft erhält man aus der Hölderschen Ungleichung:

Satz 5.4 (Höldersche Ungleichung) Ist $u \in L^p(\Omega)$, $v \in L^q(\Omega)$, $1/p + 1/q = 1$, so gilt

$$\int_\Omega u\, v \;\leq\; \|u\|_p \cdot \|v\|_q$$

(Zum Beweis siehe z.B. [20], S. 278.)

Wendet man diese Ungleichung auf a an, erhält man die **Beschränktheit** von a:

$$\begin{aligned}
a(u,\varphi) &= \int_\Omega Du \cdot D\varphi \, dt \\
&\leq \|Du\|_2 \cdot \|D\varphi\|_2 \\
&= |u|_{1,2} \cdot |\varphi|_{1,2} \\
\Rightarrow a(u,\varphi) &\leq C \cdot \|u\|_{1,2} \cdot \|\varphi\|_{1,2} \quad \text{für eine Konstante} \quad C \in \mathbb{R}.
\end{aligned}$$

Da die Bilinearität von a offensichtlich ist, sind alle Anforderungen an a erfüllt und es bleibt nun noch die Abbildung F zu betrachten.

Für $f \in L^2(\Omega)$ stellt

$$F : H_0^{1,2}(\Omega) \to \mathbb{R} \quad \text{mit } \varphi \mapsto \int_\Omega f \varphi \, dx$$

ein beschränktes lineares Funktional dar. Es ist nämlich

$$\|F\| := \sup_{\varphi \neq 0} \frac{|F\varphi|}{\|\varphi\|} \leq \|f\|_2 < \infty \,,$$

denn nach der Hölderschen und der Poincaréschen Ungleichung gilt:

$$F\varphi = \int_\Omega f \varphi \leq \|f\|_2 \cdot \|\varphi\|_2 \leq \|f\|_2 \cdot \|\varphi\|_{1,2} \,.$$

Also sind für $f \in L^2(\Omega)$ alle Voraussetzungen für den Satz von Lax-Milgram gegeben und es existiert somit eine eindeutige Lösung des Variationsproblems der Poissonschen Differentialgleichung.

5.3.3 Das Optimierungsproblem

Die Variationsgleichung $a(u,\varphi) = F(\varphi)$ kann auch als Optimierungsproblem gestellt werden, wie der folgende Satz zeigt:

Satz 5.5 Sei a eine beschränkte, symmetrische, koerzive Bilinearform auf dem Hilbert-Raum H, $F(\varphi) := (b,\varphi)$ mit $b \in H$ und J definiert als

$$J(\varphi) \;:=\; \frac{1}{2}a(\varphi,\varphi) - F(\varphi)\,.$$

Es gilt: Ein Element $u \in H$ löst das Optimierungsproblem

$$J(\varphi) \longrightarrow \quad \min ! \quad \text{bei} \quad \varphi \in H$$

genau dann, wenn u der Variationsgleichung $a(u,\varphi) = F(\varphi)$ genügt. Es existiert genau ein $u \in H$, welches $J(\varphi)$ minimiert.

Beweis: Sei zunächst u Lösung der Variationsgleichung und $\varphi \in H$. Nach Definition gilt

$$\begin{aligned}
J(\varphi) - J(u) &= \frac{1}{2}a(\varphi,\varphi) - F(\varphi) \;-\; \frac{1}{2}a(u,u) + F(u)\\
&= \frac{1}{2}a(\varphi-u,\varphi-u) \;+\; a(u,\varphi-u) \;-\; F(\varphi-u)\\
&= \frac{1}{2}a(\varphi-u,\varphi-u) \;\geq\; 0\\
\Rightarrow\; J(\varphi) &\geq J(u) \quad \text{für alle} \quad \varphi \in H\,,
\end{aligned} \tag{25}$$

also löst u das Optimierungsproblem.

Genügt nun umgekehrt u dem Optimierungsproblem, so folgt

$$J(u) \leq J(u+\lambda\varphi) \quad \text{für alle} \quad \lambda \in \mathbb{R}\,,\; \varphi \in H$$

und damit nach Definition

$$\begin{aligned}
\frac{1}{2}a(u,u) - F(u) - \frac{1}{2}a(u+\lambda\varphi,u+\lambda\varphi) + F(u+\lambda\varphi) &\leq 0\\
\Rightarrow \frac{1}{2}\lambda^2 a(\varphi,\varphi) + \lambda a(u,\varphi) - \lambda F(\varphi) &\geq 0\\
\Rightarrow \frac{1}{2}\lambda a(\varphi,\varphi) + a(u,\varphi) - F(\varphi) &\geq 0\,.
\end{aligned} \tag{26}$$

Betrachtet man nun den Grenzübergang von $\lambda \downarrow 0$, so erhält man, daß einerseits

$$a(u,\varphi) - F(\varphi) \geq 0\,,$$

und andererseits für $\lambda \uparrow 0$, daß

$$a(u,\varphi) - F(\varphi) \leq 0\,.$$

Also ist u Lösung der Variationsgleichung $a(u,\varphi) = F(\varphi)$. Die Eindeutigkeit dieser Lösung folgt direkt aus dem Satz von Lax-Milgram. (q.e.d.)

6 Diskretisierung der Probleme

Wie im vorherigen Abschnitt gezeigt, ist die Lösung einer klassischen Randwertaufgabe auch Lösung der schwachen Formulierung bzw. Variationsgleichung. Umgekehrt liefern Lösungen der Variationsgleichung unter zusätzlichen Regularitätseigenschaften (entsprechende Glattheit) auch Lösungen des passenden Randwertproblems. D. h. man löst die Problemstellung in geeigneten Hilbert-Räumen, wie $H_0^{m,p}(\Omega)$, um die Lösung einem solchen Raum zuordnen zu können und so Aussagen über die Glattheit der Lösung zu gewinnen.

Da nur in den seltensten Fällen eine Lösung explizit durch Integration o. ä. berechnet werden kann, ist man auf approximative numerische Verfahren angwiesen. Diese können nur in endlichen Dimensionen arbeiten, so daß man gezwungen ist, Probleme in unendlich dimensionalen Räumen durch solche in endlich dimensionalen zu beschreiben.

Die im Folgenden beschriebenen numerischen Techniken basieren auf der approximativen Lösung der Variationsgleichung

$$a(u, \varphi) \; = \; F(\varphi) \quad \text{für alle} \quad \varphi \in H \tag{27}$$

bzw. des Optimierungsproblems

$$J(\varphi) = \frac{1}{2} a(\varphi, \varphi) \; - \; F(\varphi) \longrightarrow \min ! \quad \text{bei} \quad \varphi \in H \, , \tag{28}$$

indem anstelle des zugrunde liegenden unendlich dimensionalen Raumes H ein Teilraum

$$H_n \subset H$$

mit $dim\, H_n < \infty$ gewählt wird. Statt des ursprünglichen Problems sind dann Aufgaben der Form

$$J(\varphi_n) = \frac{1}{2} a(\varphi_n, \varphi_n) \; - \; F(\varphi_n) \longrightarrow \min ! \quad \text{bei} \quad \varphi_n \in H_n \tag{29}$$

zu lösen. Da wegen $H_n \subset H$ auch H_n mit dem gleichen Skalarprodukt $(\cdot, \cdot)$ einen Hilbert-Raum bildet und $a(\cdot, \cdot)$ über H_n ebensolche Eigenschaften wie über H besitzt, läßt sich die abstrakte Theorie auch auf das endlich dimensionale Problem anwenden. Das diskretisierte Problem besitzt somit nach Lax-Milgram eine eindeutige Lösung $u_n \in H_n$. Diese läßt sich also auch durch die notwendige und hinreichende Bedingung

$$a(u_n, \varphi_n) \; = \; F(\varphi_n) \quad \text{für alle} \quad \varphi_n \in H_n \tag{30}$$

charakterisieren. Das auf dem diskreten Optimierungsproblem (29) basierende Verfahren heißt **Ritz-Verfahren**, das auf der diskreten Variationsgleichung (30) basierende Verfahren wird **Galerkin-Verfahren** genannt. Auf das Ritz-Verfahren wird in weiteren nicht näher eingegangen, da das Galerkin-Verfahren für das hier betrachtete Problem geeigneter ist.

6.1 Galerkin-Verfahren

Zur Konvergenz des Galerkin-Verfahrens gilt der häufig als **Lemma von Cea** bezeichnete folgende Satz:

Satz 6.1 (Cea) Sei $a(\cdot,\cdot)$ eine beschränkte, koerzive Bilinearform. Dann sind für jedes beschränkte, lineare Funktional F die Aufgaben (27) und (30) eindeutig lösbar. Für die zugehörigen Lösungen $u \in H$ und $u_n \in H_n$ gilt die Abschätzung

$$\|u - u_n\| \;\leq\; \frac{M}{\gamma} \cdot \inf_{\varphi_n \in H_n} \|u - \varphi_n\|\,,$$

für zwei Konstanten M , $\gamma \in \mathbb{R}$.

Beweis: Die eindeutige Lösbarkeit von (30) folgt mit H_n anstelle von H aus dem Satz von Lax-Milgram, da sich wegen $H_n \subset H$ die vorausgesetzten Eigenschaften von a auf H_n übertragen. Mit $H_n \subset H$ folgt aus (27) auch

$$a(u, \varphi_n) \;=\; F(\varphi_n) \quad \text{für alle} \quad \varphi_n \in H_n$$

und unter Beachtung der Linearität erhält man mit (30) hieraus die Beziehung

$$a(u - u_n, \varphi_n) = 0 \quad \text{für alle} \quad \varphi_n \in H_n\,.$$

Insbesondere gilt also auch

$$a(u - u_n, u_n) = 0$$

$$\Rightarrow\; a(u - u_n, u - u_n) = a(u - u_n, u - \varphi_n) \quad \text{für alle} \quad \varphi_n \in H_n\,.$$

Aus der Beschränktheit und Koerzivität von a folgt letztendlich

$$\gamma\|u{-}u_n\|^2 \;\leq\; M\|u{-}u_n\|{\cdot}\|u{-}\varphi_n\| \quad \text{für alle} \quad \varphi_n \in H_n \quad \text{und geeignete Konstanten} \quad \gamma\,,\, M\,,$$

was die Behauptung liefert. (q.e.d.)

Da die Dimension von H_n endlich ist, gibt es endlich viele linear unabhängige $h_i,\ i = 1, \ldots, n$, die den Teilraum H_n aufspannen, d.h.

$$H_n = span\{h_1, \ldots, h_n\} = \left\{ v \;:\; v(x) = \sum_{i=1}^{n} s_i h_i(x) \right\}\,,$$

mit rellen Koeffizienten s_i für $i = 1, \ldots, n$. Mit der Linearität von $a(\cdot, \cdot)$ und F ist damit (30) äquivalent zu

$$a(u_n, h_i) = F(h_i) \ , \quad i = 1, \ldots, n \, .$$

Beachtet man, daß die gesuchte Lösung u_n von (30) wegen $u_n \in H_n$ eine Darstellung der Form

$$u_n(x) = \sum_{j=1}^{n} s_j h_j(x) \ \ , \quad x \in \Omega$$

besitzt, dann bildet dies ein lineares Gleichungssystem

$$\sum_{j=1}^{n} a(h_j, h_i) s_j = F(h_i) \ , \ i = 1, \ldots, n \tag{31}$$

zur Bestimmung der Koeffizienten $s_j \in \mathbb{R}$, $j = 1, \ldots, n$. Die Gleichungen (31) werden **Galerkin-Gleichungen** genannt. Eine wichtige Eigenschaft dieser Gleichungen läßt sich direkt aus den Eigenschaften von a ableiten:

Lemma 6.2 Ist $a(\cdot, \cdot)$ eine koerzive, beschränkte Bilinearform, so ist die Koeffizientenmatrix des Gleichungssystems (31) regulär.

Beweis: Sei $z \in \mathbb{R}^n$ eine Lösung des homogenen Systems

$$\sum_{j=1}^{n} a(h_j, h_i) z_j = 0 \ , \quad i = 1, \ldots, n \ , \tag{32}$$

dann gilt auch

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a(h_j, h_i) z_j z_i = 0 \ .$$

Mit der Bilinearität von a erhält man hieraus

$$a(\sum_{i=1}^{n} h_i z_i, \sum_{j=1}^{n} h_j z_j) = 0$$

und wegen der Koerzivität dann schließlich

$$\sum_{i=1}^{n} h_i z_i \ = \ 0 \, .$$

Da die Funktionen h_i als linear unabhängig voraus gesetzt wurden, muß bereits $z = 0$ gelten. Also hat das homogene System (32) nur die triviale Lösung. Damit ist die Koeffizientenmatrix (im weiteren als **Steifigkeitsmatrix** bezeichnet) von (30) reglär. (q.e.d.)

Insgesamt ergibt sich die folgende schematische Übersicht zu den Variationsgleichungen und ihren diskreten Formulierungen:

Variationsgleichung		Optimierungsproblem
$a(u,\varphi) \ = \ F(\varphi)$ für alle $\varphi \in H$	$\Longleftrightarrow$ bei Symmetrie	$J(\varphi) \longrightarrow$ min! bei $\varphi \in H$
$\downarrow$		$\downarrow$
Galerkin-Verfahren $a(u_n,\varphi_n) \ = \ F(\varphi_n)$ für alle $\varphi_n \in H_n$	$\Longleftrightarrow$ bei Symmetrie	**Ritz-Verfahren** $J(\varphi_n) \longrightarrow$ min! bei $\varphi_n \in H_n$
$\searrow$		
	Galerkin-Gleichungen $\sum_{j=1}^{n} a(h_j,h_i)s_j = F(h_j)$ $i = 1,\ldots,n$	

6.2 Erweiterungen auf nichtlineare Probleme

Die im vorherigen Abschnitt mit Hilfe des Satzes von Lax-Milgram gemachten Aussagen zur Existenz von Lösungen abstrakter Variationsgleichungen setzen voraus, daß es sich bei der betrachteten Abbildung $a(\cdot,\cdot)$ um eine Bilinearform und bei F um ein lineares Funktional handelt. Die bisher gegebenen Resultate beziehen sich also nur auf lineare Randwertprobleme. Unter gewissen Bedingungen lassen sich diese aber auf nicht lineare Differentialoperatoren verallgemeinern, indem z. B. kontraktive Operatoren in Verbindung mit dem Banachschen Fixpunktsatz oder auch monotone Iterationsschemata mit zugehörigen Kompaktheitsargumenten verwendet werden.

Es bezeichne wieder H einen Hilbert-Raum mit dem zugehörigen Skalarprodukt $(\cdot, \cdot)$ und $B : H \longrightarrow H$ sei ein Operator mit den folgenden Eigenschaften:

1. Es existiert ein $\gamma > 0$ derart, daß
$$(Bu - B\varphi, u - \varphi) \geq \gamma \|u - \varphi\|^2$$
für alle $u, \varphi \in H$.
2. Es existiert ein $M > 0$ derart, daß
$$\|Bu - B\varphi\| \leq M\|u - \varphi\|$$
für alle $u, \varphi \in H$.

Als abstraktes Ausgangsproblem wird die Operatorgleichung

$$Bu \; = \; 0 \tag{33}$$

untersucht. Diese ist äquivalent zur nichtlinearen Variationsgleichung

$$(Bu\,, \varphi) \; = \; 0 \quad \text{für alle } \varphi \in H\,. \tag{34}$$

In direkter Verallgemeinerung des Satzes von Lax-Milgram gilt

Satz 6.3 Unter den Voraussetzungen 1. und 2. besitzt die Operatorgleichung (33) eine eindeutige Lösung $u \in H$. Diese ist Fixpunkt des durch

$$T_r\varphi \; := \; \varphi - rB\varphi\;, \quad \varphi \in H\,,$$

definierten Operators $T_r : H \longrightarrow H$, der für Parameterwerte $r \in (0, \frac{2\gamma}{M^2})$ kontrahierend ist.

Beweis: Wir zeigen zunächst die Kontraktivität des eingeführten Operators T_r für die angebebenen Parameterwerte. Nach Konstruktion gilt mit der durch das Skalarprodukt erzeugten Norm $\|u\| := \sqrt{(u, u)}$ des Hilbert-Raumes

$$\begin{aligned}
\|T_r y - T_r\varphi\|^2 \;&=\; \|y - rBy - [\varphi - rB\varphi]\|^2 \\
&=\; (y - \varphi - r[By - B\varphi]\,,\, y - \varphi - r[By - B\varphi]) \\
&=\; \|y - \varphi\|^2 \;-\; 2r(By - B\varphi, y - \varphi) \;+\; r^2\|By - B\varphi\|^2 \\
&\leq\; (1 - 2\gamma r + r^2 M^2)\|y - \varphi\|^2
\end{aligned}$$

für alle $y, \varphi \in H$. Also ist T_r kontrahierend für

$$\begin{aligned} & |r^2 M^2 - 2\gamma r + 1| < 1 \\ \Longleftrightarrow \quad & r \in (0, \frac{2\gamma}{M^2})\,. \end{aligned}$$

T_r besitzt damit nach dem Banachschen Fixpunktsatz einen eindeutigen Fixpunkt $u \in H$, d. h. es gilt

$$u = T_r u = u - rBu\,.$$

Folglich löst u auch die Operatorgleichung (33), so daß nun nur noch die Eindeutigkeit der Lösung zu zeigen bleibt. Es sei also $\tilde{u} \in H$ ebenfalls Lösung der Operatorgleichung, d. h.

$$B\tilde{u} = Bu = 0\,.$$

Mit der starken Monotonieeigenschaft (1., siehe oben) des Operators und der Bilinearität des Skalarproduktes folgt, daß

$$0 = (Bu - B\tilde{u}, u - \tilde{u}) \geq \gamma \|u - \tilde{u}\|^2 \geq 0\,.$$

Somit gilt $\tilde{u} = u$, so daß die Lösung von (33) eindeutig ist. (q.e.d.)

Ebenso lassen sich Gleichungen mit Operatoren $A : H^* \longrightarrow H$ anstelle von B betrachten. Dabei bezeichnet H^* den zu H gehörigen Dualraum, d.h. den Raum der beschränkten, linearen Funktionale $F : H \longrightarrow \mathbb{R}$. Es sei vorausgesetzt, daß A den folgenden Eigenschaften genüge:

1. Der Operator A ist stark monoton, d. h. es existiert ein $\gamma > 0$ derart, daß

   $$(Au - A\varphi, u - \varphi) \geq \gamma \|u - \varphi\|^2 \quad \text{für alle} \quad u, \varphi \in H\,.$$

2. Der Operator A ist Lipschitz-stetig, d. h. es existiert eine Konstante $M > 0$ derart, daß

   $$\|Au - A\varphi\| \leq M \|u - \varphi\| \quad \text{für alle} \quad u, \varphi \in H\,.$$

Dann besitzt das Problem

$$Au = F \tag{35}$$

für beliebiges $F \in H^*$ eine eindeutige Lösung $u \in H$. Diese Aussage erhält man unmittelbar aus Satz 6.3 mit Hilfe des für $\varphi \in H$ durch

$$B\varphi := T(A\varphi - F)$$

definierten Operators $B : H \longrightarrow H$. Analog zum Beweis des Satzes von Lax-Milgram bezeichnet $T : H^* \longrightarrow H$ den geeigneten Operator zum Rieszschen Darstellungssatz, der jedem beschränkten linearen Funktional $F \in H^*$ eine Element $T(F) \in H$ mit

$$F(\varphi) \;=\; (TF, \varphi) \quad \text{für alle} \quad \varphi \in H$$

zuordnet. Das Problem (35) ist äquivalent zu

$$(Au, \varphi) \;=\; F(\varphi) \tag{36}$$

für alle $\varphi \in H$.

Unter den getroffenen Voraussetzungen kann nun auch das Lemma von Cea auf die betrachteten Aufgaben übertragen werden. Hierzu gilt:

Satz 6.4 Es seien $A : H^* \longrightarrow h$ stark monoton bezüglich einer Konstanten $\gamma > 0$ und Lipschitz-stetig bezüglich einer Konstanten $M > 0$ sowie $F \in H^*$. Zu jedem linearen Unterraum $H_n \subset H$ mit $dim\, H_n < \infty$ existiert ein eindeutig bestimmtes $u_n \in H_n$, das der diskreten Variationsgleichung

$$(Au_n, \varphi_n) \;=\; F(\varphi_n) \quad \text{für alle} \quad \varphi_n \in H_n \tag{37}$$

genügt. Mit der exakten Lösung $u \in H$ der Variationsgleichung gilt hier zusätzlich die Abschätzung

$$\|u - u_n\| \leq \frac{M}{\gamma} \cdot \inf_{\varphi_n \in H_n} \|u - \varphi_n\| \,.$$

(Zum Beweis siehe [17], S. 112.)

Im Unterschied zu linearen Randwertaufgaben erhält man im vorliegenden Fall dann natürlich auch nichtlineare Galerkin-Gleichungen. Wird der Ansatz

$$u_n(x) \;=\; \sum_{j=1}^{n} s_j h_j(x)$$

gewählt, so ist (37) äquivalent zum Gleichungssystem

$$(A(\sum_{j=1}^{n} s_j h_j)\,,\, h_i) \;=\; F(h_i)\,, \quad i = 1, \ldots, n\,.$$

Auf die iterativen Methoden, die zur Lösung solcher Systeme nötig sind, wird später noch eingegangen.

6.3 Methode der finiten Elemente

6.3.1 Zusammenfassung

In diesem Abschnitt soll in Anlehnung an [32], [26] und [17] die zugrundeliegende Idee und das sich daraus ergebende Vorgehen der Methode der finiten Elemente zur Lösung von Aufgaben, wie sie in den vorherigen Abschnitten eingeführt worden sind, beschrieben werden.

Wie in Abschnitt 6.2.3 bereits skizziert, besitzt die Methode der finiten Elemente die folgenden vier typischen Merkmale:

1. Zerlegung des Grungebietes $\Omega \subset \mathbb{R}^n$ in geometrisch einfache Teilgebiete Ω_n;
2. Definition von geeigneten Ansatzfunktionen u_j und linear unabhängigen, stückweise stetigen Formfunktionen N_j über den Teilgebieten;
3. iterative Lösung des erhaltenen Gleichungssystemes liefert Approximationen u_n;
4. Auswertung des Fehlers zwischen exakter Lösung u und u_n bezüglich geeigneter Norm.

Zu 1. Die gegebene Aufgabe wird diskretisiert, indem das Grundgebiet Ω in einfache Teilgebiete, die sogenannten Elemente Ω_n, zerlegt wird. Im Fall von zweidimensionalen Problemen wird das Grundgebiet in Dreiecke, Parallelogramme, krummlinige Dreiecke oder Vierecke eingeteilt. Selbst wenn nur geradlinige Elemente verwendet werden, erreicht man mit einer entsprechend feinen Diskretisierung eine gute Approximation des Grundgebietes, wobei krummlinige Elemente selbstverständlich die Güte der Annäherung erhöhen. Bei räumlichen Problemen erfolgt eine Diskretisierung des dreidimensionalen Gebietes in Tetraederelemente, Quaderelemente oder anderen, dem Problem angepaßten, möglicherweise auch krummflächig berandeten Elementen. Die wichtigsten Finite-Elemente-Ansätze im $\mathbb{R}^2$ und $\mathbb{R}^3$ findet man z.B. in [17], S. 129ff, abgebildet und ausführlich beschrieben.

Zu 2. In jedem der Elemente wird für die gesuchte Funktion ein problemgerechter Ansatz bzw. problemgerechte Formfunktionen N_j gewählt, um so eine Approximation der gesuchten Funktion u bzw. eine Ansatzfunktion u_n für u der Form

$$u_n = \sum_{j=1}^{n} s_j \cdot N_j$$

mit unbekannten Koeffizienten s_j zu finden.

Die Art des Ansatzes hängt dabei einerseits von der Form des Elementes ab; andererseits kann auch das zu behandelnde Problem den zu wählenden Ansatz beeinflussen, da die Ansatzfunktionen beim Übergang von einem Element in ein benachbartes ganz bestimmte problemabhängige Stetigkeits- und Differenzierbarkeitsbedingungen und Randwerte erfüllen müssen. Elemente mit Ansatzfunktionen, welche den Stetigkeitsbedingungen genügen, heißen konform.

Um diese Anforderungen tatsächlich zu erfüllen, ist der Funktionsverlauf innerhalb eines betrachteten Elementes durch gegebene Funktionswerte oder auch durch Werte partieller Ableitungen in bestimmten Punkten des Elementes, den Knotenpunkten, auszudrücken. Mit Hilfe solcher Werte als Koeffizienten stellt sich die Ansatzfunktion als Linearkombination der Formfunktionen dar. Werden ausschließlich vorgegebene Funktionswerte zur Bestimmung der Ansatzfunktionen verwendet, so bezeichnet man die zugehörigen Elemente als **Lagrange-Elemente**. Im Unterschied dazu nutzen **Hermite-Elemente** dem Prinzip der Hermiteschen Interpolation folgend auch Vorgaben von Ableitungswerten zur Bestimmung der Ansatzfunktionen.

Nimmt man konkret den Fall an, daß in den m Knotenpunkten $p_j^{(e)}$ des Elementes e nur Funktionswerte $u(p_j^{(e)})$ verwendet werden, so erhält die Ansatzfunktion die Darstellung

$$u_n^{(e)} = \sum_{j=1}^{m} u(p_j^{(e)}) \cdot N_j^{(e)} .$$

Da diese Darstellung für den Funktionswert $u(p_j^{(e)})$ jedes Knotenpunktes gelten muß, so muß die Formfunktion $N_j^{(e)}$ notwendigerweise im Punkt $p_j^{(e)}$ gleich 1 sein und in den anderen Knotenpunkten des Elementes (e) verschwinden. Um an dieser Stelle die Grundlage für die Anwendung des Galerkin-Verfahrens im Sinne der Methode der finiten Elemente vorzubereiten, betrachtet man die globale Darstellung der gesuchten Funktion u_n im ganzen Grundgebiet, bestehend aus der Vereinigungsmenge der Elemente. Numeriert man die Werte in den Knotenpunkten fortlaufend von 1 bis n, dann läßt sich das Ergebnis der Zusammensetzung formulieren als

$$u_n = \sum_{j=1}^{n} u(p_j) \cdot N_j .$$

Daraus wird ersichtlich, daß die globalen Formfunktionen N_j nur in denjenigen Elementen von Null verschieden sind, welche den Knotenpunkt p_j gemeinsam haben, so daß also die Funktionen N_j nur in einem sehr beschränkten Teilgebiet von Null verschieden sind, also nur einen lokalen Träger aufweisen, was ein wesentliches Charakteristikum der Methode der finiten Elemente ist. Die Formfunktionen müssen also zwei Eigenschaften erfüllen:

- N_i hat polynomiale Gestalt über jedem Element und ist eindeutig festgelegt durch die Funktionswerte an den Knotenpunkten dieses Elementes.

- $N_i(p_j) = \delta_{ij}$, wobei δ_{ij} das Kronecker-Symbol darstellt, das für $i = j$ 1 und sonst 0 ist.

In einem Ansatz dieser Form lassen sich Randbedingungen einfach durch Vorgabe entsprechender Werte für die betreffenden Knotenpunkte verwirklichen.

Zu 3. und 4. soll in den nächsten Abschnitten genauer auf einige iterative Methoden und auf die Methode der gewichteten Residuen eingegangen werden.

6.3.2 Methode der gewichteten Residuen

Das Ziel der oben beschriebenen Methoden ist es also, eine approximative Lösung u_n der Form

$$u_n = \sum_{j=1}^{n} s_j \cdot N_j$$

mit unbekannten Koeffizienten s_j zu finden. Wird nun dieser Ansatz in eine partielle Differentialgleichung der Form

$$\mathcal{L}(u) = 0$$

mit einem geeignet definierten Differentialoperator $\mathcal{L}$ eingesetzt, so wird sie in den seltensten Fällen erfüllt sein, vielmehr resultiert ein sogenanntes **Residuum** $R = \mathcal{L}(u_n)$. Da der genaue Fehler ebenso wie die exakte Lösung meist unbekannt sind, verlangt man also, daß das Residuum im Inneren des Gebietes Ω möglichst klein werden soll. Die Idee, die hinter der Methode der Gewichteten Residuen steht, ist nun folgende:

Gesucht sind Koeffizienten $s_1, \ldots, s_n$ derart, daß das gewichtete Mittel des Residuums R über Ω verschwindet für n linear unabhängige Gewichtsfunktionen oder auch Testfunktionen W_i:

$$\int_{\Omega} R \cdot W_i = 0\,,\; i = 1, \ldots, n$$

In der **Methode von Galerkin** werden die n Gewichtsfunktionen der Reihe nach gleich den n Formfunktionen N_i gewählt, welche die Bedingung der linearen Unabhängigkeit ja bereits erfüllen. Mit dieser Wahl der Gewichtsfunktionen erreicht man, daß die Residuenfunktion orthogonal zum Funktionenunterraum ist, der von den $N_1, \ldots, N_n$ aufgespannt wird. Diese Tatsache rechtfertigt das beschriebene Vorgehen, indem die resultierende approximative Lösung u_n in diesem Sinne die bestmögliche im Raum der Ansatzfunktionen darstellt.

6.3.3 Modellproblem: Die 1D Poisson-Gleichung

Man betrachte das eindimensionale Randwertproblem

$$-u''(x) \;=\; f(x)\;, \quad \text{für} \quad x \in \Omega \;=\; [0,1]\;,\; u(0) \;=\; u_L\;,\; u(1) \;=\; u_R\,.$$

Das Gebiet $\Omega = [0,1]$ wird unterteilt in m Elemente der Länge h.

$$\Omega_e \;=\; [(e-1)h\,,\, eh] \;\text{ mit }\; h = 1/m \;\text{ und }\; e = 1,\ldots,m\,.$$

Die Knotenpunkte seien bezeichnet mit $p_i = (i-1)h\,,\, i = 1,\ldots,n = m+1$.

Jede Formfunktion N_i soll nun stückweise linear sein mit $N_i(p_j) = \delta_{ij}$, was es recht einfach macht, die vorgegebenen Randwerte zu erfüllen. Wähle dazu die Funktion

$$\psi(x) \;=\; u_L N_1(x) \;+\; u_r N_n(x)\,,$$

die folgenden Ansatz liefert:

$$u(x) \;\approx\; u_n(x) \;=\; \psi(x) \;+\; \sum_{j=2}^{n-1} s_j N_j(x) \;=\; \sum_{j=1}^{n} s_j N_j(x)\,,$$

wobei natürlich $s_1 \;=\; u_L$ und $s_n \;=\; u_R$ sein müssen.

Die Gleichung der gewichteten Residuen erhält man nun, wie schon beschrieben, durch Einsetzen der Darstellung von u_n in die Differentialgleichung, Multiplikation mit den Gewichtsfunktionen $W_i \;=\; N_i\,,\, i \;=\; 1,\ldots,n$ und anschließender Integration über das ganze Intervall $[0,1]$. Vergleichbar mit dem Variationsproblem der Poissonschen Differentialgleichung in Abschnitt 5.2.2 erhält man mittels partieller Integration folgende diskrete Formulierung:

$$\sum_{j=1}^{n}\Big(\int_0^1 N_i' N_j'\,\Big)\, s_j \;=\; \int_0^1 f N_i \;\text{ für }\; i = 1,\ldots,n\,.$$

Dieses Gleichungssystem läßt sich mit folgenden Bezeichnungen in Matrixschreibweise darstellen:

$$\begin{aligned}
A \;&:=\; (a_{ij})_{i,j=1}^n \quad \text{mit} \quad a_{ij} \;:=\; \int_0^1 N_i' N_j'\,,\\
b \;&:=\; (b_i)_{i=1}^n \quad \text{mit} \quad b_i \;:=\; \int_0^1 f N_i\,,\\
s \;&:=\; (s_i)_{i=1}^n\,.
\end{aligned}$$

Man erhält also das Gleichungssystem

$$A\,s \;=\; b\,.$$

Dabei wird A wie schon zuvor erwähnt in Anlehnung an Probleme der Elastitätstheorie **Steifigkeitsmatrix** genannt.

Eine sehr feine Zerlegung des Gebietes Ω führt zwar dazu, daß die erzeugte Matrix A eine große Dimension besitzt, allerdings ist diese Matrix dann nur schwach besetzt, da die vorher aufgezeigten Eigenschaften der Formfunktionen N_i zu diversen Nulleinträgen führen: Der Matrix-Eintrag a_{ij} ist genau dann von Null verschieden, wenn die Knotenpunkte p_i und p_j zum selben Element gehören. In anderen Worten:

$$a_{ij} \;\neq\; 0 \;\Longleftrightarrow\; j \;\in\; \{i-1, i, i+1\}\,.$$

Setzt man die betrachteten Knotenpunkte auf den Rand eines jeden Elementes fest, so könnten die Formfunktionen wie in Abbildung 13 konstruiert sein. Nimmt man nun der Einfachheit halber eine konstante Elementlänge bzw. -größe h an, so folgt direkt, daß $N_i' \;=\; \pm 1/h$. Betrachtet man nun die rechte Seite von Abbildung 13, so erhält man für die Randeinträge und weiteren Nicht-Null-Einträge von A:

$$\begin{aligned}
a_{1,1} &= \frac{1}{h}\,, \\
a_{i,i-1} &= \int_0^1 N_{i-1}' N_i' \;=\; -\frac{1}{h} \quad \text{für}\;\; i = 2, \ldots, n\,, \\
a_{i,i} &= \int_0^1 N_i' N_i' \;=\; \frac{2}{h} \quad \text{für}\;\; i = 2, \ldots, n-1\,, \\
a_{i,i+1} &= \int_0^1 N_{i+1}' N_i' \;=\; -\frac{1}{h} \quad \text{für}\;\; i = 1, \ldots, n-1\,, \\
a_{n,n} &= \frac{1}{h}\,.
\end{aligned}$$

Zur genauen Betrachtung der Ausdrücke b_i verwendet man die explizite Darstellung der N_i mit

$$N_i(x) \;=\; \begin{cases} 0 & ,\quad x \leq p_{i-1} \\ (x - p_{i-1})/h & ,\quad p_{i-1} < x \leq p_i \\ (p_{i+1} - x)/h & ,\quad p_i < x \leq p_{i+1} \\ 0 & ,\quad x > p_{i+1} \end{cases}$$

für $1 \;<\; i \;<\; n$. Die Formeln für N_1 und N_n müssen leicht abgeändert werden. Die Integrale der Ausdrücke b_i können nur in den seltensten Fällen analytisch berechnet werden, so daß häufig ein numerisches Integrationsschema nach der Trapezregel angewendet wird. Hierbei werden die Knotenpunkte auch zu den Integrationspunkten:

$$b_i = \int_0^1 f\,N_i \;\approx\; \frac{h}{2} f(p_1) N_i(p_1) \;+\; \sum_{j=2}^{n-1} h\, f(p_j) N_i(p_j) \;+\; \frac{h}{2} f(p_n) N_i(p_n)\,.$$

1 2 3 4 5 6 7 x 1 2 3 4 5 6 7 x

Abbildung 13: Skizze der stückweise linearen Formfunktionen N_i in einer Dimension für einige Elemente. Die Skalierung an der x-Achse bezieht sich auf die Knotenpunkte des Gitters. Die Graphen links illustrieren also die Formfunktionen N_3, N_4, N_5 und N_6, während die Graphen rechts die jeweiligen Ableitungen skizzieren. (aus [26], S. 126 und 128)

Das liefert $b_1 = f(0)\cdot h/2$, $b_i = f(p_i)\cdot h$ für $i = 2,\dots,n-1$ und schließlich auch $b_n = f(1)\cdot h/2$. Die Randwerte kann man allerdings auch direkt mit $s_1 = u_L$ und $s_n = u_R$ in das Gleichungssystem einfließen lassen. Also hat das globale Gleichungssystem die folgende Gestalt, die formale Ähnlichkeit mit der finiten Differenzen-Methode aufweist:

$$A\,s \;=\; b$$

$$\iff \left\{ \begin{array}{rcll} s_1 &=& u_L & \\ -\frac{1}{h}s_{i-1} + \frac{2}{h}s_i - \frac{1}{h}s_{i+1} &=& f(p_i)h & \text{für} \quad i = 2,\dots,n-1 \\ s_n &=& u_r x & \end{array} \right.$$

mit $p_i \;=\; (i-1)h\,,\; i = 2,\dots,n$.

6.3.4 Elementmatrizen

Im vorherigen Abschnitt wurde das zu lösende Gleichungssystem direkt hergeleitet, was für mehrdimensionale Probleme mit geometrisch anspruchsvolleren Gebieten nicht durchführbar ist.

Um in diesen Fällen das globale lineare Gleichungssystem aufzustellen, betrachtet man zunächst nur die jeweiligen Elementmatrizen und setzt diese zur globalen Matrix zusammen. Man schreibt

$$a_{ij} \;=\; \int_0^1 N_i' N_j'\,dx \;=\; \sum_{e=1}^{m} a_{ij}^{(e)}\,, \quad \text{wobei} \quad a_{ij}^{(e)} = \int_{\Omega_e} N_i' N_j' dx\,,$$

$$\text{und} \quad b_i \;=\; \int_0^1 f N_i\,dx \;=\; \sum_{e=1}^{m} b_i^{(e)}\,, \quad \text{wobei} \quad b_i^{(e)} = \int_{\Omega_e} f N_i dx,.$$

Nun ist, wie schon zuvor, $a_{ij}^{(e)}$ genau dann ungleich Null, wenn p_i und p_j Knotenpunkte des Elementes Ω_e sind, da sonst die Formfunktionen $N_i'\,N_j'$ verschwinden. Genau so ist auch $b_i^{(e)}$ nur dann ungleich Null, wenn p_i Knotenpunkt von Ω_e ist. Es ist also sinnvoll, nur die lokalen Knotenpunkte des gerade behandelten Elementes zur Berechnung von $a_{ij}^{(e)}$ und $b_i^{(e)}$ heran zu ziehen, was zur Notwendigkeit einer lokalen Knotennummerierung für jedes Element führt. In unserem einfachen linearen Fall mit den Elementen $\Omega_e = [p_e, p_{e+1}]$ erhält p_e die lokale Knotennummer 1 und dementsprechend p_{e+1} die lokale Knotennummer 2.

Es gibt also eine eindeutige Zuordnung $i \;=\; q(e,r)$, die zu jeder lokalen Knotennummer r des Elementes e die zugehörige globale Knotennummer i liefert und umgekehrt. Im betrachteten linearen Fall ist

$$q\,(\,e\,,\,r\,) \;=\; e \,-\, 1 + r \quad \text{für} \quad r \;= 1, 2\;.$$

Bei mehrdimensionalen Problemen läßt sich diese Zuordnung nicht mehr so einfach analytisch angeben, sondern existiert nur in Form einer Zuordnungstabelle für alle Knotenpunkte. Man sammelt also im betrachteten Fall zunächst die Nicht-Null-Einträge von $A_{ij}^{(e)}$ in einer 2×2 **Elementmatrix** $\tilde{A}_{rs}^{(e)}$, wobei $r, s = 1, 2$ die lokalen Knotennummern bezeichnen, und ebenso die **Elementvektoren** $\tilde{b}_r^{(e)}$ für $r = 1, 2$. Vor allem bei mehrdimensionalen Problem ist es dann noch zwingend notwendig, jedes pysikalische Element vor weiteren Berechnungen auf ein Referenz-Element mit fester Einheitsgröße zu transformieren. Dieses kann mit einer vollständigen Variablentransformation bei den Knotenpunkten, den Formfunktionen und sämtlichen Integralen und Ableitungen erreicht werden. Siehe dazu z.B. [26], S. 132-136.

Abbildung 14: Zusammensetzung der globalen Matrix: Die zu Element Nr. 3 und Nr. 4 gehörigen Elementmatrizen werden zur globalen Matrix hinzu gefügt. (nach [26], S. 137)

Zum Schluß bleibt dann nur noch die Aufgabe, die einzelnen Elementmatrizen und Elementvektoren zu einem globale Gleichungssystem zusammenzusetzen, was mittels der lokalen Knotennummern (r, s) und den Zuordnungen $q(e, r) = i$, $q(e, s) = j$ auf einfach Weise realisiert werden kann. Abbildung 14 zeigt, wie die 2×2 Elementmatrizen unseres linearen Modellproblemes zu einer globalen Matrix zusammen-gefügt werden. Vollendet man diesen Prozess, so erhält man das erwartete globale Gleichungssystem für konstante Elementlänge h:

$$\begin{pmatrix} \frac{1}{h} & -\frac{1}{h} & 0 & \cdots & \cdots & \cdots & 0 \\ -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & 0 & \ddots & \ddots & \vdots \\ 0 & -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} \\ 0 & \cdots & \cdots & \cdots & 0 & -\frac{1}{h} & \frac{1}{h} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ \vdots \\ \vdots \\ s_{n-1} \\ s_n \end{pmatrix} = \begin{pmatrix} u_L \\ f(p_2)h \\ \vdots \\ \vdots \\ \vdots \\ f(p_{n-1})h \\ u_R \end{pmatrix}$$

6.4 Iterative Methoden

Nachdem man partielle Differentialgleichungen mit der Finite-Elemente Methode diskretisiert hat, bleibt nun noch die Aufgabe, das wie oben erhaltene lineare Gleichungssystem effektiv und doch genau zu lösen. In diesem Abschnitt sei das Gleichungssystem ganz allgemein mit

$$A \cdot x = b$$

für eine Matrix $A \in \mathbb{R}^{(n \times n)}$ und einen gegebenen Vektor $b \in \mathbb{R}^n$ bezeichnet.

Lineare Systeme kann man einerseits direkt analytisch und andererseits iterativ lösen, wobei die direkte Lösung häufig das Eliminationsverfahren von Gauss beinhaltet. Im allgemeinen benötigt das Gaussverfahren im Laufe der Elimination also n^2 Einträge der Matrix A und führt ungefähr $2/3 \cdot n^3$ Rechenoperationen durch, d.h. Additionen, Subtraktionen, Multiplikationen und Divisionen. Gerade für mehrdimensionale Probleme und die daraus resultierenden sehr großen Gleichungssysteme sind die iterativen Methoden gegenüber den direkten Methoden klar im Vorteil: Iterative Methoden können aufgrund der schwachen Besetztheit vieler Koeffizienten-Matrizen deutlich effektiver arbeiten, da sie auch deutlich weniger Rechenoprationen benötigen. Meist basieren die iterativen Löser auf Matrix-Vektor-Multiplikationen, bei denen sich die vielen Null-Einträge dann positv auf den Rechenaufwand auswirken.

In diesem Zusammenhang ist das Maß für die Effektivität eines iterativen Lösers natürlich die Konvergenzgeschwindigkeit des Iterationsschemas. Um diese Konvergenzgeschwindigkeit zu erhöhen, gibt es verschiedenste Möglichkeiten der Vorkonditionierung eines Gleichungssystemes. Are Magnus Bruaset verdeutlicht in [5] besonders anschaulich den gravierenden Unterschied zwischen der Effektivität vorkonditionierter iterativer Methoden und direkter Methoden. Er betrachtet die Finite-Differenzen-Diskretisierung der 3D Poisson-Gleichung und vergleicht die Zeiten, die theoretisch zur Lösung des Gleichungssystemes benötigt werden. Vergleicht man die Werte der Gauss-Elimination mit denen der vorkonditionierten Methode der konjugierten Gradienten (s. [5], Tabelle 1.1, S. 5), so stellt man fest, das die Unterschiede mit zunehmender Anzahl der Unbekannten immer deutlicher werden. Bei 8 000 000 Unbekannten würde die iterative Methode theoretisch ca. 2 600 Sekunden benötigen, während die direkte Methode unpraktikable 832 Jahre beschäftigt wäre.

6.4.1 Klassische Methoden: Jacobi und Gauss-Seidel Verfahren

Im folgenden sollen in Anlehnung an die in der später noch genauer erläuterten C++-Biliothek Diffpack verwendeten Methoden (s. [26], S. 592ff) einige klassische Iterationsmethoden in ihren wesentlichen Zügen erläutert werden. Zu genaueren Details siehe [16], S. 147ff, und [5], S. 10ff.

Unter den klassischen Iterationmethoden versteht man meist Verfahren wie **Jacobi**, **Gauss-Seidel** und das **SOR**-Verfahren. Diese Verfahren basieren alle auf einer Zerlegung der Koeffizientenmatrix des Gleichungssystemes

$$A \cdot x = b$$

für eine Matrix $A \in \mathbb{R}^{(n \times n)}$ und einen gegebenen Vektor $b \in \mathbb{R}^n$. Schreibt man A in der Form $A = M - N$, wobei M eine reguläre $(n \times n)$-Matrix ist, so verändert sich das zugrunde liegende Gleichungssystem zu

$$M\,x = N\,x + b\,.$$

Diese Zerlegung der Koeffizientenmatrix wird in bezug auf die Matrix M mit **'matrix splitting'** oder **Vorkonditionierung** betitelt, wobei M dann als **Vorkonditionierer** bezeichnet wird. Die Regularität der Matrix M liefert, daß Systeme der Form $Mv = w$ durch nur $\mathcal{O}(n)$ Rechenoperationen zu lösen sind. Also ist es auch keine Schwierigkeit, das Gleichungssystem durch Multiplikation mit M^{-1} noch weiter zu modifizieren:

$$x = M^{-1}N\,x + M^{-1}b := G\,x + c$$

Diese umformulierte Variante des Gleichungssystemes legt nun offensichtlich das folgende Iterationsschema nahe: Für gegebenen Anfangsvektor x^0 berechne

$$x^k = G\,x^{k-1} + c \quad \text{für} \quad k = 1, 2, \ldots$$

Die oben definierte Matrix G spielt bei den Konvergenzbetrachtungen der iterativen Methode eine wesentliche Rolle, wie man direkt an dem Zusammenhang

$$x^k - x = G^k(x^0 - x)$$

erkennen kann. Eine iterative Methode heißt konvergent, falls die Folge $\{x^k\}$ für $k \to \infty$ und einen Startvektor x^0 gegen die exakte Lösung x konvergiert. Man sieht also, daß die Bedingung

$$\lim_{k \to \infty} \|G^k\| = 0$$

notwendig und hinreichend für die Konvergenz gegen x ist. Diese Bedingung ist äquivalent zur Formulierung

$$\rho(G) < 1$$

wobei mit $\rho(G)$ der Spektralradius von G, also das Maximum der Beträge der Eigenwerte, bezeichnet ist. Zum Beweis dieser Äquivalenz siehe z.B. [17], S. 257, Lemma 5.2. Verkleinert man $\rho(G)$, so erhöht sich die Konvergenzgeschwindigkeit.

Von besonderem Interesse ist dann außerdem die **Konvergenzrate** bzw. relative Konvergenzgeschwindigkeit R_k, die definiert ist durch

$$R_k(G) = -\frac{\ln \|G^k\|}{k},$$

und damit die zugehörige **asymptotische Konvergenzrate** liefert:

$$R_\infty(G) = \lim_{k \to \infty} R_k(G) = -\ln \rho(G)$$

Um den Anfangsfehler zum Beispiel um einen Faktor ϵ zu verringern, d.h. um

$$\|x - x^{k-1}\| \leq \epsilon \cdot \|x - x^0\|$$

zu erreichen, benötigt man $-\ln \epsilon / R_\infty(G)$ Iterationschritte.

Da die Schreibweise

$$x^k = G\,x^{k-1} + c$$

für die spätere Implementierung nicht geeignet ist, wählt man meist die etwas abgewandelte Schreibweise

$$x^k = x^{k-1} + M^{-1} r^{k-1},$$

wobei $r^{k-1} = b - Ax^{k-1}$ das Residuum nach dem Iterationsschritt $k-1$ bezeichnet. Zerlegt man nun noch A in Diagonalmatrix D, obere Dreiecksmatrix U und untere Dreiecksmatrix L, also $A = L{+}D{+}U$, dann lassen sich im folgenden die Unterschiede zwischen den klassischen Verfahren mittels dieser Notation darstellen. Desweiteren seien mit $A_{r,s}$ die Blockmatrizen der Matrix A bezeichnet, also

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,\nu} \\ \vdots & & \vdots \\ A_{\nu,1} & \cdots & A_{\nu,\nu} \end{pmatrix}, \; x = \begin{pmatrix} x_1 \\ \vdots \\ x_\nu \end{pmatrix} \quad \text{und} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_\nu \end{pmatrix},$$

wobei $A_{r,s} \in \mathbb{R}^{(n_r \times n_s)}$ und $\sum_{r=1}^{\nu} n_r = n$ sein muß.

Jacobi-Verfahren Wählt man M als die Diagonale von A, also $M = D$ und $N = -L - U$, so stellt sich das Iterationsschema fogendermaßen dar:

$$x^k = x^{k-1} + D^{-1} r^{k-1},$$

Schreibt man dieses Schema explizit aus, so folgt für den k-ten Iterationschritt und die i-te Komponente der gesuchten Lösung x^k

$$x_i^k = x_i^{k-1} + \frac{1}{a_{i,i}} \left(b_i - \sum_{j=1}^{n} a_{i,j} \cdot x_j^{k-1} \right)$$

für alle $i = 1, \ldots, n$. Zur Betrachtung des Konvergenzverhaltens wählt man wiederum die Darstellung $x^k = G_J\, x^{k-1} + c$, wobei bei diesem Verfahren

$$G_J = D^{-1}(L + U) \quad \text{und} \quad c = D^{-1} b$$

sind. Das führt nun zur Konvergenzbedingung $\rho(D^{-1}(L + U)) < 1$. Handelt es sich bei der Matrix A nun z.B. um eine positiv definite symmetrische Matrix, oder allgemeiner um eine Matrix mit Diagonaldominanz, so sieht man direkt, daß

$$\rho(G_J) = \rho(D^{-1}(L + U)) \leq \|D^{-1}(L + U)\|_\infty = \max_i \sum_{j \neq i} |a_{i,j}/a_{i,i}| < 1$$

ist. Je „stärker“ die Diagonaldominanz, desto höher die Konvergenzgeschwindigkeit.

Gauss-Seidel-Verfahren Ein lineares Gleichungssystem, dessen Koeffizientenmatrix eine obere oder untere Dreiecksgestalt besitzt, ist bekannterweise besonders einfach zu lösen, so daß der Gedanke, M als obere Dreicksmatrix zu wählen, nahe liegt. Es ist also $M = D + L$ und demnach dann $N = -U$. Die Gauss-Seidel-Methode hat also allgemein folgende Gestalt:

$$(D + L) \cdot x^k = -U\, x^{k-1} + b$$

Der Standard-Algorithmus zum Lösen eines Gleichungssystemes mit Koeffizientenmatrix in oberer Dreicksgestalt angewandt auf diese Formel liefert für den k-ten Iterationschritt und die i-te Komponente der gesuchten Lösung x^k

$$x_i^k = \frac{1}{a_{i,i}} \left(b_i - \sum_{j=1}^{i-1} a_{i,j} x_j^k - \sum_{j=i+1}^{n} a_{i,j} x_j^{k-1} \right)$$

für alle $i = 1, \ldots, n$. Um Aussagen bez. des Konvergenzverhaltens machen zu können, betrachtet man wiederum die Darstellung $x^k = G_{GS}\, x^{k-1} + c$, wobei hier

$$G_{GS} = (D - L)^{-1} \cdot U \quad \text{und} \quad c = (D - L)^{-1} \cdot b$$

sind. Das führt zur Konvergenzbedingung $\rho(G_{GS}) = \rho((D - L)^{-1} \cdot U) < 1$. Man kann nun für eine positv definite Koeffizientenmatrix A direkt Konvergenz des Verfahrens zeigen (siehe dazu z.B. [17], S. 259) oder noch allgemeiner die Gleichungungen

$$\rho(G_{GS}) = (\rho(G_J))^2 \quad \text{und} \quad R_\infty(G_{GS}) = 2 \cdot R_\infty(G_J)$$

für jede Koeffizientenmatrix A mit Nicht-Null Einträgen auf der Diagonalen. Siehe dazu z.B. [5], S. 14.

Der wesentliche Unterschied der expliziten Iterationsschemata dieser beiden Verfahren liegt also in der Verwendung bereits berechneter Werte: Das Gauss-Seidel-Verfahren benutzt die neu berechnete Näherung bereits beim nächsten Teilschritt und wird aufgrund dessen auch als **Einzelschritt-Verfahren** bezeichnet, während das Jacobi-Verfahren die neu berechnete Näherung erst nach Vollendung aller Teilschritte einbringt und somit die Bezeichnung **Gesamtschritt-Verfahren** erhält.

6.4.2 Relaxationsverfahren: SOR und SSOR

Ein erheblicher Nachteil einfacher Iterationsverfahren vom Typ der Gauss-Seidel- und Jacobi-Verfahren besteht in der asymptotisch langsamen Konvergenz bzw. in der niedrigen asymptotischen Konvergenzrate R_∞. Gewisse Beschleunigungen dieser Verfahren lassen sich durch die zusätzliche Verwendung von Ralaxations-Parametern erreichen. Im weiteren soll ausschließlich das Gauss-Seidel-Verfahren als Grundmethode betrachtet werden.

SOR-Verfahren Bezeichnet $\tilde{x}$ einen durch einen Gauss-Seidel-Schritt vorhergesagten Wert für x im k-ten Iterationsschritt, so wird x^k als gewichtete Kombination aus $\tilde{x}$ und dem vorherigen Wert x^{k-1} dargestellt:

$$x^k = \omega\tilde{x} + (1 - \omega)x^{k-1}.$$

Dabei ist $\omega > 0$ der bereits oben genannte Relaxations-Parameter. Das Ziel ist, diesen Parameter so zu wählen, daß der Spektralradius $\rho(G)$ der Iterationsmatrix minimal ist. Die resultierende Methode heißt **'successive overrelaxation'**, kurz **SOR**, und stimmt für $\omega = 1$ mit der Gauss-Seidel-Methode überein.

Die Relaxations-Methode kann ausgedrückt werden mittels

$$M = \frac{1}{\omega} D + L \quad \text{und} \quad N = \frac{1-\omega}{\omega} D - U\,.$$

Der zugehörige Algorithmus hat also offensichtlich die Form

$$x_i^k = \frac{\omega}{a_{i,i}} \left(b_i - \sum_{j=1}^{i-1} a_{i,j} x_j^k - \sum_{j=i+1}^{n} a_{i,j} x_j^{k-1} \right) + (1-\omega) x_i^{k-1}$$

für $i = 1, \ldots, n$.

SSOR-Verfahren Das SOR-Verfahren läßt sich durch Kombination von Vorwärts- und Rückwärtsvarianten symmetrisieren. Auf diese Weise erhält man das **symmetrische SOR-Verfahren**, kurz **SSOR**. Die Matrix M hat die Gestalt

$$M = \frac{1}{2-\omega} (\frac{1}{\omega} D + L) (\frac{1}{\omega} D)^{-1} (\frac{1}{\omega} D + U)\,,$$

so daß sich der zugehörige Algorithmus analog zum SOR-Verfahren erstellen läßt.

A. M. Bruaset vergleicht in [5], S. 18, die Konvergenzraten der vier erwähnten Iterationsmethoden in Anwendung auf ein Modellproblem: Betrachtet man die 2D Poisson-Gleichung $-\nabla^2 u = f$ auf einem mit Elementlänge h diskretisierten Einheitsrechteck mit $u = 0$ auf dem Rand, so erhält man folgende asymptotische Abschätzungen:

Konvergenzrate	**Jacobi**	**Gauss-Seidel**	**SOR**	**SSOR**
$R_\infty(G)$	$\pi^2 h^2/2$	$\pi^2 h^2$	$2\pi h$	$> \pi h$

Zu detaillierteren Erläuterungen der verwendeten Iterationsmethoden siehe [26], S. 591ff, und die Werke von A. Greenbaum [16] und A. M. Bruaset [5] zu iterativen Methoden und deren Analysis.

7 Existenz- und Eindeutigkeitssätze zu den Navier-Stokes Gleichungen

Viele Aussagen aus der Theorie zur Lösung partieller Differentialgleichungen lassen sich nicht auf die vollständigen Navier-Stokes-Gleichungen übertragen, da es sich hierbei um einen äußerst komplizierten instationären, nicht linearen Spezialfall handelt. Um trotzdem Existenz- und Eindeutigkeitssätze formulieren zu können, betrachtet man schwache bzw. Variationsformulierungen des Problems und versucht, mittels Konvergenz approximativer Verfahren und Abschätzungen geeigneter Hilbertraumnormen Existenz und Eindeutigkeit dieser schwachen Lösungen nachzuweisen. Im folgenden werden die klassischen Ergebnisse von Roger Temam [35] aus dem Jahre 1977 mit den aktuellen Ergebnissen von Prof. M. Wiegner [39] verglichen bzw. ergänzt, um ein vollständiges Bild des derzeitigen Wissensstandes zu erhalten.

7.1 Klassische Resultate: R. Temam, 1977

Das Problem der vollen Navier-Stokes-Gleichungen ist das folgende:

Gesucht ist eine vektorielle Funktion $u : \Omega \times [0,T] \to \mathbb{R}^n$ und eine skalare Funktion $p : \Omega \times [0,T] \to \mathbb{R}$ derart, daß

$$\begin{aligned}
\frac{\partial u}{\partial t} + (u \cdot \nabla)\, u &= -\operatorname{grad} p + \frac{1}{Re}\, \Delta\, u + f \quad \text{in } \Omega_t = \Omega \times (0,T) \\
\operatorname{div} u &= 0 \quad \text{in } \Omega_t \\
u &= 0 \quad \text{auf } \partial\Omega \times (0,T) \\
u(x,0) &= u_0(x) \quad \text{in } \Omega\,.
\end{aligned} \tag{38}$$

7.1.1 Notationen

f und u_0 sind gegeben und auf ganz $\Omega \subset \mathbb{R}^n$ bzw. $\Omega_t \subset \mathbb{R}^{n+1}$ definiert. Zur Vereinfachung sei im weiteren $\frac{1}{Re} = \nu = 1$ angenommen.

Die bis auf weiteres betrachteten Räume sind

$$\begin{aligned}
\mathcal{H} &= \{u \in \mathcal{C}_0^\infty(\Omega) \,:\, \operatorname{div} u = 0\}\,, \\
H &= \text{Vervollständigung von } \mathcal{H} \text{ in } H_0^{1,2}(\Omega)\,, \\
L &= \text{Vervollständigung von } \mathcal{H} \text{ in } L^2(\Omega)\,.
\end{aligned}$$

Die Dualräume von H und L, d.h. die Räume der linearen Funktionale auf H bzw. L seien mit H^* und L^* bezeichnet. $(\cdot,\cdot)$ sei das zum Lebesgue-Raum L gehörige Skalarprodukt

$$(u,v) = \int_\Omega u \cdot v\,,$$

$((\cdot,\cdot))$ das zum Hilbertraum H gehörige Skalarprodukt

$$((u,v)) = \int_\Omega \left[\sum_{i=1}^{n} (D_i u \cdot D_i v) \right] .$$

Zu den verwendeten Normen vergleiche Kapitel 5 und [35].

7.1.2 Schwache Lösungen des Linearen Problems

Das lineare Navier-Stokes-Problem hat folgende Gestalt:

$$\begin{aligned} \tfrac{\partial u}{\partial t} - \Delta\, u + \operatorname{grad} p &= f \quad \text{in } \Omega_t \\ \operatorname{div} u &= 0 \quad \text{in } \Omega_t \\ u &= 0 \quad \text{auf } \partial\Omega \times [0,T] \\ u(x,0) &= u_0(x) \quad \text{in } \Omega \end{aligned} \tag{39}$$

mit gegebenen Funktionen f und u_0.

Sind nun u und p klassische Lösungen von (39), also $u \in \mathcal{C}^2(\Omega)$, $p \in \mathcal{C}^1(\Omega)$, so liefert wie schon zuvor das Produkt mit einer beliebigen Funktion $\varphi \in \mathcal{H}$ in Ω und anschließende Integration über Ω

$$\int_\Omega \frac{\partial u}{\partial t} \cdot \varphi - \int_\Omega \Delta u \cdot \varphi + \int_\Omega \operatorname{grad} p \cdot \varphi \;=\; \int_\Omega f \varphi\,. \tag{40}$$

Die einzelnen Summanden kann man auf folgende Weise umformen:

$$\int_\Omega \frac{\partial u}{\partial t} \cdot \varphi \;=\; (\frac{\partial u}{\partial t}, \varphi) \;=\; (u_t, \varphi)$$

Da dies für alle $\varphi \in \mathcal{H}$ gilt, kann man dieses Skalarprodukt auch schreiben als

$$(u_t, \varphi) \;=\; \frac{d}{dt}(u, \varphi)\,.$$

Analog liefert einerseits

$$\begin{aligned}
-\int_\Omega \Delta u \cdot \varphi &= \nabla u \cdot \varphi|_{\partial\Omega} + \int_\Omega \nabla u \nabla \varphi \\
&= 0 + \int_\Omega \left[\sum_{i=1}^n (D_i u \cdot D_i v)\right] \\
&= ((u, \varphi))
\end{aligned}$$

und andererseits

$$\int_\Omega f\varphi = (f, \varphi)\,,$$

während

$$\int_\Omega \operatorname{grad} p \cdot \varphi = p \cdot \varphi|_{\partial\Omega} - \int_\Omega p \cdot \operatorname{div} \varphi = 0$$

Eingesetzt in (40) erhält man also die folgende **schwache Formulierung** des Problems (39):

Finde für gegebenes $f \in L^2(0, T\ ;\ H^*)$ und $u_0 \in L$ ein $u \in L^2(0, T\ ;\ H)$, welches der Variationsgleichung

$$\frac{d}{dt}(u, \varphi) + ((u, \varphi)) = (f, \varphi)\,. \tag{41}$$

$$\text{mit} \quad u(0) = u_0$$

für alle $\varphi \in H$ genügt.

Satz 7.1 Für gegebenes $f \in L^2(0, T\ ;\ H^*)$ und $u_0 \in L$ existiert eine eindeutige Funktion $u \in L^2(0, T\ ;\ H)$, die das Problem (41) erfüllt.

Beweis Man benutzt zum Beweis der Existenz die Galerkin-Methode. H ist separabel, d.h es existiert in H eine abzählbare dichte Teilmenge von Teilräumen $H_n = span\{h_1, \ldots, h_n\}$ mit linear unabhängigen Elementen $h_i \in H$. Also existiert eine in H vollständige Folge $(h_n)_{n\in\mathbb{N}}$ mit linear unabhängigen Folgegliedern.

Für jedes $n \in \mathbb{N}$ definiert man nun eine approximative Lösung u_n von (41) wie folgt:

$$u_n = \sum_{i=1}^n s_{in}(t) \cdot h_i \tag{42}$$

und mit $\frac{\partial u}{\partial t} = u'$

$$\begin{aligned}
(u_n', h_j) + ((u_n, h_j)) &= (f, h_j)\ ,\ j = 1, \ldots, n\,, \\
u_n(0) &= u_{0n}\,,
\end{aligned} \tag{43}$$

wobei $u_{0n} \in H_n$ mit $u_{0n} \longrightarrow u_0$ für $n \to \infty$.

Die Funktionen $s_{in}(t)$, $1 \leq i \leq n$, sind skalare Funktionen definiert auf $[0, T]$, zu deren Bestimmumg man mittels (42) und (43) ein System von Differentialgleichungen erhält.

$$\sum_{i=1}^{n}(h_i, h_j)s'_{in}(t) + \sum_{i=1}^{n}((h_i, h_j))s_{in}(t) = (f, h_j)\,,\ j = 1, \ldots, , n\,; \tag{44}$$

Die Elemente $h_1, \ldots, h_n$ sind linear unabhängig, so daß direkt folgt, daß die Matrix mit den Einträgen $(h_i, h_j)\,, 1 \leq i, j \leq n$, regulär ist. Indem man nun diese Matrix invertiert, reduziert sich das Differentialgleichungssystem (44) auf ein lineares Gleichungssystem mit konstanten Koeffizienten.

$$s'_{in}(t) + \sum_{j=1}^{n} \alpha_{ij} s_{jn}(t) = \sum_{j=1}^{n} \beta_{ij}(f(t), h_j)\,,\ i = 1, \ldots, n\,, \tag{45}$$

wobei $\alpha_{ij}, \beta_{ij} \in \mathbb{R}$. Die Anfangsbedingung von (43) ist äquivalent zu den Gleichungen

$$s_{in}(0) = \ \text{i-te Komponente von } \ u_{0n}\,. \tag{46}$$

Das lineare Gleichungssystem (45) zusammen mit den Anfangsbedingungen (46) liefert nun eindeutige Lösungen für die s_{in} auf dem ganzen Intervall $[0, T]$.

Da die skalaren Funktionen $t \mapsto (f, h_j)(t)$ zweimal differenzierbar sind, sind es die Funktionen s_{in} auch, so daß für alle $n \in \mathbb{N}$ gilt:

$$u_n \in L^2(0, T\,; H) \ \text{ und } \ u'_n \in L^2(0, T\,; H)\,. \tag{47}$$

Als nächstes sollen unabhängig von n Abschätzungen für die Funktionen u_n gefunden werden, die eine Betrachtung des Grenzüberganges für $n \to \infty$ erlauben. Dafür kehrt man wieder zurück zu den Ausgangsgleichungen

$$(u'_n, h_j) + ((u_n, h_j)) = (f, h_j)\,,\ j = 1, \ldots, n\,,$$

multipliziert mit s_{jn} und summiert die Gleichungen für $j = 1, \ldots, n$ auf. Man erhält:

$$(u'_n(t), u_n(t)) + \|u_n(t)\|^2 = (f(t), u_n(t))\,.$$

Aufgrund von (47) gilt:

$$\begin{aligned} 2(u'_n(t), u_n(t)) &= \frac{d}{dt}|u_n(t)|^2 \\ \Rightarrow \frac{d}{dt}|u_n(t)|^2 + 2\|u_n(t)\|^2 &= 2(f(t), u_n(t)) \end{aligned}$$

$$
\begin{aligned}
&& \leq\ & 2 \cdot \|f(t)\| \cdot \|u_n(t)\| \\
&& \leq\ & \|u_n(t)\|^2 + \|f(t)\|^2 \\
\Rightarrow\ & \frac{d}{dt}|u_n(t)|^2 + \|u_n(t)\|^2 & \leq\ & \|f(t)\|^2 & (48)\\
& \Rightarrow\ \frac{d}{dt}|u_n(t)|^2 & \leq\ & \|f(t)\|^2 . & (49)
\end{aligned}
$$

Integriert man (49) von 0 bis s, $T \geq s > 0$, so erhält man die Ungleichung:

$$
\begin{aligned}
& |u_n(s)|^2 & \leq\ & |u_{0n}|^2 + \int_0^s \|f(t)\|^2 \, dt \\
& & \leq\ & |u_0|^2 + \int_0^T \|f(t)\|^2 \, dt \\
\Rightarrow\ & \sup_{s\in[0,T]} |u_n(s)|^2 & \leq\ & |u_0|^2 + \int_0^T \|f(t)\|^2 \, dt < \infty \qquad (50)
\end{aligned}
$$

Die rechte Seite von (50) ist also unabhängig von n endlich, d.h.:

Die Folge u_n verbleibt in einer beschränkten Menge von $L^\infty(0, T\,;\, L)$. (51)

Ebenso erhält man durch Integration von (48) von 0 bis T die Abschätzung

$$
|u_n(T)|^2 + \int_0^T \|u_n(t)\|^2 \, dt \leq |u_0|^2 + \int_0^T \|f(t)\|^2 \, dt ,
$$

die damit zeigt:

Die Folge u_n verbleibt in einer beschränkten Menge von $L^2(0, T\,;\, H)$. (52)

Die Bedingungen (51) und (52) ermöglichen nun die Betrachtung des Grenzüberganges für $n \to \infty$:

Dank (51) existiert ein Element $u \in L^\infty(0, T\,;\, L)$ und eine Teilfolge $u_{n'}$ von u_n, derart, daß

$$
u_{n'} \longrightarrow u \quad \text{unter der schwachen Topologie von} \quad L^\infty(0, T\,;\, L)
$$

d.h. für jedes $v \in L^1(0, T\,;\, L)$ gilt:

$$
\int_0^T (u_{n'}(t) - u(t)\,,\, v(t))\, dt \longrightarrow 0 \quad \text{für } n' \to \infty . \tag{53}
$$

Durch (52) erhält man zusätzlich, daß die Teilfolge $u_{n'}$ in einem beschränkten Gebiet des $L^2(0,T\,;\,H)$ verbleibt. Das heißt ein erneuter Übergang zu einer wiedrum mit $u_{n'}$ bezeichneten Teilfolge liefert die Existenz eines $u_* \in L^2(0,T\,;\,H)$ derart, daß

$$u_{n'} \longrightarrow u_* \quad \text{unter der schwachen Topologie von} \quad L^2(0,T\,;\,H)\,,$$

d. h., es gilt für alle $v \in L^2(0,T\,;\,H')$:

$$\int_0^T (u_{n'}(t) - u_*(t)\,,\,v(t))\,dt \longrightarrow 0 \quad \text{für } n' \to \infty\,.$$

Vergleicht man dieses mit (52), so sieht man direkt, daß

$$\int_0^T (u(t) - u_*(t)\,,\,v(t)) \;=\; 0$$

für alle $v \in L^2(0,T\,;\,L)$, also

$$\Longrightarrow \quad u \;=\; u_* \;\in L^2(0,T\,;\,H) \cap L^\infty(0,T\,;\,L)\,.$$

Um jetzt den Grenzübergang von (43) für $n \to \infty$ zu betrachten, nimmt man zunächst skalare Funktionen ψ zur Hilfe, die stetig differenzierbar auf $[0,T]$ sein sollen mit

$$\psi(T) \;=\; 0\,.$$

Multipliziert man (43) mit ψ und integriert partiell, so erhält man

$$-\int_0^T (u_n\,,\,h_j \cdot \psi') \;+\; \int_0^T ((u_n\,,\,h_j \cdot \psi)) \;=\; (u_{0n}\,,\,h_j) \cdot \psi(0) \;+\; \int_0^T (f\,,\,h_j) \cdot \psi\,.$$

Betrachtet man nun an dieser Stelle den Grenzübergang für $n = n' \longrightarrow \infty$, so gilt mit den zuvor gezeigten Abschätzungen:

$$-\int_0^T (u\,,\,h_j \cdot \psi') \;+\; \int_0^T ((u\,,\,h_j \cdot \psi)) \;=\; (u_0\,,\,h_j) \cdot \psi(0) \;+\; \int_0^T (f\,,\,h_j) \cdot \psi\,.$$

Summation über alle $j = 1, \ldots, n$ ergibt

$$-\int_0^T (u\,,\,\varphi)\psi' \;+\; \int_0^T ((u\,,\,\varphi))\psi \;=\; (u_0\,,\,\varphi)\psi(0) \;+\; \int_0^T (f\,,\,\varphi)\psi$$

für jedes φ, welches endliche Linearkombination der h_j ist.

Wählt man speziell $\psi \in \mathcal{C}_0^\infty((0,T))$, so folgt

$$\frac{d}{dt}(u,\varphi) \,+\, ((u,\varphi)) \;=\; (f,\varphi) \quad \text{für alle} \quad \varphi \in H\,.$$

Also ist die als Grenznzwert gewonnene Funktion u schwache Lösung des linearen Navier-Stokes Problemes. Zum Abschluß des Existenzbeweises bleibt nur noch zu zeigen, daß für dieses u auch $u_0 = u(0)$ gilt. Dazu multipliziert wiederum die schwache Fomulierung für diese Lösung u mit oben beschriebenen ψ, integriert partiell und erhält

$$-\int_0^T (u,\varphi)\psi' \,+\, \int_0^T ((u,\varphi))\psi \;=\; (u(0),\varphi)\psi(0) \,+\, \int_0^T (f,\varphi)\psi\,.$$

Vergleicht man dieses mit der aus der Grenzbetrachtung gewonnenen Gleichung, so folgt direkt, daß

$$\begin{aligned} (u(0)-u_0\,,\,\varphi) \;&=\; 0 \quad \text{für alle} \quad \varphi \in H \\ \Longrightarrow \quad u(0) \;&=\; u_0\,, \end{aligned}$$

womit die Existenz einer Lösung bewiesen wäre. Zum Beweis der Stetigkeit und Eindeutigkeit siehe [35], S. 260-264.

7.2 Aktuelle Resultate: M.Wiegner, 1999

7.2.1 Der Stokes-Operator und die Stokes-Halbgruppe

Die analytische Behandlung partieller Differentialgleichungen basiert meist auf der Betrachtung spezifischer Differentialoperatoren in geeigneten Funktionenräumen. Bezieht man sich wie im letzten Abschnitt auf

$$\begin{aligned} \mathcal{H} \;&=\; \{u \in \mathcal{C}_0^\infty(\Omega)\,:\, \operatorname{div} u = 0\}\,, \\ H \;&=\; \text{Vervollständigung von } \mathcal{H} \text{ in } H_0^{1,2}(\Omega)\,, \\ L \;&=\; \text{Vervollständigung von } \mathcal{H} \text{ in } L^2(\Omega)\,, \end{aligned}$$

so stellt sich im Zusammenhang der Existenz- und Eindeutigkeitssätze die Frage: Gibt es eine eindeutige orthogonale Zerlegung

$$L^2(\Omega) \;=\; L \,\oplus\, G^2$$

mit einer linearen stetigen Abbildung

$$P\,:\,L^2(\Omega) \longrightarrow L$$

und wie kann man G^2 charakterisieren? Offensichtlich ist P ein von Ω abhängiger Operator, der bei Anwendung auf eine Differentialgleichung mit Lösung u weder die Randwerte von u erhält, noch mit Differentialoperatoren kommutiert, was in der Vergangenheit immer wieder zu falschen Behauptungen und Beweisen bez. der Navier-Stokes-Gleichungen geführt hat. Andererseits erhält P im allgemeinen die Stetigkeit und ist unabhängig vom Druck p. Es wird klar, daß das orthogonale Komplement zu den L-Funktionen mit div $= 0$ die Gradienten von L^2-Funktionen sein müssen. Wendet man den Operator P nun auf die Navier-Stokes-Gleichungen für $f = 0$ an, so erhält man

$$\begin{aligned} P(u_t - \Delta u + \nabla p) &= P(-u\nabla u)\,, \\ \text{wobei} \quad P\,u &= u\,, \quad \text{da } \operatorname{div} u = 0\,, \\ P\,u_t &= u_t\,, \\ P(-u\nabla u) &= -P(u\nabla u)\,, \\ P(\nabla p) &= 0\,, \\ \text{so daß} \quad u_t \underbrace{-P(\Delta u)}_{=:Au} &= -P(u\nabla u) \\ \Rightarrow\; u_t + A\,u &= -P(u\nabla u)\,. \end{aligned} \tag{54}$$

Den durch $Au = -P\Delta u$ definierten Operator bezeichnet man als **Stokes-Operator** oder **Projektor**. Mit der Hilfe dieses Operators kann das Ausgangs-Problem der partiellen Differentialgleichung auf das einfachere Problem einer gewöhnlichen Differentialgleichung zurückgeführt werden. Betrachtet man eine gewöhnliche Differentialgleichung der Form

$$y' + A\,y = f$$

mit einer Matrix $A \in Mat(n, \mathbb{R})$ und Anfangswert y_0, so liefert die Methode der Variation der Konstanten (siehe [37], S. 25) eine Lösung der Form

$$y = e^{-tA}y_0 - \int_0^t e^{-(t-s)A} f\,ds\,.$$

Die Frage, ob man diese Eigenschaften von gewöhnlichen auf partielle Differentialgleichungen übertragen kann, und wie man e^{-tA} für einen Differentialoperator A verstehen kann, läßt sich durch die **Halbgruppeneigenschaft** des Stokes-Operators A erklären, d.h., daß

$$A \text{ infinitesimaler Erzeuger einer stetigen Halbgruppe } e^{-tA}$$

ist. Diese Stokes-Halbgruppe ist beschränkt, was man z.B. über die Resolventenabschätzung gewinnen kann:

Satz 7.2 (Hille-Yosida) $A : D(A) \to L$ sei ein linearer Operator mit in dem Banachraum $L \subset L^2$ dichtem Definitionsbereich $D(A)$. Die Resolvente $R(\lambda, A) = (\lambda I - A)^{-1}$ existiere für alle $\lambda \in \mathbb{R}$ und es gelte

$$\|(I - \frac{1}{\lambda}A)^{-1}\| \leq 1 .$$

Dann erzeugt A eine eindeutig bestimmte, kontrahierende Halbgruppe.

(zum Beweis siehe [20], S. 133, und [13], S. 135)

Man hat also folgende Eigenschaften des Ausdruckes e^{-tA}:

1. Ist $u_0 \in L^2$, so ist $e^{-tA}u_0$ definiert als Lösung des Problems
$$u_t + A\,u = 0\ ,\ u(0) = u_0 .$$
2. $e^{-tA} : L \to L$ ist beschränkt.
3. e^{-tA} erfüllt die Halbgruppeneigenschaft, d.h.
$$e^{-(t+s)A}u_0 = e^{-tA}(e^{-sA}u_0) .$$

Es liegt also nahe, die Lösung der Navier-Stokes-Gleichungen durch eine Integralgleichung der Form

$$u = e^{-tA}u_0 - \int_0^t e^{-(t-s)A}P(u\nabla u)\, ds$$

zu konstruieren. Zum Beweis der Existenz der Lösung einer solchen Integralgleichung benötigt man zwei Abschätzungen der Stokes-Halbgruppe und die Höldersche Ungleichung für den Projektor P:

$$\|e^{-tA}v\|_q \leq c_0 t^{-\frac{n}{2}(\frac{1}{p}-\frac{1}{q})}\|v\|_p \quad \text{für} \quad \frac{n}{2} \leq p \leq q < \infty , \tag{55}$$

$$\|\nabla e^{-tA}v\|_n \leq c_1 t^{-\frac{n}{2p}}\|v\|_p \quad \text{für} \quad \frac{n}{2} < p \leq n , \tag{56}$$

$$\|P(v\nabla v)\|_{\frac{n}{1+\delta}} \leq c_2\|v\|_{\frac{n}{\delta}}\|\nabla v\|_n . \tag{57}$$

Vergleiche dazu [39], S. 5 und S. 9, und [22], S. 474.

7.2.2 Schwache Lösungen des Nichtlinearen Problems

Wie in Abschnitt 8.1 bereits gesehen, beschränken sich die klassischen Existenz- und Eindeutigkeitssätze auf schwache Lösungen der linearen Navier-Stokes Gleichungen. An dieser Stelle soll nun auch für die schwachen Lösungen des nichtlinearen Problems ein Existenz- und Eindeutigkeitsbeweis erbracht werden; siehe [39], S. 5-8. Die Notationen der betrachteten Funktionenräume werden aus dem vorherigen Abschnitt übernommen.

Eine schwache Lösung der Navier-Stokes-Gleichungen sei in folgendem Sinne definiert:

Definition 7.1 Sei $f \in L^2(\Omega_t)$ und $u_0 \in L$. $u \in L^2(0,T\,;\,H)$ ist schwache Lösung der Navier-Stokes-Gleichungen, falls

$$\int_0^T \int_\Omega (-u \cdot \varphi_t + \nabla u \nabla \varphi + (u\nabla)u \cdot \varphi) = \int_\Omega u_0 \cdot \varphi(0) + \int_0^T \int_\Omega f \cdot \varphi \tag{58}$$

für alle $\varphi \in H$ mit $\varphi(T) = 0$. Diese schwache Formulierung des Problems erhält man vergleichbar mit [35], S. 252, durch Multiplikation mit φ und anschließender Intergration über Ω_t.

Um die Eindeutigkeit einer solchen Lösung nachzuprüfen, bietet es sich an, die Gleichungen für zwei unterschiedliche Lösungen u_1, u_2 zu subtrahieren, mit der Differenz $v = u_1 - u_2$ zu multiplizieren und anschließend über $\Omega_t = \Omega \times [0,T]$ zu integrieren. Diese Schritte liefern nun folgende Gleichung:

$$\int_{\Omega_t} [(u_1 - u_2)_t(u_1 - u_2) - (\Delta(u_1 - u_2))(u_1 - u_2) + (u_1 \nabla u_1 - u_2 \nabla u_2)(u_1 - u_2)] = 0 \tag{59}$$

Betrachtet man die einzelnen Integrale, so erhält man:

$$\begin{aligned}
\int_{\Omega_t} (u_1 - u_2)_t(u_1 - u_2) &= \frac{1}{2}\int_\Omega \int_0^T (v^2)_t \\
&= \frac{1}{2}\int_\Omega [v^2(T) - v^2(0)] \\
&= \frac{1}{2}\int_\Omega v^2(T)\,, \\
-\int_{\Omega_t} (\Delta(u_1 - u_2))(u_1 - u_2) &= -\int_{\Omega_t} \Delta v \cdot v \\
&= -\int_0^T (\underbrace{\nabla v \cdot v|_{\partial\Omega}}_{=0} - \int_\Omega \nabla v \cdot \nabla v\,)
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\Omega_t} (\nabla v)^2\,, \\
\int_{\Omega_t} \left[(u_1\nabla)u_1 - (u_2\nabla)u_2\right](u_1-u_2) &= \int_{\Omega_t} \left[(u_1\nabla)v + (v\nabla)u_2\right] v \\
&= \int_{\Omega_t} (v\nabla)u_2\cdot v \;+\; \int_{\Omega_t} (u_1\nabla)v\cdot v \\
&= \int_{\Omega_t} (v\nabla)u_2\cdot v \;+\; \frac{1}{2}\int_{\Omega_t} u_1\nabla(v^2) \\
&= \int_{\Omega_t} (v\nabla)u_2\cdot v \;+\; \frac{1}{2}\int_0^T \Big(\underbrace{u_1 v^2\big|_{\partial\Omega}}_{=0} - \underbrace{\int_\Omega \mathrm{div} u_1\cdot v^2}_{=0}\Big) \\
&= \int_{\Omega_t} (v\nabla)u_2\cdot v\,.
\end{aligned}
$$

Also bleibt von Gleichung (59) nur noch folgendes zu betrachten :

$$
\begin{aligned}
&\frac{1}{2}\int_\Omega v^2(T) \;+\; \int_{\Omega_t} (\nabla v^2) \;+\; \int_{\Omega_t} (v\nabla)u_2\cdot v \;=\; 0 \\
\Longleftrightarrow\quad &\frac{1}{2}\|v(T)\|_2^2 \;+\; \int_0^T \|\nabla v\|_2^2 \;=\; -\int_{\Omega_t} (v\nabla)u_2\cdot v \geq 0 \qquad (60)
\end{aligned}
$$

Die rechte Seite schätzt man nun mittels Sobolev-Ungleichung und Youngscher Ungleichung ab. Die dafür notwendige Sobolev-Ungleichung ist

$$
\|v\|_4 \leq c\|v\|_2^{1-\frac{n}{4}}\|\nabla v\|_2^{\frac{n}{4}} \quad \text{für } \; n = 2, 3
$$

und die Youngsche Ungleichung liefert zuerst

$$
a\cdot b \leq \varepsilon\cdot b^p + c_\varepsilon\cdot a^q
$$

für $\frac{1}{p}+\frac{1}{q}=1$, und speziell für $p=\frac{4}{n}$, $q=\frac{4}{4-n}$

$$
a\cdot c^{\frac{4-n}{2}}\cdot b^{\frac{n}{2}} \leq \varepsilon\cdot b^2 + c_\varepsilon\cdot a^{\frac{4}{4-n}}\cdot c^2\,,
$$

so daß die folgenden Abschätzungen gerechtfertigt sind.

$$
\begin{aligned}
\left|\int_{\Omega_t} (v\nabla)u_2\cdot v\right| &= \left|\int_0^T \int_\Omega ((u_1-u_2)\nabla)u_2\cdot(u_1-u_2)\right| \\
&\leq \int_0^T \Big(\int_\Omega |\nabla u_2|^2\Big)^{\frac{1}{2}}\cdot\Big(\int_\Omega |(u_1-u_2)^2|^2\Big)^{\frac{1}{2}}
\end{aligned}
$$

$$
\begin{aligned}
& = && \int_0^T \|\nabla u_2\|_2 \cdot \|u_1 - u_2\|_4^2 \\
\text{(Sobolev Ungl.)} \quad & \leq && \int_0^T \|\nabla u_2\|_2 \cdot (c\|u_1 - u_2\|_2^{2-\frac{n}{2}} \cdot \|\nabla(u_1 - u_2)\|_2^{\frac{n}{2}}) \\
& = && c \cdot \int_0^T (\|\nabla u_2\|_2 \cdot \|u_1 - u_2\|_2^{2-\frac{n}{2}}) \cdot \|\nabla(u_1 - u_2)\|_2^{\frac{n}{2}} \\
\text{(Youngsche Ungl.)} \quad & \leq && \int_0^T \|\nabla(u_1 - u_2)\|_2^2 \; + \; \tilde{c} \int_0^T \|\nabla u_2\|_2^{\frac{4}{4-n}} \cdot \|u_1 - u_2\|_2^2
\end{aligned}
$$

Eingesetzt in Gleichung (60) erhält man

$$
\begin{aligned}
& \frac{1}{2}\|v(T)\|_2^2 \; + \int_0^T \|\nabla v\|_2^2 \\
= \; & \left| \int_{\Omega_t} (v\nabla)u_2 \cdot v \right| \\
\leq \; & \int_0^T \|\nabla v\|_2^2 \; + \; \tilde{c} \int_0^T \|\nabla u_2\|_2^{\frac{4}{4-n}} \cdot \|u_1 - u_2\|_2^2
\end{aligned}
$$

$$
\Rightarrow \; \|u_1(T) - u_2(T)\|_2^2 \; \leq \; c \int_0^T \|\nabla u_2\|_2^{\frac{4}{4-n}} \cdot \|u_1 - u_2\|_2^2 =: \int_0^T \alpha(t)\|u_1 - u_2\|_2^2 \tag{61}
$$

Diese Ungleichung würde in allen Dimensionen gelten, falls $\|\nabla u_2\|_\infty$ endlich ist, d.h. falls $\alpha(t)$ integrierbar ist. Im Fall $n = 2$ gilt für jede glatte Lösung die folgende Energie-Ungleichung

$$
\|u(t)\|_2^2 \; + \; s \int_0^1 \|\nabla u(s)\|_2^2 \, ds \; \leq \; \|u_0\|_2^2 \; + \; 2 \int_0^t \int_\Omega f \cdot u \, dx \, ds \, ,
$$

die an dieser Stelle die gesuchte Abschätzung leisten kann:

$$
\int_0^T \|\nabla u_2\|_2^2 \leq \frac{1}{2}(\|u_2(0)\|_2^2 - \|u_2(T)\|_2^2) + \int_{\Omega_t} f \cdot u_2 < \infty
$$

Dann liefert die Gronwall-Ungleichung den letzten Schritt des Beweises:

Nach Gronwall gilt: Für integrierbare α, v mit $\alpha(T) \geq 0$ und

$$
\|v(T)\|_2^2 \leq \|v(0)\|_2^2 + \int_0^T \alpha \|v\|_2^2
$$

ist

$$
\|v(T)\|_2^2 \leq \|v(o)\|_2^2 \cdot e^{N(T)} \, ,
$$

wobei

$$N(T) = \int_0^T \alpha(t)dt\,.$$

In unserem Fall ist $v = u_1 - u_2$ und $v(0) = 0$,also

$$\Rightarrow \; \|u_1(T) - u_2(T)\|_2^2 \leq 0 \cdot e^{N(T)} = 0$$

$$\Rightarrow \; u_1 \; = \; u_2\,. \quad \text{q.e.d.}$$

Der obige Beweis zeigt, daß die Existenz einer stetigen schwachen Lösung auf $(0, T)$ die Existenz weiterer Lösungen auf diesem Intervall ausschließt, also zumindest lokale Eindeutigkeit.

7.2.3 Existenz und Eigenschaften starker Lösungen

Wie vorher schon erwähnt, ist es nun das Ziel, eine starke Lösung der Navier-Stokes-Gleichungen über die Integralgleichung und die Stokes-Halbgruppe zu konstruieren. Zur Vereinfachung sei im weiteren $f = 0$ und Ω beschränktes Gebiet im $\mathbb{R}^n$ für $n \geq 3$. Um die Notationen des Problems auf die allgemeine Dimension n zu erweitern, sei die Definition des Banachraumes L auf

$$L^{n,\sigma} \;=\; \text{Vervollständigung von } \;\mathcal{H}\; \text{ in } L^n(\Omega)$$

erweitert.

Das folgende Theorem umfaßt alle zu zeigenden Eigenschaften einer Lösung (vergl. [39], Theorem 6.1, S. 9):

Theorem 7.1 Sei $a \in L^{n,\sigma}(\Omega)$ Anfangswert der vollen Navier-Stokes-Gleichungen. Dann gilt:

- Es existiert eine maximale Zeit $T > 0$ derart, daß eine Lösung $u \in C([0, T)\,;\, L^{n,\sigma})$ des Anfangs-Randwert-Problems existiert, welche eindeutig und glatt ist für $t > 0$.
- Die Normen
$$\sup_{t \leq T_1} t^{\frac{1}{2}} \|\nabla u\|_n \;\text{ und }\; \sup_{t \leq T_1} t^{(1-\frac{n}{r})/2} \|u\|_r$$
sind endlich für $T_1 < T$, $r \geq n$.
- Ist $\frac{2}{s} + \frac{n}{r} = 1$, $r > n$, so gilt
$$\int_0^{T_1} \|u\|_r^s dt \;<\; \infty\,.$$

- Ist die maximale Existenz-Zeit T endlich, so ist u gleichmäßig stetig auf $[0, T)$.
- $T = \infty$, falls $\|a\|_n$ genügend klein ist. Die oben genannten Normen mit $T_1 = \infty$ sind in diesem Falle beschränkt durch $c\|a\|_n$.

Beweis: Gemäß der bereits erwähnten Integralgleichung sei das folgende Iterationsschema definiert:

$$
\begin{aligned}
u_{j+1} &=: u_0 - F(u_j)\,, \qquad (62)\\
\text{mit } u_0(t) &= e^{-tA}a(t)\,,\\
F(v)(t) &= \int_0^t e^{-(t-s)A} \cdot P(v(s)\nabla v(s))\, ds
\end{aligned}
$$

Wähle nun ein festes $\delta \in (0, 1]$ und $0 < T \leq \infty$ und definiere

$$
\begin{aligned}
K_j &:= \sup_{t<T} t^{(1-\delta)/2}\|u_j(t)\|_{\frac{n}{\delta}}\,,\\
K_j^1 &:= \sup_{t<T} t^{1/2}\|\nabla u_j(t)\|_n \ \text{ und}\\
R_j &:= max\{K_j\,,\, K_j^1\}
\end{aligned}
$$

Wählt man nun $p = \frac{n}{1+\delta}$, $q = \frac{n}{\delta}$ in den $L^p - L^q$-Abschätzungen (55) und (57) bezüglich der Stokes-Halbgruppe, so erhält man

$$
\begin{aligned}
K_{j+1} &= \sup_{t<T} t^{(1-\delta)/2}\|u_{j+1}(t)\|_{\frac{n}{\delta}}\\
&= \sup_{t<T} t^{(1-\delta)/2}\|u_0 - F(u_j)\|_q\\
&= \sup_{t<T} t^{(1-\delta)/2}\|u_0 - \int_0^t e^{-(t-s)A} \cdot P(u_j\nabla u_j)\|_q\\
&\leq \sup_{t<T} t^{(1-\delta)/2}\|u_0\|_q + \sup_{t<T} t^{(1-\delta)/2}\| \int_0^t e^{-(t-s)A} \cdot P(u_j\nabla u_j)\|_q\\
&\leq K_0 + \sup_{t<T} t^{(1-\delta)/2} \int_0^t \|e^{-(t-s)A}P(u_j\nabla u_j)\|_q\\
\text{(Ungl. (55))} \quad &\leq K_0 + \sup_{t<T} t^{(1-\delta)/2} \int_0^t c_0(t-s)^{-\frac{n}{2}(\frac{1+\delta-\delta}{n})}\|P(u_j\nabla u_j)\|_p\\
&= K_0 + \sup_{t<T} t^{(1-\delta)/2}c_0 \cdot \int_0^t (t-s)^{-\frac{1}{2}}\|P(u_j\nabla u_j)\|_p\\
\text{(Ungl. (57))} \quad &\leq K_0 + \sup_{t<T} t^{(1-\delta)/2}c_0 \cdot c_2 \cdot \int_0^t (t-s)^{-\frac{1}{2}}\|u_j\|_q\|\nabla u_j\|_n\\
&\leq K_0 + c \cdot K_j \cdot K_j^1 \qquad (63)\\
\text{und analog } K_{j+1}^1 &\leq K_0^1 + c \cdot K_j \cdot K_j^1 \qquad (64)
\end{aligned}
$$

Ungleichungen (64) und (65) liefern direkt

$$\begin{aligned} R_{j+1} &= max\{K_{j+1}\,,\,K^1_{j+1}\} \\ &\leq max\{K_0\,,\,K^1_0\} \,+\, \gamma\cdot R_j^2 \\ &= R_0 \,+\, \gamma\cdot R_j^2 \end{aligned}$$

Setzt man nun R_0 als genügend klein voraus, z.B. $R_0 \leq \frac{1}{6\gamma}$, so erhält man induktiv

$$R_j \,\leq\, 2\,R_0 \,\leq\, \frac{1}{3\gamma}\,.$$

Betrachtet man analog die Folge

$$w_j(t) := u_j(t) - u_{j-1}(t)\;,\; w_0(t) := u_0(t)\,,$$

so erfüllen die analogen Werte $\tilde{R}_j$ die Ungleichung

$$\tilde{R}_{j+1} \,\leq\, 2\gamma\,\tilde{R}_j\,R_j \,\leq\, \frac{2}{3}\tilde{R}_j\,.$$

Für die Folge w_j heißt das:

$$\begin{aligned} \|w_{j+1}\|_n &= \|u_{j+1} - u_j\|_n \\ &= \|-\int_0^t e^{-(t-s)A}P(u_j\nabla u_j) \,+\, \int_0^t e^{-(t-s)A}P(u_{j-1}\nabla u_{j-1})\|_n \\ &\leq \int_0^t \|e^{-(t-s)A}P(u_j\nabla u_j) \,-\, e^{-(t-s)A}P(u_{j-1}\nabla u_{j-1})\|_n \\ \text{(Ungl. (55))} \quad &\leq c\cdot\int_0^t (t-s)^{-\frac{n}{2}(\frac{1+\delta-1}{n})}\|P(u_j\nabla u_j) - P(u_{j-1}\nabla u_{j-1})\|_{\frac{n}{1+\delta}} \\ &\leq c\cdot\int_0^t (t-s)^{-\frac{\delta}{2}}(\|w_j\nabla u_j\|_{\frac{n}{1+\delta}} + \|u_j\nabla w_j\|_{\frac{n}{1+\delta}}) \\ &< c\cdot\tilde{R}_j\cdot R_j \\ &\leq c\cdot(2/3)^j \end{aligned}$$

Man sieht, daß die Folge $u_k = \sum_{j=0}^k w_j$ eine Cauchyfolge in $C([0,T],L^n(\Omega))$, falls die u_j trotz der Iterationsschritte innerhalb dieses Banachraumes verbleiben. Siehe dazu [39], Beweis zu Theorem 6.1, S. 11. In diesem Fall konvergiert die Iterationsfolge gegen eine Lösung der Integralgleichung

$$u(t) \,=\, e^{-tA}a \,-\, \int_0^t e^{-(t-s)A}P(u\nabla u)\,ds\,.$$

Außerdem liefert die Ungleichung $R_j \leq 2R_0$ für $\delta = \frac{n}{r}$ die gesuchten gewichteten Abschätzungen:

$$\sup_{t<T} t^{\frac{1}{2}} \|\nabla u\|_n \;\leq\; 2R_0 \;<\; \infty \tag{65}$$

$$\sup_{t<T} t^{(1-\frac{n}{r})/2} \|u\|_r \;\leq\; 2R_0 \;<\; \infty \tag{66}$$

Der Beweis, daß $\int \|u(t)\|_r^s dt \;<\; \infty$ für $\frac{2}{s} + \frac{n}{r} = 1$, $n < r < \infty$, benötigt einen kleinen Umweg: Man definiert eine geignete Abbildung U und einen passenden Index s, so daß das Interpolationstheorem von Marcinkiewicz (siehe z.B. [34], S. 108), das im folgenden noch näher erläutert werden soll, eine geeignete Abschätzung bezüglich der L^s-Norm liefern kann. Dazu seien zunächst zwei Eigenschaften einer Abbildung

$$U \;:\; L^p(\mathbb{R}^n) \;\longrightarrow\; L^q(\mathbb{R})$$

beschrieben:

1. U heißt **stark stetig** genau dann, wenn es ein $A > 0$ gibt, so daß für alle $f \in L^p(\mathbb{R}^n)$:
$$\|U(f)\|_q \;\leq\; A \cdot \|f\|_p \,.$$
2. U ist vom **schwachen (p,q)-Typ** genau dann, wenn es ein $A > 0$ gibt, so daß für alle $f \in L^p(\mathbb{R}^n)$ und alle $\alpha > 0$ gilt:
$$meas\{t \in \mathbb{R} |\, U(f)(t) \;>\; \alpha\} \;\leq\; A^q \cdot \left(\frac{\|f\|_p}{\alpha} \right)^q .$$

Offensichtlich folgt aus 1. direkt 2., da gilt:

$$\begin{aligned} & (\alpha^q \cdot meas\{t \in \mathbb{R} |\, |U(f)(t)| \;>\; \alpha\})^{\frac{1}{q}} \\ \leq\; & (\int |U(f)(t)|^q \, dt\,)^{\frac{1}{q}} \\ \leq\; & A \cdot \|f\|_p \end{aligned}$$

Diese Eigenschaften sollen nun auf die Abbildung

$$U(f)(t) \;:=\; \left\{ \begin{array}{ccc} \|e^{-tA} P f\|_r & , & 0 \leq t \leq T \\ 0 & , & \text{sonst} \end{array} \right\} , r > n \;\; \text{fest}$$

angewendet werden. Diese ist für alle $f \in L_p$ mit $p \leq r$ definiert. Setzt man nun die benötigten Indizes als

$$p_1 \;:=\; \frac{n}{2} \;\;<\;\; p_2 \;:=\; r \quad \text{und}$$

$$s_1 \;:=\; \frac{1}{1 - \frac{n}{2r}} \;\;<\;\; s_2 \;:=\; \infty \,,$$

so gilt:

1.

$$U \,:\, L_{p_2}(\mathbb{R}^n) = L_r(\mathbb{R}^n) \longrightarrow L_\infty(\mathbb{R}) = L_{s_2}(\mathbb{R})$$

ist stark stetig, da $|U(f)(t)| \leq c_2 \|f\|_r$, denn der Operator

$$e^{-tA}P \,:\, L_r \longrightarrow L_r$$

ist, wie schon bekannt, beschränkt.

2.

$$U \,:\, L_{p_1}(\mathbb{R}^n) \longrightarrow L_{s_1}(\mathbb{R})$$

ist vom schwachen (p_1, s_1)-Typ, denn mittels Ungleichung (55) mit $p = \frac{n}{2} = p_1$, $q = r = p_2$ folgt:

$$\begin{aligned}
\|e^{-tA}Pf\|_r &\leq c \cdot t^{-\frac{n}{2}(\frac{2}{n} - \frac{1}{r})} \|f\|_{\frac{n}{2}} \\
&= c \cdot t^{-1+\frac{n}{2r}} \|f\|_{\frac{n}{2}} \\
&= c \cdot t^{-\frac{1}{s_1}} \|f\|_{\frac{n}{2}} \\
\Rightarrow \; U(f)(t) > \alpha &\Leftrightarrow \; 0 \leq t^{1-\frac{n}{2r}} \leq c \cdot \frac{\|f\|_{\frac{n}{2}}}{\alpha} \\
&\Leftrightarrow \; 0 \leq t \leq \tilde{c} \left(\frac{\|f\|_{\frac{n}{2}}}{\alpha} \right)^{s_1} ,
\end{aligned}$$

so daß $meas\{t \,|\, U(f)(t) > \alpha\} \leq \tilde{c} \cdot \left(\frac{\|f\|_{p_1}}{\alpha} \right)^{s_1}$.

Das **Interpolationstheorem von Macinkiewicz** sagt nun allgemein, daß

$$\|U(f)\|_{L_s(\mathbb{R})} \leq c \cdot \|f\|_{L_p(\mathbb{R}^n)} ,$$

das heißt

$$\left(\int_0^T \|e^{-tA}Pf\|_r^s \, dt \right)^{\frac{1}{s}} \leq c \cdot \|f\|_{L_p(\mathbb{R}^n)}$$

mit von T unabhängiger Konstante c, falls für die Indizes p,s und den Interpolationsparameter θ gilt:

$$p \in (p_1 , p_2) \qquad \text{mit}$$

$$\begin{aligned}
\frac{1-\theta}{p_1} + \frac{\theta}{p_2} &= \frac{1}{p} \quad \text{und} \\
\frac{1-\theta}{s_1} + \frac{\theta}{s_2} &= \frac{1}{s} .
\end{aligned}$$

Wählt man an dieser Stelle nun $p = n$ mit $n \in (\frac{n}{2}, r)$, so ergibt sich daraus folgendes θ:

$$\begin{aligned} \frac{1-\theta}{\frac{n}{2}} + \frac{\theta}{r} &= \frac{1}{n} \\ \Rightarrow\ 2 - 2\theta + \frac{n}{r}\theta &= 1 \\ \Rightarrow\ \theta &= \frac{1}{2 - \frac{n}{r}} \quad (<1)\,. \end{aligned}$$

Nimmt man wie gehabt $s_1 = \infty$ und $s_2 = (1 - \frac{n}{2r})^{-1}$, so berechnet sich das zugehörige s mit

$$\begin{aligned} s &= \frac{s_1}{1-\theta} = \frac{2r}{r-n} \\ \Rightarrow\ \frac{2}{s} &= \frac{r-n}{n} = 1 - \frac{n}{r} \\ \Rightarrow\ \frac{2}{s} + \frac{n}{r} &= 1\,. \end{aligned}$$

Wendet man nun letztendlich U speziell auf a an, so erhält man mit dem Theorem von Marcinkiewicz direkt, daß

$$\left(\int_0^T \| \underbrace{e^{-tA}Pa}_{=u_0} \|_r^s \, dt \right)^{\frac{1}{s}} \le c \cdot \|a\|_n$$

$$\Rightarrow \int_0^T \|u_0(t)\|_r^s dt \le \tilde{c} \cdot \|a\|_n^s < \infty \quad \text{für} \quad \frac{2}{s} + \frac{n}{r} = 1\,, \tag{67}$$

womit der lineare Teil der gesuchten Abschätzung gezeigt ist. Den Rest liefert das Iterationsschema:

Ist $r = \frac{n}{\delta}$, $s = \frac{2}{1-\delta} > 2$, so gilt:

$$\begin{aligned} \|u_{j+1}\|_{\frac{n}{\delta}} &= \|u_0 - F(u_j)\|_{\frac{n}{\delta}} \\ &= \|u_0 - \int_0^t e^{-(t-s)A} P(u_j \nabla u_j)\|_{\frac{n}{\delta}} \\ &\le \|u_0\|_{\frac{n}{\delta}} + \|\int_0^t e^{-(t-s)A} P(u_j \nabla u_j)\|_{\frac{n}{\delta}} \\ \text{(Ungl. 55)} \quad &\le \|u_0\|_{\frac{n}{\delta}} + \int_0^t \|e^{-(t-s)A} P(u_j \nabla u_j)\|_{\frac{n}{\delta}} \\ \text{(Ungl. 57)} \quad &\le \|u_0\|_{\frac{n}{\delta}} + c_0 \cdot \int_0^t (t-s)^{-\frac{n}{2}(\frac{1}{p} - \frac{\delta}{n})} \|P(u_j \nabla u_j)\|_p \end{aligned}$$

$$\begin{aligned}(p = \frac{n}{1+\delta}) \quad &\leq \quad \|u_0\|_{\frac{n}{\delta}} \,+\, c_1 \cdot \int_0^t (t-s)^{-\frac{1}{2}} \|u_j\|_{\frac{n}{\delta}} \|\nabla u_j\|_n \\ (\text{Ungl. } 66) \quad &\leq \quad \|u_0\|_{\frac{n}{\delta}} \,+\, c \cdot R_0 \cdot \int_0^t (t-s)^{-\frac{1}{2}} t^{-\frac{1}{2}} \|u_j\|_{\frac{n}{\delta}}\end{aligned}$$

Wählt man nun ein β mit $2 < \beta < s$ und wendet die Höldersche Ungleichung an, so ergibt sich daraus

$$\|u_{j+1}(t)\|^{\beta}_{\frac{n}{\delta}} \;\leq\; c\|u_0(t)\|^{\beta}_{\frac{n}{\delta}} \,+\, cR_0^{\beta} t^{-1} \int_0^t \|u_j(\tau)\|^{\beta}_{\frac{n}{\delta}} \, d\tau\,. \tag{68}$$

Den nächsten Schritt liefert Hardy's Ungleichung, die folgendes besagt:

Für $F = \int_0^t f \, d\tau$ und $p > 1$ gilt

$$\int_0^{\infty} (F \cdot t^{-1})^p dt \;<\; (\frac{p}{p-1})^p \int_0^{\infty} f^p \, d\tau\,,$$

so daß man aus $\frac{s}{\beta} > 1$ erhält:

$$\begin{aligned}&\left(\int_0^T (\underbrace{cR_0^{\beta} \int_0^t \|u_j(\tau)\|^{\beta}_{\frac{n}{\delta}} \, d\tau}_{=F} \cdot t^{-1})^{\frac{s}{\beta}} \right)^{\frac{1}{s}} \\ \leq \quad & \left(\left(\frac{\frac{s}{\beta}}{\frac{s}{\beta} - 1} \right)^{\frac{s}{\beta}} \cdot \int_0^T \tilde{c} R_0^s \|u_j(\tau)\|^{s}_{\frac{n}{\delta}} \right)^{\frac{1}{s}} \\ = \quad & \left(\frac{s}{s-\beta} \right)^{\beta} \cdot \left(\tilde{c} R_0^s \int_0^T \|u_j(\tau)\|^{s}_{\frac{n}{\delta}} \right)^{\frac{1}{s}} \\ = \quad & c \cdot R_0 \cdot \underbrace{\left(\int_0^T \|u_j(\tau)\|^s_r d\tau \right)^{\frac{1}{s}}}_{=:I_j} \\ = \quad & c \cdot R_0 \cdot I_j \end{aligned} \tag{69}$$

Wir haben also die Ungleichung

$$I_{j+1} \;\leq\; cI_0 \,+\, cR_0 I_j$$

für $I_j = (\int_0^T \|u_j(t)\|^s_r dt)^{\frac{1}{s}}$ erhalten, die uns nun durch Grenzübergang die gesuchte Abschätzung liefert.

$$\int_0^T \|u(t)\|^s_r dt \;\leq\; c \int_0^T \|u_0(t)\|^s_r dt \;<\; \infty \tag{70}$$

Also ist eine Lösung der Navier-Stokes-Gleichungen als Lösung der Integralgleichung auf dem Intervall [0,T] konstruiert worden, unter der Voraussetzung, daß R_0 hinreichend klein ist. Hierbei war R_0 folgendermaßen gewählt:

$$
\begin{aligned}
u_0(t) &= e^{-tA}a(t)\,, \\[2ex]
K_0 &= \sup_{t\leq T} t^{(1-\delta)/2}\|e^{-tA}a\|_{\frac{n}{\delta}} \\
\text{(Ungl. 55)} \quad &\leq \sup_{t\leq T} t^{(1-\delta)/2}\cdot c_0\cdot t^{-\frac{n}{2}(\frac{1}{n}-\frac{\delta}{n})}\|a\|_n \\
&= c_0\cdot\|a\|_n \\[2ex]
K_0^1 &= \sup_{t\leq T} t^{1/2}\|\nabla e^{-tA}a\|_n \\
\text{(Ungl. 56)} \quad &\leq \sup_{t\leq T} t^{1/2}\cdot c_1\cdot t^{-\frac{n}{2n}}\|a\|_n \\
&= c_1\cdot\|a\|_n \\[2ex]
\Rightarrow \quad R_0 &= \max\{K_0\,,\,K_0^1\} \\
&\leq c\cdot\|a\|_n
\end{aligned}
$$

Da unabhängig von T gilt, daß $R_0 \leq c\|a\|_n$, implizieren also kleine Anfangswerte bereits die globale Existenz einer Lösung.

Andererseits genügt R_0 auch für beliebiges $T > 0$ und $\mu > 0$ (siehe [39], S. 12) der Ungleichung

$$R_0 \;\leq\; \|a - e^{-\mu A}\|_n \;+\; \|a\|_n\left(\frac{T}{T+\mu}\right)^{\frac{1}{2}}\,.$$

Wählt man hier zuerst ein geeignet kleines μ und dann ein passendes T, so liefert diese Ungleichung auch für große Anfangswerte eine lokale Beschränkung von R_0 und damit die Existenz einer lokalen Lösung. Zum Schluß bleibt dann nur noch zu zeigen, daß die sogenannte milde Lösung der Integralgleichung bereits starke Lösung der Navier-Stokes-Gleichungen ist. H. Kozono und T. Ogawa zeigten in [24], S. 799, ähnlich wie M. Wiegner in [39], S. 12, daß für jede milde Lösung u der Navier-Stokes-Gleichungen gilt:

$$
\begin{aligned}
u(t) &= e^{-tA}a - \int_0^t e^{-(t-s)A}P(u\nabla u)\,ds \;\in\; D(A)\,, \\
Au &\in C_0((0,T)\,,\,L^{\frac{n}{1+\delta}})
\end{aligned}
$$

$$
\begin{aligned}
\text{mit } \|Au\|_n \;&\leq\; ct^{-1}(\log t)^{1-1/n} \text{ für genügend großes } t \\
&\;\; \text{und } u_t + Au + P(u\nabla u) = 0\,,
\end{aligned}
$$

womit das betrachtete Existenz- und Eindeutigkeitstheorem vollendet wäre.

An dieser Stelle sind die wesentlichen theoretischen Grundlagen zum Verhalten numerisch gewonnener Lösungen der Navier-Stokes-Gleichungen geklärt. Kehrt man jetzt wieder zurück zum eigentlichen Ausgangsproblem, den künstlichen Herzklappen, so ist es also möglich, den Blutfluß durch eine künstliche Herzklappe mittels der Navier-Stokes-Gleichungen mit geeigneten Randbedingungen so exakt wie möglich zu beschreiben. Die Anwendung numerischer Lösungverfahren liefert anschließend eine eindeutige Darstellung von Geschwindigkeit und Druck an jedem Ort innerhalb des berechneten Gebietes und zu jedem betrachteten Zeitpunkt.

Der nächste Schritt ist jetzt die Implementierung der Gleichungen, der passenden Randbedingungen, einer geeigneten Geometrie des zu diskretisierenden Gebietes und eines Finite-Elemente-Lösers, der in der Lage ist, das nötige Gleichungssystem aufzustellen und effektiv zu lösen.

8 Implementierung unter “Diffpack 3.0”

Das folgende Kapitel soll eine Möglichkeit zur Implementierung eines Finite-Elemente-Lösers (FE-Lösers) für die Navier-Stokes Gleichungen aufzeigen. Für den Aufbau eines geeigneten FE-Lösers wurde “Diffpack 3.0” benutzt, eine flexible und vielfältige Software, die speziell zur Lösung partieller Differentialgleichungen mit der Methode der finiten Elemente konstruiert wurde.

Diffpack 3.0 ist eine Sammlung von Bibliotheken, die alle notwendigen Bausteine zur Lösung komplexer Probleme und Aufgaben enthalten: Lineare und nichtlineare Gleichungssysteme, numerische Lösungsmethoden für diese Systeme, Gittererzeuger, skalare Felder und Vektorfelder auf Gittern, verschiedene Elementtypen und letztendlich einige Tools zur Visualisierung der Ergebnisse. Diffpack 3.0 ist in C++ geschrieben, was die Vor- und Nachteile der objektorientierten Programmierung mit sich bringt. Die Details werden bei genauerer Betrachtung des Programm-Codes deutlich.

Das Ziel dieses Kapitels ist es, eine Lösung der **stationären**, **inkompressiblen**, **2-dimensionalen Navier-Stokes Gleichungen** auf der Geometrie einer mechanischen Aortenklappe zu finden und zu visualisieren. Die zu lösenden Gleichungen sind also

$$
\begin{aligned}
(u \cdot \nabla)\, u &= -\mathrm{grad}\, p + \frac{1}{Re}\Delta u \quad \text{auf dem ganzen Gebiet} \quad \Omega\,, \\
\mathrm{div}\, u &= 0 \quad \text{in ganz} \quad \Omega \quad \text{und} \\
u &= 0 \quad \text{auf dem Rand} \quad \partial\Omega\,.
\end{aligned}
$$

8.1 Geometrie

Das Gebiet an einer mechanischen, zweiflügeligen Herzklappe weißt mehrere Symmetrieebenen auf, wie z. B. aus Abbildung 12, S. 26, entnehmbar ist. Ebenso wie in der Doktorarbeit von M. J. King, die im folgenden als Leitfaden zur Modellierung dient, sollen also nur auf der oberen Hälfte der Klappe und des folgenden Aortenabschnittes Geschwindigkeit und Druck berechnet werden. Auf diesem Gebiet soll bei unterschiedlichen Klappenstellungen bzw. variierenden Öffnungswinkeln (75°, 80°, 85° und 90°) ein Finite-Elemente Gitter erzeugt werden. Die Penalty-Methode, die sowohl hier als auch in der Arbeit von King verwendet wurde und im folgenden noch näher erläutert wird, liefert beste Ergebnisse auf quadratischen oder rechteckigen Elementen und unakzeptable Werte auf Triangulierungen. Viele gängige Gitter-Erzeugungs-Tools sind damit für diese Zwecke nicht geeignet und man muß z. B.

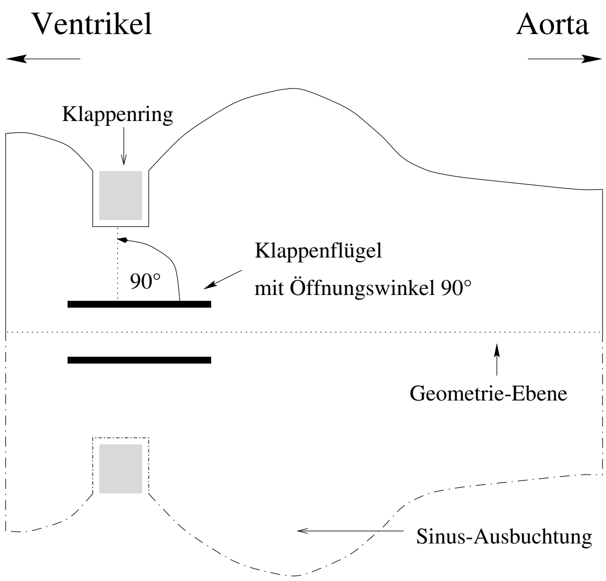


Abbildung 15: Abstrahiertes, 2-dimensionales Gebiet zur Beschreibung einer zweiflügeligen, mechanischen Herzklappe. Die weitere Modellierung beschränkt sich auf die obere Hälfte; Größenangaben aus [23], S. 44ff.

auf die Diffpack-internen Methoden zurück greifen. Das stark abstrahierte, zur Modellierung verwendete Gebiet ist in Abbildung 15 skizziert.

Zur Erzeugung eines geeigneten Gitters wurden zwei unterschiedliche Methoden verwendet: Einerseits findet man unter den Diffpack-Beispielen eine Geometrie mit dem bezeichnenden Namen "`BOX_WITH_BELL`", die ein 2-dimensionales Rohr mit einer Ausbuchtung in Form einer Gauss-Kurve beinhaltet. Benutzt man hier die unter Diffpack bereits implementierte Funktion "`addBoIndsNodes`", zur Definition zusätzlicher Randpunkte mit den dazugehörigen Randindikatoren, so kann man das in Abbildung 15 gezeigte Gebiet recht einfach konstruieren. (siehe Anhang B.6) Der Nachteil dieser Methode besteht darin, daß die Funktion "`addBoIndsNodes`" als Eingabe mit Intervallen in x- und in y-Richtung arbeitet, so daß eine Schrägstellung des Klappenflügels innerhalb dieser Methode nicht möglich war.

Um komplexere Gitter zu generieren, kann man die in Diffpack bestehende **Super-Elemente-Methode** benutzen, die das Gebiet in geometrisch einfache Teilgebiete gliedert, diese jeweils mit einer adäquaten Zerlegung versieht und schließlich ein komplettes Gitter zusammensetzt. (siehe dazu [27], Kapitel "Super-Elements") Diese Methode hat einige sehr empfindliche Parameter, die man besonders beachten sollte, wenn man den Übergang zwischen den einzelnen Super-Elementen gestaltet:

Übergibt man einem Super-Element die Randdaten des Nachbar-Elementes, so kann man zu viele oder auch zu wenige Angaben machen, was direkt zur Divergenz des iterativen Lösers führen kann. Leider sind die Gesetztmäßigkeiten hier nicht offensichtlich, so daß auch die in dieser Arbeit gelieferten Gitter und daraus resultierenden Werte nicht optimal sind. Aber die Methode der Super-Elemente birgt eine große Ausbaufähigkeit und Flexibilität in sich. Auf diese Weise sind z. B. auch verschiedene Öffnungswinkel der Klappenflügel realisierbar. Das Gitter mit Öffnungswinkel 75° hat folgende Gestalt:

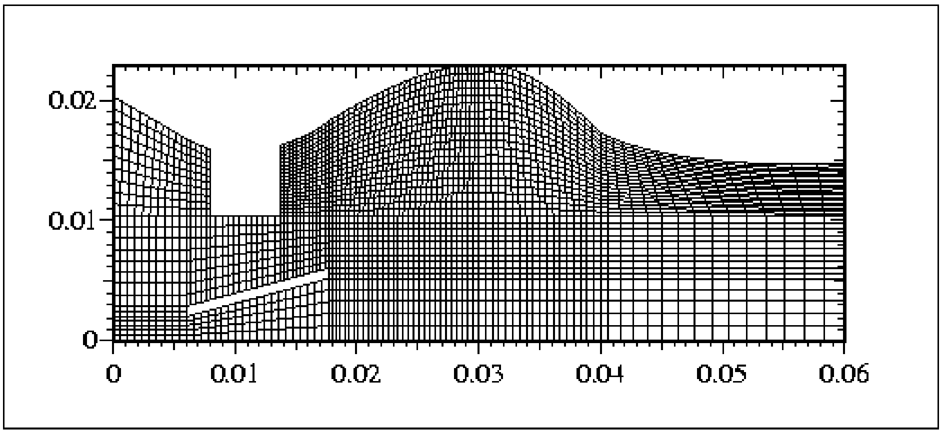


Abbildung 16: FE-Gitter bei Öffnugswinkel 75° unter Verwendung der Super-Elemente-Methode

Zu weiteren Details zur Gittergenerierung siehe Anhang B.4.

8.2 Penalty-Methode

Ein Hauptproblem bei der Lösung der Navier-Stokes Gleichungen ist die Berechnung der Druckes p bzw. die Behandlung des Termes $\operatorname{grad} p$ dar. Eine zwar sehr empfindliche, aber zur Implementierung gut geeignete Methode ist die **Penalty-Methode**, die die Kontinuitätsgleichung approximiert durch:

$$\begin{aligned} \operatorname{div} u &= -\lambda^{-1} \cdot p \quad \text{für} \quad \lambda \to \infty \, , \, \lambda \in \mathbb{R} \\ \Leftrightarrow \quad p &= -\lambda \cdot \operatorname{div} u \, , \end{aligned}$$

so daß der Druckterm aus den zu lösenden Gleichungen eliminiert werden kann. Diese auf den ersten Blick sehr einfach erscheinende Methode erfordert einige Vorsichtsmaßnahmen: Einerseits sollte für die λ-Terme in den folgenden Galerkin-Schritten eine andere Integrationsregel verwendet werden, als für die λ-freien Terme (siehe dazu

[26], S. 431, und Programmcode) und andererseits ist die Konditionszahl der Koeffizientenmatrix im zu lösenden linearen System proportional zu λ. Wählt man also ein großes λ zur guten Approximation der Gleichung $\operatorname{div} u = 0$, so sinkt die Konvergenzgeschwindigkeit der iterativen Löser stark. Da der Effekt der üblichen Vorkonditionierungsmethoden auch nur verschwindend gering ist, wird das lineare Gleichungssystem in jedem nichtlinearen Newton-Raphson- oder Successive-Substitution-Schritt durch direkte Methoden wie die Gauss-Elimination gelöst. Durch diese Schwierigkeiten sind Simulations-Programme, die auf der Penalty-Methode basieren, auf Gebiete bzw. Gitter mit wenigen Tausend Knotenpunkten beschränkt.

Setzt man also den approximierten λ-Term in die Gleichungen ein, so erhält man:

$$\begin{aligned} (u \cdot \nabla)\, u &= \lambda \operatorname{grad}(\operatorname{div} u) + \frac{1}{Re} \Delta\, u \quad \text{und} \\ \operatorname{div} u &= -\lambda^{-1} \cdot p\,. \end{aligned}$$

Verwendet man nun statt der Reynolds-Zahl explizit die Dichte ρ und die Viskosität μ, so erhält man für ein geeignet gewähltes λ:

$$\rho\, (u \cdot \nabla)\, u \;-\; \mu\, \Delta\, u \;-\; \lambda \operatorname{grad}(\operatorname{div} u) \;=\; 0$$

Die Methode der gewichteten Residuen bzw. hier die Galerkin-Methode (siehe Kapitel 6) liefert nun durch Integration über das ganze Gebiet Ω die zu implementierende dikretisierte Gleichung. Wie schon zuvor seien die Basisfunktionen mit N_i bezeichnet, so daß die Approximation der Geschwindigkeit u die Gestalt

$$\tilde{u}_k \;=\; \sum_{j=1}^{n} u_j^k\, N_j$$

erhält, wobei die (u_j^k) also die gesuchten Koeffizienten für $j = 1, \ldots, n$ und $k = 1, 2$ (bei 2 räumlichen Dimensionen) sind. Mit den gleichen N_i als Wichtungsfunktionen, unter Berücksichtigung der Randbedingungen und mittels partieller Integration ensteht das folgende Gleichungssystem:

$$\left(\int_\Omega \rho \cdot N_i \cdot \tilde{u}_k \cdot \nabla N_j \;+\; \int_\Omega \mu \cdot \nabla N_i \cdot \nabla N_j \;+\; \int_\Omega \lambda \cdot \nabla N_i \cdot \nabla \tilde{u}_k \right) \cdot (u_j^k) \;=\; 0$$

für $i, j = 1, \ldots, n$ und $k = 1, 2$.

8.3 Lösung des nichtlinearen Systemes

Im folgenden sollen die beiden verwendeten Methoden zur Lösung des oben beschriebenen nichtlinearen Systemes genauer erläutert werden. Beide werden getrennt voneinander auf die λ-Terme (siehe im Programmcode Funktion `intergrandsNonReduced`) und die λ-freien Terme (siehe Funktion `integrandsReduced`) angewendet.

8.3.1 Successive Substitutions

Eine Methode zur Lösung nichtlinearer Gleichungssysteme wie das oben beschriebene ist die einfach Iterationsmethode der "Successive Substitutions", auch bekannt unter dem Namen "Picard Iterationen". Anstelle des nichtlinearen Systemes der Form

$$A(v) \cdot v \;=\; b$$

wird in jedem Iterationsschritt das System

$$A(v^k) \cdot v^{k+1} \;=\; b$$

für $k = 0, 1, 2, \ldots$ gelöst, bis der Fehler $\|u^{k+1} - u^k\|$ genügend klein ist. Ein geeigneter Anfangsvektor ist meist $v^0 = 0$. Falls dieses Methode nicht konvergiert, so hilft in den meisten Fällen ein Ralaxationsparameter $\omega \in (0, 1)$:

Wie schon bei anderen iterativen Verfahren beschrieben, wird zuerst eine neue Approximation v^* aus $A(v^k) \cdot v^* = b$ berechnet. Den neuen Wert v^{k+1} gewinnt man dann aus dem gewichteten Mittel von v^k und v^*:

$$v^{k+1} \;=\; \omega\, v^* \;+\; (1-\omega)\, v^k \,.$$

8.3.2 Newton-Raphson-Methode

Zur Erläuterung der "Newton-Raphson-Methode" betrachten wir zunächst nur eine nichtlineare Gleichung $F(v) = 0$ mit nur einer Unbekannten v. Die zugrunde liegende Idee ist, das Gleichungssystem $F(v)$ in Abhängigkeit von einer leichter zu berechnenden approximativen Lösung v^k anzunähern. Man löst also ein lieneares Gleichungssystem $M(v; v^k) \approx F(v)$ mit der Eigenschaft, daß $M(v; v^k) = 0$ einfach zu lösen ist. Die Form dieses linearen Gleichungssytemes erinnert an den linearen Teil der Taylor-Entwicklung von F im Punkt $v = v^k$:

$$M(v; v^k) \;=\; F(v^k) \;+\; \frac{dF}{dv}\,(v^k)(v \;-\; v^k)\,.$$

Ist die nächste iterativ gewonnene Lösung v^{k+1} nun die Lösung von $M(v; v^k) = 0$, so gilt also

$$v^{k+1} \;=\; v^k \;-\; \frac{F(v^k)}{\frac{dF}{dv}(v^k)}\,.$$

Dieses Schema kann nun recht einfach auf mehrdimensionale Probleme übertragen werden. Wieder approximiert man $F(v)$ durch eine lineare Funktion $M(v; v^k)$ in Abhängigkeit von einer bestehenden Approximation v^k zu v:

$$M(v; v^k) \;=\; F(v^k) \;+\; J(v^k) \cdot (v \;-\; v^k)\,,$$

wobei $J = \nabla F$ die **Jacobi-Matrix** von F bezeichnet. Zur Verbesserung der Konvergenz kann wie bei den "Successive Substitutions" ein Relaxationsparameter $\omega \in (0, 1)$ eingefügt werden.

8.4 Programmaufbau

Zur Lösung des betrachteten Problemes wurde ein unter Diffpack bereits vorhandener FE-Löser der Navier-Stokes Gleichungen auf das betrachtete Problem angepaßt. Da nicht alle darin verwendeten Klassen für den Benutzer offensichtlich und frei zugänglich sind, war es in der vorhandenen Zeit nicht möglich, die in diesem Fall überflüssige Zeitschleife zu eliminieren.

Das Simulationsprogramm `NsPenalty.cpp` bzw. der Programmcode hat folgenden Aufbau:

→ `define`

Definition der Items des GUI-Menüs

→ `scan`

Einlesung der Parameter aus dem GUI-Menü: Gitter, Randindikatoren, Viskosität, Dichte, Penalty-Parameter usw. ⇒ Initialisierung der Variablen

→ `fillEssBC`

Randindikatoren:

bo-ind 1: inlet-Geschwindigkeitsfeld

bo-ind 2: u=0

bo-ind 3: v=0

bo-ind 4: w=0

⇒ Zuordnung der Randwerte zu den entsprechenden Randpunkten

→ `solveProblem`

Aufruf des eigentlichen Lösers mit innerer Zeitschleife

→ `setIC`
Anfangsbedingungen werden auf 0 gesetzt.

→ `timeLoop`
Zeitschleife: In jedem nächsten Zeitschritt wird Funktion `solveAtThisTimeStep` aufgerufen!

- → `inletVelocity`
  Randbedingung bei x=0: parabolisches Inlet-Geschwindigkeits-Profil
- → `integrands`
  Berechnung des Druckes p mittels der Penalty-Methode
- → `solveAtThisTimeStep`
  Aufruf der Funktion `makeAndSolveLinearSystem` mit passenden Randbedingungen zum gegenwärtigen Zeitpunkt
  - → `makeAndSolveLinearSystem`
    Initialisierung und Lösung des eigentlichen globalen Gleichungssystemes
  - → `calcElmMatVec`
    Berechnung und Zusammensetztung der Element-Matrizen und -Vektoren
  - → `integrands`
    Berechnung der einzelnen Einträge für alle Elemente mit jeweiliger Integrationsregel
    - → `integrandsNonReduced`
      Berechnung der λ-freien Terme, Implementierung von Newton Raphson und Successive Substitution
    - → `intergrandsReduced`
      Berechnung der λ-Terme mit Newton Raphson oder Successive Substitution

Details können dem Programmcode dem Anhang B entnommen werden. Zu weiteren Informationen zu den unter Diffpack benutzten C++-Klassen siehe [26], S. 431ff.

8.5 Ergebnisse

Die Lösungen für Druck p bzw. Geschwindigkeit v bei unterschiedlichen Öffnungswinkeln der Klappenflügel werden als Farb-Plot bzw. als Vektor-Plot ausgegeben. Die folgenden Abbildungen sind mit dem in Diffpack zu findenden Visualisierungs-Too "`Plotmtv`" entstanden. Siehe dazu [26], S. 232ff. Es wird stets Blutfluß von links nach rechts betrachtet.

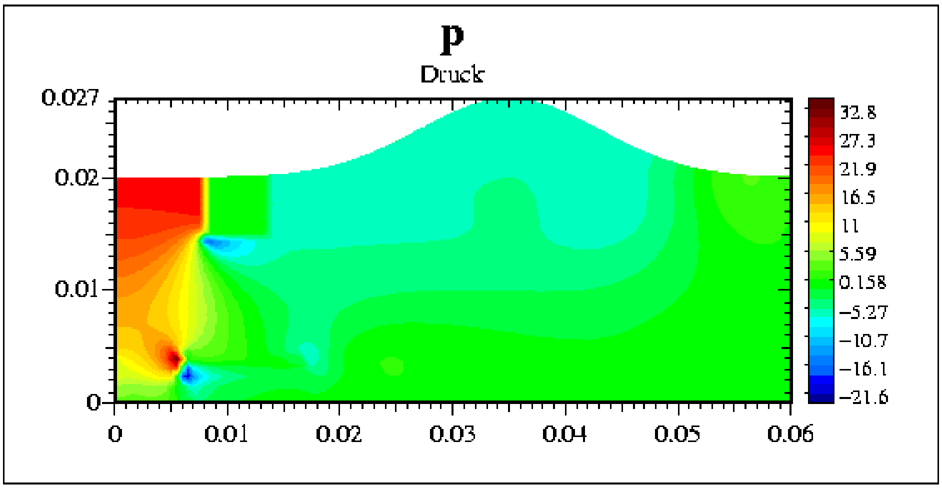


Abbildung 17: Druck p auf `BOX_WTH_BELL`-Geometrie bei Öffnungswinkel 90°

Betrachtet man die visualisierten Lösungen des FE-Lösers für p und v, so kann man deutlich die Wirbelbildungen in der Sinus-Ausbuchtung und in der Nähe der Klappenflügel erkennen. Verfolgt man nun die Entwicklung der Druck- und Geschwindigkeits-Ergebnisse bei abnehmendem Öffnungswinkel von 90° bis 75°, dann wird deutlich, daß sich außer im Sinus noch Wirbelbildungen am unteren, linken Rand des Klappenflügels finden lassen, durch die eine Region mit leichtem Unterdruck entsteht. Dieser trägt im Realfall maßgebend zur Bewegung des Klappenflügels bei, indem der "Sog" am linken Ende und der Druck durch den Sinus-Wirbel am rechten Ende des Klappenflügels diesen zurück schwingen lassen. Bei einem Öffnungswinkel von 90° ist diese Wirbelbildung direkt an dem Klappenflügel nicht zu sehen, während sie bei 75° stark ausgeprägt ist.

Im Vergleich mit den Ergebnissen von Mary J. King stellt man fest, daß auch dieses stark vereinfachte Modell schon erahnen läßt, was King mit einem 3-dimensionalen, zeitabhängigen Modell gezeigt hat: Das ideale Verhältnis zwischen den unterschiedlichen Kräften, die auf die Klappenflügel einwirken, ist bei einem Öffnungswinkel

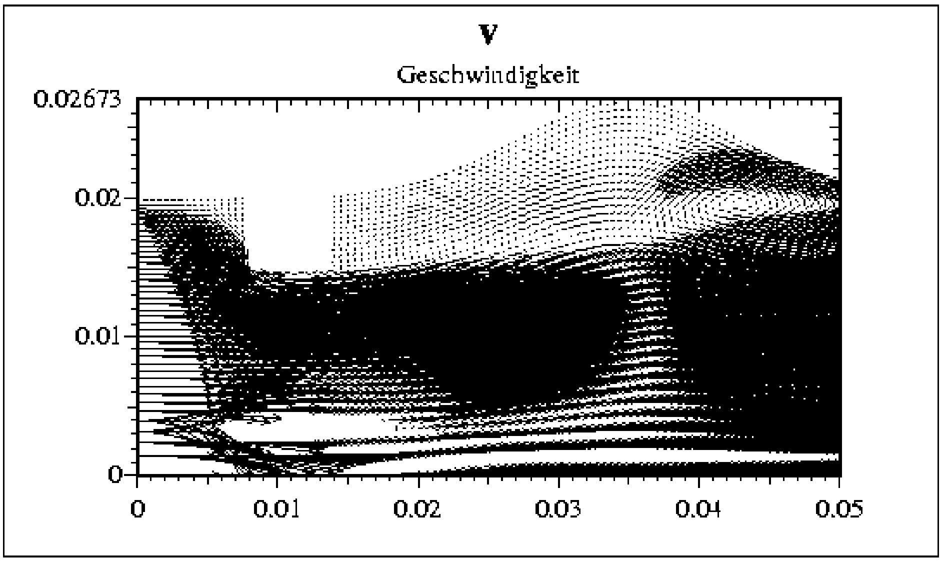


Abbildung 18: Geschwindigkeit v auf `BOX_WTH_BELL`-Geometrie bei Öffnungswinkel 90°

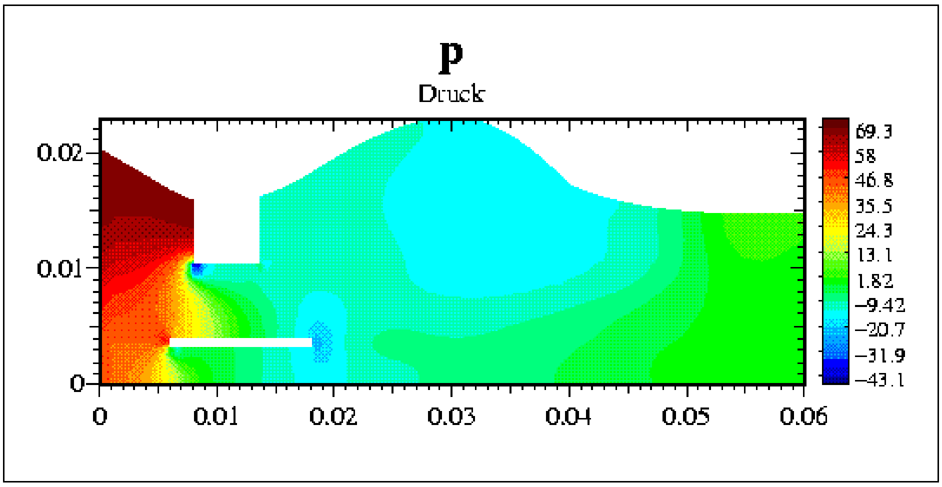


Abbildung 19: Druck p auf Superelemente-Geometrie bei Öffnungswinkel 90°

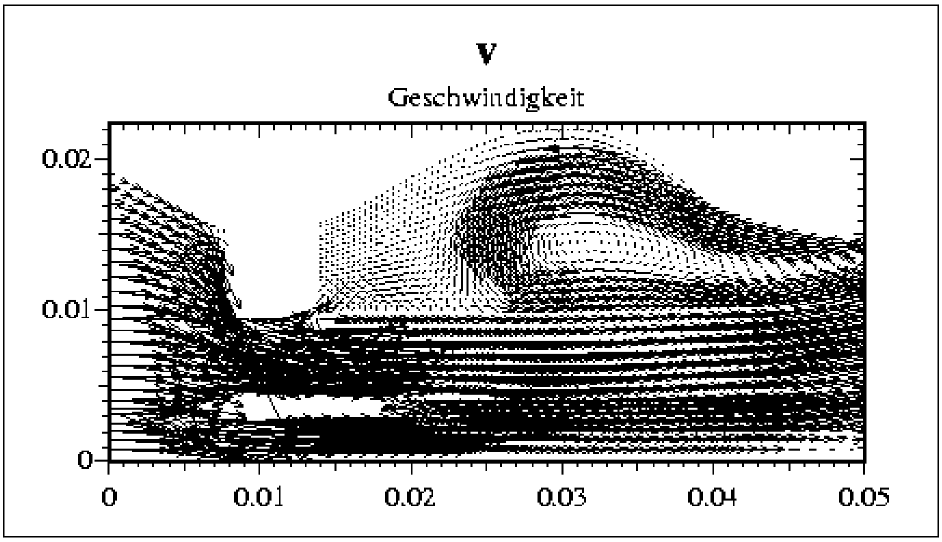


Abbildung 20: Geschwindigkeit v auf Superelemente-Geometrie bei Öffnungswinkel 90°

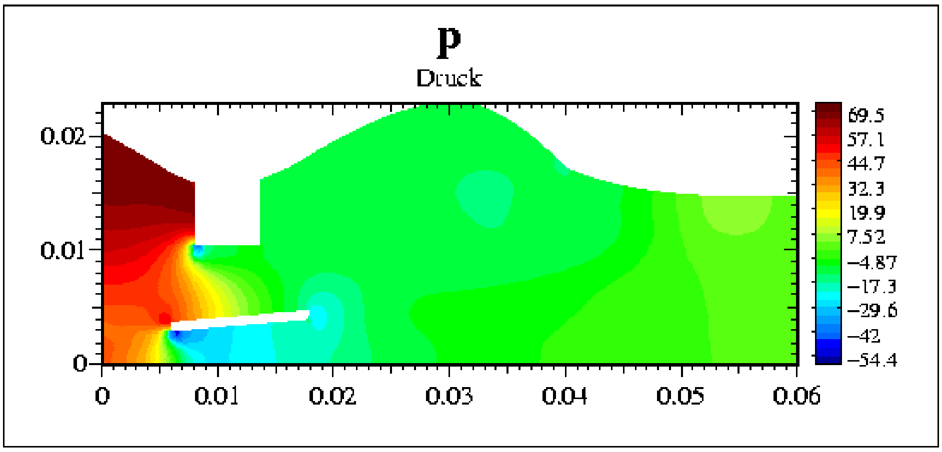


Abbildung 21: Druck p auf Superelemente-Geometrie bei Öffnungswinkel 85°

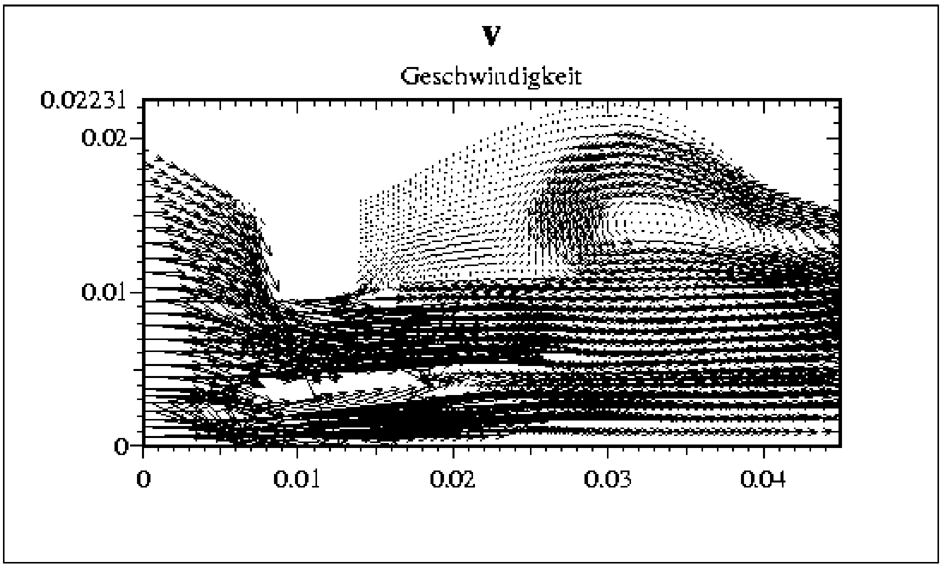


Abbildung 22: Geschwindigkeit v auf Superelemente-Geometrie bei Öffnungswinkel 85°

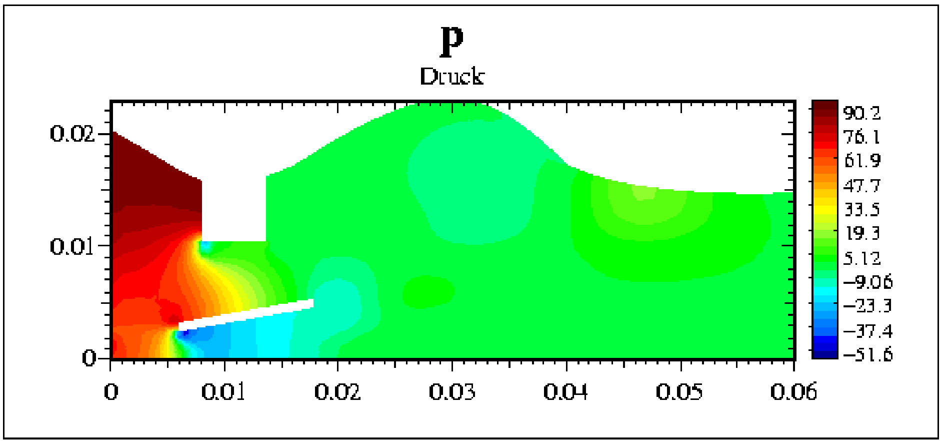


Abbildung 23: Druck p auf Superelemente-Geometrie bei Öffnungswinkel 80°

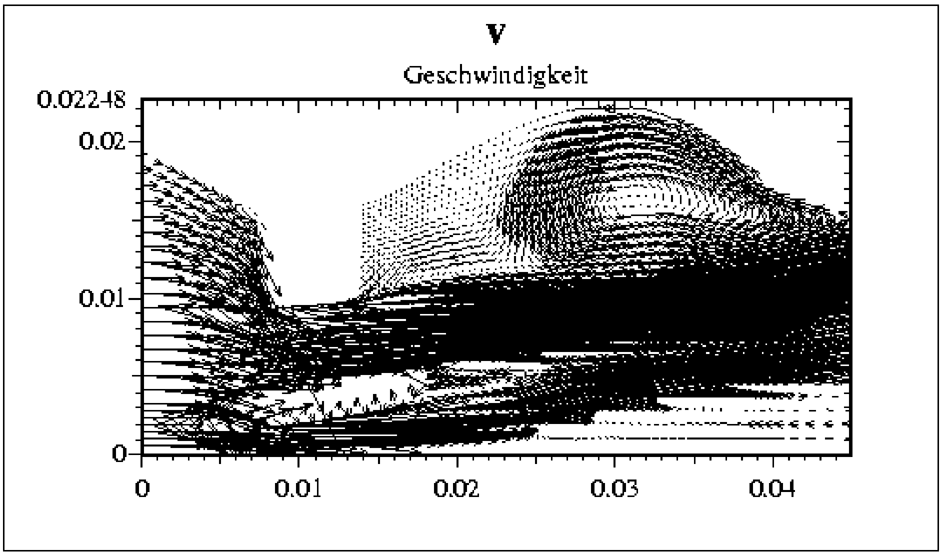


Abbildung 24: Geschwindigkeit v auf Superelemente-Geometrie bei Öffnungswinkel 80°

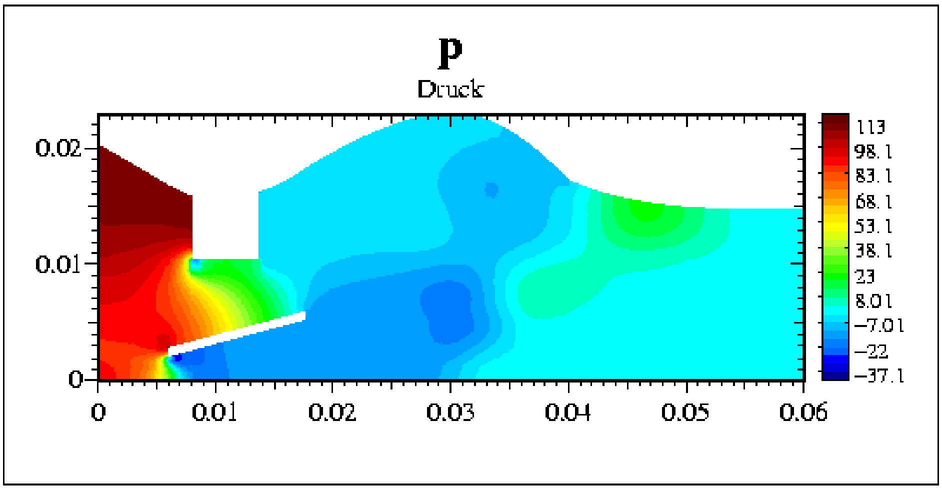


Abbildung 25: Druck p auf Superelemente-Geometrie bei Öffnungswinkel 75°

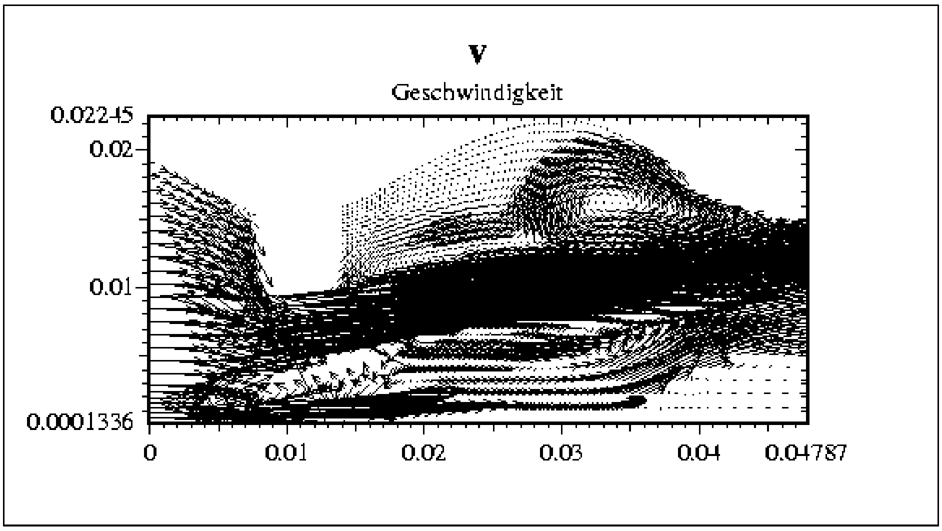


Abbildung 26: Geschwindigkeit v auf Superelemente-Geometrie bei Öffnungswinkel 75°

von ca. 85° erreicht, was der Konstruktion einer St. Jude-Herzklappenprothese entspricht. Ein maximaler Öffnungswinkel von mehr als 85° kann das reibungslose Schließen der Klappe gefährden, während ein maximaler Öffnungswinkel von weniger als 85° zu starke Turbulenzen verursacht, die sowohl dem Klappenmaterial als auch dem umliegenden Gewebe schaden, wie schon in Kapitel 2.2 beschrieben.

Zusammenfassung

Der Überblick über die biologischen, physikalischen und mathematischen Gegebenheiten zeigt, wie viele Möglichkeiten und ungeklärte Fragen man auf dem Gebiet der numerischen Simulationen noch antreffen kann. Modelliert man den Fluß einer viskosen Flüssigkeit wie den des menschlichen Blutes, so stößt man unweigerlich auf die Navier-Stokes Gleichungen, die mathematisch einige Schwierigkeiten beinhalten. Da es nicht möglich ist, eine analytische Lösung zu berechnen, ist man auf numerische Verfahren wie die Finite-Elemente-Methode zusammen mit dem Galerkin-Verfahren angewiesen. Das Hauptproblem der Navier-Stokes Gleichungen ist also ein Beweis der Existenz und Eindeutigkeit einer Lösung in mindestens drei Dimensionen, um den reell betrachteten Problemstellungen gerecht zu werden.

An die Klärung der mathematischen Hintergründe kann sich dann die Implementierung eines geeigneten Finite-Elemente-Lösers anschließen, wie es zum Beispiel mittels der C++-Bibliothek "Diffpack 3.0" möglich ist. In dieser Arbeit ist es aus Zeitgründen nicht vollständig gelungen, den vorhandenen Löser zu optimieren, d. h. es treten je nach Wahl der möglichen Parameter Konvergenzprobleme bei den iterativen Lösern des nichtlinearen globalen Gleichungssystemes, welches aus der Methode der Finite-Elemente resultiert, auf. Ein entscheidender Parameter hierbei ist die Reynoldszahl, die den Effekt von Viskosität auf das Flußverhalten beschreibt. Zur Behebung dieses Problemes könnte man für die Reynoldszahl einen "Continuation"-Parameter (siehe [26], S. 351) in die Gleichungen einfügen. Das bedeutet, man berechnet zunächst eine Lösung bezüglich einer sehr kleinen, physiologisch nicht angemessenen Reynoldszahl, um dieses Lösung als Startvektor für den nichtlinearen Löser in einer weiteren iterativen Schleife zu benutzen. Im Laufe der Iterationsschritte wird die Reynoldszahl dann durch den "Continuation"-Parameter auf die gewünschte Größe angehoben.

Trotz der beschriebenen Schwierigkeiten und starken Vereinfachungen, läßt sich der Blutfluß durch eine zweiflügelige, mechanische Herzklappe mit dem angegebenen Simulationsprogramm berechnen und skizzieren. Die Simulation des Flußverhaltens bzw. die Betrachtung von Geschwindigkeit und Druck ermöglicht einen detaillierten Einblick in die Strömungsverhältnisse, der Aufschluß über Schwachstellen in der Konstruktion einer solchen Klappenprothese geben kann. Schon kleine Änderungen der Maße von Klappenring, Klappenflügel und besonders des maximalen Öffnungswinkels führen zu einem anderen Flußbild, welches andere Auswirkungen haben kann. Dazu zählen Nebeneffekte wie z. B. Wucherungen der Aortenwand durch zu starken Aufprall des Blutstromes oder Erosionerscheinungen an den Klappenflügeln durch zu hohe Druckunterschiede. Mit der Hilfe eines Simulationsprogrammes können einzelne Parameter verändert und die Folgen direkt beobachtet werden.

Ein maximaler Öffnungswinkel von ca. 85° scheint im betrachteten Fall optimal zu sein.

Zur genauen Lokalisierung von Wirbeln und Turbulenzen ist allerdings ein 3-dimensionales, zeitabhängiges Modell nötig, das besser auf die sich ändernden Randbedingungen des Herzzyklus eingehen kann. Ergebnisse eines solchen Simulations-Programmes sind z. B. in der Arbeit von Mary J. King (s. [23]) zu finden. Der nächste Schritt ist nun ein Modell, welches den Blutfluß und die Bewegung der Klappenflügel als gekoppelte Probleme betrachtet, so daß man von tatsächlicher Interaktion zwischen Fluid, starrem Prothesen-Material und elastischer Außenwand sprechen kann. Doch selbst dann ist dieses Thema in naher Zukunft nicht ausgeschöpft, da die exakte Modellierung des Herzzyklus und des umliegenden Gewebes noch in weiter Ferne liegen.

A Medizinische Fachbegriffe

Anaemie sog. Blutarmut; Verminderung des Anteils der roten Blutkörperchen im Vollblut unter den Normalwert

Aorta die von der linken Herzhälfte abgehende Hauptschlagader als Stammgefäß des Körperkreislaufes, des sogenannten „großen“ Blutkreislaufes; mit elastisch (durch entsprechenden Wandaufbau) bedingter Windkesselfunktion

Atrium Herzvorhof

Bikuspidalität die - seltene - Fehlbildung der Aorten- oder Pulmonalklappe als nur aus 2 Klappenflügeln bestehende Gebilde

distal anatomische Richtungsweisung; weiter entfernt von der Körpermitte, Herzmitte o.ä.; Gegensatz: proximal

Embolie plötzlicher Verschluß eines Blutgefäßes (meist Arterie) durch von der Blutbahn verschlepptes Gewebe, Fremdkörper oder Luft

Endokard die alle Hohlräume (einschließlich der Herzklappen) auskleidende glatte Innenhaut des Herzens, schichtweise aufgebaut - von innen zum Herzmuskel hin - aus Endothel, feinfaserigem kollagenem Bindegewebe, elastischen Fasern und einzelnen glatten Muskelzellen

Endokarditis Entzündung der Herzinnenhaut; meist als Entzündung der Herzklappen, genauer am Schließungsrand einer Klappe, aber auch im Bereich der Vorhof- und Kammerwände, Sehnenfäden und Papillarmuskeln

Epikard das mit der äußeren Oberfläche des Herzmuskels verwachsene „innere Blatt“ des Herzbeutels als äußerste Schicht der Herzwand; erstreckt sich auch über die herznahen Teile der großen Gefäße

Haemolyse die Auflösung (Zerstörung) der roten Blutkörperchen, z.B. in Folge einer zu hohen mechanischen Beanspruchung an Herzklappenprothesen

Herzinsuffiziens akutes oder chronisches Unvermögen des Herzens, bei Belastung (=Belastungsinsuffiziens) oder schon in Ruhe (=Ruheinsuffiziens) den für den Stoffwechsel erforderlichen Blutauswurf aufzubringen

Herzminutenvolumen HMV, das vom Herzen pro Minute ausgeworfene Blutvolumen

Hypertrophie Größenzunahme eines Gewebes oder Organs nur durch Zellvergrößerung bzw. -wucherung bei normal bleibender Zellzahl

intrakardial in einer oder in eine Herzhöhle

Myocard der Herzmuskel; die mittlere, zwischen Endo- und Epikard gelegene Herzwandschicht

Perikard Herzbeutel: Schutz- und Gleithülle des Herzens

Rheumatisches Fieber spezifische Entzündungsreaktion auf Toxine der Streptokokken

serös das Blutserum betreffend; hier: serumhaltig, z.B. seröse Höhlen, wie die Bauchhöhle

Sinus latein.: Bucht, Tasche; taschenartige Körperhöhle oder Organausbuchtung; z.B. *Sinus aortae* die nahezu intrakardiale Ausweitung zwischen jedem der drei Aortenklappenflügeln und der Aortenwand

Stenose angeborene oder erworbene dauerhafte Einengung eines Kanals, z.B. Herzklappenstenose

Stenosegeräusch Herzgeräusche (evt. mit tastbarem Schwirren) bei Herzklappenstenosen

Thorax Brustkorb

Thromboembolie Embolie durch Verschleppung von Gerinselteilen, z.B. bei Herklappenerkrankungen

Thrombose die intravitale „Blutpfropfbildung" im Kreilaufsystem durch Aggregation von Thrombozyten

Thrombozyten Blutplättchen; vom Knochenmark abstammende, kernlose Blutelemente

Truncus pulmonalis von der rechten Herzhälfte abgehendes Stammgefäß des Lungenkreislaufes, des sogenannten „kleinen" Blutkreislaufes

Ventrikel Herzkammer

B Programm-Code und Input-Files

B.1 NsPenalty1.h

```
#ifndef NsPenalty1_h_IS_INCLUDED
#define NsPenalty1_h_IS_INCLUDED

#include <FEM.h>                // finite element toolbox
#include <DegFreeFE.h>          // field dof <-> linear system dof
#include <LinEqAdmFE.h>         // linear solver toolbox
#include <TimePrm.h>            // time discretization parameters
#include <NonLinEqSolver.h>     // nonlinear solver interface
#include <NonLinEqSolver_prm.h> // parameters for nonlinear solvers
#include <NonLinEqSolverUDC.h>  // user's definition of nonlinear PDEs
#include <SimCase.h>            // user's definition of menu items
#include <SaveSimRes.h>         // used in storing and plotting of
                                // results


class NsPenalty1: public FEM, public NonLinEqSolverUDC
{
public:
  Handle(GridFE)          grid;    // Finite-Elemente-Gitter
  Handle(DegFreeFE)       dof;     // Zeiger: matrix dof <-> u dof
  Handle(FieldsFE)        u;       // Geschwindigkeitsfeld
  Handle(FieldsFE)        u_prev;  // u auf vorherigem Zeitlevel
  Handle(FieldFE)         p;       // Druckfeld
  Handle(TimePrm)         tip;     // Parameter zur zeitlichen
                                   // Diskretisierung

  int  inlet_profile;     // Flußprofil bei x=0
  real theta;             // "theta"-Regel zur Zeit-Diskretisierung
  real mu;                // Viskosität
  real density;           // Dichte
  real inlet_velocity;    // Flußgeschwindigkeit bei x=0
  real lambda;            // Parameter für Penalty-Methode

  Vec(real)                   nonlin_solution; // Lösung des nichtlinearen
                                               // Systemes
  Vec(real)                   linear_solution; // Lösung des linearen
```

```
                                                // Unter-Systemes
  Handle(NonLinEqSolver_prm) nlsolver_prm;    // Parameter für nlsolver
  Handle(NonLinEqSolver)     nlsolver;        // nichtlineare Löser
  Handle(LinEqAdmFE)         lineq;           // lineare Löser
  Handle(SaveSimRes)         database;        // berechnete Daten

  // Zur Vermeidung von wiederholter Einlesung in die
  // Funktion integrands:
  Ptv(real) u_pt;                  // u am aktuellen Integrationspunkt
  Ptv(real) up_pt;                 // u_prev am aktuellen
                                   // Integrationspunkt
  VecSimple(Ptv(real)) gradu_pt;   // grad u am aktuellen
                                   // Integrationspkt.
  VecSimple(Ptv(real)) gradup_pt;  // grad u_prev am aktuellen
                                   // Integrationspkt.

  NsPenalty1 () {};
 ~NsPenalty1 () {};

  virtual void define (MenuSystem& menu, int level = MAIN);
  virtual void scan   ();
  virtual void adm    (MenuSystem& menu);

  virtual void solveProblem ();
  virtual void resultReport ();
  virtual void saveResults ();
  virtual void calcDerivedQuantities ();
  virtual void fillEssBC ();

protected:
  virtual void inletVelocity (Ptv(real)& v, const Ptv(real)& x);
  virtual void setIC ();
  virtual void timeLoop ();
  virtual void solveAtThisTimeStep ();
  virtual void makeAndSolveLinearSystem ();

  enum Integrand_type { LAMBDA_TERMS, ORDINARY_TERMS };
  Integrand_type integrands_tp;

  virtual void calcElmMatVec
    (int elm_no, ElmMatVec& elmat, FiniteElement& fe);
```

```
  virtual void integrands
          (ElmMatVec& elmat, const FiniteElement& fe);
  virtual void integrandsReduced
          (ElmMatVec& elmat, const FiniteElement& fe);
  virtual void integrandsNonReduced
          (ElmMatVec& elmat, const FiniteElement& fe);
  friend class PressureIntg;
};


class PressureIntg: public IntegrandCalc
{
  NsPenalty1* data; // Zugang zu Input-Daten und Ergebnissen
public:
  PressureIntg (NsPenalty1* data_ ) { data = data_; }
  virtual void integrands(ElmMatVec& elmat,const FiniteElement& fe);
};
#endif
```

B.2 NsPenalty.cpp

```
#include <NsPenalty1.h>
#include <ElmMatVec.h>
#include <FiniteElement.h>
#include <readOrMakeGrid.h>
#include <DegFreeFE.h>
#include <VecSimple_Ptv_real.h>
#include <MatDiag_real.h>
#include <Puttonen.h>

void NsPenalty1:: define (MenuSystem& menu, int level)
{
  menu.addItem (level, "casename", " ", "test1");
  menu.addItem (level, "gridfile",
                "filename", "gitter/test1.grid");
  menu.addItem (level, "redefine boundary indicators",
                "GridFE::redefineBoInds syntax",
                "nb=2 names D1 d2 1=(1 3) 2=(2 4)");
  menu.addItem (level, "add boundary nodes",
                "GridFE::addBoIndNodes syntax",
```

```
                "n=1 b1=[0,0]x[0,0]");
  menu.addItem (level, "time integration parameters",
                "TimePrm::scan syntax", "dt=1");
  menu.addItem (level, "inlet profile", " ", "1");
  menu.addItem (level, "characteristic inlet velocity",
                "used in inlet profile expressions", "25.0e-2");
  menu.addItem (level, "viscosity", "mu", "4.0e-3");
  menu.addItem (level, "density", "rho", "1000");
  menu.addItem (level, "penalty parameter", "lambda", "10000");
  menu.addItem (level, "bandwidth reduction", "for GaussElim", "ON");

  LinEqAdmFE        ::defineStatic (menu, level+1);
  NonLinEqSolver_prm::defineStatic (menu, level+1);
  SaveSimRes        ::defineStatic (menu, level+1);
  FEM               ::defineStatic (menu, level+1);
}

void NsPenalty1:: scan ()
{
  MenuSystem& menu = *SimCase::menu_system;
  String gridfile = menu.get ("gridfile");
  grid.rebind (new GridFE ());
  readOrMakeGrid (*grid, gridfile);
                    // Gittergenerierung

  String redef = menu.get ("redefine boundary indicators");
   if (!redef.contains ("NONE"))
    grid->redefineBoInds (redef);
                   // neue Randwertdefinierung

  String add = menu.get ("add boundary nodes");
   if (!add.contains ("NONE"))
    grid->addBoIndNodes (add);
                  // zusätzliche Randpunkte

  tip.rebind (new TimePrm());
  tip->scan (menu.get ("time integration parameters"));
  mu = menu.get ("viscosity").getReal();
  lambda = menu.get ("penalty parameter").getReal();
  density = menu.get ("density").getReal();
```

```
  const nsd = grid->getNoSpaceDim();
  database.rebind(new SaveSimRes);
  database->scan (menu, nsd);
  inlet_profile  = menu.get ("inlet profile").getInt();
  inlet_velocity = menu.get ("characteristic inlet
                              velocity").getReal();
  bool bandwidth_reduction = menu.get ("bandwidth
                                      reduction").getBool();
  if (bandwidth_reduction) {
    s_o << "\nReducing the bandwidth by
             using class Puttonen.\n";
    Puttonen bwr;  bwr.renumberNodes (*grid);
   }

  lineq.rebind(new LinEqAdmFE);
  lineq->scan (menu);


    // Initialisierung der Variablen
  p.rebind (new FieldFE (*grid,"p"));
  u.rebind (new FieldsFE (*grid,"v"));
  u_prev.rebind (new FieldsFE (*grid, "v_prev"));
  dof.rebind (new DegFreeFE (*grid, nsd));
  const int lineq_vec_length = u->getNoValues();
  nonlin_solution.redim (lineq_vec_length);
  linear_solution.redim (lineq_vec_length);
  lineq->attach (linear_solution);
  nlsolver_prm.rebind (NonLinEqSolver_prm::construct());
  nlsolver_prm->scan(menu);
  nlsolver.rebind (nlsolver_prm->create());
  nlsolver->attachUserCode (*this);
  nlsolver->attachNonLinSol (nonlin_solution);
  nlsolver->attachLinSol (linear_solution);

  u_pt.    redim (nsd);  up_pt.    redim (nsd);
  gradu_pt.redim (nsd);  gradup_pt.redim (nsd);
}


void NsPenalty1:: adm (MenuSystem& menu)
{
```

```
  SimCase::attach(menu);
  define (menu); menu.prompt (); scan ();
}


void NsPenalty1:: fillEssBC ()
{
/*   Randindikatoren:
   bo-ind 1: inlet-Geschwindigkeitsfeld
   bo-ind 2: outlet-Geschwindigkeitsfeld mit
             Normalenableitungen = 0
   bo-ind 3: u=0
   bo-ind 4: v=0
   bo-ind 5: w=0
*/

  dof->initEssBC ();
  const int nno = grid->getNoNodes();
  const int nsd = grid->getNoSpaceDim();
  Ptv(real) v (nsd);
  Ptv(real) x (nsd);
  int i, k;

  for (i = 1; i <= nno; i++) {
    if (grid->boNode (i,1)) {
      // Aufruf des inlet-Profiles
      grid->getCoor (x,i);
      inletVelocity (v, x);
      for (k=1; k <= nsd; k++)
        // k-te Komponente des i-ten
        // Knotenpunktes bekommt Wert v(k):
        dof->fillEssBC (i,k, v(k));
    }
    for (k=1; k <= nsd; k++) {
      if (grid->boNode (i, 2+k))
        dof->fillEssBC (i,k, 0.0);
    }
  }
}

void NsPenalty1:: solveProblem ()
```

```
{
  s_o->setRealFormat ("%13.6e");
  timeLoop ();
}

void NsPenalty1:: setIC ()
{
  const int nno = grid->getNoNodes();
  const int nsd = grid->getNoSpaceDim();
  int i,k;
  for (i=1; i <= nno; i++){
    for (k=1; k <= nsd; k++) {
       u_prev()(k).values()(i) = 0.0;
       u()(k).values()(i) = 0.0;
     }
  }
}

void NsPenalty1:: timeLoop ()
{
  tip->initTimeLoop();
  setIC ();
  while (!tip->finished())
    {
      tip->increaseTime();

      solveAtThisTimeStep ();
      calcDerivedQuantities ();
      saveResults ();

      *u_prev = *u;
    }
}



void NsPenalty1:: inletVelocity (Ptv(real)& v, const Ptv(real)& x)
{
  if (inlet_profile == 1)
    {
     // parabolisches Inlet-Profil in
```

```
      // x=0 mit Max. bei y=0 und Min.  bei y=0.02
       v = 0.0;
       v(1) = inlet_velocity * (1 - sqr(x(2)/0.02));
     }
 else
     errorFP("NsPenalty1::inletVelocity",
             "inlet_velocity=%d not impl.",
             inlet_velocity);
}

void NsPenalty1:: saveResults ()
{
  database->dump (*p, tip.getPtr());
  database->dump (*u, tip.getPtr());
}

void NsPenalty1:: resultReport ()
{
  // leere Funktion
}

void PressureIntg:: integrands
  (ElmMatVec& elmat, const FiniteElement& fe)
{
  const int nsd = fe.getNoSpaceDim();
  real div_v=0;
  for (int k = 1; k <= nsd; k++) {
    data->u()(k).derivativeFEM (data->gradu_pt(k), fe);
    div_v += data->gradu_pt(k)(k);
   }
  const real pressure = - data->lambda*div_v;

  const int nbf = fe.getNoBasisFunc();
  const real detJxW = fe.detJxW();
  for (int i = 1; i <= nbf; i++)
    elmat.b(i) += pressure*fe.N(i)*detJxW;
}

void NsPenalty1:: calcDerivedQuantities ()
{
  PressureIntg penalty_integrand (this);
```

```
  FEM::smoothField (*p, penalty_integrand);
}

void NsPenalty1::solveAtThisTimeStep ()
{
  fillEssBC ();

  // initialisiert die nichtlineare Lösung mit der Lösung des
  // vorherigen Zeitschrittes:
  dof->field2vec (*u_prev, nonlin_solution);
  // Aktualisierung der Randbedingungen:
  dof->insertEssBC (nonlin_solution);


#ifdef DP_DEBUG
nonlin_solution.print(s_o,"nonlin_solution before
                           nonlinear iteration");
#endif

  if (!nlsolver->solve ())
    errorFP("NsPenalty1::SolveAtThisTimeStep",
            "The nlsolver.solve call: divergence of solver \"%s\"",
            nlsolver_prm->method.c_str());
  // In jedem Iterationsschritt der Schleife  nlsolver->solve,
  // wird makeAndSolveLinearSystem aufgerufen!

  dof->vec2field (nonlin_solution, *u);
                      // weist Lösung u zu

#ifdef DP_DEBUG
  for (int k = 1; k <= grid->getNoSpaceDim (); k++)
    u()(k).values().print(s_o,oform("u(%d) after
                              nonlinear it.",k));
#endif
}

void NsPenalty1:: makeAndSolveLinearSystem ()
{
  dof->vec2field (nonlin_solution, *u);

  if (nlsolver->getCurrentState().method
```

```
                                  == NEWTON_RAPHSON)
    dof->fillEssBC2zero();
  else
    dof->unfillEssBC2zero();

  makeSystem (*dof, *lineq);
  lineq->solve();
#ifdef DP_DEBUG
  linear_solution.print(s_o,"linear_solution");
#endif
}

void NsPenalty1:: calcElmMatVec
  (int e, ElmMatVec& elmat, FiniteElement& fe)
{
  // itg_rules is inherited from base class FEM
  itg_rules.setRelativeOrder (0);
  fe.refill (e, itg_rules);
  integrands_tp = ORDINARY_TERMS;
  numItgOverElm (elmat, fe);
                 // Integration der lambda-freien Terme

  itg_rules.setRelativeOrder (-1);
  fe.refill (e, itg_rules);
  integrands_tp = LAMBDA_TERMS;
  numItgOverElm (elmat, fe);
                // "reduzierte" Integration der lambda-Terme
}

void NsPenalty1:: integrands
  (ElmMatVec& elmat, const FiniteElement& fe)
{
  if (integrands_tp == LAMBDA_TERMS)
    integrandsReduced (elmat, fe);
  else if (integrands_tp == ORDINARY_TERMS)
    integrandsNonReduced (elmat, fe);
}

void NsPenalty1::integrandsNonReduced
  (ElmMatVec& elmat, const FiniteElement& fe)
{
```

```
 int i,j;    // Indizes der Basisfunktionen N(i)
 int k,r,s;  // 1,..,nsd Indizes für
             // räumliche Dimensionen
 int ig,jg;  // Zuordnungsindex für Elementmatrizen
             // und -vektoren,
             // basierend auf i,j,r,s
 int e, jx;
 real nabla2,h1,h2,h3,h4;

 const int nsd = fe.getNoSpaceDim();
 const int nbf = fe.getNoBasisFunc();
 const real detJxW = fe.detJxW();

 for (k = 1; k <= nsd; k++) {
   gradu_pt(k).redim (nsd);  gradup_pt(k).redim (nsd);
   u_pt(k) = u()(k).valueFEM (fe);
   u()(k).derivativeFEM (gradu_pt(k), fe);
}

 real dirchl,con,dij,cij,sum;

 e = fe.getElmNo();
 if (nlsolver->getCurrentState().method
                                 == NEWTON_RAPHSON)
   {
     for (i = 1; i <= nbf; i++) {
       for (j = 1; j <= nbf; j++) {
         jx = grid->loc2glob(e,j);

         nabla2 = 0;
         for (k = 1; k <= nsd; k++)
           nabla2 += fe.dN(i,k)*fe.dN(j,k);

         dirchl = mu*nabla2;

         sum = 0;
         for (k = 1; k <= nsd; k++)
           sum += fe.dN(j,k)*u_pt(k);
         con = density*fe.N(i)*sum;

         dij = dirchl + con;
```

```
        cij = density*fe.N(i)*fe.N(j);

        for (r = 1; r <= nsd; r++)
          for (s = 1; s <= nsd; s++) {
            ig = nsd*(i-1)+r;
            jg = nsd*(j-1)+s;

            h1 = cij*gradup_pt(r)(s);
            elmat.A(ig,jg) += h1*detJxW;

            if (r == s) {
              h2 = dij;
              elmat.A(ig,jg) += h2*detJxW;

              real sli = u()(r).values()(jx);
              h3 = dij*sli;
              elmat.b(ig) += -h3*detJxW;
            }
          }
      }
    }
  }

else if (nlsolver->getCurrentState().method
                                      == SUCCESSIVE_SUBST)
  {
    for (i = 1; i <= nbf; i++) {
      for (j = 1; j <= nbf; j++) {
        jx = grid->loc2glob(e,j);

        nabla2 = 0;
        for (k = 1; k <= nsd; k++)
          nabla2 += fe.dN(i,k)*fe.dN(j,k);

        dirchl = mu*nabla2;

        sum = 0;
        for (k = 1; k <= nsd; k++)
          sum += fe.dN(j,k)*u_pt(k);
        con = density*fe.N(i)*sum;
```

```
          dij = dirchl + con;
          cij = density*fe.N(i)*fe.N(j);


          for (r = 1; r <= nsd; r++)
            for (s = 1; s <= nsd; s++) {
              ig = nsd*(i-1)+r;
              jg = nsd*(j-1)+s;

              if (r == s) {
                h1 = dij;
                elmat.A(ig,jg) += h1*detJxW;
                h2 = 0;
                elmat.b(ig) += 0;
              }
            }
        }
      }
    }

  else
    fatalerrorFP("NSPenalty1::integrands",
                 "current nonlinear method=%d, illegal value",
                 nlsolver->getCurrentState().method);
}


void NsPenalty1:: integrandsReduced
  (ElmMatVec& elmat, const FiniteElement& fe)
{
  int i,j;      // Indizes der Basisfunktionen N(i)
  int k,r,s;    // 1,..,nsd Indizes für
                // räumliche Dimensionen
  int ig,jg;    // Zuordnungsindex für Elementmatrizen
                // und -vektoren,
                // basierend auf i,j,r,s

  real h1,h2;

  const int nsd = fe.getNoSpaceDim();
  const int nbf = fe.getNoBasisFunc();
```

```
const real detJxW = fe.detJxW();


for (k = 1; k <= nsd; k++) {
  gradu_pt(k).redim (nsd);   gradup_pt(k).redim (nsd);
  u_pt(k) = u()(k).valueFEM (fe);
  u()(k).derivativeFEM (gradu_pt(k), fe);

}
real sum, div;

if (nlsolver->getCurrentState().method
                                == NEWTON_RAPHSON)
  {
    for (i = 1; i <= nbf; i++)
      for (j = 1; j <= nbf; j++)
        for (r = 1; r <= nsd; r++)
          for (s = 1; s <= nsd; s++) {
            ig = nsd*(i-1)+r;
            jg = nsd*(j-1)+s;

            h1 = lambda*fe.dN(i,r)*fe.dN(j,s);
            elmat.A(ig,jg) += h1*detJxW;
          }
    sum = 0;
    for (k = 1; k <= nsd; k++)
      sum += gradu_pt(k)(k);

    div = lambda*sum;

    for (i = 1; i <= nbf; i++)
      for (r = 1; r <= nsd; r++) {
        ig = nsd*(i-1)+r;
        h2 = div*fe.dN(i,r);
        elmat.b(ig) += - h2*detJxW;
      }
  }

else if (nlsolver->getCurrentState().method
                                     == SUCCESSIVE_SUBST)
 {
```

```
    for (i = 1; i <= nbf; i++)
      for (j = 1; j <= nbf; j++)
        for (r = 1; r <= nsd; r++)
          for (s = 1; s <= nsd; s++) {
            ig = nsd*(i-1)+r;
            jg = nsd*(j-1)+s;

             h1 = lambda*fe.dN(i,r)*fe.dN(j,s);
             elmat.A(ig,jg) += h1*detJxW;
             h2 =0;
             elmat.b(ig) += 0;
          }
    }

  else
    fatalerrorFP("NSPenalty1::integrands",
                 "current nonlinear method=%d, illegal value",
                 nlsolver->getCurrentState().method);
}
```

B.3 `main.cpp`

```
#include <NsPenalty1.h>
main (int argc, const char* argv[])
{
  initDIFFPACK (argc, argv);
  global_menu.init ("Navier-Stokes", "NsPenalty1");
  NsPenalty1 simulator;
  global_menu.multipleLoop (simulator);

}
```

B.4 Generierung der Super-Elemente-Gitter

Zu Generierung eines Super-Elemente-Gitters benötigt man ein `.geom`-File zu Festlegung der Koordinaten jedes einzelnen Super-Elementes und ein `.part`-File zur Eingabe von Elementtyp, Feinheit der Unterteilung (`divisions`) und eventuellem

Dehnungsfaktor (`grading`). Als Beispiel werden im folgenden die entsprechenden Files für den Fall mit 75° Öffnungswinkel vorgestellt.

Die Eingaben in `supels75.geom` haben die folgende Gestalt:

```
>no_of_dimensions=2; >subdomains=1; >no_of_supels=22;
>no_of_ind=5; >name links rechts oben unten klappe;

!Element 1

>SupEl; >subdomain_no=1; >elementtype = ElmB8n2D;
>boundary = [1(3)][3(2)];
>nodes = [1(0 0.0102)]+[2(0.006 0.0102)]+[3(0 0.02)]
         +[4(0.006 0.0165)]+[6(0.0035 0.018)];
>sides = ;

!Element 2

>SupEl; >subdomain_no=1; >elementtype = ElmB8n2D;
>boundary = [1(3)];
>nodes = [1(0 0.00281)]+[2(0.006 0.00281)]+[3(0 0.0102)]
         +[4(0.006 0.0102)];
>sides = [2(1 4)];

!Element 3

>SupEl; >subdomain_no=1; >elementtype = ElmB8n2D;
>boundary = [5(1)][1(3)];
>nodes = [1(0 0.00194)]+[2(0.006 0.00194)]+[3(0 0.00281)]
         +[4(0.006 0.00281)];
>sides = [2(2 4)];
```

usw. bis zu Super-Element Nr. 22. Das zugehörige `supels75.part`-File beginnt wie folgt:

```
>nsd=2; >no_of_supels=22;

!Element 1

>SupEl; >nsd=2;
>elementtype = ElmB4n2D; >divisions = [10,10]; >grading = [1,1];
```

```
!Element 2

>SupEl; >nsd=2;
>elementtype = ElmB4n2D; >divisions = [10,8]; >grading = [1,1];

!Element 3

>SupEl; >nsd=2;
>elementtype = ElmB4n2D; >divisions = [10,2]; >grading = [1,1];
```

Der Befehl

```
makegrid +iscl -c supels75 -m PreproSupElSet
               -g FILE=supels75.geom -p FILE=supels75.part -d
```

konstruiert dann das gesuchte Gitter. (`supels75.grid`)

B.5 Input-File `supels75.i` für Super-Elemente-Gitter

```
set gridfile = Verify/supels75.grid
set redefine boundary indicators = nb=4 names=inlet u0 v0 w0
                                   1=(1) 2=(3 5) 3=(3 4 5) 4=(3 4 5)
set add boundary nodes = NONE
set viscosity = 4.0e-3
set penalty parameter = 800
set density = 1000
set inlet profile = 1
set characteristic inlet velocity = 20.0e-2
! 30 würde Re=1500 liefern....
set time integration parameters = dt=0
!
sub NonLinEqSolver_prm
set nonlinear iteration method = SuccessiveSubst
set max nonlinear iterations = 100
set max estimated nonlinear error = 3.0e-3
set nonlinear relaxation prm = 0.7
ok
sub LinEqAdmFE
```

```
sub Matrix_prm
set matrix type = MatBand
set symmetric storage = false
ok
sub LinEqSolver_prm
set basic method = GaussElim
ok
ok
sub SaveSimRes
set field storage format = BINARY
set grid storage format = BINARY
ok
ok
```

B.6 Input-File king.i für BOX_WITH_BELL-Geometrie

```
set gridfile = Verify/supels75.grid
set redefine boundary indicators = nb=4 names=inlet u0 v0 w0
                                   1=(1) 2=(3 5) 3=(3 4 5) 4=(3 4 5)
set viscosity = 4.0e-3
set penalty parameter = 800
set density = 1000
set inlet profile = 1
set characteristic inlet velocity = 20.0e-2 ! 30 gives Re=1500
set time integration parameters = dt=0
!
sub NonLinEqSolver_prm
set nonlinear iteration method = SuccessiveSubst
set max nonlinear iterations = 100
set max estimated nonlinear error = 3.0e-3
set nonlinear relaxation prm = 0.7
ok
sub LinEqAdmFE
sub Matrix_prm
set matrix type = MatBand
set symmetric storage = false
ok
sub LinEqSolver_prm
set basic method = GaussElim
ok
```

```
ok
sub SaveSimRes
set field storage format = BINARY
set grid storage format = BINARY
ok
ok
```

Literatur

[1] J. Aagaard, C.N. Hansen, J. Tingleff, and I. Rygg. Seven-and-a-half Years Clinical Experience with the Carbomedics Prosthetic Heart Valve. *J. Heart Valve Dis.*, 4:628–633, 1995.

[2] B.J. Bellhouse and F.H. Bellhouse. Fluid mechanics of model normal and stenosed artic valves. *Circ Res*, 25:693–704, 1969.

[3] B.J. Bellhouse and F.H. Bellhouse. Fluid mechanics of model mitral valve and left ventrikel. *Cardiovasc Res*, 6:199–210, 1972.

[4] B.J. Bellhouse and L. Talbot. *J. Fluid Mech.*, 35:721, 1969.

[5] A. M. Bruaset. *A survey of preconditioned iterative methods.* Longman Scientific & Technical, 1. edition, 1995.

[6] Sulzer Carbomedics. Infomaterial, 1998.

[7] A.J. Chorin and J.E. Marsden. *A Mathematical Introduction to Fluid Mechanics.* Springer Verlag, 1979.

[8] L. Collatz. *Differentialgleichungen.* Teubner Verlag, 6. edition, 1981.

[9] C.W. Cryer. *Numerik Partieller Differentialgleichungen 1.* WWU Münster, 1998.

[10] C.W. Cryer. *Numerik Partieller Differentialgleichungen 2.* WWU Münster, 1998.

[11] C.W. Cryer and P.P. Lunkenheimer. *Mathematical Modelling of the Cardiovascular System.* WWU Münster, 1995.

[12] Czhiak, Langer, and Ziegler. *Biologie.* Springer Verlag, 5. edition.

[13] P. Deuring and W. von Wahl. Strong Solutions of the Navier-Stokes System in Lipschitz Bounded Domains. *Math. Nachr.*, 171:111–148, 1995.

[14] O. Forster. *Analysis 2.* Vieweg Studium, 5. edition, 1993.

[15] Y.C. Fung. *Biomechanics, Circulation.* Springer Verlag, 2. edition.

[16] A. Greenbaum. *Iterative Methods for Solving Linear Systems.* SIAM, 1. edition, 1997.

[17] Ch. Großmann and H.-G. Roos. *Numerik partieller Differentialgleichungen.* Teubner Verlag, 1. edition, 1992.

[18] Y. Henderson and F. Johnson. Two models of closure of heart valves. *Heart*, 4:69–82, 1912.

[19] D. Horstkotte and F. Loogen. Update in Heart Valve Replacement. *Proceedings of the Second European Symposium on the St.Jude Medical Heart Valve*, 1986.

[20] J. Jost. *Partielle Differentialgleichungen.* Springer Verlag, 1. edition, 1998.

[21] R. Juchems. *Herz- und Kreislaufkrankheiten.* Wiss. Buchges. Darmstadt.

[22] T. Kato. Strong L^p-Solutions of the Navier-Stokes Equation in $\mathbb{R}^m$, with Applications to Weak Solutions. *Math. Z.*, 192:135–148, 1986.

[23] Mary J. King. *Computational and experimental studies of flow through a bileaflet mechanical heart valve.* PhD thesis, University of Leeds, U.K., 1994.

[24] H. Kozono and T. Ogawa. Some L^p Estimate for the Exterior Stokes Flow and An Application to the Non-Stationary Navier-Stokes Equations. *Indiana Univ. Math. J.*, 41:789–808, 1992.

[25] W. Krahwinkel, S. Moltzahn, and M. Zeydabadinejad. *Echokardiographie der künstlichen Herzklappen.* Thieme Verlag, 1995.

[26] H. P. Langtangen. *Computational partial differential equations: Numerical methods and Diffpack Programming.* Springer Verlag, 1. edition, 1999.

[27] H. P. Langtangen. Finite element prepocessors in Diffpack. *Numerical Objects Series, Numerical Objects A.S.*, 1999.

[28] C.S.F. Lee and L. Talbot. A fluid mechanical study on the closure of heart valves. *J Fluid Mech*, 91:41–63, 1979.

[29] H. Leonhardt. *Taschenatlas der Anatomie.* Thieme Verlag, 5. edition.

[30] C.S. Peskin. Numerical Analysis of Blood Flow in the Heart. *J. Computational Physics*, 25:220–252, 1977.

[31] W. Richter. *Partielle Differentialgleichungen.* Spektrum, 1. edition, 1995.

[32] H. R. Schwarz. *Methode der finiten Elemente.* Teubner Verlag, 1. edition, 1980.

[33] J. S. Schwegler. *Der Mensch – Anatomie und Physiologie.* Thieme Verlag, 2. edition.

[34] E. M. Stein. *Singular Integrals and Differentiability of Functions.* Princeton University Press, 1. edition, 1970.

[35] R. Temam. *Navier-Stokes Equations.* North Holland Publishing, 1. edition, 1977.

[36] R. Temam. *Navier-Stokes Equations and Nonlinear Functional Analysis.* SIAM, 1. edition, 1983.

[37] Prof. W. Walter. *Gewöhnliche Differentialgleichungen.* Springer Verlag, 5. edition, 1991.

[38] K. Wieghardt. *Theoretische Strömungslehre.* Teubner Verlag, 2. edition, 1974.

[39] M. Wiegner. The Navier-Stokes Equations - a Neverending Challenge? *Jber. d. Dt. Math.-Verein.*, 101:1–25, 1999.