

Stefan Kopecz

Ein gekoppeltes Finite-Elemente/
Discontinuous-Galerkin-Verfahren zur Simulation
von Strömungs-Transport-Problemen

Stefan Kopecz

Ein gekoppeltes Finite-Elemente /
Discontinuous-Galerkin-Verfahren zur Simulation
von Strömungs-Transport-Problemen

Die vorliegende Arbeit wurde vom Fachbereich Mathematik und Naturwissenschaften der Universität Kassel als Dissertation zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) angenommen.

Erster Gutachter: Prof. Dr. Andreas Meister
Zweiter Gutachter: Prof. Dr.-Ing. Olaf Wünsch

Tag der mündlichen Prüfung

10. Juli 2012

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar

Zugl.: Kassel, Univ., Diss. 2012
ISBN print: 978-3-86219-368-4
ISBN online: 978-3-86219-369-1
URN: <http://nbn-resolving.de/urn:nbn:de:0002-33698>

© 2012, kassel university press GmbH, Kassel
www.uni-kassel.de/upress

Printed in Germany

Inhaltsverzeichnis

Vorwort	5
1 Einleitung	7
2 Gleichungen	11
2.1 Die inkompressiblen Navier-Stokes Gleichungen	12
2.2 Verallgemeinert newtonsche und nicht-isotherme Gleichungen	12
2.3 Kavitationsmodell	15
2.3.1 Beschränktheit der Dampfdichte	19
2.4 Modellproblem	20
2.5 Poiseuille-Strömung	21
3 Strömungslöser	25
3.1 Inkompressible Navier-Stokes-Gleichungen	25
3.1.1 Schwache Formulierung	25
3.1.2 Ausflussrandbedingungen	27
3.1.3 Raumdiskretisierung	28
3.1.4 Zeitdiskretisierung	30
3.2 Modellproblem	31
3.2.1 Schwache Formulierung	31
3.2.2 Raumdiskretisierung	32
3.2.3 Zeitdiskretisierung	32
3.2.4 Ausflussrandbedingungen	33
4 Diskrete Projektionsverfahren	35
4.1 Globales MPSC-Verfahren	35
4.2 Diskrete Projektionsverfahren für die inkompressiblen Gleichungen	38
4.3 Diskretes nichtlineares Projektionsverfahren für das Modellproblem . . .	39
5 Lösung linearer und nichtlinearer Gleichungssysteme	47
5.1 Fixpunkt-Defekt-Korrektur	47
5.2 Newton-Verfahren	48
5.2.1 Line-Search	49
5.3 BiCGSTAB	52
6 Advektions-Diffusions-Gleichung	53
6.1 Discontinuous-Galerkin-Verfahren	53
6.1.1 Discontinuous-Galerkin Diskretisierung für Advektion	54

6.1.2	Patankar-Verfahren	57
6.1.3	Linear-Scaling-Limiter	61
6.1.4	Patankar-DG	63
6.1.5	Maximumprinzip bei homogenen Transportgleichungen	66
6.1.6	Patankar-DG für Kavitation	71
6.2	DG für Diffusion	76
6.2.1	Recovery-based-Discontinuous-Galerkin (RDG)	78
6.3	DG für Advektions-Diffusions-Gleichungen	80
7	Kopplung von Strömung und Transport	83
7.1	Konservativität und Reproduktion konstanter Lösungen	83
7.2	Gemischte Finite Elemente für die Poisson-Gleichung	87
7.3	Conservative-Projection-Formulation	91
7.4	Local-Conservative-Projection-Formulation	92
7.5	Kopplung	95
8	Numerische Simulation	101
8.1	Hochviskose Strömungen	101
8.1.1	Poiseuille-Strömung	103
8.1.2	4:1-Kontraktion	109
8.2	Kavitation	109
9	Zusammenfassung und Ausblick	119
A	Anhang	121
A.1	Notation	121
A.2	Bestimmung einer Basis für $RT_1(\hat{K})$	122
	Literaturverzeichnis	125

Vorwort

Diese Arbeit entstand während meiner Zeit als wissenschaftlicher Mitarbeiter in der Arbeitsgruppe Analysis und Angewandte Mathematik am Institut für Mathematik der Universität Kassel. Ein Teil der Resultate wurde während meiner Zeit im Sonderforschungsbereich TR 30: *Prozessintegrierte Herstellung funktional gradierter Strukturen auf der Grundlage thermo-mechanisch gekoppelter Phänomene* erzielt.

Besonderer Dank gilt meinem Betreuer Prof. Dr. Andreas Meister, der mich bei der Anfertigung dieser Arbeit unterstützt und mich auch in schwierigen Zeiten bestärkt hat, die Arbeit fortzuführen.

Ebenso danke ich Prof. Dr.-Ing. Olaf Wunsch, der diese Arbeit aus ingenieurstechnischer Sicht betreut hat und jederzeit für auftretende Fragen zur Verfügung stand.

Weiter gilt mein Dank Prof. Dr. Stefan Turek, Dr. Abderrahim Ouazzi und dem restlichen Featflow-Team, die mit ihrem Code (www.featflow.de) die Programmgrundlage für das in dieser Arbeit entwickelte Verfahren gelegt und mir durch ihre Erklärungen den Einstieg erleichtert haben.

Außerdem bedanke ich mich bei den Mitarbeitern des Instituts und speziell bei denen der Arbeitsgemeinschaft Praktische Mathematik für die hilfreichen Diskussionen und das freundliche Arbeitsklima.

Nicht zuletzt danke ich meinen Freunden, meiner Familie und insbesondere meiner Freundin Jenny für ihre Unterstützung während aller Höhen und Tiefen, die bei der Erstellung dieser Arbeit aufgetreten sind.

1 Einleitung

Diese Arbeit befasst sich mit der numerischen Simulation von Transportprozessen innerhalb von Strömungen. Insbesondere ist dabei die Simulation der Temperaturverteilung in hochviskosen Kunststoffschmelzen, sowie die Simulation von Kavitation im Bereich von Mikroschäumen von Interesse. Beide Probleme verbindet, dass sie durch Kombination eines Transportmodells mit einem Modell für die zugrundeliegenden Strömung abgebildet werden können.

Bei den hier betrachteten Kunststoffschmelzen gilt es, die Änderung der Viskosität innerhalb einer inkompressiblen Strömung in Abhängigkeit von Scherrate und Temperaturverteilung zu erfassen. Damit sind Temperatur- und Geschwindigkeitsfeld beidseitig gekoppelt, da nicht nur das Geschwindigkeitsfeld direkten Einfluss auf den Wärmetransport hat, sondern auch die Temperatur die Viskosität beeinflusst. Das zugehörige Verhalten wird durch zwei dimensionslose Kennzahlen charakterisiert. Die Reynolds-Zahl ergibt sich aus dem Verhältnis von Trägheits- zu Zähigkeitskräften, die Prandtl-Zahl bestimmt sich aus dem Quotienten von kinematischer Viskosität und Temperaturleitfähigkeit. Aufgrund hoher Viskosität und geringer Wärmeleitfähigkeit weisen derartige Schmelzen kleine Reynolds-Zahlen und sehr große Prandtl-Zahlen auf. Beispielsweise in Schneckenextrudern liegen die Größenordnungen etwa bei 10^{-2} für die Reynolds-Zahl und 10^7 für die Prandtl-Zahl. Dies bedeutet, dass das Geschwindigkeitsfeld einen diffusiven Charakter aufweist, wohingegen innerhalb der Temperaturverteilung der Wärmetransport stark dominant ist.

Bei der Herstellung von Mikroschäumen wird ein Treibmittel in die Kunststoffschmelze injiziert, welches dann in Form winziger Gasbläschen mit der Schmelze transportiert wird. In Abhängigkeit der Druckverteilung ändert sich der Blasenradius, so dass durch eine mechanisch erzwungene Druckabsenkung ein Aufschäumen der Schmelze hin zu einem Mikroschaum erfolgen kann. Wir wollen hier das Blasenwachstum innerhalb der Schmelze im Kontext lokaler Druckvariationen simulieren. Das zugehörige Strömungsmodell muss dabei neben der vom Dampfanteil abhängigen Variation der Viskosität auch die Änderungen der Dichte erfassen. Insbesondere kann das Geschwindigkeitsfeld damit nicht mehr als inkompressibel betrachtet werden.

Ausgangspunkt für die numerische Simulation dieser beiden Prozesse sind die inkompressiblen Navier-Stokes-Gleichungen, zu deren numerischer Approximation in [Tur99] ein, auf einer Finite-Elemente-Diskretisierung basierendes, diskretes Projektionsverfahren vorgestellt wurde. Der Einsatz Finiten-Elemente bietet sich hier an, da die betrachteten Reynolds-Zahlen klein sind. Im Fall der Kunststoffschmelzen muss das Verfahren derart erweitert werden, dass Variationen der Viskosität berücksichtigt werden können. Auch zur Simulation der Kavitation durch das in [Sau00] eingeführte Kontinuumsmodell sind weitere Modifikationen nötig. Dieses Modell betrachtet nicht die einzelnen Blasen,

sondern ausschließlich den Gasanteil innerhalb der Strömung. Die daraus resultierenden Gleichungen zur Beschreibung des Strömungsfeldes erweitern die inkompressiblen Gleichungen um variable Viskosität und Dichte, sowie eine inhomogene Divergenzbedingung. Insbesondere ist die Divergenz der Geschwindigkeit dabei abhängig vom Druck und der Dampfdichte. Dies gilt es auch im Verfahren zu berücksichtigen, wobei die Abhängigkeit der Divergenz vom Druck auf ein nichtlineares Problem führt.

Zur Beschreibung der Transportprozesse koppeln wir die Gleichungen für die Strömung mit einer Advektions-Diffusions-Gleichung. Dabei ist die Wärmeverteilung innerhalb der Kunststoffschmelzen stark vom Wärmetransport dominiert, im Fall der Kavitation handelt es sich sogar um eine reine Transportgleichung. Daher bietet es sich an, ein Discontinuous-Galerkin-Verfahren zu verwenden, insbesondere können wir so die Konservativität der transportierten Größen garantieren. Ein wesentlicher Aspekt bei der Betrachtung des Kavitationsprozesses ist, dass der Dampfanteil definitionsgemäß zwischen 0% und 100% liegt und auch das numerische Verfahren die Einhaltung dieses Maximumprinzips garantieren sollte. Ein Verfahren, welches im Fall homogener Transportgleichungen ein solches Maximumprinzip gewährleisten kann, wurde in [ZS10] vorgestellt. Im Kontext des oben genannten Kavitationsmodells ist dieses Verfahren aufgrund des auftretenden nichtlinearen Quellterms allerdings nicht ohne Modifikationen anwendbar. Daher führen wir hier erstmalig das Pantkar-DG-Verfahren ein, welches die gewünschte Beschränktheit auch für das Kavitationsmodell garantiert. Der Name leitet sich von den in [BDM03] zur Positivitätserhaltung bei gewöhnlichen Differentialgleichungen vorgestellten Patankar-Verfahren ab, die hier zur Behandlung der Quellterme eingesetzt werden.

Haben wir zwei Löser, die in der Lage sind, die Approximation von Strömung und Transport in der gewünschten Weise durchzuführen, so müssen wir noch garantieren, dass die jeweiligen Diskretisierungen zueinander kompatibel sind. Eine Verletzung der Kompatibilität führt zu unphysikalischen und damit unbrauchbaren Approximationen. Durch ein geeignetes Postprocessing des Geschwindigkeitsfeldes kann die Kompatibilität nachträglich hergestellt werden. In [CDMW98] wird ein globales Postprocessing-Verfahren zur Bestimmung lokal konservativer Geschwindigkeitsfelder vorgestellt, es ermöglicht aber auch die Einhaltung der Kompatibilität. Durch eine neuartige lokale Betrachtungsweise werden wir hier den Aufwand des Verfahrens reduzieren, indem wir das Lösen großer linearer Gleichungssysteme vermeiden.

Die Arbeit ist wie folgt gegliedert. Im zweiten Kapitel führen wir die zu Grunde liegenden Gleichungen ein. Diese sind die inkompressiblen Navier-Stokes-Gleichungen, in klassischer und verallgemeinert newtonscher Form, sowie das Kavitationsmodell aus [Sau00]. Verallgemeinert newtonsch bedeutet hier, dass die Variation der Viskosität in Abhängigkeit von Scherrate und Temperatur berücksichtigt wird. Wir zeigen, dass das Kavitationsmodell ein Maximumprinzip für die Dampfdichte beinhaltet und betrachten außerdem klassische Lösungen der inkompressiblen Gleichungen, welche später zur Verifikation des entwickelten Lösert dienen.

In Kapitel 3 befassen wir uns mit der Diskretisierung der Strömungsgleichungen. Wir nutzen dabei eine auf den nichtkonformen Rannacher-Turek-Elementen ([RT92]) basierende Finite-Elemente-Diskretisierung und stellen diese zunächst für die klassischen

inkompressiblen Navier-Stokes-Gleichungen und anschließend für die allgemeineren Gleichungen vor. Außerdem wird die Handhabung von Ausflussrandbedingungen bei variabler Viskosität diskutiert und in diesem Zusammenhang eine Umformulierung des Reibungsspannungstensors vorgeschlagen.

Die Lösung des aus der Diskretisierung resultierenden Gleichungssystems wird im vierten Kapitel behandelt. Zunächst wird das aus [Tur99] stammende diskrete Projektionsverfahren für die inkompressiblen Gleichungen vorgestellt. Da die Umsetzung dieses Verfahrens für das verallgemeinerte Problem unklar ist, wird eine neuartige Herleitung des Verfahrens angegeben und anschließend auf den allgemeineren Fall übertragen. Dies führt auf ein Projektionsverfahren, bei dem eine nichtlineare Pressure-Poisson-Gleichung gelöst werden muss.

Die Lösung der nichtlinearen Pressure-Poisson-Gleichung mit einem Newton-Verfahren wird in Kapitel 5 betrachtet. Weiter werden die Fixpunkt-Defekt-Korrektur zur Lösung der diskreten Impuls-Gleichung, sowie das BiCGSTAB-Verfahren für lineare Systeme besprochen.

Damit ist die Betrachtung des Strömungslösers abgeschlossen und wir widmen uns in Kapitel 6 den Discontinuous-Galerkin-Verfahren zur Approximation der Advektions-Diffusions-Gleichung. Zunächst wird dabei ein DG-Verfahren für reine Transportprobleme betrachtet. Mit dem Ziel, die Einhaltung des Maximumprinzips für die Dampfdichte bei der Kavitationssimulation zu sichern, betrachten wir das Patankar-Verfahren aus [BDM03] und das Verfahren zur Gewährleistung von Maximumprinzipien bei homogenen Transportgleichungen aus [ZS10]. Die Kombination und Erweiterung dieser beiden Techniken im neuen Patankar-DG-Verfahren ermöglicht dann die Garantie der Beschränktheit auch in unserem allgemeineren Fall. Abschließend beschreiben wir noch die Recovery-based-Discontinuous-Galerkin Techniken aus [LLR07, Lo11], um auch die diffusiven Terme der Gleichung zu berücksichtigen.

Die Kopplung der beiden Löserkomponenten wird in Kapitel 7 behandelt. In [DSW04] wurden die Auswirkungen eines numerischen Geschwindigkeitsfeldes innerhalb der DG-Diskretisierung untersucht und gezeigt, dass es zur Reproduktion konstanter Lösungen erforderlich ist, dass Geschwindigkeitsfeld und Diskretisierung kompatibel sind. Wir wiederholen diese Resultate, bevor wir mit der Conservative-Projection-Formulation aus [CDMW98] ein Verfahren vorstellen, um ein kompatibles Geschwindigkeitsfeld in einem Postprocessing-Schritt zu berechnen. Wesentlicher Bestandteil des Verfahrens ist die Lösung eines Poisson-Problems mit $H(\text{div})$ -konformen Finiten-Elementen, was wir ebenfalls besprechen werden. Schließlich führen wir hier erstmalig die Local-CPF als effiziente Variante der CPF ein, welche ohne das Lösen großer Gleichungssysteme auskommt.

In Kapitel 8 werden die Ergebnisse verschiedener numerischer Simulationen der hochviskosen Kunststoffschmelzen, sowie der Kavitation gezeigt. Insbesondere lässt sich das entwickelte Verfahren im Fall der nicht-isothermen verallgemeinert newtonschen Fluide anhand einer Kanalströmung validieren.

2 Gleichungen

In diesem Kapitel werden wir die Gleichungen einführen, welche den weiteren Betrachtungen zugrunde liegen. Zum einen sind dies Gleichungen zur Beschreibung hochviskoser Kunststoffschmelzen, zum anderen Gleichungen zur Beschreibung von Kavitationseffekten. Während im ersten Fall insbesondere die Temperaturabhängigkeit der Viskosität des Fluids erfasst werden muss, liegt der Schwerpunkt im zweiten Fall auf der Beschreibung von Dichteunterschieden in Abhängigkeit des Drucks. In beiden Fällen beginnen wir die Herleitung zunächst mit den zweidimensionalen Bilanzgleichungen für Masse und Impuls eines Strömungsfelds in ihrer allgemeinen Form. Nach [Böh00], [RG01] gilt für die Massenbilanz

$$\partial_t \varrho + \nabla \cdot (\varrho \mathbf{u}) = 0, \quad (2.1)$$

und für die Impulsbilanz

$$\varrho (\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u}) - \nabla \cdot \mathbf{T} + \nabla p = \mathbf{f}, \quad (2.2)$$

wobei Dichte, Geschwindigkeit und Druck mit ϱ , $\mathbf{u} = (u, v)^T$ und p bezeichnet werden. Volumenkräfte werden durch $\mathbf{f} = (f_u, f_v)^T$ beschrieben und $\mathbf{T}$ stellt den Reibungsspannungstensor dar. Alle diese Größen sind von der Zeit t , dem Ort $\mathbf{x} = (x, y)^T$ und eventuell noch weiteren Größen abhängig.

Im Folgenden gehen wir davon aus, dass der Reibungsspannungstensor dem (verallgemeinerten) newtonschen Stoffgesetz

$$\mathbf{T} = 2\eta \mathbf{D} \quad (2.3)$$

genügt, wobei der Verzerrungstensor durch $\mathbf{D} = \frac{1}{2} (\nabla \mathbf{u} + \nabla \mathbf{u}^T)$ gegeben ist und wir mit η die dynamische Viskosität bezeichnen. Dabei sprechen wir vom verallgemeinerten newtonschen Stoffgesetz, falls $\eta \neq \text{const}$ gilt.

Diese Gleichungen werden entsprechend der jeweiligen Problemstellung noch um eine Transportgleichung ergänzt. Bei der Modellierung von Kavitation können wir uns auf eine reine Transportgleichung ohne Berücksichtigung diffusiver Effekte beschränken, bei der Betrachtung der Temperatur hochviskoser Fluide werden wir auch den geringen diffusiven Anteil berücksichtigen.

Da wir die Gleichungen im Weiteren ausschließlich auf beschränkten Gebieten betrachten, müssen zur vollständigen Beschreibung des Problems noch passende Anfangs- und Randbedingungen angegeben werden. Wir werden sowohl Dirichlet-Randbedingungen, als auch Neumann-Randbedingungen berücksichtigen, wobei die Beschreibung von Ausflussrandbedingungen von besonderem Interesse ist.

2.1 Die inkompressiblen Navier-Stokes Gleichungen

Basierend auf Verfahren zur numerischen Lösung der inkompressiblen Navier-Stokes Gleichungen werden wir im Folgenden entsprechende Verfahren für unsere Problemstellungen herleiten. Die Gleichungen erhalten wir aus (2.1), (2.2) und (2.3) in dem wir Dichte und Viskosität als konstant voraussetzen und dann die Identität $\nabla \cdot (\nabla \mathbf{u}^T) = \nabla(\nabla \cdot \mathbf{u})$ ausnutzen. Sie lauten somit

$$\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u} - \nu \Delta \mathbf{u} + \nabla P = \mathbf{g}, \quad (2.4a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (2.4b)$$

wobei wir die kinematische Viskosität mit $\nu = \eta/\varrho = \text{const}$ und den kinematischen Druck mit $P = p/\varrho$ bezeichnen. Für die auftretende Volumenkraft gilt $\mathbf{g} = \mathbf{f}/\varrho$.

Aus (2.1) ist ersichtlich, dass sich die Inkompressibilitätsbedingung (2.4b) schon unter der Annahme $\partial_t \varrho + \mathbf{u} \cdot \nabla \varrho = 0$, d. h. vernachlässigbarer Dichteänderungen entlang von Partikelbahnen, herleiten lässt. Trotzdem werden im Allgemeinen die Gleichungen mit konstanter Dichte als die inkompressiblen Navier-Stokes Gleichungen bezeichnet, da eine konstante Dichte die Inkompressibilität bedingt. Dies werden wir im Weiteren auch so handhaben.

Abschließend wollen wir noch die Gleichungen in entdimensionalisierter Form betrachten. Dazu wird jeder dimensionsbehafteten Größe ϕ mittels $\phi = \phi_{\text{ref}} \phi^*$ eine dimensionslose Größe ϕ^* zugeordnet, wobei ϕ_{ref} eine dimensionsbehaftete Konstante darstellt. Da Dichte und Viskosität konstant sind wählen wir $\varrho_{\text{ref}} = \varrho$ und $\eta_{\text{ref}} = \eta$. Mit $t_{\text{ref}} = x_{\text{ref}}/u_{\text{ref}}$, $p_{\text{ref}} = \varrho_{\text{ref}} u_{\text{ref}}^2$ und $f_{\text{ref}} = x_{\text{ref}}/(\varrho_{\text{ref}} u_{\text{ref}}^2)$ erhalten wir aus (2.4) dann

$$\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u} - Re^{-1} \Delta \mathbf{u} + \nabla p = \mathbf{f}, \quad (2.5a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (2.5b)$$

wobei zur Vereinfachung der Schreibweise die dimensionslosen Größen wie die dimensionsbehafteten Größen bezeichnet wurden. Die dimensionslose Reynolds-Zahl ist durch $Re = \varrho_{\text{ref}} u_{\text{ref}} x_{\text{ref}} / \eta_{\text{ref}}$ gegeben und steuert die Dominanz des diffusiven Terms gegenüber dem konvektiven Term.

2.2 Verallgemeinert newtonsche und nicht-isotherme Gleichungen

Verzichten wir auf die Annahme einer konstanten Viskosität und beschränken uns auf eine konstante Dichte, dann ergeben sich aus (2.1), (2.2) und (2.3) die Gleichungen

$$\varrho (\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u}) - \nabla \cdot (2\eta \mathbf{D}) + \nabla p = \mathbf{f}, \quad (2.6a)$$

$$\nabla \cdot \mathbf{u} = 0. \quad (2.6b)$$

Dabei kann der diffusive Term unter Ausnutzung der Identitäten

$$\begin{aligned}\nabla \cdot (\eta \nabla \mathbf{u}^T) &= \eta \nabla (\nabla \cdot \mathbf{u}) + \nabla \mathbf{u}^T \cdot \nabla \eta, \\ \nabla (\eta (\nabla \cdot \mathbf{u})) &= \eta \nabla (\nabla \cdot \mathbf{u}) + (\nabla \cdot \mathbf{u}) \nabla \eta, \\ \nabla \mathbf{u}^T \cdot \nabla \eta - (\nabla \cdot \mathbf{u}) \nabla \eta &= \nabla \mathbf{u}^* \cdot \nabla \eta,\end{aligned}$$

mit der Definition

$$\nabla \mathbf{u}^* = \begin{bmatrix} -\partial_y v & \partial_x v \\ \partial_y u & -\partial_x u \end{bmatrix} \quad (2.7)$$

auch in der Form

$$\nabla \cdot (2\eta \mathbf{D}) = \nabla \cdot (\eta \nabla \mathbf{u}) + \nabla (\eta (\nabla \cdot \mathbf{u})) + \nabla \mathbf{u}^* \cdot \nabla \eta \quad (2.8)$$

geschrieben werden. Im letzten Term treten dabei nur Ableitungen erster Ordnung auf und aufgrund der Divergenzfreiheit (2.6b) entfällt der mittlere Term. Dies werden wir in Abschnitt 3.2.4 nutzen, um auf einfache Weise Ausflussrandbedingungen beschreiben zu können. Zur Erklärung der Notation verweisen wir erneut auf A.1 im Anhang.

Um auch nicht-isotherme Effekte berücksichtigen zu können, ergänzen wir die Gleichungen (2.6) mit einer Bilanzgleichung für die Temperatur θ . Diese hat nach [Böh00], [RG01] die Form

$$\rho c (\partial_t \theta + \mathbf{u} \cdot \nabla \theta) - k \Delta \theta = q, \quad (2.9)$$

wobei k die Wärmeleitfähigkeit und c die spezifische Wärmekapazität bezeichnet. Beide Größen wollen wir im Folgenden als konstant voraussetzen. Der Quellterm

$$q = \eta (2(\partial_x u)^2 + (\partial_y u + \partial_x v)^2 + 2(\partial_y v)^2) \quad (2.10)$$

modelliert dabei die Dissipation, d.h. die Entstehung von Wärme durch Reibung.

Zur Beschreibung konkreter Materialien benötigen wir weitere Kenntnisse über die Viskosität. Für unsere Anwendungen im Bereich hochviskoser Kunststoffströmungen folgen wir [Con08] und wählen exemplarisch ein Cross-Modell

$$\frac{\eta(\dot{\gamma}, \theta)}{\eta_0} = \frac{a(\theta)}{(1 + \kappa_1 \dot{\gamma} a(\theta))^{\kappa_2}}, \quad (2.11)$$

bei dem die Temperaturabhängigkeit durch die WLF-Beziehung

$$\ln a(\theta) = -\frac{C_1(\theta - \theta_r)}{C_2 + \theta - \theta_r} \quad (2.12)$$

gegeben ist. Hierbei hängt die Viskosität neben der Temperatur auch von der Scherrate

$$\dot{\gamma} = \sqrt{2(\partial_x u)^2 + (\partial_y u + \partial_x v)^2 + 2(\partial_y v)^2} \quad (2.13)$$

ab und die weiteren konstanten Parameter $\eta_0, \kappa_1, \kappa_2, C_1, C_2$ und θ_r sind entsprechend konkreter Materialeigenschaften zu wählen. Ohne auf die konkrete Wahl dieser Parameter einzugehen, wird in Abbildung 2.1 das scherentzähende Verhalten des Viskositätsmodells (2.11) verdeutlicht. Bei wachsender Scherrate wird die Viskosität deutlich geringer.

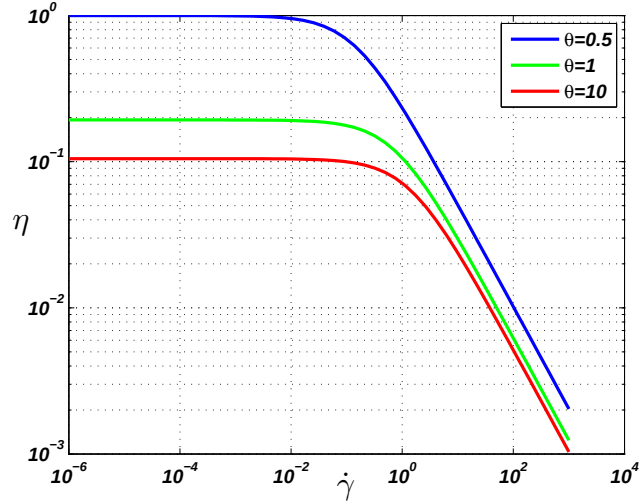


Abbildung 2.1: Veranschaulichung des Viskositätsverhaltens nach (2.11) in Abhängigkeit von Scherrate und Temperatur. Größere Temperaturen und wachsende Scherrate führen zu einer Abnahme der Viskosität.

Ebenso wird die Viskosität durch höhere Temperaturen verringert. Eine detailliertere Ausführung dazu findet sich in [Con08] und der enthaltenen Literatur.

Wir bemerken noch, dass durch die Temperaturabhängigkeit der Viskosität die Gleichungen (2.6a) und (2.9) beidseitig gekoppelt sind. Dagegen ist die Viskosität im Fall $\theta = \theta_r = \text{const}$ lediglich von der Scherrate abhängig und die Temperaturgleichung kann nach Bestimmung des Geschwindigkeitsfeldes separat gelöst werden.

Schließlich wollen wir auch diese Gleichungen in entdimensionalisierter Form betrachten. Mit den selben Ansätzen zur Entdimensionalisierung wie bei (2.5) und der zusätzlichen Wahl $\theta_{\text{ref}} = u_{\text{ref}}^2 \eta_{\text{ref}} / k_{\text{ref}}$ gilt dann

$$\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u} - Re^{-1} \nabla \cdot (2\eta \mathbf{D}) + \nabla p = \mathbf{f}, \quad (2.14a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (2.14b)$$

$$\partial_t \theta + \mathbf{u} \cdot \nabla \theta - Pe^{-1} \Delta \theta = Pe^{-1} q, \quad (2.14c)$$

wobei die Péclet-Zahl durch $Pe = \varrho_{\text{ref}} c_{\text{ref}} u_{\text{ref}} x_{\text{ref}} / k_{\text{ref}}$ gegeben ist. Mit der Péclet-Zahl kann nun auch eine Charakterisierung des Temperaturfeldes erfolgen. Sie misst das Verhältnis von transportierter und geleiteter Wärme und ergibt sich als Produkt von Reynolds-Zahl Re und Prandtl-Zahl $Pr = \eta_{\text{ref}} c_{\text{ref}} / k_{\text{ref}}$. Für hochviskose Kunststoffschmelzen gilt nach [Con08] etwa $Re \approx 10^{-2}$ und $Pe \approx 10^5$, demnach weist die Impulsgleichung einen diffusiven Charakter auf, während die Temperaturgleichung stark transportdominant ist.

2.3 Kavitationsmodell

Bei der Simulation flüssiger Mikroschäume muss Kavitationseffekten Rechnung getragen werden. Fällt der Druck der Flüssigkeit lokal unterhalb den Dampfdruck, so setzt dort ein Wachstum der in der Flüssigkeit enthaltenen Dampfblasen ein. Das Wachstum einzelner Blasen wird durch die Rayleigh-Plesset Gleichung

$$Rd_{tt}R + \frac{3}{2}(d_tR)^2 = \frac{p_D - p_F}{\varrho_F} - \frac{2\sigma}{\varrho_F R} - 4\frac{\eta_F}{R}d_tR \quad (2.15)$$

beschrieben, siehe [Bre95]. Hier steht $R(t)$ für den Blasenradius zum Zeitpunkt t und σ bezeichnet die Oberflächenspannung der Blase. Mittels der Indizes F und D können die weiteren Größen der Flüssigkeit oder dem Dampf zugeordnet werden.

Zur Beschreibung der Kavitation verwenden wir im Folgenden das in [Sau00] und [Fro04] beschriebene Modell. Wir werden zunächst die Herleitung für den dreidimensionalen Fall wiederholen und anschließend zur Behandlung zweidimensionaler Probleme modifizieren. Das Modell betrachtet anstatt jeder einzelnen Blase, das Phasengemisch aus Flüssigkeit und Dampf. Dazu sei α der Volumenanteil der dampfförmigen Phase pro Einheitsvolumen des Gemisches, nachfolgend kurz Dampfanteil genannt. Es gilt also

$$\alpha = \frac{V_D}{V}, \quad (2.16)$$

wobei V mit der zugehörigen Indizierung hier für das entsprechende Volumen steht. Weiter erhalten wir damit und der Zerlegung $V = V_F + V_D$ die Beziehung

$$V_F = V - V_D = (1 - \alpha)V. \quad (2.17)$$

Definieren wir nun die Keimzahl n_0 als die Anzahl der Kavitationskeime pro Einheitsvolumen Flüssigkeit, so erhalten wir die Anzahl N der in V enthaltenen Blasen mittels $N = n_0 V_F$, und es gilt unter der Annahme kugelförmiger Blasen

$$\alpha = \frac{V_D}{V_D + V_F} = \frac{N \frac{4}{3}\pi R^3}{N \frac{4}{3}\pi R^3 + V_F} = \frac{n_0 V_F \frac{4}{3}\pi R^3}{n_0 V_F \frac{4}{3}\pi R^3 + V_F} = \frac{n_0 \frac{4}{3}\pi R^3}{1 + n_0 \frac{4}{3}\pi R^3}.$$

Damit können wir den Dampfanteil in Abhängigkeit des Blasenradius ausdrücken, insbesondere gilt $0 \leq \alpha < 1$. Umgekehrt können wir auch den Blasenradius in Abhängigkeit der Dampfanteile ausdrücken, wir erhalten dann

$$R = \sqrt[3]{\frac{3}{4n_0\pi} \frac{\alpha}{1 - \alpha}}. \quad (2.18)$$

Mit Hilfe der Dampfdichte wollen wir im Folgenden Dichte und Viskosität des Phasengemisches modellieren. Die jeweilige Größe bestimmen wir durch die lineare Interpolation

$$\varrho = \varrho_F + \alpha(\varrho_D - \varrho_F), \quad (2.19a)$$

$$\eta = \eta_F + \alpha(\eta_D - \eta_F), \quad (2.19b)$$

2 Gleichungen

aus den entsprechen Größen der einzelnen Phasen. Zusätzlich nehmen wir im Weiteren an, dass $\varrho_D, \varrho_F, \eta_D$ und η_F konstant sind.

Wir wollen uns noch kurz mit der Definition der Keimzahl befassen. Alternativ wäre eine Definition über das Volumen des Gemisches, in der Form $N/V = \text{const}$, möglich. Den Vorteil des gewählten Ansatzes $N/V_F = \text{const}$ erläutert folgendes Beispiel. Wir betrachten $V_F=1$ und $N_0 = 100$ Blasen mit Radius $R_0 = 3 \cdot 10^{-3}$, d. h. $V_D \approx 10^{-5}$ und damit $V \approx 1.00001$. Nun wachsen die Blasen um den Faktor 25, wobei wir V konstant halten. Dann ändert sich im Fall $N/V = \text{const}$ die Anzahl der Blasen nicht und damit erhalten wir $V_D \approx 0.177$ und $V_F \approx 0.823$. Im anderen Fall dagegen gilt dann $V_F \approx 0.8498$ und $V_D \approx 0.1502$, wobei die Anzahl der Blasen auf 85 gefallen ist. Das bedeutet das ein Anwachsen der Blasen nicht wie im ersten Fall ausschließlich Flüssigkeit verdrängt, sondern mit der verdrängten Flüssigkeit auch Blasen aus V verschwinden, was sicherlich eine sinnvolle Modellierung ist.

Zur Beschreibung des Transports des Dampfanteils wollen wir nun eine Transportgleichung der Form

$$\partial_t \alpha + \nabla \cdot (\alpha \mathbf{u}) = d_t \alpha + \alpha (\nabla \cdot \mathbf{u}) = f_\alpha \quad (2.20)$$

herleiten, wobei für die totale bzw. materielle Zeitableitung $d_t = \partial_t + \mathbf{u} \cdot \nabla$ gilt. Dazu betrachten wir zunächst die totale zeitliche Änderung des Dampfanteils innerhalb eines fest gewählten Volumens V und erhalten

$$d_t V_D = d_t \left(N \frac{4}{3} \pi R^3 \right) = \frac{4}{3} \pi R^3 d_t N + N d_t \left(\frac{4}{3} \pi R^3 \right). \quad (2.21)$$

Nach (2.17) gilt

$$N = n_0 V_F = n_0 V (1 - \alpha),$$

und damit auch

$$d_t N = d_t (n_0 (1 - \alpha) V) = -n_0 V d_t \alpha,$$

da V fest gewählt wurde und sich lediglich V_D und V_F mit der Zeit ändern. Mit diesen beiden Gleichungen folgt dann aus (2.21) unter Berücksichtigung von (2.16)

$$\begin{aligned} d_t \alpha &= \frac{1}{V} d_t V_D = \frac{1}{V} \left(-\frac{4}{3} \pi R^3 n_0 V d_t \alpha + n_0 V (1 - \alpha) d_t \left(\frac{4}{3} \pi R^3 \right) \right) \\ &= -\frac{4}{3} \pi R^3 n_0 d_t \alpha + n_0 (1 - \alpha) d_t \left(\frac{4}{3} \pi R^3 \right), \end{aligned}$$

und damit schließlich

$$d_t \alpha = \frac{n_0 (1 - \alpha)}{1 + \frac{4}{3} \pi R^3 n_0} d_t \left(\frac{4}{3} \pi R^3 \right). \quad (2.22)$$

Als nächstes wollen wir eine Darstellung für die Divergenz des Geschwindigkeitsfeldes herleiten. Mit (2.1) und (2.19a) ergibt sich

$$\nabla \cdot \mathbf{u} = -\frac{d_t \varrho}{\varrho} = -\frac{(\varrho_D - \varrho_F) d_t \alpha}{\alpha \varrho_D + (1 - \alpha) \varrho_F} = -\frac{(\varrho_D - \varrho_F)}{\frac{\alpha}{1 - \alpha} \varrho_D + \varrho_F} \cdot \frac{d_t \alpha}{1 - \alpha}.$$

Wir machen nun zwei Annahmen. Erstens soll die Dichte des Dampfes viel geringer sein als die Dichte der Flüssigkeit, d. h. $\varrho_D \ll \varrho_F$. Zweitens soll die Masse des Dampfes

gegenüber der Masse der Flüssigkeit innerhalb eines Volumens vernachlässigbar sein und damit $V_D \varrho_D = \alpha V \varrho_D \ll (1 - \alpha) V \varrho_F = V_F \varrho_F$ gelten. Mit diesen Annahmen und (2.22) ergibt sich dann für die Divergenz des Geschwindigkeitsfeldes

$$\nabla \cdot \mathbf{u} = \frac{d_t \alpha}{1 - \alpha} = \frac{n_0}{1 + \frac{4}{3} \pi R^3 n_0} d_t \left(\frac{4}{3} \pi R^3 \right). \quad (2.23)$$

Damit haben wir die Transportgleichung für den Dampfanteil hergeleitet. Durch Einsetzen von (2.22) und (2.23) in (2.20) erhalten wir

$$\partial_t \alpha + \nabla \cdot (\alpha \mathbf{u}) = \frac{n_0}{1 + \frac{4}{3} \pi R^3 n_0} d_t \left(\frac{4}{3} \pi R^3 \right). \quad (2.24)$$

Dabei muss zur Bestimmung des Quellterms die Änderung des Blasenradius bekannt sein. Wir werden dazu im Folgenden eine vereinfachte Version der Rayleigh-Plesset Gleichung (2.15) benutzen. Unter Vernachlässigung der Ableitungen höherer Ordnung, sowie Viskosität und Oberflächenspannung ergibt sich die Rayleigh-Gleichung

$$\frac{dR}{dt} = \text{sign}(p_D - p_F) \sqrt{\frac{2}{3} \frac{|p_D - p_F|}{\varrho_F}}, \quad (2.25)$$

wobei der Dampfdruck p_D als konstant angenommen wird. Weiter ist die Funktion sign durch

$$\text{sign}(x) = \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

gegeben.

Mit den Gleichungen (2.2) und (2.3) zur Bestimmung der Strömung sind damit die Gleichungen für das Kavitations-Modell durch

$$\varrho (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) - \nabla \cdot (2\eta \mathbf{D}) + \nabla p = \mathbf{f}, \quad (2.26a)$$

$$\nabla \cdot \mathbf{u} = f_\alpha, \quad (2.26b)$$

$$\partial_t \alpha + \nabla \cdot (\alpha \mathbf{u}) = f_\alpha, \quad (2.26c)$$

gegeben. Unter Verwendung von (2.18) lässt sich der Quellterm aus (2.24) als

$$f_\alpha = \sqrt[3]{36\pi n_0 (1 - \alpha) \alpha^2} d_t R \quad (2.27)$$

schreiben. Im Gegensatz zu (2.18), ist in dieser Formulierung nun auch der Grenzfall $\alpha = 1$ zulässig, weshalb wir im Weiteren $0 \leq \alpha \leq 1$ verlangen.

Wir bemerken noch, dass das entstandene System sehr stark gekoppelt ist, da alle drei Gleichungen von Geschwindigkeit, Druck und Dampfanteil abhängig sind.

Um eine zweidimensionale Variante des Modell zu erhalten, nutzen wir $V_D = N\pi R^2$. Dann ergeben sich, analog zu obiger Herleitung, zwischen Dampfdichte und Blasenradius die Beziehungen

$$\alpha = \frac{n_0 \pi R^2}{1 + n_0 \pi R^2}, \quad R = \sqrt{\frac{\alpha}{1 - \alpha} \cdot \frac{1}{n_0 \pi}}. \quad (2.28)$$

2 Gleichungen

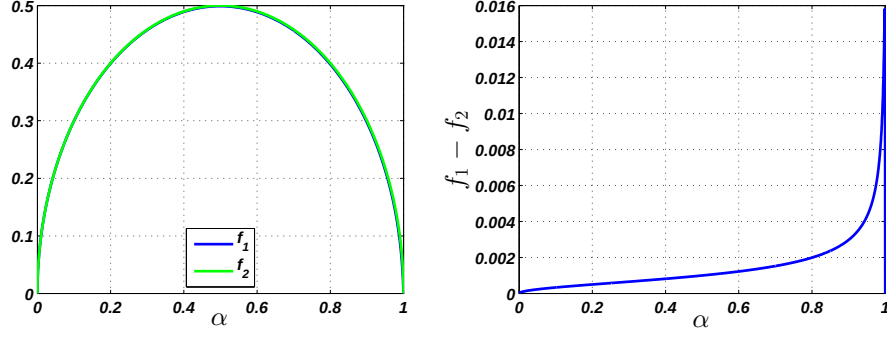


Abbildung 2.2: Vergleich des Quellterms aus (2.31) mit (2.29). Im linken Bild wurde $f_1 = \frac{\varrho_F - \varrho_D}{\varrho} \sqrt{\alpha(1-\alpha)^3}$ mit $\varrho_F = 1000$, $\varrho_D = 1$ und $f_2 = \sqrt{\alpha(1-\alpha)}$ geplottet. Die Differenz der beiden Funktionen ist rechts dargestellt.

Die einzige weitere Änderung betrifft den Quellterm. Dieser ergibt sich im zweidimensionalen Fall als

$$f_\alpha = \sqrt{4n_0\pi(1-\alpha)}\alpha d_t R. \quad (2.29)$$

Wir befassen uns nun noch kurz mit den bei der Herleitung von (2.23) gemachten Annahmen. Verzichten wir auf diese Annahmen, so ergibt sich die Divergenz im Zweidimensionalen nach einer Rechnung wie zuvor nun als

$$\nabla \cdot \mathbf{u} = \frac{\varrho_F - \varrho_D}{\varrho} \sqrt{4\pi n_0 \alpha (1-\alpha)^3} d_t R \quad (2.30)$$

und es gilt

$$\partial_t \alpha + \nabla \cdot (\alpha \mathbf{u}) = \frac{\varrho_F}{\varrho} \sqrt{4\pi n_0 \alpha (1-\alpha)^3} d_t R. \quad (2.31)$$

Im von uns betrachteten Kontext sind $\varrho_F = 1000$ und $\varrho_D = 1$ typische Werte. Ein Vergleich von (2.26b) mit (2.30) und (2.26c) mit (2.31) zeigt, dass die Unterschiede bei den gewählten Werten für ϱ_F und ϱ_D zu vernachlässigen sind. Abbildung 2.2 verdeutlicht dies und wir nutzen daher im Folgenden den Quellterm (2.29).

Auch wenn wir uns im Weiteren lediglich mit den zweidimensionalen Gleichungen befassen, so lassen sich die nachfolgend vorgestellten Techniken gleichermaßen im dreidimensionalen Fall anwenden. Daher lohnt es sich auch das Verhalten des dreidimensionalen Quellterms weiter zu untersuchen. Wir schreiben daher (2.29) und (2.27) zusammenfassend als

$$f_\alpha(\alpha, p) = \kappa(1-\alpha)^r \alpha^s \operatorname{sign}(p_D - p) \sqrt{\frac{2}{3} \frac{|p_D - p|}{\varrho_F}}, \quad (2.32)$$

mit einer Konstanten κ und $r, s < 1$.

Nach einer Entdimensionalisierung analog zu (2.5) und (2.14) erhalten wir aus (2.26) dann

$$\varrho (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) - Re^{-1} \nabla \cdot (2\eta \mathbf{D}) + \nabla p = \mathbf{f}, \quad (2.33a)$$

$$\nabla \cdot \mathbf{u} = f_\alpha, \quad (2.33b)$$

$$\partial_t \alpha + \nabla \cdot (\alpha \mathbf{u}) = f_\alpha, \quad (2.33c)$$

wobei hier $f_{\alpha, \text{ref}} = u_{\text{ref}}/x_{\text{ref}}$ genutzt wurde. Da die Dampfdichte ohnehin dimensionslos ist, bedarf es auch keiner Definition von α_{ref} .

Abschließend bemerken wir noch, dass die Verwendung des Reibungsspannungstensors in der Form (2.3) bereits die Inkompressibilität der Strömung voraussetzt. Der allgemein verwendete Zusammenhang zur Modellierung der Reibungsspannungen lautet $\mathbf{T} = 2\eta \mathbf{D} + \eta_V (\nabla \cdot \mathbf{u}) \mathbf{I}$, siehe [Böh00], wobei η_V die Volumenviskosität bezeichnet und $\mathbf{I}$ den Einheitstensor darstellt. Wir gehen also im Folgenden stets davon aus, dass $\eta_V (\nabla \cdot \mathbf{u})$ zu vernachlässigen ist.

2.3.1 Beschränktheit der Dampfdichte

Es ist klar, dass die Dampfdichte aufgrund ihrer Definition (2.16) stets $0 \leq \alpha \leq 1$ erfüllen muss. Um zu sehen, dass sich dies auch im System (2.33) widerspiegelt, betrachten wir die Änderung der Dampfdichte entlang der durch

$$\begin{aligned} x'(t) &= u(x(t), y(t)), & x(0) &= x_0, \\ y'(t) &= v(x(t), y(t)), & y(0) &= y_0, \end{aligned} \quad (2.34)$$

gegebenen Charakteristiken $(x(t), y(t))$. Wegen

$$d_t \alpha(\mathbf{x}(t), t) = \partial_t \alpha(\mathbf{x}(t), t) + \partial_x \alpha(\mathbf{x}(t), t) x'(t) + \partial_y \alpha(\mathbf{x}(t), t) y'(t)$$

ergibt sich die Änderung des Dampfanteils entlang der Charakteristiken mit (2.33c), (2.33b) und (2.34) als

$$d_t \alpha(x(t), y(t), t) = f_\alpha - \alpha (\nabla \cdot \mathbf{u}) = (1 - \alpha) f_\alpha. \quad (2.35)$$

Nach (2.32) lässt sich der Quellterm zusammenfassend in der Form

$$f_\alpha(\alpha, p) = C(p)(1 - \alpha)^r \alpha^s$$

schreiben. Setzen wir dies nun in (2.35) ein, so erhalten wir schließlich

$$d_t \alpha(x(t), y(t), t) = C(p)(1 - \alpha)^{r+1} \alpha^s. \quad (2.36)$$

Demnach ändert sich die Dampfdichte entlang der Bahnlinien nur für $\alpha \neq 0$ und $\alpha \neq 1$. Gilt zum Zeitpunkt t_0 die Beschränktheit $0 \leq \alpha(t_0) \leq 1$, so können wir aufgrund der Monotonie der rechten Seite von (2.36) die Beschränktheit $0 \leq \alpha(t) \leq 1$ für alle Zeiten t folgern.

2.4 Modellproblem

Wenn wir die Gleichungen (2.14) und (2.33) betrachten, so stellen wir fest, dass beide Systeme in

$$\varrho (\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u}) - Re^{-1} \nabla \cdot (2\eta \mathbf{D}) + \nabla p = \mathbf{f}_u, \quad (2.37a)$$

$$\nabla \cdot \mathbf{u} = g, \quad (2.37b)$$

$$\partial_t C + \nabla \cdot (C \mathbf{u}) - Pe^{-1} \Delta C = f_C, \quad (2.37c)$$

enthalten sind, wobei das in der Advektions-Diffusions-Gleichung auftretende C die Rolle von θ bzw. α einnimmt. Dieses Modellproblem dient uns zur allgemeinen Beschreibung des weiteren Vorgehens, ohne ständig zwischen den Gleichungen (2.14) und (2.33) unterscheiden zu müssen. Dabei interessieren wir uns insbesondere für den transportdominanten Fall $Pe^{-1} \ll 1$ in (2.37c), wobei wir auch $Pe^{-1} = 0$ zulassen.

An dieser Stelle wollen wir die zugehörigen Anfangs- und Randbedingungen angeben, die bei der Betrachtung des Modellproblems auf beschränkten Gebieten notwendig werden. Dazu nehmen wir im Weiteren an, dass das zugrundeliegende Gebiet $\Omega \subset \mathbb{R}^2$ ein beschränktes Polygon darstellt. Im Hinblick auf die späteren Anwendungsbereiche zerlegen wir den Rand $\partial\Omega$ in disjunkte Teilstücke in der Form $\partial\Omega = \partial\Omega_{in} \cup \partial\Omega_0 \cup \partial\Omega_{out}$. Dabei bezeichnet $\partial\Omega_{in}$ den Einflussrand, $\partial\Omega_{out}$ den Ausflussrand und $\partial\Omega_0$ steht für eine feste Wand. Es ist zweckmäßig noch die Bezeichnung $\partial\Omega_D = \partial\Omega_{in} \cup \partial\Omega_0$ einzuführen, da auf diesen beiden Randteilen Dirichlet-Randbedingungen vorgegeben werden.

Die für unsere Ziele passenden Anfangs- und Dirichlet-Randbedingungen für das Geschwindigkeitsfeld lauten

$$\mathbf{u}|_{t=0} = \mathbf{u}^0, \quad \mathbf{u}|_{\partial\Omega_{in}} = \mathbf{u}_{in}, \quad \mathbf{u}|_{\partial\Omega_0} = \mathbf{0},$$

wobei wir hier insbesondere die Haftbedingung an den festen Wänden berücksichtigen. Die beiden Dirichlet-Randbedingungen fassen wir auch als

$$\mathbf{u}|_{\partial\Omega_D} = \mathbf{u}_D \quad (2.38)$$

zusammen. Die Frage nach der „richtigen“ Ausflussrandbedingung lässt sich deutlich schwieriger beantworten. Der Vollständigkeit halber geben wir hier die nachfolgend verwendeten Ausflussrandbedingungen an. Warum diese Wahl sinnvoll ist, werden wir später in Kapitel 3 diskutieren. Wir setzen

$$(Re^{-1} (\eta \nabla \mathbf{u} \cdot \mathbf{n} + \eta (\nabla \cdot \mathbf{u}) \mathbf{n}) - p \mathbf{n})|_{\partial\Omega_{out}} = \mathbf{0}, \quad (2.39)$$

wobei hier $\mathbf{n}$ den äußeren Normalenvektor auf $\partial\Omega_{out}$ bezeichnet.

Für die transportierte Größe C lauten die Bedingungen

$$C|_{t=0} = C^0, \quad (2.40a)$$

$$C|_{\partial\Omega_D} = C_D, \quad (2.40b)$$

$$\nabla C \cdot \mathbf{n}|_{\partial\Omega_{out}} = 0, \quad (2.40c)$$

Ein Spezialfall ist der Fall reiner Kavitation, also $Pe^{-1} = 0$ in (2.37c). Denn dabei sind nur auf dem Einflussrand $\partial\Omega_{in}$ Randbedingungen für C erforderlich.

2.5 Poiseuille-Strömung

Wir wollen uns hier mit der klassischen Lösung der stationären inkompressiblen Navier-Stokes Gleichungen im Fall einer Schichtenströmung ohne Einfluss von Volumenkräften befassen. Die zugehörige Strömung nennt man auch Poiseuille-Strömung.

Wir betrachten zunächst einen unendlich langen ebenen Kanal der dimensionslosen Höhe $H = 1$, d. h. $\Omega = \mathbb{R} \times [0, 1]$. Die beiden Randstücke des Gebiets bei $y = 0$ und $y = 1$ seien feste Wände. Setzen wir nun in (2.5) die Bedingungen $\mathbf{f} = 0$, $v = 0$ und $\partial_t u = 0$ ein, so erhalten wir damit

$$u\partial_x u - Re^{-1}(\partial_{xx}u + \partial_{yy}u) + \partial_x p = 0, \quad \partial_y p = 0, \quad \partial_x u = 0.$$

Diese drei Gleichungen können wir weiter vereinfachen, so dass wir schließlich

$$\partial_{yy}u = Re\partial_x p, \tag{2.41}$$

$$\partial_y p = 0, \tag{2.42}$$

$$\partial_x u = 0, \tag{2.43}$$

erhalten. Die Geschwindigkeit variiert nach (2.43) nur über die Höhe des Kanal, während der Druck sich aufgrund von (2.42) nur in Strömungsrichtung ändert, also $u = u(y)$ und $p = p(x)$. Mit diesen Beobachtungen erkennen wir aus (2.41), dass $\partial_x p = \text{const}$ gilt und setzen daher $\partial_x p = -\delta p = \text{const}$. Der Druck ist also linear in Strömungsrichtung, d. h.

$$p(x) = -\delta p x + \text{const}.$$

Mit der Annahme von Haftbedingungen an beiden Wänden des Kanals, also $u(x, 0) = u(x, 1) = 0$ für alle $x \in \mathbb{R}$, erhalten wir schließlich für das Geschwindigkeitsfeld aus (2.41) die Darstellung

$$u(y) = \frac{Re}{2}\delta p(1-y)y. \tag{2.44}$$

Der Druck wird unter obigen Voraussetzungen wegen $\nabla(p + \text{const}) = \nabla p$ lediglich bis auf additive Konstanten bestimmt. Bei Betrachtung eines endlichen Kanals $\Omega = [0, L] \times [0, 1]$ können wir diese Mehrdeutigkeit vermeiden, indem wir am Ausflussrand explizit den Wert $p|_{x=L} = 0$ vorgeben. Diese Randbedingung ist im Einklang mit der in (2.39) angegebenen Randbedingung. Denn am Ausflussrand $x = L$ gilt

$$\mathbf{0} = Re^{-1}(\eta \nabla \mathbf{u} \cdot \mathbf{n} + \eta(\nabla \cdot \mathbf{u})\mathbf{n}) - p\mathbf{n}|_{x=L} = (-p, 0)^T|_{x=L}.$$

Damit ist Bedingung (2.39) in diesem Kontext gleichbedeutend mit $p|_{x=L} = 0$.

Ebenso können wir die Lösung der Temperaturgleichung (2.14c) betrachten. Als erstes gehen wir von einer konstanten Viskosität aus. Das zur Poiseuille-Strömung gehörige eingelaufene stationäre Temperaturfeld, erhalten wir aus (2.14c) unter den Annahmen $\partial_t \theta = \partial_x \theta = 0$. Es ergibt sich dann mit (2.10)

$$-Pe^{-1}\partial_{yy}\theta = Pe^{-1}(\partial_y u)^2$$

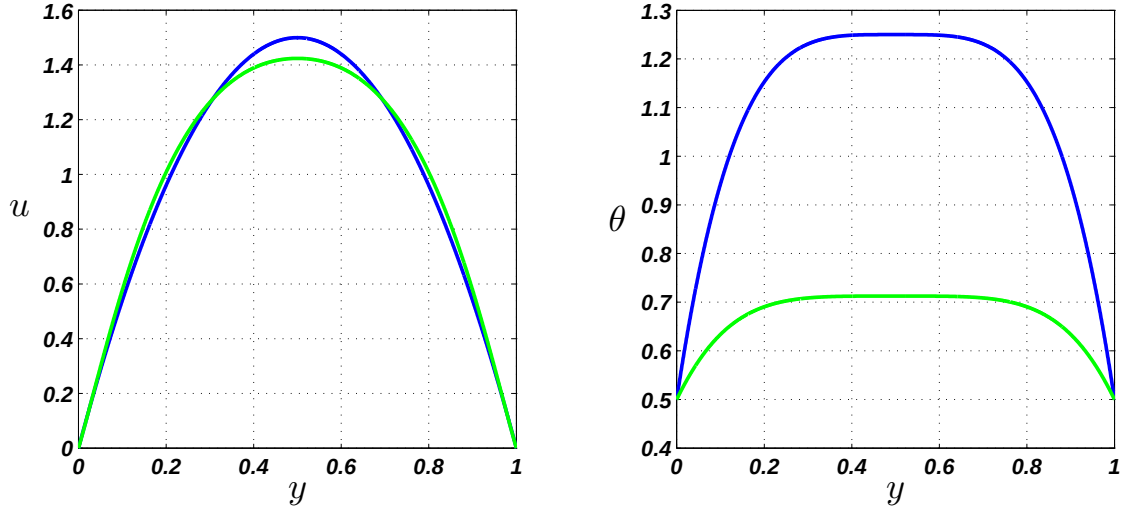


Abbildung 2.3: Exemparische x -Geschwindigkeit (links) und Temperaturverteilung für die Poiseuille-Strömung. Es werden newtonsches (blau) und nicht-newtonsches Fließverhalten (grün) gezeigt. Der Massenstrom ist in beiden Fällen identisch.

und unter Verwendung von (2.44) erhalten wir daraus

$$-\partial_{yy}\theta = \left(\frac{Re}{2} \delta p (1 - 2y) \right)^2.$$

Gehen wir weiter von Dirichlet-Randbedingungen $\theta(0) = \theta(1) = \theta_0$ an den festen Wänden aus, so ist die Temperaturverteilung durch

$$\theta(y) = \frac{1}{24} Re^2 \delta p^2 y (1 - 3y + 4y^2 - 2y^3) + \theta_0 \quad (2.45)$$

gegeben. Der grundsätzliche Verlauf von Geschwindigkeits- und Temperaturprofil sind in Abbildung (2.3) zu sehen. Die verwendeten Parameter sind momentan nicht von Bedeutung und werden in Kapitel 8 erläutert.

Im Fall variabler Viskosität können wir die Lösung allgemein nur numerisch bestimmen. Analog zu den vorherigen Betrachtungen variieren Geschwindigkeit und Temperatur ausschließlich über die Höhe des Kanals. Damit ergibt sich nach (2.13) die Scherrate hier als $\dot{\gamma} = |\partial_y u|$ und die zu lösenden Gleichungen lauten nun

$$\begin{aligned} \partial_y (\eta(|\partial_y u|, \theta) \partial_y u) &= -Re \delta p, \\ -\partial_{yy} \theta &= \eta(|\partial_y u|, \theta) (\partial_y u)^2. \end{aligned}$$

Dieses System kann auf verschiedene Weisen gelöst werden, wir geben hier eine einfache

Möglichkeit an. Mit den Substitutionen $U = \partial_y u$ und $\Theta = \partial_y \theta$ erhalten wir das System

$$\begin{aligned}\partial_y (\eta(|U|, \theta)U) &= -Re\delta p, \\ \partial_y \theta &= \Theta, \\ -\partial_y \Theta &= \eta(|U|, \theta)(\partial_y u)^2,\end{aligned}\tag{2.46}$$

von Differentialgleichungen erster Ordnung. Davon ausgehend, dass sowohl Geschwindigkeit als auch Temperatur wie im newtonschen Fall symmetrisch zur Kanalmitte sind, reicht es den halben Kanal zu betrachten. Die Randbedingungen lauten dann

$$\theta(0) = \theta_0, \quad \Theta(1/2) = 0, \quad U(1/2) = 0.\tag{2.47}$$

Insbesondere wird u auf $[0, 1/2]$ wie im newtonschen Fall monoton wachsend sein. Somit gilt $0 < \partial_y u = U = |U|$ und die Beträge in (2.46) entfallen. Das Randwertproblem (2.46), (2.47) können wir zumindest numerisch lösen. Haben wir U gefunden, so können wir auch u durch anschließende numerische Integration bestimmen.

Wir merken noch an, dass auch im Fall variabler Viskosität die Ausflussrandbedingung (2.39) ihre Gültigkeit behält. Ausschlaggebend dafür ist wieder, dass sich die Geschwindigkeit in Normalenrichtung nicht ändert.

Die Auswirkungen nicht-newtonscher Viskosität auf Geschwindigkeits- und Temperaturprofil bei gleichbleibendem Massenstrom sind ebenfalls in Abbildung 2.3 dargestellt. Die Poiseuille-Strömung mit zusätzlicher Berücksichtigung der Temperatur werden wir in Kapitel 8 als grundlegenden Testfall zur Validierung des im Weiteren beschriebenen Verfahrens zur Lösung des Modellproblems (2.37) benutzen.

3 Strömungslöser

In diesem Kapitel befassen wir uns mit der Diskretisierung und Lösung der Gleichungen für das Strömungsfeld (2.37a) und (2.37b). In Kapitel 6 werden wir dann die Advektions-Diffusions-Gleichung (2.37c) behandeln. Wir vollziehen diese Trennung, da wir zur Diskretisierung der Gleichungen verschiedene Verfahren nutzen wollen, um den jeweiligen Charakter der Gleichungen zu berücksichtigen. So werden wir das diffusive Strömungsfeld mit der Finiten-Elemente-Methode (FEM) diskretisieren und für die transportdominanten Gleichungen Discontinuous-Galerkin-Verfahren nutzen.

Zur Vereinfachung der Darstellung werden wir die im Weiteren angewandten Techniken zunächst anhand der inkompressiblen Navier-Stokes-Gleichungen (2.5) durchführen und dann die entsprechenden Verallgemeinerungen zur Behandlung der Gleichungen (2.37a) und (2.37b) besprechen.

Dabei setzen wir im Folgenden eine gewisse Vertrautheit mit der FEM voraus. Für einen Einstieg in die Thematik siehe etwa [CSS86, GS00, RG01, Cia02, ESW05, Qua09].

3.1 Inkompressible Navier-Stokes-Gleichungen

3.1.1 Schwache Formulierung

Wir wollen zunächst eine schwache Formulierung der inkompressiblen Navier-Stokes-Gleichungen (2.5) herleiten, mit deren Hilfe wir später die FEM-Diskretisierung erhalten werden. Dazu multiplizieren wir zunächst (2.5a) mit einer genügend glatten Testfunktion $\mathbf{v}$ und integrieren über Ω . Wir erhalten dann

$$\int_{\Omega} \partial_t \mathbf{u} \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} ((\mathbf{u} \cdot \nabla) \mathbf{u}) \cdot \mathbf{v} \, d\mathbf{x} - Re^{-1} \int_{\Omega} \Delta \mathbf{u} \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} \nabla p \cdot \mathbf{v} \, d\mathbf{x} = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x}. \quad (3.1)$$

Mittels partieller Integration folgt

$$- \int_{\Omega} \Delta \mathbf{u} \cdot \mathbf{v} \, d\mathbf{x} = \int_{\Omega} \nabla \mathbf{u} : \nabla \mathbf{v} \, d\mathbf{x} - \int_{\partial\Omega} (\nabla \mathbf{u} \cdot \mathbf{n}) \cdot \mathbf{v} \, ds, \quad (3.2)$$

sowie

$$\int_{\Omega} \nabla p \cdot \mathbf{v} \, d\mathbf{x} = - \int_{\Omega} p \nabla \cdot \mathbf{v} \, d\mathbf{x} + \int_{\partial\Omega} p \mathbf{v} \cdot \mathbf{n} \, ds. \quad (3.3)$$

Für die Notation verweisen wir auf A.1. Fassen wir (3.1), (3.2) und (3.3) zusammen, so ergibt sich

$$\begin{aligned} \int_{\Omega} \partial_t \mathbf{u} \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} ((\mathbf{u} \cdot \nabla) \mathbf{u}) \cdot \mathbf{v} \, d\mathbf{x} + Re^{-1} \int_{\Omega} \nabla \mathbf{u} : \nabla \mathbf{v} \, d\mathbf{x} - \int_{\Omega} p \nabla \cdot \mathbf{v} \, d\mathbf{x} \\ = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} + \int_{\partial\Omega} (Re^{-1} \nabla \mathbf{u} \cdot \mathbf{n} - p \mathbf{n}) \cdot \mathbf{v} \, d\mathbf{s}, \end{aligned} \quad (3.4)$$

wobei (2.39) als natürliche Randbedingung auftritt. In Abschnitt 3.1.2 werden wir uns noch genauer mit dieser Bedingung auseinandersetzen. Schließlich multiplizieren wir auch (2.5b) mit einer hinreichend glatten Testfunktion q und integrieren über das Gebiet. Damit erhalten wir

$$\int_{\Omega} q \nabla \cdot \mathbf{u} \, d\mathbf{x} = 0. \quad (3.5)$$

Wir definieren jetzt geeignete Funktionenräume für schwache Lösungen von (3.4) und (3.5). Wie bereits erwähnt, betrachten wir die Gleichungen stets auf beschränkten, polygonal berandeten Gebieten $\Omega \subset \mathbb{R}^2$. Mit

$$L^2(\Omega) = \left\{ q : \Omega \rightarrow \mathbb{R} \mid \int_{\Omega} q^2 \, d\mathbf{x} < \infty \right\}$$

bezeichnen wir den Raum der Funktionen, deren Quadrat über Ω Lebesgue-integrierbar ist und mit

$$H^1(\Omega) = \{ q \in L^2(\Omega) \mid \nabla q \in L^2(\Omega)^2 \}$$

den Raum der $L^2(\Omega)$ -Funktionen, deren schwache erste Ableitungen ebenfalls in $L^2(\Omega)$ liegen. Räume $H^m(\Omega)$ mit höheren Differenzierbarkeitsanforderungen werden analog definiert. Zur Berücksichtigung von Randbedingungen sei außerdem

$$\mathbf{U} = \{ \mathbf{u} \in H^1(\Omega)^2 \mid \mathbf{u}|_{\partial\Omega_D} = \mathbf{u}_D \},$$

sowie

$$\mathbf{V} = \{ \mathbf{v} \in H^1(\Omega)^2 \mid \mathbf{v}|_{\partial\Omega_D} = \mathbf{0} \}.$$

Die Skalarprodukte in $L^2(\Omega)$ bzw. $L^2(\partial\Omega)$ sind im Folgenden durch

$$(u, v)_{\Omega} = \int_{\Omega} uv \, d\mathbf{x}, \quad \langle u, v \rangle_{\partial\Omega} = \int_{\partial\Omega} uv \, d\mathbf{s}$$

gegeben, wobei wir die Indizes auch weglassen, falls es keine Unklarheiten geben kann. Im Fall vektorwertiger Funktionen verwenden wir die selbe Notation.

Unter Berücksichtigung der Randbedingungen (2.38) und (2.39) und mit den obigen Bezeichnungen ergibt sich die schwache Formulierung aus (3.4) und (3.5) wie folgt: Finde für $t > 0$ ein Paar $(\mathbf{u}(t), p(t)) \in \mathbf{U} \times L^2(\Omega)$ mit

$$(\partial_t \mathbf{u}, \mathbf{v}) + (\mathbf{u} \cdot \nabla \mathbf{u}, \mathbf{v}) + Re^{-1} (\nabla \mathbf{u}, \nabla \mathbf{v}) - (p, \nabla \cdot \mathbf{v}) = (\mathbf{f}, \mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{V}, \quad (3.6a)$$

$$(q, \nabla \cdot \mathbf{u}) = 0 \quad \forall q \in L^2(\Omega). \quad (3.6b)$$

Im Fall $\partial\Omega_{out} = \emptyset$ ist der Druck bei der Verwendung von $L^2(\Omega)$ als Testraum nur bis auf additive Konstanten bestimmt. Daher verwendet man im Fall reiner Dirichlet-Randbedingungen den Raum $L_0^2(\Omega) = \{ q \in L^2(\Omega) \mid (q, 1) = 0 \}$, um die Eindeutigkeit des Drucks zu garantieren.

In Bezug auf die Lösbarkeit von (3.6) verweisen wir auf die Diskussionen in [ESW05, Ran08, Qua09] und die darin enthaltene Literatur.

3.1.2 Ausflussrandbedingungen

In Gleichung (3.4) haben wir gesehen, dass die Randbedingung (2.39), welche sich im Fall der inkompressiblen Navier-Stokes-Gleichungen zu

$$Re^{-1} \nabla \mathbf{u} \cdot \mathbf{n} - p\mathbf{n}|_{\partial\Omega_{out}} = \mathbf{0} \quad (3.7)$$

vereinfacht, als natürliche Randbedingung in der schwachen Formulierung auftritt. Da auf $\partial\Omega_{out}$ keine weiteren Anforderungen an die Testfunktionen gestellt werden, wird sie auch als „do-nothing“-Randbedingung bezeichnet. Insbesondere erfüllt die Poiseuille-Strömung diese Bedingung in der Situation von Abschnitt 2.5, aber auch in allgemeineren Anwendungsfällen zeigt die Bedingung nur sehr wenig Einfluss auf das Strömungsfeld in der Nähe des Randes. Beispiele dazu finden sich etwa in [HRT96].

Wir wollen uns noch etwas genauer mit der Bedeutung dieser Randbedingung befassen und folgen der Diskussion in [Ran08]. Für den Rest dieses Abschnitts soll der Ausflussrand $\Gamma = \partial\Omega_{out}$ ein gerades Randstück sein und mit $\mathbf{n}$ und $\boldsymbol{\tau}$ bezeichnen wir die zugehörigen Normal- und Tangentialvektoren. Bezüglich dieses Koordinatensystems lautet (3.7) dann

$$Re^{-1} \partial_n u_n - p|_{\Gamma} = 0, \quad Re^{-1} \partial_n u_{\tau}|_{\Gamma} = 0$$

und die Divergenzfreiheit können wir durch

$$\partial_n u_n + \partial_{\tau} u_{\tau} = 0 \quad (3.8)$$

ausdrücken. Dabei bezeichnen wir mit $u_n = \mathbf{u} \cdot \mathbf{n}$ bzw. $u_{\tau} = \mathbf{u} \cdot \boldsymbol{\tau}$ die Normal- bzw. Tangentialkomponente der Geschwindigkeit, siehe auch [GS00]. Ganz allgemein betrachten wir nun die Bedingung

$$Re^{-1} \partial_n u_n - p|_{\Gamma} = -\bar{p} = \text{const.} \quad (3.9)$$

Die obige Randbedingung beinhaltet nun eine weitere Bedingung an den Mittelwert des Drucks entlang des Ausflussrands, denn mit (3.8) gilt

$$\frac{1}{|\Gamma|} \int_{\Gamma} p \, d\tau = \bar{p} + \frac{1}{|\Gamma|} \int_{\Gamma} Re^{-1} \partial_n u_n \, d\tau = \bar{p} - \frac{1}{|\Gamma|} \int_{\Gamma} Re^{-1} \partial_{\tau} u_{\tau} \, d\tau = \bar{p}. \quad (3.10)$$

Dabei haben wir insbesondere die Haftbedingungen von u_{τ} an den Endpunkten von Γ , also $u_{\tau}|_{\partial\Gamma} = 0$ und die Divergenzfreiheit (3.8) auf dem Rand ausgenutzt. Zur Klärung warum dies zulässig ist, verweisen wir auf [GS87].

Wir können also mit (3.9) den mittleren Druck am Ausflussrand vorgeben. Dies und das Vorgehen im Fall von Systemen mit mehreren Ausflusskomponenten, durch Vorgabe eines Druckabfalls, wird ausführlich in [HRT96] diskutiert. Aus Gründen der Einfachheit wählen wir im Folgenden stets $\bar{p} = 0$, andernfalls werden wir dies explizit anmerken.

3.1.3 Raumdiskretisierung

Wir stellen nun kurz die im Folgenden genutzten Finite-Elemente-Räume vor und geben anschließend eine diskrete Version der Variationsformulierung (3.6) an. Dazu sei $\mathcal{T}_h = \{K_1, \dots, K_N\}$ eine disjunkte Zerlegung von Ω durch offene Vierecke derart, dass $\bar{\Omega} = \bigcup_{K \in \mathcal{T}_h} \bar{K}$ gilt. Diese Vierecke werden wir im Weiteren auch als Zellen oder Elemente bezeichnen. Der Schnitt des Abschlusses zweier benachbarter Zellen soll dabei eine gemeinsame Kante oder ein gemeinsamer Knoten sein, wir lassen also keine hängenden Knoten zu. Für die Konvergenztheorie wird dann eine Familie von Triangulierungen $\{\mathcal{T}_h \mid h > 0\}$ betrachtet und der Fehler für $h \rightarrow 0$ untersucht. Der Parameter h steht dabei für den maximalen Durchmesser aller Elemente in $\mathcal{T}_h$, also $h = \max_{K \in \mathcal{T}_h} \text{diam}(K)$. Insbesondere wird davon ausgegangen, dass diese Familie gewissen Regularitätseigenschaften genügt, siehe [Cia02], so dass die Vierecke nicht im Laufe der Verfeinerung degenerieren.

Als zugehöriges Finite-Elemente Paar wählen wir das in [RT92] vorgestellte mittelwertorientierte nichtparametrische Rannacher-Turek-Element $\tilde{Q}_1/Q_0$. Dabei werden zur Approximation von Geschwindigkeit und Druck auf jedem Element $K \in \mathcal{T}_h$ Polynome aus $\tilde{Q}_1(K)$ bzw. $Q_0(K)$ verwendet. Mit $Q_0(K)$ wird dabei die Menge der konstanten Funktionen über K bezeichnet und wir setzen

$$\tilde{Q}_1(K) = \text{span}\{1, \xi, \eta, \xi^2 - \eta^2\}.$$

Die hier auftretenden Koordinaten (ξ, η) entstammen einem lokalem Koordinatensystem auf K , welches sich aus der Verbindung gegenüberliegender Seitenmittelpunkte ergibt. Die Konstruktion dieses Koordinatensystems wird in [Sch03] besprochen. Die zugehörigen Freiheitsgrade werden auf K durch die Knotenfunktionale

$$M_e^K(v) = \frac{1}{|e|} \int_e v|_K ds \quad (3.11)$$

bestimmt, wobei $e \subset \partial K$ eine Kante bezeichnet. Damit sind die zugehörigen Finite-Elemente-Räume durch

$$\mathbf{U}_h = \left\{ \mathbf{u}_h \in L^2(\Omega)^2 \mid \begin{array}{l} \forall K, K_1, K_2 \in \mathcal{T}_h : \mathbf{u}_h|_K \in \tilde{Q}_1(K)^2, M_e^{K_1}(\mathbf{u}_h) = M_e^{K_2}(\mathbf{u}_h) \\ \text{falls } e = K_1 \cap K_2, M_e^K(\mathbf{u}_h) = M_e^K(\mathbf{u}_D) \text{ falls } e \subset \Omega_D \end{array} \right\} \quad (3.12a)$$

$$\mathbf{V}_h = \left\{ \mathbf{v}_h \in L^2(\Omega)^2 \mid \begin{array}{l} \forall K, K_1, K_2 \in \mathcal{T}_h : \mathbf{v}_h|_K \in \tilde{Q}_1(K)^2, M_e^{K_1}(\mathbf{v}_h) = M_e^{K_2}(\mathbf{v}_h) \\ \text{falls } e = K_1 \cap K_2, M_e^K(\mathbf{v}_h) = 0 \text{ falls } e \subset \Omega_D \end{array} \right\} \quad (3.12b)$$

$$Q_h = \{ q_h \in L^2(\Omega) \mid q_h|_K = \text{const}, \forall K \in \mathcal{T}_h \} \quad (3.12c)$$

gegeben. Da Funktionen aus $\mathbf{V}_h$ im Allgemeinen nicht stetig auf Ω sind, gilt insbesondere $\mathbf{U}_h \not\subset \mathbf{U}$ und $\mathbf{V}_h \not\subset \mathbf{V}$. Es liegt also ein nichtkonformer Ansatz vor. Das Rannacher-Turek-Element wurde im FEM-Löser FEATFLOW implementiert und hat sich mehrfach bei der Lösung der inkompressiblen Navier-Stokes-Gleichungen bewährt. Wir verweisen dazu auf [Tur99, TH08] und die darin angegebene Literatur.

Nun können wir auch die diskrete Form der schwachen Formulierung (3.6) angeben: Finde für $t > 0$ ein Paar $(\mathbf{u}_h(t), p_h(t)) \in \mathbf{U}_h \times Q_h$ mit

$$(\partial_t \mathbf{u}_h + (\mathbf{u}_h \cdot \nabla) \mathbf{u}_h, \mathbf{v}_h)_h + Re^{-1} (\nabla \mathbf{u}_h, \nabla \mathbf{v}_h)_h - (p_h, \nabla \cdot \mathbf{v}_h)_h = (\mathbf{f}, \mathbf{v}_h)_h \quad \forall \mathbf{v}_h \in \mathbf{V}_h, \quad (3.13a)$$

$$(q_h, \nabla \cdot \mathbf{u}_h)_h = 0 \quad \forall q_h \in Q_h. \quad (3.13b)$$

Aufgrund der Nichtkonformität sind die auftretenden Bilinearformen zellweise zu verstehen, etwa $(\nabla \mathbf{u}_h, \nabla \mathbf{v}_h)_h = \sum_{K \in \mathcal{T}_h} (\nabla \mathbf{u}_h, \nabla \mathbf{v}_h)_{L^2(K)}$. Entsprechendes gilt auch für die daraus abgeleiteten Normen.

Für die Stokes-Gleichungen wurde in [RT92] die Fehlerabschätzung

$$\|\mathbf{u} - \mathbf{u}_h\| + h\|p - p_h\| \leq Ch^2 (\|\mathbf{u}\|_{H^2(\Omega)^2} + \|p\|_{H^1(\Omega)}) \quad (3.14)$$

bewiesen. Wir können also $L^2(\Omega)$ -Konvergenz erster Ordnung für den Druck und zweite Ordnung für die Geschwindigkeit erwarten. Für die Navier-Stokes-Gleichungen finden sich ähnliche Aussagen in [Sch97]. Entscheidend dafür ist, dass das Rannacher-Turek-Element die diskrete Babuška-Brezzi-Bedingung [GR86, BF91]

$$\inf_{0 \neq q_h \in Q_h} \sup_{0 \neq \mathbf{v}_h \in \mathbf{V}_h} \frac{(p_h, \nabla \cdot \mathbf{v}_h)_h}{\|q_h\|_{L^2(\Omega)} \|\nabla \mathbf{v}_h\|_{L^2(\Omega)}} \geq \gamma > 0$$

erfüllt, wobei die Konstante γ unabhängig von h ist. Große Reynolds-Zahlen verursachen Stabilitätsprobleme bei der obigen Diskretisierung, in diesem Fall sind zusätzliche Stabilisierungsterme in der Variationsformulierung notwendig. Da wir uns hier nur mit kleinen Reynolds-Zahlen befassen, sind derartige Stabilisierungstechniken hier nicht von Interesse und wir verweisen dazu auf [Tur99, Oua06].

Um tatsächlich diskrete Gleichungen in Form eines nichtlinearen Gleichungssystems zu erhalten, benötigen wir Basen der endlichdimensionalen Funktionenräume. Wir betrachten zunächst den Fall homogener Dirichlet-Randbedingungen, also $\mathbf{U}_h = \mathbf{V}_h$. Eine zu den einzelnen Komponenten zugehörige Basis $\{v_h^i \mid i = 1, \dots, M\}$ erhalten wir durch die Forderung $M_{e_j}(v_h^i) = \delta_{ij}$ für alle Kanten $\{e_j \mid e_j \not\subset \partial\Omega, j = 1, \dots, M\}$, die nicht zum Dirichlet-Rand gehören. Die daraus resultierende Produktbasis $\{\mathbf{v}_h^i \mid i = 1, \dots, 2M\} = \{(v_h^i, 0)^T \mid i = 1, \dots, M\} \cup \{(0, v_h^i)^T \mid i = 1, \dots, M\}$ ist dann eine Basis von $\mathbf{V}_h$. Weiter sei $\{q_h^i \mid i = 1, \dots, N\}$ die Basis von Q_h welche $q_h^i|_{K_j} = \delta_{ij}$ erfüllt. Setzen wir die Ansätze $p_h(\mathbf{x}, t) = \sum_{j=1}^N p_j(t) q_h^j(\mathbf{x})$ und $\mathbf{u}_h(\mathbf{x}, t) = \sum_{j=1}^{2M} \mathbf{u}_j(t) \mathbf{v}_h^j(\mathbf{x})$ in (3.13) ein, so erhalten wir das Gleichungssystem

$$\begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{u}} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{K}(\mathbf{u}) + \mathbf{L} & \mathbf{B} \\ -\mathbf{B}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{F} \\ \mathbf{0} \end{bmatrix} \quad (3.15)$$

mit den durch

$$\begin{aligned} \mathbf{M}_{ij} &= (\mathbf{v}_h^i, \mathbf{v}_h^j)_h, & \mathbf{K}(\mathbf{u})_{ij} &= (\mathbf{u}_h \cdot \nabla \mathbf{v}_h^j, \mathbf{v}_h^i)_h, & \mathbf{F}_i &= (\mathbf{f}, \mathbf{v}_h^i)_h, \\ \mathbf{L}_{ij} &= Re^{-1} (\nabla \mathbf{v}_h^j, \nabla \mathbf{v}_h^i)_h, & \mathbf{B}_i &= -(q_h^i, \nabla \cdot \mathbf{v}_h^i)_h, \end{aligned}$$

gegebenen Matrizen und Vektoren. Dabei sammeln wir in den Lösungsvektoren $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_{2M})$ und $\mathbf{p} = (p_1, \dots, p_N)$ die Koeffizienten aus den Basisdarstellungen. Wir bemerken noch, dass die Matrizen $\mathbf{M}$, $\mathbf{K}$ und $\mathbf{L}$ Blockdiagonalgestalt haben, da wir eine Produktbasis gewählt haben.

Im Fall inhomogener Dirichlet-Randbedingungen wird die Basis von $\mathbf{V}_h$ zu einer Basis ohne Vorgabe der Randdaten erweitert. Mit dieser Basis wird dann analog das Gleichungssystem aufgebaut. Die verglichen mit (3.15) überflüssigen Zeilen werden dann durch die Vorgabe der Randkoeffizienten ersetzt. Wir verweisen auf [Qua09] für eine Diskussion dieses Vorgehens.

Gilt $\partial\Omega_{out} = \emptyset$, so ist der Druck, wie zuvor im kontinuierlichen Fall, nur bis auf additive Konstanten bestimmt. Um dennoch einen eindeutigen Druck zu erhalten, ersetzen wir $L^2(\Omega)$ in der Definition (3.12c) durch $L_0^2(\Omega)$. Dies muss dann auch entsprechend bei der Erstellung des Gleichungssystems (3.15) berücksichtigt werden.

3.1.4 Zeitdiskretisierung

Abschließend beschäftigen wir uns mit der zeitlichen Diskretisierung der Gleichungen (3.15). Wir betrachten hier nur das implizite Euler-Verfahren. Für eine gewöhnliche Differentialgleichung $y'(t) = f(t, y(t))$ lautet es

$$y^{n+1} = y^n + \Delta t f(t^{n+1}, y^{n+1}),$$

wobei $\Delta t = t^{n+1} - t^n$ die konstante Zeitschrittweite darstellt und der hochgestellte Index n die Auswertung der entsprechenden Funktion zum Zeitpunkt t^n bezeichnet.

Anwendung des Verfahrens auf die semidiskreten Gleichungen (3.15) liefert

$$\begin{bmatrix} \mathbf{M} + \Delta t(\mathbf{K}(\mathbf{u}^{n+1}) + \mathbf{L}) & \Delta t \mathbf{B} \\ -\mathbf{B}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}^{n+1} \\ \mathbf{p}^{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix},$$

mit der rechten Seite

$$\mathbf{f} = \mathbf{M}\mathbf{u}^n + \Delta t \mathbf{F}^{n+1}. \quad (3.16)$$

Definieren wir noch

$$\mathbf{S}(\mathbf{u}^{n+1}) = \mathbf{M} + \Delta t(\mathbf{K}(\mathbf{u}^{n+1}) + \mathbf{L}) \quad (3.17)$$

so können wir das Gleichungssystem in kompakter Form als

$$\begin{bmatrix} \mathbf{S}(\mathbf{u}^{n+1}) & \Delta t \mathbf{B} \\ -\mathbf{B}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}^{n+1} \\ \mathbf{p}^{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}, \quad (3.18)$$

schreiben.

Der von uns verwendete Ansatz die Gleichungen zunächst im Raum und anschließend in der Zeit zu diskretisieren, wird auch als Linien-Methode bezeichnet. Dem gegenüber steht die Rothe-Methode, bei welcher die Raumdiskretisierung erst nach der Zeitdiskretisierung erfolgt. Bei den von uns betrachteten Diskretisierungen erhält man mit beiden Ansätzen am Ende das System (3.18). Wir haben die Präsentation als Linien-Methode

gewählt, weil sie häufiger Verwendung findet und verglichen mit der Rothe-Methode eine kompaktere Darstellung ermöglicht.

Bisher haben wir noch nicht die Umsetzung der Anfangsbedingung $\mathbf{u}|_{t=0} = \mathbf{u}^0$ besprochen. Dies geschieht mit der L^2 -Projektion $\mathcal{P} : L^2(\Omega)^2 \rightarrow \mathbf{U}_h$. Diese ist durch

$$(\mathcal{P}(\mathbf{u}^0), \mathbf{u}_h)_h = (\mathbf{u}^0, \mathbf{u}_h)_h \quad \forall \mathbf{u}_h \in \mathbf{U}_h \quad (3.19)$$

bestimmt und durch die Eigenschaft

$$\|\mathbf{u}^0 - \mathcal{P}(\mathbf{u}^0)\|_{L^2(\Omega)^2} \leq \|\mathbf{u}^0 - \mathbf{u}_h\|_{L^2(\Omega)^2} \quad \forall \mathbf{u}_h \in \mathbf{U}_h.$$

charakterisiert. Machen wir den Ansatz $\mathcal{P}(\mathbf{u}^0)(\mathbf{x}) = \sum_{j=1}^{2M} \mathbf{u}_j^0 \mathbf{v}_h^j(\mathbf{x})$ und setzen dies in (3.19) ein, so erhalten wir daraus das lineare System

$$\mathbf{M} \mathbf{u}^0 = \mathbf{r}$$

mit der Massenmatrix $\mathbf{M}$ aus (3.15) und der durch $\mathbf{r}_i = (\mathbf{u}^0, \mathbf{v}_h^i)_h$ gegebenen rechten Seite. Dieses System können wir nach $\mathbf{u}^0$ lösen und zur Berechnung von (3.16) nutzen.

3.2 Modellproblem

3.2.1 Schwache Formulierung

Ganz analog zum Vorgehen im Fall der inkompressiblen Navier-Stokes Gleichungen, können wir auch für das Modellproblem (2.37a), (2.37b) eine schwache Formulierung herleiten. Mit den Bezeichnungen aus Abschnitt 3.1.1 und $\mathbf{f} = \mathbf{f}_u$ lautet diese: Finde für $t > 0$ ein Paar $(\mathbf{u}(t), p(t)) \in \mathbf{U} \times L^2(\Omega)$ mit

$$(\varrho \partial_t \mathbf{u}, \mathbf{v}) + (\varrho(\mathbf{u} \cdot \nabla) \mathbf{u}, \mathbf{v}) + a(\mathbf{u}, \mathbf{v}) - (p, \nabla \cdot \mathbf{w}) = (\mathbf{f}, \mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{V}, \quad (3.20a)$$

$$(q, \nabla \cdot \mathbf{u}) = (g, q) \quad \forall q \in L^2(\Omega). \quad (3.20b)$$

Die dabei auftretende Bilinearform a ist durch

$$a(\mathbf{u}, \mathbf{v}) = Re^{-1} ((\eta \nabla \mathbf{u}, \nabla \mathbf{v}) + (\eta(\nabla \cdot \mathbf{u}), \nabla \cdot \mathbf{v}) - (\nabla \mathbf{u}^* \cdot \nabla \eta, \mathbf{v})) \quad (3.21)$$

gegeben, wobei $\mathbf{u}^*$ in (2.7) definiert wurde. Sie resultiert aus der partiellen Integration der Divergenz des Reibungsspannungstensors in der Form (2.8). Dabei ist zu beachten, dass wir nur die Terme mit Ableitungen zweiter Ordnung partiell integrieren. Wegen

$$-(\nabla \cdot (\eta \nabla \mathbf{u}), \mathbf{v}) = (\eta \nabla \mathbf{u}, \nabla \mathbf{v}) - \langle \eta \nabla \mathbf{u} \cdot \mathbf{n}, \mathbf{v} \rangle$$

und

$$-(\nabla(\eta(\nabla \cdot \mathbf{u})), \mathbf{v}) = (\eta(\nabla \cdot \mathbf{u}), \nabla \cdot \mathbf{v}) - \langle \eta(\nabla \cdot \mathbf{u}) \mathbf{n}, \mathbf{v} \rangle,$$

ergibt sich das zu (3.20) gehörige Randintegral als

$$\langle Re^{-1}(\eta \nabla \mathbf{u} \cdot \mathbf{n} + \eta(\nabla \cdot \mathbf{u}) \mathbf{n}) - p \mathbf{n}, \mathbf{v} \rangle_{\partial \Omega_{out}}.$$

Aufgrund der Randbedingung (2.39) entfällt es in der schwachen Formulierung (3.20). Für Spezialfälle dieser Gleichung existieren Aussagen zur Lösbarkeit, wir verweisen dazu wieder auf [Qua09] und die darin angegebene Literatur.

Die Verwendung von (2.8) innerhalb der schwachen Formulierung wird hier erstmalig vorgeschlagen und ermöglicht eine einfache Beschreibung der Ausflussrandbedingungen. Welche Vorteile die Verwendung der Bilinearform (3.21) innerhalb der schwachen Form gegenüber $(\eta(\nabla \mathbf{u} + \nabla \mathbf{u}^T), \nabla \mathbf{v})$ hat, besprechen wir im Abschnitt 3.2.4.

3.2.2 Raumdiskretisierung

Mit der schwachen Formulierung (3.20) erhalten wir auch sofort eine diskrete Variante. Finde für $t > 0$ ein Paar $(\mathbf{u}_h(t), p_h(t)) \in \mathbf{U}_h \times Q_h$ mit

$$(\varrho \partial_t \mathbf{u}_h, \mathbf{v}_h)_h + (\varrho (\mathbf{u}_h \cdot \nabla) \mathbf{u}_h, \mathbf{v}_h)_h + a_h(\mathbf{u}_h, \mathbf{v}_h) - (p_h, \nabla \cdot \mathbf{v}_h)_h = (\mathbf{f}, \mathbf{v}_h)_h \quad \forall \mathbf{v}_h \in \mathbf{V}_h, \quad (3.22a)$$

$$(q_h, \nabla \cdot \mathbf{u}_h)_h = (q_h, g)_h \quad \forall q_h \in Q_h. \quad (3.22b)$$

Dabei ist die Definition der diskreten Bilinearform a_h direkt aus (3.21) ersichtlich. Das zugehörige lineare Gleichungssystem lautet dann

$$\begin{bmatrix} \mathcal{M}(\varrho) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{u}} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathcal{K}(\varrho, \mathbf{u}) + \mathcal{L}(\eta) & \mathbf{B} \\ -\mathbf{B}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{F} \\ \mathbf{g} \end{bmatrix} \quad (3.23)$$

mit den Bezeichnungen von (3.15) und

$$\begin{aligned} \mathcal{M}(\varrho)_{ij} &= (\varrho \mathbf{v}_h^i, \mathbf{v}_h^j)_h, & \mathcal{K}(\varrho)(\mathbf{u})_{ij} &= (\varrho \mathbf{u}_h \cdot \nabla \mathbf{v}_h^j, \mathbf{v}_h^i)_h, \\ \mathcal{L}(\eta)_{ij} &= a_h(\mathbf{v}_h^j, \mathbf{v}_h^i) & \mathbf{g}_i &= (g, q_h^i)_h. \end{aligned}$$

Wir bemerken an dieser Stelle, dass sich (3.15) und (3.23) in drei wesentlichen Punkten unterscheiden. Zum einen ist das die Abhängigkeit der Massenmatrix $\mathcal{M}$ und der Transportmatrix $\mathcal{K}$ von der Dichte ϱ . Ebenso hat die variable Viskosität η nun Einfluss auf die Diffusionsmatrix $\mathcal{L}$ und schließlich liegt nun eine inhomogene Divergenzbedingung vor.

Außerdem ist $\mathcal{L}$ nicht mehr blockdiagonal, da durch die Verwendung des Reibungsspannungstensors beide Geschwindigkeitskomponenten in jeder Komponente der Impulsgleichung auftreten. Dies ist unabhängig von der Darstellung (2.8) und tritt auch mit der üblichen Formulierung der Reibungsspannungsdivergenz auf.

3.2.3 Zeitdiskretisierung

Die Anwendung des impliziten Euler-Verfahrens auf die semidiskreten Gleichungen (3.23) liefert nun

$$\begin{bmatrix} \mathcal{M}(\varrho) + \Delta t (\mathcal{K}(\varrho, \mathbf{u}) + \mathcal{L}(\eta)) & \Delta t \mathbf{B} \\ -\mathbf{B}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix}, \quad (3.24)$$

mit

$$\mathbf{f} = \mathcal{M}(\varrho)\mathbf{u}^n + \Delta t \mathbf{F}.$$

Um die Schreibweise möglichst übersichtlich zu halten, haben wir hier auf den Index $n+1$ verzichtet. Dies wollen wir auch im Weiteren so beibehalten, zeitabhängige Größen ohne Index werden also zum Zeitpunkt t^{n+1} betrachtet.

In kompakter Form können wir (3.24) auch als

$$\begin{bmatrix} \mathcal{S}(\varrho, \eta, \mathbf{u}) & \Delta t \mathbf{B} \\ -\mathbf{B}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix}, \quad (3.25)$$

schreiben, wobei wir hier

$$\mathcal{S}(\varrho, \eta, \mathbf{u}) = \mathcal{M}(\varrho) + \Delta t (\mathcal{K}(\varrho, \mathbf{u}) + \mathcal{L}(\eta)) \quad (3.26)$$

setzen. Das zu lösende System (3.25) stimmt in seiner Struktur mit dem inkompressiblen System (3.18) überein. Neu ist hier nur die Abhängigkeit von der Dichte und der Viskosität. Im nächsten Kapitel werden wir uns damit befassen wie wir derartige Systeme lösen können.

3.2.4 Ausflussrandbedingungen

Um den Vorteil der Darstellung (2.8) zu erkennen, gehen wir zunächst von der allgemeinen Darstellung des Reibungsspannungstensors (2.3), $\mathbf{T} = \eta(\nabla \mathbf{u} + \nabla \mathbf{u}^T)$, aus. Diese gilt auch schon bei der Betrachtung der inkompressiblen Navier-Stokes-Gleichungen und bringt dort die selben Probleme mit sich. Zur Bestimmung einer schwachen Formulierung von (2.37a) muss der viskose Term $-\nabla \cdot \mathbf{T}$ mit einer Testfunktion multipliziert und anschließend integriert werden. Dies führt zu

$$-(\nabla \cdot \mathbf{T}, \mathbf{v}) = (\mathbf{T}, \nabla \mathbf{v}) - \langle \mathbf{T} \cdot \mathbf{n}, \mathbf{v} \rangle_{\partial \Omega_{out}}$$

und insgesamt ergibt sich

$$Re^{-1} \eta (\nabla \mathbf{u} + \nabla \mathbf{u}^T) \cdot \mathbf{n} - p \mathbf{n} |_{\partial \Omega_{out}} = \mathbf{F} \quad (3.27)$$

als die zur schwachen Formulierung gehörige natürliche Randbedingung, wobei $\mathbf{F}$ entsprechend vorzugeben ist. Aber genau die Wahl von $\mathbf{F}$ bereitet Probleme.

Um dies einzusehen, schreiben wir (3.27), entsprechend der Situation in Abschnitt 3.1.2, in $(\mathbf{n}, \boldsymbol{\tau})$ -Koordinaten. Die Bedingungen lauten dann

$$2Re^{-1} \eta \partial_n u_n - p|_{\Gamma} = F_n, \quad Re^{-1} \eta (\partial_n u_{\tau} + \partial_{\tau} u_n) |_{\Gamma} = F_{\tau}.$$

Jetzt betrachten wir eine Poiseuille-Strömung. Für diese gilt $\partial_n u_n = \partial_n u_{\tau} = 0$, aber $\partial_{\tau} u_n \neq 0$. Damit ist also schon für diesen einfachen Fall a priori die Kenntnis des Geschwindigkeitsprofils von Nöten, um die Randbedingung sinnvoll formulieren zu können. Dies ist im Allgemeinen nicht möglich. Setzen wir mangels genauerer Informationen

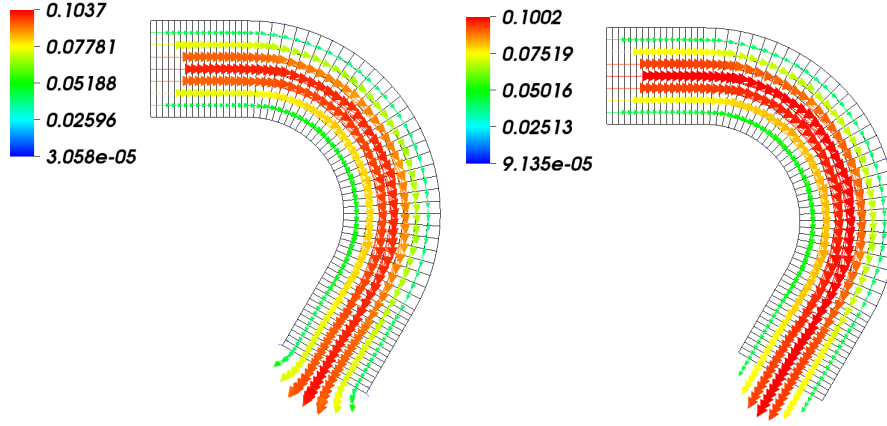


Abbildung 3.1: Unterschiedliches Verhalten am Ausflussrand aufgrund der Randbedingungen (3.27) (links) und (2.39) (rechts) mit $\mathbf{F} = \mathbf{0}$

dennoch $F_n = F_\tau = 0$, so führt dies zu einem nichtphysikalischem Verhalten am Ausflussrand, siehe Abbildung 3.1.

Um diese Problematik zu umgehen, wird in [BC07] ein Verfahren für verallgemeinert newtonsche Fluide vorgestellt, bei dem die Randbedingung durch einen zusätzlichen Penalty-Term in der Variationsformulierung vorgeschrieben werden. Dies bringt auch die Einführung eines neuen numerischen Parameters mit sich, welcher geeignet zu bestimmen ist.

Wir schlagen wie in Abschnitt 3.2.1 beschrieben einen anderen Weg vor und schreiben, motiviert durch die guten Ergebnisse mit der Bedingung (3.7), den viskosen Term in der Form (2.8). Mit der Herangehensweise des vorherigen Abschnitts ist die zugehörige Randbedingung dann durch

$$(Re^{-1}(\eta \nabla \mathbf{u} \cdot \mathbf{n} + \eta(\nabla \cdot \mathbf{u})\mathbf{n}) - p\mathbf{n})|_{\partial\Omega_{out}} = \mathbf{F}$$

gegeben. Dies entspricht im Fall einer divergenzfreien Strömung der Randbedingung (3.7), weshalb die Randbedingung (2.39) sinnvoll ist. Das Verhalten dieser Randbedingung ist ebenfalls in Abbildung 3.1 dargestellt und zeigt deutlich das für die Kanalströmung erwartete Verhalten.

Auch wenn die Lösung des Systems (3.25) erst in Kapitel 4 besprochen wird, bietet es sich an, direkt das Verhalten der beiden Ausflussrandbedingungen zu zeigen. Wie oben bereits erwähnt, ist dies in Abbildung 3.1 dargestellt. Der Testfall stammt aus [BC07] und wird hier zu Betrachtung eines inkompressiblen Fluids genutzt, wobei am Einfluss ein parabolisches Geschwindigkeitsprofil im Sinne der Poiseuille-Strömung vorgegeben wird. Es ist deutlich zu erkennen, wie sich das Geschwindigkeitsprofil am Austritt bei Verwendung der Randbedingung (3.27) ändert, während es mit der Randbedingung (2.39) unverändert bleibt.

4 Diskrete Projektionsverfahren

Zur Lösung der Systeme (3.18) und (3.25) gibt es generell zwei Herangehensweisen. Zum einen ist es möglich die Systeme direkt zu lösen ([Tur99, ESW05]), zum anderen können wir versuchen die Lösung mittels kleinerer Teilprobleme zu bestimmen. Wir gehen den zweiten Weg auf Grundlage des in [Tur99] für das System (3.18) vorgestellten diskreten Projektionsverfahrens. Dieses Verfahren ist in der Klasse der globalen MPSC-Verfahren enthalten, die Herleitung erfolgt über die Wahl optimaler Präkonditionierer für die Richardson-Iteration zur Lösung der Schur-Komplement-Gleichung für den Druck. Die Abkürzung MPSC steht dabei für Multilevel-Pressure-Schur-Complement. Da sich das so hergeleitete Verfahren nicht direkt auf das Modellproblem (3.25) übertragen lässt, werden wir zunächst eine neuartige Herleitung des Verfahrens angeben und diese dem Modellproblem entsprechend modifizieren.

4.1 Globales MPSC-Verfahren

In diesem Abschnitt präsentieren wir die Herleitung der in [Tur99] vorgestellten globalen MPSC-Verfahren. Der im Weiteren von uns genutzte Vertreter dieser Klasse von Verfahren kann als diskretes Projektionsverfahren interpretiert werden. Daher stellen wir in Abschnitt 4.2 die klassische Herleitung solcher Verfahren vor. Dies wird uns helfen, Teile des hier beschriebenen Verfahrens besser zu verstehen und eine alternative Herleitung für das Verfahren zu finden, welche wir später auf das Modellproblem übertragen können.

Nun zu den globalen MPSC-Verfahren. Das System (3.18) können wir komponentenweise als

$$\mathbf{S}(\mathbf{u})\mathbf{u} + \Delta t \mathbf{B}\mathbf{p} = \mathbf{f}, \quad (4.1)$$

$$\mathbf{B}^T \mathbf{u} = \mathbf{0} \quad (4.2)$$

schreiben. Nach formaler Multiplikation von (4.1) mit $\mathbf{B}^T \mathbf{S}(\mathbf{u})^{-1}$ erhalten wir wegen der diskreten Divergenzfreiheit (4.2) für den diskreten Druck das lineare System

$$\mathbf{B}^T \mathbf{S}(\mathbf{u})^{-1} \mathbf{B} \mathbf{p} = \frac{1}{\Delta t} \mathbf{B}^T \mathbf{S}(\mathbf{u})^{-1} \mathbf{f}. \quad (4.3)$$

Da die vorliegende Matrix $\mathbf{B}^T \mathbf{S}(\mathbf{u})^{-1} \mathbf{B}$ das Schur-Komplement der Systemmatrix aus (3.18) darstellt, wird diese Gleichung auch als Pressure-Schur-Complement-Gleichung bezeichnet. Zur Lösung dieses linearen Problems betrachten wir nun das linkspräkondi-

tionierte Richardson-Verfahren ohne Relaxationsparamter, siehe etwa [Mei99]. Die Iterationsvorschrift zur Lösung der Schur-Komplement-Gleichung (4.3) lautet dann

$$\begin{aligned}\mathbf{p}_{k+1} &= \mathbf{p}_k - \mathbf{C}^{-1} (\mathbf{B}^T \mathbf{S}(\mathbf{u})^{-1} \mathbf{B} \mathbf{p}_k - \Delta t^{-1} \mathbf{B}^T \mathbf{S}(\mathbf{u})^{-1} \mathbf{f}) \\ &= \mathbf{p}_k + \mathbf{C}^{-1} (\Delta t^{-1} \mathbf{B}^T \mathbf{S}(\mathbf{u})^{-1} (\mathbf{f} - \Delta t \mathbf{B} \mathbf{p}_k)),\end{aligned}\quad (4.4)$$

für $k = 0, 1, \dots, k_{max}$, wobei wir mit $\mathbf{C}$ den Präkonditionerer bezeichnen und $\mathbf{p}_0 = \mathbf{p}^n$ setzen. Nach Beendigung der Iteration setzen wir dann $\mathbf{p}^{n+1} = \mathbf{p}_{k_{max}}$.

Der nächste Schritt in der Verfahrensherleitung besteht nun aus der Wahl eines geeigneten Präkonditionerers. Der in [Tur99] vorgeschlagene globale Präkonditionerer hat die Form

$$\mathbf{C}^{-1} = \alpha_R \mathbf{P}^{-1} + \alpha_D \mathbf{M}_p^{-1}, \quad (4.5)$$

mit

$$\mathbf{P} = \mathbf{B}^T \mathbf{M}_l^{-1} \mathbf{B}. \quad (4.6)$$

Dabei stellt die Diagonalmatrix $\mathbf{M}_l$ eine Approximation an die Massenmatrix $\mathbf{M}$ dar (mass lumping). In unseren späteren Rechnungen erhalten wir die Diagonaleinträge von $\mathbf{M}_l$ als Zeilensummen von $\mathbf{M}$. Weiter ist die Matrix $\mathbf{M}_p$ die zu den Ansatzfunktionen für den Druck gehörige Massenmatrix und in unserem Fall stückweiser konstanter Ansatzfunktionen auf natürliche Weise von Diagonalgestalt. Außerdem sollen die Dämpfungsparamter dabei $\alpha_R \leq 1$ und $\alpha_D \leq \Delta t Re^{-1}$ erfüllen. Wir bemerken noch, dass die Matrix $\mathbf{P}$ als Schur-Komplement einer gemischten Diskretisierung eines Poisson-Problems aufgefasst werden kann, daher wird das Lösen eines Gleichungssystems mit dieser Matrix im Rahmen des weiteren Vorgehens auch als Pressure-Poisson-Problem bezeichnet.

Bei der Konstruktion des Präkonditionerers (4.5) wurde die explizite Form der Matrix

$$\mathbf{S}(\mathbf{u}^{n+1}) = \mathbf{M} + \Delta t \mathbf{K}(\mathbf{u}^{n+1}) + \Delta t \mathbf{L},$$

siehe (3.17), berücksichtigt. Ziel war es einen additiven Präkonditionerer der Form

$$\mathbf{C}^{-1} = \alpha_R \mathbf{A}_R^{-1} + \alpha_D \mathbf{A}_D^{-1} + \alpha_K \mathbf{A}_K^{-1} \quad (4.7)$$

zu konstruieren, dessen einzelne Summanden „optimale“ Präkonditionerer für die aus (3.17) resultierenden Grenzfälle dominanter Reaktion, Diffusion oder Konvektion sind. Also $\mathbf{A}_R$, $\mathbf{A}_D$ und $\mathbf{A}_K$ sind „optimal“ für $\mathbf{B}^T \mathbf{M}^{-1} \mathbf{B}$, $\mathbf{B}^T \mathbf{L}^{-1} \mathbf{B}$ und $\mathbf{B}^T \mathbf{K}^{-1} \mathbf{B}$. Dabei ist „Optimalität“ hier so zu verstehen, dass die Anwendung der Präkonditionerer auf die entsprechenden Teilprobleme (divergenzfreie L_2 -Projektion, Stokes Gleichungen, inkompressible Euler-Gleichungen) ein von äußeren Parametern und dem Gitter unabhängiges Konvergenzverhalten in Kontext eines Mehrgitterverfahrens liefert. Da in diesem Sinne kein konvektiver Präkonditionerer gefunden werden konnte, besteht der Präkonditionerer (4.5) nur aus zwei Summanden.

Der Name globales MPSC-Verfahren ergibt sich aus der Nutzung des globalen Präkonditionerers (4.5) bei der Lösung der Pressure-Schur-Complement-Gleichung (4.3). Die Optimalität des Präkonditionerers im Kontext eines Mehrgitterverfahrens erklärt das Multilevel in der Bezeichnung.

Mit der Wahl von (4.5) als Prädiktionierer und $k_{max} = 0$ können wir die Iteration (4.4) als diskretes Projektionsverfahren auffassen. Es besteht aus fünf wesentlichen Teilschritten:

1. Berechnung eines vorläufigen Geschwindigkeitsvektor $\tilde{\mathbf{u}}$ in Abhängigkeit der aktuellen Druckapproximation $\mathbf{p}^n$:

$$\mathbf{S}(\tilde{\mathbf{u}})\tilde{\mathbf{u}} = \mathbf{f} - \Delta t \mathbf{B} \mathbf{p}^n \quad (4.8)$$

2. Berechnung der rechten Seite $\mathbf{f}_{\mathbf{P}}$ für das diskrete Pressure-Poisson-Problem:

$$\mathbf{f}_{\mathbf{P}} = \frac{1}{\Delta t} \mathbf{B}^T \tilde{\mathbf{u}} \quad (4.9)$$

3. Lösung des diskreten Pressure-Poisson-Problems:

$$\mathbf{P} \mathbf{q} = \mathbf{f}_{\mathbf{P}} \quad (4.10)$$

4. Update des Drucks:

$$\mathbf{p}^{n+1} = \mathbf{p}^n + \alpha_R \mathbf{q} + \alpha_D \mathbf{M}_p^{-1} \mathbf{f}_{\mathbf{P}} \quad (4.11)$$

5. Update der Geschwindigkeit:

$$\mathbf{u}^{n+1} = \tilde{\mathbf{u}} - \Delta t \mathbf{M}_l^{-1} \mathbf{B} \mathbf{q} \quad (4.12)$$

Die aufwendigen Schritte sind dabei das Lösen der nichtlinearen Advektions-Diffusions-Gleichung in (4.8), sowie die Lösung des Pressure-Poisson Problems (4.10). Der Aufwand beim Lösen der Systeme hängt dabei natürlich von der Wahl des Löser bzw. der Linearisierung ab. Wir merken noch an, dass die zugehörige Matrix $\mathbf{P}$ über alle Zeitschritte konstant bleibt und daher bereits in einem Preprocessing-Schritt berechnet werden kann.

Für das Update der Geschwindigkeit nach (4.12) gilt aufgrund von (4.9) und (4.10) die Gleichheit

$$\begin{aligned} \mathbf{B}^T \mathbf{u}^{n+1} &= \mathbf{B}^T \tilde{\mathbf{u}} - \Delta t \mathbf{B}^T \mathbf{M}_l^{-1} \mathbf{B} \mathbf{q} \\ &= \Delta t \mathbf{f}_{\mathbf{P}} - \Delta t \mathbf{P} \mathbf{q} \\ &= \mathbf{0}, \end{aligned}$$

und wir erzwingen so die diskrete Divergenzfreiheit im nächsten Zeitschritt. Dies ist notwendig, da wir nur einen Schritt der Richardson-Iteration durchführen. Allerdings stellt sich die Frage in wie weit die so bestimmte Geschwindigkeit noch eine gute Approximation an die tatsächliche Lösung darstellt. Dies ist aus der obigen Herleitung nicht ersichtlich, wir beschäftigen uns damit in Abschnitt 4.2.

Zum Abschluss dieses Abschnitts merken wir noch an, dass in der durch (4.4) gegebenen Klasse von Verfahren, je nach Wahl des Prädiktionierers, der Anzahl der Iterationen, sowie der Berechnung der Geschwindigkeit, neben den oben angesprochenen Projektionsverfahren, noch weitere bekannte Verfahren enthalten sind. So führt die Wahl $\mathbf{C} = \sigma \mathbf{I}$ mit $\sigma \in \mathbb{R}$ zum Uzawa-Verfahren [Bra07] oder $\mathbf{C} = \mathbf{B}^T (\text{diag}(\mathbf{S}))^{-1} \mathbf{B}$ zum SIMPLE-Verfahren [HEH93].

4.2 Diskrete Projektionsverfahren für die inkompressiblen Gleichungen

Wie sich die Gleichung für das Geschwindigkeitsupdate (4.12) begründen lässt, konnten wir aus dem oben gemachten Ansatz über die Richardson-Iterationen für den Druck nicht ersehen. Der Schlüssel zum Verständnis des Updates liegt in der klassischen Herleitung von Projektionsverfahren. Damit wollen wir uns nun beschäftigen und folgen dabei der Darstellung aus [LMW02].

Nach einer Linearisierung können wir das System (3.18) in der Form

$$\mathbf{S}(\mathbf{u}^n)\mathbf{u} + \Delta t \mathbf{B}\mathbf{p} = \mathbf{f}, \quad (4.13a)$$

$$\mathbf{B}^T \mathbf{u} = \mathbf{0} \quad (4.13b)$$

schreiben. Eine erste Approximation der Lösung $\tilde{\mathbf{u}} \approx \mathbf{u}$ können wir durch Lösen von

$$\mathbf{S}(\mathbf{u}^n)\tilde{\mathbf{u}} = \mathbf{f} - \Delta t \beta \mathbf{B}\mathbf{p}^n \quad (4.14)$$

bestimmen, wobei $\beta \in \{0, 1\}$ einen konstanten Parameter bezeichnet. Dies entspricht (4.8). Da $\tilde{\mathbf{u}}$ im Allgemeinen nicht divergenzfrei sein wird, suchen wir nun eine Korrektur $\mathbf{v}$, so dass mit $\mathbf{u} = \tilde{\mathbf{u}} + \mathbf{v}$ ein divergenzfreies Geschwindigkeitsfeld vorliegt. Dazu subtrahieren wir (4.14) von (4.13a) und erhalten eine Gleichung für den Korrekturterm. Es gilt

$$\mathbf{S}(\mathbf{u}^n)\mathbf{v} + \Delta t \mathbf{B}\mathbf{q} = \mathbf{0}, \quad (4.15)$$

mit der Druckkorrektur

$$\mathbf{q} = \mathbf{p} - \beta \mathbf{p}^n.$$

Aus (4.13b) ergibt sich für die Divergenz der Geschwindigkeitskorrektur

$$\mathbf{B}^T \mathbf{v} = -\mathbf{B}^T \tilde{\mathbf{u}},$$

womit wir aus (4.15) das Pressure-Poisson-Problem

$$\mathbf{B}^T \mathbf{S}^{-1} \mathbf{B} \mathbf{q} = \frac{1}{\Delta t} \mathbf{B}^T \tilde{\mathbf{u}} \quad (4.16)$$

erhalten. Um dieses System effizient lösen zu können, ersetzen wir $\mathbf{S}$ durch eine leicht zu invertierende Approximation $\tilde{\mathbf{S}}$. Das nun resultierende System entspricht (4.10). Ist die Gleichung gelöst, ergibt sich der neue Druck als

$$\mathbf{p} = \beta \mathbf{p}^n + \mathbf{q}, \quad (4.17)$$

und wir können wegen (4.15) die neue Geschwindigkeit durch

$$\mathbf{u} = \tilde{\mathbf{u}} + \mathbf{v} = \tilde{\mathbf{u}} - \Delta t \tilde{\mathbf{S}}^{-1} \mathbf{B} \mathbf{q} \quad (4.18)$$

ermitteln.

Damit haben wir das System (4.13) gelöst und insbesondere das Geschwindigkeitsupdate (4.12) hergeleitet. Es entspricht dem Update (4.18) bei Verwendung der Approximation $\tilde{\mathbf{S}} = \mathbf{M}_l$. Der wesentliche Unterschied beider Verfahren liegt im Update des Drucks. Die Gleichungen (4.11) und (4.17) unterscheiden sich bezüglich der zusätzlichen Anwendung des diffusiven Präkonditionierers $\Delta t^{-1} \mathbf{M}_p^{-1} \mathbf{B}^T \tilde{\mathbf{u}}$. In [Tur99] wurde anhand extensiver numerischer Experimente gezeigt, dass dieser zusätzliche Term im Druck-Update deutliche Vorteile gegenüber (4.17) hat. Ebenso führt dieser zusätzliche Term zu Verfahren mit verbesserter Ordnung, siehe [GMS06].

Abschließend bemerken wir noch, dass das obige Verfahren mit $\beta = 0$ dem Chorin-Verfahren [Cho68] entspricht. Im Fall $\beta = 1$ erhalten wir eine Variante des Van Kan-Verfahrens [Kan86], bei dem klassisch das Crank-Nicolson Verfahren zur Zeitdiskretisierung genutzt wird. Damit sind die Verfahren auch in der Klasse globaler MPSC-Verfahren enthalten. Analog zur obigen Herleitung hätten wir die Projektionsverfahren auch im kontinuierlichen Fall herleiten können. Für eine tiefgehende Diskussion (kontinuierlicher) Projektionsverfahren, verweisen wir an dieser Stelle auf [Pro97, GS00, GMS06] und die enthaltene Literatur.

4.3 Diskretes nichtlineares Projektionsverfahren für das Modellproblem

Nun befassen wir uns mit der Lösung des diskreten Modellproblems (3.25). Ziel ist es das Projektionsverfahren (4.8)-(4.12) so zu modifizieren, dass wir es auf das Modellproblem anwenden können. Wir werden sehen, dass die Herleitung als globales MPSC-Verfahren Schwierigkeiten mit sich bringt. Um diese zu umgehen, werden wir hier erstmalig eine Herleitung des Projektionsverfahrens angeben, welche dem Ansatz aus Abschnitt 4.2 basiert, aber den diffusiven Präkonditionierer für das Druckupdate beinhaltet. Diese neue Herleitung werden wir dann auf das Modellproblem übertragen.

Zunächst betrachten wir den Zugang über globale MPSC-Verfahren. Dabei können wir prinzipiell so vorgehen, wie im Abschnitt 4.1. Allerdings müssen wir hier die durch (2.25) gegebene Druckabhängigkeit der Geschwindigkeitsdivergenz im Fall der Kavitation berücksichtigen. Gleichung (3.25) lautet komponentenweise

$$\begin{aligned} \mathbf{S}\mathbf{u} + \Delta t \mathbf{B}\mathbf{p} &= \mathbf{f}, \\ -\mathbf{B}^T \mathbf{u} &= \mathbf{g}(\mathbf{p}), \end{aligned}$$

und formale Elimination der Geschwindigkeit führt hier auf die nichtlineare Druck-Schur-Komplement-Gleichung

$$\mathbf{B}^T \mathbf{S}^{-1} \mathbf{B} \mathbf{p} = \frac{1}{\Delta t} (\mathbf{B}^T \mathbf{S}^{-1} \mathbf{f} + \mathbf{g}(\mathbf{p})).$$

Die zugehörige mit $\mathbf{C}^{-1}$ präkonditionierte Richardson-Iteration können wir dann als

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{C}^{-1} (\Delta t^{-1} \mathbf{B}^T \mathbf{S}^{-1} (\mathbf{f} - \Delta t \mathbf{B} \mathbf{p}_k) + \Delta t^{-1} \mathbf{g}(\mathbf{p}_k))$$

schreiben.

Nun stellt sich die Frage wie wir die Präkonditionierer wählen sollten. Während wir den reaktiven Präkonditionierer analog zu (4.6) bilden können, ist die entsprechende Wahl des diffusiven Präkonditionierers wesentlich schwieriger. Derzeit ist kein derartiger, im obigen Sinne optimaler, Präkonditionierer bekannt und wir verweisen auf die Diskussion in [TH08].

Wir gehen hier deshalb einen anderen Weg und leiten eine neue Variante des Verfahrens (4.8)-(4.12) her, welche sich auf natürliche Weise auf unser Modellproblem übertragen lässt. Ziel ist es, den diffusiven Präkonditionierer in den Ansatz von Abschnitt 4.2 zu integrieren, wir betrachten jetzt also wieder (3.18). Dabei erinnern wir uns daran, dass wir bei der Herleitung der inkompressiblen Navier-Stokes-Gleichungen (2.5) den sich aus der Divergenz des Reibungsspannungstensor ergebenden Term $\nabla(\nabla \cdot \mathbf{u})$, aufgrund der Divergenzfreiheit nicht weiter berücksichtigt haben. Es ist genau dieser Term in diskreter Form, der den diffusiven Präkonditionierer in das Verfahren einbringt. Unter Beibehaltung dieses Terms können wir die Gleichungen auch in der Form

$$\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} - Re^{-1} \Delta \mathbf{u} + \nabla p - \nabla r = \mathbf{f}, \quad (4.19)$$

$$Re^{-1} \nabla \cdot \mathbf{u} = r, \quad (4.20)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (4.21)$$

schreiben. Nach einer Diskretisierung in Raum und Zeit wie zuvor, erhalten wir damit das System

$$\mathbf{S}\mathbf{u} + \Delta t \mathbf{B}\mathbf{p} - \Delta t \mathbf{B}\mathbf{r} = \mathbf{f}, \quad (4.22)$$

$$-Re^{-1} \mathbf{B}^T \mathbf{u} = \mathbf{M}_p \mathbf{r}, \quad (4.23)$$

$$\mathbf{B}^T \mathbf{u} = \mathbf{0}. \quad (4.24)$$

Wichtig ist hier, dass wir zur Diskretisierung der neuen Gleichung (4.20) die zum Druck gehörigen Testfunktionen wählen, ebenso erfolgt der Ansatz für r über die Funktionen aus Q_h . Abschließend fassen wir (4.22) und (4.23) wieder zusammen und erhalten

$$\mathbf{S}(\mathbf{u})\mathbf{u} + \Delta t \mathbf{B}\mathbf{p} + \Delta t Re^{-1} \mathbf{B}\mathbf{M}_p^{-1} \mathbf{B}^T \mathbf{u} = \mathbf{f}, \quad (4.25)$$

$$\mathbf{B}^T \mathbf{u} = \mathbf{0}, \quad (4.26)$$

als zu lösendes System.

Nun folgen wir formal dem klassischen Vorgehen bei Projektionsverfahren aus Abschnitt 4.2. Zunächst bestimmen wir durch

$$\mathbf{S}(\tilde{\mathbf{u}})\tilde{\mathbf{u}} + \Delta t \mathbf{B}\mathbf{p}^n + \Delta t Re^{-1} \mathbf{B}\mathbf{M}_p^{-1} \mathbf{B}^T \tilde{\mathbf{u}} = \mathbf{f} \quad (4.27)$$

eine Approximation an die Geschwindigkeit. Aus (4.25) und (4.27) bestimmen wir dann eine Gleichung für die Korrektur $\mathbf{v} = \mathbf{u} - \tilde{\mathbf{u}}$. Wir erhalten

$$\begin{aligned} \tilde{\mathbf{S}}(\mathbf{v})\mathbf{v} + \Delta t \mathbf{B}\mathbf{q} &= \mathbf{0}, \\ \mathbf{B}^T \mathbf{v} &= -\mathbf{B}^T \tilde{\mathbf{u}}, \end{aligned}$$

mit

$$\mathbf{q} = \mathbf{p} - \mathbf{p}^n - Re^{-1} \mathbf{M}_p^{-1} \mathbf{B}^T \tilde{\mathbf{u}}$$

und einer geeigneten Approximation $\tilde{\mathbf{S}}(\mathbf{u}^c) \mathbf{u}^c \approx \mathbf{S}(\mathbf{u}) \mathbf{u} - \mathbf{S}(\tilde{\mathbf{u}}) \tilde{\mathbf{u}}$. Da wir das Verfahren (4.8)-(4.12) herleiten wollen, setzen wir ab jetzt $\tilde{\mathbf{S}} = \mathbf{M}_l$. Schließlich ergibt sich analog zu (4.10) nach Elimination der Geschwindigkeitskorrektur die Pressure-Poisson-Gleichung

$$\mathbf{B}^T \mathbf{M}_l^{-1} \mathbf{B} \mathbf{q} = \frac{1}{\Delta t} \mathbf{B}^T \tilde{\mathbf{u}},$$

nach deren Lösung wir Druck und Geschwindigkeit durch

$$\mathbf{p} = \mathbf{p}^n + \mathbf{q} + Re^{-1} \mathbf{M}_p^{-1} \mathbf{B}^T \tilde{\mathbf{u}}, \quad (4.28)$$

$$\mathbf{u} = \tilde{\mathbf{u}} - \Delta t \mathbf{M}_l^{-1} \mathbf{B} \mathbf{q}, \quad (4.29)$$

bestimmen können. Insbesondere ist nun der diffusive Präkonditionierer aus (4.5) im Druckupdate (4.28) enthalten. Der einzige Unterschied zwischen diesem Vorgehen und dem durch (4.8)-(4.12) gegebenen Verfahren, liegt in der Berechnung der Geschwindigkeitsapproximation. In (4.27) wird zusätzlich noch die Divergenz berücksichtigt. Gehen wir wie (4.8) vor, so gilt lediglich

$$\mathbf{S} \tilde{\mathbf{u}} = \mathbf{f} - \Delta t \mathbf{B} \mathbf{p}^n \approx \mathbf{f} - \Delta t \mathbf{B} \mathbf{p}^n - \Delta t Re^{-1} \mathbf{B} \mathbf{M}_p^{-1} \mathbf{B}^T \tilde{\mathbf{u}}.$$

Mit dieser Änderung stimmen dann beide Verfahren exakt überein.

Diese neue Betrachtungsweise können wir nun auch auf das Modellproblem (3.25) übertragen. Dazu nutzen wir die Zerlegung der Divergenz des Reibungsspannungstensors (2.8). Die Gleichungen (2.37a) und (2.37b) lauten damit

$$\varrho (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) - Re^{-1} (\nabla \cdot (\eta \nabla \mathbf{u}) + \nabla r + \nabla \mathbf{u}^* \cdot \nabla \eta) + \nabla p = \mathbf{f}_u, \quad (4.30a)$$

$$\nabla \cdot \mathbf{u} = \eta^{-1} r, \quad (4.30b)$$

$$\nabla \cdot \mathbf{u} = g(p). \quad (4.30c)$$

Nach einer Diskretisierung wie oben erhalten wir daraus das System

$$\mathcal{R} \mathbf{u} + \Delta t \mathbf{B} \mathbf{p} - \Delta t Re^{-1} \mathbf{B} \mathbf{r} = \mathbf{f}, \quad (4.31a)$$

$$-\mathbf{B}^T \mathbf{u} = \mathcal{M}_p \mathbf{r}, \quad (4.31b)$$

$$-\mathbf{B}^T \mathbf{u} = \mathbf{g}(\mathbf{p}), \quad (4.31c)$$

wobei die Matrix $\mathcal{R}$ alle Terme der linken Seite von (4.30a) außer ∇p und ∇r enthält. Damit unterscheidet sie sich von der Matrix $\mathcal{S}$ aus (3.26) die auch den ∇r Term enthält. Die auftretende gewichtete Massenmatrix $\mathcal{M}_p$ ist durch

$$(\mathcal{M}_p)_{ij} = (\eta^{-1} q_h^j, q_h^i),$$

mit $q_h \in Q_h$ gegeben. Aufgrund der Definition von Q_h (3.12c) liegt mit $\mathcal{M}_p$ wieder eine Diagonalmatrix vor. Fassen wir die Gleichungen (4.31a) und (4.31b) wieder zusammen, so erhalten wir schließlich das System

$$\mathcal{R} \mathbf{u} + \Delta t \mathbf{B} \mathbf{p} + \Delta t Re^{-1} \mathbf{B} \mathcal{M}_p^{-1} \mathbf{B}^T \mathbf{u} = \mathbf{f}, \quad (4.32a)$$

$$-\mathbf{B}^T \mathbf{u} = \mathbf{g}(\mathbf{p}). \quad (4.32b)$$

4 Diskrete Projektionsverfahren

Dies wollen wir nun wie oben lösen, wobei die Druckabhängigkeit der Divergenzbedingung auf eine nichtlineare Gleichung zur Bestimmung des Drucks führen wird.

Zunächst erhalten wir durch

$$\mathcal{R}\tilde{\mathbf{u}} + \Delta t \mathbf{B} \mathbf{p}^n + \Delta t Re^{-1} \mathbf{B} \mathcal{M}_p^{-1} \mathbf{B}^T \tilde{\mathbf{u}} = \mathbf{f} \quad (4.33)$$

eine erste Approximation $\tilde{\mathbf{u}}$ an das Geschwindigkeitsfeld. Für die zugehörige Korrektur $\mathbf{v} = \mathbf{u} - \tilde{\mathbf{u}}$ gilt dann analog zu obigem Vorgehen nach Subtraktion von (4.33) von (4.32a)

$$\begin{aligned} \mathcal{M}_l \mathbf{v} + \Delta t \mathbf{B} \mathbf{q} &= \mathbf{0}, \\ \mathbf{B}^T \mathbf{v} &= -\mathbf{g}(\mathbf{p}) - \mathbf{B}^T \tilde{\mathbf{u}}. \end{aligned} \quad (4.34)$$

Dabei ergibt sich die gelumppte Massenmatrix $\mathcal{M}_l$ aus der gewichteten Massenmatrix

$$\mathcal{M}_{ij} = (\varrho \mathbf{v}_h^j, \mathbf{v}_h^i)_h$$

und der Vektor $\mathbf{q}$ ist hierbei durch

$$\mathbf{q}(\mathbf{p}) = \mathbf{p} - \mathbf{p}^n - Re^{-1} \mathcal{M}_p^{-1} (\mathbf{g}(\mathbf{p}) + \mathbf{B}^T \tilde{\mathbf{u}})$$

gegeben. Setzen wir dann

$$\mathcal{P} = \mathbf{B}^T \mathcal{M}_l^{-1} \mathbf{B},$$

so erhalten wir aus (4.34) das bezüglich $\mathbf{p}$ nichtlineare Pressure-Poisson-Problem

$$\mathcal{P} \mathbf{q}(\mathbf{p}) = \frac{1}{\Delta t} (\mathbf{B}^T \tilde{\mathbf{u}} + \mathbf{g}(\mathbf{p})). \quad (4.35)$$

Auch wenn wir diese Gleichung direkt für den Druck lösen, müssen wir dennoch $\mathbf{q}$ bestimmen, um die Geschwindigkeit durch

$$\mathbf{u} = \tilde{\mathbf{u}} - \Delta t \mathcal{M}_l^{-1} \mathbf{B} \mathbf{q},$$

zu ermitteln.

Während variable Dichte und Viskosität lediglich zu entsprechend gewichteten Massenmatrizen führen, entsteht der Hauptunterschied verglichen mit den inkompressiblen Navier-Stokes Gleichung in der vom Druck abhängigen Divergenzbedingung. Dies erfordert bei unserer neuen Betrachtungsweise die Lösung des nichtlinearen Problems (4.35) zur Bestimmung des Drucks. Dabei muss die Matrix $\mathcal{P}$ nun aufgrund der Abhängigkeit von der Dichte jetzt in jedem Zeitschritt neu berechnet werden. Ferner bemerken wir noch, dass wir sie der gemischten Diskretisierung des Diffusionsproblems

$$\nabla \cdot (\varrho^{-1} \nabla q) = rhs,$$

welches sich auch im Fall der Anwendung klassischer Projektionstechniken auf das Modellproblem ergibt, zuordnen können.

Bei der Berechnung der Geschwindigkeitsapproximation $\tilde{\mathbf{u}}$ in (4.33) wird die Matrix $\mathcal{R} + \Delta t Re^{-1} \mathbf{B} \mathcal{M}_p^{-1} \mathbf{B}^T$ verwendet. Diese unterscheidet sich in ihrer Wirkung nicht von

$\mathcal{S}$ aus (3.26), beide Operatoren gehen auf unterschiedliche Diskretisierungen des selben Problems zurück. Daher liegt es nahe in (4.33) die Matrix $\mathcal{S}$ anstelle von $\mathcal{R} + \Delta t Re^{-1} \mathbf{B} \mathcal{M}_p^{-1} \mathbf{B}^T$ zu nutzen, da diese sich effizienter berechnen lässt. Dazu merken wir noch an, dass in numerische Experimenten keinen bedeutenden Unterschied zwischen den beiden Varianten ausgemacht werden konnten.

Der Vollständigkeit und Übersichtlichkeit halber wollen wir das Verfahren hier noch kurz in kompakter Form zusammenfassen, sowie das entsprechende iterative Verfahren angeben.

1. Berechnung eines vorläufigen Geschwindigkeitsvektors $\tilde{\mathbf{u}}$ in Abhängigkeit der aktuellen Druckapproximation $\mathbf{p}^n$:

$$\mathcal{S}(\tilde{\mathbf{u}})\tilde{\mathbf{u}} = \mathbf{f} - \Delta t \mathbf{B} \mathbf{p}^n \quad (4.36)$$

oder

$$\mathcal{R}(\tilde{\mathbf{u}})\tilde{\mathbf{u}} + \Delta t Re^{-1} \mathbf{B} \mathcal{M}_p^{-1} \mathbf{B} \tilde{\mathbf{u}} = \mathbf{f} - \Delta t \mathbf{B} \mathbf{p}^n \quad (4.37)$$

2. Lösung des diskreten nichtlinearen Pressure-Poisson Problems für den Druck:

$$\mathcal{P} \mathbf{q}(\mathbf{p}) = \mathbf{f}_{\mathcal{P}}(\mathbf{p}) \quad (4.38)$$

mit

$$\mathbf{q}(\mathbf{p}) = \mathbf{p} - \mathbf{p}^n - \Delta t Re^{-1} \mathcal{M}_p^{-1} \mathbf{f}_{\mathcal{P}}(\mathbf{p}), \quad \mathbf{f}_{\mathcal{P}}(\mathbf{p}) = \frac{1}{\Delta t} (\mathbf{B}^T \tilde{\mathbf{u}} + \mathbf{g}(\mathbf{p}))$$

3. Berechnung des Hilfsvektors für das Geschwindigkeitsupdate:

$$\mathbf{q} = \mathbf{p} - \mathbf{p}^n - Re^{-1} \mathcal{M}_p^{-1} (\mathbf{g}(\mathbf{p}) + \mathbf{B}^T \tilde{\mathbf{u}}) \quad (4.39)$$

4. Update der Geschwindigkeit:

$$\mathbf{u} = \tilde{\mathbf{u}} - \Delta t \mathcal{M}_l(\varrho)^{-1} \mathbf{B} \mathbf{q} \quad (4.40)$$

Für die Darstellung des Verfahrens als Iterationsverfahren führen wir noch den nichtlinearen Operator

$$\mathcal{T}(\mathbf{v})\mathbf{u} = \begin{cases} \mathcal{S}(\mathbf{v})\mathbf{u}, & \text{nach (4.36),} \\ (\mathcal{R}(\mathbf{v}) + \Delta t Re^{-1} \mathbf{B} \mathcal{M}_p^{-1} \mathbf{B}^T)\mathbf{u}, & \text{nach (4.37)} \end{cases} \quad (4.41)$$

ein. Damit können wir das Verfahren in der Form

$$\mathbf{p}_{k+1} = \mathbf{p}_k + (\mathcal{P}^{-1} + \Delta t Re^{-1} \mathcal{M}_p^{-1}) (\Delta t^{-1} \mathbf{B}^T \mathcal{T}^{-1} (\mathbf{f} - \Delta t \mathbf{B} \mathbf{p}_k) + \Delta t^{-1} \mathbf{g}(\mathbf{p}_{k+1}))$$

schreiben. Hierbei wird die Abhängigkeit der Geschwindigkeitsdivergenz vom Druck $\mathbf{g}(\mathbf{p})$ implizit behandelt. Numerische Tests mit einer expliziten Vorgehensweise zeigten keine Konvergenz.

Das obige Verfahren beinhaltet die Lösung zweier nichtlinearer Teilprobleme. Zunächst ergibt sich die Geschwindigkeitsapproximation $\tilde{\mathbf{u}}$ als Lösung der nichtlinearen Gleichung (4.36) oder (4.37). Die anschließende Bestimmung des Drucks erfordert das Lösen des nichtlinearen Problems (4.38). In Kapitel 5 beschäftigen wir uns mit der Lösung dieser nichtlinearen Probleme.

Anscheinend wurde ein derartiges nichtlineares Projektionsverfahren bisher nur einmal in [Sch09] vorgestellt. Dort wird es in Verbindung mit einer Finite-Volumen-Diskretisierung zur Lösung eines Granular Flow Modells genutzt. Die Grundidee ist die selbe wie bei dem hier vorgestellten Verfahren, allerdings ist ein direkter Vergleich aufgrund der unterschiedlichen Voraussetzungen kaum möglich.

Zum Ende dieses Abschnitts wollen wir anhand eines Testproblems die Konvergenz und auch die räumliche Ordnung des obigen Verfahrens zeigen. Dazu betrachten wir auf $\Omega = [0, 1]^2$ das unphysikalische Testproblem

$$\begin{aligned} \varrho (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) - \nabla \cdot (2\eta \mathbf{D}) + \nabla p &= \mathbf{f}, \\ \nabla \cdot \mathbf{u} &= g, \end{aligned} \tag{4.42}$$

mit symbolischer Dichte und Viskosität

$$\varrho = \sin(x) \sin(y), \quad \eta = \cos(x) \cos(y),$$

und den rechten Seiten

$$\begin{aligned} g &= 2 \cos(x) \cos(y), \\ f_u &= -\frac{1}{2} (2 \cos(2x) \sin(2y) + \cos(x) \sin^2(x) (\sin(y) - \sin(3y))), \\ f_v &= -\frac{1}{2} (2 \cos(2y) \sin(2x) + \cos(y) \sin^2(y) (\sin(x) - \sin(3x))). \end{aligned}$$

Der Testfall berücksichtigt variable Dichte und Viskosität, sowie die inhomogene Divergenzbedingung. Die stationäre Lösung dieses Problems ist durch

$$\begin{aligned} u &= \sin(x) \cos(y), \quad v = \cos(x) \sin(y), \\ p &= \frac{1}{2} (\cos(2y) + \cos(2x) (1 + 2 \cos(2y)) - \sin(2x) \sin(2y)) + \text{const}, \end{aligned}$$

gegeben. Wählen wir die Konstante für den Druck als $\text{const} = \frac{1}{16} (1 - 4 \cos(2) + 3 \cos(4) - 8 \sin(2))$ so sichert dies gerade, dass der Druck im Mittel verschwindet.

Zur Diskretisierung nutzen wir quadratische Gitter mit Maschenweite $\Delta x = 2^{-(l+1)}$ für $l = 0, \dots, 4$. Am Gebietsrand geben wir Dirichlet-Randbedingungen vor. Die Fehler und die daraus resultierende Konvergenzordnung sind in Tabelle 4.1 aufgelistet. Wir erkennen deutlich die nach (3.14) zu erwartende Ordnung des Verfahrens.

Tabelle 4.1: Fehler und Konvergenzordnung für das Testproblem (4.42)

Gitter l	$\ u - u_h\ _{L^2(\Omega)}$		$\ v - v_h\ _{L^2(\Omega)}$		$\ p - p_h\ _{L^2(\Omega)}$	
	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung
0	0.006381	-	0.006381	-	0.187135	-
1	0.001626	1.972592	0.001626	1.972592	0.093450	1.001817
2	0.000409	1.992388	0.000409	1.992388	0.046687	1.001156
3	0.000102	1.997792	0.000102	1.997792	0.023336	1.000465
4	0.000026	1.999332	0.000026	1.999332	0.011667	1.000159

5 Lösung linearer und nichtlinearer Gleichungssysteme

In diesem Kapitel beschreiben wir die Lösung der nichtlinearen Teilprobleme für die Geschwindigkeit (4.36) bzw. (4.37) und für den Druck (4.38). Für die Geschwindigkeit nutzen wir das in [Tur99] beschriebene Fixpunkt-Defekt-Korrektur-Verfahren, den Druck erhalten wir durch ein global konvergentes Newton-Verfahren. Die dabei auftretenden linearen Teilprobleme werden wir mit dem BiCGSTAB-Verfahren lösen.

5.1 Fixpunkt-Defekt-Korrektur

Mit dem nichtlinearen Operator (4.41) können wir (4.36) bzw. (4.37) in der Form

$$\mathcal{T}(\mathbf{u})\mathbf{u} = \mathbf{f},$$

schreiben. Um dieses System zu lösen, verwenden wir das in [Tur99] vorgestellte nicht-lineare Iterationsverfahren

$$\mathbf{u}_{k+1} = \mathbf{u}_k - \tilde{\mathcal{T}}(\mathbf{u}_k)^{-1}(\mathcal{T}(\mathbf{u}_k)\mathbf{u}_k - \mathbf{f}). \quad (5.1)$$

Hierbei bezeichnet $\tilde{\mathcal{T}}$ eine Approximation an die Jacobi-Matrix zu $\mathcal{T}(\mathbf{u}_k)\mathbf{u}_k - \mathbf{f}$. Die Wahl der exakten Jacobi-Matrix liefert dabei das Newton-Verfahren und mit $\tilde{\mathcal{T}}(\mathbf{u}_k) = \mathcal{T}(\mathbf{u}_k)$ ergibt sich eine Fixpunkt-Iteration. Wir nutzen im Folgenden eine Variante der Fixpunkt-Iteration, bei welcher wir $\mathcal{T}(\mathbf{u}_k)$ durch die zugehörigen Diagonalblöcke approximieren.

Die Iteration (5.1) können wir in drei Teilschritte unterteilen.

1. Berechnung des Defekts:

$$\mathbf{d}_k = \mathcal{T}(\mathbf{u}_k)\mathbf{u}_k - \mathbf{f}$$

2. Lösen des linearen Teilproblems:

$$\tilde{\mathcal{T}}(\mathbf{u}_k)\mathbf{y}_k = \mathbf{d}_k \quad (5.2)$$

3. Update:

$$\mathbf{u}_{k+1} = \mathbf{u}_k - \mathbf{y}_k$$

Der wesentliche Aufwand des Verfahrens liegt dabei in der Behandlung des linearen Teilproblems (5.2). Da $\tilde{\mathcal{T}}$ in unserem Fall von Diagonalblockgestalt ist, lösen wir zwei Systeme der Größe M , anstatt eines Systems der Dimension $2M$. Außer der schwachen

Besetzungsstruktur weisen diese Systeme keine weiteren Eigenschaften auf, die sich bei ihrer Lösung vorteilhaft ausnutzen lassen. Daher verwenden wir zu ihrer Lösung das in Abschnitt 5.3 beschriebene BiCGSTAB-Verfahren. Die Iteration wird gestoppt, falls $\|\mathcal{T}(\mathbf{u}_k)\mathbf{u}_k - \mathbf{f}\|$ nach dem k -ten Schritt hinreichend klein ist.

Eine ausführliche Diskussion dieses Verfahren im Kontext der inkompressiblen Navier-Stokes-Gleichungen findet sich ebenfalls in [Tur99].

5.2 Newton-Verfahren

Wir befassen uns nun mit der Lösung der nichtlinearen Pressure-Poisson Gleichung (4.38), wobei wir jetzt explizit das Kavitationsmodell (2.33) betrachten. Um dazu ein Newton-Verfahren zu nutzen, setzen wir $\mathbf{F}(\mathbf{p}) = f_{\mathcal{P}}(\mathbf{p}) - \mathcal{P}\mathbf{q}(\mathbf{p})$. Dann gilt

$$\begin{aligned}\mathbf{F}(\mathbf{p}) &= f_{\mathcal{P}}(\mathbf{p}) - \mathcal{P}\mathbf{q}(\mathbf{p}) \\ &= \frac{1}{\Delta t} (\mathbf{B}^T \tilde{\mathbf{u}} + \mathbf{g}(\mathbf{p})) - \mathcal{P}(\mathbf{p} - \mathbf{p}^n - \Delta t Re^{-1} \mathcal{M}_p^{-1} \mathbf{f}_{\mathcal{P}}(\mathbf{p})) \\ &= \frac{1}{\Delta t} (\mathbf{B}^T \tilde{\mathbf{u}} + \mathbf{g}(\mathbf{p})) - \mathcal{P}(\mathbf{p} - \mathbf{p}^n - Re^{-1} \mathcal{M}_p^{-1} (\mathbf{B}^T \tilde{\mathbf{u}} + \mathbf{g}(\mathbf{p}))) \\ &= \frac{1}{\Delta t} (\mathbf{I} + \Delta t Re^{-1} \mathcal{P} \mathcal{M}_p^{-1}) \mathbf{B}^T \tilde{\mathbf{u}} + \mathcal{P} \mathbf{p}^n + \frac{1}{\Delta t} (\mathbf{I} + \Delta t Re^{-1} \mathcal{P} \mathcal{M}_p^{-1}) \mathbf{g}(\mathbf{p}) - \mathcal{P} \mathbf{p}\end{aligned}$$

und die zugehörige Jacobi-Matrix ist durch

$$\mathbf{D}\mathbf{F}(\mathbf{p}) = \frac{1}{\Delta t} (\mathbf{I} + \Delta t Re^{-1} \mathcal{P} \mathcal{M}_p^{-1}) \mathbf{D}\mathbf{g}(\mathbf{p}) - \mathcal{P} \quad (5.3)$$

gegeben. Da wir stückweise konstante Ansatzfunktionen für den Druck verwenden, ist $\mathbf{D}\mathbf{g}$ eine Diagonalmatrix und die Jacobi-Matrix $\mathbf{D}\mathbf{F}$ hat die selbe Besetzungsstruktur wie $\mathcal{P}$. Um dies einzusehen, betrachten wir zunächst die Komponenten von $\mathbf{g}$. Wegen (2.32) gilt

$$\mathbf{g}(\mathbf{p})_i = \int_{K_i} C_\alpha \operatorname{sign}(p_D - p_h) \sqrt{\frac{2}{3} \frac{|p_D - p_h|}{\varrho_F}} d\mathbf{x},$$

mit einem von p unabhängigen Faktor C_α . Da wir stückweise konstante Ansatzfunktionen für den Druck nutzen, d. h. $p_h|_{K_i} = p_i = \text{const}$, ergibt sich daraus

$$\mathbf{g}(\mathbf{p})_i = \operatorname{sign}(p_D - p_i) \sqrt{\frac{2}{3} \frac{|p_D - p_i|}{\varrho_F}} \int_{K_i} C_\alpha d\mathbf{x}. \quad (5.4)$$

Die i -te Komponente des Vektors $\mathbf{g}$ hängt also ausschließlich von der i -ten Komponente der Druckapproximation ab, weshalb sich für die zugehörige Jacobi-Matrix eine Diagonalgestalt ergibt. Insbesondere gilt

$$\mathbf{D}\mathbf{g}(\mathbf{p})_{ii} = \partial_{p_i} \mathbf{g}(\mathbf{p})_i = -\frac{1}{\sqrt{6\varrho_F |p_D - p_i|}} \int_{K_i} C_\alpha d\mathbf{x}. \quad (5.5)$$

Entspricht der Druck in der i -ten Zelle dem Dampfdruck, also $p_D - p_i = 0$, so ist die Ableitung (5.5) nicht definiert. In diesem Fall hat es sich praktisch bewährt p_i durch $p_i + r$ mit $|r| \ll 1$ zu ersetzen. Wir nehmen also eine minimale Änderung der letzten Iterierten vor. In unseren Rechnungen verwenden wir im Weiteren stets $r = 10^{-6}$.

Die Newton-Iteration zur Lösung von $\mathbf{F}(\mathbf{p}) = \mathbf{0}$ ist durch die folgenden Schritte gegeben.

1. Lösung des linearen Teilproblems:

$$\mathbf{D}\mathbf{F}(\mathbf{p}_k)\mathbf{y}_k = -\mathbf{F}(\mathbf{p}_k) \quad (5.6)$$

2. Update mit Line-Search:

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \lambda_k \mathbf{y}_k \quad (5.7)$$

Die Iteration wird gestoppt, falls $\|\mathbf{F}(\mathbf{p}_{k+1})\|$ oder $\|\mathbf{p}_{k+1} - \mathbf{p}_k\|$ hinreichend klein ist. Die in (5.6) auftretende Jacobi-Matrix ist ebenso schwach besetzt wie $\mathcal{P}$. Auch wenn $\mathcal{P}$ symmetrisch positiv definit ist, gilt dies nicht für $\mathbf{D}\mathbf{F}$. Daher lösen wir auch diese System mit dem BiCGSTAB-Verfahren.

Mit Hilfe der Parameter λ_k in (5.7) wird, wie wir im Folgenden sehen werden, die globale Konvergenz des Verfahrens gesichert. In praktischen Rechnungen werden wir zunächst ein Geschwindigkeitsfeld ohne Kavitation bestimmen und die Kavitation erst anschließend dazuschalten. In dieser Situation ist ein global konvergentes Verfahren notwendig, da wir sicher nicht innerhalb des Konvergenzbereichs des Newton-Verfahrens starten werden. Zusätzlich wird durch die Verwendung der Line-Search Strategie die Konvergenz des Verfahrens beschleunigt.

5.2.1 Line-Search

Zur Bestimmung der Schrittlängen-Parameter λ_k nutzen wir die Armijo-Regel, siehe etwa [Kel95]. Dabei wählen wir die Parameter so, dass

$$\|\mathbf{F}(\mathbf{p}_k + \lambda_k \mathbf{y}_k)\| < (1 - \omega \lambda_k) \|\mathbf{F}(\mathbf{p}_k)\|$$

erfüllt ist, wobei $\omega = 10^{-4}$ eine gebräuchliche Wahl ist. Die einfachste Methode einen passenden Wert für $\lambda_k = \lambda_k^m$ zu ermitteln, ist sukzessives Verringern des letzten Wertes ausgehend vom Startwert $\lambda_k^0 = 1$. Da wir dazu den optimalen Faktor zur Verringerung nicht kennen, nutzen wir im Folgenden eine Interpolationstechnik aus [Kel95]. Wir approximieren

$$f(\lambda) = \|\mathbf{F}(\mathbf{p}_k + \lambda \mathbf{y}_k)\| \quad (5.8)$$

durch ein quadratisches Polynom, und wählen dann λ so, dass das Polynom minimiert wird. Das genutzte Polynom ist durch

$$p(\lambda) = f(0) + \lambda c_1 + \lambda^2 c_2$$

mit

$$c_1 = \frac{(\lambda_k^{m-1})^2 (f(\lambda_k^{m-2}) - f(0)) - (\lambda_k^{m-2})^2 (f(\lambda_k^{m-1}) - f(0))}{\lambda_k^{m-1} \lambda_k^{m-2} (\lambda_k^{m-1} - \lambda_k^{m-2})}$$

und

$$c_2 = \frac{\lambda_k^{m-2}(f(\lambda_k^{m-1}) - f(0)) - \lambda_k^{m-1}(f(\lambda_k^{m-2}) - f(0))}{\lambda_k^{m-1}\lambda_k^{m-2}(\lambda_k^{m-1} - \lambda_k^{m-2})}$$

gegeben. Das Polynom stimmt mit der Funktion (5.8) an den Stellen $\lambda = 0$, $\lambda = \lambda_k^{m-1}$ und $\lambda = \lambda_k^{m-2}$ überein und das Extremum wird bei

$$\lambda_t = -\frac{1}{2} \frac{c_1}{c_2}$$

angenommen. Falls das Extremum ein Minimum ist, so wird der aktuellen Wert für λ_k durch

$$\lambda_k^m = \begin{cases} \sigma_0 \lambda_k^{m-1}, & \lambda_t < \sigma_0 \lambda_k^{m-1} \\ \sigma_1 \lambda_k^{m-1}, & \lambda_t > \sigma_1 \lambda_k^{m-1} \\ \lambda_t, & \text{sonst} \end{cases}$$

festgesetzt. Dieses sogenannte Safeguarding mit $0 < \sigma_0 < \sigma_1 < 1$ soll dabei insbesondere die Stagnation der Newton-Iteration im Fall $\lambda_t = 0$ verhindern. In unseren Rechnungen nutzen wir stets $\sigma_0 = 1/10$ und $\sigma_1 = 1/2$. Liegt bei λ_t ein Maximum vor, so setzen wir einfach $\lambda_k^m = \sigma_1 \lambda_k^{m-1}$. Weiter setzen wir $\lambda_k^0 = 1$ und $\lambda_k^1 = \sigma_1$, um die Suche nach einem geeigneten λ_k zu starten.

Abschließend wollen wir nun die Vorteile der Line-Search Strategie anhand eines eindimensionalen Testproblems aufzeigen. Wir betrachten das nichtlineare Problem

$$0 = F(p) = c - g(p), \quad (5.9)$$

mit einer Konstanten c und der nichtlinearen Funktion

$$g(p) = \text{sign}(p_D - p) \sqrt{|p_D - p|}.$$

Diese ist dabei den einzelnen Komponenten von (5.4) nachempfunden, das Testproblem ist somit eine eindimensionale Vereinfachung des eigentlichen Problems $\mathbf{F}(\mathbf{p}) = \mathbf{0}$. Abbildung 5.1 veranschaulicht die Newton-Iterationen, Fehler und Defekt, sowie die Ordnung des Verfahrens bei der Lösung des Testproblems mit und ohne Line-Search. Als Parameter wurden hier $p_D = 0$ und $c = g(1)$ gewählt. Startwert für die Newton-Iteration ist $p_0 = 100$. Ohne Line-Search sind 92 Newton-Iterationen zur Lösung des Testproblems notwendig, mit Line-Search dagegen nur 7. Dabei wurde die Iteration beendet, falls $\|\mathbf{F}\|_2 < 10^{-6}$ gilt. Der zusätzliche Aufwand beim Line-Search entsteht im wesentlichen durch weitere Auswertungen der Funktion F . Vergleichen wir die Anzahl aller Funktionsauswertungen mit und ohne Line-Search für dieses Testproblem, so stehen den 92 Auswertungen ohne Line-Search lediglich 9 Auswertungen im Fall mit Line-Search gegenüber. Insbesondere zeigen die Graphiken wie durch Line-Search diese deutliche Reduktion der Iterationen erreicht wird. Numerische Tests zeigen die gleichen Resultate auch für das mehrdimensionale Problem $\mathbf{F}(\mathbf{p}) = \mathbf{0}$.

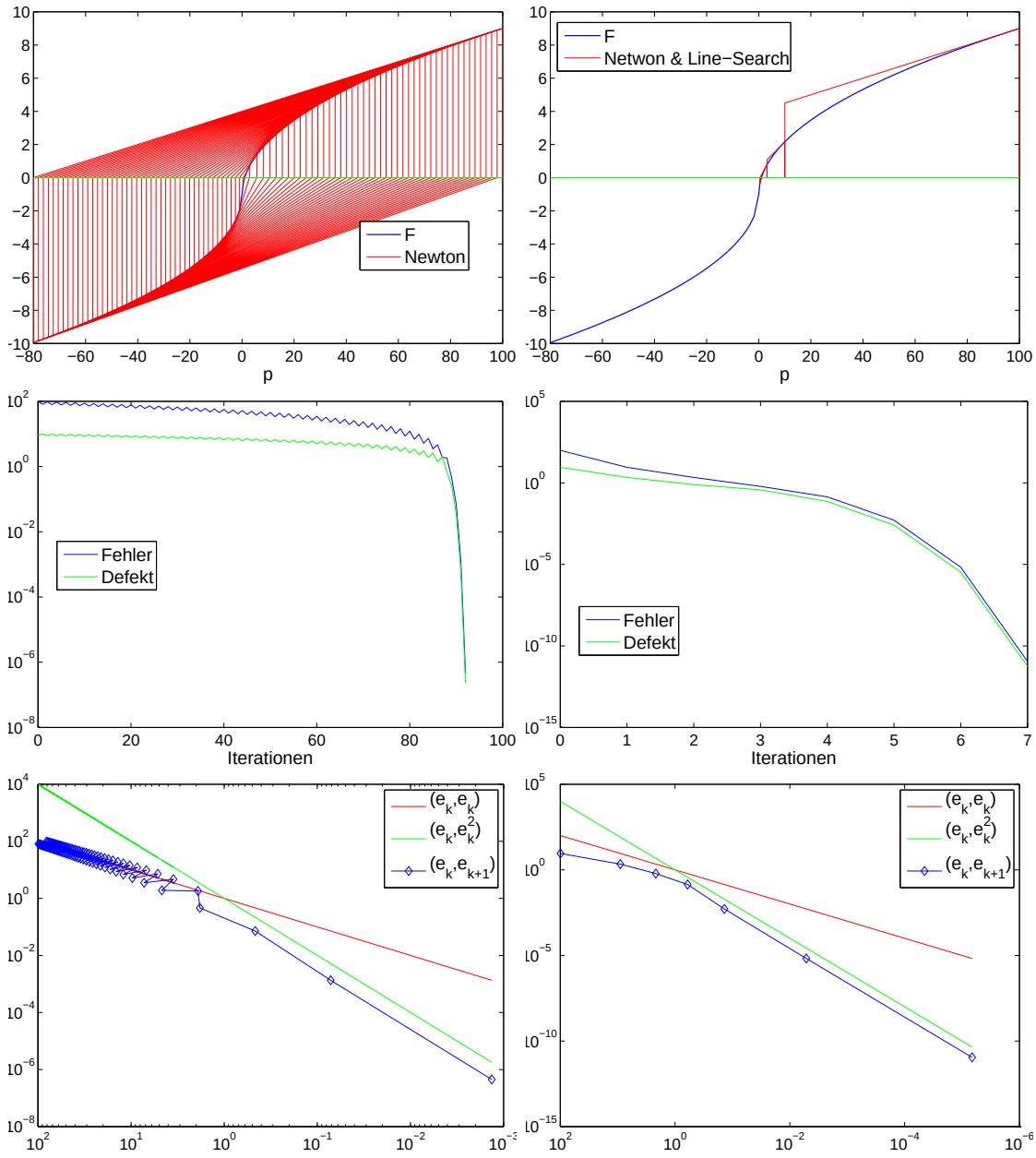


Abbildung 5.1: Newton-Verfahren ohne (links) und mit Line-Search (rechts) für das Testproblem (5.9). Veranschaulichung der Berechnung der Iterierten (oben), Fehler und Defekt (mitte) und Ordnung des Verfahrens (unten). Zur graphischen Bestimmung der Ordnung wurde der Fehler nach der $k + 1$ -ten Iteration e_{k+1} über dem Fehler e_k logarithmisch aufgetragen.

5.3 BiCGSTAB

Wie schon oben beschrieben, lösen wir die linearen Probleme (5.2) und (5.6) mit dem BiCGSTAB-Verfahren. Dies ist eine Krylov-Unterraum-Methode, die sich durch kurze Rekursion und damit konstantem Rechen- bzw. Speicheraufwand pro Iteration auszeichnet. In jedem Durchlauf müssen vier Skalarprodukte und zwei Matrix-Vektor-Produkte berechnet werden. Da die in unserem Fall auftretenden linearen Systeme schwach besetzt sind, kann die Berechnung der Matrix-Vektor-Produkte effizient durchgeführt werden. Dem BiCGSTAB-Verfahren liegt keine Minimierung des Residuums zugrunde, dennoch werden die auftretenden Oszillationen durch eine integrierte eindimensionale Minimierung gedämpft.

Zur Beschreibung des Verfahrens sei das System $\mathbf{Ax} = \mathbf{b}$ mit $\mathbf{A} \in \mathbb{R}^{n \times n}$ und $\mathbf{x}, \mathbf{b} \in \mathbb{R}^n$ gegeben. Die Schritte des BiCGSTAB-Verfahrens lauten dann:

1. Wähle $\mathbf{x}_0 \in \mathbb{R}^n$ und $\varepsilon > 0$
2. Setze $\mathbf{r}_0 = \mathbf{p}_0 = \mathbf{b} - \mathbf{Ax}_0$, $\rho_0 = \mathbf{r}_0^T \mathbf{r}_0$, $j = 0$
3. Solange $\|\mathbf{r}\|_2 > \varepsilon$
 - a) $\mathbf{v}_j = \mathbf{Ap}_j$, $\alpha_j = \rho_j / (\mathbf{v}_j^T \mathbf{r}_0)$
 - b) $\mathbf{s}_j = \mathbf{r}_j - \alpha_j \mathbf{v}_j$, $\mathbf{t}_j = \mathbf{As}_j$
 - c) $\omega_j = (\mathbf{t}_j^T \mathbf{s}_j) / (\mathbf{t}_j^T \mathbf{t}_j)$
 - d) $\mathbf{x}_{j+1} = \mathbf{x}_j + \alpha_j \mathbf{p}_j + \omega_j \mathbf{s}_j$
 - e) $\mathbf{r}_{j+1} = \mathbf{s}_j - \omega_j \mathbf{t}_j$
 - f) $\rho_{j+1} = \mathbf{r}_{j+1}^T \mathbf{r}_0$, $\beta_j = (\alpha_j / \omega_j)(\rho_{j+1} / \rho_j)$
 - g) $\mathbf{p}_{j+1} = \mathbf{r}_{j+1} + \beta_j(\mathbf{p}_j - \omega_j \mathbf{v}_j)$, $j = j + 1$

Nachteil des Verfahrens ist ein möglicher frühzeitiger Verfahrensabbruch in 3f). In unseren praktischen Rechnungen konnte dieser Fall allerdings nicht beobachtet werden. Eine ausführliche Herleitung und Diskussion des Verfahrens findet sich in [Mei99].

6 Advektions-Diffusions-Gleichung

In diesem Kapitel beschäftigen wir uns mit der numerischen Lösung der Advektions-Diffusions-Gleichung (2.37c), wobei wir davon ausgehen, dass das Geschwindigkeitsfeld $\mathbf{u}$ analytisch gegeben ist. Die Auswirkungen eines numerischen Geschwindigkeitsfeld untersuchen wir in Kapitel 7. Da wir uns speziell für den Fall transportdominanter Advektions-Diffusions-Gleichungen oder sogar reiner Advektions-Gleichungen interessieren, behandeln wir die Gleichung mit Discontinuous-Galerkin-Verfahren (DG-Verfahren).

6.1 Discontinuous-Galerkin-Verfahren

Die DG-Verfahren wurden ursprünglich für hyperbolische Erhaltungsgleichungen entwickelt und später auf andere Gleichungen übertragen. Sie können als Verallgemeinerung von Finite-Volumen-Verfahren aufgefasst werden und sind formal von beliebiger Ordnung. Uns interessiert hier nicht der Aspekt der hohen Ordnung, sondern ihr Potential zur Lösung von Transportproblemen, welches sich aus der lokalen Konservativität der Verfahren ergibt. Im Gegensatz zu FE-Verfahren sind hier keine zusätzlichen Stabilisierungsterme notwendig. Wir verweisen auf [CKS00, CS01, HW08, Riv08, Shu09], um einen Überblick über DG-Verfahren zu erhalten. Arbeiten die sich speziell mit der Advektions-(Diffusions)-Gleichung befassen sind etwa [LS00, ADCC01, DGM09].

Der Abschnitt ist wie folgt unterteilt. Zunächst stellen wir ein explizites DG-Verfahren für reine Transportprobleme vor. Ziel ist die Lösung der Transportgleichung (2.33c) für die Dampfdichte aus dem Kavitationsmodell. Wie wir in Abschnitt 2.3.1 gesehen haben, erfüllt die Dampfdichte das Maximumprinzip $0 \leq \alpha \leq 1$. Dies sollte auch unsere numerische Verfahren erfüllen. In [ZS10] wird diesbezüglich eine entsprechende Vorgehensweise im Fall der homogenen Konvektionsgleichung bei divergenzfreiem Geschwindigkeitsfeld vorgestellt. Dabei wird zunächst über eine CFL-Bedingung das Maximumprinzip für die Mittelwerte der DG-Approximationen garantiert und in einem zweiten Schritt durch die Anwendung eines Limiters auf die eigentliche Approximation übertragen. Wir werden dieses Verfahren auf divergenzbehaftete Geschwindigkeitsfelder und nichtlineare Quellterme erweitern. Dies geschieht, indem wir die Idee des in [BDM03] für gewöhnliche Differentialgleichungen vorgestellten Euler-Patankar-Verfahrens auf die DG-Diskretisierung transferieren. Daher führen wir nun zunächst das Patankar-Verfahren ein, bevor wir uns mit den Resultaten aus [ZS10] befassen. Durch Kombination der beiden Techniken erhalten wir dann das neuartige Patankar-DG-Verfahren, welches auch für das Kavitationsmodell die Beschränktheit der Dampfdichte sichern wird.

Danach betrachten wir die diffusiven Anteile der Gleichung. Die Konstruktion von DG-Verfahren für elliptische Probleme ist ebenso mit Schwierigkeiten verbunden, wie die

Entwicklung von FEM für hyperbolische Gleichungen. Eine Möglichkeit zur Behandlung diffusiver Terme innerhalb eines DG-Verfahrens sind die Recovery-based DG-Verfahren [LLR07]. Diese werden wir nutzen, um ein DG-Verfahren zur Lösung von (2.37c) zu erhalten.

6.1.1 Discontinuous-Galerkin Diskretisierung für Advektion

Bevor wir uns der gesamten Gleichung (2.37c) widmen, beschränken wir uns zunächst auf den rein konvektiven Fall und betrachten also die Gleichung

$$\partial_t C + \nabla \cdot (C\mathbf{u}) = f. \quad (6.1)$$

Die zugehörigen Anfangs- und Randbedingungen lauten, wie schon in Abschnitt 2.4 bemerkt,

$$C|_{t=0} = C^0, \quad C|_{\partial\Omega_{in}} = C_D.$$

Der Einflusstrand $\partial\Omega_{in}$ ist insbesondere dadurch charakterisiert, dass es in das Gebiet hineinströmt, also $\mathbf{u} \cdot \mathbf{n}|_{\partial\Omega_{in}} < 0$ gilt.

Zunächst werden wir wieder eine schwache Formulierung bestimmen. Anders als bei den Finite-Elemente-Diskretisierungen zuvor, integrieren wir die Gleichung nach Multiplikation mit einer hinreichend glatten Testfunktion w nun nicht über das gesamte Gebiet, sondern lediglich über eine einzelne Zelle K . Somit erhalten wir nach partieller Integration von (6.1) die Gleichung

$$(\partial_t C, w)_K + (\nabla \cdot (C\mathbf{u}), 1)_K - (C\mathbf{u}, \nabla w)_K = (f, w)_K.$$

Anschließende Anwendung des Gaußschen Integralsatzes liefert dann

$$(\partial_t C, w)_K + \langle C\mathbf{u} \cdot \mathbf{n}, w \rangle_{\partial K} - (C\mathbf{u}, \nabla w)_K = (f, w)_K. \quad (6.2)$$

Im Gegensatz zur FEM wollen wir hier zur Approximation der Lösung von (6.2) Funktionen zulassen, die keinen Stetigkeitsbedingungen unterliegen. Dazu sei

$$W_h = \{ w_h \mid w_h|_K \text{ ist für } K \in \mathcal{T}_h \text{ ein Polynom vom Grad } \leq N \}$$

der Raum der elementweise polynomialen Funktionen über dem Gitter $\mathcal{T}_h$. Derartige Funktionen sind im Allgemeinen über die Zellränder hinweg unstetig und es ist unklar wie das Randintegral in (6.2) zu behandeln ist. Wir nutzen daher eine konsistente und konservative numerische Flussfunktion $H(C^-, C^+, \mathbf{u} \cdot \mathbf{n})$ zur Berechnung der Randflüsse, wobei wir mit C^- die Werte innerhalb und mit C^+ die Werte außerhalb der Zelle K bezeichnen. Genauer gilt für $\varepsilon > 0$

$$C^-(\mathbf{x}) = \lim_{\varepsilon \rightarrow 0} C(\mathbf{x} - \varepsilon \mathbf{n}), \quad C^+(\mathbf{x}) = \lim_{\varepsilon \rightarrow 0} C(\mathbf{x} + \varepsilon \mathbf{n}),$$

mit dem äußeren Normalenvektor $\mathbf{n}$.

Eine Flussfunktion heißt konsistent, falls sie $H(C, C, \mathbf{u} \cdot \mathbf{n}) = C \mathbf{u} \cdot \mathbf{n}$ erfüllt. Von Konservativität sprechen wir, wenn $H(C^-, C^+, \mathbf{u} \cdot \mathbf{n}) = -H(C^+, C^-, -\mathbf{u} \cdot \mathbf{n})$ gilt. Wir nutzen im Folgenden stets den lokalen Lax-Friedrichs-Fluss

$$H(C^-, C^+, \mathbf{u} \cdot \mathbf{n}) = \frac{1}{2} (\mathbf{u} \cdot \mathbf{n} (C^+ + C^-) - a(C^+ - C^-)), \quad (6.3)$$

wobei $a = \max |\mathbf{u} \cdot \mathbf{n}|$ die maximale Geschwindigkeit in Normalenrichtung über die jeweilige Kante bezeichnet. Bei konstanter Geschwindigkeit unterscheidet sich dieser nicht vom Upwind-Fluss. Denn gilt $\mathbf{u} \cdot \mathbf{n} < 0$, strömt es also in die Zelle hinein, so ergibt sich $H(C^-, C^+, \mathbf{u} \cdot \mathbf{n}) = \mathbf{u} \cdot \mathbf{n} C^+$. Strömt es aus der Zelle hinaus, so gilt entsprechend $H(C^-, C^+, \mathbf{u} \cdot \mathbf{n}) = \mathbf{u} \cdot \mathbf{n} C^-$.

Insgesamt ergibt sich das DG-Verfahren somit als: Finde für $t > 0$ ein $C_h(t) \in W_h$, so dass

$$(\partial_t C_h, w_h)_K + \langle H(C_h^-, C_h^+, \mathbf{u} \cdot \mathbf{n}), w_h^- \rangle_{\partial K} - (C_h \mathbf{u}, \nabla w_h)_K = (f, w_h)_K \quad \forall w_h \in W_h \quad (6.4)$$

erfüllt ist. Wir sehen, dass der Austausch von Informationen zwischen den einzelnen Zellen ausschließlich über die Ränder verläuft. Demnach können auch die Randbedingungen durch Vorgabe von C_h^+ entlang der Randkanten des Einflussrandes $\partial\Omega_{in}$ realisiert werden.

Insbesondere setzen wir in diesem Kapitel voraus, dass $\mathbf{u} \cdot \mathbf{n}$ über Zellränder hinweg stetig ist. Diese Bedingung wird von einem diskreten Geschwindigkeitsfeld aus dem in Abschnitt 3.1.3 gegebenen FE-Raum $\mathbf{U}_h$, siehe (3.12a), nicht erfüllt. Allerdings werden wir in Kapitel 7 sehen, dass die Kopplung von Strömung und Transport spezielle Anforderungen an die gewählte Diskretisierung stellt. Werden diese, wie in unserem Fall, nicht erfüllt, so ist ein entsprechendes Postprocessing des Geschwindigkeitsfelds nötig. Die daraus resultierende angepasste Geschwindigkeitsverteilung wird dann aber insbesondere die gewünschte Stetigkeit in Normalenrichtung erfüllen.

Konkret verwenden wir in unseren Rechnungen parametrische bilineare Ansatzfunktionen, d. h. Funktionen aus

$$Q_1(K) = \{ w_h \mid w_h = \hat{w}_h \circ F_K^{-1}, \hat{w}_h \in \text{span}\langle 1, x, y, xy \rangle \}. \quad (6.5)$$

Hierbei bezeichnen wir mit $\hat{K} = [-1, 1] \times [-1, 1]$ das Referenzquadrat und mit F_K die zugehörige Referenzabbildung $F_K : \hat{K} \rightarrow K$. Insbesondere wählen wir als Basis von $\langle 1, x, y, xy \rangle$ die übliche Langrange-Basis. Also genügt die j -te Basisfunktion $\hat{w}_h^j$ den Bedingungen $\hat{w}_h^j(\mathbf{x}_j) = 1$ und $\hat{w}_h^j(\mathbf{x}_i) = 0$ für $i \neq j$, wobei $\mathbf{x}_i$ für $i = 1, \dots, 4$ die Ecken des Elements K bezeichnen. Damit sind die Elemente von $Q_1(K)$ also durch ihre Werte an den Ecken eines Elements eindeutig bestimmt. Insbesondere gilt die Gleichheit

$$\hat{w}_h^1(\mathbf{x}) + \hat{w}_h^2(\mathbf{x}) + \hat{w}_h^3(\mathbf{x}) + \hat{w}_h^4(\mathbf{x}) = 1. \quad (6.6)$$

Bei Verwendung dieser Elemente können wir zweite Ordnung im Raum erwarten. Siehe dazu [Shu09] und die angegebene Literatur.

Wie bei den Finiten-Elementen in Abschnitt 3.1.3 wollen wir nun aus (6.4) ein lineares Gleichungssystem herleiten. Dazu sei $w_h^1, \dots, w_h^4$ die Basis von $Q_1(K)$. Wir weisen

darauf hin, dass sich die nachfolgenden Funktionen, Matrizen und Vektoren stets auf das Element K beziehen. Wir verzichten aber aus Gründen der Übersichtlichkeit auf eine entsprechende Kennzeichnung.

Setzen wir den Ansatz $C_h(\mathbf{x}, t) = \sum_{j=1}^4 C_j(t) w_h^j(\mathbf{x})$ in (6.4) ein und lassen wir w_h die einzelnen Basisfunktionen durchlaufen, so erhalten wir das räumlich diskrete System

$$\mathbf{M}\dot{\mathbf{C}} - \mathbf{K}(\mathbf{u})\mathbf{C} + \mathbf{H}(\mathbf{C}^+, \mathbf{C}^-, \mathbf{u} \cdot \mathbf{n}) = \mathbf{f}. \quad (6.7)$$

Hier bezeichnen wir mit $\mathbf{C} = (C_1, \dots, C_4)^T$ den zu C_h zugehörigen Koeffizientenvektor. Die anderen Vektoren und Matrizen sind durch

$$\begin{aligned} \mathbf{M}_{ij} &= (w_h^j, w_h^i)_K, & \mathbf{K}(\mathbf{u})_{ij} &= (w_h^j \mathbf{u}, \nabla w_h^i)_K, \\ \mathbf{H}(\mathbf{C}^+, \mathbf{C}^-, \mathbf{u} \cdot \mathbf{n})_i &= \langle H(C_h^-, C_h^+, \mathbf{u} \cdot \mathbf{n}), w_h^i \rangle_{\partial K}, & \mathbf{f}_i &= (f, w_h^i)_K, \end{aligned}$$

definiert. Im Gegensatz zu den entsprechenden Matrizen der Finite-Elemente-Diskretisierung sind die hier auftretenden Matrizen voll besetzt. Allerdings kann die Massenmatrix $\mathbf{M} \in \mathbb{R}^{4 \times 4}$ aufgrund ihrer geringen Größe problemlos invertiert werden.

Der Vollständigkeit halber geben wir hier noch die vollständig diskrete Form nach Zeitdiskretisierung mit dem expliziten Euler-Verfahren an. Gleichung (6.7) lautet dann

$$\mathbf{C}^{n+1} = \mathbf{C}^n + \Delta t \mathbf{M}^{-1} (\mathbf{K}(\mathbf{u}^n) \mathbf{C}^n - \mathbf{H}(\mathbf{C}^{+,n}, \mathbf{C}^{-,n}, \mathbf{u}^n \cdot \mathbf{n}) + \mathbf{f}^n).$$

Um den Vektor der Anfangsdaten $\mathbf{C}^0$ zu erhalten, verwenden wir wie in Abschnitt 3.1.4 die L^2 -Projektion. Hier nutzen wir allerdings die L^2 -Projektion $\mathcal{P} : L^2(K) \rightarrow Q_1(K)$. Das explizite Euler-Verfahren wird sich im Folgenden als für unsere Zwecke unzureichend erweisen, da es die Einhaltung des Maximumprinzips für die Dampfdichte aus Abschnitt 2.3.1 nicht garantieren kann. Durch welche Modifikationen wir dies ändern können, sehen wir in den nächsten Abschnitten.

Im Kontext von DG-Verfahren müssen wir noch auf die starke Zeitschrittweitenrestriktion des expliziten Euler-Verfahrens, siehe [CC89, CS01], hinweisen. Abhilfe schaffen spezielle Zeitintegrationsverfahren höherer Ordnung [GKS09], welche auch im Zusammenhang mit den nachfolgend beschriebenen Techniken problemlos einsetzbar sind. Da die Zeitintegration nicht im Vordergrund dieser Arbeit steht, wird dies hier nicht weiter ausgeführt.

Abschließend wollen wir noch numerisch bestätigen, dass das Verfahren für hinreichend glatte Lösungen räumlich von zweiter Ordnung ist. Dazu betrachten wir auf $\Omega = [0, 3]^2$ das Testproblem

$$\partial_t C + \nabla \cdot (C \mathbf{u}) = 0, \quad (6.8)$$

mit dem Geschwindigkeitsfeld $\mathbf{u}(x, y) = (-y, x)^T$. Die Randdaten wählen wir dabei so, dass sich

$$C(x, y) = \sin(0.5(x^2 + y^2)) \quad (6.9)$$

als stationäre Lösung des Testproblems ergibt. Zur Diskretisierung betrachten wir quadratische Gitter mit der Maschenweite $\Delta x = 0.6/2^l$ für $l = 0, \dots, 4$. Die zugehörigen Fehler, sowie die daraus resultierende Ordnung, sind für die einzelnen Gitter in Tabelle 6.1 aufgelistet und zeigen wie erwartet die zweite Ordnung. Als Anfangsdaten wurde dabei jeweils die Lösung, also $C^0(x, y) = \sin(0.5(x^2 + y^2))$ gewählt.

Tabelle 6.1: Fehler und Konvergenzordnung für das Testproblem (6.8)

Gitter l	$\ C - C_h\ _{L^2(\Omega)}$ Fehler	Ordnung
0	0.316899	-
1	0.074277	2.093035
2	0.017359	2.097270
3	0.004256	2.028122
4	0.001064	1.999516

6.1.2 Patankar-Verfahren

Wir führen nun das Patankar-Verfahren ein, mit welchem die Positivität entsprechender gewöhnlicher Differentialgleichungen numerisch gesichert werden kann. Allerdings nutzen wir es hier sogar zur Einhaltung eines Maximumprinzips. Für eine gewöhnliche Differentialgleichung der Form

$$\alpha'(t) = P(\alpha(t)) - D(\alpha(t)), \quad \alpha(t_0) = \alpha_0 \quad (6.10)$$

mit $P, D \geq 0$, ist das Euler-Patankar-Verfahren nach [BDM03] durch

$$\alpha^{n+1} = \alpha^n + \Delta t P(\alpha^n) - \Delta t D(\alpha^n) \frac{\alpha^{n+1}}{\alpha^n} \quad (6.11)$$

gegeben. Im Weiteren sprechen wir kurz vom Patankar-Verfahren. Das Verfahren ist linear implizit, so dass keine Lösung einer nichtlinearen Gleichung erforderlich ist. Weiter stimmt es für $\alpha' = -D(\alpha)$ und linearer rechter Seite mit dem impliziten Euler-Verfahren überein. Es wurde eingeführt, um bei entsprechenden Differentialgleichungen die Positivität der numerischen Lösung für beliebige Zeitschrittweiten zu sichern und stimmt für $D = 0$ mit dem expliziten Euler-Verfahren überein. Für $\alpha^n > 0$ ist das Verfahren wohldefiniert, spezielle Berücksichtigung erfordert der Fall $\alpha^n = 0$. Um die Positivität der Lösung von (6.10) zu gewährleisten, muss mit $\alpha \rightarrow 0$ auch $D(\alpha) \rightarrow 0$ gelten. Somit verschwindet der destruktive Term D für $\alpha^n = 0$ und eine Gewichtung wie in (6.11) wird unnötig.

Eine Umformulierung von (6.11) liefert

$$\alpha^{n+1} = \frac{\alpha^n (\alpha^n + \Delta t P(\alpha^n))}{\alpha^n + \Delta t D(\alpha^n)}, \quad (6.12)$$

womit sich die unbedingte Positivität des Verfahrens unter der Voraussetzung $\alpha^0 > 0$ ablesen lässt. Das durch die Gewichtung des destruktiven Terms D die Konsistenz des Verfahrens nicht verloren geht, zeigt das folgende Lemma.

Lemma 6.1. *Das Patankar-Verfahren (6.11) zur Lösung von (6.10) ist von erster Ordnung.*

Beweis. Zunächst erhalten wir durch Taylorentwicklung der exakten Lösung von (6.10)

$$\alpha(t^{n+1}) = \alpha(t^n) + \alpha'(t^n)\Delta t + \mathcal{O}(\Delta t^2), \quad \Delta t \rightarrow 0.$$

Anschließendes Ersetzen der Ableitung durch die Differentialgleichung (6.10) liefert

$$\alpha(t^{n+1}) = \alpha(t^n) + \Delta t [P(\alpha(t^n)) - D(\alpha(t^n))] + \mathcal{O}(\Delta t^2), \quad \Delta t \rightarrow 0.$$

Durch Nulladdition und Ausnutzung der Verfahrensvorschriften (6.11), sowie (6.12), erhalten wir dann

$$\begin{aligned} & \alpha(t^n) + \Delta t [P(\alpha(t^n)) - D(\alpha(t^n))] \\ &= \alpha(t^n) + \Delta t \left[P(\alpha(t^n)) - D(\alpha(t^n)) \frac{\alpha^{n+1}}{\alpha(t^n)} \right] + \Delta t \left[D(\alpha(t^n)) \frac{\alpha^{n+1}}{\alpha(t^n)} - D(\alpha(t^n)) \right] \\ &= \alpha^{n+1} + \Delta t^2 \frac{D(\alpha(t^n))(P(\alpha(t^n)) - D(\alpha(t^n)))}{\alpha(t^n) + \Delta t D(\alpha(t^n))}. \end{aligned}$$

Damit haben wir die behauptete Konsistenz

$$\alpha(t^{n+1}) = \alpha^{n+1} + \mathcal{O}(\Delta t^2), \quad \Delta t \rightarrow 0$$

gezeigt. □

Wir wollen das Patankar-Verfahren nun zur Zeitintegration der gewöhnlichen Differentialgleichung

$$\alpha'(t) = f_\alpha(\alpha), \tag{6.13}$$

mit der rechten Seite aus dem Kavitationsmodell, nutzen. Wie bereits erwähnt, lohnt sich die Betrachtung des 3D-Quellterms (2.27) auch dann, wenn wir in dieser Arbeit nur zwei Raumdimensionen betrachten, da sich die vorgestellten Verfahren gleichermaßen im Dreidimensionalen anwenden lassen.

Wir betrachten also den verallgemeinerten Quellterm (2.32) in der Form

$$f_\alpha(\alpha) = C(p)(1 - \alpha)^r \alpha^s,$$

mit $0 < r, s < 1$ und einer Konstanten $C(p)$. Damit erfüllt die Lösung der Gleichung (6.13) unter der Annahme $0 \leq \alpha(t_0) \leq 1$ das Maximumprinzip $0 \leq \alpha \leq 1$. Weiter ist α monoton wachsend für $C(p) \geq 0$ und monoton fallend, für $C(p) \leq 0$.

Für diese Differentialgleichung ist das explizite Euler-Verfahren

$$\alpha^{n+1} = \alpha^n + \Delta t C(p)(1 - \alpha^n)^r (\alpha^n)^s$$

ungeeignet, denn um für $C(p) < 0$ die Positivität zu sichern, muss die Zeitschrittweite der Bedingung

$$\Delta t \leq -\frac{1}{C(p)} \frac{(\alpha^n)^{1-s}}{(1 - \alpha^n)^r} \tag{6.14}$$

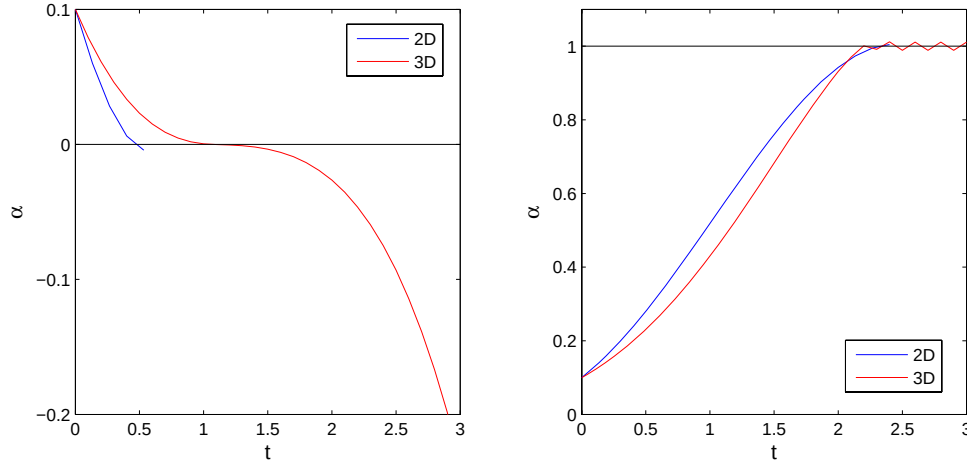


Abbildung 6.1: Explizites Euler-Verfahren zur Lösung von (6.13) mit $C(p) = -1$ (links) und $C(p) = 1$ (rechts). Die Zeitschrittweite beträgt $\Delta t = 0.1$, der Anfangswert ist durch $\alpha^0 = 0.1$ gegeben. Rechte Seite (2.29) in blau und (2.27) in rot.

genügen. Da aber die rechte Seite der Ungleichung für $\alpha^n \rightarrow 0$ wegen $s < 1$ ebenfalls verschwindet, ist diese Schrittweitenrestriktion unpraktikabel. Eine zu (6.14) analoge Bedingung gilt wegen $r < 1$ auch im Fall $C(p) > 0$, um für $\alpha^n \rightarrow 1$ die Einhaltung der oberen Schranke zu sichern. Die Einhaltung beider Schranken ist aber besonders wichtig, da die Iterierten für (2.29) bei Missachtung der Zeitschrittweitenrestriktion sehr schnell komplex werden und das Verfahren abbricht. Ähnlich verhält es sich auch für den Quellterm (2.27), wobei allerdings negative Werte unter der Wurzel vermieden werden können. Im Fall $C(p) < 0$ gehen die Iterierten dann gegen $-\infty$, für $C(p) > 0$ treten unerwünschte Oszillationen auf. Dieses Verhalten des expliziten Euler-Verfahrens ist in Abbildung 6.1 für beide Quellterme dargestellt.

Um diese Problematik zu umgehen, nutzen wir daher im Fall $C(p) < 0$ das Patankar-Verfahren um die Positivität zu sichern. Dazu zerlegen wir $f_\alpha = P - D$ mit

$$P = 0, \quad D = -f_\alpha. \quad (6.15)$$

Allerdings ist das Verfahren (6.11) für $\alpha^n = 0$ nicht definiert. Dies stellt aber kein Problem dar, da in diesem Fall auch $\alpha^{n+1} = 0$ gelten muss, was wir direkt aus (6.13) ablesen können. Dieser Spezialfall kann also durch eine entsprechende Abfrage abgefangen werden.

Lemma 6.2. *Wir betrachten die Patankar-Näherung (6.11) mit der Wahl (6.15) zur Lösung von (6.13). Falls $C(p) < 0$ und $0 < \alpha^0 \leq 1$, so erfüllen die Iterierten die Monotonie und Beschränktheit*

$$0 < \alpha^{n+1} \leq \alpha^n \leq \alpha^0.$$

Beweis. Der Beweis erfolgt induktiv über n und für $n = 0$ gilt die Behauptung nach Voraussetzung. Gilt die Behauptung für α^n so ergibt sich nach (6.15) mit $C(p) < 0$ die Positivität $D > 0$. Aus (6.12) und $P = 0$ erhalten wir dann

$$\alpha^{n+1} = \alpha^n \left(\frac{\alpha^n}{\alpha^n + \Delta t D(\alpha^n)} \right).$$

In dieser Form können wir die Positivität, als auch die Monotonie direkt ablesen. $\square$

Um nun für $C(p) > 0$ die Beschränktheit nach oben zu garantieren, führen wir dies auf die Einhaltung einer Positivitätsbedingung zurück. Wir setzen

$$\alpha = 1 - \beta \tag{6.16}$$

und erhalten aus (6.13) die Differentialgleichung

$$\beta' = -f_\alpha(1 - \beta). \tag{6.17}$$

Dabei ist $\alpha \leq 1$ äquivalent zu $\beta \geq 0$. Wie zuvor können wir mit dem Patankar-Verfahren die Positivität der Hilfsgrößen β^n sichern, anschließende Rücktransformation liefert dann $\alpha^n \leq 1$. Dieses Vorgehen bezeichnen wir im Folgenden als Patankar-Verfahren mit Transformation. Weiter bezeichnen wir β als transformierte Variable und die zugehörige Differentialgleichung (6.17) als transformierte Gleichung. Die zur Patankar-Formulierung gehörigen Quellterme für die transformierte Gleichung wählen wir als

$$P_\beta(\beta) = 0, \quad D_\beta(\beta) = f_\alpha(1 - \beta). \tag{6.18}$$

Dies sichert, analog zu $C(p) < 0$, neben der Positivität auch die Monotonie der transformierten Variablen und damit auch die gewünschten Eigenschaften von α .

Lemma 6.3. *Wir betrachten die Patankar-Näherung (6.11) mit Transformation (6.16) und der Wahl (6.18) zur Lösung von (6.13). Falls $C(p) > 0$ und $0 \leq \alpha^0 < 1$, so erfüllen die Iterierten die Monotonie und Beschränktheit*

$$\alpha^0 \leq \alpha^n \leq \alpha^{n+1} < 1.$$

Beweis. Analog zu Lemma 6.2 ergibt sich $0 < \beta^{n+1} \leq \beta^n \leq \beta^0$. Anwendung der Transformation (6.16) liefert dann wie gewünscht

$$\alpha^0 \leq \alpha^n \leq \alpha^{n+1} < 1.$$

$\square$

Durch Einführung eines anderen Quellterm-Splittings können wir das Patankar-Verfahren mit Transformation auch direkt als Patankar-Verfahren für α formulieren. Wir schreiben zunächst nach (6.12) und (6.18) die Iterierten der transformierten Variable in der Form

$$\beta^{n+1} = \frac{(\beta^n)^2}{\beta^n + \Delta t C(p)(\beta^n)^r(1 - \beta^n)^s}.$$

Transformation mittels (6.16) und Ausklammern liefert dann

$$1 - \alpha^{n+1} = \frac{(1 - \alpha^n)^2}{1 - \alpha^n + \Delta t C(p)(1 - \alpha^n)^r (\alpha^n)^s} = \frac{(1 - \alpha^n)^2}{(1 - \alpha^n)(1 + \Delta t C(p)(1 - \alpha^n)^{r-1} (\alpha^n)^s)}.$$

Schließlich erhalten wir damit nach weiteren Umformungen

$$\alpha^{n+1} = 1 - \frac{1 - \alpha^n}{1 + \Delta t C(p)(1 - \alpha^n)^{r-1} (\alpha^n)^s} = \frac{\alpha^n + \Delta t C(p)(1 - \alpha^n)^{r-1} (\alpha^n)^s}{1 + \Delta t C(p)(1 - \alpha^n)^{r-1} (\alpha^n)^s}.$$

Ein Vergleich mit (6.12) zeigt uns, dass das Patankar-Verfahren mit Transformation im Fall $C(p) > 0$ äquivalent ist zum direkten Verfahren mit

$$P = C(p)(1 - \alpha^n)^{r-1} (\alpha^n)^s, \quad D = C(p)(1 - \alpha^n)^{r-1} (\alpha^n)^{s+1}. \quad (6.19)$$

Insbesondere gilt tatsächlich $f_\alpha = P - D$.

Für den Quellterm (2.27) mit $r = 1/3$ und $s = 2/3$ ergibt dies

$$P = C(p) \left(\frac{\alpha}{1 - \alpha} \right)^{2/3}, \quad D = C(p) \alpha \left(\frac{\alpha}{1 - \alpha} \right)^{2/3}.$$

Dieses Quellterm-Splitting wurde von uns in [KMW11] für die Behandlung der partiellen Differentialgleichung vorgeschlagen. Es hat sich aber gezeigt, dass die Realisierung der Einhaltung der oberen Schranke durch die Sicherstellung der Positivität der Transformaten numerische Vorteile gegenüber diesem Splitting hat. Diese werden wir in Abschnitt 6.1.4 diskutieren, nachdem wir die Idee des Patankar-Verfahrens auf die DG-Diskretisierung übertragen haben.

Abschließend zeigen wir noch in Abbildung 6.2 das Verhalten des Patankar-Verfahrens mit Transformation zur Lösung von (6.13) für $C(p) = 1$ und $C(p) = -1$. Dabei wurde die Zeitschrittweite $\Delta t = 0.1$ verwendet und es gilt $\alpha^0 = 0.1$. Die Einhaltung des Maximumprinzips ist dabei deutlich sichtbar. Wie in Abbildung 6.1 gezeigt, kann das explizite Euler-Verfahren in diesem Fall nicht verwendet werden.

6.1.3 Linear-Scaling-Limiter

Bevor wir uns damit befassen, das Patankar-Verfahren auf die DG-Diskretisierung zu übertragen, führen wir noch den eingangs erwähnten Limiter, siehe auch [LO96], in seiner allgemeinen Form ein. Der Limiter sichert die Einhaltung eines Maximumprinzips unter der Voraussetzung, dass der Zellmittelwert dieses bereits erfüllt.

Wir betrachten zur Einführung des Limiters das auf der Zelle K definierte DG-Polynom $p(\mathbf{x})$ und nehmen an, dass der zugehörige Mittelwert $\bar{p}$ das Maximumprinzip $m \leq \bar{p} \leq M$ erfüllt. Unter dieser Voraussetzung wird dann auch das limitierte Polynom $\hat{p}$ innerhalb dieser Schranken liegen. Um dieses definieren zu können, bezeichnen wir mit

$$M_K = \max_{\mathbf{x} \in K} p(\mathbf{x}), \quad m_K = \min_{\mathbf{x} \in K} p(\mathbf{x})$$

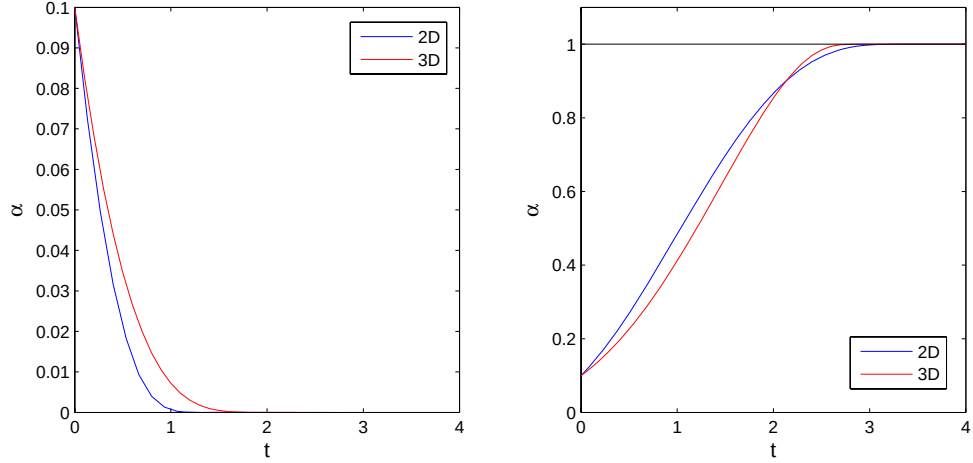


Abbildung 6.2: Patankar-Verfahren mit Transformation für (6.13) mit $C(p) = -1$ (links) und $C(p) = 1$ (rechts). Rechte Seite (2.29) in blau und (2.27) in rot.

das Maximum bzw. Minimum von p auf der Zelle K . Das limitierte Polynom $\hat{p}$ erhalten wir dann durch

$$\hat{p}(\mathbf{x}) = \bar{p} + \theta(p(\mathbf{x}) - \bar{p}), \quad \theta = \min \left\{ \left| \frac{M - \bar{p}}{M_K - \bar{p}} \right|, \left| \frac{m - \bar{p}}{m_K - \bar{p}} \right|, 1 \right\}. \quad (6.20)$$

Es wird gerade so konstruiert, dass es den Mittelwert erhält und das Maximumprinzip $\hat{p}(\mathbf{x}) \in [m, M]$ für $\mathbf{x} \in K$ erfüllt. Im Fall $|M - \bar{p}| < |M_K - \bar{p}|$ werden Overshoots unterdrückt, falls $|m - \bar{p}| < |m_K - \bar{p}|$ werden Undershoots vermieden. Für konstantes $p = \bar{p}$ ist der Gewichtungssparameter θ nicht definiert. Da in diesem Fall das Maximumprinzip bereits erfüllt ist, setzen wir dann $\hat{p} = p$.

Da wir im Weiteren das Patankar-Verfahren mit Transformation nutzen wollen, um die Einhaltung solcher Maximumprinzipien zu garantieren, gilt es insbesondere die Positivität zu garantieren, also Undershoots zu verhindern. In diesem Fall gilt die Abschätzung $|m - \bar{p}| < |m_K - \bar{p}|$ und damit

$$\hat{p}(\mathbf{x}) = \bar{p} + \left| \frac{m - \bar{p}}{m_K - \bar{p}} \right| (p(\mathbf{x}) - \bar{p}).$$

Da in unserem konkreten Fall p ein bilineares Polynom ist, werden Maximum und Minimum an einer der vier Ecken des Elements angenommen. Sei $\mathbf{x}_m$ die Koordinaten der Ecke mit $p(\mathbf{x}_m) = m_K < m$. Dann gilt wegen $m_K < m \leq \bar{p}$ die Gleichheit

$$\hat{p}(\mathbf{x}_m) = \bar{p} + \left| \frac{m - \bar{p}}{m_K - \bar{p}} \right| (m_K - \bar{p}) = m.$$

Da $\hat{p}$ durch die Werte an den vier Eckpunkten bestimmt ist, können wir durch explizites Setzen von $\hat{p}(\mathbf{x}_m) = m$ vermeiden, dass Rundungsfehler die Einhaltung der Positivität stören. Die Werte an den anderen Ecken können einfach gemäß (6.20) berechnet werden.

Entscheidend für den Einsatz des Limiters ist, dass die Approximationsordnung des DG-Verfahrens dadurch nicht verringert wird. Das dies tatsächlich nicht der Fall ist, wird in [ZS10] für den eindimensionalen Fall bewiesen. Der Beweis lässt sich allerdings direkt auf den zweidimensionalen Fall übertragen.

Mit dem Limiter (6.20) haben wir nun die Möglichkeit das Maximumprinzip zu erfüllen, solange es uns gelingt die Schranken im Zellmittel einzuhalten. Wie wir dazu die Idee des Patankar-Verfahrens auf die DG-Diskretisierung übertragen können, sehen wir im nächsten Abschnitt.

6.1.4 Patankar-DG

Jetzt betrachten wir statt der gewöhnlichen Differentialgleichung (6.10) die partielle Differentialgleichung

$$\partial_t \alpha(\mathbf{x}, t) = P(\alpha(\mathbf{x}, t), \mathbf{x}) - D(\alpha(\mathbf{x}, t), \mathbf{x}),$$

wobei die einzelnen Größen nun auch räumlich variieren. Wie bei (6.2) ergibt sich die zugehörige DG-Diskretisierung dann als

$$(\partial_t \alpha_h, w_h)_K = (P, w_h)_K - (D, w_h)_K.$$

Nun übertragen wir die Idee des Patankar-Verfahrens mit dem Ziel den Mittelwert innerhalb geforderter Schranken zu halten. Dies gelingt mit dem Verfahren: Finde für $t > 0$ ein $\alpha_h(t) \in W_h$ mit

$$(\alpha_h^{n+1}, w_h)_K = (\alpha_h^n, w_h)_K + \Delta t (P^n, w_h)_K - \Delta t (D^n, w_h)_K \frac{(\alpha_h^{n+1}, w_h)_K}{(\alpha_h^n, w_h)_K} \quad \forall w_h \in W_h, \quad (6.21)$$

welches wir im Folgenden als Patankar-DG-Verfahren bezeichnen werden. Nach einer Umformung wie zuvor lautet die direkte Verfahrensvorschrift dann

$$(\alpha_h^{n+1}, w_h)_K = \left(1 + \Delta t \frac{(D^n, w_h)_K}{(\alpha_h^n, w_h)_K} \right)^{-1} ((\alpha_h^n, w_h)_K + \Delta t (P^n, w_h)_K). \quad (6.22)$$

Ebenso können wir das Verfahren in Matrix-Vektor-Notation schreiben. Mit dem Ansatz $\alpha_h^{n+1} = \sum_{j=1}^4 \alpha_j^{n+1} w_h^j$ lautet es dann

$$\boldsymbol{\alpha}^{n+1} = \mathbf{M}^{-1} \mathbf{F}(D^n) (\mathbf{M} \boldsymbol{\alpha}^n + \Delta t \mathbf{P}^n), \quad (6.23)$$

wobei $\boldsymbol{\alpha}^{n+1}$ den Koeffizientenvektor, $\mathbf{M}$ die Massenmatrix und $\mathbf{P}$ die rechte Seite wie in (6.7) bezeichnen. Die Diagonalmatrix $\mathbf{F}$ ist durch

$$\mathbf{F}(D) = \text{diag}\{f_1(D), \dots, f_4(D)\}$$

mit

$$f_j(D) = \left(1 + \Delta t (D^n, w_h^j)_K / (\alpha_h^n, w_h^j)_K \right)^{-1}$$

für $j = 1, \dots, 4$ gegeben. Auch hier sind die einzelnen Matrizen und Vektoren wieder vom Element K abhängig, auch wenn die Bezeichnungen dies nicht widerspiegeln.

Wir werden nun zeigen, dass das Patankar-DG-Verfahren (6.21) in Kombination mit dem Limiter (6.20) die Einhaltung von Maximumprinzipien garantieren kann. Wichtig für die Nutzung des Limiters ist, dass das Patankar-DG-Verfahren die Beschränktheit im Mittel garantiert. Mit der abkürzenden Schreibweise $\bar{\alpha} = \int_K \alpha_h d\mathbf{x} / |K|$ für den Mittelwert und $w_h = 1$ in (6.22) ergibt sich das Verfahren für die Mittelwerte als

$$\bar{\alpha}^{n+1} = \frac{\bar{\alpha}^n (\bar{\alpha}^n + \Delta t \bar{P}^n)}{\bar{\alpha}^n + \Delta t \bar{D}^n}. \quad (6.24)$$

Außerdem bemerken wir noch, dass sich die Transformation (6.16) wegen

$$\bar{\alpha} = \frac{1}{|K|} \int_K \alpha d\mathbf{x} = \frac{1}{|K|} \int_K (1 - \beta) d\mathbf{x} = \frac{1}{|K|} \int_K 1 d\mathbf{x} - \frac{1}{|K|} \int_K \beta d\mathbf{x} = 1 - \bar{\beta} \quad (6.25)$$

auch direkt auf die Mittelwerte überträgt.

Wie in Abschnitt 6.1.2 betrachten wir nun das Verfahren zur Lösung von

$$\partial_t \alpha(\mathbf{x}, t) = f_\alpha(\alpha(\mathbf{x}, t), p(\mathbf{x})). \quad (6.26)$$

Den Quellterm (2.32) schreiben wir dazu in der Form

$$f_\alpha(\alpha(\mathbf{x}, t), p(\mathbf{x})) = C(p(\mathbf{x}))(1 - \alpha(\mathbf{x}, t))^r \alpha(\mathbf{x}, t)^s.$$

Analog zu Lemma 6.2 und Lemma 6.3 ist dann auch hier die Einhaltung des Maximumprinzips $0 \leq \alpha_h^n \leq 1$ für alle Zeiten t^n sichergestellt.

Satz 6.1. *Wir betrachten das Patankar-Verfahren (6.21) mit Transformation (6.16) und anschließender Limitierung nach (6.20) zur Lösung von (6.26). Das zugehörigen Quellterm-Splitting wählen wir gemäß (6.15) und (6.18). Außerdem sei $C(p)$ zellweise konstant. Gilt $0 \leq \bar{\alpha}_h^0 \leq 1$, so erfüllt das Verfahren für $C(p) \geq 0$ die Eigenschaft*

$$1 \geq \bar{\alpha}_h^{n+1} \geq \bar{\alpha}_h^n \geq 0$$

und für $C(p) < 0$ gilt

$$0 \leq \bar{\alpha}_h^{n+1} \leq \bar{\alpha}_h^n \leq 1.$$

Insbesondere ist damit die Einhaltung des Maximumprinzips $0 \leq \alpha_h^n \leq 1$ für alle Zeitpunkte t^n gesichert.

Beweis. Der Beweis erfolgt wieder induktiv. Wir zeigen hier, dass aus $0 \leq \bar{\alpha}_h^n \leq 1$ auch $0 \leq \bar{\alpha}_h^{n+1} \leq 1$ folgt. Dabei ist der Induktionsanfang durch die Voraussetzung $0 \leq \bar{\alpha}_h^0 \leq 1$ gesichert. Mit $0 \leq \bar{\alpha}_h^n \leq 1$ gilt auch für die Transformierte $0 \leq \bar{\beta}_h^n \leq 1$. Nach Definition von P und D durch (6.15) und P_β und D_β durch (6.18) sind, alle in (6.24) auftretenden Terme positiv, unabhängig von $C(p)$. Der erste Teil der Behauptung folgt für $C(p) \leq 0$ wie in Lemma 6.2. Anderenfalls wie in Lemma 6.3, wobei sich die Transformation nach (6.25) direkt auf die Mittelwerte überträgt. Da so die Einhaltung des Maximumprinzips für die Mittelwerte gesichert ist, folgt nach Anwendung des Limiters (6.20) wie behauptet $0 \leq \alpha_h^{n+1} \leq 1$. $\square$

Die im Satz genutzte Voraussetzung, dass $C(p)$ elementweise konstant sein soll, ist in unserem konkreten Fall mit zellweise konstanten Druckapproximationen nach (3.12c) keine Einschränkung. Falls $C(p)$ sowohl positive als auch negative Werte über einer Zelle annimmt, so muss die Entscheidung über das Umschalten zur transformierten Gleichung anhand von $\bar{f}_\alpha$ getroffen werden. Ist dieser Wert negativ, so gilt es die Positivität zu sichern, ist der Wert positiv, so nutzen wir die transformierte Gleichung zur Einhaltung der oberen Schranke.

Wir wollen nun noch Anmerkungen zur Umsetzung und Implementierung des hier vorgestellten Patankar-Verfahrens machen. Zunächst einmal ist das Verfahren (6.21) für $\alpha_h^n = 0$ nicht definiert. Wie im Fall der gewöhnlichen Differentialgleichung (6.13) ist dies aber unproblematisch. Denn in diesem Fall gilt nach Definition des Quellterms auch $\alpha_h^{n+1} = 0$. Dies kann also durch eine entsprechende Abfrage gesetzt werden.

Des weiteren wollen wir diskutieren, warum die etwas aufwendigere Variante des Verfahrens mit Transformation gegenüber dem von uns in [KMW11] vorgestellten Splitting zu bevorzugen ist. Dazu betrachten wir die diskrete Form des Patankar-Verfahrens (6.23) und betrachten zunächst den Fall der zu erhaltenden Positivität. Dabei gilt

$$\mathbf{M}\boldsymbol{\alpha}^{n+1} = \mathbf{F}(D^n) (\mathbf{M}\boldsymbol{\alpha}^n + \Delta t \mathbf{P}^n), \quad (6.27)$$

mit $\mathbf{M}, \mathbf{F}(D^n), \mathbf{P}^n, \boldsymbol{\alpha}^n \geq 0$, aufgrund von (6.15) und der Positivität der verwendeten Basis (6.6). Hier bezeichnen wir Matrizen und Vektoren als positiv, falls alle Komponenten positiv sind. Für den Zellmittelwert gilt

$$\frac{1}{|K|} \int_K \alpha_h^{n+1} d\mathbf{x} = \frac{1}{|K|} \sum_{i=1}^4 \int_K \alpha_h^{n+1} w_h^i d\mathbf{x} = \frac{1}{|K|} \sum_{j=1}^4 [\mathbf{M}\boldsymbol{\alpha}^{n+1}]_j,$$

wonach er sich durch Summation der Komponenten der rechten Seite von (6.27) und anschließender Multiplikation mit dem positiven Zellvolumen ergibt. Damit ist die Positivität des Mittelwerts auch numerisch, also unter Berücksichtigung von Rundungsfehlern, garantiert. Bei entsprechender Umsetzung des Limiters gilt dann damit $\alpha_h^{n+1} \geq 0$. Die Einhaltung der oberen Schranke sichern wir, indem wir mit dem Limiter die Positivität der Transformaten β_h^{n+1} garantieren und erst nach der Limitierung die Rücktransformation auf α_h^{n+1} durchführen.

Nutzen wir dagegen das Quellterm-Splitting (6.19) zur Einhaltung der oberen Schranke, so können wir nicht garantieren, dass der Mittelwert $\bar{\alpha}_h^{n+1} \leq 1$ erfüllt. Aufgrund von Rundungsfehlern gilt hier lediglich $\sum_{j=1}^4 [\mathbf{F}(D^n) (\mathbf{M}\boldsymbol{\alpha}^n + \Delta t \mathbf{P}^n)]_j \approx |K|$ und damit $|K|^{-1} \sum_j [\mathbf{M}\boldsymbol{\alpha}^{n+1}]_j = 1 + \delta$. Ob dieses δ positiv oder negativ ist, können wir nicht kontrollieren. Da aber die Einhaltung der Schranken essentiell für die Limitierung ist, verwenden wir zur Einhaltung der oberen Schranke die Transformation (6.16). Denn hier können auch Rundungsfehler nichts an der Positivität des Mittelwerts ändern.

Abschließend zeigen wir in Abbildung 6.3 noch das Ergebnis einer Testrechnung mit dem Patankar-DG-Verfahren zur Lösung von (6.26). Dabei wurde $\Omega = [-1.5, 1.5]^2$ durch ein Netz mit 40×40 äquidistanten Zellen diskretisiert. Für $x < 0$ wurde für den Quellterm in Anlehnung an (2.27) $r = 1/3$ und $s = 2/3$ gewählt, für $x \geq 0$ wurde wie in (2.29)

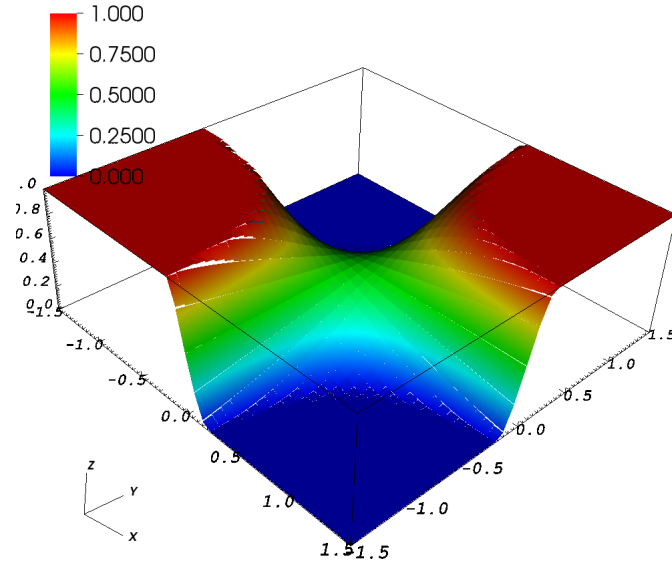


Abbildung 6.3: Testrechnung mit dem Patankar-DG-Verfahren für (6.26).

$r = s = 1/2$ gesetzt. Innerhalb des Quellterms wurde $C(p(x, y)) = xy$ gesetzt. Die Lösung wird zum Zeitpunkt $t = 5$ dargestellt, wobei als Zeitschrittweite $\Delta t = 0.1$ und als Anfangsdaten $\alpha_h^0(\mathbf{x}) = 0.5$ verwendet wurde. Wir erkennen deutlich das erwartete Verhalten, ein Anwachsen von α_h falls $xy > 0$ und ein Fallen für $xy < 0$. Die Einhaltung des Maximumprinzips ist dabei sichtlich gewährleistet.

6.1.5 Maximumprinzip bei homogenen Transportgleichungen

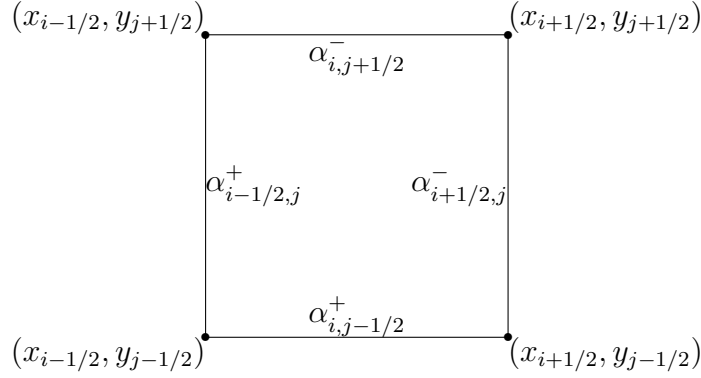
Wie wir gesehen haben, können wir mit dem Limiter (6.20) ein Maximumprinzip erfüllen, falls der zugehörige Mittelwert das Maximumprinzip bereits erfüllt. Eine Bedingung, die dies im Fall der DG-Diskretisierung für homogene Transportgleichungen mit divergenzfreiem Geschwindigkeitsfeld sichert, wurde in [ZS10] vorgestellt. Bevor wir dieses Verfahren unseren Anforderungen entsprechend erweitern, werden wir nun die wichtigsten Resultate aus [ZS10] wiederholen. Dazu betrachten wir die homogene Transportgleichung

$$\partial_t \alpha + \mathbf{u} \cdot \nabla \alpha = 0, \quad \nabla \cdot \mathbf{u} = 0 \quad (6.28)$$

mit divergenzfreiem Geschwindigkeitsfeld. Wie zuvor in Abschnitt 2.3.1 erhalten wir aus der Charakteristiken-Theorie, dass die Lösung dieser Gleichung entlang der Charakteristiken konstant verläuft, womit sie auch ein entsprechendes Maximumprinzip erfüllt.

Mit dem expliziten Euler-Verfahren zur Zeitintegration ergibt sich die Verfahrensvorschrift des DG-Verfahrens für (6.28) nach (6.4) als

$$(\alpha_h^{n+1}, w_h)_K = (\alpha_h^n, w_h)_K - \Delta t \langle H(\alpha_h^{-,n}, \alpha_h^{+,n}, \mathbf{u}^n \cdot \mathbf{n}), w_h^- \rangle_{\partial K} + \Delta t (\alpha_h^n \mathbf{u}^n, \nabla w_h)_K. \quad (6.29)$$

Abbildung 6.4: Einschränkung von α_{ij} entlang der Zellränder.

Die Resultate aus [ZS10] setzen die Verwendung von Rechtecksgittern voraus. Da dies die Notation vereinfacht und um eine bessere Vergleichbarkeit mit [ZS10] zu erzielen, verwenden wir ab jetzt die dort eingeführte Notation. Dazu sei $K = K_{ij} = [x_{i-1/2}, x_{i+1/2}] \times [y_{j-1/2}, y_{j+1/2}]$ und $\alpha_{ij}^n = \alpha_h^n|_{K_{ij}}$. Die Einschränkungen von α_{ij}^n entlang der Zellkanten ist aus Abbildung 6.4 ersichtlich. Dabei verzichten wir zur Verkürzung der Schreibweise auf den Index n . Der Mittelwert von α_h über K_{ij} ist in dieser Notation durch

$$\bar{\alpha}_{ij} = \frac{1}{\Delta x \Delta y} \int_{y_{j-1/2}}^{y_{j+1/2}} \int_{x_{i-1/2}}^{x_{i+1/2}} \alpha_h(x, y) dx dy$$

gegeben. Weiter bezeichnen wir die Stützstellen der Gaußquadratur auf $[x_{i-1/2}, x_{i+1/2}]$ mit x_i^β für $\beta = 1, \dots, L$. Analog nennen wir die Stützstellen auf $[y_{j-1/2}, y_{j+1/2}]$ dann y_j^β für $\beta = 1, \dots, L$. Damit sind die Stützstellen der Gauß-Quadratur etwa auf der linken Kante durch $(x_{i-1/2}, y_j^\beta)$ gegeben, die zugehörigen Werte in diesen Punkten bezeichnen wir mit $\alpha_{i-1/2,\beta}^+ = \alpha_{i-1/2,j}^+(y_j^\beta)$. Die Notation für die anderen Kanten folgt entsprechend. Die zugehörigen Gewichte ω_β haben die Eigenschaft $\sum_{\beta=1}^L \omega_\beta = 1$, wodurch sich für hinreichend gewähltes L die Gleichheit

$$\int_{y_{j-1/2}}^{y_{j+1/2}} \alpha_{i-1/2,j}^+(y) dy = \Delta y \sum_{\beta=1}^L \omega_\beta \alpha_{i-1/2,\beta}^+$$

ergibt und Polynome bis zum Grad $2L - 1$ exakt integriert werden. Neben der Gauß-Quadratur, benötigen wir im Folgenden auch die Gauß-Lobatto-Quadratur. Genauer gesagt, sogar nur die entsprechenden Gewichte $\hat{\omega}_\gamma$. Im Unterschied zur Gauß-Quadratur ist die Gauß-Lobatto-Quadratur nur für Polynome vom Grad $2L - 3$ exakt. Dennoch spielt sie im Folgenden eine wichtige Rolle, da die zugehörigen Stützstellen auch die Intervallgrenzen enthalten.

Setzen wir nun $w_h = 1$ in (6.29) so erhalten wir die Evolution der Mittelwerte als

$$\begin{aligned}\bar{\alpha}_{ij}^{n+1} &= \bar{\alpha}_{ij}^n - \frac{\Delta t}{\Delta x \Delta y} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \left[H(\alpha_{i+\frac{1}{2},j}^-, \alpha_{i+\frac{1}{2},j}^+, u_{i+\frac{1}{2},j}) - H(\alpha_{i-\frac{1}{2},j}^-, \alpha_{i-\frac{1}{2},j}^+, u_{i-\frac{1}{2},j}) \right] dy \\ &\quad - \frac{\Delta t}{\Delta x \Delta y} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left[H(\alpha_{i,j+\frac{1}{2}}^-, \alpha_{i,j+\frac{1}{2}}^+, v_{i,j+\frac{1}{2}}) - H(\alpha_{i,j-\frac{1}{2}}^-, \alpha_{i,j-\frac{1}{2}}^+, v_{i,j-\frac{1}{2}}) \right] dx.\end{aligned}$$

Vorausgesetzt die Geschwindigkeit ist polynomial, so sind die hier auftretenden Integranden Polynome und können durch Anwendung einer hinreichend genauen Quadratur-Formel exakt berechnet werden. Die Gleichung lautet dann unter Verwendung der Gauß-Quadratur

$$\begin{aligned}\bar{\alpha}_{ij}^{n+1} &= \bar{\alpha}_{ij}^n - \lambda_1 \sum_{\beta=1}^k \omega_{\beta} \left[H(\alpha_{i+\frac{1}{2},\beta}^-, \alpha_{i+\frac{1}{2},\beta}^+, u_{i+\frac{1}{2},\beta}) - H(\alpha_{i-\frac{1}{2},\beta}^-, \alpha_{i-\frac{1}{2},\beta}^+, u_{i-\frac{1}{2},\beta}) \right] \\ &\quad - \lambda_2 \sum_{\beta=1}^k \omega_{\beta} \left[H(\alpha_{\beta,j+\frac{1}{2}}^-, \alpha_{\beta,j+\frac{1}{2}}^+, v_{\beta,j+\frac{1}{2}}) - H(\alpha_{\beta,j-\frac{1}{2}}^-, \alpha_{\beta,j-\frac{1}{2}}^+, v_{\beta,j-\frac{1}{2}}) \right],\end{aligned}\tag{6.30}$$

wobei wir hier die Abkürzungen $\lambda_1 = \Delta t / \Delta x$ und $\lambda_2 = \Delta t / \Delta y$ verwenden. Den Mittelwert $\bar{\alpha}_{ij}^n$ auf der rechten Seite dieser Gleichung wollen wir nun weiter umformen. Dazu bezeichnen wir $\bar{\alpha}_{\beta,j}$ bzw. $\bar{\alpha}_{i,\beta}$ die Mittelwerte entlang der durch die Stützstellen verlaufenden und zu den Kanten parallelen Strecken. Es gilt

$$\begin{aligned}\bar{\alpha}_{\beta,j} &= \frac{1}{\Delta y} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \alpha_{ij}(x_i^{\beta}, y) dy, \\ \bar{\alpha}_{i,\beta} &= \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \alpha_{ij}(x, y_j^{\beta}) dx,\end{aligned}$$

womit wir die zwei Darstellungen

$$\begin{aligned}\bar{\alpha}_{ij} &= \sum_{\beta=1}^k \omega_{\beta} \bar{\alpha}_{\beta,j}, \\ \bar{\alpha}_{ij} &= \sum_{\beta=1}^k \omega_{\beta} \bar{\alpha}_{i,\beta}\end{aligned}\tag{6.31}$$

für den Mittelwert erhalten. Die erste Gleichung gilt wegen

$$\begin{aligned}\bar{\alpha}_{ij} &= \frac{1}{\Delta x \Delta y} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \alpha_{ij} dx dy = \frac{1}{\Delta x \Delta y} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \left(\sum_{\beta=1}^k \Delta x \omega_{\beta} \alpha_{ij}(x_i^{\beta}, y) \right) dy \\ &= \sum_{\beta=1}^k \omega_{\beta} \left(\frac{1}{\Delta y} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \alpha_{ij}(x_i^{\beta}, y) dy \right) = \sum_{\beta=1}^k \omega_{\beta} \bar{\alpha}_{\beta,j}\end{aligned}$$

und die zweite folgt analog. Setzen wir zur Verkürzung der Schreibweise noch $\lambda = a_1\lambda_1/(a_1\lambda_1 + a_2\lambda_2)$ und $\mu = a_2\lambda_2/(a_1\lambda_1 + a_2\lambda_2)$ mit $a_1 = \max |u|$ und $a_2 = \max |v|$, so können wir mit (6.31) die Entwicklung der Mittelwerte (6.30) in der Form

$$\begin{aligned}\bar{\alpha}_{ij}^{n+1} &= \lambda \sum_{\beta=1}^k \omega_{\beta} \bar{\alpha}_{i,\beta} - \lambda_1 \sum_{\beta=1}^k \omega_{\beta} \left[H(\alpha_{i+\frac{1}{2},\beta}^-, \alpha_{i+\frac{1}{2},\beta}^+, u_{i+\frac{1}{2},\beta}) - H(\alpha_{i-\frac{1}{2},\beta}^-, \alpha_{i-\frac{1}{2},\beta}^+, u_{i-\frac{1}{2},\beta}) \right] \\ &\quad + \mu \sum_{\beta=1}^k \omega_{\beta} \bar{\alpha}_{\beta,j} - \lambda_2 \sum_{\beta=1}^k \omega_{\beta} \left[H(\alpha_{\beta,j+\frac{1}{2}}^-, \alpha_{\beta,j+\frac{1}{2}}^+, v_{\beta,j+\frac{1}{2}}) - H(\alpha_{\beta,j-\frac{1}{2}}^-, \alpha_{\beta,j-\frac{1}{2}}^+, v_{\beta,j-\frac{1}{2}}) \right] \\ &= \lambda \sum_{\beta=1}^k \omega_{\beta} \left[\bar{\alpha}_{i,\beta} - \frac{\lambda_1}{\lambda} \left(H(\alpha_{i+\frac{1}{2},\beta}^-, \alpha_{i+\frac{1}{2},\beta}^+, u_{i+\frac{1}{2},\beta}) - H(\alpha_{i-\frac{1}{2},\beta}^-, \alpha_{i-\frac{1}{2},\beta}^+, u_{i-\frac{1}{2},\beta}) \right) \right] \\ &\quad + \mu \sum_{\beta=1}^k \omega_{\beta} \left[\bar{\alpha}_{\beta,j} - \frac{\lambda_2}{\mu} \left(H(\alpha_{\beta,j+\frac{1}{2}}^-, \alpha_{\beta,j+\frac{1}{2}}^+, v_{\beta,j+\frac{1}{2}}) - H(\alpha_{\beta,j-\frac{1}{2}}^-, \alpha_{\beta,j-\frac{1}{2}}^+, v_{\beta,j-\frac{1}{2}}) \right) \right]\end{aligned}$$

schreiben. Mit der abkürzenden Schreibweise

$$\begin{aligned}H_x^{i,\beta} &= \bar{\alpha}_{i,\beta} - \frac{\lambda_1}{\lambda} \left(H(\alpha_{i+\frac{1}{2},\beta}^-, \alpha_{i+\frac{1}{2},\beta}^+, u_{i+\frac{1}{2},\beta}) - H(\alpha_{i-\frac{1}{2},\beta}^-, \alpha_{i-\frac{1}{2},\beta}^+, u_{i-\frac{1}{2},\beta}) \right), \\ H_y^{\beta,j} &= \bar{\alpha}_{\beta,j} - \frac{\lambda_2}{\mu} \left(H(\alpha_{\beta,j+\frac{1}{2}}^-, \alpha_{\beta,j+\frac{1}{2}}^+, v_{\beta,j+\frac{1}{2}}) - H(\alpha_{\beta,j-\frac{1}{2}}^-, \alpha_{\beta,j-\frac{1}{2}}^+, v_{\beta,j-\frac{1}{2}}) \right)\end{aligned}$$

erhalten wir schließlich die Darstellung

$$\bar{\alpha}_{ij}^{n+1} = \lambda \sum_{\beta=1}^k \omega_{\beta} H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_{\beta} H_y^{\beta,j}. \quad (6.32)$$

Nun werden wir sehen, unter welcher Bedingung an die Zeitschrittweite wir die Beschränktheit des Mittelwerts in der Form $\bar{\alpha}_{ij} \in [m, M]$ erfüllen können. Dazu benötigen wir noch die folgenden zwei Lemmas. In Hinblick auf Abschnitt 6.1.6 bemerken hier bereits, dass die Beweise beider Lemmas unabhängig von der Divergenzfreiheit des Geschwindigkeitsfeldes sind. Darüber hinaus verdoppelt sich zudem die zulässige Zeitschrittweite im Fall konstanter Geschwindigkeit.

Lemma 6.4. *Wir betrachten das eindimensionale Verfahren*

$$\alpha_j^{n+1} = \alpha_j - \nu (H(\alpha_j, \alpha_{j+1}, u_j) - H(\alpha_{j-1}, \alpha_j, u_{j-1}))$$

mit $\nu = \Delta t / \Delta x$, dem Lax-Friedrichs-Fluss H aus (6.3) und $-a \leq u_{j-1}, u_j \leq a$. Weiter gelte $m \leq \alpha_{j-1}, \alpha_j, \alpha_{j+1} \leq M$. Dann ist die Ungleichung

$$m - \nu (H(m, m, u_j) - H(m, m, u_{j-1})) \leq \alpha_j^{n+1} \leq M - \nu (H(M, M, u_j) - H(M, M, u_{j-1})),$$

erfüllt, falls die Zeitschrittweite der CFL-Bedingung $a\nu \leq \frac{1}{2}$ genügt.

Beweis. Schreiben wir α_j^{n+1} als Linearkombination der α_i so sind die zugehörigen Linearfaktoren positiv, falls die CFL-Bedingung erfüllt wird. Die Abschätzung folgt dann aus der Beschränktheit der α_i , siehe [ZS10]. $\square$

Lemma 6.5. *Unter der CFL-Bedingung $a_1\lambda_1 + a_2\lambda_2 \leq \frac{1}{2} \min_{\gamma=1,\dots,N} \hat{\omega}_\gamma$ gelten*

$$\begin{aligned} \left[1 - \frac{\lambda_1}{\lambda}(u_{i+\frac{1}{2},\beta} - u_{i-\frac{1}{2},\beta})\right] m &\leq H_x^{i,\beta} \leq \left[1 - \frac{\lambda_1}{\lambda}(u_{i+\frac{1}{2},\beta} - u_{i-\frac{1}{2},\beta})\right] M, \\ \left[1 - \frac{\lambda_2}{\mu}(v_{\beta,j+\frac{1}{2}} - v_{\beta,j-\frac{1}{2}})\right] m &\leq H_y^{\beta,j} \leq \left[1 - \frac{\lambda_2}{\mu}(v_{\beta,j+\frac{1}{2}} - v_{\beta,j-\frac{1}{2}})\right] M. \end{aligned}$$

Hier bezeichnet $\hat{\omega}_\gamma$ für $i = 1, \dots, N$ die Gewichte der Gauß-Lobatto-Quadratur und N ist so gewählt, dass α_{ij}^n exakt integriert werden kann.

Beweis. Mit der Gauß-Lobatto-Quadratur lassen sich $H_x^{i,\beta}$ und $H_y^{\beta,j}$ als Konvexkombination eindimensionaler Verfahren schreiben. Die Abschätzungen folgen dann aus Lemma 6.4, siehe [ZS10]. $\square$

Nun können wir das wesentliche Resultat aus [ZS10] wiedergeben. Es besagt, dass unter Einhaltung der CFL-Bedingung

$$a_1\lambda_1 + a_2\lambda_2 \leq \frac{1}{2} \min_{\gamma=1,\dots,N} \hat{\omega}_\gamma, \quad (6.33)$$

das Maximumprinzip erfüllt wird. Konkret sind die von uns verwendeten bilinearen Polynome bei Restriktion auf eine Variable linear, weshalb $N = 2$ (Trapezregel) für die Gauß-Lobatto-Quadratur ausreichend ist. Demnach gilt $\hat{\omega}_1 = \hat{\omega}_2 = 1/2$ und die CFL-Bedingung wird zu

$$a_1\lambda_1 + a_2\lambda_2 \leq \frac{1}{4}. \quad (6.34)$$

Satz 6.2. *Angenommen $m \leq \bar{\alpha}_{ij} \leq M$ und $m \leq \alpha_{ij}(x, y) \leq M$ für alle $(x, y) \in K_{ij}$. Dann ist das Maximumprinzip $m \leq \bar{\alpha}_{ij}^{n+1} \leq M$ für das Verfahren (6.30) unter der CFL-Bedingung $a_1\lambda_1 + a_2\lambda_2 \leq \frac{1}{2} \min_{\gamma=1,\dots,N} \hat{\omega}_\gamma$ erfüllt.*

Beweis. Wir zeigen die Einhaltung der oberen Schranke. Nach (6.32) gilt

$$\bar{\alpha}_{ij}^{n+1} = \lambda \sum_{\beta=1}^k \omega_\beta H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_\beta H_y^{\beta,j}.$$

Mit Lemma 6.5 und Umformungen erhalten wir die Abschätzung

$$\begin{aligned} \bar{\alpha}_{ij}^{n+1} &\leq \lambda \sum_{\beta=1}^k \omega_\beta \left(1 - \frac{\lambda_1}{\lambda}(u_{i+\frac{1}{2},\beta} - u_{i-\frac{1}{2},\beta})\right) M + \mu \sum_{\beta=1}^k \omega_\beta \left(1 - \frac{\lambda_2}{\mu}(v_{\beta,j+\frac{1}{2}} - v_{\beta,j-\frac{1}{2}})\right) M \\ &= M - M \sum_{\beta=1}^k \omega_\beta \lambda_1 (u_{i+\frac{1}{2},\beta} - u_{i-\frac{1}{2},\beta}) - M \sum_{\beta=1}^k \omega_\beta \lambda_2 (v_{\beta,j+\frac{1}{2}} - v_{\beta,j-\frac{1}{2}}) \\ &= M \left[1 - \frac{\Delta t}{\Delta x \Delta y} \left(\sum_{\beta=1}^k (u_{i+\frac{1}{2},\beta} - u_{i-\frac{1}{2},\beta}) \omega_\beta \Delta y + \sum_{\beta=1}^k (v_{\beta,j+\frac{1}{2}} - v_{\beta,j-\frac{1}{2}}) \omega_\beta \Delta x \right) \right]. \end{aligned}$$

Die Summation der Geschwindigkeiten können wir auch über die Divergenz des Geschwindigkeitsfeldes ausdrücken. Es gilt

$$\begin{aligned}
 & \sum_{\beta=1}^k (u_{i+\frac{1}{2},\beta} - u_{i-\frac{1}{2},\beta}) \omega_\beta \Delta y + \sum_{\beta=1}^k (v_{\beta,j+\frac{1}{2}} - v_{\beta,j-\frac{1}{2}}) \omega_\beta \Delta x \\
 &= \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} (u(x_{i+\frac{1}{2}}, y) - u(x_{i-\frac{1}{2}}, y)) dy + \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} (v(x, y_{j+\frac{1}{2}}) - v(x, y_{j-\frac{1}{2}})) dx \\
 &= \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \partial_x u \, dx dy + \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \partial_y v \, dy dx \\
 &= \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \nabla \cdot \mathbf{u} \, dx dy.
 \end{aligned}$$

Damit erhalten wir die Abschätzung

$$\bar{\alpha}_{ij}^{n+1} \leq M \left[1 - \frac{\Delta t}{\Delta x \Delta y} \left(\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \nabla \cdot \mathbf{u} \, dx dy \right) \right] \quad (6.35)$$

und $\bar{\alpha}_{ij}^{n+1} \leq M$ folgt aus $\nabla \cdot \mathbf{u} = 0$. Der Beweis für $m \leq \bar{\alpha}_{ij}^{n+1}$ verläuft analog, indem die Abschätzung

$$m \left[1 - \frac{\Delta t}{\Delta x \Delta y} \left(\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \nabla \cdot \mathbf{u} \, dx dy \right) \right] \leq \bar{\alpha}_{ij}^{n+1} \quad (6.36)$$

hergeleitet wird. $\square$

Abschließend wollen wir auch hier den obigen Testfall (6.8) betrachten, bei dem die Lösung (6.9) das Maximumprinzip $-1 \leq C \leq 1$ erfüllt. Unter Einhaltung der CFL-Bedingung (6.33) wird dann auch bei Verwendung des Limiters (6.20) die numerische Lösung entsprechend beschränkt sein. Das dabei die Ordnung des Verfahrens erhalten bleibt können wir Tabelle (6.2) entnehmen.

6.1.6 Patankar-DG für Kavitation

Wir wollen jetzt das Resultat aus Satz 6.2 in modifizierter Form nutzen, um in Kombination mit dem Patankar-DG-Verfahren, die Einhaltung des Maximumprinzips für unseren Fall der Kavitation (2.33) zu sichern. Erweiterungen der in [ZS10] vorgestellten Idee für andere Gleichungen, finden sich etwa im Übersichtsartikel [ZS11]. Allerdings ist uns keine derartige Erweiterung bekannt, die in unserem Fall anwendbar wäre.

Nun betrachten wir die inhomogene Transportgleichung (2.33c) mit $\nabla \cdot \mathbf{u} = f_\alpha$. Den Quellterm schreiben wir dabei wieder in der Form

$$f_\alpha(\alpha(\mathbf{x}, t), p(\mathbf{x})) = C(p(\mathbf{x}))(1 - \alpha(\mathbf{x}, t))^r \alpha(\mathbf{x}, t)^s,$$

Tabelle 6.2: Fehler und Konvergenzordnung für das Testproblem (6.8) unter Verwendung des Limiters (6.20)

Gitter	$\ C - C_h\ _{L^2(\Omega)}$	
1	Fehler	Ordnung
0	0.471261	-
1	0.121371	1.957103
2	0.025741	2.237296
3	0.005872	2.132134
4	0.001401	2.067223

wobei wir wieder annehmen, dass $C(p)$ zellweise konstant ist. Das zugehörige Patankar-DG-Verfahren lautet in allgemeiner Form nach (6.4) und (6.21),

$$\begin{aligned}
 (\alpha_h^{n+1}, w_h)_K &= (\alpha_h^n, w_h)_K - \Delta t \langle H(\alpha_h^{-,n}, \alpha_h^{+,n}, \mathbf{u}^n \cdot \mathbf{n}), w_h^- \rangle_{\partial K} \\
 &\quad + \Delta t (\alpha_h^n \mathbf{u}^n, \nabla w_h)_K + \Delta t (P(\alpha_h^n), w_h)_K - \Delta t (D(\alpha_h^n), w_h)_K \frac{(\alpha_h^{n+1}, w_h)_K}{(\alpha_h^n, w_h)_K}, \quad (6.37)
 \end{aligned}$$

mit positiven Quelltermen P und D . In der Notation aus Abschnitt 6.1.5 ergibt sich der neue Mittelwert $\bar{\alpha}_{ij}^{n+1}$ innerhalb der Zelle K_{ij} , analog zu (6.30), nun durch

$$\begin{aligned}
 \bar{\alpha}_{ij}^{n+1} &= \bar{\alpha}_{ij}^n - \frac{\Delta t}{\Delta x \Delta y} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \left[H(\alpha_{i+\frac{1}{2},j}^-, \alpha_{i+\frac{1}{2},j}^+, u_{i+\frac{1}{2},j}) - H(\alpha_{i-\frac{1}{2},j}^-, \alpha_{i-\frac{1}{2},j}^+, u_{i-\frac{1}{2},j}) \right] dy \\
 &\quad - \frac{\Delta t}{\Delta x \Delta y} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left[H(\alpha_{i,j+\frac{1}{2}}^-, \alpha_{i,j+\frac{1}{2}}^+, v_{i,j+\frac{1}{2}}) - H(\alpha_{i,j-\frac{1}{2}}^-, \alpha_{i,j-\frac{1}{2}}^+, v_{i,j-\frac{1}{2}}) \right] dx \\
 &\quad + \Delta t \bar{P}^n - \Delta t \bar{D}^n \frac{\bar{\alpha}_{ij}^{n+1}}{\bar{\alpha}_{ij}^n}.
 \end{aligned}$$

Liegt wie im konkreten Fall (2.32) ein Quellterm vor, der sich nicht als Polynom darstellen lässt, so können die zugehörigen Integrale nur approximativ durch Quadraturformeln bestimmt werden. Allerdings wird auch bei der Approximation durch Gauß-Quadratur, aufgrund der Positivität der einzelnen Terme, stets $\bar{P}^n, \bar{D}^n \geq 0$ gelten. Die obige Darstellung unterscheidet sich von (6.30) nur durch das Auftreten der gemittelten Quellterme und kann analog zu (6.32) kurz als

$$\bar{\alpha}_{ij}^{n+1} = \lambda \sum_{\beta=1}^k \omega_{\beta} H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_{\beta} H_y^{\beta,j} + \Delta t \bar{P}^n - \Delta t \bar{D}^n \frac{\bar{\alpha}_{ij}^{n+1}}{\bar{\alpha}_{ij}^n} \quad (6.38)$$

geschrieben werden. Der neue Mittelwert lässt sich dann explizit durch

$$\bar{\alpha}_{ij}^{n+1} = \frac{\bar{\alpha}_{ij}^n (\lambda \sum_{\beta=1}^k \omega_{\beta} H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_{\beta} H_y^{\beta,j} + \Delta t \bar{P}^n)}{\bar{\alpha}_{ij}^n + \Delta t \bar{D}^n} \quad (6.39)$$

berechnen.

Wesentlich für das weitere Vorgehen ist die Abschätzung

$$0 \leq \lambda \sum_{\beta=1}^k \omega_{\beta} H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_{\beta} H_y^{\beta,j} \leq 1 - \frac{\Delta t}{\Delta x \Delta y} \int_{y_j-\frac{1}{2}}^{y_j+\frac{1}{2}} \int_{x_i-\frac{1}{2}}^{x_i+\frac{1}{2}} \nabla \cdot \mathbf{u} \, dx dy. \quad (6.40)$$

Diese ergibt sich mit $m = 0$ und $M = 1$ ohne zusätzliche Voraussetzungen aus (6.35) und (6.36). Unter Berücksichtigung von $\nabla \cdot \mathbf{u} = f_{\alpha}$ können wir dies kürzer als

$$0 \leq \lambda \sum_{\beta=1}^k \omega_{\beta} H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_{\beta} H_y^{\beta,j} \leq 1 - \Delta t \bar{f}_{\alpha}^n \quad (6.41)$$

schreiben.

Wir wollen zunächst die Einhaltung des Maximumprinzips mit dem expliziten Euler-Verfahren untersuchen. Dazu setzen wir in (6.39) die Quellterme als $P = f_{\alpha}$ und $D = 0$. Im Fall $C(p) \leq 0$ gilt es die Positivität des Verfahrens zu sichern. Mit (6.41) erhalten wir aus (6.39) die Abschätzung

$$\bar{\alpha}_{ij}^{n+1} \geq \Delta t \bar{f}_{\alpha}^n.$$

Da $\bar{f}_{\alpha}^n \leq 0$, sichert dies nicht die Positivität. Betrachten wir nun $C(p) \geq 0$ und die Einhaltung der oberen Schranke. Mit (6.39) und (6.41) gilt jetzt

$$\bar{\alpha}_{ij}^{n+1} \leq 1 - \Delta t \bar{f}_{\alpha} + \Delta t \bar{f}_{\alpha}^n = 1.$$

Theoretisch ist die Beschränktheit nach oben also bereits mit dem expliziten Euler-Verfahren gesichert. Aufgrund auftretender Rundungsfehler können wir dies aber numerisch nicht sichern. Vor allem wollen wir es in der Umsetzung des Verfahrens vermeiden Werte die größer 1 sind, explizit auf 1 zu setzen. Daher nutzen wir auch zur Einhaltung der oberen Schranke das Patankar-DG-Verfahren mit Transformation.

Wir bemerken noch, dass das nachfolgend vorgestellte Verfahren auch für Probleme bei denen sich der Quellterm der Advektions-Gleichung und die Divergenzbedingung unterscheiden genutzt werden kann.

Jetzt werden wir die Einhaltung des Maximumprinzips mit dem Patankar-DG-Verfahren in Abhängigkeit des Quellterms f_{α} beweisen. Bleiben die Zellmittelwerte beschränkt, so gilt dies auch für das DG-Polynom nach Anwendung des Limiters (6.20). Um das Maximumprinzip für die Mittelwerte zu zeigen, nehmen wir wieder an, dass bereits $0 \leq \bar{\alpha}_{ij}^n \leq 1$ gilt. Im Fall $C(p) = 0$ befinden wir uns in der Situation von Abschnitt 6.1.5, womit das Maximumprinzip $0 \leq \alpha_{ij}^{n+1} \leq 1$ nach Satz 6.2 mit $m = 0$ und $M = 1$ erfüllt ist.

Falls $C(p) \leq 0$ wählen wir die Quellterme des Patankar-Verfahrens wie zuvor als $P = 0$ und $D = -f_{\alpha}$. Mit (6.39) erhalten wir

$$\bar{\alpha}_{ij}^{n+1} = \frac{\bar{\alpha}_{ij}^n (\lambda \sum_{\beta=1}^k \omega_{\beta} H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_{\beta} H_y^{\beta,j})}{\bar{\alpha}_{ij}^n + \Delta t \bar{D}^n}.$$

Um den Term in Klammern abzuschätzen, nutzen wir (6.41). Mit $f_\alpha = -D$ ergibt sich

$$0 \leq \lambda \sum_{\beta=1}^k \omega_\beta H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_\beta H_y^{\beta,j} \leq 1 + \Delta t \bar{D}^n$$

und wir erhalten nach Multiplikation mit $\bar{\alpha}_{ij}^n / (\bar{\alpha}_{ij}^n + \bar{D}^n) \geq 0$ daraus

$$0 \leq \frac{\bar{\alpha}_{ij}^n (\lambda \sum_{\beta=1}^k \omega_\beta H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_\beta H_y^{\beta,j})}{\bar{\alpha}_{ij}^n + \Delta t \bar{D}^n} \leq \frac{\bar{\alpha}_{ij}^n + \bar{\alpha}_{ij}^n \Delta t \bar{D}^n}{\bar{\alpha}_{ij}^n + \Delta t \bar{D}^n} \leq 1.$$

Damit haben wir $0 \leq \bar{\alpha}_{ij}^{n+1} \leq 1$ für $C(p) < 0$ gezeigt. Besonders die Positivität ist hier wichtig. Die Monotonie der Mittelwerte wie in Abschnitt 6.1.4 lässt sich mit der Abschätzung (6.41) nicht mehr zeigen.

Es bleibt der Fall $C(p) > 0$. Hier transformieren wir die Gleichung wieder gemäß $\alpha = 1 - \beta$ und erhalten so

$$\partial_t \beta - \nabla \cdot (\alpha \mathbf{u}) = -f_\alpha.$$

Entsprechend (6.37) lautet die Verfahrensvorschrift des Patankar-DG-Verfahrens für die transformierte Gleichung

$$\begin{aligned} (\beta_h^{n+1}, w_h)_K &= (\beta_h^n, w_h)_K + \Delta t \langle H(\alpha_h^{-,n}, \alpha_h^{+,n}, \mathbf{u}^n \cdot \mathbf{n}), w_h^- \rangle_{\partial K} \\ &\quad - \Delta t (\alpha_h^n \mathbf{u}^n, \nabla w_h)_K + \Delta t (P_\beta(\alpha_h^n), w_h)_K - \Delta t (D_\beta(\alpha_h^n), w_h)_K \frac{(\beta_h^{n+1}, w_h)_K}{(\beta_h^n, w_h)_K}. \end{aligned}$$

Für die Mittelwerte erhalten wir damit

$$\begin{aligned} \bar{\beta}_{ij}^{n+1} &= \bar{\beta}_{ij}^n + \frac{\Delta t}{\Delta x \Delta y} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \left[H(\alpha_{i+\frac{1}{2},j}^-, \alpha_{i+\frac{1}{2},j}^+, u_{i+\frac{1}{2},j}) - H(\alpha_{i-\frac{1}{2},j}^-, \alpha_{i-\frac{1}{2},j}^+, u_{i-\frac{1}{2},j}) \right] dy \\ &\quad + \frac{\Delta t}{\Delta x \Delta y} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left[H(\alpha_{i,j+\frac{1}{2}}^-, \alpha_{i,j+\frac{1}{2}}^+, v_{i,j+\frac{1}{2}}) - H(\alpha_{i,j-\frac{1}{2}}^-, \alpha_{i,j-\frac{1}{2}}^+, v_{i,j-\frac{1}{2}}) \right] dx \\ &\quad + \Delta t \bar{P}_\beta^n - \Delta t \bar{D}_\beta^n \frac{\bar{\beta}_{ij}^{n+1}}{\bar{\beta}_{ij}^n}. \end{aligned}$$

Mit (6.32) können wir dies wegen $\bar{\beta}_{ij}^n = 1 - \bar{\alpha}_{ij}^n$ auch in der Form

$$\bar{\beta}_{ij}^{n+1} = 1 - (\lambda \sum_{\beta=1}^k \omega_\beta H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_\beta H_y^{\beta,j}) + \Delta t \bar{P}_\beta^n - \Delta t \bar{D}_\beta^n \frac{\bar{\beta}_{ij}^{n+1}}{\bar{\beta}_{ij}^n}$$

schreiben. Äquivalent dazu ergibt sich die Darstellung

$$\bar{\beta}_{ij}^{n+1} = \frac{\bar{\beta}_{ij}^n (1 - (\lambda \sum_{\beta=1}^k \omega_\beta H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_\beta H_y^{\beta,j}) + \Delta t \bar{P}_\beta^n)}{\bar{\beta}_{ij}^n + \Delta t \bar{D}_\beta^n}.$$

Die Patankar-Quellterme wählen wir als $P_\beta = 0$ und $D_\beta = f_\alpha$. Damit erhalten wir aus (6.41) die Abschätzung

$$0 \leq \lambda \sum_{\beta=1}^k \omega_\beta H_x^{i,\beta} + \mu \sum_{\beta=1}^k \omega_\beta H_y^{\beta,j} \leq 1 - \Delta t \bar{D}_\beta^n$$

und diese führt auf

$$1 \geq \frac{\bar{\beta}_{ij}^n}{\bar{\beta}_{ij}^n + \Delta t \bar{D}^n} \geq \bar{\beta}_{ij}^{n+1} \geq \frac{\bar{\beta}_{ij}^n \Delta t \bar{D}^n}{\bar{\beta}_{ij}^n + \Delta t \bar{D}^n} \geq 0.$$

Damit haben wir $0 \leq \bar{\beta}_{ij}^{n+1} \leq 1$ gezeigt und Rücktransformation liefert dann auch für $C(p) > 0$ die Einhaltung des Maximumprinzips $0 \leq \bar{\alpha}_{ij}^{n+1} \leq 1$. Nach Anwendung des Limiters (6.20) gilt dann auch $0 \leq \alpha_{ij}^{n+1} \leq 1$, und wir haben den folgenden Satz bewiesen.

Satz 6.3. *Wir betrachten das Patankar-DG Verfahren (6.37) mit Transformation (6.16) und Limiter (6.20) zur Lösung von (2.33c) mit (2.33b), also*

$$\partial_t \alpha + \nabla \cdot (\alpha \mathbf{u}) = f_\alpha, \quad \nabla \cdot \mathbf{u} = f_\alpha$$

Die Wahl der Quellterme für das Patankar-DG-Verfahren erfolgt gemäß (6.15) und (6.18), dabei erfolgt die Anwendung der Transformation falls $C(p) > 0$. Mit Anfangsdaten $0 \leq \alpha_{ij}^0 \leq 1$ und unter Einhaltung der CFL-Bedingung (6.33) garantiert das Verfahren die Einhaltung des Maximumprinzips $0 \leq \alpha^n \leq 1$ für alle Zeiten t^n .

Auch hier wollen wir wieder abschließend die Anwendbarkeit des Verfahrens anhand eines numerischen Testfalls verdeutlichen. Dazu betrachten wir die inhomogene Transportgleichung

$$\partial_t \alpha + \nabla \cdot (\alpha \mathbf{u}) = f(\alpha) \tag{6.42}$$

mit divergenzbehaftetem Geschwindigkeitsfeld

$$u(x, y) = x, \quad v(x, y) = y,$$

und Quellterm

$$f(\alpha) = 20\alpha(1 - \alpha) + 2\alpha.$$

Analog zu (2.35) lässt sich die Lösung entlang der Charakteristiken bestimmen und wir erhalten

$$\alpha(x, y, t) = \frac{e^{20t} \alpha_0(e^{-t}x, e^{-t}y)}{1 + (e^{20t} - 1) \alpha_0(e^{-t}x, e^{-t}y)}.$$

Mit den Anfangsdaten

$$\alpha^0(x, y) = e^{-50(x^2+y^2)}$$

ergibt sich die Lösung

$$\alpha(x, y, t) = \frac{e^{20t}}{e^{20t} + e^{50e^{-2t}(x^2+y^2)} - 1}.$$

Tabelle 6.3: Fehler und Konvergenzordnung für das Testproblem (6.42) unter Verwendung des Patankar-DG-Verfahrens mit Limiter

Gitter l	$\ \alpha - \alpha_h\ _{L^2(\Omega)}$ Fehler	Ordnung
0	0.207460	-
1	0.025975	2.997646
2	0.005040	2.365663
3	0.001258	2.001838
4	0.000315	1.998648

Diese erfüllt für alle Zeiten t das Maximumprinzip $0 \leq \alpha \leq 1$.

Der Testfall ist so konstruiert, dass neben der Beschränktheit auch eine exakte Lösung vorliegt. Numerische Experimente zeigen, dass wir mit dem expliziten Euler-Verfahren die Einhaltung der oberen Schranke auch bei diesem einfachen Testfall (hier gilt $r = s = 1$) nicht für alle Zeiten erfüllen können. Diese Problematik verschärft sich etwa mit dem Quellterm $f(\alpha) = 20\sqrt{(1-\alpha)\alpha} + 2\alpha$.

Zur Diskretisierung und Ordnungsanalyse verwenden wir wieder verschiedene quadratische Gitter, wobei das l -te Gitter gerade die Maschenweite $\Delta x_l = 0.2/2^l$ hat. In Abbildung 6.5 zeigen wir die numerische Lösung für $t = 0.1, 0.25, 0.5$ auf dem Gitter $l = 2$ und erkennen deutlich die geforderte Beschränktheit. Das das Verfahren weiterhin von zweiter Ordnung im Raum ist, können wir der Tabelle 6.3 entnehmen. Dabei betrachten wir die Fehler jeweils zum Zeitpunkt $t = 0.25$. Die dabei verwendete Zeitschrittweite wurde so klein gewählt, dass der zeitliche Fehler keinen Einfluss auf den räumlichen Fehler hat.

6.2 DG für Diffusion

Wir kommen nun zur Behandlung der diffusiven Terme der Advektions-Diffusionsgleichung (2.37c). Dazu betrachten wir die Gleichung

$$\partial_t C - \Delta C = f.$$

Wie zuvor bei (6.2) erhalten wir nach Multiplikation mit einer hinreichend glatten Testfunktion und partieller Integration die Gleichung

$$(\partial_t C, w)_K - \langle \nabla C \cdot \mathbf{n}, w \rangle_{\partial K} + (\nabla C, \nabla w)_K = (f, w)_K. \quad (6.43)$$

Auch hier erfolgt die Kommunikation der einzelnen Zellen ausschließlich über die Randflüsse und es stellt sich die Frage wie diese bei unstetigem C zu wählen sind. Wir könnten nun versuchen, wie im konvektiven Fall zu verfahren und hier einen zentralen Fluss für die

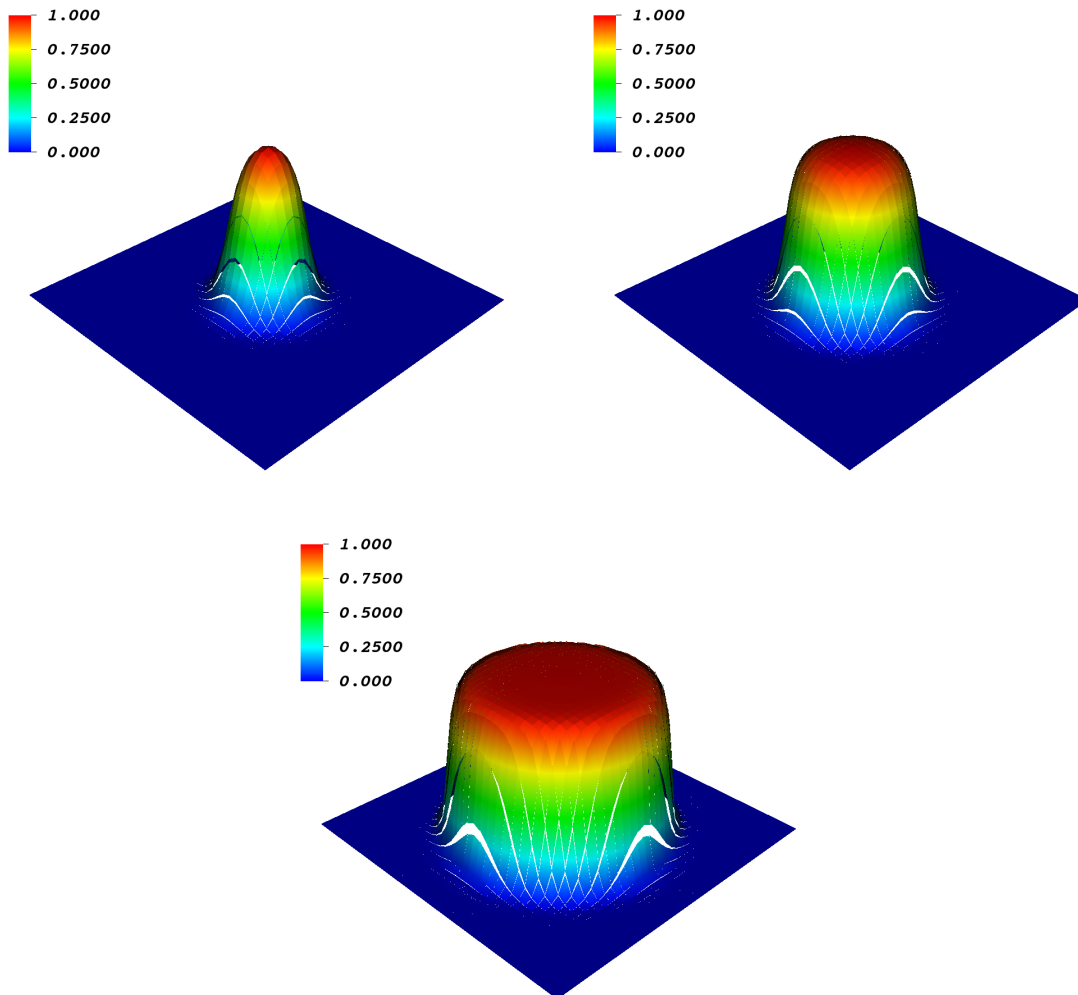


Abbildung 6.5: Numerische Lösung des Testproblems 6.42 für $t = 0.1$ (oben links), $t = 0.25$ (oben rechts) und $t = 0.5$ (unten) unter Verwendung des Patankar-DG-Verfahrens mit Limiter

Gradienten zu nutzen, da es keine ausgezeichnete Richtung gibt. Das dies kein konsistentes Verfahren liefert, wurde in [ZS03] gezeigt. Sinnvolle Verfahrensklassen stellen etwa die Interior-Penalty Verfahren (IPG) [Riv08] und die recovery-based Discontinuous-Galerkin Verfahren (RDG) [LLR07, LL09, Lo11] dar. Bei den IPG Verfahren wird die Stabilität durch die Hinzunahme sogenannter Penalty-Terme zu (6.43) garantiert. Zur Kontrolle dieser Strafterme werden neue, ausschließlich numerische Parameter eingeführt, die entsprechend dem jeweiligen Problem gewählt werden müssen. Um diese Problematik zu umgehen, wurden die RDG-Verfahren entwickelt. Diese werden wir auch im Weiteren nutzen.

6.2.1 Recovery-based-Discontinuous-Galerkin (RDG)

In Anlehnung an (6.43) lautet das RDG-Verfahren in der von uns genutzten Form: Finde für $t > 0$ ein $C_h(t) \in W_h$, so dass

$$(\partial_t C_h, w_h)_K - \langle \nabla R_h \cdot \mathbf{n}, w_h^- \rangle_{\partial K} + (\nabla C_h, \nabla w_h)_K = (f, w_h)_K \quad \forall w_h \in W_h \quad (6.44)$$

erfüllt ist. Um die Problematik der unstetigen Daten bei der Berechnung der Randintegrale zu umgehen, wird bei den RDG-Verfahren zu jeder Kante $e_{ij} = K_i \cap K_j$ eine über $K_i \cup K_j$ glatte Recovery-Funktion R_h bestimmt und mit dieser das entsprechende Integral berechnet. Bei den von uns verwendeten Vierecksgittern sind also vier verschiedene R_h zur Auswertung des Randterms in (6.44) nötig.

Die zu e_{ij} gehörige Recovery-Funktion erhalten wir durch die Forderungen

$$\int_{K_i} R_h w_h d\mathbf{x} = \int_{K_i} C_h w_h d\mathbf{x} \quad \forall w_h \in W_h, \quad (6.45)$$

$$\int_{K_j} R_h w_h d\mathbf{x} = \int_{K_j} C_h w_h d\mathbf{x} \quad \forall w_h \in W_h. \quad (6.46)$$

Damit lässt sich die Recovery-Funktion im schwachen Sinn nicht von der eigentlichen DG-Approximation C_h unterscheiden. Bei Verwendung der bilinearen Testfunktionen (6.5) ergeben sich 8 Gleichungen und somit auch 8 Freiheitsgrade zur Bestimmung von R_h . Daher machen wir den Ansatz $R_h(\mathbf{x}, t) = \sum_{j=1}^8 R_j(t) v_h^j(\mathbf{x})$ und setzen dies, sowie $C_h(\mathbf{x}, t)|_{K_l} = \sum_{j=1}^4 C_j^l(t) w_h^j(\mathbf{x})$, in (6.45) ein. Mit den Bezeichnungen

$$\begin{aligned} \mathbf{c} &= (C_1^i, \dots, C_4^i, C_1^j, \dots, C_4^j)^T, \\ \mathbf{r} &= (R_1, \dots, R_8)^T, \\ \mathbf{M}_{kl} &= \begin{cases} (w_h^k, w_h^l)_{K_i}, & k, l \in \{1, \dots, 4\}, \\ 0, & k \in \{1, \dots, 4\}, l \in \{5, \dots, 8\}, \\ 0, & k \in \{5, \dots, 8\}, l \in \{1, \dots, 4\}, \\ (w_h^{k-4}, w_h^{l-4})_{K_j}, & k, l \in \{5, \dots, 8\}, \end{cases} \\ \mathbf{S}_{kl} &= \begin{cases} (v_h^l, w_h^k)_{K_i}, & k \in \{1, \dots, 4\}, l \in \{1, \dots, 8\}, \\ (v_h^l, w_h^{k-4})_{K_j}, & k \in \{5, \dots, 8\}, l \in \{1, \dots, 8\}, \end{cases} \end{aligned}$$

ergibt sich das lineare Gleichungssystem

$$\mathbf{S}\mathbf{r} = \mathbf{M}\mathbf{c}.$$

Setzen wir noch $\mathbf{Q} = \mathbf{S}^{-1}\mathbf{M}$, so erhalten wir die Koeffizienten für R_h aus $\mathbf{r} = \mathbf{Q}\mathbf{c}$. Die Matrizen $\mathbf{Q}$ können wir in einem Preprocessing-Schritt vor dem Beginn einer Rechnung speichern.

Damit die Matrix $\mathbf{Q}$ wohldefiniert ist, muss $\mathbf{S}$ invertierbar sein. Nicht jede beliebige Wahl der Basisfunktionen v_h^j liefert diese Eigenschaft. Wir nutzen daher die in [Lo11] vorgeschlagene Basis. Dazu definieren wir zunächst das in Abbildung 6.6 veranschaulichte Recovery-Koordinatensystem (r, s) . Es wird wie folgt aus dem ursprünglichen Koordinatensystem konstruiert. Zunächst wird der Ursprung in den Mittelpunkt (x_m, y_m) der Kante e_{ij} verschoben. Anschließend werden die Achsen um den Winkel α gedreht, so dass die s -Achse entlang von e_{ij} verläuft. Die Skalierung erfolgt so, dass die Ecken der Kante die Koordinaten $(0, 1)$ bzw. $(0, -1)$ haben. Aus den ursprünglichen Koordinaten (x, y) ergeben sich die Recovery-Koordinaten damit durch

$$\begin{bmatrix} r \\ s \end{bmatrix} = \frac{2}{|e_{ij}|} \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} x - x_m \\ y - y_m \end{bmatrix},$$

wobei hier $|e_{ij}|$ die Länge der Kante darstellt. Wählen wir dann die Basisfunktionen aus dem Raum $\text{span}\{1, r, r^2, r^3, s, sr, sr^2, sr^3\}$, so sichert dies die Regularität von $\mathbf{S}$ und damit die Existenz von $\mathbf{Q}$. Am einfachsten ist es natürlich, die auftretenden Monome direkt als Basisfunktionen zu wählen. Wir bemerken noch, dass die Basis so gewählt wurde, dass die Recovery-Funktion normal zur Kante, also in r -Richtung, bessere Approximationseigenschaften besitzt.

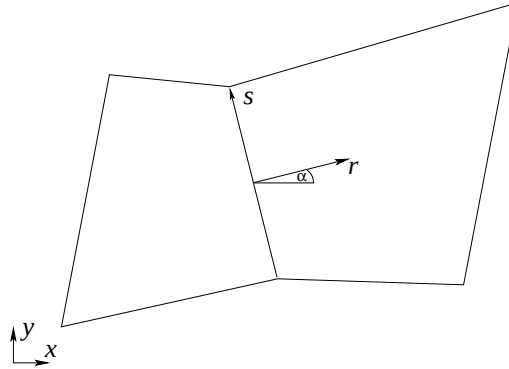


Abbildung 6.6: Recovery-Koordinatensystem

Damit können wir die Recovery-Funktion für alle inneren Kanten berechnen. An Randkanten ohne periodische Randbedingungen müssen wir einen anderen Weg gehen. Wir nutzen hier eine verkleinerte Recovery-Basis, bei der wir nur die Monome $\{1, r, r^2, s, sr, sr^2\}$ als Basisfunktionen verwenden. Sei e die zu K gehörige Randkante,

so lauten die zu Dirichlet-Randbedingungen $C|_e = g$ gehörigen Forderungen

$$\int_K R_h w_h d\mathbf{x} = \int_K C_h w_h d\mathbf{x} \quad \forall w_h \in W_h, \quad (6.47)$$

$$\int_e R_h w_h d\mathbf{s} = \int_e g w_h d\mathbf{s} \quad \forall w_h \in W_h, w_h|_e \neq 0. \quad (6.48)$$

In (6.47) durchläuft w_h wieder alle Basisfunktionen, in (6.48) allerdings nur die beiden, welche nicht auf e verschwinden. Damit haben wir 6 Gleichungen für die zu bestimmen 6 Freiheitsgrade.

Einfacher ist die Behandlung von Neumann-Randbedingungen. Ist die Randbedingung $\nabla C \cdot \mathbf{n}|_e = g$ gegeben, so können wir $\nabla R_h \cdot \mathbf{n}|_e = g$ direkt in das Randintegral in (6.44) einsetzen. An diesen Kanten müssen wir also keine Recovery-Funktion berechnen.

Mit RDG-Verfahren können bei Diffusionsproblemen auch unter Verwendung der bilinearen Ansatzfunktionen Approximationsordnungen größer als zwei erzielt werden. Dazu muss unter anderem die volle Recovery-Basis auch an Zellrändern verwendet werden, wozu Daten weiter innen liegender Zellen benötigt werden. Da wir im Weiteren Diffusion stets mit Konvektion zusammen betrachten, beschränken wir uns hier auf dieses einfache Verfahren zweiter Ordnung.

Nun wollen wir auch hier numerisch bestätigen, dass das Verfahren für hinreichend glatte Lösungen räumlich von zweiter Ordnung ist. Dazu betrachten wir auf $\Omega = [0, 1]^2$ das Testproblem

$$\partial_t C - \Delta C = f. \quad (6.49)$$

Die Randdaten und rechte Seite wählen wir dabei so, dass sich

$$C(x, y) = \frac{1}{2}(\cos(2\pi x) + \cos(2\pi y) + 2xy^2 - 1)$$

als stationäre Lösung des Testproblems ergibt, wobei wir auf der Kante mit $x = 1$ Neumann-Randbedingungen und auf den restlichen Kanten Dirichlet-Randbedingungen vorgeben. Zur Diskretisierung betrachten wir quadratische Gitter mit der Maschenweite $\Delta x = 1/2^{l+1}$ für $l = 0, \dots, 4$. Die zugehörigen Fehler, sowie die daraus resultierende Ordnung, sind für die einzelnen Gitter in Tabelle 6.4 aufgelistet und zeigen wie erhofft die zweite Ordnung. Als Anfangsdaten wurde dabei jeweils die Lösung $C^0(x, y) = C(x, y)$ gewählt.

6.3 DG für Advektions-Diffusions-Gleichungen

Die Kombination der Verfahren (6.4) und (6.44) liefert uns ein Verfahren zur Lösung der Advektions-Diffusions-Gleichung (2.37c). Das Verfahren lautet: Finde für $t > 0$ ein $C_h(t) \in W_h$, so dass

$$\begin{aligned} (\partial_t C_h, w_h)_K + \langle H(C_h^-, C_h^+, \mathbf{u} \cdot \mathbf{n}), w_h^- \rangle_{\partial K} - \langle \nabla R_h \cdot \mathbf{n}, w_h^- \rangle_{\partial K} \\ - (C_h \mathbf{u}, \nabla w_h)_K + (\nabla C_h, \nabla w_h)_K = (f, w_h)_K \quad \forall w_h \in W_h. \end{aligned} \quad (6.50)$$

Tabelle 6.4: Fehler und Konvergenzordnung für das Testproblem (6.49)

Gitter l	$\ C - C_h\ _{L^2(\Omega)}$ Fehler	Ordnung
0	0.187249	-
1	0.049253	1.926667
2	0.012057	2.030405
3	0.003007	2.003498
4	0.000751	2.000609

Auch für die Advektions-Diffusions-Gleichung

$$\partial_t C + \nabla \cdot (C \mathbf{u}) - \Delta C = f \quad (6.51)$$

betrachten wir noch ein Testproblem, um die Konvergenz des Verfahrens zu bestätigen. Dazu betrachten wir wieder das Gebiet $\Omega = [0, 1] \times [0, 1]$ und das Geschwindigkeitsfeld

$$u(x, y) = \sin(2x) \cos(2y), \quad v(x, y) = -\sin(2y) \cos(2x)$$

Mit $f = 8 \sin(2x) \sin(2y)$ und entsprechenden Randdaten ergibt sich

$$C(x, y) = \sin(2x) \sin(2y)$$

als stationäre Lösung des Testproblems. An der Kante mit $x = 1$ geben wir Neumann-Randbedingungen vor, auf den restlichen Kanten Dirichlet-Randbedingungen. Zur Diskretisierung betrachten wir wieder quadratische Gitter mit der Maschenweite $\Delta x = 1/2^{l+1}$ für $l = 0, \dots, 4$. Die zugehörigen Fehler, sowie die daraus resultierende Ordnung, sind für die einzelnen Gitter in Tabelle 6.5 aufgelistet und bestätigen die zweite Ordnung des Verfahrens. Auch hier wurde als Anfangsdaten die Lösung, also $C^0(x, y) = C(x, y)$ gewählt.

Tabelle 6.5: Fehler und Konvergenzordnung für Testproblem (6.51) (rechts)

Gitter	$\ C - C_h\ _{L^2(\Omega)}$	
l	Fehler	Ordnung
0	0.032631	-
1	0.008033	2.022293
2	0.001988	2.014655
3	0.000495	2.006264
4	0.000123	2.002798

7 Kopplung von Strömung und Transport

In diesem Kapitel befassen wir uns mit der Lösung des gekoppelten Systems (2.37). Wir werden im Folgenden sehen, dass eine sinnvolle Kopplung nur möglich ist, falls das numerische Geschwindigkeitsfeld mit der Transportdiskretisierung in einem gewissen Sinn kompatibel ist. Geschwindigkeiten aus (3.12a) in Verbindung mit der DG-Diskretisierung (6.50) erfüllen diese Kompatibilität nicht, was dazu führt, dass konstante Lösungen der Advektions-Diffusions-Gleichung mit dem DG-Verfahren nicht reproduziert werden können.

Daher werden wir das Geschwindigkeitsfeld in einem zusätzlichen Postprocessing-Schritt in ein kompatibles Geschwindigkeitsfeld überführen. Dies geschieht indem wir die Geschwindigkeit durch eine lokale Update-Technik aufdatieren. Auf den einzelnen Zellen ergibt sich der zugehörige Korrekturterm aus der Lösung eines Poisson-Problems, zu dessen Lösung wir $H(\text{div})$ -konforme gemischte Finte-Elemente verwenden. Durch die Wahl dieser Elemente erhalten wir zusätzlich noch die Stetigkeit der Geschwindigkeitsnormalkomponenten über die Zellränder hinweg. Dies ist essentiell für die Anwendbarkeit des DG-Verfahrens.

Das Kapitel ist wie folgt gegliedert. In Abschnitt 7.1 untersuchen wir die Auswirkung eines numerischen Geschwindigkeitsfeldes innerhalb des Verfahrens (6.50). Dies wird uns zeigen, was unter Kompatibilität von Strömung und Transport zu verstehen ist. In Abschnitt 7.2 befassen wir uns mit der Lösung von Poisson-Gleichungen mit gemischten FEM. Diese sind wesentlicher Bestandteil der in [CDMW98] eingeführten Conservative Projection Formulation (CPF) zur Bestimmung eines lokal konservativen Geschwindigkeitsfeldes. Wir wiederholen dieses Verfahren in Abschnitt 7.3, bevor wir es in Abschnitt 7.4 unserer Situation anpassen und die Local-CPF (LCPF) zur Berechnung eines kompatiblen Geschwindigkeitsfeldes vorstellen. In Abschnitt 7.5 betrachten wir dann die Lösung des gekoppelten Systems (2.37).

7.1 Konservativität und Reproduktion konstanter Lösungen

In Kapitel 6 haben wir ein DG-Verfahren zur Lösung der Advektions-Diffusions-Gleichung (2.37c) betrachtet. Dabei wurde stets angenommen, dass das Geschwindigkeitsfeld analytisch gegeben ist. In diesem Abschnitt untersuchen wir nun die Auswirkungen eines numerischen Geschwindigkeitsfeldes auf die Diskretisierung. Dies wurde bereits in

[DSW04] diskutiert und wir wiederholen hier die für uns wichtigen Resultate.

In diesem Abschnitt ist es zweckmäßig, die Gleichungen (2.14c) und (2.33c) in der Form

$$\partial_t C + \nabla \cdot (C \mathbf{u}) - \Delta C = \tilde{C} g, \quad (7.1)$$

mit

$$\nabla \cdot \mathbf{u} = g \quad (7.2)$$

zusammenzufassen. Eine homogene Form von (2.14c) ist darin für $g = 0$ enthalten. Mit $\tilde{C} \in \{0, 1\}$ erhalten wir eine Modifikation von (2.33c), wobei der zusätzliche Diffusions-term für das Folgende unwesentlich ist. Die Rand- und Anfangsbedingungen sollen wie in (2.40) durch $C|_{\Omega_D} = C_D$, $\nabla C \cdot \mathbf{n}|_{\partial\Omega_{out}} = 0$ sowie $C(\mathbf{x}, 0) = C^0(\mathbf{x})$ gegeben sein. Wir werden jetzt Eigenschaften der Lösung dieser Advektions-Diffusions-Gleichung betrachten, welche ein numerisches Verfahren entsprechend wiedergeben sollte.

Die Lösung von (7.1) erfüllt die globale Konservativität

$$\frac{d}{dt} \int_{\Omega} C \, d\mathbf{x} + \int_{\partial\Omega_{out}} C \mathbf{u} \cdot \mathbf{n} \, d\mathbf{x} + \int_{\partial\Omega_D} (C_D \mathbf{u} - \nabla C) \cdot \mathbf{n} \, d\mathbf{x} = \int_{\Omega} \tilde{C} g \, d\mathbf{x}. \quad (7.3)$$

Demnach erfolgt die zeitliche Änderung von C innerhalb des Gebietes ausschließlich durch die Randflüsse und den Quellterm. Außerdem gilt aufgrund von (7.2), dass die Konstante $\tilde{C}$ für alle Zeiten Lösung von (7.1) ist, falls wir $C = C_D = C^0 = \tilde{C} = \hat{C}$ setzen.

Unter welchen Voraussetzungen das DG-Verfahren (6.50) diese Eigenschaften reproduzieren kann, werden wir jetzt untersuchen. Mit einem exakt gegebenen Geschwindigkeitsfeld $\mathbf{u}$ lautet das Verfahren zur Approximation von (7.1)

$$\begin{aligned} (\partial_t C_h, w_h)_K + \langle H(C_h^-, C_h^+, \mathbf{u} \cdot \mathbf{n}), w_h^- \rangle_{\partial K} - \langle \nabla R_h \cdot \mathbf{n}, w_h^- \rangle_{\partial K} \\ - (C_h \mathbf{u}, \nabla w_h)_K + (\nabla C_h, \nabla w_h)_K = (\tilde{C} g, w_h)_K \quad \forall w_h \in W_h. \end{aligned} \quad (7.4)$$

Setzen wir $w_h = 1$, so erhalten wir nach Summation über alle Zellen und unter Berücksichtigung der Randbedingungen

$$\frac{d}{dt} \int_{\Omega} C_h \, d\mathbf{x} + \int_{\partial\Omega_{out}} H(C_h^-, C_h^+, \mathbf{u} \cdot \mathbf{n}) \, d\mathbf{s} + \int_{\partial\Omega_D} H(C_h^-, C_D, \mathbf{u} \cdot \mathbf{n}) - \nabla R_h \cdot \mathbf{n} \, d\mathbf{s} = \int_{\Omega} \tilde{C} g \, d\mathbf{x}. \quad (7.5)$$

Dies entspricht (7.3) und damit ist auch das DG-Verfahren global konservativ. Um die Reproduktion konstanter Lösungen zu überprüfen, setzen wir $C_h = C_D = C^0 = \tilde{C} = \hat{C}$ in (7.4), mit konstantem $\hat{C}$. Wegen der Konsistenz der numerischen Flussfunktion ergeben sich die konvektiven Randflüsse als

$$\langle H(\hat{C}, \hat{C}, \mathbf{u} \cdot \mathbf{n}), w_h^- \rangle_{\partial K} = \langle \hat{C} \mathbf{u} \cdot \mathbf{n}, w_h^- \rangle_{\partial K} = (\hat{C}, \nabla \cdot (w_h \mathbf{u}))_K.$$

Die linke Seite lautet dann

$$(\hat{C}, \nabla \cdot (w_h \mathbf{u}))_K - (\hat{C} \mathbf{u}, \nabla w_h)_K = (\hat{C} \nabla \cdot \mathbf{u}, w_h)_K$$

und führt auf die Bedingung

$$(\nabla \cdot \mathbf{u}, w_h)_K = (g, w_h)_K \quad \forall w_h \in W_h.$$

Für ein exakt gegebenes Geschwindigkeitsfeld mit (7.2) ist dies erfüllt und damit erhält das DG-Verfahren konstante Lösungen.

Nun ersetzen wir das exakte Feld $\mathbf{u}$ durch die numerische Approximation $\mathbf{u}_h$ aus (3.22). Da Funktionen aus $\mathbf{U}_h$ nach Definition in (3.12a) nur im Mittel über Zellgrenzen hinweg stetig sind, ist die Anwendung innerhalb der numerischen Flussfunktion unklar. Nutzen wir dagegen $\{\mathbf{u}_h\} = (\mathbf{u}_h^+ + \mathbf{u}_h^-)/2$ als Approximation, so entfällt diese Problematik und die Konservativität der numerischen Flussfunktion bleibt gesichert. Damit erhalten wir dann analog zu (7.5) die globale Konservativität des DG-Verfahrens mit Geschwindigkeitsfeld $\{\mathbf{u}_h\}$.

Die Reproduktion konstanter Lösungen gelingt aber nicht mehr. Die erforderliche Bedingung lautet nun

$$\langle \{\mathbf{u}_h\} \cdot \mathbf{n}, w_h^- \rangle_{\partial K} - (\mathbf{u}_h, \nabla w_h)_K = (g, w_h)_K \quad \forall w_h \in W_h. \quad (7.6)$$

Wir betrachten zunächst den Fall $w_h = 1$. Nach (3.12a) gilt über einer beliebigen Kante $e \in \mathcal{T}_h$ die Identität $\int_e \mathbf{u}_h^- ds = \int_e \mathbf{u}_h^+ ds$ und damit folgt

$$\int_{\partial K} \{\mathbf{u}_h\} \cdot \mathbf{n} ds = \int_{\partial K} \mathbf{u}_h^- \cdot \mathbf{n} ds = \int_K \nabla \cdot \mathbf{u}_h d\mathbf{x} = \int_K g d\mathbf{x}. \quad (7.7)$$

Die Bedingung (7.6) ist also erfüllt, falls W_h nur die konstanten Funktionen enthält. Im Allgemeinen gilt dies aber nicht, insbesondere mit den elementweise konstanten Druckapproximationen aus (3.12c) und den bilinearen Ansatzfunktionen für das DG-Verfahren (6.5) ist die Bedingung (7.6) verletzt.

Die Bedingung (7.6) können wir auch in der Form

$$(\nabla \cdot \mathbf{u}_h, w_h)_K = (g, w_h)_K \quad \forall w_h \in W_h \quad (7.8)$$

schreiben. Ein Vergleich mit (3.22) zeigt uns, dass der gewählte Ansatzraum für den Druck, die Ansatzfunktionen der transportierten Größe enthalten muss, um die Bedingung zu erfüllen und damit konstante Lösungen erhalten zu können.

Damit hat (7.6) eine essentielle Bedeutung für die Kopplung von Strömung und Transport und wird im Folgenden als Kompatibilitätsbedingung bezeichnet. Weiter nennen wir zwei Diskretisierungen für Strömung und Transport kompatibel, falls die Kompatibilitätsbedingung erfüllt ist. Geschwindigkeitsfelder für die die Kompatibilitätsbedingung mit $w_h = 1$ erfüllt sind, werden auch als lokal konservativ bezeichnet. Demnach ist die Bedingung (7.3) eine Verallgemeinerung der lokalen Konservativität. Insbesondere ist das Geschwindigkeitsfeld aus (3.22) lokal konservativ, aber nicht ausreichend kompatibel zur gewählten DG-Diskretisierung.

Um die Auswirkungen einer solchen Inkompatibilität zu verdeutlichen, betrachten wir den durch

$$\partial_t C + \nabla \cdot (C\mathbf{u}) = 0, \quad \nabla \cdot \mathbf{u} = 0 \quad (7.9)$$

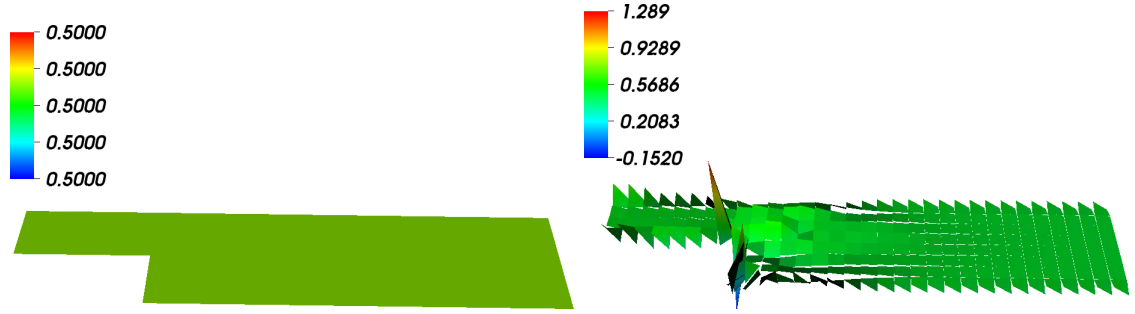


Abbildung 7.1: Numerische Lösung C_h des linearen Transportproblems (7.9) zur Zeit $t = 0$ (links) und $t = 1$ (rechts) bei Verwendung des inkompatiblen Geschwindigkeitsfeldes

gegebenen linearen Transport. Die verwendete Geometrie eines Backward-Facing-Step ist aus Abbildung 7.1 ersichtlich. Zur Bestimmung des Geschwindigkeitsfeldes geben wir am linken Rand ein parabolisches Geschwindigkeitsprofil vor, rechts setzen wir Ausflussrandbedingungen und die anderen Randstücke werden als feste Wände behandelt. Mit den in Kapitel 3 und 4 beschriebenen Techniken erhalten wir ein inkompressibles Geschwindigkeitsfeld $\mathbf{u}_h$ aus (3.12a). Als Dirichlet Randbedingungen und Anfangsdaten für die zu transportierende Größe wählen wir $C|_{\partial\Omega} = C^0 = 1/2$, womit die analytische Lösung $C = 1/2$ für alle Zeiten konstant bleibt. Aufgrund der zuvor beschriebenen Inkompatibilität des Geschwindigkeitsfeldes mit der verwendeten DG-Diskretisierung (6.50) gilt dies allerdings nicht für die numerische Lösung C_h . Wie groß deren tatsächlicher Einfluss ist, zeigt sich in Abbildung 7.1, wo besonders in der Nähe des Steps unphysikalisches Verhalten ersichtlich ist. Der relative Fehler beträgt über 150%, womit die Approximation inakzeptabel ist.

Um die Kompatibilität von Strömung und Transport zu erreichen, haben wir nun verschiedene Möglichkeiten. Wir könnten den Ansatzraum der Druckapproximationen vergrößern. Dies beinhaltet im Allgemeinen, sofern keine zusätzlichen Stabilisierungstechniken genutzt werden, auch eine entsprechende Erweiterung des Ansatzraumes für die Geschwindigkeiten. Damit steigt allerdings auch der Aufwand zur Berechnung der Geschwindigkeit, obwohl dies hier eine untergeordnete Rolle spielt. Andererseits können wir auch einen zellweise konstanten Ansatz innerhalb des DG-Verfahrens nutzen, was zu einer geringeren räumlichen Ordnung führt. Schließlich bleibt noch die Aufbereitung des Geschwindigkeitsfeldes in einem zusätzlichen Postprocessing-Schritt. Dies erlaubt uns die gegebenen Ansatzräume der jeweiligen Diskretisierung weiter zu nutzen.

Wir werden im Weiteren das Geschwindigkeits-Postprocessing betrachten. Ziel ist die Berechnung eines zur Diskretisierung (6.50) kompatiblen Geschwindigkeitsfeldes mit stetiger Normalkomponente. In [SW06] wird dazu ein Algorithmus eingeführt, welcher unter der ausschließlichen Berücksichtigung der auf den Zellrändern gegebenen Geschwindigkeitsnormalkomponenten ein kompatibles Geschwindigkeitsfeld bestimmt. In einem ersten Schritt werden dazu lokal-konservative Normalkomponenten aus den gegebenen Normal-

komponenten bestimmt. Aus diesen wird dann in einem zweiten Schritt das kompatible Geschwindigkeitsfeld elementweise berechnet, indem die Gleichungen zur Bestimmung des Geschwindigkeitsfeld elementweise, mit den neu bestimmten Normalkomponenten als Randbedingungen, gelöst werden. Die Vorgabe der Normalkomponenten wird ermöglicht indem eine gemischte FEM zur Berechnung der Geschwindigkeit verwendet wird. Hierbei ist anzumerken, dass in [SW06] die Darcy-Gleichungen der Strömungsberechnung zu Grunde liegen. Der Aufwand zur Berechnung der Strömung entspricht damit lediglich dem Lösen eines Poisson-Problems.

Auf unseren Fall übertragen vereinfacht sich der erste Schritt dieser Vorgehensweise. Wegen (7.7) haben wir mit $\{\mathbf{u}_h\} \cdot \mathbf{n}$ bereits geeignete Normalkomponenten vorliegen. Bleibt also der zweite Schritt, der von uns das Lösen der Modellgleichungen (2.37a), (2.37b) auf den einzelnen Elementen verlangen würde. Aufgrund der Komplexität der Gleichungen sehen wir davon aber ab.

Stattdessen nutzen wir die vom obigen Verfahren innerhalb der einzelnen Elemente nicht berücksichtigten Geschwindigkeiten und bestimmen ein Update, so dass sich die Kompatibilität durch Aufaddieren ergibt. Das Update erhalten wir als Lösung eines Poisson-Problems, zu deren Bestimmung wir ebenfalls eine gemischte FEM verwenden. Diese Idee stammt aus [CDMW98] und wird dort nicht lokal, sondern global als Conservative Projection Formulation (CPF) zur Bestimmung eines lokal konservativen Geschwindigkeitsfeldes vorgestellt. Allerdings kann die CPF auch genutzt werden, um ein kompatibles Geschwindigkeitsfeld zu berechnen. Der Vorteil unserer lokalen Betrachtung liegt darin, dass wir das Lösen großer Gleichungssysteme vermeiden.

7.2 Gemischte Finite Elemente für die Poisson-Gleichung

Bei den oben erwähnten Verfahren zur Bestimmung eines kompatiblen Geschwindigkeitsfeldes kommt der Lösung von Poisson-Problemen mit gemischten FEM eine wesentliche Rolle zu. Daher werden wir diese hier kurz einführen, für einen ausführlicheren Überblick verweisen wir auf die Darstellungen in [BF91, RT91, Mon03, BBD⁺08].

Wir betrachten auf Ω das Poisson-Problem $-\Delta p = g$ mit homogenen Dirichlet- und Neumann-Randbedingungen, also $p|_{\partial\Omega_D} = 0$ und $\nabla p \cdot \mathbf{n}|_{\partial\Omega_N} = 0$. Dabei soll sich der Gebietsrand $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$ aus der disjunkten Vereinigung der beiden Randstücke ergeben. Einführung der neuen Unbekannten $\mathbf{v} = -\nabla p$ liefert auf Ω das gemischte System

$$\mathbf{v} + \nabla p = 0, \quad (7.10a)$$

$$\nabla \cdot \mathbf{v} = g, \quad (7.10b)$$

mit den Randbedingungen

$$p|_{\partial\Omega_D} = 0, \quad (7.11)$$

$$\mathbf{v} \cdot \mathbf{n}|_{\partial\Omega_N} = 0. \quad (7.12)$$

Um eine schwache Formulierung dieses Gleichungssystems zu bestimmen, multiplizieren wir zunächst (7.10a) mit einer hinreichend glatten Testfunktion $\mathbf{w}$ und integrieren über das gesamte Gebiet. Nach anschließender partieller Integration erhalten wir damit

$$(\mathbf{v}, \mathbf{w}) - (p, \nabla \cdot \mathbf{w}) + \langle p, \mathbf{w} \cdot \mathbf{n} \rangle_{\partial\Omega} = 0.$$

Das dabei auftretende Randintegral verschwindet wegen (7.11) auf dem Dirichlet-Rand $\partial\Omega_D$. Damit dies auch auf $\partial\Omega_N$ eintritt, müssen wir von den Testfunktionen explizit $\mathbf{w} \cdot \mathbf{n}|_{\partial\Omega_N} = 0$ fordern. Im Unterschied zu den in Kapitel 3 betrachteten schwachen Formulierungen gehen hier also Dirichlet-Randbedingungen als natürliche Randbedingungen ein und Neumann-Randbedingungen müssen als wesentliche Randbedingungen gesondert eingebracht werden. Entsprechend verfahren wir auch mit (7.10b) unter Verwendung einer hinreichend glatten Testfunktion q und erhalten so

$$(\nabla \cdot \mathbf{v}, q) = (g, q).$$

Da dabei nur die Divergenz der Lösung auftritt, definieren wir durch

$$H(\text{div}, \Omega) = \{\mathbf{w} | \mathbf{w} \in L^2(\Omega), \nabla \cdot \mathbf{w} \in L^2(\Omega)\}$$

den Raum der vektorwertigen Funktionen mit Divergenz in $L^2(\Omega)$. Außerdem sei

$$H_0(\text{div}, \Omega) = \{\mathbf{w} | \mathbf{w} \in H(\text{div}, \Omega), \mathbf{w} \cdot \mathbf{n}|_{\partial\Omega_N} = 0\}$$

der Raum der Funktionen aus $H(\text{div}, \Omega)$, deren Normalkomponente auf $\partial\Omega_N$ verschwindet. Somit ergibt sich eine schwache Formulierung des Systems (7.10) wie folgt: Finde $\mathbf{v} \in H_0(\text{div}, \Omega)$ und $p \in L^2(\Omega)$ mit

$$(\mathbf{v}, \mathbf{w}) - (p, \nabla \cdot \mathbf{w}) = 0 \quad \forall \mathbf{w} \in H_0(\text{div}, \Omega), \quad (7.13a)$$

$$(\nabla \cdot \mathbf{v}, q) = (g, q) \quad \forall q \in L^2(\Omega). \quad (7.13b)$$

Um die Lösung dieses Systems mittels konformer Finiten-Elemente zu approximieren, benötigen wir endlich dimensionale Teilräume $\mathbf{W}_{h,0} \subset H_0(\text{div}, \Omega)$ und $Q_h \subset L^2(\Omega)$. Als erstes wollen wir einen endlichen Teilraum von $H(\text{div}, \Omega)$ angeben, dazu führen wir die divergenz-konformen Räume nach Raviart-Thomas [RT77, BF91] auf Viereckselementen ein. Dabei folgen wir im Wesentlichen den Darstellungen aus [BF91] und [Mon03]. Insbesondere bemerken wir noch, dass Funktionen aus Teilräumen von $H(\text{div}, \Omega)$, welche polynomial auf den einzelnen Elementen sind, stetig bezüglich ihrer Normalkomponente sein müssen.

Auf dem Referenzquadrat $\hat{K}$ definieren wir den Polynomraum

$$RT_k(\hat{K}) = P_{k+1,k}(\hat{K}) \times P_{k,k+1}(\hat{K}),$$

wobei wir mit $P_{k,j}(\hat{K})$ den Raum der Polynome vom Grad k bezüglich $\hat{x}$ und Grad j bezüglich $\hat{y}$ bezeichnen. Statt $P_{k,k}$ schreiben wir kurz Q_k . Es gilt $\dim RT_k = 2(k+1)(k+2)$ und die Räume sind so konstruiert, dass $\nabla \cdot \hat{\mathbf{w}} \in Q_k(\hat{K})$ für $\hat{\mathbf{w}} \in RT_k(\hat{K})$ gilt. Für eine

Kante $\hat{e} \subset \partial \hat{K}$ mit äußerem Normalenvektor $\hat{\mathbf{n}}$ gilt außerdem $\hat{\mathbf{w}} \cdot \hat{\mathbf{n}}|_{\hat{e}} \in P_k(\hat{e})$, wobei hier P_k den Raum der Polynome vom Grad k in einer Variablen bezeichnet. Insbesondere handelt es sich bei RT_k um vektorielle Räume, welche sich nicht als Tensorprodukt eines skalaren Raumes ergeben. Die Freiheitsgrade werden mit Hilfe der Räume $P_k(\hat{e})$, sowie $P_{k-1,k}(\hat{K}) \times P_{k,k-1}(\hat{K})$ definiert. Dazu sei $\{\hat{q}_1, \dots, \hat{q}_{k+1}\}$ eine Basis von $P_k(\hat{e})$ und $\{\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_{2k(k+1)}\}$ eine Basis von $P_{k-1,k}(\hat{K}) \times P_{k,k-1}(\hat{K})$. Dann lauten die Freiheitsgrade

$$M_{\hat{e},j}^{\hat{K}}(\hat{\mathbf{w}}) = \int_{\hat{e}} \hat{\mathbf{w}} \cdot \hat{\mathbf{n}}|_{\hat{K}} \hat{q}_j d\hat{\mathbf{s}}, \quad j = 1, \dots, k+1, \forall \hat{e} \subset \hat{K} \quad (7.14a)$$

$$M_j^{\hat{K}}(\hat{\mathbf{w}}) = \int_{\hat{K}} \hat{\mathbf{w}} \cdot \hat{\mathbf{v}}_j d\hat{\mathbf{x}}, \quad j = 1, \dots, 2k(k+1), \quad (7.14b)$$

wobei $\hat{\mathbf{n}}$ den zu $\hat{e}$ gehörigen äußeren Normalenvektor bezeichnet. Wegen $\hat{\mathbf{w}} \cdot \hat{\mathbf{n}}|_{\hat{e}} \in P_k(\hat{e})$ können wir die Bedingung (7.14a) auch so verstehen, dass die Normalkomponente auf jeder Kante $\hat{e}$ an $k+1$ Stellen vorgegeben wird und damit eindeutig bestimmt ist. Die Wahl von (7.14a) als Freiheitsgrad erfolgt, um die Stetigkeit der Normalkomponenten über Zellränder hinweg zu gewährleisten. Die Bestimmung einer Basis für $RT_1(\hat{K})$ ist im Anhang A.2 beschrieben, die Freiheitsgrade für diesen und den Fall $k=0$ sind in Abbildung 7.2 dargestellt.

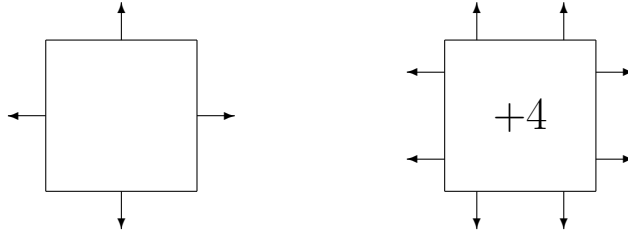


Abbildung 7.2: Die Freiheitsgrade für das RT_0 -Element (links) und das RT_1 -Element (rechts).

Um den Raum $RT_k(K)$ auf beliebigen Vierecken K zu definieren, verwenden wir die Piola-Transformation P_F , siehe auch [Cia88]. Bezeichnet F die Referenzabbildung von $\hat{K}$ nach K , also $\mathbf{x} = F(\hat{\mathbf{x}})$, so ist die Piola-Transformation durch

$$\mathbf{w}(\mathbf{x}) = P_F(\hat{\mathbf{w}})(\mathbf{x}) = \frac{1}{\det DF(\hat{\mathbf{x}})} DF(\hat{\mathbf{x}}) \hat{\mathbf{w}}(\hat{\mathbf{x}}) \quad (7.15)$$

gegeben, wobei DF für die zugehörige Jacobi-Matrix steht. Außerdem gehen wir davon aus, dass $\det DF$ stets positiv ist. Mit $\mathbf{w} = P_F(\hat{\mathbf{w}})$ und $q = \hat{q} \circ F^{-1}$ gelten die Identitäten

$$\int_K q \nabla \cdot \mathbf{w} d\mathbf{x} = \int_{\hat{K}} \hat{q} \hat{\nabla} \cdot \hat{\mathbf{w}} d\hat{\mathbf{x}} \quad (7.16)$$

und

$$\int_{\partial K} q \mathbf{w} \cdot \mathbf{n} d\mathbf{s} = \int_{\partial \hat{K}} \hat{q} \hat{\mathbf{w}} \cdot \hat{\mathbf{n}} d\hat{\mathbf{s}}. \quad (7.17)$$

Erklären wir die zu K gehörigen Freiheitsgrade durch

$$M_{e,j}^K(\mathbf{w}) = \int_e \mathbf{w} \cdot \mathbf{n}|_K q_j ds, \quad q_j = \hat{q}_j \circ F^{-1}, j = 1, \dots, k+1, \forall e \subset \partial K, \quad (7.18a)$$

$$M_j^K(\mathbf{w}) = \int_K \mathbf{w} \cdot \mathbf{v}_j d\mathbf{x}, \quad \mathbf{v}_j = DF^{-T} \hat{\mathbf{v}}_j \circ F^{-1}, j = 1, \dots, 2k(k+1), \quad (7.18b)$$

so sind diese invariant unter der Piola-Transformation (7.15). Für M_e gilt dies nach (7.17) und für M_K ergibt sich die Invarianz aus dem Transformationssatz. Funktionen aus einem globalen Finite-Elemente Raum $\mathbf{W}_h$ erhalten wir dann durch Vereinigung der lokalen Freiheitsgrade (7.18) über alle $K \in \mathcal{T}_h$ und Vorgabe entsprechender Werte. Dabei ist $\mathbf{w}_h \cdot \mathbf{n}$ für $\mathbf{w}_h \in \mathbf{W}_h$ nach Konstruktion stetig über die Zellränder hinweg, womit sich

$$\mathbf{W}_h = \left\{ \mathbf{w}_h \in H(\operatorname{div}, \Omega) \mid \mathbf{w}_h|_K \in P_F(RT(\hat{K})), \forall K \in \mathcal{T}_h \right\}$$

als der gesuchte endlichdimensionale Teilraum von $H(\operatorname{div}, \Omega)$ ergibt.

Für hinreichend reguläres $\mathbf{w}$, siehe [BF91], können wir einen Interpolationsoperator Π_K mit $\Pi_K \mathbf{w} \in P_F(RT(\hat{K}))$ definieren. Dieser ist durch die Bedingungen

$$\begin{aligned} M_{e,j}^K(\mathbf{w} - \Pi_K \mathbf{w}) &= 0, & j &= 1, \dots, k+1, \forall e \subset K, \\ M_K(\mathbf{w} - \Pi_K \mathbf{w}) &= 0, & j &= 1, \dots, 2k(k+1), \end{aligned}$$

eindeutig festgelegt und es gilt $\widehat{\Pi_K \mathbf{w}} = \Pi_{\hat{K}} \hat{\mathbf{w}}$. Den zugehörigen globalen Interpolator $\Pi_h \mathbf{w}$ erhalten wir dann durch die Bedingung

$$(\Pi_h \mathbf{w})|_K = \Pi_K(\mathbf{w}|_K) \quad \forall K \in \mathcal{T}_h. \quad (7.19)$$

Der Interpolationsfehler genügt bei hinreichend glattem $\mathbf{w}$ der Abschätzung

$$\|\mathbf{w} - \Pi_h \mathbf{w}\|_{L^2(\Omega)} \leq Ch^{k+1} |\mathbf{w}|_{H^{k+1}(\Omega)}, \quad (7.20)$$

falls $\widehat{\Pi_h \mathbf{w}} \in RT_k(\hat{K})$. Für affine Transformationen des Referenzquadrats findet sich der Beweis etwa in [BF91], der allgemeinere Fall wird in [ABF05] behandelt.

Wir kommen nun zurück zur Diskretisierung der schwachen Formulierung (7.13). Setzen wir

$$\mathbf{W}_{h,0} = \{ \mathbf{w}_h \in \mathbf{W}_h \mid \mathbf{w}_h \cdot \mathbf{n}|_{\partial\Omega_N} = 0 \}$$

und

$$Q_h = \left\{ q \in L^2(\Omega) \mid (q \circ F)|_{\hat{K}} \in Q_k(\hat{K}) \right\},$$

dann lautet die diskrete Form: Finde $\mathbf{v}_h \in \mathbf{W}_{h,0}$ und $p \in Q_h$ mit

$$\begin{aligned} (\mathbf{v}_h, \mathbf{w}_h) - (p_h, \nabla \cdot \mathbf{w}_h) &= 0 & \forall \mathbf{w}_h \in \mathbf{W}_{h,0} \\ (\nabla \cdot \mathbf{v}_h, q_h) &= (g, q_h) & \forall q_h \in Q_h. \end{aligned}$$

Dieses System ist eindeutig lösbar, der Beweis findet sich in [BF91]. Dort wird ebenfalls der Fall inhomogener Randdaten behandelt. Diese stellen keine größeren Probleme dar, verkomplizieren lediglich die Präsentation. Aus praktischer Sicht können inhomogene Neumann-Randbedingungen direkt durch Vorgabe der entsprechenden Freiheitsgrade implementiert werden.

7.3 Conservative-Projection-Formulation

Wir führen nun die Conservative Projection Formulation (CPF) nach [CDMW98] ein. Ziel des Verfahrens ist es aus einem gegebenen Geschwindigkeitsfeld ein neues Geschwindigkeitsfeld zu bestimmen, welches lokal konservativ ist und dessen Geschwindigkeitsnormalkomponenten über die Zellränder hinweg stetig sind. Dabei soll die Approximationsordnung erhalten bleiben. Gegebenenfalls können hierbei sogar unterschiedliche Gitter verwendet werden.

Sei nun $\mathbf{u}_h \in \mathbf{W}_h$ eine Approximation an das exakte Geschwindigkeitsfeld $\mathbf{u}$, welches den Bedingungen $\nabla \cdot \mathbf{u} = g$ auf Ω und $\mathbf{u} \cdot \mathbf{n} = b$ auf $\partial\Omega_N$ genügt, wobei der Gebietsrand $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$ wieder in zwei disjunkte Teilstücke zerlegt ist. Das Problem besteht nun darin, bei gegebenem Raum $\mathbf{W}_h^*$ ein $\mathbf{u}_h^* \in \mathbf{W}_h^*$ zu finden, welches

$$\nabla \cdot \mathbf{u}_h^* = g \text{ auf } \Omega, \quad \mathbf{u}_h^* \cdot \mathbf{n} = b \text{ auf } \partial\Omega_N, \quad (7.21)$$

in einem hinreichend schwachen Sinne erfüllt. Zur Bestimmung des neuen Geschwindigkeitsfeldes machen wir den Ansatz

$$\mathbf{u}_h^* = \mathcal{P}\mathbf{u}_h + \mathbf{v}_h, \quad (7.22)$$

wobei wir $\mathbf{v}_h \in \mathbf{W}_h^*$ wählen und mit $\mathcal{P}\mathbf{u}_h \in \mathbf{W}_h^*$ die L^2 -Projektion von $\mathbf{u}_h \in \mathbf{W}_h$ in den Raum $\mathbf{W}_h^*$ bezeichnen. Mit dem Ansatz (7.22) erhalten wir aus (7.21) die Gleichungen

$$\nabla \cdot \mathbf{v}_h = \tilde{g} \text{ auf } \Omega, \quad \tilde{g} = g - \nabla \cdot \mathcal{P}\mathbf{u}_h, \quad (7.23a)$$

$$\mathbf{v}_h \cdot \mathbf{n} = \tilde{b} \text{ in } \partial\Omega_N, \quad \tilde{b} = b - \mathcal{P}\mathbf{u}_h \cdot \mathbf{n}. \quad (7.23b)$$

Gehen wir weiter davon aus, dass wir den Korrekturterm als Gradient einer skalaren Funktion, also in der Form

$$\mathbf{v}_h = -\nabla p_h, \quad (7.24)$$

schreiben können, so erhalten wir aus (7.23) das Poisson-Problem

$$-\Delta p_h = \tilde{g} \text{ in } \Omega, \quad -\nabla p_h \cdot \mathbf{n} = \tilde{b} \text{ auf } \partial\Omega_N, \quad p_h = 0 \text{ auf } \partial\Omega_D. \quad (7.25)$$

Um das Problem so einfach wie möglich zu halten, haben wir hier homogene Dirichlet-Randbedingungen für die Hilfsgröße p_h gewählt. Da wir uns lediglich für den Korrekturterm $\mathbf{v}_h$ interessieren, hat die genaue Wahl der Dirichlet-Randbedingungen keinen weiteren Einfluss.

Das System partieller Differentialgleichungen (7.25) wird nun wie in Abschnitt 7.2 mit einer gemischten FEM gelöst: Finde $\mathbf{v}_h \in \mathbf{W}_h^*$ mit $\mathbf{v}_h \cdot \mathbf{n}|_{\partial\Omega_N} = \tilde{b}$ und $p_h \in Q_h^*$, so dass

$$(\mathbf{v}_h, \mathbf{w}_h) - (p_h, \nabla \cdot \mathbf{w}_h) = 0 \quad \forall \mathbf{w}_h \in \mathbf{W}_h^* \text{ mit } \mathbf{v}_h \cdot \mathbf{n}|_{\partial\Omega_N} = 0, \quad (7.26a)$$

$$(\nabla \cdot \mathbf{v}_h, q_h) = (\tilde{g}, q_h) \quad \forall q_h \in Q_h^*. \quad (7.26b)$$

Eine mögliche Wahl der Räume haben wir mit den RT_k -Räumen in Abschnitt 7.2 kennengelernt. Bei der Verwendung derartiger konformer Ansatzräume wird die lokale Konservativität, als auch die Stetigkeit der Normalkomponenten gesichert. Die lokale Konservativität ergibt sich aus (7.26b), was äquivalent ist zu

$$(\nabla \cdot \mathbf{u}^*, q_h) = (g, q_h) \quad \forall q_h \in Q_h^*. \quad (7.27)$$

Die Stetigkeit der Normalkomponenten von $\mathbf{v}$ ergibt sich aus der $H(\text{div}, \Omega)$ -Konformität des verwendeten Raums. Da $\mathcal{P}\mathbf{u}$ ebenfalls in diesem Raum liegt, weist damit auch $\mathbf{u}^*$ eine stetige Normalkomponente auf. Tatsächlich ist $\mathbf{u}_h^*$ nicht nur lokal konservativ, sondern bei hinreichender Größe von Q_h^* folgt aus (7.27) die Kompatibilität zur DG-Diskretisierung (6.50) für Ansatzfunktionen beliebiger Ordnung.

Die Gleichungen (7.22) und (7.24) in Verbindung mit (7.26) werden zusammen als Conservative Projection Formulation bezeichnet. Dabei gilt nach [CDMW98] die Fehlerabschätzung

$$\|\mathbf{u}_h^* - \mathbf{u}\| \leq C(\|\Pi_h \mathbf{u} - \mathbf{u}\| + \|\mathbf{u}_h - \mathbf{u}\|), \quad (7.28)$$

wobei Π_h den globalen Interpolationsoperator (7.19) bezeichnet. Bei Verwendung der RT_k -Elemente ist der Interpolationsfehler nach (7.20) von der Ordnung $k + 1$. Ist die ursprüngliche numerische Approximation von der selben Ordnung, so gilt nach (7.28) die Abschätzung $\|\mathbf{u}_h^* - \mathbf{u}\| \leq Ch^{k+1}$ und die Approximationsordnung des ursprünglichen Geschwindigkeitsfeldes bleibt erhalten.

In ihrer ursprünglichen globalen Form erfordert die CPF das Lösen von zwei großen linearen Gleichungssystemen. Eines zur Bestimmung der L^2 -Projektion von $\mathbf{u}_h$, das andere zur Berechnung des Korrekturterms $\mathbf{v}_h$ aus (7.26). Wir werden gleich eine lokale Variante des Verfahrens vorstellen, innerhalb der wir die Lösung dieser großen Systeme vermeiden können.

7.4 Local-Conservative-Projection-Formulation

Wir wollen jetzt die CPF auf unseren Fall übertragen. Um die Lösung großer linearer Gleichungssysteme zu vermeiden, werden wir die CPF elementweise anwenden und somit eine neuartige Variante, welche wir als Local-CPF (LCPF) bezeichnen, herleiten. Die Idee das kompatible Geschwindigkeitsfeld lokal zu berechnen stammt aus [SW06], dort werden allerdings die dem Geschwindigkeitsfeld zu Grunde liegenden Gleichungen elementweise gelöst. Als Randbedingungen dienen dabei zuvor bestimmte lokal konservative Geschwindigkeitsnormalkomponenten. Wir kombinieren diesen Ansatz mit der CPF und berechnen das CPF-Update lokal, auch dazu benötigen wir die lokal konservativen Normalkomponenten. Um die Darstellung konkret und kurz zu halten, betrachten wir hier direkt das Postprocessing der Geschwindigkeiten aus (3.22) mit den RT_1 -Räumen aus Abschnitt 7.2. Die Vorgehensweise lässt sich aber entsprechend auf andere Fälle übertragen.

Für das exakte Geschwindigkeitsfeld $\mathbf{u}$ gelte wieder $\nabla \cdot \mathbf{u} = g$. Daraus ergibt sich auch eine Bedingung an die Randdaten, wenn wir die CPF lokal einsetzen wollen. Für $K \in \mathcal{T}_h$ sei $b = \mathbf{u} \cdot \mathbf{n}|_{\partial K}$, dann gilt

$$\int_{\partial K} b \, ds = \int_{\partial K} \mathbf{u} \cdot \mathbf{n} \, ds = \int_K \nabla \cdot \mathbf{u} \, d\mathbf{x} = \int_K g \, d\mathbf{x}.$$

Liegt mit $\mathbf{u}_h$ eine lokal konservative Approximation an $\mathbf{u}$ vor, so können wir $b = \mathbf{u}_h \cdot \mathbf{n}|_{\partial K}$ als geeignete Randbedingungen wählen. Für Fälle in denen $\mathbf{u}_h$ nicht lokal konservativ

ist, verweisen wir auf [LN04, SW06, CGW07]. Wegen (7.7) nehmen wir auf den Kanten $\{\mathbf{u}_h\} \cdot \mathbf{n}$ für die Geschwindigkeit aus (3.22). Zwar ist das Geschwindigkeitsfeld auch ohne Mittelwertbildung lokal konservativ, so erhalten wir aber auch einen eindeutigen Normalfluss auf den Kanten.

Analog zu (7.22) machen wir auf $K \in \mathcal{T}_h$ den Ansatz $\mathbf{u}_h^* = \mathcal{P}_K \mathbf{u}_h + \mathbf{v}_h$ mit $\mathbf{v}_h = -\nabla p_h$, wobei $\mathcal{P}_K$ die L^2 -Projektion auf $RT_1(K) = P_F(RT_1(\hat{K}))$ bezeichnet. Dies führt dann auf die Gleichungen

$$\nabla \cdot \mathbf{v}_h = \tilde{g} \text{ auf } K, \quad \tilde{g} = g - \nabla \cdot \mathcal{P}_K \mathbf{u}_h, \quad (7.29)$$

$$\mathbf{v}_h \cdot \mathbf{n} = \tilde{b} \text{ auf } \partial K, \quad \tilde{b} = \mathcal{P}_K (\{\mathbf{u}_h\} \cdot \mathbf{n} - \mathbf{u}_h \cdot \mathbf{n})|_{\partial K} \quad (7.30)$$

und die entsprechende schwache Formulierung lautet: Finde $(\mathbf{v}_h, p_h) \in RT_1(K) \times Q_1(K)$ mit $\mathbf{v}_h \cdot \mathbf{n}|_{\partial K} = \tilde{b}$ und $\int_K p_h d\mathbf{x} = 0$, so dass

$$(\mathbf{v}_h, \mathbf{w}_h)_K - (p_h, \nabla \cdot \mathbf{w}_h)_K = 0 \quad \forall \mathbf{w}_h \in RT_1(K) \text{ mit } \mathbf{w}_h \cdot \mathbf{n}|_{\partial K} = 0, \quad (7.31a)$$

$$(\nabla \cdot \mathbf{v}_h, q_h)_K = (\tilde{g}, q_h)_K \quad \forall q_h \in Q_1(K), \quad (7.31b)$$

gilt. Die Forderung $\int_K p_h d\mathbf{x} = 0$ nutzen wir, um eine eindeutige Lösung zu erhalten. Ohne weitere Bedingung ist p_h nur bis auf additive Konstanten eindeutig bestimmt.

Um aus (7.31) ein lineares Gleichungssystem zu erhalten, benötigen wir eine Basis von $RT_1(K)$. Im Anhang A.2 wird eine Basis von $RT_1(\hat{K})$ angegeben, für die sich die ersten 8 Koeffizienten der Basisdarstellung von $\hat{\mathbf{w}}_h \in RT_1(\hat{K})$ aus den Werten von $\hat{\mathbf{w}}_h \cdot \hat{\mathbf{n}}$ an zwei speziellen Stellen pro Kante ergeben. Die genauen Bedingungen sind in (A.2) gegeben. Umgekehrt können durch Vorgabe dieser 8 Linearkoeffizienten Randbedingungen vorgegeben werden, denn $\hat{\mathbf{w}}_h \cdot \hat{\mathbf{n}}$ ist linear auf den Kanten und damit durch die Vorgabe zweier Werte eindeutig bestimmt. Mittels der Piola-Transformation (7.15) kann die Basis (A.1) in eine Basis von $RT_1(K)$ überführt werden.

Um auch auf $RT_1(K)$ Randbedingungen vorgeben zu können, müssen wir die Transformation der Normalkomponenten betrachten. Da Normalenvektoren $\hat{\mathbf{n}}$ unter F auf $\mathbf{n} = DF^{-T} \hat{\mathbf{n}} / \|DF^{-T} \hat{\mathbf{n}}\|_2$ abgebildet werden, gilt

$$(\mathbf{w} \cdot \mathbf{n}) \circ F(\hat{\mathbf{x}}) = \frac{DF \hat{\mathbf{w}}}{\det(DF)} \cdot \frac{DF^{-T} \hat{\mathbf{n}}}{\|DF^{-T} \hat{\mathbf{n}}\|}(\hat{\mathbf{x}}) = \frac{\hat{\mathbf{w}} \cdot \hat{\mathbf{n}}}{\det(DF) \|DF^{-T} \hat{\mathbf{n}}\|}(\hat{\mathbf{x}}). \quad (7.32)$$

Dies können wir nutzen, um Randbedingungen in $RT_1(K)$ einzubauen. Machen wir den Ansatz $\mathbf{w}_h = \sum_{j=1}^{12} v_j \mathbf{w}_h^j$, so gilt beispielsweise mit $\hat{\mathbf{n}} = (0, -1)^T$ und $\hat{\mathbf{x}}_0 = (-1/3, -1)^T$ nach (7.32) und (A.2) die Identität

$$(\mathbf{w}_h \cdot \mathbf{n}) \circ F(\hat{\mathbf{x}}_0) = \sum_{j=1}^{12} v_j \frac{\hat{\mathbf{w}}_h^j(\hat{\mathbf{x}}_0) \cdot \hat{\mathbf{n}}}{\det(DF) \|DF^{-T} \hat{\mathbf{n}}\|} = \frac{v_1}{\det(DF) \|DF^{-T} \hat{\mathbf{n}}\|}$$

und damit

$$v_1 = \det(DF) \|DF^{-T} \hat{\mathbf{n}}\| (\mathbf{w}_h \cdot \mathbf{n}) \circ F(\hat{\mathbf{x}}_0).$$

Entsprechendes gilt für $v_2, \dots, v_8$. Der dabei auftretende Ausdruck $\det(DF) \|DF^{-T} \hat{\mathbf{n}}\|$ ist auf den einzelnen Kanten konstant. Da $F \in Q_1(\hat{K})^2$, sind die Komponenten der ersten

Spalte von DF^{-T} linear in $\hat{x}$ und die der zweiten Spalte linear in $\hat{y}$. Auswertung auf einer Kante und Multiplikation mit dem zugehörigen Normalenvektor ergibt einen konstanten Vektor. Insbesondere ist damit auch $\mathbf{w}_h \cdot \mathbf{n}$ ein lineares Polynom auf den Kanten von K . Daher ist es zweckmäßig die Randdaten zunächst in den Raum der linearen Polynome zu projizieren, um die Randbedingung (7.30) umzusetzen.

Jetzt stellen wir das aus (7.31) resultierende Gleichungssystem auf. Dazu setzen wir $\mathbf{v}_h(\mathbf{x}) = \sum_{j=1}^{12} v_j \mathbf{w}_h^j(\mathbf{x})$ und $p_h = \sum_{j=1}^4 p_j q_h^j(\mathbf{x})$. Die ersten acht Koeffizienten der Basisdarstellung von $\mathbf{v}_h$ sind durch die Randbedingung (7.30) gegeben, daher schreiben wir auch $\mathbf{v}_h(\mathbf{x}) = \tilde{\mathbf{v}}_h(\mathbf{x}) + \sum_{j=9}^{12} v_j \mathbf{w}_h^j(\mathbf{x})$. Einsetzen der beiden Ansätze in (7.31) und die Variation der Testfunktionen über alle Basisfunktionen mit verschwindender Normalkomponente führt auf das lineare System

$$\begin{bmatrix} \mathbf{M} & -\mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix}. \quad (7.33)$$

Dabei sind die Matrizen und Vektoren durch

$$\begin{aligned} \mathbf{M} &= [(\mathbf{w}_h^j, \mathbf{w}_h^i)_K]_{i,j=9,\dots,12}, & \mathbf{f} &= [-(\tilde{\mathbf{v}}_h, \mathbf{w}_h^i)_K]_{i=9,\dots,12}, \\ \mathbf{B} &= [(\nabla \cdot \mathbf{w}_h^j, q_h^i)_K]_{i=1,\dots,4, j=9,\dots,12}, & \mathbf{g} &= [(\tilde{g} - \nabla \cdot \tilde{\mathbf{v}}_h, q_h^i)_K]_{i=1,\dots,4}, \end{aligned}$$

gegeben. Die Vektoren $\mathbf{v} = (v_9, \dots, v_{12})^T$ und $\mathbf{p}$ ergeben sich aus den gesuchten Koeffizienten der Basisdarstellung.

Nach (7.16) ist die Matrix $\mathbf{B}$ unabhängig vom tatsächlichen Element K und bei Verwendung der Basis aus Abschnitt 6.1.1 für $Q_1(K)$ gilt

$$\mathbf{B} = \begin{bmatrix} 1/4 & 1/4 & -1/4 & -1/4 \\ -1/4 & 1/4 & 1/4 & 1/4 \\ -1/4 & -1/4 & -1/4 & -1/4 \\ 1/4 & -1/4 & 1/4 & 1/4 \end{bmatrix}.$$

Da die letzten Spalten der Matrix identisch sind, hat sie keinen vollen Rang. Tatsächlich gilt $\text{rang}(\mathbf{B}) = 3$. Dies liegt daran, dass wir bisher die Bedingung $\int p_h d\mathbf{x} = 0$ nicht berücksichtigt haben. Bei der hier genutzten Basis wird daraus die Bedingung $\sum_{j=1}^4 p_j = 0$, wobei p_j die Koeffizienten der Basisdarstellung bezeichnet. Ersetzen wir damit dann etwa die letzte Zeile der Systemmatrix von (7.33), so ist das System eindeutig lösbar.

Für uns ist dabei nur die Lösungskomponente $\mathbf{v}$ von Interesse. Mit den Abkürzungen

$$\begin{aligned} \Delta_1 &= M_{34} - M_{44}, & \Delta_2 &= M_{13} - M_{14}, \\ \Delta_3 &= M_{23} - M_{24}, & \Delta_4 &= M_{33} - M_{14}, \\ \delta &= f_4 - f_3, & \sigma_1 &= 2(g_1 + g_3), \\ \sigma_2 &= 2(g_2 + g_3), & \sigma_3 &= 2(g_1 + g_2), \end{aligned}$$

erhalten wir diese durch

$$\begin{aligned} v_9 &= -\sigma_2, & v_{11} &= \frac{\delta - \sigma_1 \Delta_1 - \sigma_2 \Delta_2 + \sigma_3 \Delta_3}{\Delta_1 - \Delta_4}, \\ v_{10} &= \sigma_3, & v_{12} &= \frac{\delta - \sigma_1 \Delta_4 - \sigma_2 \Delta_2 + \sigma_3 \Delta_3}{\Delta_4 - \Delta_1}. \end{aligned}$$

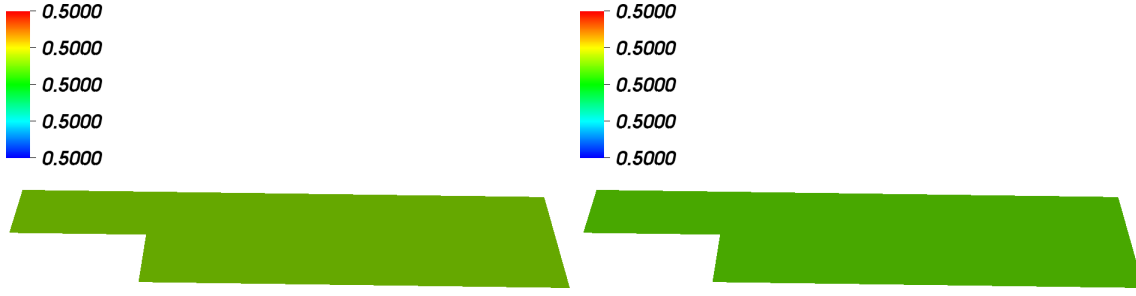


Abbildung 7.3: Numerische Lösung C_h des linearen Transportproblems (7.9) zur Zeit $t = 0$ (links) und $t = 1$ (rechts) bei Verwendung, des mit der LCPF bestimmten, kompatiblen Geschwindigkeitsfeldes

Haben wir $\mathbf{v}_h$ auf den einzelnen Zellen bestimmt, so erhalten wir damit das neue Geschwindigkeitsfeld $\mathbf{u}_h^* = \mathcal{P}_K \mathbf{u}_h + \mathbf{v}_h$. Nach Wahl der Randbedingungen bei der Bestimmung von $\mathbf{v}_h$ ist die Normalkomponente von $\mathbf{u}_h^*$ über Zellgrenzen hinweg stetig. Außerdem gilt

$$(\nabla \cdot \mathbf{u}_h^*, q_h)_K = (g, q_h)_K \quad \forall q_h \in Q_1(K),$$

womit $\mathbf{u}_h^*$ wie gewünscht zur DG-Diskretisierung (6.50) kompatibel ist.

Um zu demonstrieren, dass wir mit der LCPF tatsächlich ein kompatibles Geschwindigkeitsfeld bestimmen können, kommen wir zurück auf den Testfall (7.9) aus Abschnitt 7.1. Analog zur Abbildung 7.1 zeigen wir in Abbildung 7.3 die transportierte Größe C_h wieder für $t = 0$ und $t = 1$. Hier sehen wir allerdings, dass die konstante Lösung wie gewünscht erhalten bleibt.

Zusätzlich veranschaulicht Abbildung 7.4 die bei diesem Testfall berechneten Geschwindigkeitskomponenten vor und nach dem Postprocessing. Auf dem verwendeten Rechtecksgitter ist die Stetigkeit von $\mathbf{u}_h^* \cdot \mathbf{n}$ äquivalent zur Stetigkeit von u_h^* in x -Richtung und der Stetigkeit von v_h^* in y -Richtung, wobei $\mathbf{u}_h^* = (u_h^*, v_h^*)^T$ gilt. Zur Veranschaulichung dieser Eigenschaften zeigt Abbildung 7.4 eine Vergrößerung der verwendeten Geometrie im Bereich der Ecke. Während beim ursprünglichen Geschwindigkeitsfeld $\mathbf{u}_h$ keine Stetigkeit in irgendeine Richtung zu erkennen ist, zeigt das kompatible Geschwindigkeitsfeld die geforderte Stetigkeit von u_h^* und v_h^* in die jeweilige Richtung.

7.5 Kopplung

Nun sind wir in der Lage, das gekoppelte System (2.37) zu lösen. Bei gegebenen Daten $\mathbf{u}_h^n$, p_h^n und C_h^n erfolgt der Ablauf in drei wesentlichen Schritten.

1. Berechnung eines kompatiblen Geschwindigkeitsfeldes $\mathbf{u}_h^*$ aus $\mathbf{u}_h^n$ mittels der LCPF aus Abschnitt 7.4
2. Bestimmung von C_h^{n+1} mit dem DG-Verfahren (6.50) auf Basis des kompatiblen Geschwindigkeitsfeldes $\mathbf{u}_h^*$

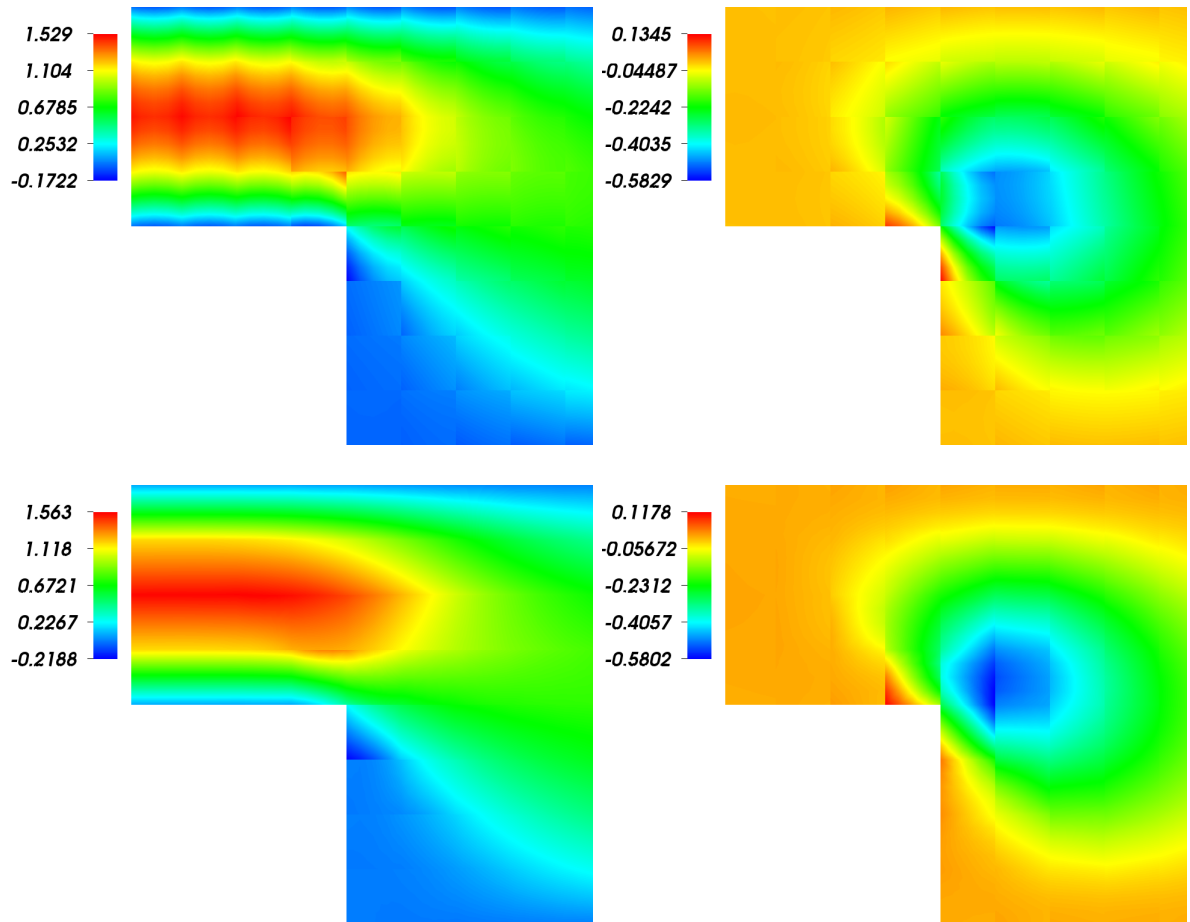


Abbildung 7.4: Die Komponenten des ursprünglichen Geschwindigkeitsfeldes $\mathbf{u}_h$ (oben) und die des kompatiblen Geschwindigkeitsfeldes $\mathbf{u}_h^*$ (unten). Die x -Komponente ist jeweils links, die y -Komponente rechts geplottet.

3. Berechnung von $\mathbf{u}_h^{n+1}$ und p_h^{n+1} durch das diskrete nichtlineare Projektions-Verfahren (4.36)-(4.40)

Bei numerischen Experimenten konnte kein signifikanter Unterschied zwischen den beiden Varianten (4.36) und (4.37) festgestellt werden. Daher nutzen wir im Folgenden stets das Verfahren mit (4.36), da hier die Berechnung und Addition einer zusätzlichen Matrix entfällt.

Anhand eines konstruierten Testfalls wollen wir zeigen, dass sich die Konvergenz der Einzelverfahren ohne Ordnungsverlust auf das gekoppelte Verfahren überträgt. Dazu betrachten wir auf $\Omega = [0, 2]^2$ in Anlehnung an (2.33) das System

$$\begin{aligned} \varrho(\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u}) - \nabla \cdot (2\eta \mathbf{D}) + \nabla p &= \mathbf{f}_u, \\ \nabla \cdot \mathbf{u} &= g, \\ \partial_t \alpha + \nabla \cdot (\alpha \mathbf{u}) &= f_\alpha, \end{aligned} \tag{7.34}$$

mit Dichte und Viskosität

$$\varrho = 2 + 9\alpha/10, \quad \eta = 3 + 9\alpha/10.$$

Die rechten Seiten sind so gewählt, dass sich

$$\begin{aligned} u &= \sin(x + y), & v &= \cos(x - y), \\ \alpha &= \sin(x - y), & p &= \cos(x + y) - \cos(2) \sin(1)^2, \end{aligned}$$

als stationäre Lösung des Problems ergibt. Dabei verschwindet der Druck im Mittel über Ω , aufgrund der gewählten Konstanten. Neben der inhomogenen Divergenzbedingung, sowie variabler Dichte und Viskosität, steht hier die Kopplung der drei Gleichungen im Vordergrund. Dazu bestimmen wir mit der LCPF ein zur Diskretisierung (6.50) kompatibles Geschwindigkeitsfeld und nutzen dies innerhalb des DG-Verfahrens.

Zur Diskretisierung nutzen wir quadratische Gitter mit Maschenweite $\Delta x = 2^{-l}$ für $l = 0, \dots, 4$ und geben auf dem Gebietsrand Dirichlet-Randbedingungen vor. Die Fehler, sowie die daraus resultierende Konvergenzordnung, sind in Tabelle 7.1 aufgelistet. Die Tabelle zeigt, dass die Ordnung der Einzelverfahren erhalten wurde. Insbesondere hat die Verwendung des durch die LCPF bestimmten kompatiblen Geschwindigkeitsfeldes keine nachteilige Auswirkung. Dies erklärt sich durch die Fehlerabschätzung (7.28) die nun lokale Gültigkeit besitzt. Voraussetzung dafür ist aber, dass die verwendeten Randdaten bereits die erforderliche Ordnung aufweisen.

Verglichen mit (2.33), vernachlässigt der zuvor beschriebene Testfall (7.34) die Abhängigkeit der Geschwindigkeitsdivergenz von p und α . Des Weiteren stimmen die rechten Seite von (2.33b) und (2.33c) überein, was in (7.34) ebenfalls nicht berücksichtigt wurde. Daher betrachten wir noch einen zweiten Testfall, um die Konvergenz des Verfahrens auch für diese Fälle zu zeigen. Die Gleichungen für den weiteren Testfall lauten

$$\begin{aligned} \varrho(\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u}) - \nabla \cdot (2\eta \mathbf{D}) + \nabla p &= \mathbf{f}_u, \\ \nabla \cdot \mathbf{u} &= 12\alpha\sqrt{p}, \\ \partial_t \alpha + \nabla \cdot (\alpha \mathbf{u}) &= 12\alpha\sqrt{p}, \end{aligned} \tag{7.35}$$

Tabelle 7.1: Fehler und Konvergenzordnung für das Testproblem (7.34)

Gitter l	$\ u - u_h\ _{L^2(\Omega)}$		$\ v - v_h\ _{L^2(\Omega)}$	
	Fehler	Ordnung	Fehler	Ordnung
0	0.1945	-	0.1831	-
1	0.0502	1.9540	0.0444	2.0447
2	0.0127	1.9869	0.0110	2.0099
3	0.0032	1.9960	0.0027	2.0027
4	0.0008	1.9951	0.0007	1.9995

Gitter l	$\ p - p_h\ _{L^2(\Omega)}$		$\ \alpha - \alpha_h\ _{L^2(\Omega)}$	
	Fehler	Ordnung	Fehler	Ordnung
0	0.6843	-	0.1601	-
1	0.3427	0.9978	0.0297	2.4308
2	0.1629	1.0724	0.0068	2.1300
3	0.0794	1.0363	0.0017	2.0177
4	0.0393	1.0147	0.0004	1.9936

auf $\Omega = [0, 1] \times [0, 1]$ mit $\varrho = \eta = 1$. Die rechte Seite $\mathbf{f}_u$ wählen wir dabei so, dass sich

$$\begin{aligned}
 u(x, y, t) &= \frac{6e^{8t^{3/2}}\sqrt{t}x}{9 + e^{8t^{3/2}}}, & v(x, y, t) &= \frac{6e^{8t^{3/2}}\sqrt{t}y}{9 + e^{8t^{3/2}}}, \\
 p(x, y, t) &= t, & \alpha(x, y, t) &= 1 - \frac{9}{9 + e^{8t^{3/2}}}
 \end{aligned}$$

als Lösung des Systems ergibt. In den folgenden Rechnungen wurden am rechten Randstück Neumann-Randbedingungen gemäß (2.39) und auf dem verbleibenden Rand Dirichlet-Randdaten vorgegeben. Ausgehend vom Startzeitpunkt $t_0 = 0.1$ wird die Lösung zum Zeitpunkt $t_1 = 1.1$ unter Verwendung der Zeitschrittweiten $\Delta t = 5120^{-1} \times 2^{-l}$ für $l = 0, \dots, 4$ berechnet. Fehler und Ordnung sind in Abbildung 7.2 dargestellt. Wie erwartet ist das Verfahren zeitlich von erster Ordnung.

Tabelle 7.2: Fehler und Konvergenzordnung für das Testproblem (7.35)

Δt	$\ u(t_1) - u_h(t_1)\ _{L^2(\Omega)}$		$\ v(t_1) - v_h(t_1)\ _{L^2(\Omega)}$	
l	Fehler	Ordnung	Fehler	Ordnung
0	2.0628e-04	-	4.1972e-05	-
1	1.1256e-04	0.8738	2.1047e-05	0.9958
2	6.0436e-05	0.8973	1.0547e-05	0.9967
3	3.1828e-05	0.9251	5.2817e-06	0.9978
4	1.6462e-05	0.9511	2.6434e-06	0.9986

Δt	$\ p(t_1) - p_h(t_1)\ _{L^2(\Omega)}$		$\ \alpha(t_1) - \alpha_h(t_1)\ _{L^2(\Omega)}$	
l	Fehler	Ordnung	Fehler	Ordnung
0	3.2860e-03	-	1.4497e-03	-
1	1.8926e-03	0.7960	8.3657e-04	0.7932
2	1.0624e-03	0.8330	4.7025e-04	0.8311
3	5.7878e-04	0.8763	2.5640e-04	0.8750
4	3.0635e-04	0.9178	1.3580e-04	0.9170

8 Numerische Simulation

In diesem Kapitel zeigen wir Simulationen hochviskoser Kunststoffschmelzen nach (2.14), sowie von Kavitation mit dem Modell (2.33). Dabei geht es uns nicht darum ingenieurstechische Anwendungen genau wiederzugeben, sondern anhand einfacher Testfälle das in den vorhergehenden Kapiteln entwickelte Verfahren zu validieren und das prinzipielle Verhalten der Modelle aufzuzeigen.

Zur Untersuchung hochviskoser Kunststoffschmelzen betrachten wir in Abschnitt 8.1 die Poiseuille-Strömung und eine 4:1-Kontraktion als Testfälle. Für die Poiseuille-Strömung im eingelaufen Zustand können wir nach der Diskussion in Paragraph 2.5 eine numerische Referenzlösung bestimmen, ist die Viskosität konstant, so liegt sogar die analytische Lösung vor.

Anschließend werden in Abschnitt 8.2 Simulationen der Kavitation betrachtet. Als Testfall dient dabei die lokale Ausgasung über einem Backward-Facing-Step.

8.1 Hochviskose Strömungen

Wie in Abschnitt 2.2 ausgeführt, nutzen wir das Cross-Modell (2.11) zur Beschreibung der Viskosität. Die in (2.11) gegebene Darstellung des Modells ist dimensionsbehaftet und muss somit noch entdimensionalisiert werden. Wir wählen dazu $\eta_{\text{ref}} = \eta_0$, womit das dimensionslose Cross-Modell

$$\eta(\dot{\gamma}, \theta) = \frac{a(\theta)}{(1 + \kappa_1 \dot{\gamma} a(\theta))^{\kappa_2}}, \quad (8.1)$$

lautet. Weiter behält das WLF-Modell (2.12) zur Beschreibung der Temperaturabhängigkeit seine Form auch im dimensionslosen Fall. Insbesondere halten wir die Schreibweise wie zuvor einfach, indem wir bei der Bezeichnung dimensionsloser und dimensionsbehafteter Größen keine Unterschiede machen. Innerhalb des Cross-Modells (8.1) nutzen wir im Weiteren stets die Parameter

$$\kappa_1 = \kappa_2 = 0.7 \quad (8.2)$$

und im WLF-Modell (2.12) verwenden wir

$$C_1 = 1, \quad C_2 = 0.2, \quad \theta_s = 0.5. \quad (8.3)$$

Diese Wahl der Parameter bezieht sich nicht auf ein konkretes Material, sondern wurde so getroffen, dass bei der in Abschnitt 8.1.1 betrachteten Poiseuille-Strömung deutliche Unterschiede zwischen newtonschem und nicht-newtonschem Verhalten auftreten.

Das zu diesen Parametern gehörige Verhalten der Viskositätsfunktion wurde bereits in Abbildung 2.1 dargestellt.

Bei der Betrachtung des Reibungsspannungstensors in der Form (2.8) werden auch Ableitungen der Viskosität benötigt. Diese lauten

$$\begin{aligned}\partial_x \eta &= \frac{(1 - (\kappa_1 - 1)\kappa_2 a(\theta)\dot{\gamma})a'(\theta)\partial_x \theta - \kappa_1 \kappa_2 a(\theta)^2 \partial_x \dot{\gamma}}{(1 + \kappa_1 a(\theta)\dot{\gamma})^{\kappa_2 + 1}}, \\ \partial_y \eta &= \frac{(1 - (\kappa_1 - 1)\kappa_2 a(\theta)\dot{\gamma})a'(\theta)\partial_y \theta - \kappa_1 \kappa_2 a(\theta)^2 \partial_y \dot{\gamma}}{(1 + \kappa_1 a(\theta)\dot{\gamma})^{\kappa_2 + 1}},\end{aligned}$$

wobei die Ableitung des WLF-Modells (2.12) bezüglich der Temperatur durch

$$a'(\theta) = -\frac{C_1 C_2 a(\theta)}{(C_2 + \theta - \theta_s)^2}$$

gegeben ist. Während die dabei auftretenden partiellen Ableitungen der Temperatur direkt ausgewertet werden können, muss bei den Ableitungen der Scherrate eine Fallunterscheidung getroffen werden. Die Scherrate ist in (2.13) definiert und die partiellen Ableitungen ergeben sich als

$$\begin{aligned}\partial_x \dot{\gamma} &= \frac{2(\partial_y v \partial_{xy} v + \partial_x u \partial_{xx} u) + (\partial_y u + \partial_x v)(\partial_{xy} u + \partial_{xx} v)}{\dot{\gamma}}, \\ \partial_y \dot{\gamma} &= \frac{2(\partial_y v \partial_{yy} v + \partial_x u \partial_{xy} u) + (\partial_y u + \partial_x v)(\partial_{yy} u + \partial_{xy} v)}{\dot{\gamma}}.\end{aligned}$$

Im Fall $\dot{\gamma} = 0$ sind die obigen Ableitungen undefiniert. Ein derartiger Fall tritt insbesondere bei konstanten Anfangsdaten $\mathbf{u}^0(\mathbf{x}) = \text{const}$ auf. Dann gilt aber neben $\dot{\gamma}(\mathbf{x}) = 0$ auch $\partial_x \dot{\gamma}(\mathbf{x}) = \partial_y \dot{\gamma}(\mathbf{x}) = 0$, womit dieser Fall explizit abgefangen werden kann. Natürlich kann die Scherrate auch in anderen Situationen verschwinden, da die Auswertung der Ableitungen aber nur an wenigen diskreten Stellen benötigt wird und aufgrund der auftretenden Rundungsfehler, sind im Allgemeinen keine Probleme zu erwarten. Bei den nachfolgenden Rechnungen traten in der Tat keine Probleme dieser Art auf.

Konkret wird die Auswertung der Viskosität und damit auch die Berechnung der partiellen Ableitungen der Scherrate an zwei Stellen benötigt. Während im Strömungslöser das Geschwindigkeitsfeld $\mathbf{u}_h \in \mathbf{U}_h$ zur Berechnung von $\mathcal{L}(\eta)$ in (3.23) verwendet wird, muss innerhalb des DG-Verfahrens das kompatible Geschwindigkeitsfeld $\mathbf{u}_h^*$ zur Integration des Quellterms (2.10) genutzt werden. Lokal gilt $\mathbf{u}_h^*|_K \in RT_1(K)$ und wir wollen nun beschreiben, wie Ableitungen dieser Funktionen berechnet werden können. Dabei beschränken wir uns auf Rechtecksgitter, womit die Referenzabbildung F affin wird und die zugehörige Jacobi-Matrix DF konstante Einträge hat. Damit erhalten wir nach (7.15) für $\mathbf{w} \in RT_1(K)$ die Darstellung

$$\mathbf{w}(\mathbf{x}) = \frac{1}{\det DF} DF \hat{\mathbf{w}}(\hat{\mathbf{x}}) \Big|_{\hat{\mathbf{x}}=F^{-1}(\mathbf{x})}$$

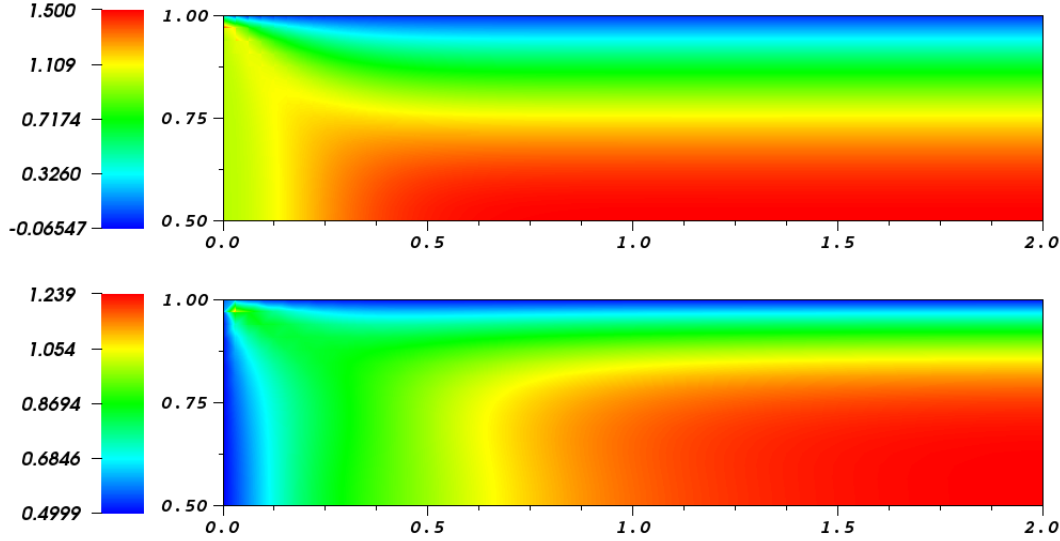


Abbildung 8.1: Die x -Geschwindigkeit (oben) und Temperatur (unten) bei konstanter Viskosität für $Re = 0.01$ und $Pe = 1$

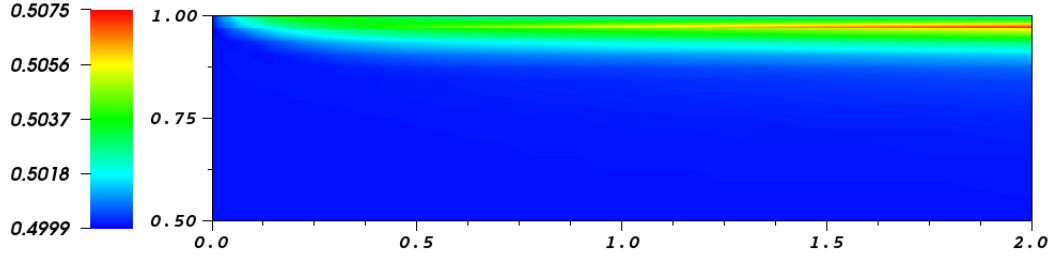
und die gesuchten Ableitungen ergeben sich aus

$$D\mathbf{w}(\mathbf{x}) = \frac{1}{\det DF} (DF) D\hat{\mathbf{w}}(\hat{\mathbf{x}}) (DF^{-1}).$$

8.1.1 Poiseuille-Strömung

Wir betrachten in diesem Abschnitt newtonsche wie auch nicht-newtonsche Poiseuille-Strömungen im endlichen Kanal $\Omega = [0, 2] \times [0, 1]$, wobei wir unterschiedliche Randbedingungen auf den einzelnen Randstücken vorschreiben. Für das Geschwindigkeitsfeld setzen wir links am Einfluss $\mathbf{u} = (1, 0)^T$, an der oberen Wand $\mathbf{u} = 0$ und rechts am Austritt Ausflussrandbedingungen gemäß (2.39). Das verbleibende Randstück wird als Symmetrierand behandelt, wir setzen dort $\partial_y u = 0$ und $v = 0$. Für die Temperatur geben wir die Dirichlet-Bedingung $\theta = 0.5$ am Eintritt und der Wand vor, am Austritt und auf dem Symmetrierand setzen wir die homogene Neumann-Bedingung $\nabla \theta \cdot \mathbf{n} = 0$. Weiter beträgt die Reynolds-Zahl im gesamten Abschnitt $Re = 0.01$ und zur Diskretisierung des Kanals wurde ein quadratisches Gitter mit 64×16 Zellen verwendet, was einer Mascheinweite von $\Delta x = 0.0625$ entspricht.

In Abbildung 8.1 sind die stationären Zustände der x -Geschwindigkeit und des Temperaturfeldes bei konstanter Viskosität und Péclet-Zahl $Pe = 1$ dargestellt. Bei Betrachtung des Geschwindigkeitsfeldes erkennen wir, dass sich das Geschwindigkeitsprofil nach einer kurzen Einlaufphase in x -Richtung nicht mehr ändert, man spricht dann vom hydrodynamisch eingelaufenen Zustand. Entsprechendes wird beim Temperaturfeld als thermisch eingelaufener Zustand bezeichnet. Die jeweilige Einlauflänge bis ein derartig eingelaufener Zustand erreicht ist, verhält sich dabei proportional zur Reynolds- bzw.

Abbildung 8.2: Temperaturverteilung bei konstanter Viskosität und $Pe = 10^5$

Péclet-Zahl. Da hier $Re = 0.01 < 1 = Pe$ gilt, ist die hydrodynamische Einlaufänge deutlich geringer als die thermischen Einlaufänge. Erhöhen wir die Péclet-Zahl so wird die thermische Einlaufänge weiter ausgedehnt und es zeigt sich, das in Abbildung 8.2 für $Pe = 10^5$ geplottete, Grenzschichtverhalten.

Unter der Annahme konstanter Viskosität konnten wir in Abschnitt 2.5 analytische Darstellungen für die eingelaufenen Geschwindigkeits- und Temperaturprofile bestimmen. Diese wollen wir nun mit numerisch ermittelten Profilen vergleichen. Um den eingelaufenen Zustand numerisch zu bestimmen, berechnen wir zunächst eine stationäre Lösung wie in Abbildung 8.1 dargestellt. Anschließend nutzen wir die daraus resultierenden Daten am Ausflussrand als neue Randbedingungen am Einfluss. Sollten der daraus ermittelte stationäre Zustand weiterhin unterschiedliche Werte am Einfluss- und Ausflussrand aufweisen, so wiederholen wir diese Prozedur, bis die Differenzen von Geschwindigkeit und Temperatur am Ein- und Ausfluss zu vernachlässigen sind. Die so bei $Pe = 1$ numerisch ermittelten eingelaufenen Zustände werden in Abbildung 8.3 gezeigt. Wir verzichten auf die Darstellung bei höheren Péclet-Zahlen, denn aus Abschnitt 2.5 wissen wir, dass Geschwindigkeit und Temperatur im eingelaufenen Zustand unabhängig von dieser Kennzahl sind. Den Vergleich der numerisch ermittelten Daten mit den analytischen Lösungen (2.44) und (2.45) sehen wir in Abbildung 8.4, wobei sich sehr gute Übereinstimmungen zeigen.

Zur Bestimmung der analytischen Lösung aus (2.44) und (2.45) wird der richtige Wert für δp benötigt. Dieser kann über den Massenstrom

$$\mu = \int_0^1 \varrho u \, dy = \int_0^1 u \, dy$$

bestimmt werden, wobei für die Dichte im hier betrachteten dimensionslosen Fall $\varrho = 1$ gilt. Da μ unabhängig von x ist und am Einfluss durch

$$\int_0^1 u \, dy = \int_0^1 dy = 1$$

gegeben ist, ergibt sich δp nach (2.44) als Lösung der Gleichung

$$1 = \int_0^1 \left[\frac{1}{2} Re \delta p y (1 - y) \right] dy$$

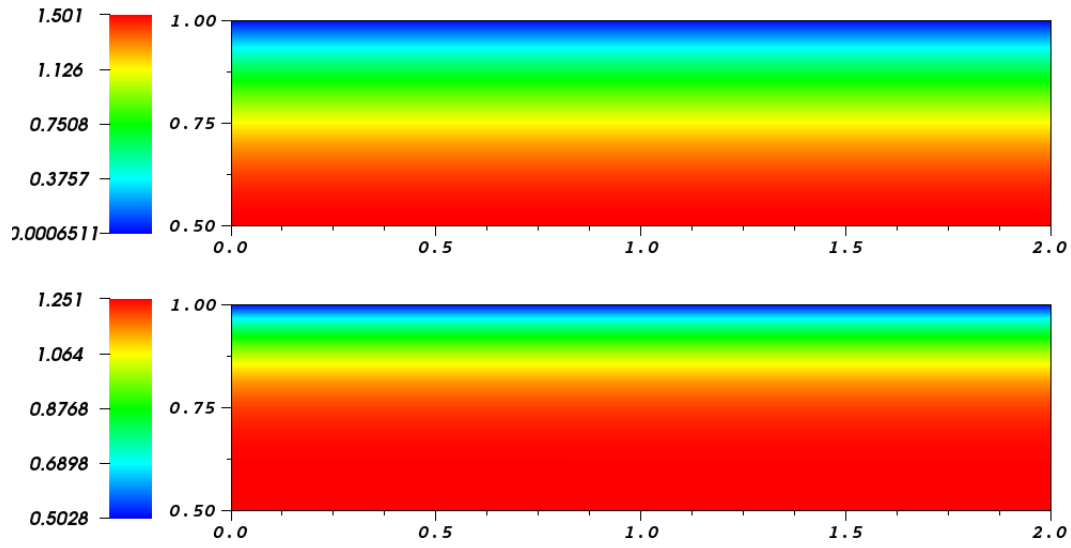


Abbildung 8.3: Hydrodynamisch eingelaufener Zustand der x -Geschwindigkeit (oben) und thermisch eingelaufener Zustand des Temperaturfeldes (unten) bei $Re = 0.01$

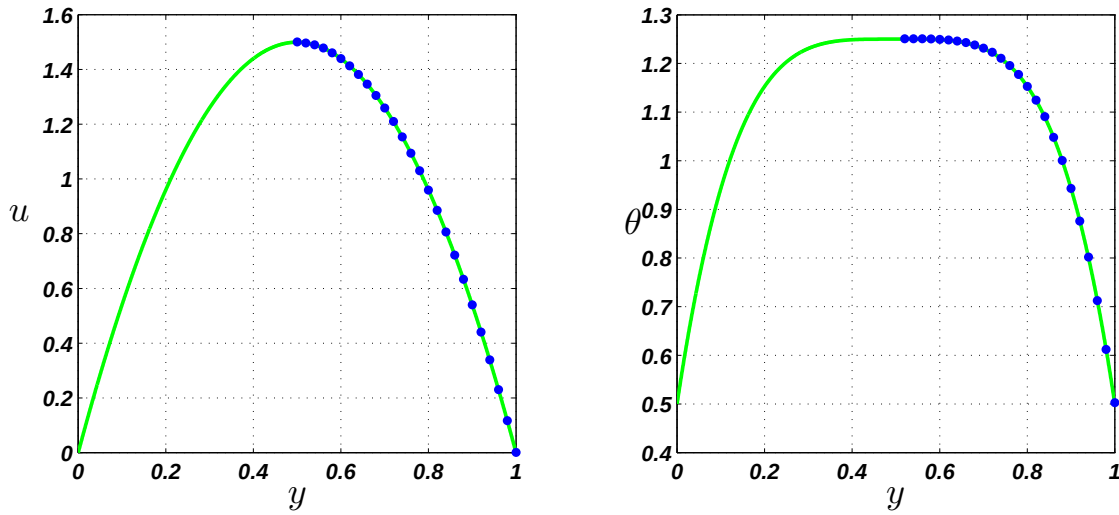


Abbildung 8.4: Analytische Lösung (grün) und numerische Approximation (blau) des hydrodynamisch eingelaufenen Geschwindigkeitsprofils (links) und des thermisch eingelaufenen Temperaturprofils (rechts)

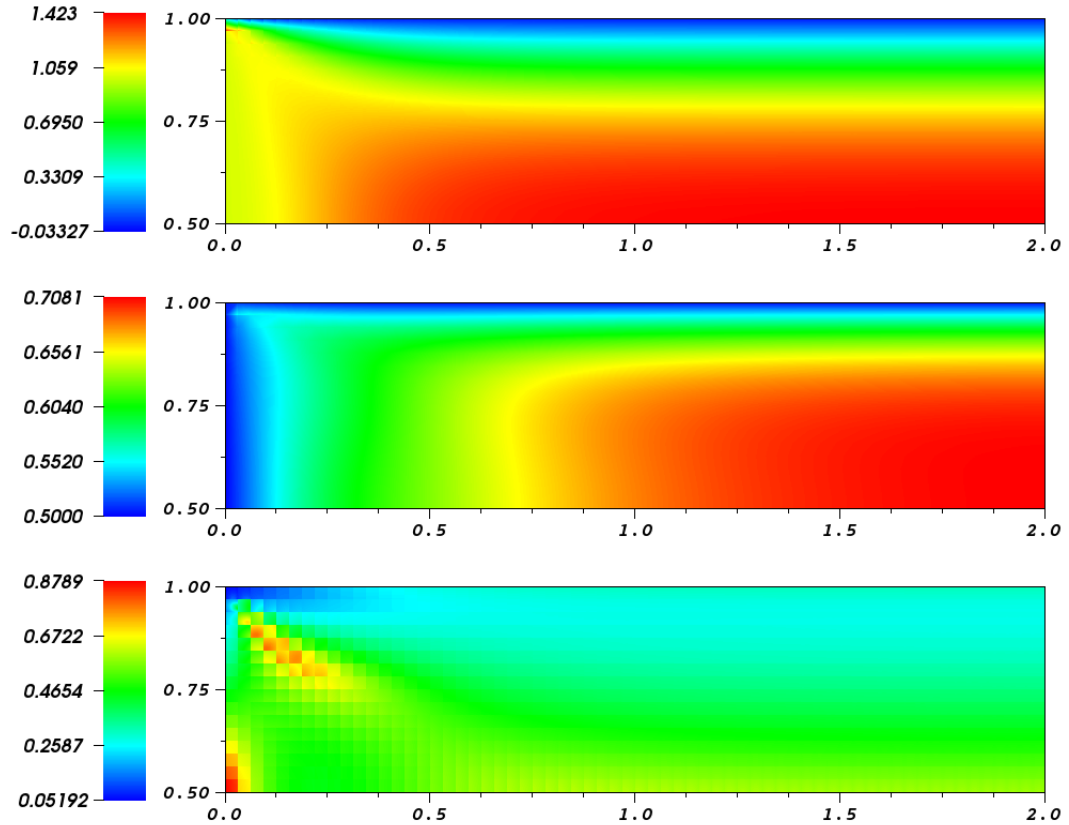


Abbildung 8.5: Die x -Geschwindigkeit (oben), Temperatur (mitte) und Viskosität (unten) bei nicht-newtonscher Viskosität für $Re = 0.01$ und $Pe = 1$

und somit gilt $\delta p = 12/Re$.

Nun wollen wir den entsprechenden nicht-newtonschen Fall betrachten. Die im dabei verwendeten Cross-Modell (8.1) genutzten Parameter sind in (8.2) und (8.3) gegeben und in Abbildung 8.5 werden x -Geschwindigkeit, Temperatur und Viskosität im stationären Zustand gezeigt. Verglichen mit den in Abbildung 8.1 dargestellten Daten, erkennen wir einen leichten Abfall der Geschwindigkeit und deutlich geringere Temperaturen. Diese Änderungen gehen auf die im Vergleich zum newtonschen Fall mit $\eta = 1$ wesentlich geringere Viskosität zurück. Zusätzlich werden noch die zugehörigen eingelaufen Zustände in Abbildung 8.6 dargestellt.

Mit dem in 2.5 beschriebenen Verfahren können wir im nicht-newtonschen Fall eine numerische Referenzlösung für die eingelaufenen Zustände bestimmen. Bei der Wahl von δp ist dabei wie oben zu beachten, dass der Massenstrom $\mu = 1$ erfüllt. Die so erzeugte Referenzlösung wird in Abbildung 8.7 mit den numerischen Daten verglichen, wobei sich auch in diesem Fall eine sehr gute Übereinstimmung zeigt.

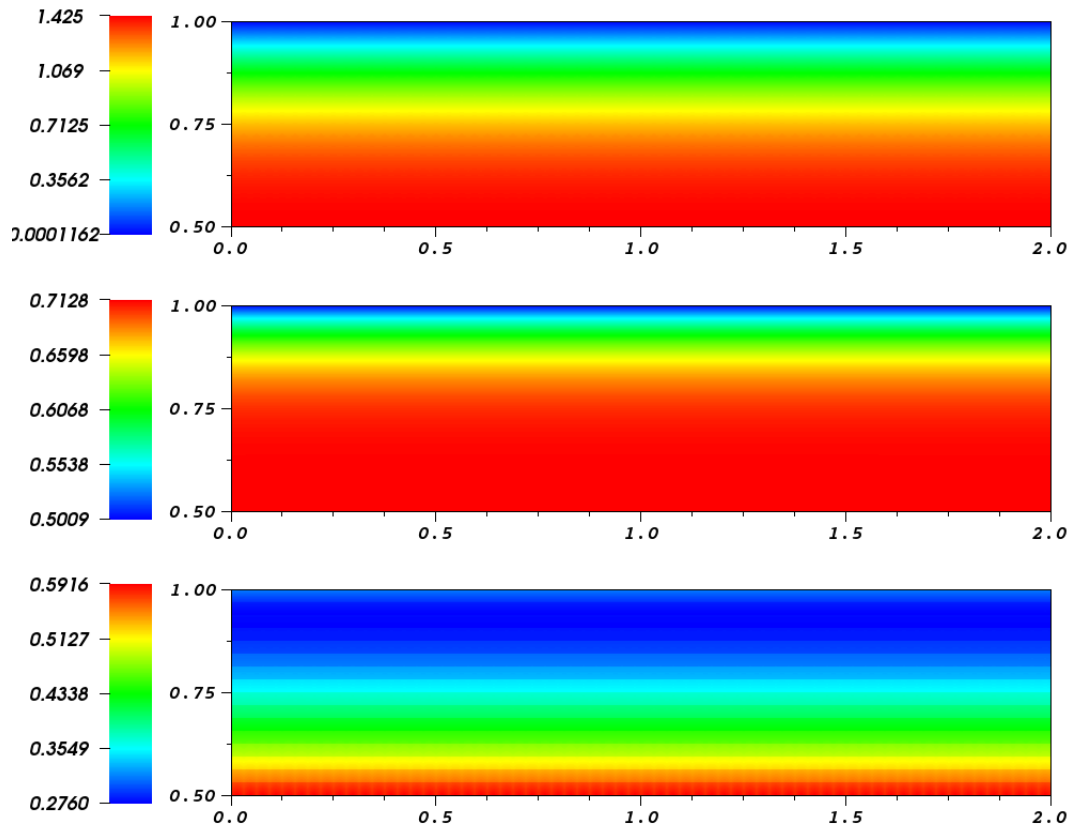


Abbildung 8.6: Nicht-newtonsch eingelaufene Zustände unabhängig von der Péclet-Zahl bei $Re = 0.01$. x -Geschwindigkeit (oben), Temperatur (mitte), Viskosität (unten)

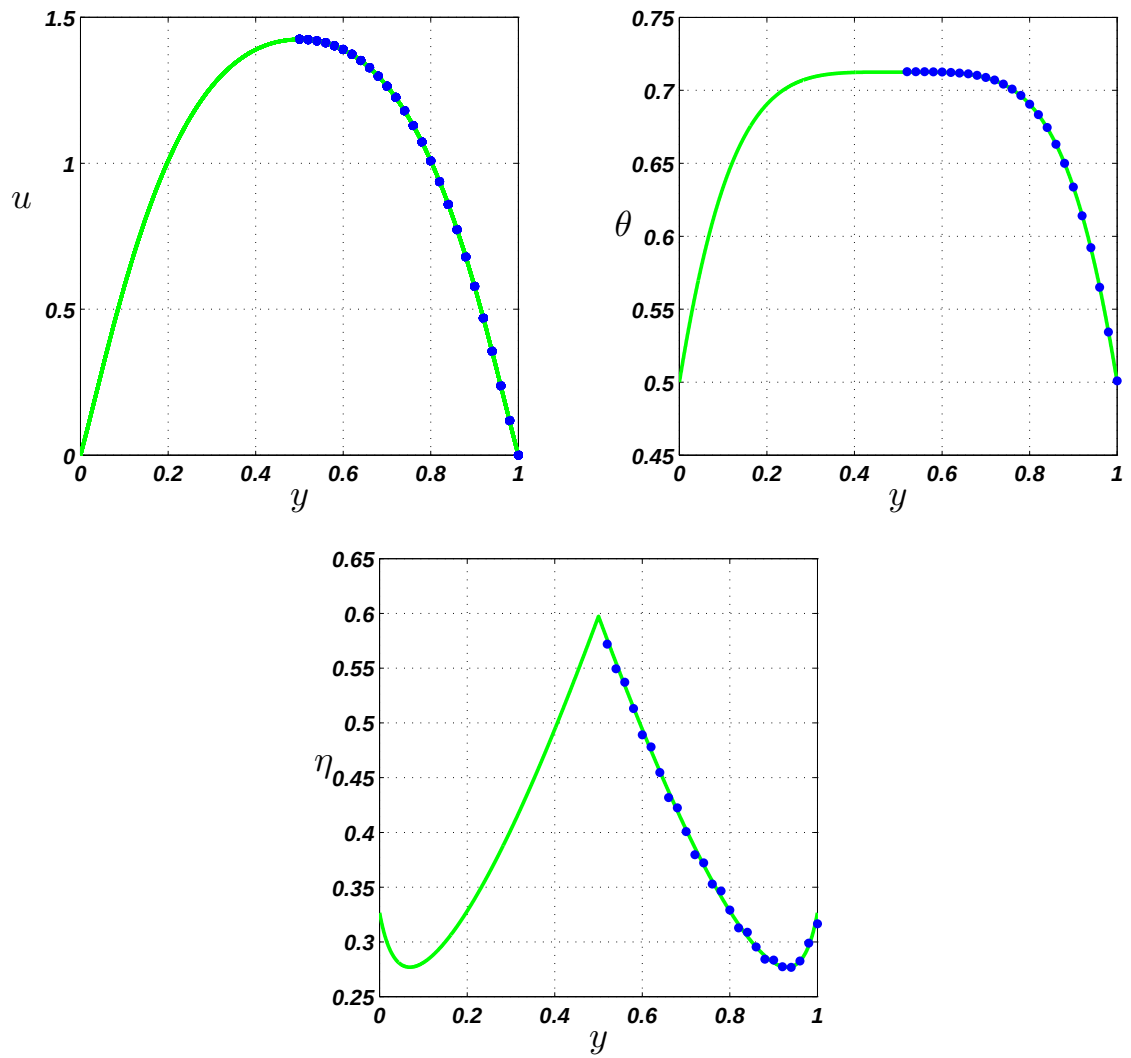


Abbildung 8.7: Referenzlösung (grün) und numerische Approximation (blau) des hydrodynamisch eingelaufenen Zustands (links) und des thermisch eingelaufenen Temperaturprofils (rechts) im nicht-newtonschen Fall

8.1.2 4:1-Kontraktion

Als weiteren Testfall betrachten eine nicht-newtonsche Strömung in einer 4:1-Kontraktion, deren Geometrie aus Abbildung 8.8 ersichtlich ist. Auch hier geben wir wieder unterschiedliche Randbedingungen vor, wobei wir links am Eintritt $\mathbf{u} = (1, 0)^T$, am unteren Randstück Symmetrie- und rechts Ausflussrandbedingungen setzen. Die verbleibenden Randstücke modellieren feste Wände, weshalb wir dort $\mathbf{u} = 0$ vorschreiben. Für das Temperaturfeld schreiben wir am Symmetrie- und am Ausflussrand homogene Neumann-Randbedingungen vor, während die restlichen Randstücke unterschiedlich beheizt werden. Dazu setzen wir $\theta = 0.63$ am oberen Rand und $\theta = 0.5$ auf den verbleibenden Teilstücken. Das zur Diskretisierung verwendete quadratische Gitter hat eine Maschenweite von $\Delta x = 0.03125$ und die Kennzahlen wurden als $Re = 0.01$ und $Pe = 10^3$ gewählt. Für eine bessere Visualisierung betrachten wir hier keine höheren Péclet-Zahlen, da in diesen Fällen das Grenzschichtverhalten zu stark wird.

Die Wahl der Parameter in (8.3) enthält die Annahme, dass die Differenz $\theta - \theta_s$ nicht negativ wird. Diese Eigenschaft gilt es auch mit den numerischen Verfahren zu sichern, da die Viskosität ansonsten exponentiell anwächst anstatt zu fallen. Mit der Wahl von $\theta^0 = 0.5$ darf die Temperatur bei den gegebenen Randbedingungen im zeitlichen Verlauf aus physikalischer Sicht nicht geringer werden, denn durch den dissipativen Quellterm (2.10) und die beheizte Wand am oberen Rand kann es nur zu Temperaturerhöhungen kommen. Aber gerade die beheizte Wand kann zu numerischen Problemen führen, da auf groben Gittern zu Beginn der Zeititeration starke Gradienten auftreten, die zu Unterschwingern führen können. Mit dem Limiter (6.20) aus Abschnitt 6.1.3 können wir aber die Einhaltung der unteren Schranke $\theta \geq 0.5$ sichern.

Für diesen Testfall werden die stationären Zustände der einzelnen Größen in Abbildung 8.8 gezeigt. Bei Betrachtung des Geschwindigkeitsfeldes erkennt man die deutliche Erhöhung der Geschwindigkeit in der Verengung, sowie das Totwassergebiet in der rechten oberen Ecke des aufgeweiteten Kanalstücks. Das Temperaturfeld verdeutlicht die durch die hohe Péclet-Zahl gegebene Dominanz des Wärmetransports am Einfluss, sowie den durch die beheizte Wand gegebenen überwiegenden Wärmefluss im Totwassergebiet. Zusätzlich zeigt sich bedingt durch die Dissipation eine starke Temperaturgrenzschicht an der oberen Wand nach der Verengung des Kanals, was in Kombination mit den hohen Scherraten dort einen starken Abfall der Viskosität verursacht. Ebenso sinkt die Viskosität aufgrund der hohen Scherraten im Bereich der Kontraktion.

8.2 Kavitation

In diesem Abschnitt betrachten wir Simulationen von Kavitation auf Basis des Modells (2.33), um zu zeigen, dass die zuvor an Spezialfällen getesteten Komponenten des Verfahrens auch im komplexen Zusammenspiel funktionieren. Dazu nutzen wir die aus Abbildung 8.9 ersichtliche Geometrie eines Backward-Facing-Step. Ausgangspunkt für die Simulation der Kavitation ist dabei der stationäre Zustand eines newtonschen inkompressiblen Fluids, zu dessen Bestimmung am linken Rand ein parabolisches Geschwindig-

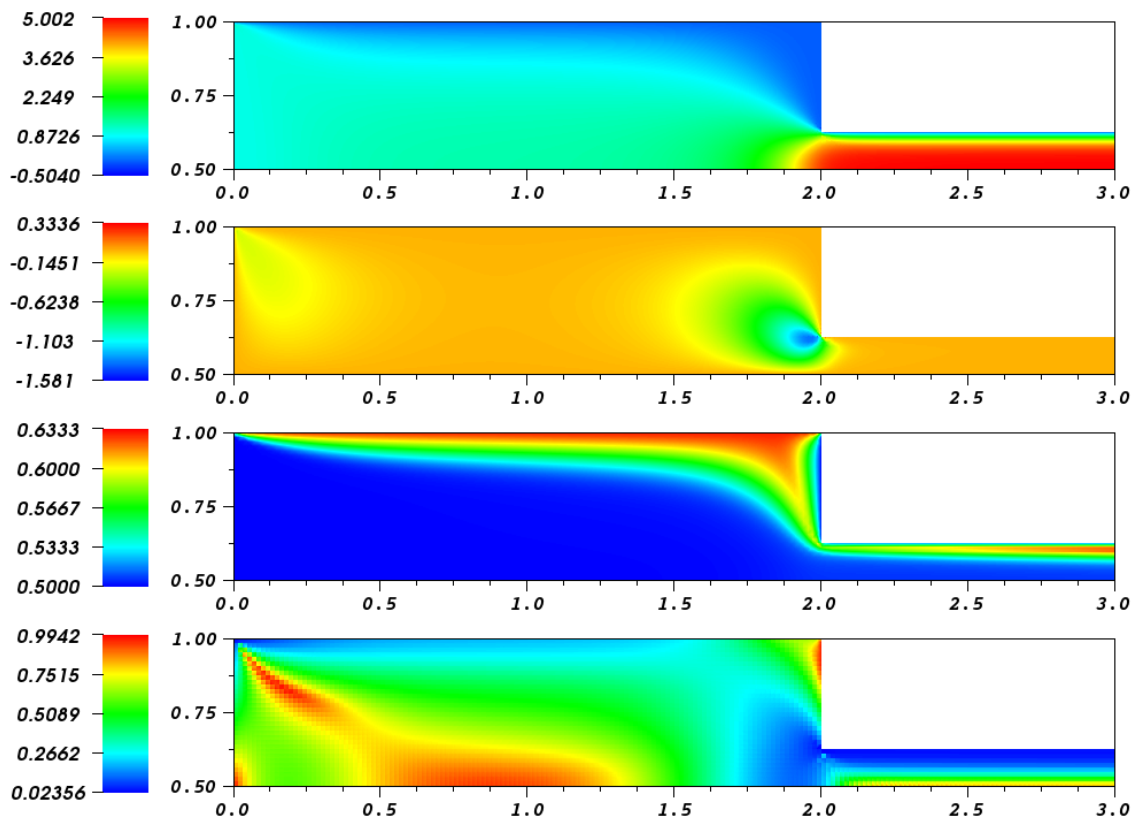


Abbildung 8.8: Stationärer Zustand für die 4:1-Kontraktion bei $Pe = 10^3$. Von oben nach unten sind x -Geschwindigkeit, y -Geschwindigkeit, Temperatur und Viskosität dargestellt

keitsprofil und am rechten Rand Ausflussrandbedingungen (2.39) vorgeschrieben werden. Während der obere Rand als Symmetrierand ($\partial_y u = 0, v = 0$) behandelt wird, werden die restlichen Randstücke als feste Wände ($\mathbf{u} = 0$) betrachtet. In Abbildung 8.9 ist der bei einer Reynolds-Zahl von $Re = 0.05$ so bestimmte stationäre Zustand der Geschwindigkeitskomponenten und des Drucks dargestellt. Die hierbei verwendete Maschenweite des äquidistanten Gitters beträgt $\Delta x = 0.0625$ und die instationäre Rechnung wurde abgebrochen als die approximierte zeitliche Änderung der einzelnen Größen kleiner als 10^{-6} war. Für den Druck etwa wurde die zeitliche Änderung durch $\partial_t p \approx (\mathbf{p}^{n+1} - \mathbf{p}^n)/\Delta t$ angenähert, wobei die Differenz mit der diskreten L^2 -Norm gemessen wurde.

Die Diskussion in Abschnitt 3.1.2 hat gezeigt, dass durch die Randbedingung (2.39) der gemittelte Druck am Austritt Γ verschwindet. Folglich gilt

$$\frac{1}{|\Gamma|} \int_{\Gamma} p \, d\mathbf{s} = 0,$$

womit sich in der gewählten Konfiguration sogar $p|_{\Gamma} \approx 0$ ergibt. Da der lokale Druckabfall an der Aufweitung des Kanal stärker ausfällt als der Druckabfall in x -Richtung, kommt es somit zu negativen Werten des Drucks im Bereich des Steps. Dies stellt für das weitere Vorgehen aber keine Einschränkung dar und könnte durch eine entsprechende Modifikation der Ausfluss-Randbedingung vermieden werden.

Weiter merken wir an, dass das aus Abbildung 8.9 ersichtliche Minimum von u , sowie das Maximum von v vom Betrag her zu groß sind. Die Ursache hierfür liegt in den verwendeten Ansatzfunktionen und die angegebenen Werte werden tatsächlich nur in den direkt an der Ecke liegenden Zellen angenommen. Dies lässt sich insbesondere am Plot der y -Geschwindigkeit erkennen.

Ausgehend vom stationären Zustand der newtonschen inkompressiblen Strömung wollen wir nun Kavitationseffekte simulieren. Die Idee dahinter ist, dass, nachdem zunächst keine Blasen in der Strömung vorhanden sind, ein mit kleinen Blasen versetztes Fluid in die Geometrie einströmt. Anzahl und Größe der Blasen setzen wir auf $n_0 = 10^8$ und $R = 2 \times 10^{-5}$ und erhalten damit nach (2.28) $\alpha \approx 0.11$, was wir als Dirichlet-Randbedingung für die Dampfdichte am Einfluss vorgeben. An allen weiteren Randstücken verwenden wir homogene Neumann-Randbedingungen, was insbesondere an den festen Wänden nötig ist, da das verwendete Geschwindigkeitsfeld die Randbedingungen nur im Mittel erfüllt und es sonst zu unerwünschten Flüssen durch die Wände kommen kann. Außerdem nutzen wir im Weiteren stets den 2D-Quellterm (2.29) und die für die Berechnung von Dichte und Viskosität nach (2.19) erforderlichen Parameter lauten $\varrho_F = 1, \varrho_D = 10^{-3}, \eta_F = 1$ und $\eta_D = 5 \times 10^{-4}$. Des Weiteren verwenden wir (4.36) zur Bestimmung der vorläufigen Geschwindigkeit innerhalb des FEM-Lösers für das Strömungsfeld, da dies verglichen mit (4.37) weniger Rechenaufwand bedeutet und bei numerischen Experimenten keine Nachteile dieses Vorgehens beobachtet wurden.

Der Rayleigh-Gleichung (2.25) können zwei Sachverhalte entnommen werden. Erstens kommt es zu Ausgasung, wenn der gewählte Dampfdruck über dem Druck des Fluids liegt und zweitens nimmt das Ausmaß der Kavitation mit wachsender Druckdifferenz zu. Dies wollen wir nun durch drei Simulationen mit verschiedenen Dampfdrücken überprüfen. Betrachten wir die Anfangswerte des Drucks, so liegt das Minimum mit $p_{min} \approx -55$ an

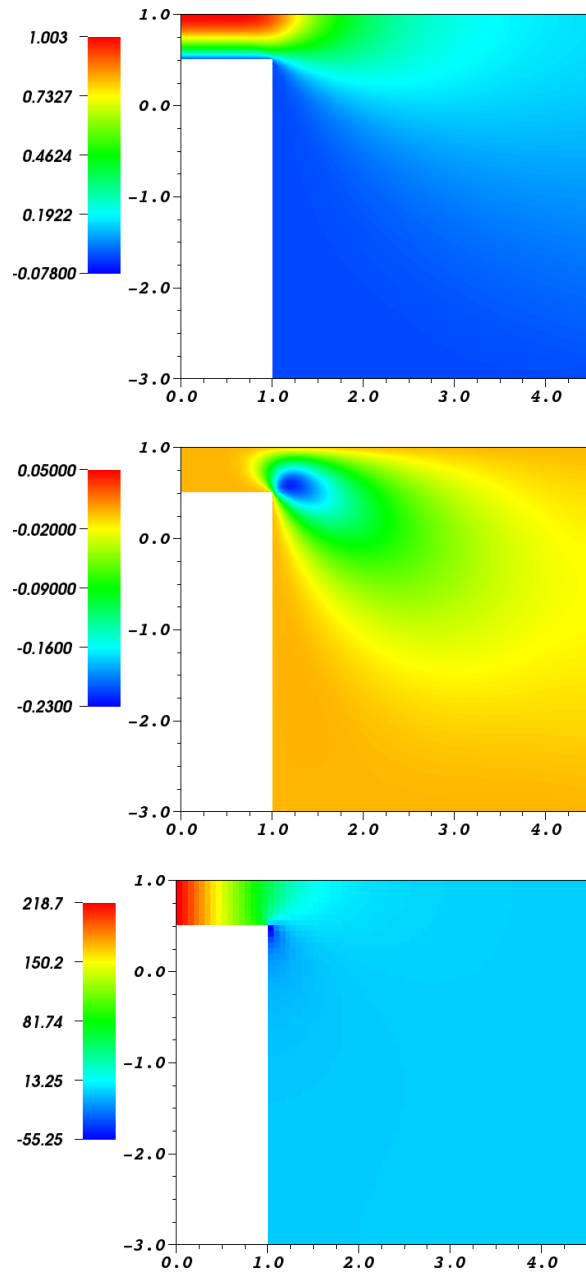


Abbildung 8.9: Anfangszustand ($t = 0$) für die Kavitationssimulation. Geplottet sind x -Geschwindigkeit (oben), y -Geschwindigkeit (mitte) und Druck (unten).

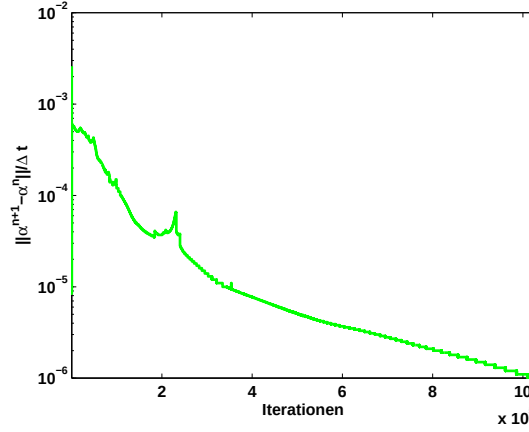


Abbildung 8.10: Zeitliche Änderung der Dampfdichte

der Ecke der Geometrie, während am Ausfluss $p_{out} \approx 0$ gilt. Wir wollen den Dampfdruck p_D so wählen, dass es ausschließlich lokal am Step zur Kavitation kommt. Durch $p_{min} < p_D < p_{out}$ sollte dies gewährleistet werden und wir wählen hier $p_D = -15, -10, -5$. Auch wenn sich p_{min} im Laufe der Zeit ändert, sollte dies auf die prinzipielle Situation keinen wesentlichen Einfluss haben. Wir erwarten dann in allen drei Fällen, dass sich an der Stufe eine Kavitationswolke bildet, welche in Richtung Ausfluss transportiert wird. Dabei sollte das Kavitationsgebiet im Fall $p_D = -5$ größer ausfallen, als bei $p_D = -15$, da in dieser Situation die Differenz $p_D - p_{min}$ betragsmäßig größer ist. In den Abbildungen 8.11, 8.12 und 8.13 sind die Resultate mit den unterschiedlich gewählten Dampfdrücken dargestellt.

Die bei den Rechnungen verwendete Zeitschrittweite beträgt $\Delta t = 10^{-4}$. Nach (6.34) wäre zur Einhaltung des Maximumprinzips etwa $\Delta t < 10^{-2}$ erforderlich. Der Grund für die Wahl eines kleineren Zeitschritts liegt am Newton-Verfahren für das nichtlineare Pressure-Poisson Problem (4.38). Es hat sich gezeigt, dass die ausreichende Reduktion des zugehörigen Residuums $\|\mathbf{F}(\mathbf{p}_k)\|$ von höchster Bedeutung ist. Bei der Verwendung zu großer Zeitschrittweiten konnte dies in einzelnen Zeitschritt nicht gewährleistet werden. Auch die Wiederholung des Zeitschritts mit kleiner Schrittweite führte zu keiner Verbesserung. Die Zusammenhänge scheinen hier sehr komplex und sind Gegenstand weiterer Forschung. Eine Ursache für dieses Problematik könnte die, verglichen mit den anderen Verfahrensteilen, unterschiedliche Gewichtung des Quellterms im Patankar-DG-Verfahren sein. Es bleibt noch zu erwähnen, dass die verwendete Zeitschrittweite keineswegs optimal gewählt wurde. Weitere numerische Tests haben gezeigt, dass das Verfahren auch für größere Zeitschrittweiten sinnvolle Ergebnisse liefert.

Die Ausgabezeit $t = 100$ wurde gewählt, da die zeitliche Änderung der Dampfdichte im Fall $p_D = -5$ um über drei Größenordnungen gefallen ist und als stationär angesehen werden kann. Die Approximation der Zeitableitung $\partial_t \alpha$ für diesen Fall ist in Abbildung (8.10) dargestellt.

Nun vergleichen wir die in den Abbildungen 8.11, 8.12 und 8.13 dargestellten Resul-

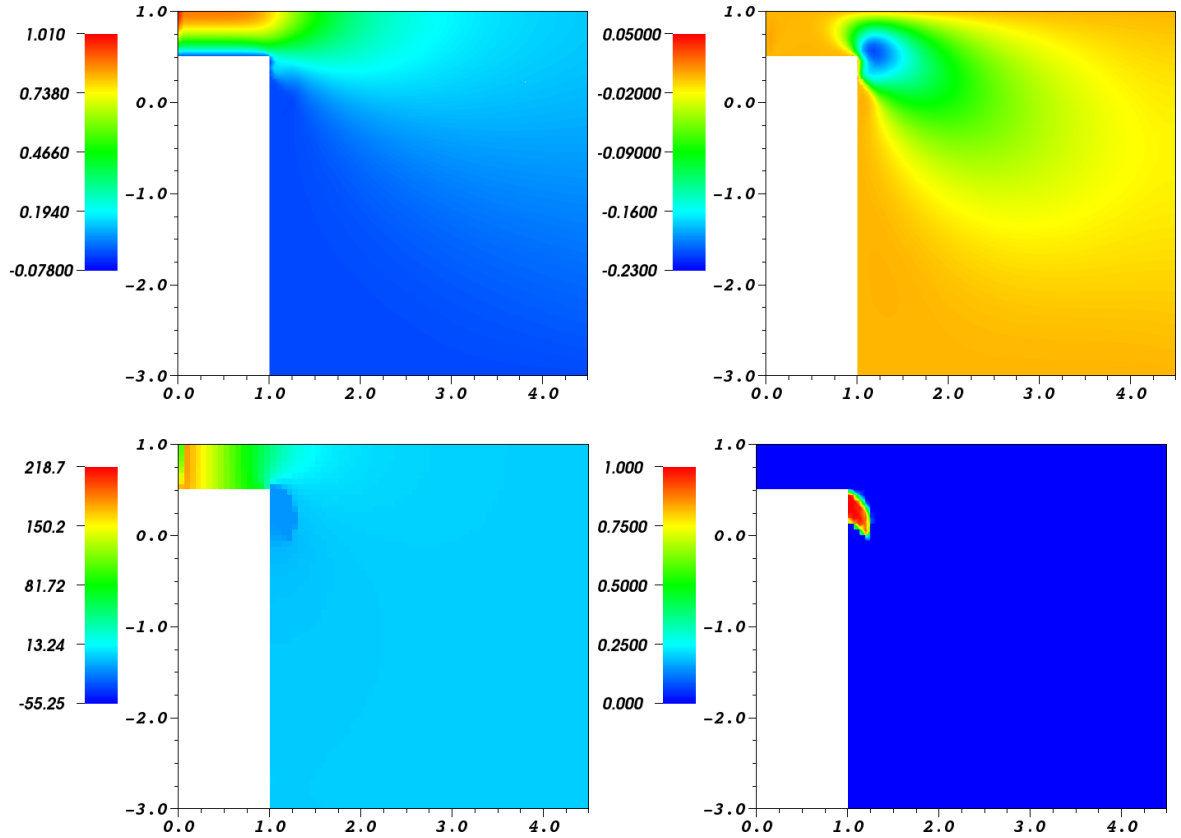


Abbildung 8.11: Simulation mit dem Kavitationsmodell für $p_D = -15$ bei $t = 100$. Geplottet sind x -Geschwindigkeit (oben links), y -Geschwindigkeit (oben rechts), Druck (unten links) und Dampfdichte (unten rechts).

tate der Simulationen. Wir erkennen, dass es jeweils zu lokaler Ausgasung an der Stufe kommt, wobei das Kavitationsgebiet wie erwartet um so größer ist, je höher der Dampfdruck gewählt wurde. In allen drei Fällen wächst der Dampfanteil bis auf 100%, was sich dadurch erklären lässt, dass aufgrund der Haftbedingung an der Wand die Geschwindigkeiten in Wandnähe extrem niedrig sind und die dort entstehende Ausgasung durch die Strömung nur sehr schlecht abtransportiert werden kann. Der Druck innerhalb des Kavitationsbereiches entspricht weitgehend dem gewählten Dampfdruck und der lokale Druckabfall am Einfluss wird durch die Randbedingung für die Dampfdichte verursacht. Weiter fällt auf, dass sich im Kavitationsgebiet starke Gradienten in den Geschwindigkeitskomponenten gebildet haben. Diese teilen das Kavitationsgebiet in zwei Bereiche, wobei die Geschwindigkeiten im wandnahen Bereich verschwindend gering sind. Es hat sich gezeigt, dass die Grenze zwischen diesen beiden Bereichen auch das Kavitationsgebiet halbiert.

In Abbildung 8.14 wird zusätzlich noch die zeitliche Entwicklung des Kavitationsgebietes mit $p_D = -5$ gezeigt. Daraus wird ersichtlich, dass sich zunächst die Form des Gebietes ausbildet und dann der Dampfanteil innerhalb des Gebietes anwächst.

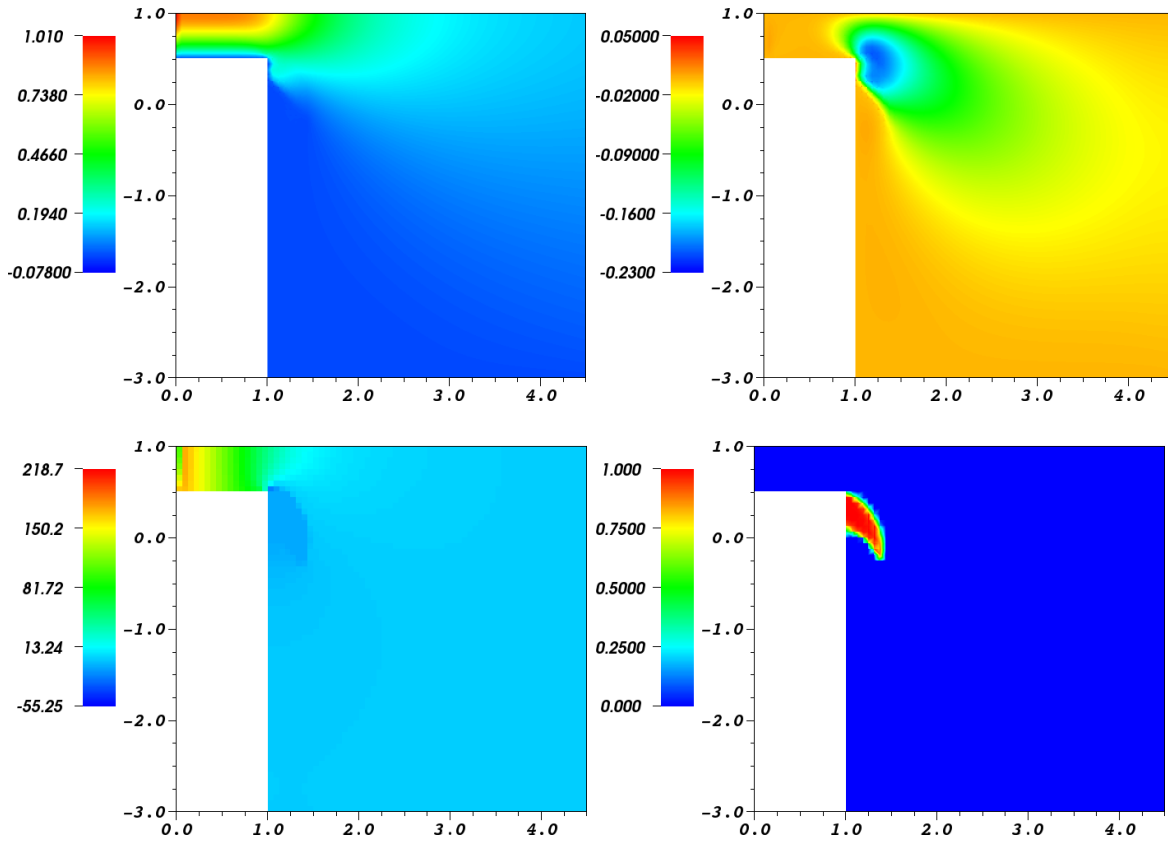


Abbildung 8.12: Simulation mit dem Kavitationsmodell für $p_D = -10$ bei $t = 100$. Geplottet sind x -Geschwindigkeit (oben links), y -Geschwindigkeit (oben rechts), Druck (unten links) und Dampfdichte (unten rechts).

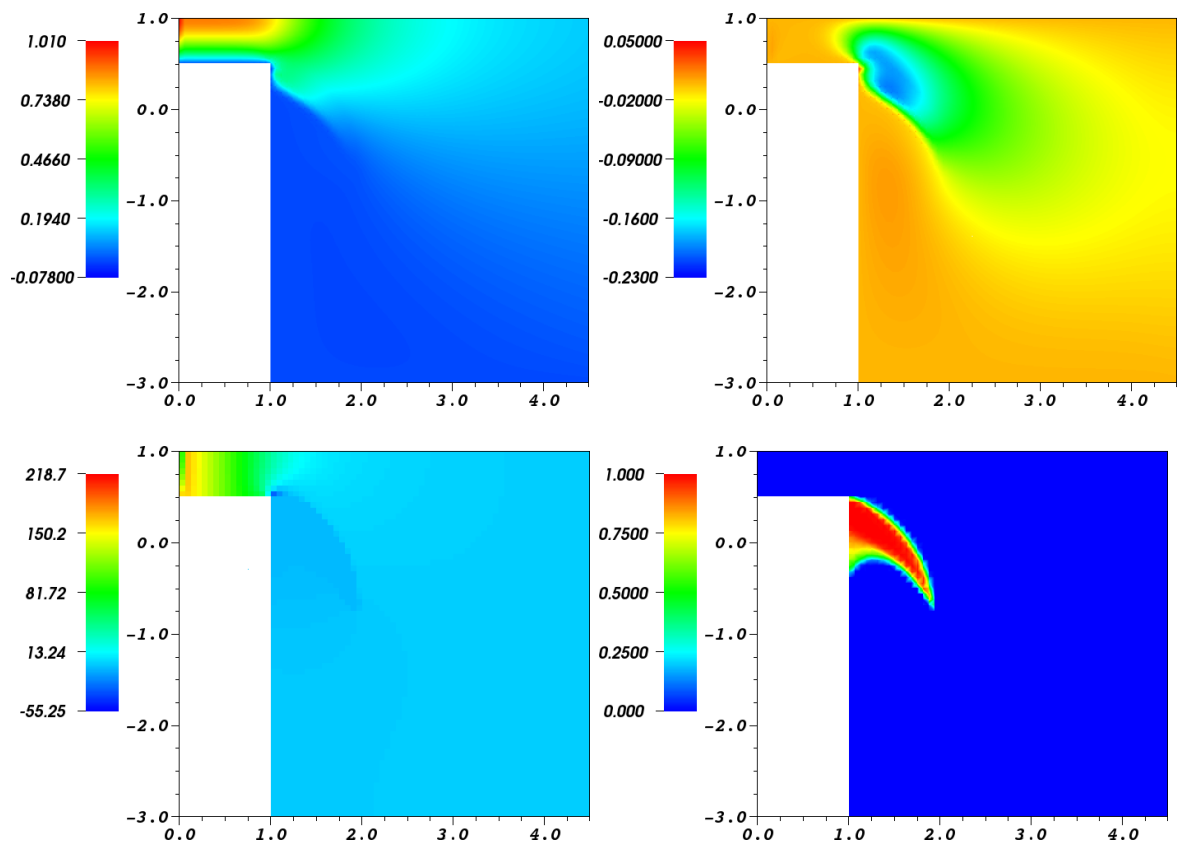


Abbildung 8.13: Simulation mit dem Kavitationsmodell für $p_D = -5$ bei $t = 100$. Geplottet sind x -Geschwindigkeit (oben links), y -Geschwindigkeit (oben rechts), Druck (unten links) und Dampfdichte (unten rechts).

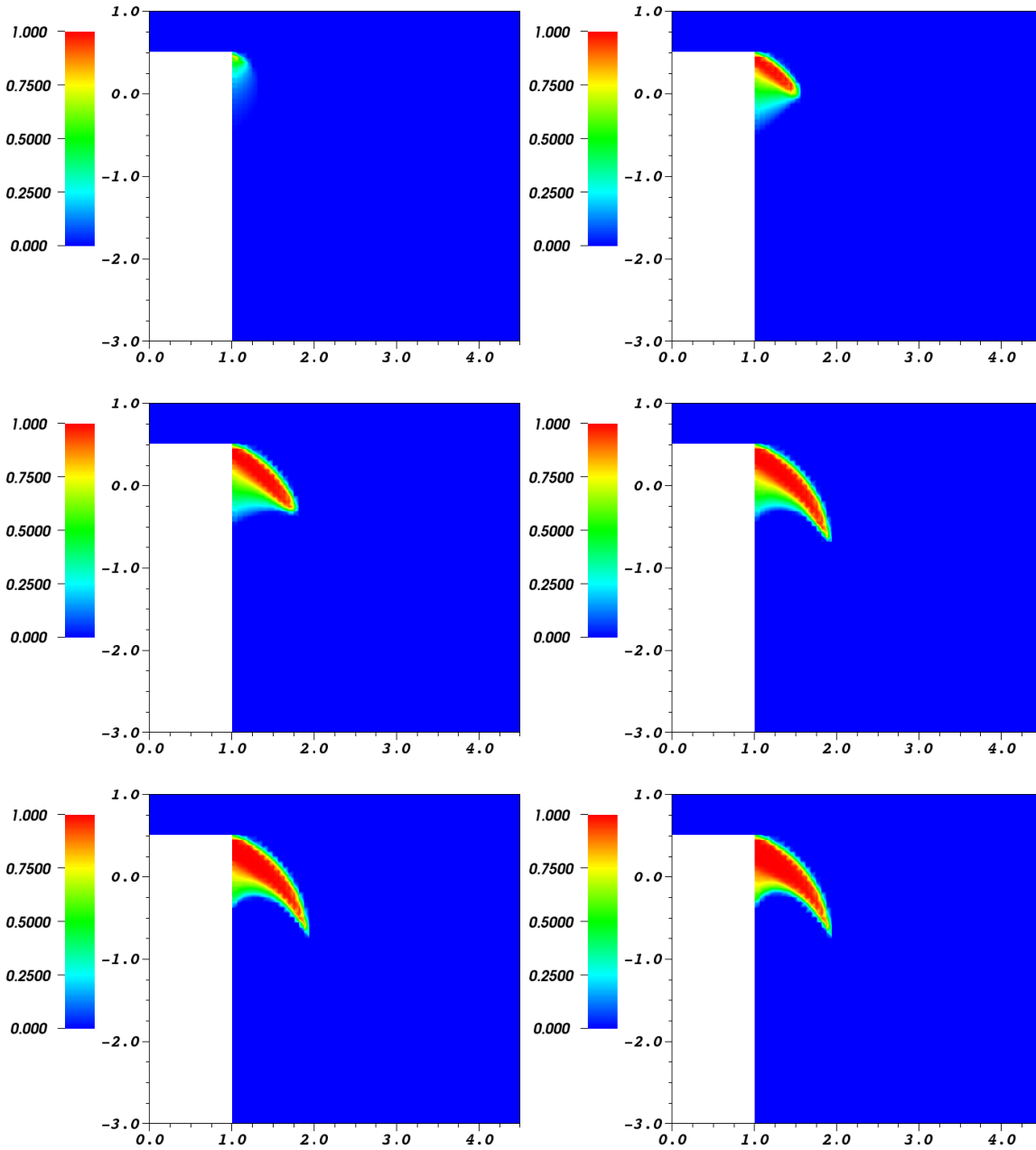


Abbildung 8.14: Zeitliche Entwicklung des Kavitationsgebietes bei Dampfdruck $p_D = -5$ für $t = 1, 5, 10, 25, 50, 100$. Die Abfolge geht von links nach rechts und dann von oben nach unten.

9 Zusammenfassung und Ausblick

In dieser Arbeit wurde ein Verfahren für gekoppelte Strömungs-Transport-Probleme entwickelt. Dazu wurde im Bereich der Strömung ein für inkompressible Gleichungen entwickeltes Projektionsverfahren neu interpretiert und erfolgreich auf die hier betrachtete Problemstellung übertragen. Als Resultat ergab sich ein neuartiges Projektionsverfahren innerhalb dessen ein nichtlineares Pressure-Poisson Problem zu lösen ist. Des weiteren wurde eine neue Formulierung des Reibungsspannungstensors vorgeschlagen, um unphysikalisches Verhalten der Strömung an Ausflussrändern zu vermeiden. Diese Formulierung ist mit geringem Programmieraufwand umsetzbar und vermeidet die Einführung zusätzlicher numerischer Penaltyterme.

Zur Lösung der zugehörigen Transportgleichung wurde ein Discontinuous-Galerkin-Verfahren hergeleitet. Dabei ist speziell für die Simulation von Kavitation die Einhaltung eines Maximumprinzips von wesentlicher Bedeutung. Um diese Eigenschaft im numerischen Verfahren gewährleisten zu können, wurde in dieser Arbeit das Patankar-DG-Verfahren entwickelt. Dieses Verfahren sichert die Einhaltung eines Maximumprinzips unter Erfüllung einer CFL-Bedingung bei beliebigem nichtlinearem Quellterm. Die Idee stammt aus dem zur Positivitätserhaltung im Kontext gewöhnlicher Differentialgleichungen verwendeten Patankar-Verfahren, wobei es durch den Einsatz von Transformationen zur Einhaltung beliebiger Schranken genutzt wird.

Im Allgemeinen können Strömung und Transport aufgrund der Inkompatibilität der Diskretisierungen nicht sinnvoll gekoppelt werden, was ein zusätzliches Postprocessing des Geschwindigkeitsfeldes verlangt. Dies ist auch bei den oben verwendeten Diskretisierungen der Fall und mit der Local-Conservative-Projection-Formulation wurde eine an die gegebene Geschwindigkeitsdiskretisierung angepasste und effiziente Methode zur Bestimmung eines kompatiblen Geschwindigkeitsfeldes vorgestellt. Der Vorteil dieses Verfahrens liegt darin, das Postprocessing elementweise durchzuführen und somit das Lösen großer linearer Gleichungssysteme zu vermeiden.

Schließlich wurde die Verwendbarkeit des Verfahrens anhand mehrerer Testfälle erfolgreich demonstriert.

Für zukünftige Arbeiten gibt es drei wesentliche Ziele. Eines ist die Weiterentwicklung des Patankar-DG-Verfahrens in Bezug auf Zeitdiskretisierungen höherer Ordnung, welche in dieser Arbeit nicht betrachtet wurden. Weiter hat sich gezeigt, dass dem hinreichend genauen Lösen des nichtlinearen Pressure-Poisson-Problems eine zentrale Rolle zukommt. Dies noch effizienter zu gestalten ist zukünftig ebenso von großem Interesse. Außerdem wurde in [KKMW10] gezeigt, dass die Berechnung von Strömung und Transport auf unterschiedlichen angepassten Gittern viel Potential für Effizienzsteigerungen bietet. Der Einsatz des hier vorgestellten Verfahrens sollte in diesem Zusammenhang einen noch deutlicheren Effizienzgewinn hervorbringen.

A Anhang

A.1 Notation

Wir betrachten hier eine skalare Funktion q , vektorwertige Funktionen $\mathbf{v} = (v_1, v_2)^T$ und $\mathbf{w} = (w_1, w_2)^T$, sowie einen Tensor zweiter Stufe $\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$. Alle Größen hängen dabei von den Raumkoordinaten $\mathbf{x} = (x, y)^T$ und der Zeit t ab.

Kennzeichnen wir die partiellen Ableitungen in die jeweilige Raumrichtung mit ∂_x und ∂_y , dann ist der Gradient einer skalaren Größe

$$\nabla q = \begin{bmatrix} \partial_x q \\ \partial_y q \end{bmatrix}$$

als Vektor der Richtungsableitungen gegeben. Bei Tensoren ist dies komponentenweise aufzufassen und es gilt

$$\nabla \mathbf{v} = \begin{bmatrix} \partial_x v_1 & \partial_y v_1 \\ \partial_x v_2 & \partial_y v_2 \end{bmatrix}.$$

Mit $\nabla \mathbf{v}^T$ bezeichnen wir die zugehörige Transponierte und mit $\nabla \mathbf{v}^{-T}$ deren Inverse. Der Laplace-Operator lautet im skalaren Fall

$$\Delta q = \partial_{xx} q + \partial_{yy} q,$$

bei Vektoren wird der Laplace-Operator auf die einzelnen Komponenten angewandt, also

$$\Delta \mathbf{v} = \begin{bmatrix} \partial_{xx} v_1 + \partial_{yy} v_1 \\ \partial_{xx} v_2 + \partial_{yy} v_2 \end{bmatrix}.$$

Die Divergenz eines Vektors ist durch

$$\nabla \cdot \mathbf{v} = \partial_x v_1 + \partial_y v_2$$

gegeben, bei Tensoren lautet sie

$$\nabla \cdot \mathbf{A} = \begin{bmatrix} \partial_x A_{11} + \partial_y A_{12} \\ \partial_x A_{21} + \partial_y A_{22} \end{bmatrix}.$$

Weiter gilt

$$\mathbf{u} \cdot \nabla \mathbf{v} = (\mathbf{u} \cdot \nabla) \mathbf{v} = \begin{bmatrix} u_1 \partial_x v_1 + u_2 \partial_y v_1 \\ u_1 \partial_x v_2 + u_2 \partial_y v_2 \end{bmatrix}$$

und

$$\nabla \mathbf{u} : \nabla \mathbf{v} = \partial_x u_1 \partial_x v_1 + \partial_y u_1 \partial_y v_1 + \partial_x u_2 \partial_x v_2 + \partial_y u_2 \partial_y v_2.$$

A.2 Bestimmung einer Basis für $RT_1(\hat{K})$

Um eine Basis für $RT_1(\hat{K})$ angeben zu können, benötigen wir zunächst eine Basis von $P_1([-1, 1])$. Wir nutzen im Weiteren die Lagrange-Basis $\{\hat{q}_1, \hat{q}_2\}$ mit

$$\hat{q}_1(s) = -\frac{1}{2}(s-1), \quad \hat{q}_2(s) = \frac{1}{2}(s+1).$$

Setzen wir $\hat{\mathbf{w}} = (\hat{w}_1, \hat{w}_2)^T \in RT_1(\hat{K})$, so erhalten wir aus (7.14a) die 8 Freiheitsgrade

$$\begin{aligned} M_1(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= -\int_{-1}^1 \hat{w}_2(\hat{x}, -1) \hat{q}_1(\hat{x}) d\hat{x}, \\ M_2(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= -\int_{-1}^1 \hat{w}_2(\hat{x}, -1) \hat{q}_2(\hat{x}) d\hat{x}, \\ M_3(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= \int_{-1}^1 \hat{w}_1(1, \hat{y}) \hat{q}_1(\hat{y}) d\hat{y}, \\ M_4(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= \int_{-1}^1 \hat{w}_1(1, \hat{y}) \hat{q}_2(\hat{y}) d\hat{y}, \\ M_5(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= \int_{-1}^1 \hat{w}_2(\hat{x}, 1) \hat{q}_2(\hat{x}) d\hat{x}, \\ M_6(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= \int_{-1}^1 \hat{w}_2(\hat{x}, 1) \hat{q}_1(\hat{x}) d\hat{x}, \\ M_7(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= -\int_{-1}^1 \hat{w}_1(-1, \hat{y}) \hat{q}_2(\hat{y}) d\hat{y}, \\ M_8(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= -\int_{-1}^1 \hat{w}_1(-1, \hat{y}) \hat{q}_1(\hat{y}) d\hat{y}. \end{aligned}$$

Um die restlichen 4 Freiheitsgrade zu bestimmen, benötigen wir eine Basis des Polynomraumes $P_{0,1}(\hat{K}) \times P_{1,0}(\hat{K}) = \text{span}\{y, 1\} \times \text{span}\{x, 1\}$. Nutzen wir die Monom-Basis $\{(1, 0)^T, (0, 1)^T, (\hat{y}, 0)^T, (0, \hat{x})^T\}$, so erhalten wir die übrigen Freiheitsgrade als

$$\begin{aligned} M_9(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= \int_{-1}^1 \int_{-1}^1 \hat{w}_1(\hat{x}, \hat{y}) d\hat{x} d\hat{y}, \\ M_{10}(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= \int_{-1}^1 \int_{-1}^1 \hat{w}_2(\hat{x}, \hat{y}) d\hat{x} d\hat{y}, \\ M_{11}(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= \int_{-1}^1 \int_{-1}^1 \hat{w}_1(\hat{x}, \hat{y}) \hat{y} d\hat{x} d\hat{y}, \\ M_{12}(\hat{\mathbf{w}}(\hat{x}, \hat{y})) &= \int_{-1}^1 \int_{-1}^1 \hat{w}_2(\hat{x}, \hat{y}) \hat{x} d\hat{x} d\hat{y}. \end{aligned}$$

Mit diesen Freiheitsgraden können wir eine Basis $\{\hat{\mathbf{w}}^1, \dots, \hat{\mathbf{w}}^{12}\}$ von $RT_1(\hat{K})$ bestimmen, indem wir für die einzelnen Basisfunktionen die Bedingung $M_j(\hat{\mathbf{w}}^i) = \delta_{ij}$ für $i, j =$

$1, \dots, 12$ fordern. Die so ermittelten Basisfunktionen sind dann durch

$$\begin{aligned}
\hat{\mathbf{w}}^1(\hat{x}, \hat{y}) &= \left(0, \frac{1}{8}(-1 + 3\hat{x})(-1 - 2\hat{y} + 3\hat{y}^2)\right)^T, \\
\hat{\mathbf{w}}^2(\hat{x}, \hat{y}) &= \left(0, \frac{1}{8}(1 - 3\hat{x})(-1 - 2\hat{y} + 3\hat{y}^2)\right)^T, \\
\hat{\mathbf{w}}^3(\hat{x}, \hat{y}) &= \left(\frac{1}{8}(-1 + 2\hat{x} + 3\hat{x}^2)(1 - 3\hat{y}), 0\right)^T, \\
\hat{\mathbf{w}}^4(\hat{x}, \hat{y}) &= \left(\frac{1}{8}(-1 + 2\hat{x} + 3\hat{x}^2)(1 + 3\hat{y}), 0\right)^T, \\
\hat{\mathbf{w}}^5(\hat{x}, \hat{y}) &= \left(0, \frac{1}{8}(1 + 3\hat{x})(-1 + 2\hat{y} + 3\hat{y}^2)\right)^T, \\
\hat{\mathbf{w}}^6(\hat{x}, \hat{y}) &= \left(0, \frac{1}{8}(1 - 3\hat{x})(-1 + 2\hat{y} + 3\hat{y}^2)\right)^T, \\
\hat{\mathbf{w}}^7(\hat{x}, \hat{y}) &= \left(\frac{1}{8}(-1 - 2\hat{x} + 3\hat{x}^2)(-1 - 3\hat{y}), 0\right)^T, \\
\hat{\mathbf{w}}^8(\hat{x}, \hat{y}) &= \left(\frac{1}{8}(-1 - 2\hat{x} + 3\hat{x}^2)(-1 + 3\hat{y}), 0\right)^T, \\
\hat{\mathbf{w}}^9(\hat{x}, \hat{y}) &= \left(\frac{3}{8}(1 - \hat{x}^2), 0\right)^T, \\
\hat{\mathbf{w}}^{10}(\hat{x}, \hat{y}) &= \left(0, \frac{3}{8}(1 - \hat{y}^2)\right)^T, \\
\hat{\mathbf{w}}^{11}(\hat{x}, \hat{y}) &= \left(\frac{9}{8}(1 - \hat{x}^2)\hat{y}, 0\right)^T, \\
\hat{\mathbf{w}}^{12}(\hat{x}, \hat{y}) &= \left(0, \frac{9}{8}\hat{x}(1 - \hat{y}^2)\right)^T,
\end{aligned} \tag{A.1}$$

gegeben. Für $\hat{\mathbf{w}}(\hat{\mathbf{x}}) = \sum_{j=1}^{12} v_j \hat{\mathbf{w}}^j(\hat{\mathbf{x}})$ sind die ersten acht Koeffizienten der Basisdarstellung durch

$$\begin{aligned}
v_1 &= \hat{\mathbf{w}}(-1/3, -1) \cdot (0, -1)^T, \\
v_2 &= \hat{\mathbf{w}}(1/3, -1) \cdot (0, -1)^T, \\
v_3 &= \hat{\mathbf{w}}(1, -1/3) \cdot (1, 0)^T, \\
v_4 &= \hat{\mathbf{w}}(1, 1/3) \cdot (1, 0)^T, \\
v_5 &= \hat{\mathbf{w}}(1/3, 1) \cdot (0, 1)^T, \\
v_6 &= \hat{\mathbf{w}}(-1/3, 1) \cdot (0, 1)^T, \\
v_7 &= \hat{\mathbf{w}}(-1, 1/3) \cdot (-1, 0)^T, \\
v_8 &= \hat{\mathbf{w}}(-1, -1/3) \cdot (-1, 0)^T
\end{aligned} \tag{A.2}$$

A Anhang

festgelegt. Sie ergeben sich also nach Auswertung der einzelnen Normalkomponenten an den Stellen $+1/3$ und $-1/3$. Dies können wir nutzen um Randbedingungen vorzugeben.

Literaturverzeichnis

- [ABF05] ARNOLD, Douglas N. ; BOFFI, Daniele ; FALK, Richard S.: Quadrilateral $H(\text{div})$ finite elements. In: *SIAM J. Numer. Anal.* 42 (2005), Nr. 6, S. 2429–2451
- [ADCC01] AIZINGER, Vadym ; DAWSON, Clint ; COCKBURN, Bernardo ; CASTILLO, Paul: The local discontinuous Galerkin method for contaminant transport. In: *Adv. Water Res.* 24 (2001), S. 73–87
- [BBD⁺08] BOFFI, Daniele ; BREZZI, Franco ; DEMKOWICZ, Leszek F. ; DURÁN, Ricardo G. ; FALK, Richard S. ; FORTIN, Michel: *Lecture Notes in Mathematics*. Bd. 1939: *Mixed finite elements, compatibility conditions, and applications*. Springer-Verlag, 2008. – Lectures given at the C.I.M.E. Summer School held in Cetraro, June 26–July 1, 2006, Edited by Boffi and Lucia Gastaldi
- [BC07] BARTH, William L. ; CAREY, Graham F.: On a boundary condition for pressure-driven laminar flow of incompressible fluids. In: *Internat. J. Numer. Methods Fluids* 54 (2007), Nr. 11, S. 1313–1325
- [BDM03] BURCHARD, Hans ; DELEERSNIJDER, Eric ; MEISTER, Andreas: A high-order conservative Patankar-type discretisation for stiff systems of production-destruction equations. In: *Appl. Numer. Math.* 47 (2003), Nr. 1, S. 1–30
- [BF91] BREZZI, Franco ; FORTIN, Michel: *Springer Series in Computational Mathematics*. Bd. 15: *Mixed and hybrid finite element methods*. Springer-Verlag, 1991
- [Böh00] BÖHME, Gert: *Strömungsmechanik nichtnewtonscher Fluide*. Teubner, 2000
- [Bra07] BRAESS, Dietrich: *Finite Elemente Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie*. Springer, 2007
- [Bre95] BRENNEN, Christopher E.: *Cavitation and bubble dynamics*. Oxford Univ. Press, 1995
- [CC89] CHAVENT, Guy ; COCKBURN, Bernardo: The local projection P^0P^1 -discontinuous-Galerkin finite element method for scalar conservation laws. In: *RAIRO Modél. Math. Anal. Numér.* 23 (1989), Nr. 4, S. 565–592

- [CDMW98] CHIPPAD, S. ; DAWSON, C. N. ; MARTÍNEZ, M. L. ; WHEELER, M. F.: A projection method for constructing a mass conservative velocity field. In: *Comput. Methods Appl. Mech. Engrg.* 157 (1998), S. 1–10
- [CGW07] COCKBURN, Bernardo ; GOPALAKRISHNAN, Jayadeep ; WANG, Haiying: Locally conservative fluxes for the continuous Galerkin method. In: *SIAM J. Numer. Anal.* 45 (2007), Nr. 4, S. 1742–1776
- [Cho68] CHORIN, Alexandre J.: Numerical solution of the Navier-Stokes equations. In: *Math. Comp.* 22 (1968), S. 745–762
- [Cia88] CIARLET, Philippe G.: *Studies in Mathematics and its Applications*. Bd. 20: *Mathematical elasticity. Vol. I.* Amsterdam : North-Holland Publishing Co., 1988. – Three-dimensional elasticity
- [Cia02] CIARLET, Philippe G.: *The finite element method for elliptic problems*. Society for Industrial and Applied Mathematics (SIAM), 2002
- [CKS00] COCKBURN, Bernardo ; KARNIADAKIS, George E. ; SHU, Chi-Wang: The development of discontinuous Galerkin methods. In: *Discontinuous Galerkin Methods. Theory, Computation and Applications* Bd. 11. Springer, 2000, S. 3–50
- [Con08] CONZEN, Carsten: *Numerische und experimentelle Untersuchungen zu Transportvorgängen in Schneckenmaschinen*. Kassel Univ. Press, 2008
- [CS01] COCKBURN, Bernardo ; SHU, Chi-Wang: Runge-Kutta discontinuous Galerkin methods for convection-dominated problems. In: *J. Sci. Comput.* 16 (2001), Nr. 3, S. 173–261
- [CSS86] CUVELIER, C. ; SEGAL, A. ; STEENHOVEN, A. A.: *Finite element methods and Navier-Stokes equations*. Dordrecht : D. Reidel Publishing Co., 1986 (Mathematics and its Applications)
- [DGM09] DUMBSER, Michael ; GASSNER, Gregor ; MUNZ, Claus-Dieter: Discontinuous-Galerkin-Verfahren für zeitabhängige Advektions-Diffusions-Gleichungen. In: *JB DMV* 111 (2009), Nr. 3, S. 97–123
- [DSW04] DAWSON, Clint ; SUN, Shuyu ; WHEELER, Mary F.: Compatible algorithms for coupled flow and transport. In: *Comput. Methods Appl. Mech. Engrg.* 193 (2004), Nr. 23-26, S. 2565–2580
- [ESW05] ELMAN, Howard C. ; SILVESTER, David J. ; WATHEN, Andrew J.: *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Oxford University Press, 2005 (Numerical Mathematics and Scientific Computation)

- [Fro04] FROBENIUS, Moritz: *Numerische Simulation kavitierender Strömungen in hydraulischen Strömungsmaschinen*, Technische Universität München, Diss., 2004
- [GKS09] GOTTLIEB, Sigal ; KETCHESON, David I. ; SHU, Chi-Wang: High order strong stability preserving time discretizations. In: *J. Sci. Comput.* 38 (2009), Nr. 3, S. 251–289
- [GMS06] GUERMOND, J. L. ; MINEV, P. ; SHEN, Jie: An overview of projection methods for incompressible flows. In: *Comput. Methods Appl. Mech. Engrg.* 195 (2006), Nr. 44-47, S. 6011–6045
- [GR86] GIRAULT, Vivette ; RAVIART, Pierre-Arnaud: *Finite element methods for Navier-Stokes equations*. Springer-Verlag, 1986
- [GS87] GRESHO, Philip M. ; SANI, Robert L.: On pressure boundary conditions for the incompressible Navier-Stokes equations. In: *International Journal for Numerical Methods in Fluids* 7 (1987), Nr. 10
- [GS00] GRESHO, P. M. ; SANI, R. L.: *Incompressible Flow and the Finite Element Method: Isothermal Laminar Flow*. Wiley, 2000
- [HEH93] HAROUTUNIAN, Vah ; ENGELMAN, Michael S. ; HASBANI, Isaac: Segregated finite element algorithms for the numerical solution of large-scale incompressible flow problems. In: *International Journal for Numerical Methods in Fluids* 17 (1993), Nr. 4, S. 323–348
- [HRT96] HEYWOOD, John G. ; RANNACHER, Rolf ; TUREK, Stefan: Artificial boundaries and flux and pressure conditions for the incompressible Navier-Stokes equations. In: *Internat. J. Numer. Methods Fluids* 22 (1996), Nr. 5, S. 325–352
- [HW08] HESTHAVEN, Jan S. ; WARBURTON, Tim: *Texts in Applied Mathematics*. Bd. 54: *Nodal discontinuous Galerkin methods*. Springer, 2008. – Algorithms, analysis, and applications
- [Kan86] KAN, J. van: A second-order accurate pressure-correction scheme for viscous incompressible flow. In: *SIAM J. Sci. Statist. Comput.* 7 (1986), Nr. 3, S. 870–891
- [Kel95] KELLEY, C. T.: *Iterative methods for linear and nonlinear equations*. SIAM, 1995
- [KKMW10] KOPECZ, Stefan ; KREBS, Mathias ; MEISTER, Andreas ; WÜNSCH, Olaf: A fast numerical approach for the simulation of highly viscous non-isothermal non-Newtonian fluids. In: *Z. Angew. Math. Phys.* 61 (2010), Nr. 4, S. 673–684

- [KMW11] KOPECZ, Stefan ; MEISTER, Andreas ; WÜNSCH, Olaf: A coupled FEM/DG approach for the simulation of cavitating micro foams. In: *PAMM* 11 (2011), S. 779–780
- [LL09] LEER, B. van ; LO, M.: Unification of Discontinuous Galerkin Methods for Advection and Diffusion. In: *47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition, Orlando, Florida*, 2009
- [LLR07] LEER, B. van ; LO, M. ; RAALTE, M. van: A Discontinuous Galerkin Method for Diffusion Based on Recovery. In: *18th AIAA Computational Fluid Dynamics Conference, Miami, Florida, June 25-28*, 2007
- [LMW02] LANGTANGEN, Hans P. ; MARDAL, Kent-Andre ; WINTHER, Ragnar: Numerical methods for incompressible viscous flow. In: *Advances in Water Resources* 25 (2002), Nr. 8-12, S. 1125 – 1146
- [LN04] LARSON, M. G. ; NIKLASSON, A. J.: A conservative flux for the continuous Galerkin method based on discontinuous enrichment. In: *Calcolo* 41 (2004), Nr. 2, S. 65–76
- [LO96] LIU, Xu-Dong ; OSHER, Stanley: Nonoscillatory high order accurate self-similar maximum principle satisfying shock capturing schemes. I. In: *SIAM J. Numer. Anal.* 33 (1996), Nr. 2, S. 760–779
- [Lo11] LO, Marcus: *A Space-Time Discontinuous Galerkin Method for Navier-Stokes with Recovery*, University of Michigan, Diss., 2011
- [LS00] LIU, Jian-Guo ; SHU, Chi-Wang: A high-order discontinuous Galerkin method for 2D incompressible flows. In: *J. Comput. Phys.* 160 (2000), Nr. 2, S. 577–596
- [Mei99] MEISTER, Andreas: *Numerik linearer Gleichungssysteme*. Vieweg, 1999
- [Mon03] MONK, Peter: *Finite element methods for Maxwell's equations*. Oxford University Press, 2003 (Numerical Mathematics and Scientific Computation)
- [Oua06] OUAZZI, Abderrahim: *Finite Element Simulation of Nonlinear Fluids*. Shaker, 2006
- [Pro97] PROHL, Andreas: *Projection and quasi-compressibility methods for solving the incompressible Navier-Stokes equations*. Teubner, 1997 (Advances in Numerical Mathematics)
- [Qua09] QUARTERONI, Alfio: *Numerical models for differential problems*. Springer, 2009 (MS&A. Modeling, Simulation and Applications)

- [Ran08] RANNACHER, Rolf: Methods for numerical flow simulation. In: *Hemodynamical flows*. Birkhäuser, 2008, S. 275–332
- [RG01] REDDY, J. N. ; GARTLING, D. K.: *The finite element method in heat transfer and fluid dynamics*. CRC Press, 2001
- [Riv08] RIVIÈRE, Béatrice: *Frontiers in Applied Mathematics*. Bd. 35: *Discontinuous Galerkin methods for solving elliptic and parabolic equations*. Society for Industrial and Applied Mathematics (SIAM), 2008. – Theory and implementation
- [RT77] RAVIART, P.-A. ; THOMAS, J. M.: A mixed finite element method for 2nd order elliptic problems. In: *Mathematical aspects of finite element methods. Lecture Notes in Math., Vol. 606*. Berlin : Springer, 1977, S. 292–315
- [RT91] ROBERTS, J. E. ; THOMAS, J.-M.: Mixed and hybrid methods. In: *Handbook of numerical analysis, Vol. II*. Amsterdam : North-Holland, 1991 (Handb. Numer. Anal., II), S. 523–639
- [RT92] RANNACHER, R. ; TUREK, S.: Simple nonconforming quadrilateral Stokes element. In: *Numer. Methods Partial Differential Equations* 8 (1992), Nr. 2, S. 97–111
- [Sau00] SAUER, Jürgen: *Instationär kavitierende Strömungen. ein neues Modell, basierend auf Front Capturing (VoF) und Blasendynamik*, Universität Karlsruhe, Diss., 2000
- [Sch97] SCHIEWECK, Friedhelm: *Parallele Lösung der inkompressiblen Navier-Stokes Gleichungen*, Universität Magdeburg, Habilitation, 1997
- [Sch03] SCHMACHTEL, R.: *Robuste lineare und nichtlineare Lösungsverfahren für die inkompressiblen Navier–Stokes-Gleichungen*, Universität Dortmund, Diss., 2003
- [Sch09] SCHMIDT, Sebastian: *On numerical simulation of granular flow*, Universität Kaiserslautern, Diss., 2009
- [Shu09] SHU, Chi-Wang: Discontinuous Galerkin methods: general approach and stability. In: *Numerical solutions of partial differential equations*. Birkhäuser, 2009 (Adv. Courses Math. CRM Barcelona), S. 149–201
- [SW06] SUN, Shuyu ; WHEELER, Mary F.: Projections of velocity data for the compatibility with transport. In: *Comput. Methods Appl. Mech. Engrg.* 195 (2006), Nr. 7-8, S. 653–673
- [TH08] TUREK, Stefan ; HRON, Jaroslav: Numerical techniques for multiphase flow with liquid-solid interaction. In: *Hemodynamical flows*. Birkhäuser, 2008, S. 379–501

- [Tur99] TUREK, Stefan: *Lecture Notes in Computational Science and Engineering*. Bd. 6: *Efficient solvers for incompressible flow problems*. Springer, 1999
- [ZS03] ZHANG, Mengping ; SHU, Chi-Wang: An analysis of three different formulations of the discontinuous Galerkin method for diffusion equations. In: *Math. Models Methods Appl. Sci.* 13 (2003)
- [ZS10] ZHANG, Xiangxiong ; SHU, Chi-Wang: On maximum-principle-satisfying high order schemes for scalar conservation laws. In: *J. Comput. Phys.* 229 (2010), Nr. 9, S. 3091–3120
- [ZS11] ZHANG, Xiangxiong ; SHU, Chi-Wang: Maximum-principle-satisfying and positivity-preserving high-order schemes for conservation laws: survey and new developments. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* (2011)

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig und ohne unerlaubte Hilfe angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden.

ISBN 978-3-86219-368-4