

Variationelle Zeitdiskretisierungen höherer Ordnung der Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten

Dissertation
zur
Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

im Fachbereich Mathematik und Naturwissenschaften
der Universität Kassel

vorgelegt von
Khalid Adrigal
aus Kassel
geboren am 29.09.1983 in Driouch/Marokko

1. Gutachter: Prof. Dr. rer. nat. Gunar Matthies (TU Dresden)
2. Gutachter: Prof. Dr. rer. nat. habil. Markus Bause (HSU Hamburg)

Kassel, den 15.07.2016

Tag der mündlichen Prüfung:

15.07.2016

1. Prüfer:

Prof. Dr. Gunar Matthies (Dresden)

2. Prüfer:

Prof. Dr. Markus Bause (Hamburg)

3. Prüfer:

Prof. Dr. Andreas Meister (Kassel)

4. Prüfer:

Prof. Dr. Andreas Bley (Kassel)

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter in der Arbeitsgruppe Analysis und angewandte Mathematik am Institut für Mathematik der Universität Kassel.

Ich möchte mich allen voran besonders bei meinem Betreuer Herrn Prof. Dr. Gunar Matthies für die fachliche Unterstützung und für den zuverlässigen Einsatz während der gesamten Erstellungsphase dieser Arbeit recht herzlich bedanken.

Ebenso richtet sich mein Dank an Herrn Prof. Dr. habil. Markus Bause aus Hamburg für sein Interesse und das Begutachten dieser Dissertation. Außerdem möchte ich mich bei Herrn Prof. Dr. Andreas Meister für seine hilfreichen Ratschläge während meiner Zeit an der Universität Kassel bedanken.

Meinen Kollegen danke ich für die schöne Zeit. Weiterhin bedanke ich mich auch bei unserer Sekretärin Christina Franke für die organisatorische Unterstützung.

Insbesondere möchte ich meinem Sohn Hamza und meiner Tochter Manar für die abwechslungsreiche Zeit danken.

Mein größter Dank gebührt nicht zuletzt meiner Frau für ihre Geduld und ihre Unterstützung.

Abstrakt

Abstrakt:

Wir betrachten zeitabhängige Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten, wobei die Bewegung des Gebietsrandes bekannt ist. Die zeitliche Entwicklung des Gebietes wird durch die ALE-Formulierung behandelt, die die Nachteile der klassischen Euler- und Lagrange-Betrachtungsweisen behebt. Die Position des Randes und seine Geschwindigkeit werden dabei so in das Gebietsinnere fortgesetzt, dass starke Gitterdeformationen verhindert werden.

Als Zeitdiskretisierungen höherer Ordnung werden stetige Galerkin-Petrov-Verfahren (cGP) und unstetige Galerkin-Verfahren (dG) auf Probleme in zeitabhängigen Gebieten angewendet. Weiterhin werden das C^1 -stetige Galerkin-Petrov-Verfahren und das C^0 -stetige Galerkin-Verfahren vorgestellt. Deren Lösungen lassen sich auch in zeitabhängigen Gebieten durch ein einfaches einheitliches Postprocessing aus der Lösung des cGP-Problems bzw. dG-Problems erhalten.

Für Problemstellungen in festen Gebieten und mit zeitlich konstanten Konvektions- und Reaktionsstermen werden Stabilitätsresultate sowie optimale Fehlerabschätzungen für die nachbereiteten Lösungen der cGP-Verfahren und der dG-Verfahren angegeben.

Für zeitabhängige Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten präsentieren wir konservative und nicht-konservative Formulierungen, wobei eine besondere Aufmerksamkeit der Behandlung der Zeitableitung und der Gittergeschwindigkeit gilt. Stabilität und optimale Fehlerschätzungen für die in der Zeit semi-diskretisierten konservativen und nicht-konservativen Formulierungen werden vorgestellt.

Abschließend wird das vlldiskretisierte Problem betrachtet, wobei eine Finite-Elemente-Methode zur Ortsdiskretisierung der Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten im ALE-Rahmen einbezogen wurde. Darüber hinaus wird eine lokale Projektionsstabilisierung (LPS) eingesetzt, um der Konvektionsdominanz Rechnung zu tragen. Weiterhin wird numerisch untersucht, wie sich die Approximation der Gebietsgeschwindigkeit auf die Genauigkeit der Zeitdiskretisierungsverfahren auswirkt.

abstract

Abstract:

We consider time-dependent convection-diffusion-reaction equations in time-dependent domains where the movement of the domain boundary is a priori known. The temporal development of the domain is treated by the arbitrary Lagrangian-Eulerian formulation (ALE) to overcome the disadvantages of the classical Euler and Lagrangian approaches. The position of the domain boundary and its velocity are extended into the domain in such a way that strong mesh deformations are prevented.

As higher order time discretizations, continuous Galerkin-Petrov methods (cGP) and discontinuous Galerkin methods (dG) are applied to problems in time-dependent domains. Furthermore, the C^1 -continuous Galerkin-Petrov method and the C^0 -continuous Galerkin methods are presented. Their solutions could be obtained, also in time-dependent domains, by a simple uniform post-processing from the solution of the cGP-problem or dG-problem, respectively.

For problems in fixed domain and with temporally constant convection and reaction terms, stability results and optimal error estimates for the post-processed solutions of cGP and dG methods are given.

We present for time-dependent convection-diffusion-reaction equations in time-dependent domains both conservative and non-conservative formulations with special attention to the treatment of the time derivative and the mesh velocity. Stability and optimal error estimates for the temporally semi-discretized conservative and non-conservative formulations are presented.

Finally, the fully discrete problem with a finite element method as spatial discretization of the convection-diffusion-reaction equation in time-dependent domains was considered in the ALE framework. In addition, a local projection stabilization (LPS) is applied to handle the dominant convection. Furthermore, we investigate numerically how the approximation of the mesh velocity influences the accuracy of the time discretizations.

1 Einleitung

Das grundlegende Ziel dieser Arbeit ist die Lösung von Konvektions-Diffusions-Reaktions-Gleichungen in einem zeitabhängigen Gebiet $\Omega_t, t \in (0, T]$, welche durch das Modell

$$\partial_t u - \varepsilon \Delta_{\mathbf{x}} u + \mathbf{b} \cdot \nabla_{\mathbf{x}} u + cu = f \quad (1.1)$$

mit einem vorbeschriebenen Rand $\Gamma_t := \partial\Omega_t$ dargestellt werden. Dabei ist $0 < \varepsilon \ll 1$ eine kleine positive Diffusionskonstante. \mathbf{b} und c sind zeitabhängige Funktionen, die von \mathbf{x} abhängen können und die Konvektion und die Reaktion bezeichnen. Das Problem (1.1) wird mit Anfangsbedingungen sowie Randbedingungen versehen.

Die numerische Lösung eines solchen Problems in zeitabhängigen Geometrien benötigt eine genauere Betrachtung als die üblichen kontinuumsmechanischen Betrachtungsweisen von Lagrange und Euler, da diese bei einer Ortsdiskretisierung mittels der Finite-Elemente-Methode ungeeignet sind.

In der reinen Euler-Formulierung, die in der Fluidodynamik eine breite Anwendung findet, wird die Strömung des Kontinuums, die beispielsweise in Rohrleitungen bzw. durch die Schwerkraft entsteht, in einem raumfesten Bezugssystem beschrieben. Ein Nachteil der Euler-Betrachtungsweise liegt z.B. in der Auflösung der sich bewegenden Ränder bzw. freien Oberflächen. Bei einer reinen Lagrange-Formulierung, die vor allem in der Strukturmechanik anzuwenden ist, wird die Bewegung eines Kontinuums in einem mitbewegten Bezugssystem beschrieben. Somit erlaubt diese Beschreibung eine einfache Verfolgung von freien Oberflächen. Allerdings können bei dieser Betrachtungsweise starke Gitterdeformationen (Elementverzerrungen) auftreten, wenn die gegebene Geschwindigkeit des Gebietsrandes auf die Gittergeschwindigkeit im Inneren des Gebietes erweitert wird. Dies kann die Stabilität und Genauigkeit der Finite-Elemente-Methode stark beeinflussen [4].

Mit dem Ziel die Mängel beider Formulierungen zu beheben, ist die ALE-Formulierung (Arbitrary Lagrangian-Eulerian Formulation) ein wichtiges technisches Werkzeug, wobei die jeweiligen Vorteile der beiden klassischen Betrachtungen kombiniert werden. Es wird ein Referenzgebiet betrachtet, das vom Partikel und dessen räumlicher Konfiguration unabhängig ist. Demzufolge werden die bei der Behandlung von freien Oberflächen in einer reinen Euler-Betrachtungsweise auftretenden Probleme sowie die Gitterdeformationen vermieden. Die ALE-Formulierung stellt einen Startpunkt für die numerische Approximation verschiedener Problemen in zeitabhängigen Gebieten dar, vgl. [24, 13, 4].

In diesem Rahmen werden konservative sowie nichtkonservative Formulierungen für Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten präsentiert, wobei eine besondere Aufmerksamkeit auf die Behandlung der Zeitableitung und auf die Gittergeschwindigkeit gerichtet wird.

Der Kern der vorliegenden Arbeit ist die Verwendung der unstetigen Galerkin-Verfahren (dG) und der stetigen Galerkin-Petrov-Verfahren (cGP) als variationelle Zeitdiskretisierungen höherer Ordnung. Bezüglich der konservativen und der nichtkonservativen Formulierung werden die Konvergenzeigenschaften der dG-Verfahren in zeitabhängigen Gebieten sowohl analytisch als auch numerisch analysiert, wohingegen wir uns bei den cGP-Verfahren hauptsächlich auf die numerische Untersuchung konzentrieren werden.

Im Falle eines festen Gebietes werden die Konvergenzeigenschaften der cGP- und dG-Verfahren in Kombination mit einer lokalen Projektionsstabilisierung (LPS) analytisch und ebenso numerisch untersucht.

Nach dieser kurzen Einführung folgt nun ein Überblick der Struktur dieser Arbeit.

In Kapitel 2 werden wir die Grundgedanken der ALE-Formulierung darlegen, zudem werden wir auf die Zeit-Semidiskretisierung eingehen und führen die Gauß–Radau- und Gauß–Lobatto-Quadraturformeln ein, die für die gesamte Arbeit, insbesondere für eine variationelle Zeitdiskretisierung höherer Ordnung wesentlich sind. Außerdem soll eine relevante vergleichende Literatur helfen den Inhalt dieser Arbeit einzuordnen.

Kapitel 3 dient als Zugang zur variationellen Zeitdiskretisierung höherer Ordnung in zeitabhängigen Gebieten, wobei wir die von Matthies und Schieweck in [29] neu eingeführten Galerkin-Verfahren auf zeitabhängige Gebiete verallgemeinern werden. Zudem werden wir zeigen, dass das in [29] aufgeführte Postprocessing ebenso für Probleme in zeitabhängigen Gebieten gilt. Der abschließende Abschnitt widmet sich dem verallgemeinerten einheitlichen Postprocessing für die betrachteten cGP- und dG-Verfahren in zeitabhängigen Gebieten.

Im vierten Kapitel werden relevante Ergebnisse von Ern und Schieweck in [14] bezüglich der Stabilität und Fehlerabschätzung für dG-Verfahren, angewendet auf das betrachtete Problem in festen Gebieten, auf cGP-Verfahren verallgemeinert. In Bezug auf dG-Verfahren unterscheidet sich unsere Arbeit von der von Ern und Schieweck zum einen durch das in Kapitel 3 eingeführte Postprocessing und zum anderen durch die definierten Interpolationsoperatoren, die bei unserer Analyse eine wichtige Rolle spielen. Anschließend werden die aus der theoretischen Analyse gewonnenen Resultate numerisch durch Beispiele veranschaulicht.

Das fünfte Kapitel befasst sich mit der nichtkonservativen und der konservativen schwachen Formulierung von (1.1) im ALE-Rahmen. Wir führen zum einen die zeitdiskrete ALE-Abbildung und zum anderen die diskrete Gittergeschwindigkeit ein, wobei wir die exakte und approximative Gittergeschwindigkeit vorstellen. Außerdem werden wir auf eine weitere wichtige Variante, nämlich auf die approximative Gittergeschwindigkeit mit Korrektur, eingehen. Bei der Stabilitäts- und Fehleranalyse stehen vor allem die dG-Verfahren im Fokus unserer Untersuchung.

Das volldiskrete dG- bzw. cGP-Problem im ALE-Rahmen wird in Kapitel 6 betrachtet. Für beide Probleme werden die konservative Formulierung und die nicht konservative Formulierung zunächst im Ort mittels der FEM im ALE-Rahmen diskretisiert. Die daraus erhaltenen algebraischen Formulierungen werden sowohl mit cGP- als auch mit dG-Verfahren in der Zeit diskretisiert. Im Anschluss daran wird kurz die Implementierung beschrieben. Abschließend werden wir anhand vergleichender numerischer Untersuchungen sehen, inwiefern sich die konservative und die nichtkonservative Formulierungen unterscheiden. Desweiteren veranschaulichen wir die unterschiedlichen Approximationen der Gittergeschwindigkeit und ihren Einfluss auf die Genauigkeit der Zeitdiskretisierungsverfahren. Für die Implementierung der numerischen Ergebnisse dient eine auf zeitabhängige Gebiete angepasste Variante von MooNMD [26], die parallel zu dieser Arbeit entwickelt wurde.

Am Ende dieser Ausarbeitung befindet sich eine kurze Zusammenfassung der vorgestellten Ergebnisse. Zudem werden mögliche Fortsetzungen der vorliegenden Arbeit aufgeführt.

Inhaltsverzeichnis

1	Einleitung	9
2	Grundlagen	13
2.1	ALE-Formulierung	13
2.1.1	Kurze Einführung	13
2.1.2	Warum die ALE-Formulierung?	16
2.1.3	Konstruktion der ALE-Abbildung in der Praxis	16
2.2	Räume und Normen	17
2.3	Regularität der ALE-Abbildung	18
2.4	Geometrische Erhaltungssätze (GCL)	18
2.5	Vergleich zur existierenden Literatur	19
2.6	Zeit-Semidiskretisierung	20
2.7	Quadraturformeln höherer Ordnung	20
2.7.1	Gauß–Radau-Quadraturformel	20
2.7.2	Gauß–Lobatto-Quadraturformel	21
3	Variationelle Zeitdiskretisierungen höherer Ordnung	23
3.1	Problemstellung	23
3.2	Stetige Galerkin–Petrov-Verfahren	24
3.2.1	cGP(1)-Verfahren	26
3.2.2	cGP(2)-Verfahren	26
3.3	Die cGP- $C^1(k)$ -Verfahren	27
3.4	Unstetige Galerkin-Verfahren	28
3.4.1	dG(0)-Verfahren	29
3.4.2	dG(1)-Verfahren	29
3.5	Die dG- $C^0(k)$ -Verfahren	29
3.6	Postprocessing	30
3.6.1	Verbindung zwischen cGP(k) und cGP- $C^1(k+1)$	30
3.6.2	Verbindung zwischen dG(k) und dG- $C^0(k+1)$	33
4	Zeitabhängige Konvektions-Diffusions-Reaktions-Gleichungen in festen Gebieten	37
4.1	Problemstellung und Ortsstabilisierung	37
4.1.1	Lokale Stabilisierung (LPS)	38
4.2	Stetige Galerkin-Verfahren (cGP)	41
4.2.1	Projektions- und Interpolationsoperatoren	42
4.2.2	Eigenschaften der Interpolationsoperatoren	44
4.2.3	Normen	47
4.2.4	Volldiskretes Problem mit Ortsstabilisierung	48
4.2.5	Volldiskretes Problem und Postprocessing	48
4.2.6	Fehleranalyse	49
4.2.7	Fehlerabschätzung für die Zeitableitung	55
4.3	Unstetige Galerkin-Verfahren (dG)	66
4.3.1	Projektions- und Interpolationsoperatoren	66
4.3.2	Eigenschaften der Interpolationsoperatoren	67
4.3.3	Normen	69

4.3.4	Volldiskretes Problem mit Ortsstabilisierung	69
4.3.5	Volldiskretes Problem und Postprocessing	69
4.3.6	Fehleranalyse	70
4.3.7	Fehlerabschätzung für die Zeitableitung	74
4.4	Numerische Ergebnisse	79
5	Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten	85
5.1	Problemstellung	85
5.1.1	Stabilitätsanalyse der Differentialgleichung im ALE-Rahmen	86
5.2	Schwache Formulierung im ALE-Rahmen	87
5.2.1	Nichtkonservative ALE-Formulierung	87
5.2.2	Konservative ALE-Formulierung	87
5.3	Zeitdiskrete ALE-Abbildung und Gittergeschwindigkeit	88
5.3.1	Exakte Gittergeschwindigkeit	88
5.3.2	Approximative Gittergeschwindigkeit	88
5.3.3	Approximative Gittergeschwindigkeit mit Korrektur	89
5.3.4	Approximatives Gebiet	89
5.4	Unstetige Galerkin-Verfahren	90
5.4.1	Schwache Formulierung im ALE-Rahmen	91
5.4.2	Stabilität	93
5.4.3	A-priori Fehleranalyse	95
5.4.4	Quadraturformel und Stabilität	100
5.4.5	Quadraturformel und Fehlerabschätzung	103
6	FE-Diskretisierung der Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten	113
6.1	FE-Semidiskretisierung im ALE-Rahmen	113
6.1.1	FE-Semidiskretisierung auf dem Referenzgebiet	113
6.1.2	FE-Semidiskretisierung der ALE-Abbildung	113
6.1.3	FE-Räume und Semi-diskrete Formulierung	114
6.1.4	Nichtkonservative ALE-Formulierung	115
6.1.5	Konservative ALE-Formulierung	115
6.2	FE-Approximierung	115
6.2.1	Algebraische Formulierung der nichtkonservativen Form	115
6.2.2	Algebraische Formulierung der konservativen Form	116
6.3	Implementierung der konservativen Form	116
6.3.1	Konservatives cGP-Problem	117
6.3.2	cGP(1)-Verfahren	117
6.3.3	cGP(2)-Verfahren	117
6.3.4	Konservatives dG-Problem	118
6.3.5	dG(0)-Verfahren	118
6.3.6	dG(1)-Verfahren	118
6.4	Implementierung der nichtkonservativen Form	118
6.4.1	Nichtkonservatives cGP-Problem	119
6.4.2	cGP(1)-Verfahren	119
6.4.3	cGP(2)-Verfahren	119
6.4.4	Nichtkonservatives dG-Problem	119
6.4.5	dG(0)-Verfahren	119
6.4.6	dG(1)-Verfahren	120
6.5	Numerische Ergebnisse	120
7	Zusammenfassung und Ausblick	131

2 Grundlagen

In diesem Kapitel werden wir sämtliche Grundlagen, Notationen und Annahmen zusammenfassen. Diese sind für eine bessere Verständlichkeit notwendig und liegen dieser Arbeit zugrunde. Zunächst werden wir die ALE-Formulierung (Arbitrary Lagrangian-Eulerian Formulation) und ihre Konstruktion in der Praxis vorstellen. Da wir eine variationelle Zeitdiskretisierung höherer Ordnung anstreben, werden wir anschließend sowohl die Gauß–Radau- als auch die Gauß–Lobatto-Quadraturformeln kennenlernen.

2.1 ALE-Formulierung

Dieser einführende Abschnitt befasst sich mit den grundlegenden Konzepten der ALE-Formulierung sowie der Konstruktion und Regularität der ALE-Abbildung.

2.1.1 Kurze Einführung

Die ALE-Formulierung basiert grundsätzlich auf die Einführung adäquater Abbildungen \mathcal{A}_t , die zu jedem Zeitpunkt $t \in [0, T]$ die ALE-Koordinate $\mathbf{Y} \in \widehat{\Omega}$ mit der räumlichen Koordinate $\mathbf{x} = \mathbf{x}(t, \mathbf{Y}) \in \Omega_t$ verbinden. Dabei sind $\widehat{\Omega}$ und Ω_t zwei beschränkte Gebiete, die sich nicht unbedingt unterscheiden müssen und $T > 0$. Zu Beginn betrachten wir die folgende Definition:

Definition 2.1. [ALE-Abbildung] Die ALE-Abbildung zwischen den Gebieten $\widehat{\Omega} \subset \mathbb{R}^d$ und $\Omega_t \in \mathbb{R}^d$ wird definiert durch

$$\mathcal{A}_t : \widehat{\Omega} \rightarrow \Omega_t, \quad \mathcal{A}_t(\mathbf{Y}) = \mathbf{x}(t, \mathbf{Y}).$$

$\widehat{\Omega}$ wird in der gesamten Arbeit als Referenzgebiet bezeichnet.

In einiger Literatur wie bei Gastaldi in [21] und bei Nobile in [32] wird das Gebiet Ω_0 an der Zeit $t = 0$ als Referenzgebiet bezeichnet. Um einige Ausdrücke vereinfacht darzustellen, setzen wir $I = [0, T]$. Mit \mathcal{S} bezeichnen wir den Zeit-Raum-Zylinder

$$\left\{ (t, \mathbf{x}) \in \mathbb{R}^{d+1} : t \in I, \mathbf{x} \in \Omega_t \right\}. \quad (2.1)$$

Wir nehmen an, dass die ALE-Abbildung \mathcal{A}_t homöomorph ist, d.h., $\mathcal{A}_t \in \mathcal{C}^0(\widehat{\Omega})$ und die Inverse $\mathcal{A}_t^{-1} : \Omega_t \rightarrow \widehat{\Omega}$ existiert mit $\mathcal{A}_t^{-1} \in \mathcal{C}^0(\Omega_t)$. Desweiteren sei die Funktion

$$t \mapsto \mathbf{x}(t, \mathbf{Y}), \quad \mathbf{Y} \in \widehat{\Omega}$$

fast überall in I differenzierbar.

Weiterhin verwenden wir die folgende Notation, um die ALE-Abbildung zwischen zwei Zeitpunkten t_1 und t_2 in I zu bezeichnen

$$\mathcal{A}_{t_1 \rightarrow t_2} : \Omega_{t_1} \rightarrow \Omega_{t_2}, \quad \mathcal{A}_{t_1 \rightarrow t_2} := \mathcal{A}_{t_2} \circ \mathcal{A}_{t_1}^{-1}.$$

Eine wesentliche Bedingung, auf der die gesamte Arbeit basiert, ist die Voraussetzung, dass die Beschreibung des Randes $\partial\Omega_t$ zu jedem Zeitpunkt $t \in [0, T]$ bekannt ist, d.h.,

$$\mathcal{A}_{0 \rightarrow t}(\partial\Omega_0) = \partial\Omega_t \quad \text{für alle } t \in I,$$

wobei $\mathcal{A}_0(\widehat{\Omega}) = \Omega_0$.

Einen erklärenden Hintergrund der ALE-Abbildung können wir anhand der folgenden Illustration veranschaulichen. Wir betrachten hierbei das Referenzgebiet $\widehat{\Omega}$ und das zeitabhängige Gebiet Ω_t an zwei verschiedenen Zeitpunkten t_1 und t_2 .

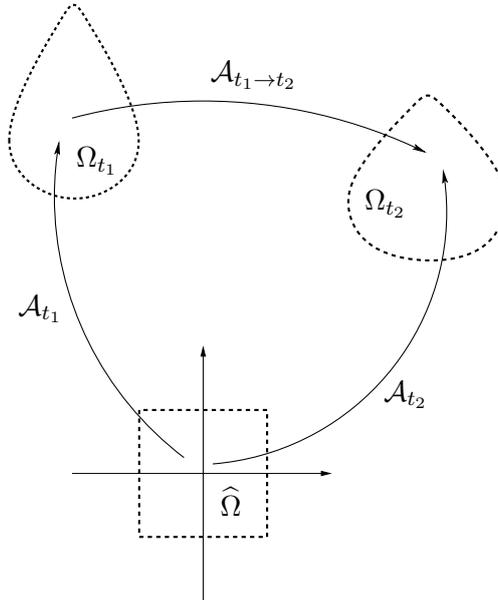


Abbildung 2.1: ALE-Abbildung

Sei nun $f : \mathcal{S} \rightarrow \mathbb{R}$ die im Euler-Rahmen definierte Funktion und $\hat{f} = f \circ \mathcal{A}_t : \widehat{\mathcal{S}} \rightarrow \mathbb{R}$ die dazugehörige Funktion, die im ALE-Rahmen definiert ist. Dabei ist $\widehat{\mathcal{S}} := I \times \widehat{\Omega}$. Der Zusammenhang zwischen den Funktionen f und \hat{f} wird wie folgt erklärt:

$$\begin{aligned} \hat{f} : \widehat{\mathcal{S}} &\rightarrow \mathbb{R}, & \hat{f}(t, \mathbf{Y}) &= f(t, \mathcal{A}_t(\mathbf{Y})) = f(t, \mathbf{x}), \\ f : \mathcal{S} &\rightarrow \mathbb{R}, & f(t, \mathbf{x}) &= \hat{f}(t, \mathcal{A}_t^{-1}(\mathbf{x})). \end{aligned}$$

Außerdem ist die Zeitableitung von f im ALE-Rahmen definiert durch

$$\left. \frac{\partial f}{\partial t} \right|_{\mathbf{Y}} : \mathcal{S} \rightarrow \mathbb{R}^d, \quad \left. \frac{\partial f}{\partial t} \right|_{\mathbf{Y}}(t, \mathbf{x}) = \left. \frac{\partial \hat{f}}{\partial t} \right|_{\mathbf{x}}(t, \mathbf{Y}), \quad \mathbf{Y} = \mathcal{A}_t^{-1}(\mathbf{x}).$$

Für die Analogie benennen wir die räumliche Zeitableitung $\left. \frac{\partial f}{\partial t} \right|_{\mathbf{x}}$, zudem definieren wir die Gebietsgeschwindigkeit \mathbf{w} durch

$$\mathbf{w}(t, \mathbf{x}) := \left. \frac{\partial \mathbf{x}}{\partial t} \right|_{\mathbf{Y}}(t, \mathbf{Y}) = \left. \frac{\partial \mathbf{x}}{\partial t} \right|_{\mathbf{Y}}(t, \mathcal{A}_t^{-1}(\mathbf{x})).$$

Mit Hilfe der Kettenregel erhalten wir eine Relation zwischen den Zeitableitungen von f im ALE-Rahmen und im Euler-Rahmen

$$\begin{aligned} \left. \frac{\partial f}{\partial t} \right|_{\mathbf{Y}}(t, \mathbf{x}) &= \left. \frac{\partial f}{\partial t} \right|_{\mathbf{x}}(t, \mathbf{x}) + \left. \frac{\partial \mathbf{x}}{\partial t} \right|_{\mathbf{Y}}(t, \mathbf{Y}) \cdot \nabla_{\mathbf{x}} f(t, \mathbf{x}) \\ &= \left. \frac{\partial f}{\partial t} \right|_{\mathbf{x}}(t, \mathbf{x}) + \mathbf{w}(t, \mathbf{x}) \cdot \nabla_{\mathbf{x}} f(t, \mathbf{x}), \end{aligned} \tag{2.2}$$

wobei $\nabla_{\mathbf{x}}$ die Ortsableitung bezüglich \mathbf{x} bezeichnet.

Abschließend wollen wir noch für die ALE-Abbildung \mathcal{A}_t die Jacobi-Matrix $\mathbf{J}_{\mathcal{A}_t}$ und ihre Determinante $J_{\mathcal{A}_t}$ folgendermaßen definieren:

$$(\mathbf{J}_{\mathcal{A}_t})_{i,j} := \frac{\partial \mathcal{A}_{ti}}{\partial \mathbf{Y}_j} \quad \text{und} \quad J_{\mathcal{A}_t} := \det \mathbf{J}_{\mathcal{A}_t}.$$

Wir betrachten die folgende Transformation der Zeitableitungen der Terme in $J_{\mathcal{A}_t}$

$$\frac{\partial}{\partial t} \left(\frac{\partial \mathcal{A}_{ti}}{\partial \mathbf{Y}_j} \right) = \frac{\partial}{\partial \mathbf{Y}_j} \left(\frac{\partial \mathcal{A}_{ti}}{\partial t} \right) = \frac{\partial \mathbf{w}_i}{\partial \mathbf{Y}_j} = \sum_{k=1}^d \frac{\partial \mathbf{w}_i}{\partial \mathbf{x}_k} \frac{\partial \mathbf{x}_k}{\partial \mathbf{Y}_j}.$$

Eine ausführliche Entwicklung der Determinante in $\left. \frac{\partial J_{\mathcal{A}_t}}{\partial t} \right|_{\mathbf{Y}}$ führt zu

$$\left. \frac{\partial J_{\mathcal{A}_t}}{\partial t} \right|_{\mathbf{Y}} = \sum_{k=1}^d \left(\frac{\partial \mathbf{w}_k}{\partial \mathbf{x}_k} \right) J_{\mathcal{A}_t},$$

d.h., die zeitliche Entwicklung von $J_{\mathcal{A}_t}$ kann durch die Divergenz der Gebietsgeschwindigkeit gemäß der folgenden Differentialgleichung beschrieben werden (Euler-Expansionsformel) [36]:

$$\left. \frac{\partial J_{\mathcal{A}_t}}{\partial t} \right|_{\mathbf{Y}} (\mathbf{Y}) = J_{\mathcal{A}_t} \nabla_{\mathbf{x}} \cdot \mathbf{w}(t, \mathcal{A}_t(\mathbf{Y})). \quad (2.3)$$

Hierbei ist erwähnenswert, dass im Falle $\Omega_t = \widehat{\Omega}$, $J_{\mathcal{A}_t} = 1$ gilt. Aus (2.3) folgt

$$\det \mathbf{J}_{\mathcal{A}_t}(\mathbf{Y}) = e^{\int_0^t \nabla_{\mathbf{x}} \cdot \mathbf{w}(s, \mathcal{A}_s(\mathbf{Y})) ds} \det \mathbf{J}_{\mathcal{A}_0}(\mathbf{Y}). \quad (2.4)$$

Aus dieser kann man entnehmen, dass $\det \mathbf{J}_{\mathcal{A}_t}(\mathbf{Y})$ ein konstantes Vorzeichen hat und zwar das von $\det \mathbf{J}_{\mathcal{A}_0}(\mathbf{Y})$. Der Einfachheit halber bezeichnen wir im Folgenden die Zeitableitung im ALE-Rahmen mit $D_t f$ und mit $\partial_t f$ die gewöhnliche Zeitableitung im Euler-Rahmen, d.h.,

$$D_t f := \left. \frac{\partial f}{\partial t} \right|_{\mathbf{Y}}, \quad \partial_t f := \left. \frac{\partial f}{\partial t} \right|_{\mathbf{x}}. \quad (2.5)$$

Vorsehen mit der aus dem Skalarprodukt $(u, v) = \int_{\Omega_t} uv \, d\mathbf{x}$ induzierten Norm $\|\cdot\|_{L^2(\Omega_t)}$ ist der Lebesgue-Raum $L^2(\Omega_t)$ der quadratisch integrierbaren Funktionen ein Hilbert-Raum. Die partielle Integration im ALE-Rahmen ergibt sich aus dem folgenden Satz:

Satz 2.2. [Reynolds-Identität] Seien u und v zwei Funktionen, die auf dem Euler-System \mathcal{S} definiert sind. Es gilt

$$(D_t u, v)_{\Omega_t} = \partial_t (u, v)_{\Omega_t} - (u, D_t v)_{\Omega_t} - (u, v \nabla_{\mathbf{x}} \cdot \mathbf{w})_{\Omega_t}. \quad (2.6)$$

Beweis. Nach der Transformation auf dem Referenzgebiet liefert eine partielle Integration

$$\begin{aligned} \int_{\Omega_t} D_t uv \, d\mathbf{x} &= \int_{\widehat{\Omega}} \partial_t uv \det \mathbf{J}_{\mathcal{A}_t} \, d\widehat{\mathbf{x}} \\ &= \frac{\partial}{\partial t} \int_{\widehat{\Omega}} uv \det \mathbf{J}_{\mathcal{A}_t} \, d\widehat{\mathbf{x}} - \int_{\widehat{\Omega}} u \partial_t (v \det \mathbf{J}_{\mathcal{A}_t}) \, d\widehat{\mathbf{x}} \\ &= \frac{\partial}{\partial t} \int_{\widehat{\Omega}} uv \det \mathbf{J}_{\mathcal{A}_t} \, d\widehat{\mathbf{x}} - \int_{\widehat{\Omega}} u \partial_t v \det \mathbf{J}_{\mathcal{A}_t} \, d\widehat{\mathbf{x}} - \int_{\widehat{\Omega}} uv \partial_t \det \mathbf{J}_{\mathcal{A}_t} \, d\widehat{\mathbf{x}} \\ &= \frac{\partial}{\partial t} \int_{\Omega_t} uv \, d\mathbf{x} - \int_{\Omega_t} u D_t v \, d\mathbf{x} - \int_{\Omega_t} uv \nabla_{\mathbf{x}} \cdot \mathbf{w} \, d\mathbf{x}, \end{aligned}$$

wobei wir im letzten Schritt die Euler-Expansionsformel (2.3) verwendet haben. Abschließend folgt aus der Rücktransformation auf Ω_t die Behauptung. \square

Im folgenden Unterabschnitt werden wir eine Begründung für die Wahl der ALE-Formulierung darlegen.

2.1.2 Warum die ALE-Formulierung?

Bei einer Ortsdiskretisierung eines Problems in zeitabhängigen Gebieten mittels der Finite-Elemente-Methode in den konventionellen Darstellungen von Lagrange und Euler entstehen gewisse Nachteile bezüglich des Rechengitters.

Die Lagrange-Formulierung, die vor allem in der Strukturmechanik angewendet wird, besteht darin, die Partikel in ihrer Bewegung zu folgen. Zu diesem Zweck führt man ein Rechengitter ein, das dem Partikel in seiner Bewegung folgt. Hierbei sind die Gitterknoten permanent mit demselben Partikel verbunden. Somit erlaubt die Lagrange-Beschreibung eine einfache Verfolgung von freien Oberflächen, jedoch hat diese Sicht den Nachteil nicht die großen Deformationen des Rechengebietes behandeln zu können. Ohne zusätzlichen Rückgriff auf neue Vernetzungen bzw. Netzkorrektur, führt dies zu einem Genauigkeitsverlust, vgl. [13, 4].

Bei der reinen Euler-Formulierung, die in der Fluidströmung häufig verwendet wird, kann man die oben erwähnten Gitterdeformationen überwinden. Die Grundidee der Euler-Sicht ist die Behandlung der Bewegung eines Partikels in einem festen Gebiet. Die Materialgeschwindigkeit an einem gegebenen Gitterknoten entspricht an jeder Zeit t der Geschwindigkeit des Partikels, der mit dem Knoten verbunden ist. Die Gittergeschwindigkeit wird somit in Bezug auf ein festes Gitter und ohne Referenz auf die erste Konfiguration beschrieben. Infolge der Entkopplung der Gitterknoten von den Partikeln erscheinen Konvektionseffekte aufgrund der relativen Bewegung zwischen dem Partikel und dem Rechengitter. Ein weiterer Nachteil der Euler-Beschreibung ist in der Auflösung der sich bewegenden Rändern und freien Oberflächen, vgl. [13, 4].

Aufgrund der Defizite der reinen Lagrange- und Euler-Beschreibungen erweist sich die Einführung der ALE-Formulierung als eine Kombination beider Ansätze als vorteilhaft, da hierbei nur die besten Eigenschaften der beiden Formulierungen ausgenutzt werden. Dabei werden sowohl die Freien-Oberflächen-Probleme gelöst als auch die Gitterdeformationen verhindert, indem ein Referenzgebiet betrachtet wird, das vom Partikel (Lagrange-Sicht) und dessen räumlicher Konfiguration (Euler-Sicht) unabhängig ist, siehe [24, 13, 4]. In diesem Referenzgebiet wird die referenzielle bzw. die ALE-Koordinate \mathbf{Y} der Gitterpunkte definiert. In diesem Zusammenhang wird die ALE-Abbildung \mathcal{A}_t eingeführt, damit zwischen den unterschiedlichen Konfigurationen transformiert werden kann.

Die vorliegende Arbeit basiert grundsätzlich auf der ALE-Formulierung, wobei die Gebiets- bzw. Gittergeschwindigkeit eine besonders große Rolle spielt. Das Gebiet Ω_t lässt sich als das zu $\widehat{\Omega}$ durch die ALE-Abbildung \mathcal{A}_t zugeordnete Bild darstellen. Insbesondere haben wir die Eigenschaft $\partial\Omega_t = \mathcal{A}_t(\partial\widehat{\Omega})$, die bei Problemen mit vorbeschriebenen Ränder bei der Konstruktion von Ω_t ausgenutzt wird.

2.1.3 Konstruktion der ALE-Abbildung in der Praxis

Um die ALE-Abbildung zu konstruieren, betrachten wir das folgende Problem:

Problem 2.3. *Gegeben sei die Entwicklung des Gebietsrandes*

$$h : I \times \partial\widehat{\Omega} \rightarrow \partial\Omega_t.$$

Finde eine ALE-Abbildung \mathcal{A}_t , sodass an jeder Zeit $t \in I$,

$$\mathcal{A}_t(\mathbf{Y}) = h(t, \mathbf{Y}), \quad \forall \mathbf{Y} \in \partial\widehat{\Omega},$$

gilt.

Es gibt in der Literatur verschiedene Techniken, die dieses Problem behandeln, wie z.B. Farhat, Lesoinne und Maman, die in [17] ein elastisches Problem betrachtet haben, um die Bewegung des Gebietes zu beschreiben. Dabei stellt die ALE-Abbildung die Deformation des Gebietes dar. Diese wird konstruiert, indem ein Elastizitätsproblem gelöst wird. Dieselbe Technik wurde von Gastaldi in [21] verwendet, wobei das Gebiet als ein elastischer Körper angesehen wird und die ALE-Abbildung die Umformung zu dem aktuellen Gebiet darstellt. Nobile hingegen hat in [32] zwei andere Varianten der Konstruktion der ALE-Abbildung vorgestellt. Die erste besteht darin, ein parabolisches Problem zu betrachten, das die Evolution des Randes Γ_t entlang der Seitenfläche von \mathcal{S} beschreibt. Als vereinfachte Version der ersten wird bei der zweiten Variante die Position des Randes an der Zeit t durch eine Funktion $h : \partial\hat{\Omega} \rightarrow \partial\Omega_t$ beschrieben. Das aktuelle Gebiet zu der Zeit t ergibt sich dann durch eine harmonische Erweiterung des Randes auf das gesamte Gebiet. Dies liefert die Lösung des Problems:

Problem 2.4. Seien $\hat{\Omega}$ und h gegeben. Finde $\mathbf{x} : \hat{\Omega} \rightarrow \Omega_t$, sodass

$$\begin{cases} \nabla_{\mathbf{Y}} \cdot (\kappa \nabla_{\mathbf{Y}} \mathbf{x}) = 0 & \mathbf{Y} \in \hat{\Omega}, \\ \mathbf{x}(\mathbf{Y}) = h(\mathbf{Y}) & \mathbf{Y} \in \partial\hat{\Omega}, \end{cases}$$

wobei κ eine Konstante ist.

Diese Variante wurde beispielsweise in [25, 33] verwendet. Eine weitere Technik, die sich von den oben genannten Vorgehensweisen unterscheidet und von der beispielsweise beim Bau von Flugzeugen und Turbinen (Schwingungen der Flügel) Gebrauch gemacht wird, wurde von Cesenek und Feistauer in [18] beschrieben.

In der gesamten Arbeit konzentrieren wir uns auf den Fall, bei dem die Bewegung des Gebietsrandes vorbeschrieben ist. Die Position des Randes ist nur an diskreten Stellen bekannt, daher ist die zweite erwähnte Methode von Nobile, siehe Problem 2.4, am geeignetsten für diese Arbeit.

2.2 Räume und Normen

In diesem Unterabschnitt definieren wir einige Bochner-Räume und -Normen, die wir in der gesamten Arbeit für das allgemeine Verständnis benötigen werden. Mit $H^r(\Omega)$ bezeichnen wir den Sobolevraum der Funktionen in $L^2(\Omega)$, deren schwachen partiellen Ableitungen bis zur Ordnung r ebenso in $L^2(\Omega)$ liegen. Für einen Multiindex $\alpha \in \mathbb{N}_0^d$ vereinbaren wir $|\alpha| := \alpha_1 + \dots + \alpha_d$ und

$$\partial^{\alpha} v := \frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}.$$

Wir definieren durch

$$|v|_k := \left(\sum_{|\alpha|=k} \|\partial^{\alpha} v\|^2 \right)^{1/2}$$

und

$$\|v\|_r := \left(\sum_{k=0}^r |v|_k^2 \right)^{1/2}$$

Halbnormen und Normen. Weiterhin sei J ein Zeitintervall und V ein Banachraum mit der Norm $\|\cdot\|_V$. Wir definieren die Räume

$$\begin{aligned} C(J, V) &:= \{v : J \rightarrow V, \quad v \text{ stetig}\}, \\ L^2(J, V) &:= \left\{ v : J \rightarrow V, \quad \int_J \|v(t)\|^2 dt < \infty \right\}, \\ H^r(J, V) &:= \left\{ v : J \rightarrow V, \quad v \in L^2(J, V), \quad \frac{\partial^j v}{\partial t^j} \in L^2(J, V), \quad 1 \leq j \leq r \right\}. \end{aligned}$$

Abschließend bezeichnen wir mit $W^{r,\infty}(V)$ den Sobolevraum derjenigen Funktionen $v \in L^\infty(V)$, deren schwachen Ableitungen bis zur Ordnung r auch im Lebesgue-Raum $L^\infty(V)$ liegen. Dabei ist

$$L^\infty(V) := \left\{ v : V \rightarrow \mathbb{R}, \quad : \sup_{\mathbf{x} \in V} |v(\mathbf{x})| < \infty \right\}.$$

2.3 Regularität der ALE-Abbildung

Hier werden wir auf die Regularität der ALE-Abbildung eingehen. In dieser Arbeit interessieren wir uns für Strömungen und elastische Strukturen, deshalb wollen wir mit den Funktionen $v(\cdot, t) \rightarrow V(\Omega_t) \subset H^1(\Omega_t)$ arbeiten, die sich in $H^1(I, H^1(\Omega_t))$ befinden. Die ALE-Abbildung \mathcal{A}_t wird, wie bereits erwähnt, mithilfe der Randbewegung konstruiert. Wir werden in Kapitel 6 sehen, dass bei der Ortsdiskretisierung eine Finite-Elemente-Methode herangezogen wird. Dabei wird \mathcal{A}_t durch eine Kombination von Basisfunktionen dargestellt, die sich in $H^1(\Omega_t)$ befinden müssen. Im Folgenden werden wir hinreichende Bedingungen an \mathcal{A}_t vorstellen, die die Regularität der Funktionen $v = \hat{v} \circ \mathcal{A}_t^{-1} : \Omega_t \rightarrow \mathbb{R}$ garantieren, d.h., für diese gilt $\hat{v} \in H^1(\hat{\Omega})$ genau dann, wenn $v = \hat{v} \circ \mathcal{A}_t^{-1} \in H^1(\Omega_t)$. Dazu benutzen wir die folgenden zwei Sätze, vgl. [19, prop 2.1, prop 2.2]:

Satz 2.5. *Sei $\hat{\Omega}$ ein beschränktes Gebiet mit einem Lipschitz-Rand. Zu $\mathcal{A}_t : \hat{\Omega} \rightarrow \Omega_t$ existiert eine Inverse auf $\overline{\hat{\Omega}}$, sodass die folgenden Bedingungen für alle $t \in I$ erfüllt sind:*

- $\Omega_t = \mathcal{A}_t(\hat{\Omega})$ ist beschränkt und besitzt einen Lipschitz-Rand,
- $\mathcal{A}_t \in W^{1,\infty}(\hat{\Omega})$, $\mathcal{A}_t^{-1} \in W^{1,\infty}(\Omega_t)$.

Dann gilt $v \in H^1(\Omega_t)$ genau dann, wenn $\hat{v} = v \circ \mathcal{A}_t \in H^1(\hat{\Omega})$ erfüllt ist. Außerdem ist $\|\cdot\|_{H^1(\Omega_t)}^2$ äquivalent zu $\|\cdot\|_{H^1(\hat{\Omega})}^2$ für alle $v \in H^1(\Omega_t)$, d.h. es existiert eine nur von \mathcal{A}_t abhängige Konstante $C > 0$ mit

$$\frac{1}{C} \|\hat{v}\|_{H^1(\hat{\Omega})}^2 \leq \|\hat{v} \circ \mathcal{A}_t^{-1}\|_{H^1(\Omega_t)}^2 \leq C \|\hat{v}\|_{H^1(\hat{\Omega})}^2 \quad \forall \hat{v} \in H^1(\hat{\Omega}).$$

Aus diesem Satz kann man schlußfolgern, dass wir einen Isomorphismus zwischen $H^1(\hat{\Omega})$ und $H^1(\Omega_t)$ konstruiert haben. Sei nun $\hat{v} \in H^1(I, H^1(\hat{\Omega}))$ mit $\hat{v}(t, \mathbf{Y}) := \sum_{i=1}^{\infty} v_i(t) \psi_i(\mathbf{Y})$, wobei $\{\psi_i(\mathbf{Y})\}$ eine orthonormale Basis von $H^1(\hat{\Omega})$ ist und $v_i \in H^1(I)$. Dann gilt

$$D_t v \circ \mathcal{A}_t(\mathbf{Y}) = \frac{\partial \hat{v}}{\partial t}(\mathbf{Y}) = \sum_{i=1}^{\infty} \frac{\partial v_i}{\partial t} \psi_i(\mathbf{Y}).$$

Unter Anwendung der Normäquivalenz aus Satz 2.5 ergibt sich, dass die Menge $\{\phi_i : \phi_i(\mathbf{x}) := \psi_i \circ \mathcal{A}_t^{-1}(\mathbf{x})\}$ eine Basis von $H^1(\Omega_t)$ darstellt, daher die folgende zeitliche Regularität:

Satz 2.6. *Unter der Annahme $\mathbf{x} \in H^1(I, W^{1,\infty}(\hat{\Omega}))$ folgt, dass wenn $\hat{v} \in H^1(I, H^1(\hat{\Omega}))$, dann auch $v = \hat{v} \circ \mathcal{A}_t^{-1} \in H^1(I, H^1(\Omega_t))$ und*

$$D_t v \in L^2(I, H^1(\Omega_t)).$$

2.4 Geometrische Erhaltungssätze (GCL)

Beim Umgang mit den partiellen Differentialgleichungen, die auf beweglichen Gebieten zu lösen sind, muss darauf geachtet werden, dass das zu lösende Problem die sogenannten „Geometrischen Erhaltungssätze“ (geometric conservation law (GCL)) erfüllt. Unter Verwendung der

ALE-Formulierung auf einem sich bewegenden Gitter tauchen geometrische Größen auf, wie die Jacobi-Matrix und die Oberflächennormale, die sich an der Berechnung der Positionen und Geschwindigkeiten beteiligen. Die Nichtbeachtung der (GCL) bedeutet, dass die Stabilität der Zeitintegration nicht gewährleistet wird und somit könnte die Konvergenz nicht beibehalten werden. Die (GCL) sind erfüllt, wenn der Algorithmus eine konstante Lösung exakt reproduzieren kann. Ein Problem ist konservativ, wenn es in Abwesenheit von äußeren Kräften fähig ist, eine konstante Lösung zu reproduzieren. Dementsprechend sind die Konstanten zulässige Testfunktionen.

Die GCL wurden zum ersten Mal von Thomas und Lombard in [38], von Guillard und Farhat in [23] und von Farhat in [16] für Finite-Differenzen- und Finite-Volumen-Verfahren als minimales Kriterium für unbedingte Stabilität eingeführt. Diese GCL sind jedoch nicht notwendig, um Verfahren zu erhalten, deren Stabilität von der ALE-Abbildung frei ist, vgl. Geuzaine, Grandmont und Farhat [22]. Hier wurde zum Beispiel von Geuzaine und Co. ein Finite-Volumen-Verfahren zweiter Ordnung im ALE-Rahmen betrachtet, das dieselben Stabilitätseigenschaften besitzt, wie auch das stetige Problem.

2.5 Vergleich zur existierenden Literatur

Bei der Zeitdiskretisierung der Advektions-Diffusionsprobleme auf beweglichen Gebieten wurden im Anschluss von Finite-Differenzen- und Finite-Volumen-Methoden verschiedene ALE-Verfahren bis zur zweiten Ordnung, wie Euler-, Crank–Nicolson- und BDF-Verfahren betrachtet, vgl. [17, 16, 15, 32]. Hierbei wurden ALE-frei stabile Verfahren bis zur zweiten Ordnung entwickelt. Bezüglich der FEM-Diskretisierung im ALE-Rahmen wurden in [32, 5, 19, 20, 21] verschiedene Verfahren auf Stabilität untersucht. Mit ALE-frei wurde nur das Rückwärts-Euler-Verfahren nachgewiesen. Hingegen haben Bonito, Kyza und Nochetto die dG-Verfahren als Diskretisierungsverfahren höherer Ordnung in der Zeit eingesetzt. Dabei wurde die ALE-freie Stabilität in [7] und die ALE-freie Fehlerabschätzung in [6] sowohl für den stetigen Fall als auch im diskreten Fall gezeigt, wobei eine „unpraktische“ und ALE-abhängige Reynolds-Quadraturformel verwendet wurde. Die RKR-Verfahren, bei denen eine Radau-Quadratur eingesetzt wird, sind unter einer milden Bedingung über die Zeitschrittweite ALE-stabil. Diese Bedingung ist nach Bonito und Co. keine CFL-Bedingung.

In dieser Arbeit werden wir denselben Weg von Bonito und Co. folgen, um variationelle Zeitdiskretisierungen höherer Ordnung der Konvektions-Diffusions-Reaktions-Gleichungen (1.1) in zeitabhängigen Gebieten zu erhalten. Dabei werden wir die Gauß–Radau- und Gauß–Lobatto-Quadratur jeweils mit den dG- bzw. cGP-Verfahren kombinieren. Als Zugang dazu werden wir die von Matthies und Schieweck in [29] eingeführten variationellen Zeitdiskretisierungen höherer Ordnung auf zeitabhängige Gebiete verallgemeinern und auf unser Problem anwenden. Dazu werden wir auf die von Bonito und Co. bzw. von Nobile und Co. betrachteten Bedingung $\nabla_{\mathbf{x}} \cdot \mathbf{b} = 0$ verzichten. Stattdessen werden wir in unserem Fall den Reaktionsterm c ausnutzen und die Annahme

$$c(t, \mathbf{x}) - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}(t, \mathbf{x}) \geq c_0 > 0 \quad \text{in } \Omega_t, t \in I,$$

betrachten. Dabei ist c_0 eine Konstante. Dies stellt, wie wir sehen werden, keine Einschränkung der Allgemeinheit dar. Ein weiterer wichtiger Gewinn ist, dass wir Stabilitäts- sowie Fehlerabschätzungen erhalten, die im Gegensatz zu beispielsweise Bonito und Co., den Term $\frac{1}{\varepsilon}$ nicht enthalten.

2.6 Zeit-Semidiskretisierung

Sei

$$t_0 = 0 < t_1 < \dots < t_N = T$$

eine nicht notwendig äquidistante Unterteilung des Zeitintervalls I in halboffene Teilintervalle $I_n = (t_{n-1}, t_n]$, $n = 1, \dots, N$. Wir setzen $\tau := \max_{1 \leq n \leq N} \tau_n$ mit $\tau_n := t_n - t_{n-1}$. τ wird Zeitdiskretisierungsparameter genannt. Wir bezeichnen mit $\mathcal{M}_\tau := \{I_n, n = 1, \dots, N\}$ das Zeitgitter. Sei $\mathcal{V} := \mathbb{R}^m$. Unser Ziel ist die Approximation der exakten Lösung $\mathbf{u} : I \rightarrow \mathcal{V}$ mit einer Funktion $\mathbf{u}_\tau : I \rightarrow \mathcal{V}$, die stückweise aus Polynomen höchstens k -ten Grades zusammengesetzt ist. Dazu definieren wir die Räume

$$\mathbb{P}_k(I_n, \mathcal{V}) := \left\{ \mathbf{v} : I_n \rightarrow \mathcal{V} : \mathbf{v}(t) = \sum_{j=0}^k \mathbf{V}^j t^j \quad \forall t \in I_n, \quad \mathbf{V}^j \in \mathcal{V} \quad \forall j \right\}$$

der vektorwertigen Polynome der Ordnung k in der Zeit über dem Intervall I_n .

Wir betrachten $\hat{I} := (-1, 1]$ als Referenzintervall und definieren eine affine Referenztransformation $T_n : \hat{I} \rightarrow I_n$ durch

$$t := T_n(\hat{t}) := (t_{n-1} + t_n)/2 + \frac{\tau_n}{2} \hat{t}.$$

Für $\mathbf{v} \in \mathbb{P}_k(I_n, \mathcal{V})$ gilt $\mathbf{v}(t) = \hat{\mathbf{v}}(T_n^{-1}(t))$, wobei $\hat{\mathbf{v}} \in \mathbb{P}_k(\hat{I}, \mathcal{V})$.

2.7 Quadraturformeln höherer Ordnung

In diesem Unterabschnitt werden wir die Gauß–Radau- und Gauß–Lobatto-Quadraturformeln vorstellen. Diese werden in dieser Arbeit benötigt, um die Integrale, die aus den variationellen Formulierungen entstehen, mit einem höchstmöglichen Genauigkeitsgrad anzunähern.

Die Gauß–Radau- und Gauß–Lobatto-Formeln sind Quadraturformeln, die durch die Gewichtungsfunktion $W(\hat{t}) = 1$ für alle $\hat{t} \in \hat{I}$ und die Funktionswerte an speziellen Stützstellen konstruiert sind. Dabei werden nicht nur die inneren Punkte des Intervalls der Integration verwendet, sondern ebenso ein oder beide Endpunkte dieses Intervalls. Mit ihren Eigenschaften stellen die Gauß–Radau- und Gauß–Lobatto-Quadraturformeln bei der Integration der Differentialgleichungen oder die Lösung der Integralgleichungen, die Basis für viele Verfahren höherer Ordnung dar.

2.7.1 Gauß–Radau-Quadraturformel

Die Gauß–Radau-Quadraturformel unterscheidet sich von der Gauß-Quadraturformel dadurch, dass zu den inneren Stützstellen noch zusätzlich ein Randpunkt, der Anfangspunkt t_{n-1} (linksseitige Quadraturformel) bzw. der Endpunkt t_n (rechtsseitige Quadraturformel), als Quadraturpunkt des Intervalls hinzugenommen wird. Die beiden Varianten sind für Polynome vom Grad höchstens $2k$ exakt.

Auf dem Referenzintervall \hat{I} und für $\hat{\mathbf{v}} \in C(\hat{I})$ hat die rechtsseitige Gauß–Radau-Quadraturformel, die in dieser Arbeit verwendet wird, die Form

$$\hat{Q}[\hat{\mathbf{v}}] := \sum_{\mu=1}^{k+1} \hat{\omega}_\mu \hat{\mathbf{v}}(\hat{t}_\mu) = \int_{-1}^1 \hat{\mathbf{v}}(\hat{t}) d\hat{t} - \hat{R}_k. \quad (2.7)$$

Dabei ist die Stützstelle \hat{t}_μ die μ -te Nullstelle von

$$P_k(-\hat{t}) + P_{k+1}(-\hat{t}),$$

wobei P_k das k -te Legendre-Polynom mit $P_k(1) = 1$ ist. Die Stützstellen \hat{t}_μ , $\mu = 1, \dots, k+1$, werden wie folgt sortiert:

$$-1 \leq \hat{t}_1 < \dots < \hat{t}_{k+1} = 1$$

und die dazugehörigen Gewichte $\hat{\omega}_\mu > 0$ sind durch

$$\hat{\omega}_{k+1} = \frac{2}{(k+1)^2} \quad \text{und} \quad \hat{\omega}_\mu = \frac{1}{(k+1)^2} \frac{1 - \hat{t}_\mu}{[P_k(\hat{t}_\mu)]^2} = \frac{1}{1 - \hat{t}_\mu} \frac{1}{[P'_k(\hat{t}_\mu)]^2} \quad \text{für } \mu \neq k+1$$

definiert. Sowohl die Stützstellen als auch die zugehörigen Gewichte bis $k = 23$ kann man z.B. dem Buch „Handbook of Mathematical Functions“ [1] von Abramowitz und Stegun entnehmen. Der Fehleranteil R_k ist gegeben durch

$$\hat{R}_k := \frac{2^{2k+1}(k+1)}{[(2k+1)!]^3} [(k+1)!]^4 \hat{\mathbf{v}}^{(2k+1)}(\hat{\xi}),$$

wobei $\hat{\xi}$ eine Zwischenstelle aus \hat{I} ist. Mit Q_n wird die Gauß–Radau-Quadraturformel auf dem Zeitintervall I_n bezeichnet. Diese Formel lässt sich auf jedem Teilintervall I_n für alle $\mathbf{v} \in \mathbb{P}_k(I_n, \mathcal{V})$ wie folgt definieren:

$$Q_n[\mathbf{v}] := \frac{\tau_n}{2} \sum_{\mu=1}^{k+1} \omega_{n,\mu} \mathbf{v}(t_{n,\mu}) = \int_{I_n} \mathbf{v}(t) dt - R_n.$$

Es gilt dabei $t_{n,\mu} := T_n(\hat{t}_\mu)$, $\mu = 1, \dots, k+1$. Der Fehleranteil R_n auf jedem Teilintervall I_n lässt sich mit \hat{R}_k anhand der Referenztransformation T_n gewinnen. Dieser Fehleranteil kann auf jedem Teilintervall folgendermaßen abgeschätzt werden:

Lemma 2.7. [Quadraturfehler] Sei $v : I_n \rightarrow \mathbb{R}$ eine $(2k+1)$ -mal differenzierbare Funktion. Es gilt

$$|R_n| \leq C \tau_n^j \sup_{I_n} |\partial_t^j v(\xi)|, \quad \xi \in I_n, \quad 0 \leq j \leq 2k+1,$$

d.h., mit $(k+1)$ Stützstellen hat eine Gauß–Radau-Quadraturformel den höchsten algebraischen Genauigkeitsgrad $2k$.

2.7.2 Gauß–Lobatto-Quadraturformel

Bei der Gauß–Lobatto-Quadraturformel werden die beiden Randpunkte als Stützstellen betrachtet. Hier werden Polynome vom Grad höchstens $2k-1$ exakt integriert. Die allgemeine Form lautet:

$$\hat{Q}[\hat{\mathbf{v}}] := \sum_{i=0}^k \hat{\omega}_i \hat{\mathbf{v}}(\hat{t}_i) = \int_{-1}^1 \hat{\mathbf{v}}(\hat{t}) d\hat{t} - \hat{E}_k. \quad (2.8)$$

Dabei ist die Stützstelle \hat{t}_μ die $(\mu-1)$ -te Nullstelle von $(1-\hat{t}^2)P'_k(\hat{t})$, wobei P_k das k -te Legendre-Polynom ist. Die Stützstellen \hat{t}_μ , $\mu = 0, \dots, k$, werden wir folgendermaßen auflisten:

$$-1 = \hat{t}_0 < \hat{t}_2 < \dots < \hat{t}_k = 1$$

und die dazugehörigen Gewichte $\hat{\omega}_\mu > 0$ sind durch

$$\hat{\omega}_0 = \hat{\omega}_{k+1} = \frac{2}{k(k+1)} \quad \text{und} \quad \hat{\omega}_\mu = \frac{2}{k(k+1)[P_k(\hat{t}_\mu)]^2} \quad \text{für } \mu = 1, \dots, k-1$$

definiert. Die Stützstellen und die zugehörigen Gewichte bis $k = 9$ kann man z.B. dem Buch [1] von Abramowitz, Stegun entnehmen. Der Fehleranteil \hat{E}_k ist hier definiert durch

$$\hat{E}_k := \frac{-2^{2k+1}k(k+1)[(k-1)!]^4}{(2k+1)[(2k)!]^3} \hat{\mathbf{v}}^{(2k)}(\hat{t}),$$

wobei $\hat{\xi}$ eine Zwischenstelle aus \hat{I} ist. Die Gauß–Lobatto-Quadraturformel Q_n definieren wir auf jedem Teilintervall I_n durch

$$Q_n[\mathbf{v}] := \frac{\tau_n}{2} \sum_{\mu=0}^k \omega_{n,\mu} \mathbf{v}(t_{n,\mu}) = \int_{I_n} \mathbf{v}(t) dt - E_n,$$

wobei $t_{n,\mu} := T_n(\hat{t}_\mu)$, $\mu = 0, \dots, k$. Wir erhalten den Fehleranteil E_n mit Hilfe der Referenztransformation T_n auf jedem Teilintervall I_n aus \hat{E}_k .

Weiterhin lässt sich der Fehler der Gauß–Lobatto-Quadraturformel auf jedem Teilintervall I_n wie folgt abschätzen:

Lemma 2.8. *[Quadraturfehler] Sei $v : I_n \rightarrow \mathbb{R}$ eine $2k$ -mal differenzierbare Funktion. Es gilt*

$$|E_n| \leq C \tau_n^{j+1} \sup_{I_n} |\partial_t^j v(t)|, \quad \xi \in I_n, \quad 0 \leq j \leq 2k,$$

demzufolge hat eine Gauß–Radau-Quadraturformel mit $(k + 1)$ Stützstellen den höchsten algebraischen Genauigkeitsgrad $2k - 1$.

3 Variationelle Zeitdiskretisierungen höherer Ordnung

In diesem Kapitel werden Zeit-Semidiskretisierungen höherer Ordnung mithilfe der Galerkin-Verfahren geschaffen, indem wir die von Matthies und Schieweck in [29] dargelegten Ergebnisse für die Galerkin-Verfahren und deren Postprocessing auf zeitabhängige Gebiete verallgemeinern werden. Infolgedessen ist die Massematrix auch zeitabhängig. Dabei folgen wir ähnlichen Notationen und Definitionen.

3.1 Problemstellung

Das Ziel dieser Arbeit ist die Untersuchung parabolischer Differentialgleichungen in zeitabhängigen Gebieten. Ebenso wie bei Problemen in festen Gebieten erhält man hier die Ortsdiskretisierung der Gleichungen, indem man Standardverfahren einsetzt, wie Differenzenverfahren oder Finite-Elemente-Methoden. Dadurch erhält man ein Anfangswertproblem für ein System gewöhnlicher Differentialgleichungen. Abgesehen davon, dass die Steifigkeitsmatrix bei bewegtem Gitter zusätzliche Terme enthält, die von der Gittergeschwindigkeit abhängig sind, ist die Massematrix in diesem Fall zeitabhängig.

Für die Herleitung der Galerkin-Verfahren betrachten wir die Anfangswertaufgabe:

$$\begin{aligned} \partial_t (\mathbf{M}(t)\mathbf{u}(t)) &= \mathbf{F}(t, \mathbf{u}(t)) \quad \forall t \in (0, T), \\ \mathbf{u}(0) &= \mathbf{u}_0, \end{aligned} \tag{3.1}$$

wobei $\mathbf{u}(t) = (u_j(t)) \in \mathcal{V}$ mit $\mathcal{V} := \mathbb{R}^m$. \mathbf{u}_0 ist der Anfangswert an der Zeit $t = 0$. \mathbf{M} und \mathbf{F} sind jeweils die aus einer Ortsdiskretisierung gelieferten zeitabhängigen Massematrix und die rechte Seite. Mit ∂_t bezeichnen wir die gewöhnliche Ableitung in der Zeit. Ausnahmsweise wird diese für die Basisfunktionen durch einen Strich ersetzt. Die Funktion $\mathbf{F} : [0, T] \times \mathcal{V} \rightarrow \mathcal{V}$ sei hinreichend glatt und kann nichtlinear sein. In den weiteren Kapiteln wird sich herausstellen, dass wir genau diese Art von Problemen erhalten werden, wenn wir Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten betrachten, deren schwache Formulierung im ALE-Rahmen herleiten und im Ort diskretisieren.

Um die Ergebnisse aus [29] auf zeitabhängige Gebiete verallgemeinern zu können, schreiben wir das Problem (3.1) folgendermaßen um:

$$\begin{aligned} \partial_t \mathbf{v}(t) &= \tilde{\mathbf{F}}(t, \mathbf{v}(t)) \quad \forall t \in (0, T), \\ \mathbf{v}(0) &= \mathbf{v}_0, \end{aligned} \tag{3.2}$$

wobei $\mathbf{v}(t) := \mathbf{M}(t)\mathbf{u}(t)$. $\mathbf{v}_0 = \mathbf{M}(0)\mathbf{u}_0$ ist der Anfangswert an der Zeit $t = 0$ und die neue rechte Seite ergibt sich mit $\tilde{\mathbf{F}}(t, \mathbf{v}(t)) := \mathbf{F}(t, \mathbf{u}(t))$ für alle $t \in I$.

Im Weiteren betrachten wir die in Kapitel 2 eingeführte Zeitdiskretisierung, den Raum $\mathbb{P}_k(I_n, \mathcal{V})$ der vektorwertigen Polynome von der Ordnung k in der Zeit über dem Intervall I_n und die Referenztransformation T_n . Unser Ziel ist die Approximation der exakten Lösung $\mathbf{u} : I \rightarrow \mathcal{V}$ mit einer Funktion $\mathbf{u}_\tau : I \rightarrow \mathcal{V}$, die stückweise aus Polynomen höchstens k -ten Grades zusammengesetzt ist. Dazu werden wir sowohl stetige als auch unstetige Verfahren betrachten. Diese unterscheiden sich grundsätzlich im betrachteten Ansatz- und Testraum. Dazu definieren wir den unstetigen zeitdiskreten Raum

$$\mathbb{P}_k^d(I, \mathcal{V}) := \left\{ \mathbf{w} \in L^2(I, \mathcal{V}) : \mathbf{w}|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V}) \quad \forall I_n \in \mathcal{M}_\tau \right\},$$

den stetigen zeitdiskreten Raum

$$\mathbb{P}_k^c(I, \mathcal{V}) := \left\{ \mathbf{w} \in C(I, \mathcal{V}) : \mathbf{w}|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V}) \quad \forall I_n \in \mathcal{M}_\tau \right\}$$

und den C^1 -stetigen zeitdiskreten Raum

$$\mathbb{P}_k^{C^1}(I, \mathcal{V}) := \left\{ \mathbf{w} \in C^1(I, \mathcal{V}) : \mathbf{w}|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V}) \quad \forall I_n \in \mathcal{M}_\tau \right\}.$$

In Bezug auf die nicht stetigen Funktionen aus $\mathbb{P}_k^d(I, \mathcal{V})$ definieren wir an den Zeitpunkten $t_n, n = 0, \dots, N$, die folgenden Werte:

$$\mathbf{v}_n^- := \lim_{t \nearrow t_n} \mathbf{v}(t), \quad \mathbf{v}_n^+ := \lim_{t \searrow t_n} \mathbf{v}(t), \quad [v]_n := v_n^+ - v_n^-, \quad (3.3)$$

dabei sind \mathbf{v}_n^- , \mathbf{v}_n^+ und $[v]_n$ der links-, der rechtsseitige Wert und der Sprung von \mathbf{v} bei t_n . Desweiteren definieren wir den Funktionswert $\mathbf{v}(t_n)$ für $n \geq 1$ als den Wert aus dem Teilintervall $I_n = (t_{n-1}, t_n]$, d.h. durch $\mathbf{v}(t_n) := \mathbf{v}_n^-$, während für $\mathbf{v}(0)$ der Startwert aufgenommen wird, d.h. $\mathbf{v}(0) := \mathbf{v}_0$. Als Zugang zu unseren diskreten Galerkin-Methoden führen wir die folgende Aufgabenstellung ein:

Finde $\mathbf{v} \in C^1(I, \mathcal{V})$, sodass $\mathbf{v}(t_0) = \mathbf{v}_0$ und

$$\int_0^T (\partial_t \mathbf{v}(t) - \tilde{\mathbf{F}}(t, \mathbf{v}(t)), \mathbf{w}(t)) dt = 0 \quad \forall \mathbf{w} \in C(I, \mathcal{V}). \quad (3.4)$$

Diese Formulierung erhalten wir durch Multiplikation der Gleichung (3.2) mit einer Testfunktion \mathbf{w} und Integration über das gesamte Zeitintervall. Diese variationelle Formulierung ist äquivalent zum Problem (3.2). Dies lässt sich mit einfachen Argumenten zeigen, siehe dazu [34].

3.2 Stetige Galerkin–Petrov-Verfahren

Der Ausgangspunkt für die stetigen Galerkin–Petrov-Verfahren ist die Benutzung des Raumes $\mathbb{P}_k^c(I, \mathcal{V})$ als Ansatzraum und des Testraumes $\mathbb{P}_{k-1}^d(I, \mathcal{V})$. Das „**globale cGP(k)-Verfahren**“ lautet:

Finde $\mathbf{v}_\tau \in \mathbb{P}_k^c(I, \mathcal{V})$, sodass $\mathbf{v}_\tau(0) = \mathbf{v}_0$ und

$$\int_0^T (\partial_t \mathbf{v}_\tau(t), \mathbf{w}_\tau(t)) dt = \int_0^T (\tilde{\mathbf{F}}(t, \mathbf{v}_\tau(t)), \mathbf{w}_\tau(t)) dt \quad \forall \mathbf{w}_\tau \in \mathbb{P}_{k-1}^d(I, \mathcal{V}). \quad (3.5)$$

Die Testfunktionen in $\mathbb{P}_{k-1}^d(I, \mathcal{V})$ dürfen unstetig sein und das globale Problem (3.5) lässt sich in eine Folge von zeitlokalen Problemen auf den einzelnen Zeitintervallen $I_n, n = 1, \dots, N$, überführen. Dementsprechend wird (3.5) zu einem sukzessiven Zeitschritt-Verfahren umgewandelt.

In diesem Rahmen wählen wir die Testfunktionen $\mathbf{w}_\tau(t) := \mathbf{w}\psi(t)$ mit einem zeitunabhängigen Vektor $\mathbf{w} \in \mathcal{V}$. Die skalare Funktion $\psi : I \rightarrow \mathbb{R}$ ist ein Polynom vom Grad kleiner oder gleich $k - 1$ auf I_n und außerhalb von I_n ist sie 0. Auf diese Weise ergibt sich das lokal diskretisierte Problem in der Zeit:

Finde $\mathbf{v}_\tau|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V})$, sodass $\mathbf{v}_\tau|_{I_n}(t_{n-1}) = \mathbf{v}_{n-1}^-$ und

$$\int_{I_n} (\partial_t \mathbf{v}_\tau(t), \mathbf{w}) \psi(t) dt = \int_{I_n} (\tilde{\mathbf{F}}(t, \mathbf{v}_\tau(t)), \mathbf{w}) \psi(t) dt \quad \forall \mathbf{w} \in \mathcal{V} \quad \forall \psi \in \mathbb{P}_{k-1}(I_n), \quad (3.6)$$

wobei $\mathbf{v}_{n-1}^- := \mathbf{v}_\tau|_{I_{n-1}}(t_{n-1})$ für alle $n \geq 2$ und $\mathbf{v}_{n-1}^- := \mathbf{v}_0$ für $n = 1$.

Im Allgemeinen, d.h. bei einer nichtlinearen rechten Seite, muss die Funktion $\tilde{\mathbf{F}}$ numerisch integriert werden. Unter der Verwendung einer $(k+1)$ -punktigen Gauß–Lobatto-Quadraturformel können Polynome vom Grad kleiner oder gleich $2k-1$ exakt integriert werden. Seien $\hat{\omega}_j$ und $\hat{t}_j, j=0, \dots, k$, die dazugehörigen Gewichte und Knoten auf dem Referenzintervall $(-1, 1]$. Das numerisch integrierte „lokale cGP(k)-Problem“ lautet:

Finde $\mathbf{v}_\tau|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V})$, sodass

$$\begin{aligned} \mathbf{v}_\tau(t_{n-1}) &= \mathbf{v}_{n-1}, \\ \sum_{j=0}^k \hat{\omega}_j \partial_t \mathbf{v}_\tau(t_{n,j}) \psi(t_{n,j}) &= \sum_{j=0}^k \hat{\omega}_j \tilde{\mathbf{F}}(t_{n,j}, \mathbf{v}_\tau(t_{n,j})) \psi(t_{n,j}) \quad \forall \psi \in \mathbb{P}_{k-1}(I_n). \end{aligned} \quad (3.7)$$

Nun betrachten wir die spezielle Darstellung von \mathbf{v}_τ auf I_n :

$$\mathbf{v}_\tau(t) := \sum_{j=0}^k \mathbf{V}_n^j \phi_{n,j}(t) \quad \forall t \in I_n, \quad (3.8)$$

wobei die Koeffizienten \mathbf{V}_n^j in $\mathcal{V} = \mathbb{R}^m$ enthalten sind. Mit $\phi_{n,j} \in \mathbb{P}_k(I_n, \mathbb{R}^m)$ bezeichnen wir die Lagrange-Basis-Funktionen bezüglich der $k+1$ Knoten der vorher betrachteten Gauß–Lobatto-Formel auf \bar{I}_n . Diesbezüglich ist $t_{n,0} = t_{n-1}$ und $t_{n,k} = t_n$. Demzufolge ist die Anfangsbedingung in (3.7) äquivalent zu

$$\mathbf{V}_n^0 = \begin{cases} \mathbf{v}_0, & \text{falls } n = 1, \\ \mathbf{v}_\tau|_{I_{n-1}}(t_{n-1}), & \text{falls } n \geq 2. \end{cases}$$

Ein unmittelbarer Vorteil der betrachteten Lagrange-Basis-Funktionen ist die Eigenschaft $\mathbf{V}_n^j = \mathbf{v}_\tau(t_{n,j})$. Wir gehen wie in [29] vor und erhalten durch eine affine Referenztransformation aus Abschnitt 2.6 $T_n : \hat{I} \rightarrow I_n, t = T_n(\hat{t})$ die Basisfunktionen $\hat{\phi}_j \in \mathbb{P}_k(\hat{I}), j=0, \dots, k$, die sich an den Quadraturknoten $\hat{t}_j, j=0, \dots, k$, ergeben. Es gilt

$$\phi_{n,j}(t) := \hat{\phi}_j(\hat{t}) \quad \text{mit } \hat{t} := T_n^{-1}(t) = \frac{2}{\tau_n} \left(t - \frac{t_n + t_{n-1}}{2} \right) \in \hat{I}.$$

In Bezug auf die Gauß–Lobatto-Punkte, mit Ausnahme von dem linken Punkt, definieren wir analog hierzu die Testbasisfunktionen $\psi_{n,i}$ anhand der Referenzbasisfunktionen $\hat{\psi}_i \in \mathbb{P}_{k-1}(\hat{I})$ durch

$$\psi_{n,i}(t) := \hat{\psi}_i(T_n^{-1}(t)) \quad \forall t \in I_n, i = 1, \dots, k,$$

dabei sind die Testfunktionen $\psi_{n,i} \in \mathbb{P}_{k-1}(I_n)$ so gewählt, dass die Beziehung

$$\hat{\psi}_i(\hat{t}_\mu) := (\hat{\omega}_\mu)^{-1} \delta_{i,\mu}$$

für alle $i, \mu = 1, \dots, k$ erfüllt ist.

Aus der Definition

$$\mathbf{u}_\tau(t) = \mathbf{M}^{-1}(t) \mathbf{v}_\tau(t) \quad (3.9)$$

erhält man die diskrete Lösung \mathbf{u}_τ des ursprünglichen Problems (3.1) in der Form einer Funktion, die im allgemeinen Fall kein Polynom mehr ist. Alternativ könnte man sich überlegen, \mathbf{u}_τ durch die Koeffizienten gemäß

$$\mathbf{V}_n^j = \mathbf{M}(t_{n,j}) \mathbf{U}_n^j$$

zu definieren, wobei hier die Koeffizienten \mathbf{U}_n^j in $\mathcal{V} = \mathbb{R}^m$ enthalten sind. Dies führt zu einer anderen Funktion \mathbf{u}_τ der Form

$$\mathbf{u}_\tau(t) := \sum_{j=0}^k \mathbf{U}_n^j \phi_{n,j}(t) \quad \forall t \in I_n. \quad (3.10)$$

Ausgehend von der Darstellung (3.8) ist die Zeitableitung von \mathbf{v}_τ auf I_n gegeben durch

$$\partial_t \mathbf{v}_\tau(t) := \sum_{j=0}^k \mathbf{V}_n^j \phi'_{n,j}(t) \quad \forall t \in I_n. \quad (3.11)$$

Infolgedessen erhalten wir für jede Basisfunktion $\psi \in \mathbb{P}_{k-1}(I_n)$ und alle $\mathbf{w} \in \mathcal{V}$ die Formulierung

$$\int_0^T (\partial_t \mathbf{v}_\tau(t), \mathbf{w}) \psi(t) dt = \sum_{\mu=0}^k \hat{\omega}_\mu \sum_{j=0}^k (\mathbf{V}_n^j, \mathbf{w}) \hat{\phi}'_j(\hat{t}_\mu) \hat{\psi}(\hat{t}_\mu).$$

Abschließend erhalten wir auf diese Weise eine „**spezielle Form des numerisch integrierten cGP(k)-Verfahrens**“:

Finde die Koeffizienten $\mathbf{V}_n^j \in \mathcal{V}$, $j = 1, \dots, k$, als Lösung des nichtlinearen $(k \times k)$ -Block-Systems:

$$\sum_{j=0}^k \alpha_{i,j} \mathbf{V}_n^j = \frac{\tau_n}{2} \left\{ \tilde{\mathbf{F}}(t_{n,i}, \mathbf{V}_n^i) + \beta_i \tilde{\mathbf{F}}(t_{n,0}, \mathbf{V}_n^0) \right\}, \quad i = 1, \dots, k,$$

wobei $\mathbf{V}_n^0 := \mathbf{v}_{n-1}^-$ der Startwert ist. Die Koeffizienten $\alpha_{i,j}$ bzw. β_i definieren sich wie in [29] durch:

$$\alpha_{i,j} := \hat{\phi}'_j(\hat{t}_i) + \beta_i \hat{\phi}'_j(\hat{t}_0), \quad \beta_i := \hat{\omega}_0 \hat{\psi}_i(\hat{t}_0), \quad i = 1, \dots, k, \quad j = 0, \dots, k. \quad (3.12)$$

Man beachte, dass diese Koeffizienten von τ_n unabhängig sind. Durch die Übertragung auf das ursprüngliche Problem (3.1) ist dies äquivalent zu

Finde die Koeffizienten $\mathbf{U}_n^j \in \mathcal{V}$, $j = 1, \dots, k$, als Lösung des nichtlinearen $(k \times k)$ -Block-Systems:

$$\sum_{j=0}^k \alpha_{i,j} \mathbf{M}(t_{n,j}) \mathbf{U}_n^j = \frac{\tau_n}{2} \left\{ \mathbf{F}(t_{n,i}, \mathbf{U}_n^i) + \beta_i \mathbf{F}(t_{n,0}, \mathbf{U}_n^0) \right\}, \quad i = 1, \dots, k, \quad (3.13)$$

wobei $\mathbf{U}_n^0 := \mathbf{u}_{n-1}^-$ der Startwert ist. Für eine bessere Verständlichkeit der erhaltenen speziellen Formulierung (3.13), präsentieren wir nun die cGP(k)-Verfahren für die Fälle $k = 1$ und $k = 2$.

3.2.1 cGP(1)-Verfahren

Die Anwendung der 2-punktigen Gauß–Lobatto-Formel mit den Stützstellen $t_{n,0} = t_{n-1}$, $t_{n,1} = t_n$ und den Referenzgewichten $\hat{\omega}_0 = \hat{\omega}_1 = 1$ liefert die bekannte Trapezregel

$$\mathbf{M}(t_n) \mathbf{U}_n^1 - \mathbf{M}(t_{n-1}) \mathbf{U}_n^0 = \frac{\tau_n}{2} \left\{ \mathbf{F}(t_n, \mathbf{U}_n^1) + \mathbf{F}(t_{n-1}, \mathbf{U}_n^0) \right\},$$

wobei sich die Lösung durch $u_\tau(t_n) = \mathbf{U}_n^1$ ergibt.

3.2.2 cGP(2)-Verfahren

Nun verwenden wir die 3-punktige Gauß–Lobatto-Formel mit den Stützstellen $t_{n,0} = t_{n-1}$, $t_{n,1} = (t_{n-1} + t_n)/2$, $t_{n,2} = t_n$ und den Referenzgewichten $\hat{\omega}_0 = \hat{\omega}_2 = 1/3$, $\hat{\omega}_1 = 4/3$. Hier erhalten wir die bekannte Simpsonregel

$$\begin{aligned} \mathbf{M}(t_{n,1}) \mathbf{U}_n^1 + \frac{1}{4} \mathbf{M}(t_{n,2}) \mathbf{U}_n^2 &= \frac{5}{4} \mathbf{M}(t_{n,0}) \mathbf{U}_n^0 + \frac{\tau_n}{2} \left\{ \mathbf{F}(t_{n,1}, \mathbf{U}_n^1) + \frac{1}{2} \mathbf{F}(t_{n,0}, \mathbf{U}_n^0) \right\}, \\ -4 \mathbf{M}(t_{n,1}) \mathbf{U}_n^1 + 2 \mathbf{M}(t_{n,2}) \mathbf{U}_n^2 &= -2 \mathbf{M}(t_{n,0}) \mathbf{U}_n^0 + \frac{\tau_n}{2} \left\{ \mathbf{F}(t_{n,2}, \mathbf{U}_n^2) - \mathbf{F}(t_{n,0}, \mathbf{U}_n^0) \right\} \end{aligned}$$

auf dem Zeitintervall $I_n = (t_{n-1}, t_n]$, wobei $\mathbf{u}_\tau(t_{n,j}) = \mathbf{U}_n^j$, $j = 1, 2$.

3.3 Die cGP- $C^1(k)$ -Verfahren

Das Ziel in diesem Abschnitt ist eine C^1 -stetige Lösung \mathbf{u}_τ für das ursprüngliche Problem (3.1) bzw. \mathbf{v}_τ für das äquivalente Problem (3.2) als Polynom vom Grad k mit $k \geq 3$ zu erhalten. Dazu benötigen wir auf jedem Teilintervall $k + 1$ Bedingungen. Beispielsweise sind \mathbf{v}_τ als Lösung der Differentialgleichung (3.2) auf I_1 und ihre Zeitableitung bei $t_0 = 0$ durch $\mathbf{v}_\tau(0) = \mathbf{v}_0$ und

$$\partial_t \mathbf{v}_\tau(t_0) = \tilde{\mathbf{F}}(t_0, \mathbf{v}_\tau(t_0)) = \tilde{\mathbf{F}}(t_0, \mathbf{v}_0)$$

bekannt. Eine weitere quasi-explizite Bedingung, die $\mathbf{v}_1 := \mathbf{v}_\tau(t_1)$ und die Zeitableitung $\partial_t \mathbf{v}_\tau(t_1)$ koppelt, ist gegeben durch

$$\partial_t \mathbf{v}_\tau(t_1) = \tilde{\mathbf{F}}(t_1, \mathbf{v}_\tau(t_1)).$$

Demzufolge verbleiben nur noch $k - 2$ Unbekannten, die implizit gekoppelt sind.

Das „cGP- $C^1(k)$ -Verfahren“ für $k \geq 3$ lautet :

Finde $\mathbf{v}_\tau \in \mathbb{P}_k^{C^1}(I, \mathcal{V})$, sodass $\mathbf{v}_\tau(0) = \mathbf{v}_0$ und

$$\int_0^T (\partial_t \mathbf{v}_\tau(t), \mathbf{w}_\tau(t)) dt = \int_0^T (\tilde{\mathbf{F}}(t, \mathbf{v}_\tau(t)), \mathbf{w}_\tau(t)) dt \quad \forall \mathbf{w}_\tau \in \mathbb{P}_{k-3}^d(I, \mathcal{V}),$$

$$\partial_t \mathbf{v}_\tau(t_n) = \tilde{\mathbf{F}}(t_n, \mathbf{v}_\tau(t_n)), \quad n = 0, \dots, N.$$

Analog zum cGP(k)-Verfahren ergibt sich das lokale und in der Zeit diskrete Problem:

Finde $\mathbf{v}_\tau|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V})$, sodass

$$\begin{aligned} \mathbf{v}_\tau(t_{n-1}) &= \mathbf{v}_{n-1}, \\ \partial_t \mathbf{v}_\tau(t_{n-1}) &= \tilde{\mathbf{F}}(t_{n-1}, \mathbf{v}_\tau(t_{n-1})), \\ \partial_t \mathbf{v}_\tau(t_n) &= \tilde{\mathbf{F}}(t_n, \mathbf{v}_\tau(t_n)), \\ \int_{I_n} (\partial_t \mathbf{v}_\tau(t), \mathbf{w}) \psi(t) dt &= \int_{I_n} (\tilde{\mathbf{F}}(t, \mathbf{v}_\tau(t)), \mathbf{w}) \psi(t) dt \quad \forall \psi \in \mathbb{P}_{k-3}(I_n) \quad \forall \mathbf{w} \in \mathcal{V}. \end{aligned}$$

Die Anwendung der k -punktigen Gauß-Lobatto-Quadratur, die Polynome vom Grad kleiner oder gleich $2k - 3$ exakt integriert, liefert das „numerisch integrierte cGP- $C^1(k)$ -Verfahren“ :

Finde $\mathbf{v}_\tau|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V})$, sodass

$$\begin{aligned} \mathbf{v}_\tau(t_{n-1}) &= \mathbf{v}_{n-1}, \\ \partial_t \mathbf{v}_\tau(t_{n-1}) &= \tilde{\mathbf{F}}(t_{n-1}, \mathbf{v}_\tau(t_{n-1})), \\ \partial_t \mathbf{v}_\tau(t_n) &= \tilde{\mathbf{F}}(t_n, \mathbf{u}_\tau(t_n)), \\ \sum_{j=0}^{k-1} \hat{\omega}_j (\partial_t \mathbf{v}_\tau(t_{n,j}), \mathbf{w}) \psi(t_{n,j}) &= \sum_{j=0}^k \hat{\omega}_j (\tilde{\mathbf{F}}(t_{n,j}, \mathbf{v}_\tau(t_{n,j})), \mathbf{w}) \psi(t_{n,j}) \quad \forall \psi \in \mathbb{P}_{k-3}(I_n). \end{aligned} \quad (3.14)$$

Aufgrund der Bedingung $\mathbf{v}_\tau|_{I_{n+1}}(t_n) = \mathbf{v}_\tau|_{I_n}(t_n) =: \mathbf{v}_\tau(t_n)$, erhalten wir aus 3.14

$$\partial_t^- \mathbf{v}_\tau(t_n) = \partial_t \mathbf{v}_\tau|_{I_n}(t_n) = \tilde{\mathbf{F}}(t_n, \mathbf{v}_\tau(t_n)) = \partial_t \mathbf{v}_\tau|_{I_{n+1}}(t_n) = \partial_t^+ \mathbf{v}_\tau(t_n).$$

Darausfolgend ist die Lösung \mathbf{v}_τ C^1 -stetig.

Die Koeffizienten \mathbf{U}_n^j ergeben sich in der Form

$$\mathbf{U}_n^j = \mathbf{M}(t_{n,j})^{-1} \mathbf{V}_n^j, \quad j = 1, \dots, k, \quad n = 1, \dots, N.$$

Die Funktion \mathbf{u}_τ wird somit durch die Darstellung (3.10) geliefert. Eine weitere Möglichkeit wäre die direkte Anwendung der Definition (3.9).

3.4 Unstetige Galerkin-Verfahren

Im Gegensatz zu den cGP-Verfahren entspricht hier der Testraum dem Ansatzraum $\mathbb{P}_k^d(I, \mathcal{V})$. Das globale „**dG(k)-Verfahren**“ lautet:

Finde $\mathbf{v}_\tau \in \mathbb{P}_k^d(I, \mathcal{V})$, sodass $\mathbf{v}_\tau(0) = \mathbf{v}_0^- = \mathbf{v}_0$ und

$$\begin{aligned} \sum_{n=1}^N \int_{I_n} (\partial_t \mathbf{v}_\tau(t), \mathbf{w}_\tau(t)) dt + \sum_{n=1}^N ([\mathbf{v}_\tau]_{n-1}, \mathbf{w}_{n-1}^+) \\ = \int_0^T (\tilde{\mathbf{F}}(t, \mathbf{v}_\tau(t)), \mathbf{w}_\tau(t)) dt \quad \forall \mathbf{w}_\tau \in \mathbb{P}_k^d(I, \mathcal{V}). \end{aligned} \quad (3.15)$$

Für eine zeitliche Aufspaltung des Problems betrachten wir die Testfunktionen $\mathbf{w}_\tau(t) = \mathbf{w}\psi(t)$. Dabei ist $\mathbf{w} \in \mathcal{V}$ beliebig und zeitunabhängig. $\psi(t)$ sei eine skalare Basisfunktion aus $\mathbb{P}_k(I_n)$, die man außerhalb \bar{I}_n durch 0 erweitern kann. Anhand der bekannten Werte $\mathbf{V}_n^0 := \mathbf{v}_{n-1}^-$ lässt sich die Lösung auf jedem Teilintervall I_n nacheinander bestimmen. Das „**lokale dG(k)-Verfahren**“ lautet:

Finde $\mathbf{v}_\tau|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V})$, sodass

$$\begin{aligned} \int_{I_n} (\partial_t \mathbf{v}_\tau(t), \mathbf{w}) \psi(t) dt + ([\mathbf{v}_\tau]_{n-1}, \mathbf{w}_{n-1}^+) \psi(t_{n-1}) \\ = \int_{I_n} (\tilde{\mathbf{F}}(t, \mathbf{v}_\tau(t), \mathbf{w}) \psi(t) dt \quad \forall \mathbf{w} \in \mathcal{V} \quad \forall \psi \in \mathbb{P}_k(I_n). \end{aligned} \quad (3.16)$$

Analog zum cGP-Verfahren definieren wir die diskrete Lösung \mathbf{v}_τ auf I_n durch

$$\mathbf{v}_\tau(t) := \sum_{j=1}^{k+1} \mathbf{V}_n^j \phi_{n,j}(t) \quad \forall t \in I_n, \quad (3.17)$$

wobei die Koeffizienten \mathbf{V}_n^j in $\mathcal{V} = \mathbb{R}^m$ enthalten sind. Hier bezeichnen wir mit $\phi_{n,j} \in \mathbb{P}_k(I_n, \mathbb{R}^m)$ die Lagrange-Basis-Funktionen bezüglich der $k+1$ Knoten der in Kapitel 2 betrachteten rechtsseitigen Gauß–Radau-Formel auf I_n . Diese hat den Genauigkeitsgrad $2k$. Es gilt $\phi_{n,j}(t_{n,i}) = \delta_{i,j}$ für alle $i, j = 1, \dots, k+1$. Die Anwendung der Quadraturformel auf das Problem (3.16) auf I_n liefert das „**lokale numerisch integrierte dG(k)-Verfahren**“:

Finde $\mathbf{v}_\tau|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V})$, sodass

$$\begin{aligned} \frac{\tau_n}{2} \sum_{j=1}^{k+1} \hat{\omega}_j \partial_t \mathbf{v}_\tau(t_{n,j}) \psi(t_{n,j}) + (\mathbf{v}_\tau(t_{n-1}) - \mathbf{v}_{n-1}^-) \psi(t_{n-1}) \\ = \frac{\tau_n}{2} \sum_{j=1}^{k+1} \hat{\omega}_j \tilde{\mathbf{F}}(t_{n,j}, \mathbf{v}_\tau(t_{n,j})) \psi(t_{n,j}) \quad \forall \psi \in \mathbb{P}_k(I_n). \end{aligned} \quad (3.18)$$

Die diskrete Lösung des ursprünglichen Problems ist, wie auch im cGP-Fall, durch $\mathbf{U}_n^j = \mathbf{M}(t_{n,j})^{-1} \mathbf{V}_n^j$, $j = 1, \dots, k+1$, $n = 1, \dots, N$, zu erhalten. Die Testfunktionen $\psi_{n,i}$ sind mittels der Referenzfunktionen mit $\hat{\psi}_i := \hat{\omega}_i^{-1} \hat{\phi}_i$, $i = 1, \dots, k+1$, definiert, wobei $\hat{\omega}_i$ die Gewichte der Gauß–Radau-Formel auf \hat{I} sind. $\hat{\phi}_i$ sind Basisfunktionen, die mittels der affinen Referenztransformation T_n durch $\phi_{n,i}(t) := \hat{\phi}_i(\hat{t})$, $i = 1, \dots, k+1$, erhalten werden. Wir setzen (3.17) in das Problem (3.18) ein und transformieren die Integrale auf das Referenzintervall \hat{I} . Es entsteht eine „**spezielle Form des lokalen numerisch integrierten dG(k)-Verfahrens**“:

Finde die Koeffizienten $\mathbf{V}_n^j \in \mathcal{V}$, $j = 1, \dots, k+1$, sodass

$$\sum_{j=1}^{k+1} \alpha_{i,j} \mathbf{V}_n^j = d_i \mathbf{V}_n^0 + \frac{\tau_n}{2} \tilde{\mathbf{F}}(t_{n,i}, \mathbf{V}_n^i), \quad i = 1, \dots, k+1, \quad (3.19)$$

wobei $\alpha_{i,j} := \hat{\phi}'_j(\hat{t}_j) + c_j d_i$, $c_j := \hat{\phi}'_j(-1)$, $d_i := (\hat{\omega}_i)^{-1} \hat{\phi}'_i(-1)$ und $\mathbf{V}_n^0 := \mathbf{v}_{n-1}^-$, vgl. [29]. Auch hier sind die Koeffizienten $\alpha_{i,j}$, c_j und d_i von τ_n unabhängig.

Durch die Übertragung auf das ursprüngliche Problem (3.1) ist dies äquivalent zu

Finde die Koeffizienten $\mathbf{U}_n^j \in \mathcal{V}$, $j = 1, \dots, k+1$, sodass

$$\sum_{j=1}^{k+1} \alpha_{i,j} \mathbf{M}(t_{n,j}) \mathbf{U}_n^j = d_i \mathbf{M}(t_{n-1}) \mathbf{U}_n^0 + \frac{\tau_n}{2} \mathbf{F}(t_{n,i}, \mathbf{U}_n^i), \quad i = 1, \dots, k+1, \quad (3.20)$$

wobei $\mathbf{U}_n^0 := \mathbf{u}_{n-1}^-$.

Nach dem Lösen des Problems (3.20) auf I_n wird das Problem auf $I_{n+1} = (t_n, t_{n+1}]$ auf die gleiche Weise gelöst. Als Startwert setzt man nun $\mathbf{U}_{n+1}^0 := \mathbf{u}_n^-$. Diesen haben wir bereits auf I_n berechnet. Diesen Vorgang wiederholen wir bis zum letzten Teilintervall. Wir erhalten die Lösung \mathbf{u}_τ auf dem gesamten Zeitintervall I . Im Folgenden werden wir die Schemata für die numerisch integrierten dG(0)- und dG(1)-Verfahren präsentieren.

3.4.1 dG(0)-Verfahren

Die Anwendung der 1-punktigen Gauß–Radau-Formel mit der Stützstelle $t_{n,1} = t_n$ und dem Referenzgewicht $\hat{\omega}_1 = 2$ liefert das bekannte implizite Euler-Verfahren

$$\mathbf{M}(t_n) \mathbf{U}_n^1 - \mathbf{M}(t_{n-1}) \mathbf{U}_n^0 = \tau_n \mathbf{F}(t_n, \mathbf{U}_n^1),$$

wobei wir die diskrete Lösung durch $\mathbf{U}_n^1 = \mathbf{u}_\tau(t_n) \in \mathcal{V}$ erhalten.

3.4.2 dG(1)-Verfahren

In dem dG(1)-Verfahren verwenden wir die 2-punktige Gauß–Radau-Formel mit den Stützstellen $t_{n,1} = (t_{n-1} + t_n)/3$, $t_{n,2} = t_n$ und den Referenzgewichten $\hat{\omega}_1 = 3/2$, $\hat{\omega}_2 = 1/2$. Hierbei erhalten wir das folgende (2×2) -Block-System:

$$\begin{aligned} \frac{3}{4} \mathbf{M}(t_{n,1}) \mathbf{U}_n^1 + \frac{1}{4} \mathbf{M}(t_{n,2}) \mathbf{U}_n^2 &= \mathbf{M}(t_{n,0}) \mathbf{U}_n^0 + \frac{\tau_n}{2} \mathbf{F}(t_{n,1}, \mathbf{U}_n^1), \\ -\frac{9}{4} \mathbf{M}(t_{n,1}) \mathbf{U}_n^1 + \frac{5}{4} \mathbf{M}(t_{n,2}) \mathbf{U}_n^2 &= -\mathbf{M}(t_{n,0}) \mathbf{U}_n^0 + \frac{\tau_n}{2} \mathbf{F}(t_{n,2}, \mathbf{U}_n^2) \end{aligned}$$

auf dem Zeitintervall $I_n = (t_{n-1}, t_n]$, wobei $\mathbf{u}_\tau(t_{n,j}) = \mathbf{U}_n^j$, $j = 1, 2$.

3.5 Die dG- $C^0(k)$ -Verfahren

Analog zu dem Übergang vom cGP(k)-Verfahren zum cGP- $C^1(k)$ -Verfahren werden wir hier die dG- $C^0(k)$ -Verfahren konstruieren. Wir werden die variationelle Formulierung der dG(k)-Verfahren übernehmen und machen die diskrete Lösung um eine Stufe glatter, sodass im Zeitprozess weitere Informationen der Lösung aus dem vorherigen Zeitintervall in das neue übertragen werden. Dabei wurde eine Senkung der Rechenkosten beachtet. Dies bedeutet in unserem Fall, dass wir die unstetige Lösung in den dG-Verfahren durch eine stetige Lösung im dG- $C^0(k)$ -Verfahren ersetzen.

Das „globale dG- $C^0(k)$ -Verfahren“ für das modifizierte Problem (3.2) für $k \geq 2$ lautet:

Finde $\mathbf{v}_\tau \in \mathbb{P}_k^c(I, \mathcal{V})$, sodass $\mathbf{v}_\tau(0) = \mathbf{v}_0$ und

$$\int_0^T (\partial_t \mathbf{v}_\tau(t), \mathbf{w}_\tau(t)) dt = \int_0^T (\tilde{\mathbf{F}}(t, \mathbf{v}_\tau(t)), \mathbf{w}_\tau(t)) dt \quad \forall \mathbf{w}_\tau \in \mathbb{P}_{k-2}^d(I, \mathcal{V})$$

$$\partial_t \mathbf{v}_n^- = \tilde{\mathbf{F}}(t_n, \mathbf{v}_n^-), \quad n = 1, \dots, N,$$

wobei $\partial_t \mathbf{v}_n^- = \lim_{t \nearrow t_n} \partial_t \mathbf{v}_\tau|_{I_n}(t)$.

Die Unstetigkeit des Testraumes ermöglicht uns das Problem in Teilprobleme auf den Teilintervallen I_n aufzuspalten. Somit erhalten wir das „lokale dG- $C^0(k)$ -Verfahren“ :

Finde $\mathbf{v}_\tau|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V})$, sodass

$$\mathbf{v}_\tau(t_{n-1}) = \mathbf{v}_{n-1}^-,$$

$$\partial_t \mathbf{v}_\tau(t_n) = \tilde{\mathbf{F}}(t_n, \mathbf{v}_\tau(t_n)),$$

$$\int_{I_n} (\partial_t \mathbf{v}_\tau(t), \mathbf{w}) \psi(t) dt = \int_{I_n} (\tilde{\mathbf{F}}(t, \mathbf{v}_\tau(t)), \mathbf{w}) \psi(t) dt \quad \forall \psi \in \mathbb{P}_{k-2}(I_n) \quad \forall \mathbf{w} \in \mathcal{V}.$$

Die Anwendung der rechtsseitigen k -punktigen Gauß–Radau-Quadraturformel, die Polynome vom Grad kleiner oder gleich $2k - 2$ exakt integriert, liefern das „numerisch integrierte dG- $C^0(k)$ -Verfahren“ :

Finde $\mathbf{v}_\tau|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V})$, sodass

$$\mathbf{v}_\tau(t_{n-1}) = \mathbf{v}_{n-1}^-,$$

$$\partial_t \mathbf{v}_\tau(t_n) = \tilde{\mathbf{F}}(t_n, \mathbf{v}_\tau(t_n)),$$

$$\sum_{j=1}^k \hat{\omega}_j (\partial_t \mathbf{v}_\tau(t_{n,j}), \mathbf{w}) \psi(t_{n,j}) = \sum_{j=1}^k \hat{\omega}_j (\tilde{\mathbf{F}}(t_{n,j}, \mathbf{v}_\tau(t_{n,j})), \mathbf{w}) \psi(t_{n,j}) \quad \forall \psi \in \mathbb{P}_{k-2}(I_n). \quad (3.21)$$

Wie auch bei festen Gebieten sind hier die Kosten der Lösung eines Blocksystems von cGP- $C^1(k+1)$ - bzw. dG- $C^0(k+1)$ -Verfahren vergleichbar mit denjenigen des cGP(k)- bzw. dG-(k)-Verfahrens.

Mit diesen einführenden Abschnitten haben wir die Zeit-Semidiskretisierungen höherer Ordnung mithilfe der Galerkin-Verfahren geschaffen, die in dieser Arbeit eine wesentliche Rolle spielen.

3.6 Postprocessing

3.6.1 Verbindung zwischen cGP(k) und cGP- $C^1(k+1)$

In diesem Unterabschnitt werden wir die wichtige Verbindung zwischen den Lösungen des diskreten cGP(k)- und cGP- $C^1(k+1)$ -Problems in der Arbeit [29] auch für zeitabhängige Gebiete bestätigen.

Satz 3.1. *Seien \mathbf{v}_τ^C und $\mathbf{v}_\tau^{C^1}$ jeweils die Lösungen der numerisch integrierten Varianten des diskreten cGP(k)-Verfahrens (3.7) und des cGP- $C^1(k+1)$ -Verfahrens (3.14). Es sei angenommen, dass die Zeitschritte τ_n ausreichend klein sind, sodass die Probleme beider Verfahren eine eindeutige Lösung für alle $n = 1, \dots, N$ haben. Dann gilt auf jedem Teilintervall $I_n = (t_{n-1}, t_n]$*

$$\mathbf{v}_\tau^{C^1}(t) = \mathbf{v}_\tau^C(t) + \mathbf{a}_n \zeta_n(t) \quad \forall t \in I_n,$$

wobei der Vektor $\mathbf{a}_n \in \mathcal{V}$ durch

$$\mathbf{a}_n := \mathbf{F}(t_n, \mathbf{v}_\tau^C(t_n)) - \partial_t \mathbf{v}_\tau^C(t_n)$$

gegeben ist und

$$\zeta_n(t) := \frac{\tau_n}{2} \hat{\zeta}(\hat{t}) \quad \text{mit} \quad \hat{t} := T_n^{-1}(t).$$

Dabei ist $T_n : \hat{I} \rightarrow I_n$ die affine Referenztransformation und $\hat{\zeta} \in \mathbb{P}_{k+1}(\hat{I})$ ist das Polynom, welches durch die Bedingungen $\hat{\zeta}'(1) = 1$ und $\hat{\zeta}(\hat{t}_j) = 0$ für alle $j = 0, \dots, k$ definiert ist, wobei \hat{t}_j die Integrationspunkte der $(k+1)$ -punktigen Gauß–Lobatto-Formel auf \hat{I} sind.

Beweis. Wir definieren $\mathbf{z}(t) := \mathbf{v}_\tau^C + \mathbf{a}_n \zeta_n(t)$. Da die Lösungen des cGP(k)- und cGP- $C^1(k+1)$ -Problems eindeutig sind, genügt es zu zeigen, dass \mathbf{z} die Bedingungen für die Lösung des diskreten cGP- $C^1(k+1)$ -Problems in (3.14) erfüllt. Für alle Gauß–Lobatto-Stützstellen $t_{n,j}$ gilt $\zeta_n(t_{n,j}) = 0$. Folglich ist

$$\mathbf{z}(t_{n,j}) = \mathbf{v}_\tau^C(t_{n,j}) \quad \forall j = 0, \dots, k.$$

Aus $\mathbf{v}_\tau^C(t_{n-1}) = \mathbf{v}_{n-1}$ und $t_{n,0} = t_{n-1}$ folgt $\mathbf{z}(t_{n-1}) = \mathbf{v}_{n-1}$. Nun betrachten wir die numerisch integrierten variationellen Bedingungen. Sei $\psi \in \mathbb{P}_{k-2}(I_n)$ beliebig. Dann gilt

$$\begin{aligned} & \sum_{j=0}^k \hat{\omega}_j \partial_t \mathbf{z}(t_{n,j}) \psi(t_{n,j}) \\ &= \sum_{j=0}^k \hat{\omega}_j \partial_t \mathbf{v}_\tau^C(t_{n,j}) \psi(t_{n,j}) + \sum_{j=0}^k \hat{\omega}_j \mathbf{a}_n \zeta_n'(t_{n,j}) \psi(t_{n,j}) \\ &= \sum_{j=0}^k \hat{\omega}_j \tilde{\mathbf{F}}(t_{n,j}, \mathbf{v}_\tau^C(t_{n,j})) \psi(t_{n,j}) + \frac{2}{\tau_n} \mathbf{a}_n \int_{I_n} \zeta_n'(t) \psi(t) dt, \end{aligned} \quad (3.22)$$

wobei \mathbf{v}_τ^C die Lösung des numerisch integrierten Problems (3.7) ist. In der letzten Gleichung haben wir die exakte Integrierbarkeit der Polynome vom maximalen Grad $2k-1$ mittels der Gauß–Lobatto-Formel benutzt. Eine partielle Integration des letzten Terms liefert

$$\begin{aligned} \int_{I_n} \zeta_n'(t) \psi(t) dt &= - \int_{I_n} \zeta_n(t) \psi'(t) dt + \zeta_n(t_n) \psi(t_n) - \zeta_n(t_{n-1}) \psi(t_{n-1}) \\ &= - \frac{\tau_n}{2} \sum_{j=0}^k \hat{\omega}_j \zeta_n(t_{n,j}) \psi'(t_{n,j}) = 0. \end{aligned} \quad (3.23)$$

Dabei haben wir die Definition von ζ_n verwendet. Die Kombination von (3.22) und (3.23) liefert

$$\sum_{j=0}^k \hat{\omega}_j \partial_t \mathbf{z}(t_{n,j}) \psi(t_{n,j}) = \sum_{j=0}^k \hat{\omega}_j \tilde{\mathbf{F}}(t_{n,j}, \mathbf{v}_\tau^C(t_{n,j})) \psi(t_{n,j}) = \sum_{j=0}^k \hat{\omega}_j \tilde{\mathbf{F}}(t_{n,j}, \mathbf{z}(t_{n,j})) \psi(t_{n,j}).$$

Daraus kann man schließen, dass die Funktion \mathbf{z} die variationelle Bedingung in (3.14) erfüllt. Im weiteren Verlauf werden wir zeigen, dass \mathbf{z} die Differentialgleichung in t_n erfüllen wird. Aus der Definition von \mathbf{a}_n und $\partial_t \zeta_n(t_n) = \partial_{\hat{t}} \hat{\zeta}(1) = 1$ folgt

$$\begin{aligned} \partial_t \mathbf{z}(t_n) &= \partial_t \mathbf{v}_\tau^C(t_n) + \mathbf{a}_n \zeta_n'(t_n) \\ &= \partial_t \mathbf{v}_\tau^C(t_n) + \tilde{\mathbf{F}}(t_n, \mathbf{v}_\tau^C(t_n)) - \partial_t \mathbf{v}_\tau^C(t_n) \\ &= \tilde{\mathbf{F}}(t_n, \mathbf{z}(t_n)). \end{aligned}$$

Um zu beweisen, dass \mathbf{z} die Differentialgleichung in t_{n-1} erfüllt, verwenden wir eine zusätzliche Beziehung. Zu diesem Zweck wählen wir beim diskreten cGP(k)-Problem als Testfunktion das Polynom $\varrho_n \in \mathbb{P}_{k-1}$, das in allen inneren Gauß–Lobatto-Stützstellen $t_{n,j}, j = 1, \dots, k-1$, verschwindet und welches $\varrho_n(t_n) = 1$ erfüllt. Man beachte, dass ϱ_n eine zulässige Testfunktion im

cGP(k)-Problem ist. Infolgedessen erhalten wir aus (3.7) und der Tatsache, dass alle Terme mit dem Index $j = 1, \dots, k-1$ verschwinden, die Gleichung

$$\begin{aligned} \hat{\omega}_0 \partial_t \mathbf{v}_\tau^C(t_{n-1}) \varrho_n(t_{n-1}) + \hat{\omega}_k \partial_t \mathbf{v}_\tau^C(t_n) \varrho_n(t_n) \\ = \hat{\omega}_0 \tilde{\mathbf{F}}(t_{n-1}, \mathbf{v}_\tau^C(t_{n-1})) \varrho_n(t_{n-1}) + \hat{\omega}_k \tilde{\mathbf{F}}(t_n, \mathbf{v}_\tau^C(t_n)) \varrho_n(t_n). \end{aligned}$$

Diese lässt sich mittels $\hat{\omega}_0 = \hat{\omega}_k \neq 0$ und $\varrho_n(t_n) = 1$ vereinfachen. Es gilt

$$\varrho_n(t_{n-1}) \left\{ \partial_t \mathbf{v}_\tau^C(t_{n-1}) - \tilde{\mathbf{F}}(t_{n-1}, \mathbf{v}_\tau^C(t_{n-1})) \right\} = \tilde{\mathbf{F}}(t_n, \mathbf{v}_\tau^C(t_n)) \varrho_n(t_n) - \partial_t \mathbf{v}_\tau^C(t_n) \varrho_n(t_n).$$

Aus der Definition von \mathbf{z} und der Definition von \mathbf{a}_n folgt

$$\begin{aligned} \partial_t \mathbf{z}(t_{n-1}) &= \partial_t \mathbf{v}_\tau^C(t_{n-1}) + \mathbf{a}_n \zeta'_n(t_{n-1}) \\ &= \partial_t \mathbf{v}_\tau^C(t_{n-1}) + \zeta'_n(t_{n-1}) \varrho_n(t_{n-1}) \left\{ \partial_t \mathbf{v}_\tau^C(t_{n-1}) - \tilde{\mathbf{F}}(t_{n-1}, \mathbf{v}_\tau^C(t_{n-1})) \right\}. \end{aligned} \quad (3.24)$$

Nun werden wir zeigen, dass $\zeta'_n(t_{n-1}) \varrho_n(t_{n-1}) = -1$. Aus einer partiellen Integration folgt

$$\begin{aligned} 0 &= \frac{\tau_n}{2} \sum_{j=0}^k \hat{\omega}_j \zeta_n(t_{n,j}) \varrho'_n(t_{n,j}) = \int_{I_n} \zeta_n(t) \varrho'_n(t) dt \\ &= - \int_{I_n} \zeta'_n(t) \varrho_n(t) dt + \zeta_n(t_n) \varrho_n(t_n) - \zeta_n(t_{n-1}) \varrho_n(t_{n-1}) \\ &= - \frac{\tau_n}{2} \hat{\omega}_0 \zeta'_n(t_{n-1}) \varrho_n(t_{n-1}) - \frac{\tau_n}{2} \hat{\omega}_k \zeta'_n(t_n) \varrho_n(t_n) \\ &= - \frac{\tau_n}{2} \hat{\omega}_0 \zeta'_n(t_{n-1}) \varrho_n(t_{n-1}) - \frac{\tau_n}{2} \hat{\omega}_k. \end{aligned}$$

Dabei haben wir von den Eigenschaften von ζ_n und ϱ_n Gebrauch gemacht. Weiterhin folgt aus $\hat{\omega}_0 = \hat{\omega}_k \neq 0$, dass $\zeta'_n(t_{n-1}) \varrho_n(t_{n-1}) = -1$. Abschließend kann man aus (3.24) und $\mathbf{z}(t_{n-1}) = \mathbf{v}_\tau^C(t_{n-1})$ entnehmen, dass die Differentialgleichung an dem Punkt t_{n-1} erfüllt wird, was den Beweis abschließt. \square

Da die Probleme (3.1) und (3.2) äquivalent sind, können wir dieses Ergebnis ohne Weiteres auf das ursprüngliche Problem übertragen, d.h., es gilt auf jedem Teilintervall I_n

$$\mathbf{u}_\tau^{C^1}(t) = \mathbf{u}_\tau^C(t) + \mathbf{M}^{-1}(t) \mathbf{a}_n \zeta_n(t) \quad \forall t \in I_n, \quad (3.25)$$

wobei $\mathbf{u}_\tau^C(t) := \mathbf{M}^{-1}(t) \mathbf{v}_\tau^C(t)$ und $\mathbf{u}_\tau^{C^1}(t) := \mathbf{M}^{-1}(t) \mathbf{v}_\tau^{C^1}(t)$ für alle $t \in I_n$ jeweils die Lösungen der numerisch integrierten Varianten der diskreten cGP(k)- und cGP- $C^1(k+1)$ -Verfahren für das ursprüngliche Problem (3.1) sind und der Vektor $\mathbf{a}_n \in \mathcal{V}$ durch

$$\mathbf{a}_n := \mathbf{F}(t_n, \mathbf{u}_\tau^C(t_n)) - \partial_t(\mathbf{M} \mathbf{u}_\tau^C)(t_n)$$

gegeben ist.

Im nächsten Unterabschnitt werden wir nicht nur eine Verallgemeinerung der Ergebnisse in [29] vorstellen, sondern ebenso zeigen, dass der Vektor \mathbf{a}_n für die Verbindung der diskreten dG(k)- und dG- $C^0(k+1)$ -Verfahren geeignet ist, wenn diese mittels der Lösungen der numerisch integrierten Varianten des dG-Verfahrens definiert ist.

3.6.2 Verbindung zwischen $dG(k)$ und $dG-C^0(k+1)$

Bevor wir uns mit der Übertragung des Postprocessings aus dem cGP-Fall auf die dG-Verfahren befassen, führen wir das folgende Lemma, das wir in unserer nachfolgenden Analyse benötigen werden, ein.

Lemma 3.2. *Seien \mathbf{v}_τ^{dG} und $\mathbf{v}_\tau^{C^0}$ jeweils die Lösungen der numerisch integrierten Varianten der diskreten $dG(k)$ -Verfahren (3.18) und $dG-C^0(k+1)$ -Verfahren (3.21). Den Vektor \mathbf{b}_n definieren wir durch*

$$\mathbf{b}_n := \tilde{\mathbf{F}}(t_n, \mathbf{v}_\tau^{dG}(t_n)) - \partial_t \mathbf{v}_\tau^{dG}(t_n).$$

Dann gilt

$$\mathbf{b}_n = \frac{1}{\vartheta_n(t_{n-1})} \{ \mathbf{v}_{n-1}^- - \mathbf{v}_\tau^{dG}(t_{n-1}) \} \quad (3.26)$$

und

$$\mathbf{v}^{dG}(t_{n-1}) - \mathbf{v}_{n-1}^- + \mathbf{b}_n \vartheta_n(t_{n-1}) = 0, \quad (3.27)$$

wobei

$$\vartheta_n(t) := \frac{\tau_n}{2} \hat{\theta}(\hat{t}) \quad \text{mit} \quad \hat{t} := T_n^{-1}(t)$$

und $T_n : \hat{I} \rightarrow I_n$ die affine Referenztransformation ist. $\hat{\theta} \in \mathbb{P}_{k+1}(\hat{I})$ ist das Polynom, welches durch die Bedingungen $\hat{\theta}'(1) = 1$ und $\hat{\theta}(\hat{t}_j) = 0$ für alle $j = 1, \dots, k+1$ definiert ist, wobei \hat{t}_j die Integrationspunkte der $(k+1)$ -punktigen rechtsseitigen Gauß-Radau-Formel auf \hat{I} sind.

Beweis. Da (3.27) unmittelbar aus (3.26) folgt, reicht es nur diese zu beweisen. Dazu führen wir zunächst die spezielle Testfunktion $\varrho_n \in \mathbb{P}_k$ ein, die mit den Eigenschaften $\varrho_n(t_n) = 1$ und $\varrho_n(t_{n,j}) = 0$ für alle $j = 1, \dots, k$ ausgestattet ist. Das Einsetzen von ϱ_n als Testfunktion in das $dG(k)$ -Verfahren (3.18) ergibt

$$\begin{aligned} \frac{\tau_n}{2} \sum_{j=1}^{k+1} \hat{\omega}_j \partial_t \mathbf{v}^{dG}(t_{n,j}) \varrho_n(t_{n,j}) + (\mathbf{v}^{dG}(t_{n-1}) - \mathbf{v}_{n-1}^-) \varrho_n(t_{n-1}) \\ = \frac{\tau_n}{2} \sum_{j=1}^{k+1} \hat{\omega}_j \tilde{\mathbf{F}}(t_{n,j}, \mathbf{v}^{dG}(t_{n,j})) \varrho_n(t_{n,j}). \end{aligned}$$

Nach der Definition von ϱ_n und aufgrund von $t_{n,k+1} = t_n$ ist diese Gleichung äquivalent zu

$$\frac{\tau_n}{2} \hat{\omega}_{k+1} \partial_t \mathbf{v}^{dG}(t_n) + (\mathbf{v}^{dG}(t_{n-1}) - \mathbf{v}_{n-1}^-) \varrho_n(t_{n-1}) = \frac{\tau_n}{2} \hat{\omega}_{k+1} \tilde{\mathbf{F}}(t_n, \mathbf{v}^{dG}(t_n)). \quad (3.28)$$

Ein weiterer Nutzen aus den Eigenschaften von ϑ_n ist

$$\begin{aligned} 0 &= \frac{\tau_n}{2} \sum_{j=1}^{k+1} \hat{\omega}_j \vartheta_n(t_{n,j}) \varrho_n'(t_{n,j}) = \int_{I_n} \vartheta_n(t) \varrho_n'(t) dt \\ &= - \int_{I_n} \vartheta_n'(t) \varrho_n(t) dt + \vartheta_n(t_n) \varrho_n(t_n) - \vartheta_n(t_{n-1}) \varrho_n(t_{n-1}) \\ &= - \frac{\tau_n}{2} \sum_{j=1}^{k+1} \hat{\omega}_j \vartheta_n'(t_{n,j}) \varrho_n(t_{n,j}) - \vartheta_n(t_{n-1}) \varrho_n(t_{n-1}) \\ &= - \frac{\tau_n}{2} \hat{\omega}_{k+1} \vartheta_n'(t_n) \varrho_n(t_n) - \vartheta_n(t_{n-1}) \varrho_n(t_{n-1}). \end{aligned}$$

Daraus folgt

$$\vartheta_n(t_{n-1}) \varrho_n(t_{n-1}) = - \frac{\tau_n}{2} \hat{\omega}_{k+1}. \quad (3.29)$$

Nun betrachten wir die folgenden Umformulierungen:

$$\begin{aligned}
 & \left[\mathbf{v}^{dG}(t_{n-1}) - \mathbf{v}_{n-1}^- + \mathbf{b}_n \vartheta_n(t_{n-1}) \right] \varrho_n(t_{n-1}) \\
 &= \left[\mathbf{v}^{dG}(t_{n-1}) - \mathbf{v}_{n-1}^- \right] \varrho_n(t_{n-1}) + \mathbf{b}_n \vartheta_n(t_{n-1}) \varrho_n(t_{n-1}) \\
 &= \frac{\tau_n}{2} \hat{\omega}_{k+1} \tilde{\mathbf{F}}(t_n, \mathbf{v}^{dG}(t_n)) - \frac{\tau_n}{2} \hat{\omega}_{k+1} \partial_t \mathbf{v}^{dG}(t_n) - \mathbf{b}_n \frac{\tau_n}{2} \hat{\omega}_{k+1} \\
 &= \frac{\tau_n}{2} \hat{\omega}_{k+1} \left[\tilde{\mathbf{F}}(t_n, \mathbf{v}^{dG}(t_n)) - \partial_t \mathbf{v}^{dG}(t_n) - \left[\tilde{\mathbf{F}}(t_n, \mathbf{v}^{dG}(t_n)) - \partial_t \mathbf{v}^{dG}(t_n) \right] \right] \\
 &= 0.
 \end{aligned}$$

Aus diesem Ergebnis folgt unmittelbar die Behauptung. \square

Satz 3.3. Seien \mathbf{v}_τ^{dG} und $\mathbf{v}_\tau^{C^0}$ jeweils die Lösungen der numerisch integrierten Varianten der diskreten $dG(k)$ -Verfahren (3.18) und $dG-C^0(k+1)$ -Verfahren (3.21). Es sei angenommen, dass die Zeitschritte τ_n ausreichend klein sind, sodass die Probleme der beiden Verfahren eine eindeutige Lösung für alle $n = 1, \dots, N$ haben. Dann gilt auf jedem Teilintervall $I_n = (t_{n-1}, t_n]$

$$\mathbf{v}_\tau^{C^0}(t) = \mathbf{v}_\tau^{dG}(t) + \mathbf{b}_n \vartheta_n(t) \quad \forall t \in I_n, \quad (3.30)$$

wobei der Vektor $\mathbf{b}_n \in \mathcal{V}$ durch

$$\mathbf{b}_n := \tilde{\mathbf{F}}(t_n, \mathbf{v}_\tau^{dG}(t_n)) - \partial_t \mathbf{v}_\tau^{dG}(t_n) \quad (3.31)$$

gegeben ist und

$$\vartheta_n(t) := \frac{\tau_n}{2} \hat{\vartheta}(\hat{t}) \quad \text{mit} \quad \hat{t} := T_n^{-1}(t).$$

Dabei ist $T_n : \hat{I} \rightarrow I_n$ die affine Referenztransformation und $\hat{\vartheta} \in \mathbb{P}_{k+1}(\hat{I})$ das Polynom, welches durch die Bedingungen $\hat{\vartheta}'(1) = 1$ und $\hat{\vartheta}(\hat{t}_j) = 0$ für alle $j = 1, \dots, k+1$ definiert ist. Hierbei sind \hat{t}_j die Integrationspunkte der $(k+1)$ -punktigen rechtsseitigen Gauß–Radau-Formel auf \hat{I} .

Beweis. Sei \mathbf{v}_τ^{dG} die Lösung des numerisch integrierten Problems (3.18). Wir definieren $\mathbf{z}(t) := \mathbf{v}_\tau^{dG} + \mathbf{b}_n \vartheta_n(t)$. Wie bereits im cGP-Fall reicht es zu zeigen, dass \mathbf{z} die Bedingungen für die Lösung des diskreten $dG-C^0(k+1)$ -Problems in (3.21) erfüllt. Für alle Gauß–Radau-Stützstellen $t_{n,j}$ gilt $\vartheta_n(t_{n,j}) = 0$. Daraus folgt

$$\mathbf{z}(t_{n,j}) = \mathbf{v}_\tau^{dG}(t_{n,j}) \quad \forall j = 1, \dots, k+1. \quad (3.32)$$

Mithilfe von Lemma 3.2 werden wir zeigen, dass die Funktion \mathbf{z} die Differentialgleichung in t_{n-1} erfüllt. Es gilt

$$\begin{aligned}
 \mathbf{z}(t_{n-1}) &= \mathbf{v}_\tau^{dG}(t_{n-1}) + \mathbf{b}_n \vartheta_n(t_{n-1}) \\
 &= \mathbf{v}_\tau^{dG}(t_{n-1}) + \{ \mathbf{v}_{n-1}^- - \mathbf{v}_\tau^{dG}(t_{n-1}) \} \\
 &= \mathbf{v}_{n-1}^-.
 \end{aligned}$$

Im weiteren Verlauf werden wir zeigen, dass \mathbf{z} die numerisch integrierte variationelle Bedingung erfüllt. Sei $\psi \in \mathbb{P}_{k-1}(I_n)$ beliebig. Dann gilt

$$\begin{aligned}
 \sum_{j=1}^{k+1} \hat{\omega}_j \partial_t \mathbf{z}_\tau(t_{n,j}) \psi(t_{n,j}) &= \sum_{j=1}^{k+1} \hat{\omega}_j \partial_t \mathbf{v}_\tau^{dG}(t_{n,j}) \psi(t_{n,j}) + \sum_{j=1}^{k+1} \hat{\omega}_j \mathbf{b}_n \vartheta_n'(t_{n,j}) \psi(t_{n,j}) \\
 &= \sum_{j=1}^{k+1} \hat{\omega}_j \tilde{\mathbf{F}}(t_{n,j}, \mathbf{v}_\tau^{dG}(t_{n,j})) \psi(t_{n,j}) - \frac{2}{\tau_n} (\mathbf{v}^{dG}(t_{n-1}) - \mathbf{v}_{n-1}^-) \psi(t_{n-1}) \\
 &\quad + \frac{2}{\tau_n} \mathbf{b}_n \int_{I_n} \vartheta_n'(t) \psi(t) dt, \quad (3.33)
 \end{aligned}$$

wobei \mathbf{v}_τ^{dG} die Lösung des numerisch integrierten Problems (3.18) mit dem dG(k)-Verfahren ist. In der letzten Gleichung haben wir die exakte Integrierbarkeit der Polynome vom maximalen Grad $2k$ mittels der $(k+1)$ -punktigen Gauß–Radau-Formel angewendet. Eine partielle Integration liefert

$$\begin{aligned} \int_{I_n} \vartheta'_n(t)\psi(t) dt &= - \int_{I_n} \vartheta_n(t)\psi'(t) dt + \vartheta_n(t_n)\psi(t_n) - \vartheta_n(t_{n-1})\psi(t_{n-1}) \\ &= -\frac{\tau_n}{2} \sum_{j=1}^{k+1} \hat{\omega}_j \vartheta_n(t_{n,j})\psi'(t_{n,j}) - \vartheta_n(t_{n-1})\psi(t_{n-1}) \\ &= -\vartheta_n(t_{n-1})\psi(t_{n-1}), \end{aligned} \quad (3.34)$$

da ϑ_n in allen Quadraturstützstellen verschwindet und $t_n = t_{n,k+1}$ eine Quadraturstützstelle ist. Weiterhin liefert die Kombination von (3.33) und (3.34)

$$\begin{aligned} \sum_{j=1}^{k+1} \hat{\omega}_j \partial_t \mathbf{z}(t_{n,j})\psi(t_{n,j}) &= \sum_{j=1}^{k+1} \hat{\omega}_j \tilde{\mathbf{F}}(t_{n,j}, \mathbf{v}_\tau^{dG}(t_{n,j}))\psi(t_{n,j}) \\ &\quad - \frac{2}{\tau_n} \left[\mathbf{v}^{dG}(t_{n-1}) - \mathbf{v}_{n-1}^- + \mathbf{b}_n \vartheta_n(t_{n-1}) \right] \psi(t_{n-1}). \end{aligned}$$

Um zu zeigen, dass die Funktion \mathbf{z} die variationelle Bedingung in (3.21) erfüllt, muss bewiesen werden, dass der letzte Term gleich 0 ist. Dies liefert Lemma 3.2.

Bezüglich der Differentialgleichung in t_n haben wir

$$\begin{aligned} \partial_t \mathbf{z}(t_n) &= \partial_t (\mathbf{v}_\tau^{dG} + \tilde{\mathbf{b}}_n \vartheta_n)(t_n) \\ &= \partial_t \mathbf{v}_\tau^{dG}(t_n) + \tilde{\mathbf{b}}_n \vartheta'_n(t_n) \\ &= \partial_t \mathbf{v}_\tau^{dG}(t_n) + \tilde{\mathbf{b}}_n \\ &= \partial_t \mathbf{v}_\tau^{dG}(t_n) + \left\{ \mathbf{F}(t_n, \mathbf{v}_\tau^{dG}(t_n)) - \partial_t \mathbf{v}_\tau^{dG}(t_n) \right\} \\ &= \tilde{\mathbf{F}}(t_n, \mathbf{v}_\tau^{dG}(t_n)) = \tilde{\mathbf{F}}(t_n, \mathbf{z}(t_n)). \end{aligned}$$

Dies zeigt, dass die Funktion \mathbf{z} alle Bedingungen in (3.21) erfüllt. Aus der Eindeutigkeit der Lösung folgt, dass \mathbf{z} mit $\mathbf{u}_\tau^{C^0}$ übereinstimmt, infolgedessen ist der Beweis vom Satz 3.3 vollständig beendet. \square

Daraufbauend gilt die Behauptung für das ursprüngliche Problem (3.1). Analog zum Satz 3.1 erhalten wir auf jedem Teilintervall $I_n = (t_{n-1}, t_n]$

$$\mathbf{u}_\tau^{C^0}(t) = \mathbf{u}_\tau^{dG}(t) + \mathbf{M}^{-1}(t)\mathbf{b}_n \vartheta_n(t) \quad \forall t \in I_n,$$

wobei $\mathbf{u}_\tau^{dG}(t) := \mathbf{M}^{-1}(t)\mathbf{v}_\tau^{dG}(t)$ und $\mathbf{u}_\tau^{C^0}(t) := \mathbf{M}^{-1}(t)\mathbf{v}_\tau^{C^0}(t)$ für alle $t \in I_n$ jeweils die Lösungen, die aus den numerisch integrierten Varianten der diskreten dG(k)- und dG- $C^0(k+1)$ -Verfahren stammen und der Vektor $\mathbf{b}_n \in \mathcal{V}$ durch

$$\mathbf{b}_n := \mathbf{F}(t_n, \mathbf{u}_\tau^{dG}(t_n)) - \partial_t \left(\mathbf{M}\mathbf{v}_\tau^{dG} \right)(t_n) \quad (3.35)$$

bzw.

$$\mathbf{b}_n := \frac{\mathbf{M}(t_{n-1})}{\vartheta_n(t_{n-1})} \left\{ \mathbf{u}_{n-1}^- - \mathbf{u}_\tau^{dG}(t_{n-1}) \right\} \quad (3.36)$$

definiert ist.

In diesem Unterabschnitt wurde gezeigt, dass für $k \geq 2$ die numerisch integrierten Versionen der dG(k)- und der dG- $C^0(k+1)$ -Verfahren die gleichen linksseitigen Werte an den Endpunkten

der Zeitintervalle erzeugen. Dies impliziert, dass die $dG-C^0(k+1)$ -Verfahren auch A-stabil und L-stabil sind. Darüber hinaus kann die numerische Lösung auf der Basis der $dG-C^0(k)$ -Verfahren aus der von der dG-Verfahren erzeugten Lösung auf einfache Weise berechnet werden, siehe [29, Remark 6].

Schließlich haben wir das Ziel erreicht eine einheitliche Postprocessing-Methode für die Galerkin-Verfahren in zeitabhängigen Gebieten zu beweisen. Diese verallgemeinerte Methode lässt sich ebenso für Galerkin-Verfahren in festen Gebieten anwenden. Besonders auffällig ist, dass sich diese einheitliche Variante vom Postprocessing auf andere Probleme, wie beispielsweise auf die Navier–Stokes-Gleichungen, verallgemeinern lässt.

4 Zeitabhängige Konvektions-Diffusions-Reaktions-Gleichungen in festen Gebieten

Dieses Kapitel behandelt die zeitabhängigen Konvektions-Diffusions-Reaktions-Gleichungen in festen Gebieten. Für die Zeitdiskretisierung werden wir zum einen die stetigen Galerkin-Petrov-Verfahren (cGP) und zum anderen unstetigen Galerkin-Verfahren (dG) als variationelle Zeitdiskretisierungen höherer Ordnung anwenden. Die beiden Semidiskretisierungen werden mit einer Finite-Elemente-Methode (FEM) kombiniert, die durch eine lokale Projektion stabilisiert wird. Diesbezüglich untersuchen wir sowohl die Stabilität als auch eine a-priori Fehlerabschätzung. Abschließend werden wir die wichtigen Ergebnisse durch Beispiele numerisch illustrieren.

4.1 Problemstellung und Ortsstabilisierung

Sei Ω ein beschränktes Gebiet in $\mathbb{R}^d, d = 2, 3$, mit Lipschitz-stetigem Rand $\partial\Omega$ und sei $I := (0, T]$ mit $T > 0$. Wir betrachten das folgende zeitabhängige Konvektions-Diffusions-Reaktions-Problem:

Finde $u : I \times \Omega \rightarrow \mathbb{R}$, sodass

$$\begin{aligned} \partial_t u - \varepsilon \Delta_{\mathbf{x}} u + \mathbf{b} \cdot \nabla_{\mathbf{x}} u + cu &= f && \text{in } I \times \Omega, \\ u &= 0 && \text{auf } I \times \partial\Omega, \\ u(0, \cdot) &= u_0 && \text{in } \Omega, \end{aligned} \tag{4.1}$$

wobei $0 < \varepsilon \ll 1$ eine kleine positive Diffusionskonstante ist. \mathbf{b} und c sind zeitunabhängige Funktionen, die von \mathbf{x} abhängen können und die die Bedingung

$$c(\mathbf{x}) - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}(\mathbf{x}) \geq c_0 > 0 \quad \forall \mathbf{x} \in \Omega \tag{4.2}$$

erfüllen. Hierbei ist c_0 eine Konstante. Die äußere Kraft $f \in L^2(\Omega)$ und der ursprüngliche Anfangswert $u_0 \in H_0^1(\Omega)$ sind gegeben. Mit $\nabla_{\mathbf{x}}$ bzw. $\Delta_{\mathbf{x}}$ bezeichnen wir jeweils den Gradienten bzw. den Laplace-Operator bezüglich der Ortsvariable \mathbf{x} . Die gesetzte Bedingung (4.2) stellt für zeitabhängige Konvektions-Diffusions-Reaktions-Gleichungen keine Einschränkung der Allgemeinheit dar. Man kann nämlich bei Gleichungen, die die Bedingung nicht erfüllen, die Transformation $v(t, \mathbf{x}) = e^{\alpha t} u(t, \mathbf{x})$ einsetzen. Wir erhalten $\nabla_{\mathbf{x}} v = e^{\alpha t} \nabla_{\mathbf{x}} u$, $\Delta_{\mathbf{x}} v = e^{\alpha t} \Delta_{\mathbf{x}} u$ und $\partial_t v = \alpha e^{\alpha t} u + e^{\alpha t} \partial_t u$. So wird (4.1) folgendermaßen umgeschrieben

$$\partial_t v - \varepsilon \Delta_{\mathbf{x}} v + \mathbf{b} \cdot \nabla_{\mathbf{x}} v + \tilde{c} v = \tilde{f},$$

wobei $\tilde{c} := c + \alpha$ und $\tilde{f} := e^{-\alpha t} f$. Während (4.2) für diese Gleichung wie folgt aussieht:

$$\tilde{c}(\mathbf{x}) - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}(\mathbf{x}) \geq c_0 > 0 \Leftrightarrow c(\mathbf{x}) + \alpha - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}(\mathbf{x}) \geq c_0 > 0.$$

Diese lässt sich immer für ein hinreichend großes α erfüllen. Allerdings darf α nicht zu groß gewählt werden, damit die Kontrolle über das Verhalten von v nicht verloren geht.

Der Einfachheit halber besitzt das Gebiet Ω einen polygonalen ($d = 2$) bzw. polyhedralen ($d = 3$) Rand $\partial\Omega$ und das Problem ist mit einer homogenen Dirichlet-Randbedingung ausgestattet. Ein Zugang zum Lösen partieller Differentialgleichungen ist die Überführung in ein Variationsproblem. Aus einer Multiplikation der Differentialgleichung (4.1) mit einer geeigneten Funktion $v \in H_0^1(\Omega)$, der Integration der resultierenden Gleichung über Ω und einer anschließenden partiellen Integration folgt

$$(\partial_t u, v)_\Omega + (\mathbf{b} \cdot \nabla_{\mathbf{x}} u, v)_\Omega + \varepsilon(\nabla_{\mathbf{x}} u, \nabla_{\mathbf{x}} v)_\Omega + (cu, v)_\Omega = (f, v)_\Omega \quad \forall v \in H_0^1(\Omega). \quad (4.3)$$

Dabei ist $(u, v)_\Omega := \int_\Omega uv \, d\mathbf{x}$ und $\|\cdot\|_0$ die zugehörige L^2 -Norm.

Nun definieren wir die Bilinearform $a(u, v) : H_0^1(\Omega) \times H_0^1(\Omega) \rightarrow \mathbb{R}$ durch

$$a(u, v) := \varepsilon(\nabla_{\mathbf{x}} u, \nabla_{\mathbf{x}} v)_\Omega + (\mathbf{b} \cdot \nabla_{\mathbf{x}} u, v)_\Omega + (cu, v)_\Omega,$$

somit kann die schwache Formulierung von (4.1) gemäß

Finde $u \in H^1(H_0^1(\Omega))$ mit $u(0) = u_0$, sodass

$$(\partial_t u, v)_\Omega + a(u, v) = (f, v)_\Omega \quad \forall v \in L^2(H_0^1(\Omega)) \quad (4.4)$$

umgeschrieben werden. Sei nun $f \in L^2(L^2(\Omega))$. Man beachte, dass aus $u \in H^1(H_0^1(\Omega))$ die Stetigkeit von $u : I \rightarrow L^2(\Omega)$ folgt. Dies garantiert die Wohlgestelltheit der Anfangsbedingung $u(0) = u_0$. Wir integrieren das Problem (4.4) über I und erhalten das neue Problem

Finde $u \in H^1(H_0^1(\Omega))$ mit $u(0) = u_0$, sodass

$$\int_0^T \left((\partial_t u(t), v(t))_\Omega + a(u(t), v(t)) \right) dt = \int_0^T (f(t), v(t))_\Omega dt \quad \forall v \in L^2(H_0^1(\Omega)). \quad (4.5)$$

Die Lösung des Problems (4.5) erfüllt $u \in L^2(H_0^1(\Omega))$ und $\partial_t u \in L^2(H_0^1(\Omega))$.

Der Einfachheit halber werden wir im Folgenden das Skalarprodukt $(v, w)_\Omega$ ohne den Index Ω schreiben.

4.1.1 Lokale Stabilisierung (LPS)

Bei der Finite-Elemente-Methode (FEM) wird das Gebiet Ω in ein Ortsgitter Ω_h zerlegt. Dabei sei $h := \max_{1 \leq i \leq M} h_i$ mit h_i der Durchmesser des i -ten Elements K , welches wir im Folgenden definieren. Wir bezeichnen mit \mathcal{T}_h eine Triangulierung des Gebietes Ω , sodass $\overline{\Omega}_h = \cup_{K \in \mathcal{T}_h} K$. Seien $\mathcal{V}_h \subset H^1(\Omega)$ der Raum von stetigen stückweise polynomialen Funktionen vom Höchstgrad r und

$$\mathcal{V}_h^0 := \{v_h \in \mathcal{V}_h : v_h|_{\partial\Omega} = 0\} \subset H_0^1(\Omega)$$

der Finite-Elemente-Ansatz- und Testraum auf Ω .

Im Falle einer Konvektionsdominanz ($\varepsilon \ll 1$) können in der diskreten Lösung durch eine Galerkin-Diskretisierung von (4.5) unphysikalische Oszillationen auftreten. Zur Vermeidung dieses Problems, betrachten wir eine Stabilisierung durch die lokale Projektion. Dazu seien $\mathcal{D}_h(K)$, $K \in \mathcal{T}_h$ endlichdimensionale Räume und $\pi_K : L^2(K) \rightarrow \mathcal{D}_h(K)$ die lokale Projektion in $\mathcal{D}_h(K)$. Der Fluktuationoperator $\kappa_K : L^2(K) \rightarrow L^2(K)$ ist gegeben durch $\kappa_K v_h := v_h - \pi_K v_h$. Wir definieren den Stabilisierungsterm

$$s_h(u_h, v_h) := \sum_{K \in \mathcal{T}_h} \beta_K (\kappa_K(\nabla_{\mathbf{x}} u_h), \kappa_K(\nabla_{\mathbf{x}} v_h))_K, \quad (4.6)$$

wobei $\beta_K, K \in \mathcal{T}_h$ nicht-negative und zu h_K proportionale Konstanten sind. Dies stellt die Fluktuation des Gradienten dar. Der Fluktuationsoperator wird für vektorwertige Funktionen komponentenweise angewendet. Im Vergleich zum Standard-Galerkin-Verfahren, gibt die lokale Projektionsstabilisierung eine zusätzliche Kontrolle über die Fluktuation des Gradienten. Eine weitere Möglichkeit erhält man, indem man den Gradienten ∇u_h durch den Gradienten in Stromlinienrichtung $\mathbf{b} \cdot \nabla_{\mathbf{x}} u_h$ ersetzt oder noch besser durch $\mathbf{b}_K \cdot \nabla_{\mathbf{x}} u_h$, wobei \mathbf{b}_K eine stückweise konstante Approximation von \mathbf{b} ist, siehe dazu [27, 28]. Man nehme beispielsweise $\mathbf{b}_K = \frac{1}{|K|} \int_K \mathbf{b}(\mathbf{x}) d\mathbf{x}$. Aus der Definition von s_h folgt die Symmetrieeigenschaft $s_h(u_h, v_h) = s_h(v_h, u_h)$. Für mehr Details über Stabilität, Konvergenz, sowie weitere Varianten der LPS-Methode verweisen wir auf [2] und [31]. Sei $u_{0,h}$ eine geeignete Approximation von u_0 in \mathcal{V}_h^0 , die wir später konkret angeben werden. Anschließend erhalten wir die stabilisierte Formulierung von (4.5) gemäß:

Finde $u_h \in H^1(\mathcal{V}_h^0)$ mit $u_h(0) = u_0$, sodass

$$\int_0^T \left((\partial_t u_h(t), v_h)_\Omega + a_h(u_h(t), v_h) \right) dt = \int_0^T (f(t), v_h)_\Omega dt \quad \forall v_h \in L^2(\mathcal{V}_h^0), \quad (4.7)$$

wobei die diskrete Bilinearform a_h durch

$$a_h(u_h, v_h) := a(u_h, v_h) + s_h(u_h, v_h) \quad (4.8)$$

definiert ist. Die Voraussetzung (4.2) garantiert die Koerzivität der Bilinearform a_h . Es gilt

$$a_h(v, v) \geq \|v\|^2 \quad \forall v \in \mathcal{V}_h, \quad (4.9)$$

wobei

$$\|v\| := \left\{ \varepsilon |v|_1^2 + c_0 \|v\|_0^2 + s_h(u, v) \right\}^{1/2}.$$

Die Stabilitäts- und Konvergenzeigenschaften der lokalen Projektionsverfahren basieren auf den folgenden Annahmen in Bezug auf den diskreten Raum \mathcal{V}_h und bezüglich der lokalen Projektionsräume $\mathcal{D}_h(K), K \in \mathcal{T}_h$, vgl. [30],[35, Theorem 3.74].

Annahme 4.1. *Es existiert ein Interpolationsoperator $\tilde{P}_h : H_0^1(\Omega) \rightarrow \mathcal{V}_h \cap H_0^1(\Omega)$, sodass für alle $K \in \mathcal{T}_h, v \in H^l(\omega(K))$ und $1 \leq l \leq r+1$*

$$\|v - \tilde{P}_h v\|_{0,K} + h_K |v - \tilde{P}_h v|_{1,K} \leq Ch^l \|v\|_{l,\omega(K)}$$

gilt, wobei $\omega(K)$ eine Umgebung von K ist.

Für mehr Details über solche Operatoren verweisen wir auf [37, 12, 30].

Annahme 4.2. *Der Fluktuationsoperator κ_K erfüllt die folgende Eigenschaft:*

$$\|\kappa_K q\|_{0,K} \leq Ch_K^l |q|_{l,K} \quad \forall K \in \mathcal{T}_h, \quad \forall q \in H^l(K), \quad 0 \leq l \leq r.$$

Der Einfachheit halber definieren wir den zur Bilinearform a gehörigen Operator A durch

$$\begin{aligned} A &: H_0^1(\Omega) \rightarrow H^{-1}(\Omega), \\ \langle Au, v \rangle &:= a(u, v) \quad \forall v \in H_0^1(\Omega). \end{aligned}$$

Analog ergibt sich der zur Bilinearform a_h gehörige diskrete Operator A_h durch

$$\begin{aligned} A_h &: H_0^1(\Omega) \rightarrow H^{-1}(\Omega), \\ \langle A_h u, v \rangle &:= a_h(u, v) \quad \forall v \in H_0^1(\Omega). \end{aligned}$$

Um die Verständlichkeit der Arbeit zu verbessern, verwenden wir die gleiche Notation für den zur Bilinearform s_h gehörigen Operator S_h durch

$$\begin{aligned} S_h &: H_0^1(\Omega) \rightarrow H^{-1}(\Omega), \\ \langle S_h u, v \rangle &:= s_h(u, v) \quad \forall v \in H_0^1(\Omega). \end{aligned}$$

Somit ist (4.8) äquivalent zu $A_h = A + S_h$. Aus der Cauchy–Schwarz-Ungleichung und der Definition (4.6) folgt die Stetigkeit des Stabilisierungsoperators S_h . Es gilt

$$\begin{aligned} |\langle S_h v, w \rangle| &= |s_h(v, w)| \leq (s_h(v, v))^{1/2} (s_h(w, w))^{1/2} \\ &\leq \|v\| \|w\|. \end{aligned} \quad (4.10)$$

Der Operator A_h erfüllt die folgende Abschätzung:

Satz 4.3. *Es gilt*

$$|\langle A_h u, v \rangle| \leq C \|u\|_1 \|v\|. \quad (4.11)$$

Beweis. Anhand der Definition kann man die Stetigkeit des Operators A_h zeigen. Es gilt

$$\langle A_h u, v \rangle = \varepsilon(\nabla_{\mathbf{x}} u, \nabla_{\mathbf{x}} v) + (\mathbf{b} \cdot \nabla_{\mathbf{x}} u, v) + (cu, v) + s_h(u, v). \quad (4.12)$$

Die Abschätzung der einzelnen Terme ergibt

$$\begin{aligned} |\varepsilon(\nabla_{\mathbf{x}} u, \nabla_{\mathbf{x}} v)| &\leq \varepsilon^{1/2} \|\nabla_{\mathbf{x}} u\|_0 \varepsilon^{1/2} \|\nabla_{\mathbf{x}} v\|_0, \\ |(cu, v)| &\leq \|c\|_{\infty} \|u\|_0 \|v\|_0, \\ |(\mathbf{b} \cdot \nabla_{\mathbf{x}} u, v)| &\leq \|\mathbf{b}\|_{\infty} \|\nabla_{\mathbf{x}} u\|_0 \|v\|_0 \end{aligned}$$

und anhand der Abschätzung (4.10) und der Annahme 4.2 lässt sich der letzte Term wie folgt abschätzen:

$$|\langle S_h v, w \rangle| = |s_h(v, w)| \leq C \|\nabla_{\mathbf{x}} u\|_0 \|w\|.$$

Die Behauptung folgt aus der Kombination der Abschätzungen der jeweiligen Terme in (4.12). \square

In Bezug auf den Interpolationsoperator \tilde{P}_h haben wir die folgende Fehlerabschätzung:

Korollar 4.4. [\tilde{P}_h -Fehler] Sei $\varphi \in H_0^1(\Omega) \cap H^{r+1}(\Omega)$. Es gilt

$$|a_h(\tilde{P}_h \varphi - \varphi, v)| \leq C(\varepsilon^{1/2} + h^{1/2}) h^r \|\varphi\|_{r+1} \|v\| \quad \forall v \in H_0^1(\Omega).$$

Beweis.

$$\begin{aligned} a_h(\tilde{P}_h \varphi - \varphi, v) &= \varepsilon(\nabla_{\mathbf{x}}(\tilde{P}_h \varphi - \varphi), \nabla_{\mathbf{x}} v) + (\mathbf{b} \cdot \nabla_{\mathbf{x}}(\tilde{P}_h \varphi - \varphi), v) \\ &\quad + (c(\tilde{P}_h \varphi - \varphi), v) + s_h(\tilde{P}_h \varphi - \varphi, v). \end{aligned} \quad (4.13)$$

Wir werden nun die Terme auf der rechten Seite einzeln abschätzen. Aus der Annahme 4.1 folgt

$$\begin{aligned} |\varepsilon(\nabla_{\mathbf{x}}(\tilde{P}_h \varphi - \varphi), \nabla_{\mathbf{x}} v)| &\leq \varepsilon^{1/2} |\tilde{P}_h \varphi - \varphi|_1 \varepsilon^{1/2} \|v\|_1 \\ &\leq \varepsilon^{1/2} h^r \|\varphi\|_{r+1} \|v\|. \end{aligned} \quad (4.14)$$

Aus der Abschätzung (4.10), der Annahme 4.2 und $\beta_K \leq Ch_K \leq Ch$ folgt

$$\begin{aligned} |s_h(\tilde{P}_h \varphi - \varphi, v)| &\leq C \sum_{K \in \mathcal{T}_h} h_K^{1/2} (\kappa_K (\nabla_{\mathbf{x}}(\tilde{P}_h \varphi - \varphi))_K) \|v\| \\ &\leq C \sum_{K \in \mathcal{T}_h} h_K^{1/2} h_K^r |\varphi|_{r+1, K} \|v\| \\ &\leq Ch^{r+1/2} |\varphi|_{r+1} \|v\|. \end{aligned} \quad (4.15)$$

Aus einer partiellen Integration im Ort folgt

$$(\mathbf{b} \cdot \nabla_{\mathbf{x}}(\tilde{P}_h \varphi - \varphi), v) = -(\mathbf{b} \cdot \nabla_{\mathbf{x}} v, \tilde{P}_h \varphi - \varphi) - (v \nabla_{\mathbf{x}} \cdot \mathbf{b}, \tilde{P}_h \varphi - \varphi). \quad (4.16)$$

Sei nun $\bar{\mathbf{b}}$ die L^2 -Projektion von \mathbf{b} in den Raum der stückweise konstanten Funktionen. Aus der L^2 -Stabilität des Operators κ_K , einer inversen Ungleichung und der Identität $\kappa_K(\bar{\mathbf{b}} \cdot \nabla_{\mathbf{x}})v = \bar{\mathbf{b}} \cdot \kappa_K(\nabla_{\mathbf{x}}v)$ erhalten wir

$$\begin{aligned} \|\kappa_K(\mathbf{b} \cdot \nabla_{\mathbf{x}})v\|_{0,K} &\leq \|\kappa_K((\mathbf{b} - \bar{\mathbf{b}}) \cdot \nabla_{\mathbf{x}})v\|_{0,K} + \|\kappa_K(\bar{\mathbf{b}} \cdot \nabla_{\mathbf{x}})v\|_{0,K} \\ &\leq Ch_K |\mathbf{b}|_{1,\infty,K} \|\nabla_{\mathbf{x}}v\|_{0,K} + \|\mathbf{b}\|_{0,\infty,K} \|\kappa_K(\nabla_{\mathbf{x}}v)\|_{0,K} \\ &\leq C |\mathbf{b}|_{1,\infty,K} \|v\|_{0,K} + \|\mathbf{b}\|_{0,\infty,K} \|\kappa_K(\nabla_{\mathbf{x}}v)\|_{0,K}, \end{aligned}$$

vgl. [30, Seite 726]. Weiterhin folgt aus den Definitionen der Stabilisierung s_h und der Norm $\|\cdot\|$ die Beziehung $\beta_K^{1/2} \|v\|_{0,K} \leq \|v\|_K$. Somit lässt sich der erste Term auf der rechten Seite von (4.16) folgendermaßen abschätzen:

$$\begin{aligned} |(\mathbf{b} \cdot \nabla_{\mathbf{x}} v, \tilde{P}_h \varphi - \varphi)| &= |(\kappa_K(\mathbf{b} \cdot \nabla_{\mathbf{x}})v, \tilde{P}_h \varphi - \varphi)| \\ &\leq \|\kappa_K(\mathbf{b} \cdot \nabla_{\mathbf{x}})v\|_0 \|\tilde{P}_h \varphi - \varphi\|_0 \\ &\leq (|\mathbf{b}|_{1,\infty} \|v\|_0 + \|\kappa_K(\nabla_{\mathbf{x}})v\|_0) h^{r+1} \|\varphi\|_{r+1} \\ &\leq Ch^{r+1} \|\varphi\|_{r+1} (\|v\|_0 + \|\kappa_K(\nabla_{\mathbf{x}})v\|_0) \\ &\leq Ch^{r+1} \|\varphi\|_{r+1} (\|v\|_0 + h^{-1/2} \|v\|) \\ &\leq Ch^{r+1} h^{-1/2} \|\varphi\|_{r+1} \|v\|. \end{aligned} \quad (4.17)$$

Weiterhin gilt

$$\begin{aligned} |((c - \nabla_{\mathbf{x}} \cdot \mathbf{b})(\tilde{P}_h \varphi - \varphi), v)| &\leq \|c - \nabla_{\mathbf{x}} \cdot \mathbf{b}\|_{\infty} \|\tilde{P}_h \varphi - \varphi\|_0 \|v\|_0 \\ &\leq Ch^{r+1} \|\varphi\|_{r+1} \|v\|. \end{aligned} \quad (4.18)$$

Die Behauptung folgt aus der Kombination der Abschätzungen (4.13)–(4.18). \square

Im Folgenden sei P_h die L^2 -Projektion in den Finite-Elemente-Raum \mathcal{V}_h . Um die Glattheit der Funktionen $t \mapsto v(t)$ zu charakterisieren, führen wir den Raum $C^j(J, W)$ der j -mal stetig differenzierbaren Funktionen ein, den wir mit der Norm

$$\|v\|_{C^l(J, W)} := \max_{0 \leq k \leq l} \sup_{t \in J} \|\partial_t^k v(t)\|_W$$

ausstatten. Dabei seien J ein Zeitintervall und W ein Banachraum ausgestattet mit einer Norm $\|\cdot\|_W$.

Um die schwache Formulierung des volldiskreten Problems herzuleiten, betrachten wir die in Kapitel 2 eingeführte Zeitdiskretisierung. Wir werden im Folgenden sowohl stetige als auch unstetige Galerkin-Verfahren einbeziehen, die sich grundsätzlich im betrachteten Ansatz- und Testraum unterscheiden. Der Einfachheit halber verwenden wir im Folgenden statt $(\cdot, \cdot)_{\Omega}$ die Notation (\cdot, \cdot) und $w(t)$ ansatt $w(t, \mathbf{x})$.

4.2 Stetige Galerkin-Verfahren (cGP)

Aufbauend auf die in der Arbeit von Ern und Schieweck [14] verwendeten Techniken werden wir in diesem Abschnitt relevante Ergebnisse bezüglich cGP-Verfahren vorstellen.

4.2.1 Projektions- und Interpolationsoperatoren

Wir betrachten den Banachraum \mathcal{W} . Für eine hinreichend glatte Funktion v definieren wir sämtliche Operatoren, die für unsere Fehleranalyse notwendig sein werden.

Operatoren auf $\hat{I} = [-1, 1]$

Zunächst definieren wir eine L^2 -Projektion, die wir für die Analyse der cGP-Verfahren benötigen werden.

L^2 -Projektion $\hat{\pi}_{k-1}$:

$$\begin{aligned} \hat{\pi}_{k-1} : L^2(\hat{I}, \mathcal{W}) &\rightarrow \mathbb{P}_{k-1}(\hat{I}, \mathcal{W}), \\ \int_{-1}^1 (\hat{\pi}_{k-1} \hat{v}(\hat{t}), \hat{\varphi}(\hat{t})) \, d\hat{t} &= \int_{-1}^1 (\hat{v}(\hat{t}), \hat{\varphi}(\hat{t})) \, d\hat{t} \end{aligned} \quad (4.19)$$

für alle $\hat{\varphi} \in \mathbb{P}_{k-1}(\hat{I}, \mathcal{W})$.

Gauß-Lobatto-Interpolation \hat{I}_k^c :

Weiterhin betrachten wir die folgende Gauß-Lobatto-Interpolation:

$$\begin{aligned} \hat{I}_k^c : C(\hat{I}, \mathcal{W}) &\rightarrow \mathbb{P}_k(\hat{I}, \mathcal{W}), \\ \hat{I}_k^c \hat{v}(\hat{t}) &:= \sum_{\mu=0}^k \hat{v}(\hat{t}_\mu) \hat{L}_\mu(\hat{t}), \end{aligned} \quad (4.20)$$

wobei $\hat{t}_\mu, \mu = 0, \dots, k$, die $(k+1)$ Knoten der Gauß-Lobatto-Quadraturformel auf \hat{I} sind und \hat{L}_μ die dazugehörigen Lagrange-Basispolynome darstellen.

Hermite/Lagrange-Interpolation \hat{L}_{k+1}^c :

Nun definieren wir die Operatoren \hat{L}_{k+1}^c durch

$$\begin{aligned} \hat{L}_{k+1}^c : C^1(\hat{I}, \mathcal{W}) &\rightarrow \mathbb{P}_{k+1}(\hat{I}, \mathcal{W}), \\ \hat{L}_{k+1}^c \hat{v}(\hat{t}_\mu) &= \hat{v}(\hat{t}_\mu), \quad \mu = 0, \dots, k, \\ \partial_t \hat{L}_{k+1}^c \hat{v}(1) &= \partial_t \hat{v}(1). \end{aligned}$$

Hermite/Lagrange-Interpolation \hat{J}_{k+2}^c :

Wir definieren die Hermite/Lagrange-Interpolation $\hat{J}_{k+2}^c: C^1(\hat{I}, \mathcal{W}) \rightarrow \mathbb{P}_{k+2}(\hat{I}, \mathcal{W})$ durch

$$\begin{aligned} \hat{J}_{k+2}^c \hat{v}(-1) &= \hat{v}(-1), \\ \partial_{\hat{t}} \hat{J}_{k+2}^c \hat{v}(-1) &= \partial_{\hat{t}} \hat{v}(-1), \\ \hat{J}_{k+2}^c \hat{v}(1) &= \hat{v}(1), \\ \partial_{\hat{t}} \hat{J}_{k+2}^c \hat{v}(1) &= \partial_{\hat{t}} \hat{v}(1), \\ \hat{J}_{k+2}^c \hat{v}(\hat{t}_\mu) &= \hat{v}(\hat{t}_\mu), \quad \mu = 1, \dots, k-1. \end{aligned}$$

Dabei sind \hat{t}_μ die Gauß-Lobatto-Punkte auf \hat{I} .

Spezielle Interpolatoren \hat{R}_{k+1}^c :

Zunächst führen wir den speziellen Hilfsoperator $\hat{R}_{k+1}^c : C^1(\hat{I}, \mathcal{W}) \rightarrow \mathbb{P}_{k+1}(\hat{I}, \mathcal{W})$ ein, definiert auf dem Referenzintervall \hat{I} durch

$$\begin{aligned} \partial_{\hat{t}} \hat{R}_{k+1}^c \hat{v}(\hat{t}_\mu) &= \partial_{\hat{t}} \hat{J}_{k+2}^c \hat{v}(\hat{t}_\mu), \quad \mu = 0, \dots, k, \\ \hat{R}_{k+1}^c \hat{v}(-1) &= \hat{J}_{k+2}^c \hat{v}(-1) = \hat{v}(-1). \end{aligned} \quad (4.21)$$

Operatoren auf $I_n = [t_{n-1}, t_n]$.

Nun definieren wir Operatoren auf jedem Teilintervall I_n , die sich auf I verallgemeinern lassen.

L^2 -Projektion π_{k-1} :

$$\begin{aligned} \pi_{k-1} : L^2(I_n, \mathcal{W}) &\rightarrow \mathbb{P}_{k-1}(I_n, \mathcal{W}), \\ \int_{I_n} (\pi_{k-1}v(t), \varphi(t))_{\mathcal{W}} dt &= \int_{I_n} (v(t), \varphi(t))_{\mathcal{W}} dt \end{aligned} \quad (4.22)$$

für alle $\varphi \in \mathbb{P}_{k-1}(I_n, \mathcal{W})$.

Gauß–Lobatto-Interpolation I_k^c :

$$\begin{aligned} I_k^c : C(I_n, \mathcal{W}) &\rightarrow \mathbb{P}_k(I_n, \mathcal{W}) \\ I_k^c v(t) &:= \sum_{\mu=0}^k v(t_{n,\mu}) L_{n,\mu}(t) \quad \forall t \in I_n, \end{aligned} \quad (4.23)$$

wobei $t_{n,\mu}, \mu = 0, \dots, k$, die transformierten Gauß–Lobatto-Punkte auf I_n sind und $L_{n,\mu}$ die dazugehörigen Lagrange-Basispolynome bezeichnen. Die Gauß–Lobatto-Quadraturformel ist, wie in Kapitel 2 bereits erwähnt wurde, für Polynome vom Grad höchstens $(2k - 1)$ exakt, d.h.

$$Q_n[p] = \frac{\tau_n}{2} \sum_{\mu=0}^k \omega_{\mu} p(t_{n,\mu}) = \int_{I_n} p(t) dt \quad \forall p \in \mathbb{P}_{2k-1}. \quad (4.24)$$

Hermite/Lagrange-Interpolation J_{k+2}^c :

Wir definieren die Hermite/Lagrange-Interpolation J_{k+2}^c , sodass $J_{k+2}^c v \in \mathbb{P}_{k+2}^{C^1}(I, \mathcal{W})$ für alle $v \in C^1(I, \mathcal{W})$ durch

$$\begin{aligned} J_{k+2}^c : C^1(I, \mathcal{W}) &\rightarrow \mathbb{P}_{k+2}^{C^1}(I, \mathcal{W}), \\ J_{k+2}^c v(t_{n-1}) &= v(t_{n-1}), \\ \partial_t J_{k+2}^c v(t_{n-1}) &= \partial_t v(t_{n-1}), \\ J_{k+2}^c v(t_n) &= v(t_n), \\ \partial_t J_{k+2}^c v(t_n) &= \partial_t v(t_n), \\ J_{k+2}^c v(t_{n,\mu}) &= v(t_{n,\mu}), \quad \mu = 1, \dots, k-1, \quad n = 1, \dots, N. \end{aligned}$$

Dabei sind $t_{n,\mu}$ die Gauß–Lobatto-Punkte auf I_n . Die globale C^1 -Stetigkeit folgt unmittelbar aus der Definition an den Punkten t_n .

Spezielle Interpolatoren R_{k+1}^n :

Zunächst betrachten wir die speziellen Hilfsoperatoren R_{k+1}^n , definiert auf jedem Teilintervall I_n durch

$$\begin{aligned} R_{k+1}^n : C^1(I_n, \mathcal{W}) &\rightarrow \mathbb{P}_{k+1}(I_n, \mathcal{W}), \\ \partial_t R_{k+1}^n v(t_{n,\mu}) &= \partial_t J_{k+2}^c v(t_{n,\mu}), \quad \mu = 0, \dots, k, \\ R_{k+1}^n v(t_{n-1}) &= J_{k+2}^c v(t_{n-1}) = v(t_{n-1}). \end{aligned} \quad (4.25)$$

Hermite/Lagrange-Interpolation L_{k+1}^c :

Nun definieren wir den Operator L_{k+1}^c auf I_n durch

$$\begin{aligned} L_{k+1}^c : C^1(I_n, \mathcal{W}) &\rightarrow \mathbb{P}_{k+1}(I_n, \mathcal{W}), \\ L_{k+1}^c v(t_{n,\mu}) &= v(t_{n,\mu}), \quad \mu = 0, \dots, k, \\ \partial_t L_{k+1}^c v(t_n) &= \partial_t v(t_n). \end{aligned} \quad (4.26)$$

Globaler Operator R_{k+1}^c :

Im Weiteren definieren wir den globalen Operator R_{k+1}^c durch die Operatoren R_{k+1}^n , $n = 1, \dots, N$, mit

$$R_{k+1}^c|_{I_n} := R_{k+1}^n.$$

4.2.2 Eigenschaften der Interpolationsoperatoren

Nach Konstruktion ist J_{k+2}^c eine C^1 -Funktion mit den Werten in \mathcal{W} . Dies folgt aus den Bedingungen an den Funktionswerten und den Ableitungen. Da R_{k+1}^c die Ableitung von J_{k+2}^c , insbesondere in den Endpunkten der Intervalle I_n abgreift, ist die stückweise definierte Funktion $\partial_t(R_{k+1}^c v|_{I_n})$ stetig. Zusätzlich ist $R_{k+1}^c v$ selbst auch stetig, da

$$\begin{aligned} R_{k+1}^c v(t_n) &= R_{k+1}^c v(t_{n-1}) + \int_{I_n} \partial_t R_{k+1}^c v(t) dt \\ &= R_{k+1}^n v(t_{n-1}) + Q_n[\partial_t R_{k+1}^n v] \\ &= R_{k+1}^n v(t_{n-1}) + Q_n[\partial_t J_{k+2}^c v] \\ &= R_{k+1}^n v(t_{n-1}) + \int_{I_n} \partial_t J_{k+2}^c v(t) dt \\ &= R_{k+1}^n v(t_{n-1}) + J_{k+2}^c v(t_n) - J_{k+2}^c v(t_{n-1}) \\ &= v(t_{n-1}) + v(t_n) - v(t_{n-1}) \\ &= v(t_n). \end{aligned}$$

Dabei haben wir sowohl die Definition von J_{k+2}^c und die von R_{k+1}^c über J_{k+2}^c , als auch die exakte Integrierbarkeit von $\partial_t R_{k+1}^c v \in \mathbb{P}_k^c(I)$ und $\partial_t J_{k+2}^c v \in \mathbb{P}_{k+1}^c(I)$ durch die Gauß–Lobatto-Formel Q_n aus Kapitel 2. Somit bildet R_{k+1}^c auf C^1 -Funktionen ab.

Im Folgenden machen wir Gebrauch von dem Polynom $\hat{\zeta} \in \mathbb{P}_{k+1}(\hat{I}, \mathcal{V}_h)$, welches durch die Bedingungen $\partial_t \hat{\zeta}(1) = 1$ und $\hat{\zeta}(\hat{t}_j) = 0$ für alle $j = 0, \dots, k$, definiert ist. Dabei sind \hat{t}_j die Integrationspunkte der $(k+1)$ -punktigen Gauß–Lobatto-Formel auf \hat{I} . Weiterhin betrachten wir auf jedem Teilintervall I_n das Polynom $\zeta_n \in \mathbb{P}_{k+1}(\hat{I}, \mathcal{V}_h)$ durch $\zeta_n(t) := \frac{\tau_n}{2} \hat{\zeta}(\hat{t})$ mit $\hat{t} := T_n^{-1}(t)$. Dabei ist T_n die affine Referenztransformation definiert durch

$$\begin{aligned} T_n : \hat{I} &\rightarrow I_n \\ T_n(\hat{t}) &:= (t_{n-1} + t_n)/2 + \frac{\tau_n}{2} \hat{t}. \end{aligned} \quad (4.27)$$

Basierend auf die in (4.27) definierte Referenztransformation T_n können wir einen konkreten Zusammenhang zwischen den auf \hat{I} definierten Operatoren und den auf I_n definierten Operatoren herstellen, nämlich $\hat{\pi}_{k-1} \hat{v} = \pi_{k-1} v$, $\hat{I}_k^c \hat{v} = I_k^c v$ und $\hat{L}_{k+1}^c \hat{v} = L_{k+1}^c v$ für jede hinreichend glatte Funktion v und \hat{v} mit $\hat{v}(\hat{t}) = v(t)$.

Lemma 4.5. *Sei $w \in \mathbb{P}_{k+1}(I_n, \mathcal{V}_h)$. Es gilt*

$$w(t) = I_k^c w(t) + l_{n-1} \zeta_n(t) \quad \forall t \in I_n. \quad (4.28)$$

Weiterhin gilt für $w \in C^1(I_n, \mathcal{V}_h)$

$$L_{k+1}^c w(t) = I_k^c w(t) + \tilde{l}_{n-1} \zeta_n(t) \quad \forall t \in I_n. \quad (4.29)$$

Beweis. Durch die in 4.27 definierte Referenztransformation T_n kann (4.28) umgeschrieben werden als

$$\hat{w}(\hat{t}) = \hat{I}_k^c \hat{w}(\hat{t}) + \hat{l} \hat{\zeta}(\hat{t}) \quad \forall \hat{t} \in \hat{I}. \quad (4.30)$$

\hat{w} und $\hat{I}_k^c \hat{w}$ sind Polynome vom Grad $\leq k+1$, die in den $(k+1)$ Gauß-Lobatto-Punkten übereinstimmen. Daher muss ihre Differenz ein Vielfaches von $\hat{\zeta}$ sein. Dies zeigt die Eigenschaft (4.30) und daraus folgt die Behauptung. Es gilt $l_{n-1} := w(t_{n-1}) - I_k^c w(t_{n-1}^+) = [I_k^c w]_{n-1}$. Die zweite Aussage lässt sich ähnlich zeigen, indem man sie mithilfe der Referenztransformation T_n in \hat{I} schreibt. Diesbezüglich erhalten wir

$$\hat{L}_{k+1}^c(\hat{t}) = \hat{I}_k^c \hat{w}(\hat{t}) + \hat{l}_L \hat{\zeta}(\hat{t}) \quad \forall \hat{t} \in \hat{I}. \quad (4.31)$$

Die Terme $L_{k+1}^c w$ und $I_k^c w$ unterscheiden sich auf I_n nur um ein Vielfaches von ζ_n . \square

Sei nun $v \in C^{k+3}(I_n, \mathcal{V}_h)$ und $r \in \mathbb{N}$, dann erhalten wir für den oben definierten Interpolationsoperator J_{k+2}^c die Fehlerabschätzung

$$\sup_{t \in I_n} |J_{k+2}^c v(t) - v(t)|_r \leq C \tau_n^{k+3} \sup_{t \in I_n} |\partial_t^{k+3} v(t)|_r \quad (4.32)$$

und für die Zeitableitungen von J_{k+2}^c gilt

$$\sup_{t \in I_n} |\partial_t J_{k+2}^c v(t) - \partial_t v(t)|_r \leq C \tau_n^{k+2} \sup_{t \in I_n} |\partial_t^{k+3} v(t)|_r \quad (4.33)$$

und

$$\sup_{t \in I_n} |\partial_t^2 J_{k+2}^c v(t) - \partial_t^2 v(t)|_r \leq C \tau_n^{k+1} \sup_{t \in I_n} |\partial_t^{k+3} v(t)|_r. \quad (4.34)$$

Diese folgen unmittelbar aus den $(k+3)$ Bedingungen in der Definition von J_{k+2}^c . Aus den Definitionen von R_{k+1}^c und J_{k+2}^c kann man entnehmen, dass $\partial_t R_{k+1}^c v(t_{n,\mu}) = \partial_t J_{k+2}^c v(t_{n,\mu})$, $\mu = 0, \dots, k$, d.h. $\partial_t R_{k+1}^c$ greift $\partial_t J_{k+2}^c$ an den Stellen $t_{n,\mu}$, $\mu = 0, \dots, k$, ab. Wir erhalten somit die Interpolationsfehlerabschätzung auf I_n

$$\sup_{t \in I_n} \|\partial_t J_{k+2}^c v(t) - \partial_t R_{k+1}^c v(t)\|_0 \leq C \tau_n^{k+1} \sup_{t \in I_n} \|\partial_t^{k+2} v(t)\|_0. \quad (4.35)$$

Im Folgenden definieren wir den zu R_{k+1}^c gehörigen Operator

$$\hat{R}_{k+1}^c : C^1(\hat{I}, H^{r+1}(\Omega)) \rightarrow \mathbb{P}_{k+1}(\hat{I}, H^{r+1}(\Omega))$$

auf dem Referenzintervall \hat{I} durch

$$\hat{R}_{k+1}^c \hat{v}(\hat{t}) := R_{k+1}^c v(t) \quad \forall \hat{t} \in \hat{I}$$

und den zu J_{k+2}^c gehörigen Operator

$$\hat{J}_{k+2}^c : C^1(\hat{I}, H^{r+1}(\Omega)) \rightarrow \mathbb{P}_{k+2}(\hat{I}, H^{r+1}(\Omega))$$

durch

$$\hat{J}_{k+2}^c \hat{v}(\hat{t}) := R_{k+2}^c v(t) \quad \forall \hat{t} \in \hat{I}.$$

Dabei ist $t := T_n(\hat{t}) := (t_{n-1} + t_n)/2 + \frac{\tau_n}{2} \hat{t}$. Darausfolgend gilt für $j = 1, \dots, k+2$

$$\partial_t^j \hat{v}(\hat{t}) = \partial_t^j v(T_n(\hat{t})) \left(\frac{\tau_n}{2}\right)^j \quad \forall \hat{t} \in \hat{I}.$$

Bei den Endpunkten $\hat{t} = \pm 1$ wird nur eine einseitige Ableitung in Betracht gezogen.

Lemma 4.6. [R_{k+1}^c -Fehlerabschätzung] Für alle $n = 1, \dots, N$ sei $v \in C^{k+2}(I_n, H^{r+1}(\Omega))$. Es gilt

$$\sup_{t \in I_n} \|R_{k+1}^c v(t) - v(t)\|_{r+1} \leq C \tau_n^j \sup_{t \in I_n} \|\partial_t^j v(t)\|_{r+1} \quad \forall j = 1, \dots, k+2. \quad (4.36)$$

Darüber hinaus gilt

$$\sup_{t \in I_n} \|R_{k+1}^c v(t)\|_{r+1} \leq C \|v\|_{C^1(I_n, H^{r+1}(\Omega))} \quad \forall v \in C^1(I_n, H^{r+1}(\Omega)). \quad (4.37)$$

Beweis. Aufgrund der Umrechnung zwischen den Zeitableitungen auf dem Intervall I_n und dem Referenzzeitintervall \hat{I} ist (4.36) äquivalent zur Aussage

$$\sup_{\hat{t} \in \hat{I}} \|\hat{R}_{k+1}^c \hat{v}(\hat{t}) - \hat{v}(\hat{t})\|_{r+1} \leq C \sup_{\hat{t} \in \hat{I}} |\partial_{\hat{t}}^j \hat{v}(\hat{t})|_{r+1}, \quad j = 1, \dots, k+2.$$

Nun betrachten wir das Funktional $F : \hat{v} \mapsto \hat{R}_{k+1}^c \hat{v} - \hat{v}$ auf dem Raum $C^1(\hat{I}, H^{r+1}(\Omega))$. F ist wohldefiniert und stetig, da \hat{R}_{k+1}^c über \hat{J}_{k+2}^c definiert ist und \hat{J}_{k+2}^c nur auf den Funktionswert und die erste Ableitung von \hat{v} zurückgreift. Es gilt daher

$$\|F(\hat{v})\|_{C^0(\hat{I}, H^{r+1}(\Omega))} \leq C \|\hat{v}\|_{C^1(\hat{I}, H^{r+1}(\Omega))} \leq C \|\hat{v}\|_{C^j(\hat{I}, H^{r+1}(\Omega))},$$

wobei $j \geq 1$ ist.

Aus $\hat{v} \in \mathbb{P}_{k+1}(\hat{I}, H^{r+1}(\Omega))$ und $\hat{R}_{k+1}^c \hat{v} \in \mathbb{P}_{k+1}(\hat{I}, H^{r+1}(\Omega))$ folgt weiterhin $\partial_{\hat{t}} \hat{v} \in \mathbb{P}_k(\hat{I}, H^{r+1}(\Omega))$ und $\partial_{\hat{t}} \hat{R}_{k+1}^c \hat{v} \in \mathbb{P}_k(\hat{I}, H^{r+1}(\Omega))$. Aus der Definition von \hat{R}_{k+1}^c kann man $\partial_{\hat{t}} \hat{R}_{k+1}^c \hat{v}(t_{n,\mu}) = \partial_{\hat{t}} \hat{J}_{k+2}^c \hat{v}(\hat{t}_\mu) = \partial_{\hat{t}} \hat{v}(\hat{t}_\mu)$ für alle $\mu = 0, \dots, k$ entnehmen. Somit sind $\partial_{\hat{t}} \hat{v}$ und $\partial_{\hat{t}} \hat{R}_{k+1}^c \hat{v}$ Polynome vom Grad $\leq k$, die an $(k+1)$ Punkten übereinstimmen. Dies ergibt $\partial_{\hat{t}} \hat{R}_{k+1}^c \hat{v} = \partial_{\hat{t}} \hat{v}$. Darausfolgend gilt $\hat{R}_{k+1}^c \hat{v} = \hat{v} + c$, wobei c eine Konstante ist und aus $\hat{R}_{k+1}^c \hat{v}(0) = \hat{v}(0)$ erhalten wir $\hat{R}_{k+1}^c \hat{v} = \hat{v}$ für alle $\hat{v} \in \mathbb{P}_{k+1}(\hat{I}, H^{r+1}(\Omega))$. Daraufbauend verschwindet das Funktional F auf allen Polynomen bis zum Grad $(k+1)$. Mithilfe der Arbeit von Bramble-Hilbert, vgl. [9, Kapitel 3], angewendet auf dem Referenzzeitintervall \hat{I} , erhalten wir die gewünschte Abschätzung, da die Norm $\|\hat{v}\|_{C^j(\hat{I}, H^{r+1}(\Omega))}$ durch die zugehörige Halbnorm ersetzt werden darf.

Bezüglich der Stabilitätsaussage erhalten wir mithilfe der Dreiecksungleichung, sowie (4.36) mit $j = 1$

$$\begin{aligned} \sup_{t \in I_n} \|R_{k+1}^c v(t)\|_{r+1} &\leq \sup_{t \in I_n} \|R_{k+1}^c v(t) - v(t)\|_{r+1} + \sup_{t \in I_n} \|v(t)\|_{r+1} \\ &\leq C \tau_n \sup_{t \in I_n} \|\partial_t v(t)\|_{r+1} + \sup_{t \in I_n} \|v(t)\|_{r+1} \\ &\leq C \|v\|_{C^1(I_n, H^{r+1}(\Omega))}. \end{aligned}$$

□

Korollar 4.7. [$\partial_t R_{k+1}^c$ -Fehlerabschätzung] Für alle $n = 0, \dots, N$ und für alle $v \in C^{k+2}(I_n, H^{r+1}(\Omega))$ gilt

$$\sup_{t \in I_n} \|\partial_t R_{k+1}^c v(t) - \partial_t v(t)\|_{r+1} \leq C \tau_n^j \sup_{t \in I_n} \|\partial_t^{j+1} v(t)\|_{r+1} \quad \forall j = 0, \dots, k+1. \quad (4.38)$$

Darüber hinaus gilt

$$\sup_{t \in I_n} \|\partial_t R_{k+1}^c v(t)\|_{r+1} \leq C \|v\|_{C^1(I_n, H^{r+1}(\Omega))} \quad \forall v \in C^1(I_n, H^{r+1}(\Omega)). \quad (4.39)$$

Beweis. Sei L ein Interpolationsoperator (z.B. Lagrange-Interpolation) im Raum $\mathbb{P}_{k+1}(I_n, H^{r+1}(\Omega))$. Aus der Dreiecksungleichung erhalten wir

$$\|\partial_t R_{k+1}^c v(t) - \partial_t v(t)\|_{r+1} \leq \|\partial_t R_{k+1}^c v(t) - \partial_t Lv(t)\|_{r+1} + \|\partial_t Lv(t) - \partial_t v(t)\|_{r+1}.$$

Der erste Summand lässt sich mit einer inversen Ungleichung abschätzen. Es gilt

$$\begin{aligned} \sup_{t \in I_n} \|\partial_t R_{k+1}^c v(t) - \partial_t Lv(t)\|_{r+1} &\leq C \tau_n^{-1} \sup_{t \in I_n} \|R_{k+1}^c v(t) - Lv(t)\|_{r+1} \\ &\leq C \tau_n^{-1} \sup_{t \in I_n} \|R_{k+1}^c u(t) - v(t)\|_{r+1} + C \tau_n^{-1} \sup_{t \in I_n} \|v(t) - Lv(t)\|_{r+1}, \end{aligned}$$

wobei im letzten Schritt nochmals eine Dreiecksungleichung verwendet wurde. Die Differenz $R_{k+1}^c v - v$ lässt sich mit (4.36) abschätzen und liefert für $0 \leq j \leq k+2$ die Schranke $C \tau_n^j |\partial_t^j v|_{r+1}$. Die gleiche Abschätzung erhalten wir auch für $v - Lv$. Dies gilt ebenso für den verbleibenden Term $\|\partial_t Lv - \partial_t v\|_{r+1}$, da dort eine gewöhnliche Interpolation verwendet wird. Zusammengefasst ergibt sich (4.38).

Die Stabilitätsaussage kann wie folgt gezeigt werden:

$$\begin{aligned} \sup_{t \in I_n} \|\partial_t R_{k+1}^c v(t)\|_{r+1} &\leq \sup_{t \in I_n} \|\partial_t R_{k+1}^c v(t) - \partial_t v(t)\|_{r+1} + \sup_{t \in I_n} \|\partial_t v(t)\|_{r+1} \\ &\leq C \sup_{t \in I_n} \|\partial_t v\|_{r+1} + C \sup_{t \in I_n} \|\partial_t v\|_{r+1} \\ &\leq C \|v\|_{C^1(I_n, H^{r+1}(\Omega))}. \end{aligned}$$

Den ersten Term haben wir mit (4.38) und $j = 0$ abgeschätzt. \square

Bezüglich der Hermite/Lagrange-Interpolation L_{k+1}^c gilt die Fehlerabschätzung

$$\sup_{t \in I_n} \|L_{k+1}^c v(t) - v(t)\|_0 \leq C \tau_n^{k+2} \sup_{t \in I_n} \|\partial_t^{k+2} v(t)\|_0. \quad (4.40)$$

4.2.3 Normen

Satz 4.8. *Die Ausdrücke*

$$\left(\int_{I_n} \|v(t)\|_W^2 dt \right)^{1/2} \quad (4.41)$$

und

$$\left(\int_{I_n} \|\pi_{k-1} I_k^c v(t)\|_W^2 dt + \frac{\tau_n}{2} \|v(t_n)\|_W^2 + \left(\frac{\tau_n}{2}\right)^3 \|\partial_t v(t_n)\|_W^2 \right)^{1/2} \quad (4.42)$$

sind äquivalente Normen auf $\mathbb{P}_{k+1}(I_n, W)$, wobei die Äquivalenzkonstanten nicht von τ_n abhängen und W einen beliebigen Banachraum darstellt.

Beweis. Mit dem Gebrauch der Transformation T_n^{-1} von I_n auf \hat{I} ist die Behauptung äquivalent zu:

Die Ausdrücke

$$\left(\int_{-1}^1 \|\hat{v}(\hat{t})\|_W^2 d\hat{t} \right)^{1/2} \quad (4.43)$$

und

$$\left(\int_{-1}^1 \|\widehat{\pi}_{k-1} \widehat{I}_k^c \widehat{v}(\hat{t})\|_W^2 d\hat{t} + \|\widehat{v}(1)\|_W^2 + \|\partial_{\hat{t}} \widehat{v}(1)\|_W^2 \right)^{1/2} \quad (4.44)$$

sind äquivalente Normen auf dem endlich dimensionalen Raum $\mathbb{P}_{k+1}(\hat{I}, W)$. Der Ausdruck (4.43) ist offensichtlich eine Norm. Zudem erfüllt der Ausdruck (4.44) nach Konstruktion alle Eigenschaften einer Halbnorm. Mit dem Nachweis der verbleibenden Normeigenschaft folgt die Aussage mittels der Normäquivalenz auf dem endlich dimensionalen Raum $\mathbb{P}_{k+1}(\hat{I}, W)$. Es bleibt folglich zu zeigen, dass sich aus

$$0 = \int_{-1}^1 \|\widehat{\pi}_{k-1} \widehat{I}_k^c \widehat{v}(\hat{t})\|_W^2 d\hat{t} + \|\widehat{v}(1)\|_W^2 + \|\partial_{\hat{t}} \widehat{v}(1)\|_W^2 \quad (4.45)$$

die Folgerung $\widehat{v} \equiv 0$ ergibt. Aus (4.45) folgt, dass $\widehat{\pi}_{k-1} \widehat{I}_k^c \widehat{v} = 0$. Da $\widehat{I}_k^c \widehat{v} \in \mathbb{P}_k(\hat{I}, W)$ gilt, ist $\widehat{I}_k^c \widehat{v}$ ein Vielfaches des k -ten Legendre-Polynoms P_k , d.h. $\widehat{I}_k^c \widehat{v}(\hat{t}) = \hat{a} P_k(\hat{t})$ auf \hat{I} . Da \widehat{v} und $\widehat{I}_k^c \widehat{v}$ nach Definition von \widehat{I}_k^c an den Gauß–Lobatto-Punkten $\hat{t}_\mu, \mu = 0, \dots, k$, übereinstimmen und $\widehat{v} \in \mathbb{P}_{k+1}(\hat{I}, W)$ gilt, folgt aus Lemma 4.5 die äquivalente Form auf \hat{I}

$$\widehat{v}(\hat{t}) = \widehat{I}_k^c \widehat{v}(\hat{t}) + \hat{l} \hat{\zeta}(\hat{t}) \quad \forall \hat{t} \in \hat{I},$$

wobei $\hat{\zeta} \in \mathbb{P}_{k+1}(\hat{I}, W)$ an den Gauß–Lobatto-Punkten verschwindet und mittels $\partial_{\hat{t}} \hat{\zeta}(1) = 1$ normiert ist. Zusammen ergibt sich hier

$$\widehat{v}(\hat{t}) = \hat{a} P_k(\hat{t}) + \hat{l} \hat{\zeta}(\hat{t}).$$

Aus $\widehat{v}(1) = 0$, $P_k(1) = 1$ und $\hat{\zeta}(1) = 0$ folgt nun $\hat{a} = 0$. Schließlich liefert die Zeitableitung

$$\partial_{\hat{t}} \widehat{v}(\hat{t}) = \hat{l} \partial_{\hat{t}} \hat{\zeta}(\hat{t}).$$

Aus $\partial_{\hat{t}} \widehat{v}(1) = 0$ und $\partial_{\hat{t}} \hat{\zeta}(1) = 1$ ergibt sich $\hat{l} = 0$. Somit haben wir $\widehat{v} \equiv 0$ erhalten und der Ausdruck (4.44) ist eine Norm. Die Behauptung folgt unmittelbar daraus. \square

4.2.4 Volldiskretes Problem mit Ortsstabilisierung

Das volldiskrete cGP-Verfahren kombiniert mit der Ortsstabilisierung lässt sich folgendermaßen formulieren:

Finde $u_{\tau h}|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V}_h^0)$ mit $u_{\tau h}(t_{n-1}) = u_{\tau h}^{n-1}$ und

$$\int_{I_n} \langle \partial_t u_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt = Q_n[(f, v_{\tau h})] \quad \forall v_{\tau h} \in \mathbb{P}_{k-1}(I_n, \mathcal{V}_h). \quad (4.44)$$

4.2.5 Volldiskretes Problem und Postprocessing

In diesem Abschnitt betrachten wir das Postprocessing aus Kapitel 3. Das Postprocessing wird nur für die volldiskrete Lösung $u_{\tau h}$ benötigt, damit es über die Differentialgleichung definiert werden kann. Das cGP-Problem formulieren wir gemäß:

Finde $u_{\tau h}|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V}_h^0)$ mit $u_{\tau h}(t_{n-1}) = u_{\tau h}^{n-1}$ und

$$\int_{I_n} \langle \partial_t u_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt = Q_n[(f, v_{\tau h})] \quad \forall v_{\tau h} \in \mathbb{P}_{k-1}(I_n, \mathcal{V}_h). \quad (4.45)$$

Hierbei wird $u_{\tau h}(0) = \tilde{P}_h u_0$ als Approximation an der Anfangsbedingung u_0 betrachtet. Auf dem Intervall I_n setzen wir

$$\tilde{u}_{\tau h}(t) = u_{\tau h}(t) + c_n \zeta_n(t), \quad (4.46)$$

wobei $c_n \in \mathcal{V}_h$ eine ortsdiskrete Funktion darstellt und $\zeta_n \in \mathbb{P}_{k+1}(I_n, \mathcal{V}_h)$, mit $\zeta_n(t_{n,\mu}) = 0$ für alle $\mu = 0, \dots, k$ und $\partial_t \zeta_n(t_n) = 1$. Dabei sind $t_{n,\mu}$ die Gauß-Lobatto-Punkte.

Die nachbereitete Lösung $\tilde{u}_{\tau h}$ soll im L^2 -Ortssinne die zeitliche Differentialgleichung an $t = t_n$ erfüllen, d.h.,

$$(\partial_t \tilde{u}_{\tau h}(t_n), v_h) = \langle f(t_n) - A_h \tilde{u}_{\tau h}(t_n), v_h \rangle \quad \forall v_h \in \mathcal{V}_h.$$

Aufgrund von $\zeta_n(t_n) = 0$ ist $\tilde{u}_{\tau h}(t_n) = u_{\tau h}(t_n)$, deshalb ist die Forderung äquivalent zu

$$(\partial_t u_{\tau h}(t_n), v_h) + (c_n, v_h) \partial_t \zeta_n(t_n) = \langle f(t_n) - A_h u_{\tau h}(t_n), v_h \rangle \quad \forall v_h \in \mathcal{V}_h.$$

Aus $\partial_t \zeta_n(t_n) = 1$ folgt

$$(c_n, v_h) = \langle f(t_n) - A_h u_{\tau h}(t_n) - \partial_t u_{\tau h}(t_n), v_h \rangle.$$

Dies liefert

$$c_n = P_h f(t_n) - A_h u_{\tau h}(t_n) - \partial_t u_{\tau h}(t_n),$$

wobei P_h die L^2 -Projektion in den Finite-Elemente-Raum \mathcal{V}_h darstellt. Man beachte, dass $u_{\tau h}$ und $\tilde{u}_{\tau h}$ in den Gauß-Lobatto-Punkten übereinstimmen.

4.2.6 Fehleranalyse

Wir beginnen mit dem folgenden Hilfssatz, um die Ergebnisse bezüglich der cGP-Verfahren zu veranschaulichen.

Lemma 4.9. *Es gilt für $n = 1, \dots, N$*

$$\int_{I_n} (\partial_t \zeta_n(t), p(t)) dt = 0 \quad \forall p \in \mathbb{P}_{k-1}(I_n, \mathcal{V}_h).$$

Beweis. Aus einer partiellen Integration folgt

$$\begin{aligned} \int_{I_n} (\partial_t \zeta_n(t), p(t)) dt &= - \int_{I_n} (\zeta_n(t), \partial_t p(t)) dt + [(\zeta_n(t), p(t))]_{t_{n-1}}^{t_n} \\ &= - \sum_{\mu=0}^k \hat{\omega}_\mu \zeta_n(t_{n,\mu}) \partial_t p(t_{n,\mu}) + [(\zeta_n(t), p(t))]_{t_{n-1}}^{t_n} \\ &= 0. \end{aligned}$$

Dies gilt, da ζ_n an allen Gauß-Lobatto-Punkten verschwindet, insbesondere an den Endpunkten, die zu den Gauß-Lobatto-Stützstellen gehören. \square

Bezüglich der durch Postprocessing nachbearbeiteten Lösung ist das folgende Lemma zu betrachten:

Lemma 4.10. *Für alle $n = 1, \dots, N$ gilt*

$$\int_{I_n} \langle \partial_t \tilde{u}_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt = Q_n[(f, v_{\tau h})] \quad \text{für alle } v_{\tau h} \in \mathbb{P}_{k-1}(I_n, \mathcal{V}_h^0).$$

Beweis. Aus der Definition vom Postprocessing folgt

$$\int_{I_n} (\partial_t \tilde{u}_{\tau h}(t), v_{\tau h}(t)) dt = \int_{I_n} (\partial_t u_{\tau h}(t), v_{\tau h}(t)) dt + \int_{I_n} (c_n \partial_t \zeta_n(t), v_{\tau h}(t)) dt.$$

Da $v_{\tau h} \in \mathbb{P}_{k-1}(I_n, \mathcal{V}_h)$, liefert das Lemma 4.9

$$\int_{I_n} (c_n \partial_t \zeta_n(t), v_{\tau h}(t)) dt = 0 \quad (4.47)$$

und es folgt unmittelbar die Behauptung. \square

Um weitere Ergebnisse bezüglich des volldiskreten Problems zu erläutern, definieren wir auf jedem Teilintervall I_n , $n = 1, \dots, N$, die Bilinearform B_n durch

$$B_n(w, v) := Q_n[(\partial_t w, v)] + Q_n[\langle A_h w, v \rangle].$$

Lemma 4.11. *Die nachbereitete Lösung erfüllt für alle $n = 1, \dots, N$:*

$$B_n(\tilde{u}_{\tau h}, v_{\tau h}) = Q_n[(f, v_{\tau h})] \quad \forall v_{\tau h} \in \mathbb{P}_{k-1}(I_n, \mathcal{V}_h).$$

Beweis. Aus der Definition der Bilinearform B_n folgt

$$B_n(\tilde{u}_{\tau h}, v_{\tau h}) = Q_n[(\partial_t \tilde{u}_{\tau h}, v_{\tau h})] + Q_n[\langle A_h \tilde{u}_{\tau h}, v_{\tau h} \rangle].$$

Im ersten Summanden auf der rechten Seite liegt das Produkt in $\mathbb{P}_{2k-1}(I_n)$. Damit ist die Gauß–Lobatto-Formel exakt und wir erhalten

$$Q_n[(\partial_t \tilde{u}_{\tau h}, v_{\tau h})] = \int_{I_n} (\partial_t \tilde{u}_{\tau h}(t), v_{\tau h}(t)) dt = \int_{I_n} (\partial_t u_{\tau h}(t), v_{\tau h}(t)) dt,$$

wobei wir Lemma 4.9 verwendet haben. Für den zweiten Summanden ergibt sich

$$Q_n[\langle A_h \tilde{u}_{\tau h}, v_{\tau h} \rangle] = Q_n[\langle A_h u_{\tau h}, v_{\tau h} \rangle] = \int_{I_n} \langle A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt,$$

da $u_{\tau h}$ und $\tilde{u}_{\tau h}$ in den Gauß–Lobatto-Punkten übereinstimmen. Der letzte Schritt ist möglich, da $\langle A_h u_{\tau h}, v_{\tau h} \rangle \in \mathbb{P}_{2k-1}(I_n)$ gilt. Die Quadratur ist dementsprechend exakt und es folgt

$$B_n(\tilde{u}_{\tau h}, v_{\tau h}) = \int_{I_n} \langle \partial_t u_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt = Q_n[(f, v_{\tau h})]$$

für alle $v_{\tau h} \in \mathbb{P}_{k-1}(I_n, \mathcal{V}_h)$. \square

Lemma 4.12. *[Stabilität] Sei $w \in \mathbb{P}_{k+1}^{C^1}(I, \mathcal{V}_h)$, dann gilt*

$$B_n(w, \pi_{k-1} I_k^c w) \geq \frac{1}{2} \|w(t_n)\|_0^2 - \frac{1}{2} \|w(t_{n-1})\|_0^2 + \frac{1}{2} \|l_{n-1}\|_0^2 \|\zeta_n(t_{n-1})\|_0^2 + Q_n[\|\pi_{k-1} I_k^c w\|^2].$$

Dabei ist l_{n-1} wie in Lemma 4.5.

Beweis. Aus der Definition der Bilinearform B_n erhalten wir

$$B_n(w, \pi_{k-1} I_k^c w) = Q_n[(\partial_t w, \pi_{k-1} I_k^c w)] + Q_n[\langle A_h w, \pi_{k-1} I_k^c w \rangle]. \quad (4.48)$$

Im ersten Summanden auf der rechten Seite liegt das Produkt in $\mathbb{P}_{2k-1}(I_n)$. Somit ist die Gauß–Lobatto-Formel exakt. Dies ergibt $Q_n[(\partial_t w, \pi_{k-1} I_k^c w)] = \int_{I_n} (\partial_t w(t), \pi_{k-1} I_k^c w(t)) dt$. Weiterhin erhalten wir aus Lemma 4.5

$$Q_n[(\partial_t w, \pi_{k-1} I_k^c w)] = \int_{I_n} (\partial_t I_k^c w(t), \pi_{k-1} I_k^c w(t)) dt + \int_{I_n} (l_{n-1} \partial_t \zeta_n(t), \pi_{k-1} I_k^c w(t)) dt. \quad (4.49)$$

Einerseits liegt der Term $\partial_t I_k^c w$ in $\mathbb{P}_{k-1}(I_n, \mathcal{V}_h)$, dementsprechend stellt er eine zulässige Testfunktion für die L^2 -Projektion π_{k-1} dar, d.h. es gilt

$$\begin{aligned} \int_{I_n} (\partial_t I_k^c w(t), \pi_{k-1} I_k^c w(t)) dt &= \int_{I_n} (\partial_t I_k^c w(t), I_k^c w(t)) dt \\ &= \frac{1}{2} \int_{I_n} \partial_t \|I_k^c w(t)\|_0^2 dt \\ &= \frac{1}{2} \|I_k^c w(t_n)\|_0^2 - \frac{1}{2} \|I_k^c w(t_{n-1})\|_0^2 \\ &= \frac{1}{2} \|w(t_n)\|_0^2 - \frac{1}{2} \|w(t_{n-1})\|_0^2. \end{aligned} \quad (4.50)$$

Andererseits folgt aus Lemma 4.9 $\int_{I_n} ((l_{n-1} \partial_t \zeta_n(t), \pi_{k-1} I_k^c w(t))) = 0$, da $\pi_{k-1} I_k^c w$ in $\mathbb{P}_{k-1}(I_n, \mathcal{V}_h)$ liegt. Weiterhin haben wir

$$\begin{aligned} Q_n [\langle A_h w, \pi_{k-1} I_k^c w \rangle] &= Q_n [\langle A_h I_k^c w, \pi_{k-1} I_k^c w \rangle] + Q_n [\langle A_h l_{n-1} \zeta_n, \pi_{k-1} I_k^c w \rangle] \\ &= \int_{I_n} \langle A_h I_k^c w(t), \pi_{k-1} I_k^c w(t) \rangle dt, \end{aligned} \quad (4.51)$$

wobei wir erneut die Eigenschaft, dass ζ_n an den Gauß-Lobatto-Punkten verschwindet, ausgenutzt haben. Man beachte dabei, dass A_h nicht zeitabhängig ist. Da $\pi_{k-1} I_k^c w \in \mathbb{P}_{k-1}(I_n, \mathcal{V}_h)$ und somit eine zulässige Testfunktion für die L^2 -Projektion π_{k-1} ist, gilt

$$\begin{aligned} \int_{I_n} \langle A_h I_k^c w(t), \pi_{k-1} I_k^c w(t) \rangle dt &= \int_{I_n} \langle \pi_{k-1} A_h I_k^c w(t), \pi_{k-1} I_k^c w(t) \rangle dt \\ &= \int_{I_n} \langle A_h \pi_{k-1} I_k^c w(t), \pi_{k-1} I_k^c w(t) \rangle dt \\ &\geq \int_{I_n} \|\pi_{k-1} I_k^c w(t)\|^2 dt = Q_n [\|\pi_{k-1} I_k^c w\|^2]. \end{aligned} \quad (4.52)$$

Dabei haben wir die Koerzivität (4.9) des Operators A_h verwendet. Zusammengefasst ergibt sich aus der Kombination von (4.48)-(4.52) die Behauptung. \square

Als Zugang zur Fehlerabschätzung zerlegen wir den Fehler $\tilde{e}(t) := u(t) - \tilde{u}_{\tau h}(t)$, indem wir

$$\begin{aligned} \eta(t) &:= u(t) - \tilde{P}_h R_{k+1}^c u(t), \\ \tilde{e}_{\tau h}(t) &:= \tilde{P}_h R_{k+1}^c u(t) - \tilde{u}_{\tau h}(t) \end{aligned}$$

setzen. Es folgt daraus

$$\tilde{e} = \eta + \tilde{e}_{\tau h}. \quad (4.53)$$

Lemma 4.13. [Fehlerabschätzung] Es gilt

$$|B_n(\tilde{e}_{\tau h}, v_{\tau h})|^2 \leq C \tau_n \left\{ E_1^n(u) + E_2^n(u) + E_3^n(u) \right\} \left\{ Q_n [\|v_{\tau h}\|^2] \right\}, \quad (4.54)$$

wobei die Fehler im Ort durch

$$\begin{aligned} E_1^n(u) &:= h^{2r+1} \sup_{t \in I_n} \|u(t)\|_{r+1}^2, \\ E_2^n(u) &:= (\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2 \end{aligned}$$

gegeben sind, wohingegen der Fehler in der Zeit durch

$$E_3^n(u) := \tau_n^{2(k+2)} \left\{ \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + \sup_{t \in I_n} \|\partial_t^{k+3} u(t)\|_0^2 \right\}$$

gegeben ist.

Beweis. Nach Lemma 4.11 gilt

$$B_n(\tilde{u}_{\tau h}, v_{\tau h}) = Q_n[(f, v_{\tau h})] \quad \forall v_{\tau h} \in \mathbb{P}_{k-1}(I_n, \mathcal{V}_h).$$

Aus $A_h u = Au + S_h u$ folgt

$$\begin{aligned} B_n(u, v_{\tau h}) &= Q_n[\langle \partial_t u + A_h u, v_{\tau h} \rangle] \\ &= Q_n[\langle \partial_t u + Au + S_h u, v_{\tau h} \rangle] \\ &= Q_n[(f, v_{\tau h})] + Q_n[\langle S_h u, v_{\tau h} \rangle]. \end{aligned}$$

Zusammen ergibt sich

$$\begin{aligned} B_n(u - \tilde{u}_{\tau h}, v_{\tau h}) &= Q_n[(f, v_{\tau h})] + Q_n[\langle S_h u, v_{\tau h} \rangle] - Q_n[(f, v_{\tau h})] \\ &= Q_n[\langle S_h u, v_{\tau h} \rangle]. \end{aligned} \tag{4.55}$$

Aus der umgestellten Fehlergleichung (4.55) erhalten wir

$$B_n(\tilde{e}_{\tau h}, v_{\tau h}) = -B_n(\eta, v_{\tau h}) + Q_n[\langle S_h u, v_{\tau h} \rangle].$$

Um den Fehler $\tilde{e}_{\tau h}$ nach oben abzuschätzen, betrachten wir die Zerlegung der Terme auf der rechten Seite

$$\begin{aligned} \left| -B_n(\eta, v_{\tau h}) + Q_n[\langle S_h u, v_{\tau h} \rangle] \right| &\leq \left| Q_n[(\partial_t u - \partial_t R_{k+1}^c u, v_{\tau h})] \right| \\ &\quad + \left| Q_n[(\partial_t(R_{k+1}^c u - \tilde{P}_h R_{k+1}^c u), v_{\tau h})] \right| \\ &\quad + \left| Q_n[a_h((u - R_{k+1}^c u), v_{\tau h})] \right| \\ &\quad + \left| Q_n[a_h((R_{k+1}^c u - \tilde{P}_h R_{k+1}^c u), v_{\tau h})] \right| \\ &\quad + \left| Q_n[s_h(u, v_{\tau h})] \right| \\ &=: T_1 + T_2 + T_3 + T_4 + T_5. \end{aligned}$$

Im Folgenden werden die Terme $T_i, i = 1, \dots, 5$, einzeln abgeschätzt. Zunächst betrachten wir den Term T_1 . Mithilfe der Cauchy-Schwarz-Ungleichung erhalten wir

$$\left| Q_n[(\partial_t u - \partial_t R_{k+1}^c u), v_{\tau h}] \right| \leq \left\{ Q_n[\|\partial_t u - \partial_t R_{k+1}^c u\|_0^2] \right\}^{1/2} \left\{ Q_n[\|v_{\tau h}\|_0^2] \right\}^{1/2}.$$

Aus der Definition von R_{k+1}^c und J_{k+2}^c folgt

$$\begin{aligned} Q_n[\|\partial_t u - \partial_t R_{k+1}^c u\|_0^2] &= \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \|\partial_t u(t_{n,\mu}) - \partial_t R_{k+1}^c u(t_{n,\mu})\|_0^2 \\ &= \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \|\partial_t u(t_{n,\mu}) - \partial_t J_{k+2}^c u(t_{n,\mu})\|_0^2 \\ &\leq C \tau_n \tau_n^{2(k+2)} \sum_{\mu=0}^k \hat{\omega}_\mu \sup_{t \in I_n} \|\partial_t^{k+3} u(t)\|_0^2. \end{aligned}$$

Hierbei ergibt sich der letzte Schritt aus (4.33). Es folgt die Abschätzung

$$T_1 \leq C \tau_n^{1/2} \tau_n^{k+2} \sup_{t \in I_n} \|\partial_t^{k+3} u(t)\|_0 \left\{ Q_n[\|v_{\tau h}\|_0^2] \right\}^{1/2}.$$

In Bezug auf T_2 liefert die Cauchy–Schwarz-Ungleichung

$$\left| Q_n [(\partial_t(R_{k+1}^c u - \tilde{P}_h R_{k+1}^c u), v_{\tau h})] \right| \leq \left\{ Q_n \left[\|\partial_t(R_{k+1}^c u - \tilde{P}_h R_{k+1}^c u)\|_0^2 \right] \right\}^{1/2} \left\{ Q_n [\|v_{\tau h}\|_0^2] \right\}^{1/2},$$

wobei

$$\begin{aligned} Q_n \left[\|\partial_t(R_{k+1}^c u - \tilde{P}_h R_{k+1}^c u)\|_0^2 \right] &= \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \|\partial_t(R_{k+1}^c u(t_{n,\mu}) - \tilde{P}_h R_{k+1}^c u(t_{n,\mu}))\|_0^2 \\ &\leq \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu C h^{2r+2} \|\partial_t R_{k+1}^c u(t_{n,\mu})\|_{r+1}^2 \\ &\leq C \tau_n h^{2r+2} \sup_{t \in I_n} \|\partial_t R_{k+1}^c u(t)\|_{r+1}^2 \\ &\leq C \tau_n h^{2r+2} \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2. \end{aligned}$$

Dabei haben wir sowohl die Annahme 4.1 als auch die Fehlerabschätzung (4.39) verwendet. Daraus folgt

$$T_2 \leq C \tau_n^{1/2} h^{r+1} \|u\|_{C^1(I_n, H^{r+1}(\Omega))} \left\{ Q_n [\|v_{\tau h}\|_0^2] \right\}^{1/2}.$$

Bei der Abschätzung von T_3 gilt Folgendes:

$$\begin{aligned} \left| Q_n [(a_h(u - R_{k+1}^c u), v_{\tau h})] \right| &= \left| \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \langle A_h(R_{k+1}^c u(t_{n,\mu}) - u(t_{n,\mu})), v_{\tau h}(t_{n,\mu}) \rangle \right| \\ &\leq C \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \|R_{k+1}^c u(t_{n,\mu}) - u(t_{n,\mu})\|_1 \|v_{\tau h}(t_{n,\mu})\|. \end{aligned}$$

Hierbei wurde die Stetigkeit von A_h ausgenutzt, siehe (4.11). Aus (4.36) und der Cauchy–Schwarz-Ungleichung folgt

$$\begin{aligned} T_3 &\leq C \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \tau_n^{k+2} \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1 \|v_{\tau h}(t_{n,\mu})\| \\ &\leq C \left\{ \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \tau_n^{2(k+2)} \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 \right\}^{1/2} \left\{ Q_n [\|v_{\tau h}\|^2] \right\}^{1/2} \\ &\leq C \tau_n^{1/2} \tau_n^{k+2} \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1 \left\{ Q_n [\|v_{\tau h}\|^2] \right\}^{1/2}. \end{aligned}$$

Weiterhin wird T_4 wie folgt abgeschätzt. Aus Korollar 4.4 und der Stetigkeit von A_h folgt

$$\begin{aligned} \left| Q_n [a_h((\tilde{P}_h R_{k+1}^c u - R_{k+1}^c u), v_{\tau h})] \right| &= \left| \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \langle A_h(\tilde{P}_h R_{k+1}^c u(t_{n,\mu}) - R_{k+1}^c u(t_{n,\mu})), v_{\tau h}(t_{n,\mu}) \rangle \right| \\ &\leq C \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu (\varepsilon^{1/2} + h^{1/2}) h^r \|R_{k+1}^c u(t_{n,\mu})\|_{r+1} \|v_{\tau h}(t_{n,\mu})\|. \end{aligned}$$

Hierbei entsteht der letzte Schritt durch die Fehlerabschätzung von \tilde{P}_h im Korollar 4.4. Es folgt aus der Cauchy–Schwarz-Ungleichung weiterhin

$$\begin{aligned} T_4 &\leq C \left\{ \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu (\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \|R_{k+1}^c u(t_{n,\mu})\|_{r+1}^2 \right\}^{1/2} \left\{ \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \|v_{\tau h}(t_{n,\mu})\|^2 \right\}^{1/2} \\ &\leq C \tau_n^{1/2} (\varepsilon^{1/2} + h^{1/2}) h^r \sup_{t \in I_n} \|R_{k+1}^c u(t)\|_{r+1} \left\{ Q_n [\|v_{\tau h}\|^2] \right\}^{1/2}. \end{aligned}$$

Anhand der Stabilität von R_{k+1}^c bezüglich der H^1 -Norm im Ort und der L^∞ -Norm in der Zeit, ergibt sich aus (4.37)

$$T_4 \leq C \tau_n^{1/2} (\varepsilon^{1/2} + h^{1/2}) h^r \|u(t)\|_{C^1(I_n, H^{r+1}(\Omega))} \left\{ Q_n [\|v_{\tau h}\|^2] \right\}^{1/2}.$$

Schließlich wird die Abschätzung von T_5 betrachtet. Hierbei liefert die Cauchy–Schwarz-Ungleichung

$$\left| Q_n [s_h(u, v_{\tau h})] \right| \leq \left\{ Q_n [s_h(u, u)] \right\}^{1/2} \left\{ Q_n [s_h(v_{\tau h}, v_{\tau h})] \right\}^{1/2}.$$

Wir erhalten mithilfe des definierten Stabilisierungsterms

$$\begin{aligned} Q_n [s_h(u, u)] &= \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu s_h(u(t_{n,\mu}), u(t_{n,\mu})) \\ &= \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \left(\sum_K \beta_K \|\kappa_K(\nabla u(t_{n,\mu}))\|_{0,K}^2 \right). \end{aligned}$$

Aus $\beta_K = Ch_K \leq Ch$ und der Annahme 4.2 folgt

$$\begin{aligned} Q_n [s_h(u, u)] &\leq C \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu h^{2r+1} \|u(t_{n,\mu})\|_{r+1,K}^2 \\ &\leq C \tau_n h^{2r+1} \sup_{t \in I_n} \|u(t)\|_{r+1,K}^2. \end{aligned}$$

Dies ergibt die folgende Abschätzung für T_5 :

$$T_5 \leq C \tau_n^{1/2} h^{r+1/2} \sup_{t \in I_n} \|u(t)\|_{r+1} \left\{ Q_n [s_h(v_{\tau h}, v_{\tau h})] \right\}^{1/2}.$$

Abschließend erhält man durch die jeweiligen Teilabschätzungen die Behauptung. Dabei ergeben sich die Terme E_1^n aus der Abschätzung von T_5 , E_2^n aus T_2 und T_4 und zuletzt ergibt sich E_3^n aus der Kombination der Fehlerabschätzungen in der Zeit von T_1 und T_3 . Man beachte, dass $\|v_{\tau h}\|_0 \leq C \|v_{\tau h}\|^2$ bzw. $Q_n [s_h(v_{\tau h}, v_{\tau h})] \leq Q_n [\|v_{\tau h}\|^2]$ für alle $v_{\tau h}$ in $\mathbb{P}_k^c(I, \mathcal{V}_h)$ und die Gauß–Lobatto-Formel für Polynome vom Grad höchstens $(2k-1)$ exakt ist. Daraus kann man entnehmen, dass

$$Q_n [\|v_{\tau h}\|^2] = \int_{I_n} \|v_{\tau h}(t)\|^2 dt$$

gilt. Somit ist der Beweis vollendet. \square

Lemma 4.14. *[Fehlerabschätzung] Es gilt für $m = 1, \dots, N$*

$$\|\tilde{e}_{\tau h}(t_m)\|_0^2 + \sum_{n=1}^m Q_n [\|\pi_{k-1} I_k^c \tilde{e}_{\tau h}\|^2] \leq C \sum_{n=1}^m \tau_n \left\{ E_1^n(u) + E_2^n(u) + E_3^n(u) \right\}, \quad (4.56)$$

wobei E_i^n , $i = 1, \dots, 3$, dem Lemma 4.13 zu entnehmen sind.

Beweis. Man setze $w = \tilde{e}_{\tau h}$ in der Stabilitätsungleichung in Lemma 4.12, $v_{\tau h} = \pi_{k-1} I_k^c \tilde{e}_{\tau h}$ in der Fehlerabschätzung in Lemma 4.13 und kombiniere die Abschätzungen in beiden Lemmata, dann folgt

$$\begin{aligned} \frac{1}{2} \|\tilde{e}_{\tau h}(t_n)\|_0^2 - \frac{1}{2} \|\tilde{e}_{\tau h}(t_{n-1})\|_0^2 + Q_n [\|\pi_{k-1} I_k^c \tilde{e}_{\tau h}\|^2] \\ \leq C\tau_n \left\{ E_1^n(u) + E_2^n(u) + E_3^n(u) \right\}^{1/2} \left\{ Q_n [\|\pi_{k-1} I_k^c \tilde{e}_{\tau h}\|^2] \right\}^{1/2}. \end{aligned}$$

Aus der Young-Ungleichung erhalten wir

$$\begin{aligned} \frac{1}{2} \|\tilde{e}_{\tau h}(t_n)\|_0^2 - \frac{1}{2} \|\tilde{e}_{\tau h}(t_{n-1})\|_0^2 + \frac{1}{2} Q_n [\|\pi_{k-1} I_k^c \tilde{e}_{\tau h}\|^2] \\ \leq C\tau_n \left\{ E_1^n(u) + E_2^n(u) + E_3^n(u) \right\}. \end{aligned} \quad (4.57)$$

Weiterhin liefert das Aufsummieren der Abschätzung (4.57) für $n = 1, \dots, m$ die Behauptung (4.56), wobei wir die Eigenschaft $\tilde{P}_h u_0 = u_{\tau h}(0)$ ausgenutzt haben. \square

Lemma 4.15. [*L²-Fehlerabschätzung*] Es gilt für $m = 1, \dots, N$,

$$\|\tilde{e}(t_m)\|_0^2 \leq C\tau_m \max_{1 \leq n \leq m} \left\{ C_{T,n}(u) \tau_n^{2(k+2)} + C_{S,n}(u) (\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \right\} + C'_{S,m}(u) h^{2r+2},$$

wobei

$$\begin{aligned} C_{T,n}(u) &:= \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + \sup_{t \in I_n} \|\partial_t^{k+3} u(t)\|_0^2, \\ C_{S,n}(u) &:= \sup_{t \in I_n} \|u(t)\|_{r+1}^2 + \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2, \\ C'_{S,m}(u) &:= \|u(t_m)\|_{r+1}^2. \end{aligned}$$

Beweis. Unter der Berücksichtigung des Fehlersplittings (4.53) betrachten wir die folgenden Abschätzungen für η . Es folgt aus der Annahme 4.1 für $m = 1, \dots, N$

$$\begin{aligned} \|\eta(t_m)\|_0^2 &= \|u(t_m) - \tilde{P}_h R_{k+1}^c u(t_m)\|_0^2 \leq \|u(t_m) - \tilde{P}_h u(t_m)\|_0^2 \\ &\leq Ch^{2r+2} \|u(t_m)\|_{r+1}^2, \end{aligned}$$

dabei wurde in Betracht gezogen, dass $R_{k+1}^c u$ und u in allen Punkten $t_m, m = 1, \dots, N$, übereinstimmen, siehe Unterabschnitt 4.2.2. Dies kombiniert mit der Abschätzung (4.56) liefert die Behauptung. \square

4.2.7 Fehlerabschätzung für die Zeitableitung

Aus dem Postprocessing für die cGP-Verfahren kann man entnehmen, dass die nachbereitete diskrete Lösung $\tilde{u}_{\tau h}$ in $C^1(I, \mathcal{V}_h) \cap \mathbb{P}_{k+1}^c(I, \mathcal{V}_h)$ liegt, siehe Unterabschnitt 3.6.1 in Kapitel 3. Darausfolgend ist die Zeitableitung von $\tilde{u}_{\tau h}$ über I wohldefiniert. Es gilt $\partial_t \tilde{u}_{\tau h}(t_n) \in \mathbb{P}_k^c(I, \mathcal{V}_h)$.

Lemma 4.16. Sei $u_{\tau h}$ die volldiskrete Lösung und $\tilde{u}_{\tau h}$ die durch das Postprocessing nachbearbeitete Lösung, dann gilt

$$\partial_t \tilde{u}_{\tau h}(t) + A_h u_{\tau h}(t) = P_h I_k^c f(t) \quad \forall t \in I,$$

wobei I_k^c der in (4.23) definierte Gauß-Lobatto-Zeitinterpolator ist.

Beweis. Aus Lemma 4.10 wissen wir für alle $n = 1, \dots, N$

$$\int_{I_n} \langle \partial_t \tilde{u}_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt = Q_n[(f, v_{\tau h})]$$

für alle $v_{\tau h} \in \mathbb{P}_{k-1}(I_n, \mathcal{V}_h)$. Die Quadraturformel Q_n wertet die zu integrierende Funktion nur an den Gauß-Lobatto-Punkten aus. Es gilt daher

$$Q_n[(f, v_{\tau h})] = Q_n[(I_k^c f, v_{\tau h})].$$

Desweiteren liegt das Produkt $(I_k^c f, v_{\tau h})$ in $\mathbb{P}_{2k-1}(I_n)$. Wir erhalten somit

$$\begin{aligned} \int_{I_n} \langle \partial_t \tilde{u}_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt &= Q_n[(I_k^c f, v_{\tau h})] = \int_{I_n} (I_k^c f(t), v_{\tau h}(t)) dt \\ &= \int_{I_n} (P_h I_k^c f(t), v_{\tau h}(t)) dt = Q_n[(P_h I_k^c f, v_{\tau h})], \end{aligned} \quad (4.58)$$

wobei der vorletzte Schritt aus den Eigenschaften der L^2 -Projektion P_h folgt. Wir wählen nun eine spezielle Testfunktion $v_{\tau h}$. Sei $l \in \{0, \dots, k-1\}$ und sei $v_{\tau h}(t_{n,\mu}) := 0$ für $\mu \in \{0, \dots, k-1\} \setminus \{l\}$. Es folgt einerseits

$$Q_n[(P_h I_k^c f, v_{\tau h})] = \frac{\tau_n}{2} [\hat{\omega}_l (P_h I_k^c f(t_{n,l}), v_{\tau h}(t_{n,l})) + \hat{\omega}_k (P_h I_k^c f(t_n), v_{\tau h}(t_n))], \quad (4.59)$$

wobei $t_n = t_{n,k}$ und andererseits erhalten wir

$$\begin{aligned} \int_{I_n} \langle \partial_t \tilde{u}_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt &= Q_n[\langle \partial_t \tilde{u}_{\tau h} + A_h u_{\tau h}, v_{\tau h} \rangle] \\ &= \frac{\tau_n}{2} [\hat{\omega}_l \langle \partial_t \tilde{u}_{\tau h}(t_{n,l}) + A_h u_{\tau h}(t_{n,l}), v_{\tau h} \rangle \\ &\quad + \hat{\omega}_k \langle \partial_t \tilde{u}_{\tau h}(t_n) + A_h u_{\tau h}(t_n), v_{\tau h} \rangle]. \end{aligned} \quad (4.60)$$

Dabei wurde ausgenutzt, dass $\langle \partial_t \tilde{u}_{\tau h} + A_h u_{\tau h}, v_{\tau h} \rangle \in \mathbb{P}_{2k-1}(I_n)$ mit der Gauß-Lobatto-Formel exakt integriert wird. Nach Konstruktion von $\tilde{u}_{\tau h}$ gilt

$$\langle \partial_t \tilde{u}_{\tau h}(t_n) + A_h u_{\tau h}(t_n), v_{\tau h} \rangle = (f(t_n), v_{\tau h}) = (P_h I_k^c f(t_n), v_{\tau h}).$$

Dies in (4.58) eingesetzt und mit (4.59) und (4.60) kombiniert liefert

$$\langle \partial_t \tilde{u}_{\tau h}(t_{n,l}) + A_h u_{\tau h}(t_{n,l}), v_{\tau h} \rangle = (P_h I_k^c f(t_{n,l}), v_{\tau h}) \quad \forall l \in \{0, \dots, k-1\}.$$

Somit stimmen die \mathbb{P}_k -Polynome $P_h I_k^c f$ und $\partial_t \tilde{u}_{\tau h} + A_h u_{\tau h}$ in den $(k+1)$ Punkten $t_{n,l}, l = 0, \dots, k$, überein, infolgedessen müssen sie identisch sein. Somit ist die Behauptung auf das ganze Intervall I gültig. Es gilt

$$\langle \partial_t \tilde{u}_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h} \rangle = (P_h I_k^c f(t), v_{\tau h}) \quad \forall t \in I.$$

□

Die Bilinearform B_n erfüllt die nachfolgende Abschätzung:

Lemma 4.17. *Sei $w \in \mathbb{P}_k(I_n, \mathcal{V}_h)$. Es gilt*

$$B_n(w, \pi_{k-1} w) \geq \frac{1}{2} \|w(t_n)\|_0^2 - \frac{1}{2} \|w(t_{n-1})\|_0^2 + \int_{I_n} \|\pi_{k-1} w(t)\|^2 dt.$$

Beweis. Aus der Definition von B_n erhalten wir

$$B_n(w, \pi_{k-1}w) = Q_n[(\partial_t w, \pi_{k-1}w)] + Q_n[a_h(w, \pi_{k-1}w)].$$

In den beiden Quadraturformeln liegt der Integrand in $\mathbb{P}_{2k-1}(I_n)$, daher lassen sich die Terme als Integrale schreiben. Wir erhalten für den ersten Term

$$\begin{aligned} Q_n[(\partial_t w, \pi_{k-1}w)] &= \int_{I_n} (\partial_t w(t), \pi_{k-1}w(t)) dt = \int_{I_n} (\partial_t w(t), w(t)) dt = \frac{1}{2} \int_{I_n} \frac{d}{dt} \|w(t)\|_0^2 dt \\ &= \frac{1}{2} \|w(t_n)\|_0^2 - \frac{1}{2} \|w(t_{n-1})\|_0^2. \end{aligned}$$

Dabei wurde ausgenutzt, dass $\partial_t w \in \mathbb{P}_{k-1}(I_n, \mathcal{V}_h)$ eine zulässige Testfunktion für die L^2 -Projektion π_{k-1} ist. Weiterhin gilt

$$\begin{aligned} Q_n[a_h(w, \pi_{k-1}w)] &= \int_{I_n} \langle A_h w(t), \pi_{k-1}w(t) \rangle dt = \int_{I_n} \langle \pi_{k-1} A_h w(t), \pi_{k-1}w(t) \rangle dt \\ &= \int_{I_n} \langle A_h \pi_{k-1}w(t), \pi_{k-1}w(t) \rangle dt \\ &\geq \int_{I_n} \|\pi_{k-1}w(t)\|^2 dt. \end{aligned}$$

Wir haben die Zeitunabhängigkeit von A_h verwendet. Die beiden Ergebnisse liefern zusammen die Aussage des Lemmas. \square

Herleitung einer Fehlergleichung für die Zeitableitung:

Den Fehler bei der Zeitableitung werden wir wie folgt zerlegen. Wir setzen

$$\begin{aligned} \hat{e}(t) &:= \partial_t u(t) - \partial_t \tilde{u}_{\tau h}(t), \\ \hat{\eta}(t) &:= \partial_t u(t) - \tilde{P}_h \partial_t R_{k+1}^c u(t), \\ \hat{e}_{\tau h}(t) &:= \tilde{P}_h \partial_t R_{k+1}^c u(t) - \partial_t \tilde{u}_{\tau h}(t), \end{aligned}$$

somit erhalten wir

$$\hat{e} = \hat{\eta} + \hat{e}_{\tau h}. \quad (4.61)$$

Aus der Definition der Bilinearform B_n folgt

$$\begin{aligned} B_n(\partial_t u, v_{\tau h}) &= Q_n[\langle \partial_t^2 u + A_h \partial_t u, v_{\tau h} \rangle] \\ &= Q_n[\langle \partial_t(\partial_t u + A_h u), v_{\tau h} \rangle]. \end{aligned}$$

Aus $A_h u = Au + S_h u$ folgt

$$\begin{aligned} B_n(\partial_t u, v_{\tau h}) &= Q_n[\langle \partial_t(\partial_t u + Au + S_h u), v_{\tau h} \rangle] \\ &= Q_n[\langle \partial_t(f + S_h u), v_{\tau h} \rangle] \\ &= Q_n[\langle \partial_t f, v_{\tau h} \rangle] + Q_n[\langle S_h \partial_t u, v_{\tau h} \rangle]. \end{aligned} \quad (4.62)$$

Hierbei wurde ausgenutzt, dass A bzw. S_h von der Zeit unabhängig ist und u die Differentialgleichung (4.7) erfüllt. Andererseits haben wir

$$\begin{aligned} B_n(\partial_t \tilde{u}_{\tau h}, v_{\tau h}) &= Q_n[\langle \partial_t^2 \tilde{u} + A_h \partial_t \tilde{u}, v_{\tau h} \rangle] \\ &= Q_n[\langle \partial_t(\partial_t \tilde{u}_{\tau h} + A_h(u_{\tau h} + c_n \zeta_n)), v_{\tau h} \rangle] \\ &= Q_n[\langle \partial_t(\partial_t \tilde{u}_{\tau h} + A_h u_{\tau h}), v_{\tau h} \rangle] + Q_n[\langle A_h c_n \partial_t \zeta_n, v_{\tau h} \rangle] \\ &= Q_n[\langle \partial_t P_h I_k^c f, v_{\tau h} \rangle], \end{aligned}$$

wobei wir (4.47) ausgenutzt und die Beziehung aus Lemma 4.16 nach der Zeit differenziert und verwendet haben. Die Terme $I_k^c f$ und $L_{k+1}^c f$ unterscheiden sich nur um ein Vielfaches von $\zeta_n \in \mathbb{P}_{k+1}$. Es gilt

$$\begin{aligned} Q_n [(\partial_t P_h I_k^c f, v_{\tau h})] &= Q_n [(\partial_t I_k^c f, v_{\tau h})] \\ &= Q_n [(\partial_t L_{k+1}^c f, v_{\tau h})] - Q_n [(\partial_t \zeta_n \tilde{l}_{n-1}, v_{\tau h})] \\ &= Q_n [(\partial_t L_{k+1}^c f, v_{\tau h})]. \end{aligned}$$

Der letzte Schritt folgt aus Lemma 4.9. Zusammengefasst ergibt sich

$$B_n(\partial_t(u - \tilde{u}_{\tau h}), v_{\tau h}) = Q_n [(\partial_t f, v_{\tau h})] + Q_n [\langle S_h \partial_t u, v_{\tau h} \rangle] - Q_n [(\partial_t L_{k+1}^c f, v_{\tau h})].$$

Das Umstellen der Fehlergleichung (4.61) liefert

$$B_n(\hat{e}_{\tau h}, v_{\tau h}) = -B_n(\hat{\eta}, v_{\tau h}) + Q_n [\langle S_h \partial_t u, v_{\tau h} \rangle] + Q_n [(\partial_t(f - L_{k+1}^c f), v_{\tau h})]. \quad (4.63)$$

Im Folgenden betrachten wir für die Zeitableitung die gekoppelte Fehlerabschätzung im Ort und in der Zeit.

Lemma 4.18. *[Fehlerabschätzung für die Zeitableitung] Es gilt*

$$|B_n(\hat{e}_{\tau h}, v_{\tau h})|^2 \leq C\tau_n \left\{ \hat{E}_1^n(u) + \hat{E}_2^n(u) + \hat{E}_3^n(u) + \hat{E}_4^n(f) \right\} Q_n [\|v_{\tau h}\|^2],$$

wobei die Fehler im Ort durch

$$\begin{aligned} \hat{E}_1^n(u) &:= h^{2r+1} \sup_{t \in I_n} \|\partial_t u\|_{r+1}^2 + h^{2r+1} \sup_{t \in I_n} \|\partial_t^2 u\|_{r+1}^2, \\ \hat{E}_2^n(u) &:= (\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2 \end{aligned}$$

gegeben sind, während die Fehler in der Zeit durch

$$\begin{aligned} \hat{E}_3^n(u) &:= \tau_n^{2(k+1)} \left\{ \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + \sup_{t \in I_n} \|\partial_t^{k+3} u(t)\|_1^2 \right\}, \\ \hat{E}_4^n(f) &:= \tau_n^{2(k+1)} \sup_{t \in I_n} \|\partial_t^{k+2} f(t)\|_0^2 \end{aligned}$$

gegeben sind.

Beweis. Um den Fehler nach oben abzuschätzen, betrachten wir die Zerlegung der Terme auf der rechten Seite

$$\begin{aligned} |B_n(\hat{e}_{\tau h}, v_{\tau h})| &\leq \left| -B_n(\hat{\eta}, v_{\tau h}) \right| + \left| Q_n [s_h(\partial_t u, v_{\tau h})] \right| + \left| Q_n [(\partial_t(f - L_{k+1}^c f), v_{\tau h})] \right| \\ &\leq \left| Q_n [(\partial_t(\partial_t u - \tilde{P}_h \partial_t R_{k+1}^c u), v_{\tau h})] \right| + \left| Q_n [\langle A_h(\partial_t u - \partial_t R_{k+1}^c u), v_{\tau h} \rangle] \right| \\ &\quad + \left| Q_n [\langle A_h(\partial_t R_{k+1}^c u - \tilde{P}_h \partial_t R_{k+1}^c u), v_{\tau h} \rangle] \right| + \left| Q_n [s_h(\partial_t u, v_{\tau h})] \right| \\ &\quad + \left| Q_n [(\partial_t(f - L_{k+1}^c f), v_{\tau h})] \right| \\ &=: T_1 + T_2 + T_3 + T_4 + T_5. \end{aligned}$$

Im Folgenden werden wir die Terme $T_i, i = 1, \dots, 5$, einzeln abschätzen. Zunächst einmal betrachten wir den Term T_5 . Es gilt

$$\left| Q_n [(\partial_t(f - L_{k+1}^c f), v_{\tau h})] \right| \leq \left\{ Q_n [\|\partial_t(f - L_{k+1}^c f)\|_0^2] \right\}^{1/2} \left\{ Q_n [\|v_{\tau h}\|_0^2] \right\}^{1/2}.$$

Aus der Fehlerabschätzung (4.40) von L_{k+1}^c folgt

$$\begin{aligned} Q_n [\|\partial_t(f - L_{k+1}^c f)\|_0^2] &= \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \|\partial_t(f(t_{n,\mu}) - L_{k+1}^c f(t_{n,\mu}))\|_0^2 \\ &\leq \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu C \tau_n^{2(k+1)} \sup_{t \in I_n} \|\partial_t^{k+2} f(t)\|_0^2 \\ &\leq C \tau_n \tau_n^{2(k+1)} \sup_{t \in I_n} \|\partial_t^{k+2} f(t)\|_0^2, \end{aligned}$$

somit erhalten wir

$$T_5 \leq C \tau_n^{1/2} \tau_n^{k+1} \sup_{t \in I_n} \|\partial_t^{k+2} f(t)\|_0 \left\{ Q_n [\|v_{\tau h}\|_0^2] \right\}^{1/2}.$$

In Bezug auf T_1 betrachten wir die Umformulierung

$$\partial_t(\partial_t u - \tilde{P}_h \partial_t R_{k+1}^c u) = \partial_t^2 u - \tilde{P}_h \partial_t^2 u + \tilde{P}_h \partial_t^2 u - \tilde{P}_h \partial_t^2 J_{k+2}^c u + \tilde{P}_h \partial_t^2 J_{k+2}^c u - \tilde{P}_h \partial_t^2 R_{k+1}^c u.$$

Darauf aufbauend haben wir

$$\left| Q_n [(\partial_t^2 u - \tilde{P}_h \partial_t^2 u, v_{\tau h})] \right| \leq \left\{ Q_n [\|\partial_t^2 u - \tilde{P}_h \partial_t^2 u\|_0^2] \right\}^{1/2} \left\{ Q_n [\|v_{\tau h}\|_0^2] \right\}^{1/2},$$

wobei

$$\begin{aligned} Q_n [\|\partial_t^2 u - \tilde{P}_h \partial_t^2 u\|_0^2] &= \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \|\partial_t^2 u(t_{n,\mu}) - \tilde{P}_h \partial_t^2 u(t_{n,\mu})\|_0^2 \\ &\leq \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu C h^{2r+2} \sup_{t \in I_n} \|\partial_t^2 u(t)\|_{r+1}^2 \\ &\leq C \tau_n h^{2r+2} \sup_{t \in I_n} \|\partial_t^2 u(t)\|_{r+1}^2. \end{aligned}$$

Wir haben hier die Fehlerabschätzung aus der Annahme 4.1 verwendet.

Das Produkt $(\tilde{P}_h \partial_t^2 J_{k+2}^c u - \tilde{P}_h \partial_t^2 R_{k+1}^c u, v_{\tau h})$ liegt in $\mathbb{P}_{2k-1}(I_n)$, so kann die Quadraturformel durch das Integral ersetzt werden. Aus einer partiellen Integration folgt

$$\begin{aligned} &Q_n [(\tilde{P}_h \partial_t^2 J_{k+2}^c u - \tilde{P}_h \partial_t^2 R_{k+1}^c u, v_{\tau h})] \\ &= \int_{I_n} (\tilde{P}_h \partial_t^2 J_{k+2}^c u(t) - \tilde{P}_h \partial_t^2 R_{k+1}^c u(t), v_{\tau h}(t)) dt \\ &= - \int_{I_n} (\tilde{P}_h \partial_t J_{k+2}^c u(t) - \tilde{P}_h \partial_t R_{k+1}^c u(t), \partial_t v_{\tau h}(t)) dt + [(\tilde{P}_h \partial_t J_{k+2}^c u - \tilde{P}_h \partial_t R_{k+1}^c u, v_{\tau h})]_{t_{n-1}}^{t_n} \\ &= -Q_n [(\tilde{P}_h \partial_t J_{k+2}^c u - \tilde{P}_h \partial_t R_{k+1}^c u, \partial_t v_{\tau h})] + [(\tilde{P}_h \partial_t J_{k+2}^c u - \tilde{P}_h \partial_t R_{k+1}^c u, v_{\tau h})]_{t_{n-1}}^{t_n} \\ &= 0. \end{aligned}$$

Der letzte Schritt folgt unmittelbar aus der Definition von R_{k+1}^c über J_{k+2}^c , dabei stimmen die Zeitableitungen der beiden Operatoren in allen Gauß–Lobatto-Punkten überein. Weiterhin folgt aus der Cauchy–Schwarz-Ungleichung

$$\left| Q_n [(\tilde{P}_h \partial_t^2 u - \tilde{P}_h \partial_t^2 J_{k+2}^c u, v_{\tau h})] \right| \leq \left\{ Q_n [\|\tilde{P}_h \partial_t^2 u - \tilde{P}_h \partial_t^2 J_{k+2}^c u\|_0^2] \right\}^{1/2} \left\{ Q_n [\|v_{\tau h}\|_0^2] \right\}^{1/2}.$$

Der Interpolationsoperator \tilde{P}_h lässt sich mit der Zeitableitung vertauschen, da er zeitunabhängig ist. Daher erhalten wir anhand der Fehlerabschätzung (4.34) von J_{k+2}^c Folgendes:

$$\begin{aligned}
 Q_n \left[\left\| \tilde{P}_h \partial_t^2 u - \tilde{P}_h \partial_t^2 J_{k+2}^c \right\|_0^2 \right] &= \left[\left\| \partial_t^2 (\tilde{P}_h u) - \partial_t^2 J_{k+2}^c (\tilde{P}_h u) \right\|_0^2 \right] \\
 &= \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \left\| (\partial_t^2 (\tilde{P}_h u) - \partial_t^2 J_{k+2}^c (\tilde{P}_h u))(t_{n,\mu}) \right\|_0^2 \\
 &\leq \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu C \tau_n^{2(k+1)} \left\| (\tilde{P}_h u(t))^{(k+3)} \right\|_0^2 \\
 &\leq C \tau_n \tau_n^{2(k+1)} \sup_{t \in I_n} \left\| \tilde{P}_h \partial_t^{k+3} u(t) \right\|_0^2 \\
 &\leq C \tau_n \tau_n^{2(k+1)} \sup_{t \in I_n} \left\| \partial_t^{k+3} u(t) \right\|_1^2.
 \end{aligned}$$

Der letzte Schritt folgt aus $\|\tilde{P}_h w\|_0 \leq \|\tilde{P}_h w - w\|_0 + \|w\|_0 \leq Ch\|w\|_1 + \|w\|_0 \leq C\|w\|_1$, vgl. Annahme 4.1, daher gilt

$$\begin{aligned}
 T_1 &\leq C \tau_n^{1/2} \tau_n^{k+1} \sup_{t \in I_n} \left\| \partial_t^{k+3} u(t) \right\|_1 \left\{ Q_n \left[\left\| v_{\tau h} \right\|_0^2 \right] \right\}^{1/2} \\
 &\quad + C \tau_n^{1/2} h^{r+1} \sup_{t \in I_n} \left\| \partial_t^2 u(t) \right\|_{r+1} \left\{ Q_n \left[\left\| v_{\tau h} \right\|_0^2 \right] \right\}^{1/2}.
 \end{aligned}$$

Die restlichen Terme lassen sich wie im Beweis von Lemma 4.13 behandeln. Bei der Abschätzung von T_2 gilt

$$\begin{aligned}
 &\left| Q_n \left[\langle A_h (\partial_t u - \partial_t R_{k+1}^c u), v_{\tau h} \rangle \right] \right| \\
 &= \left| \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \langle A_h (\partial_t u(t_{n,\mu}) - \partial_t R_{k+1}^c u(t_{n,\mu})), v_{\tau h}(t_{n,\mu}) \rangle \right| \\
 &\leq C \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \left\| \partial_t u(t_{n,\mu}) - \partial_t R_{k+1}^c u(t_{n,\mu}) \right\|_1 \|v_{\tau h}(t_{n,\mu})\|.
 \end{aligned}$$

Dabei wurde die Stetigkeit des Operators A_h ausgenutzt, siehe (4.11). Aus (4.38) und der Cauchy-Schwarz-Ungleichung folgt

$$\begin{aligned}
 T_2 &\leq C \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \tau_n^{k+1} \sup_{t \in I_n} \left\| \partial_t^{k+2} u(t) \right\|_1 \|v_{\tau h}(t_{n,\mu})\| \\
 &\leq C \left\{ \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \tau_n^{2(k+1)} \sup_{t \in I_n} \left\| \partial_t^{k+2} u(t) \right\|_1^2 \right\}^{1/2} \left\{ Q_n \left[\left\| v_{\tau h} \right\|^2 \right] \right\}^{1/2} \\
 &\leq C \tau_n^{1/2} \tau_n^{k+1} \sup_{t \in I_n} \left\| \partial_t^{k+2} u(t) \right\|_1 \left\{ Q_n \left[\left\| v_{\tau h} \right\|^2 \right] \right\}^{1/2}.
 \end{aligned}$$

Weiterhin wird T_3 wie folgt abgeschätzt:

$$\begin{aligned}
 &\left| Q_n \left[\langle A_h (\tilde{P}_h \partial_t R_{k+1}^c u - \partial_t R_{k+1}^c u), v_{\tau h} \rangle \right] \right| \\
 &= \left| \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \langle A_h (\tilde{P}_h \partial_t R_{k+1}^c u(t_{n,\mu}) - \partial_t R_{k+1}^c u(t_{n,\mu})), v_{\tau h}(t_{n,\mu}) \rangle \right| \\
 &\leq C \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu (\varepsilon^{1/2} + h^{1/2}) h^r \left\| \partial_t R_{k+1}^c u(t_{n,\mu}) \right\|_{r+1} \|v_{\tau h}(t_{n,\mu})\|.
 \end{aligned}$$

Hierbei entsteht der letzte Schritt durch die Fehlerabschätzung aus Korollar 4.4. Es folgt aus der Cauchy–Schwarz-Ungleichung weiterhin

$$\begin{aligned} T_3 &\leq C \left\{ \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu (\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \|\partial_t R_{k+1}^c u(t_{n,\mu})\|_{r+1}^2 \right\}^{1/2} \left\{ Q_n [\|v_{\tau h}\|^2] \right\}^{1/2} \\ &\leq C \tau_n^{1/2} (\varepsilon^{1/2} + h^{1/2}) h^r \sup_{t \in I_n} \|\partial_t R_{k+1}^c u(t)\|_{r+1} \left\{ Q_n [\|v_{\tau h}\|^2] \right\}^{1/2} \\ &\leq C \tau_n^{1/2} (\varepsilon^{1/2} + h^{1/2}) h^r \sup_{t \in I_n} \|\partial_t R_{k+1}^c u(t)\|_{r+1} \left\{ Q_n [\|v_{\tau h}\|^2] \right\}^{1/2}. \end{aligned}$$

Anhand der Stabilität von R_{k+1}^c bezüglich der H^1 -Norm im Ort und der L^∞ -Norm in der Zeit, vgl. Lemma 4.7, ergibt sich aus

$$\sup_{t \in I_n} \|\partial_t R_{k+1}^c u(t)\|_{r+1} \leq C \|u\|_{C^1(I_n, H^1(\Omega))}$$

die Abschätzung

$$T_3 \leq C \tau_n^{1/2} (\varepsilon^{1/2} + h^{1/2}) h^r \|u\|_{C^1(I_n, H^{r+1}(\Omega))} \left\{ Q_n [\|v_{\tau h}\|^2] \right\}^{1/2}.$$

Schließlich wird die Abschätzung von T_4 betrachtet. Hierbei liefert die Cauchy–Schwarz-Ungleichung

$$\left| Q_n [s_h(\partial_t u, v_{\tau h})] \right| \leq \left\{ Q_n [s_h(\partial_t u, \partial_t u)] \right\}^{1/2} \left\{ Q_n [s_h(v_{\tau h}, v_{\tau h})] \right\}^{1/2}.$$

Mithilfe des definierten Stabilisierungsterms erhalten wir

$$\begin{aligned} Q_n [s_h(\partial_t u, \partial_t u)] &= \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu s_h(\partial_t u(t_{n,\mu}), \partial_t u(t_{n,\mu})) \\ &= \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu \left(\sum_{K \in \mathcal{T}_h} \beta_K \|\kappa_K(\nabla_{\mathbf{x}}(\partial_t u)(t_{n,\mu}))\|_{0,K}^2 \right). \end{aligned}$$

Aus $\beta_K = Ch_k \leq Ch$ und der Annahme 4.2 folgt

$$\begin{aligned} Q_n [s_h(\partial_t u, \partial_t u)] &\leq C \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu h^{2r+1} \sum_{K \in \mathcal{T}_h} \|\nabla_{\mathbf{x}}(\partial_t u)(t_{n,\mu})\|_{r,K}^2 \\ &\leq C \frac{\tau_n}{2} \sum_{\mu=0}^k \hat{\omega}_\mu h^{2r+1} \|\partial_t u(t_{n,\mu})\|_{r+1}^2 \\ &\leq C \tau_n h^{2r+1} \sup_{t \in I_n} \|\partial_t u(t)\|_{r+1}^2. \end{aligned}$$

Dies ergibt die folgende Abschätzung für T_4 :

$$T_4 \leq C \tau_n^{1/2} h^{r+1/2} \sup_{t \in I_n} \|\partial_t u(t)\|_{r+1} \left\{ Q_n [(S_h v_{\tau h}, v_{\tau h})] \right\}^{1/2}.$$

Dabei lässt sich der Term $Q_n [s_h(v_{\tau h}, v_{\tau h})]$ links absorbieren. Man beachte, dass $\|v\|_0 \leq C \|v\|$. Schließlich erhält man durch die jeweiligen Teilabschätzungen die Behauptung. Dabei ergeben sich die Terme E_1^n aus der Abschätzung von T_4 , E_2^n aus T_1 und T_3 . E_3^n liefert die Kombination der Fehlerabschätzungen von T_1 und T_2 . Zuletzt erhalten wir die Abschätzung von T_5 aus E_4^n . \square

Lemma 4.19. [$\hat{e}_{\tau h}$ -Fehlerabschätzung] Es gilt für $m = 1, \dots, N$

$$\|\hat{e}_{\tau h}(t_m)\|_0^2 + \sum_{n=1}^m Q_n [\|\pi_{k-1}\hat{e}_{\tau h}\|^2] \leq C \sum_{n=1}^m \tau_n \left\{ \hat{E}_1^n(u) + \hat{E}_2^n(u) + \hat{E}_3^n(u) + \hat{E}_4^n(f) \right\}, \quad (4.64)$$

wobei \hat{E}_i^n , $i = 1, \dots, 4$, dem Lemma 4.18 zu entnehmen sind. Weiterhin erhalten wir die L^2 -Abschätzung

$$\|\hat{e}_{\tau h}\|_{L^2(I, L^2(\Omega))}^2 \leq CT \sum_{n=1}^N \tau_n \left\{ \hat{E}_1^n(u) + \hat{E}_2^n(u) + \hat{E}_3^n(u) + \hat{E}_4^n(f) \right\}.$$

Beweis. Aus dem Einsetzen von $w = \hat{e}_{\tau h}$ in der Stabilitätsungleichung in Lemma 4.17, dem Einsetzen von $v_{\tau h} = \pi_{k-1}\hat{e}_{\tau h}$ in der Fehlerabschätzung in Lemma 4.18 und die Kombination der Abschätzungen in beiden Lemmata folgt

$$\begin{aligned} & \frac{1}{2} \|\hat{e}_{\tau h}(t_n)\|_0^2 - \frac{1}{2} \|\hat{e}_{\tau h}(t_{n-1})\|_0^2 + Q_n [\|\pi_{k-1}\hat{e}_{\tau h}\|^2] \\ & \leq C \tau_n \left\{ \hat{E}_1^n(u) + \hat{E}_2^n(u) + \hat{E}_3^n(u) + \hat{E}_4^n(f) \right\}^{1/2} \left\{ Q_n [\|\pi_{k-1}\hat{e}_{\tau h}\|^2] \right\}^{1/2}. \end{aligned}$$

Aus der Young-Ungleichung erhalten wir

$$\begin{aligned} & \frac{1}{2} \|\hat{e}_{\tau h}(t_n)\|_0^2 - \frac{1}{2} \|\hat{e}_{\tau h}(t_{n-1})\|_0^2 + \frac{1}{2} Q_n [\|\pi_{k-1}\hat{e}_{\tau h}\|^2] \\ & \leq C \tau_n \left\{ \hat{E}_1^n(u) + \hat{E}_2^n(u) + \hat{E}_3^n(u) + \hat{E}_4^n(f) \right\}. \end{aligned} \quad (4.65)$$

Weiterhin liefert das Aufsummieren der Abschätzung (4.65) für $n = 1, \dots, m$ die Behauptung (4.64), wobei wir die Eigenschaft $\tilde{P}_h u_0 = u_h(0)$ ausgenutzt haben.

Nun multiplizieren wir die Abschätzung (4.64) mit τ_m , vernachlässigen die nicht-negative Quadraturformel auf der linken Seite und summieren die Abschätzungen über $m = 1, \dots, N$ auf. Wir erhalten

$$\begin{aligned} \sum_{m=1}^N \tau_m \|\hat{e}_{\tau h}(t_m)\|_0^2 & \leq C \left(\sum_{m=1}^N \tau_m \right) \max_{1 \leq m \leq N} \sum_{n=1}^m \tau_n \left\{ (\hat{E}^n(u, f)) \right\} \\ & \leq CT \sum_{n=1}^N \tau_n \left\{ (\hat{E}^n(u, f)) \right\}, \end{aligned} \quad (4.66)$$

wobei $\hat{E}^n(u, f) := \hat{E}_1^n(u) + \hat{E}_2^n(u) + \hat{E}_3^n(u) + \hat{E}_4^n(f)$.

Es ist offensichtlich, dass $\hat{e}_{\tau h}$ eine stetige stückweise polynomiale Funktion in der Zeit vom Grad kleiner oder gleich k ist. Aus der Normäquivalenz auf dem endlichdimensionalen Raum $\mathbb{P}_k(I_n, \mathcal{V}_h)$ erhalten wir

$$\|\hat{e}_{\tau h}\|_{L^2(I_n, L^2(\Omega))}^2 \leq C \left(Q_n [\|\hat{e}_{\tau h}\|_0^2] + \tau_n \|\hat{e}_{\tau h}(t_n)\|_0^2 \right). \quad (4.67)$$

Aus der Kombination von (4.64), (4.66) und (4.67) folgt

$$\begin{aligned} \|\hat{e}_{\tau h}\|_{L^2(I, L^2(\Omega))}^2 & = \sum_{n=1}^N \|\hat{e}_{\tau h}\|_{L^2(I_n, L^2(\Omega))}^2 \\ & \leq C \sum_{n=1}^N \left(Q_n [\|\pi_{k-1}\hat{e}_{\tau h}\|_0^2] + \tau_n \|\hat{e}_{\tau h}(t_n)\|_0^2 \right) \\ & \leq C \sum_{n=1}^N \tau_n \left\{ (\hat{E}^n(u, f)) \right\}. \end{aligned}$$

□

Lemma 4.20. [L^2 -Fehlerabschätzung] Es gilt für $m = 1, \dots, N$

$$\begin{aligned} \|\hat{e}(t_m)\|_0^2 &\leq C\tau_m \max_{1 \leq n \leq m} \left\{ \widehat{C}_{T,n}(u)\tau_n^{2(k+1)} + \widehat{C}_{T,n}(f)\tau_n^{2(k+1)} + \widehat{C}_{S,n}(u)(\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \right\} \\ &\quad + C'_{S,m}(\partial_t u)h^{2r+2}, \end{aligned}$$

wobei

$$\begin{aligned} \widehat{C}_{T,n}(u) &:= \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + \sup_{t \in I_n} \|\partial_t^{k+3} u(t)\|_1^2, \\ \widehat{C}_{T,n}(f) &:= \sup_{t \in I_n} \|\partial_t^{k+2} f(t)\|_0^2, \\ \widehat{C}_{S,n}(u) &:= \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2 + \sup_{t \in I_n} \|\partial_t u(t)\|_{r+1}^2 + \sup_{t \in I_n} \|\partial_t^2 u(t)\|_{r+1}^2. \end{aligned}$$

$C'_{S,m}$ ist in Lemma 4.15 definiert. Weiterhin gilt

$$\|\hat{e}\|_{L^2(I, L^2(\Omega))}^2 \leq CT \max_{1 \leq n \leq N} \left\{ \widehat{C}_{T,n}(u)\tau_n^{2(k+1)} + \widehat{C}_{T,n}(f)\tau_n^{2(k+1)} + \widehat{C}_{S,n}(u)(\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \right\}.$$

Beweis. Unter der Berücksichtigung des Fehlersplittings (4.61) betrachten wir die folgenden Abschätzungen für $\hat{\eta}$. Es gilt für $m = 1, \dots, N$

$$\begin{aligned} \|\hat{\eta}(t_m)\|_0^2 &= \|\partial_t u(t_m) - \tilde{P}_h \partial_t R_{k+1}^c u(t_m)\|_0^2 \\ &\leq \|\partial_t u(t_m) - \tilde{P}_h \partial_t u(t_m)\|_0^2 \\ &\leq Ch^{2r+2} \|\partial_t u(t_m)\|_{r+1}^2, \end{aligned} \tag{4.68}$$

dabei wurde in Betracht gezogen, dass $\partial_t R_{k+1}^c u(t_m) = \partial_t u(t_m)$ in allen Punkten $t_m, m = 1, \dots, N$. Weiterhin gilt

$$\begin{aligned} \|\hat{\eta}\|_{L^2(I_n, L^2(\Omega))}^2 &= \|\partial_t u - \tilde{P}_h \partial_t R_{k+1}^c u\|_{L^2(I_n, L^2(\Omega))}^2 \\ &\leq \|\partial_t u - \partial_t R_{k+1}^c u\|_{L^2(I_n, L^2(\Omega))}^2 + \|\partial_t R_{k+1}^c u - \tilde{P}_h(\partial_t R_{k+1}^c u)\|_{L^2(I_n, L^2(\Omega))}^2 \\ &\leq \|\partial_t u - \partial_t R_{k+1}^c u\|_{L^2(I_n, L^2(\Omega))}^2 + C\tau_n h^{2r+2} \sup_{t \in I_n} \|\partial_t R_{k+1}^c u\|_{r+1}^2 \\ &\leq C\tau_n \left\{ \tau_n^{2(k+1)} \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + Ch^{2r+2} \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2 \right\}, \end{aligned}$$

wobei wir die Fehlerabschätzung (4.38) und Korollar 4.4 ausgenutzt haben. Aus dem Aufsummieren über $n = 1, \dots, N$ erhalten wir

$$\begin{aligned} \|\hat{\eta}\|_{L^2(I, L^2(\Omega))}^2 &\leq C \left(\sum_{n=1}^N \tau_n \right) \left\{ \tau_n^{2(k+1)} \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + h^{2r+2} \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2 \right\} \\ &\leq CT \max_{1 \leq n \leq N} \left\{ \tau_n^{2(k+1)} \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + h^{2r+2} \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2 \right\}. \end{aligned} \tag{4.69}$$

Unter Verwendung der Dreiecksungleichung liefert die Kombination der Abschätzungen (4.68) und (4.69) mit den Ergebnissen aus Lemma 4.19 die Aussagen. \square

Lemma 4.21. [$L^2(L^2)$ -Fehlerabschätzung] Es gilt

$$\begin{aligned} \|\tilde{e}_{\tau h}\|_{L^2(I, L^2(\Omega))}^2 &\leq CT \sum_{n=1}^N \tau_n \left\{ E_1^n(u) + E_2^n(u) + E_3^n(u) \right\} \\ &\quad + CT\tau^2 \sum_{n=1}^N \tau_n \left\{ \widehat{E}_1^n(u) + \widehat{E}_2^n(u) + \widehat{E}_3^n(u) + \widehat{E}_4^n(f) \right\}, \end{aligned} \tag{4.70}$$

wobei E_i^n , $i = 1, \dots, 3$, und \widehat{E}_i^n , $i = 1, \dots, 4$, den Lemmata 4.13 und 4.18 zu entnehmen sind und $\tau := \max_{1 \leq n \leq N} \tau_n$. Weiterhin gilt

$$\begin{aligned} \|\tilde{e}\|_{L^2(I, L^2(\Omega))}^2 &\leq CT\tau^{2(k+2)} \max_{1 \leq n \leq N} \left\{ C_{T,n}(u) + \widehat{C}_{T,n}(u) + \widehat{C}_{T,n}(f) \right\} \\ &\quad + CT(\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \max_{1 \leq n \leq N} \left\{ C_{S,n}(u) + \tau^2 \widehat{C}_{S,n}(u) \right\}, \end{aligned} \quad (4.71)$$

wobei $C_{T,n}$, $C_{S,n}$, $\widehat{C}_{T,n}$ und $\widehat{C}_{S,n}$ in den Lemmata 4.15 und 4.20 definiert sind.

Beweis. Wir betrachten nun die Abschätzungen (4.56) bzw. (4.64) und vernachlässigen die nicht-negativen Quadraturformeln auf den linken Seiten der beiden Ungleichungen. Zudem multiplizieren wir den Rest, einerseits der ersten Ungleichung mit τ_m andererseits der zweiten Ungleichung mit τ_m^3 . Die Addition der beiden Ergebnisse liefert

$$\begin{aligned} \sum_{m=1}^N \tau_m \|\tilde{e}_{\tau h}(t_m)\|_0^2 + \sum_{m=1}^N \tau_m^3 \|\hat{e}_{\tau h}(t_m)\|_0^2 &\leq C \left(\sum_{m=1}^N \tau_m \right) \sum_{n=1}^m \tau_n \left\{ (E^n(u)) \right\} \\ &\quad + C \left(\sum_{m=1}^N \tau_m^3 \right) \sum_{n=1}^m \tau_n \left\{ (\widehat{E}^n(u, f)) \right\} \\ &\leq CT \sum_{n=1}^N \tau_n \left\{ (E^n(u)) \right\} + CT\tau^2 \sum_{n=1}^N \tau_n \left\{ (\widehat{E}^n(u, f)) \right\}, \end{aligned} \quad (4.72)$$

wobei $E^n(u) := E_1^n(u) + E_2^n(u) + E_3^n(u)$ und $\widehat{E}^n(u, f) := \widehat{E}_1^n(u) + \widehat{E}_2^n(u) + \widehat{E}_3^n(u) + \widehat{E}_4^n(f)$. Aus der Normäquivalenz in Satz 4.8 auf den endlichdimensionalen Räumen erhalten wir

$$\|\tilde{e}_{\tau h}\|_{L^2(I_n, L^2(\Omega))}^2 \leq C \left\{ Q_n [\|\pi_{k-1} \tilde{e}_{\tau h}\|_0^2] + \frac{\tau_n}{2} \|\tilde{e}_{\tau h}(t_n)\|_0^2 + \left(\frac{\tau_n}{2}\right)^3 \|\hat{e}_{\tau h}(t_n)\|_0^2 \right\}. \quad (4.73)$$

Aus der Kombination von (4.56), (4.72) und (4.73) folgt

$$\begin{aligned} \|\tilde{e}_{\tau h}\|_{L^2(I, L^2(\Omega))}^2 &= \sum_{n=1}^N \|\tilde{e}_{\tau h}\|_{L^2(I_n, L^2(\Omega))}^2 \\ &\leq C \sum_{n=1}^N \left\{ Q_n [\|\pi_{k-1} \tilde{e}_{\tau h}\|_0^2] + \frac{\tau_n}{2} \|\tilde{e}_{\tau h}(t_n)\|_0^2 + \left(\frac{\tau_n}{2}\right)^3 \|\hat{e}_{\tau h}(t_n)\|_0^2 \right\}. \end{aligned}$$

Demzufolge ist

$$\|\tilde{e}_{\tau h}\|_{L^2(I, L^2(\Omega))}^2 \leq CT \sum_{n=1}^N \tau_n \left\{ (E^n(u)) \right\} + CT\tau^2 \sum_{n=1}^N \tau_n \left\{ (\widehat{E}^n(u, f)) \right\}.$$

Weiterhin gilt

$$\begin{aligned} \|\eta\|_{L^2(I_n, L^2(\Omega))}^2 &= \|u - \widetilde{P}_h R_{k+1}^c u\|_{L^2(I_n, L^2(\Omega))}^2 \\ &\leq \|u - R_{k+1}^c u\|_{L^2(I_n, L^2(\Omega))}^2 + \|R_{k+1}^c u - \widetilde{P}_h(R_{k+1}^c u)\|_{L^2(I_n, L^2(\Omega))}^2 \\ &\leq \|u - R_{k+1}^c u\|_{L^2(I_n, L^2(\Omega))}^2 + C\tau_n h^{2r+2} \sup_{t \in I_n} \|R_{k+1}^c u\|_{r+1}^2 \\ &\leq C\tau_n \left\{ \tau_n^{2(k+2)} \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + Ch^{2r+2} \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2 \right\}. \end{aligned}$$

Dabei haben wir (4.36), (4.4) sowie (4.37) verwendet. Aus dem Aufsummieren über $n = 1, \dots, N$ ergibt sich

$$\begin{aligned} \|\eta\|_{L^2(I, L^2(\Omega))}^2 &\leq C \left(\sum_{n=1}^N \tau_n \right) \left\{ \tau_n^{2(k+2)} \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + h^{2r+2} \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2 \right\} \\ &\leq CT \max_{1 \leq n \leq N} \left\{ \tau_n^{2(k+2)} \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + h^{2r+2} \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2 \right\}. \end{aligned}$$

Mit der Verwendung der Dreiecksungleichung erhalten wir aus der Kombination der Abschätzung mit (4.70) die Aussage. \square

4.3 Unstetige Galerkin-Verfahren (dG)

Im folgenden Abschnitt werden wir die erhaltenen Ergebnisse bezüglich der cGP-Verfahren auf die unstetigen Galerkin-Verfahren verallgemeinern. Dies erfolgt mit dem gleichen Vorgehen. Hierbei muss darauf geachtet werden, dass der Testraum bei den dG-Verfahren mit dem Ansatzraum übereinstimmt. Darausfolgend sind einige Änderungen bei der Fehleranalyse zu betrachten, wie z.B. die L^2 -Projektion π_{k-1} , die im dG-Fall nicht mehr benötigt wird. Die theoretischen Ergebnisse, die wir im Folgenden vorstellen werden, entsprechen denen von Ern und Schieweck in [14]. Im Gegensatz zu [14] werden wir das Postprocessing auf eine andere Weise definieren. Darüber hinaus werden einige der verwendeten Operatoren modifiziert.

Analog zu dem vorherigen Abschnitt werden wir zunächst sämtliche Operatoren definieren, die für die Analyse der dG-Verfahren notwendig sein werden. Um Wiederholungen zu vermeiden werden wir dabei Ergebnisse zusammenfassen, die mit denen im cGP-Fall identisch sind.

4.3.1 Projektions- und Interpolationsoperatoren

Sei \mathcal{W} ein Banachraum. Für eine Funktion $v \in C^1(I, \mathcal{W})$ definieren wir sämtliche Operatoren, die für unsere Fehleranalyse notwendig sind.

Operatoren auf $\hat{I} = (-1, 1]$

Gauß–Radau-Interpolation \hat{I}_k^d :

Zunächst definieren wir die folgende Gauß–Radau-Interpolation:

$$\begin{aligned} \hat{I}_k^d : C^0(\hat{I}, \mathcal{W}) &\rightarrow \mathbb{P}_k(\hat{I}, \mathcal{W}), \\ \hat{I}_k^d \hat{v}(t) &= \sum_{\mu=1}^{k+1} \hat{v}(\hat{t}_\mu) \hat{L}_\mu(t), \end{aligned} \quad (4.74)$$

wobei $\hat{t}_\mu, \mu = 1, \dots, k+1$, die $(k+1)$ Knoten der Gauß–Radau-Quadraturformel auf $[-1, 1]$ sind und \hat{L}_μ die dazugehörigen Lagrange-Basispolynome darstellen.

Hermite/Lagrange-Interpolation \hat{L}_{k+1}^d :

Wir betrachten den Operatoren \hat{L}_{k+1}^d , den wir mittels

$$\begin{aligned} \hat{L}_{k+1}^d : C^1(\hat{I}, \mathcal{W}) &\rightarrow \mathbb{P}_{k+1}(\hat{I}, \mathcal{W}), \\ \hat{L}_{k+1}^d \hat{v}(\hat{t}_\mu) &= \hat{v}(\hat{t}_\mu), \quad \mu = 1, \dots, k+1, \\ \partial_t \hat{L}_{k+1}^d \hat{v}(1) &= \partial_t \hat{v}(1) \end{aligned} \quad (4.75)$$

definieren.

Operatoren auf $I_n = (t_{n-1}, t_n]$.

Nun definieren wir Operatoren auf jedem Teilintervall I_n , die sich auf \hat{I} verallgemeinern lassen.

Gauß–Radau-Interpolation I_k^d :

$$\begin{aligned} I_k^d : C(I_n, \mathcal{W}) &\rightarrow \mathbb{P}_k(I_n, \mathcal{W}) \\ I_k^d|_{I_n} v(t) &= \sum_{\mu=1}^{k+1} v(t_{n,\mu}) L_{n,\mu}(t), \end{aligned} \quad (4.76)$$

wobei $t_{n,\mu}, \mu = 1, \dots, k+1$, die transformierten Gauß–Radau-Punkte auf I_n sind und $L_{n,\mu}$ die zugehörigen Lagrange-Basispolynome bezeichnen. Die Gauß–Radau-Quadraturformel ist, wie in

Kapitel 2 erwähnt wurde, für Polynome vom Grad höchstens $2k$ exakt, d.h.

$$Q_n[p] = \frac{\tau_n}{2} \sum_{\mu=1}^{k+1} \omega_\mu p(t_{n,\mu}) = \int_{I_n} p(t) dt \quad \forall p \in \mathbb{P}_{2k}.$$

Hermite/Lagrange-Interpolation J_{k+2}^c :

Die Hermite/Lagrange-Interpolation J_{k+2}^d definieren wir, wie im cGP-Fall, durch

$$\begin{aligned} J_{k+2}^d : C^1(I, \mathcal{W}) &\rightarrow \mathbb{P}_{k+2}^{C^1}(I, \mathcal{W}), \\ \partial_t J_{k+2}^d v(t_{n-1}) &= \partial_t v(t_{n-1}), \\ J_{k+2}^d v(t_n) &= v(t_n), \\ \partial_t J_{k+2}^d v(t_n) &= \partial_t v(t_n), \\ J_{k+2}^d v(t_{n,\mu}) &= v(t_{n,\mu}), \quad \mu = 1, \dots, k, \quad n = 1, \dots, N, \end{aligned}$$

wobei $J_{k+2}^d v|_{I_n} \in \mathbb{P}_{k+2}(I_n, \mathcal{W})$ und $t_{n,\mu}$ die Gauß–Radau-Punkte auf $(t_{n-1}, t_n]$ sind. Die globale C^1 -Stetigkeit folgt unmittelbar aus der Definition an den Punkten t_n .

Spezielle Interpolatoren R_{k+1}^n :

Auf jedem Teilintervall I_n definieren wir den Hilfsoperator R_{k+1}^n durch

$$\begin{aligned} R_{k+1}^n : C^1(I_n, \mathcal{W}) &\rightarrow \mathbb{P}_{k+1}(I_n, \mathcal{W}) \\ \partial_t R_{k+1}^n v(t_{n,\mu}) &= \partial_t J_{k+2}^d v(t_{n,\mu}), \quad \mu = 1, \dots, k+1, \\ R_{k+1}^n v(t_{n-1}) &= J_{k+2}^d v(t_{n-1}) = v(t_{n-1}). \end{aligned} \tag{4.77}$$

Hermite/Lagrange-Interpolation L_{k+1}^d :

Wir betrachten desweiteren den Operator L_{k+1}^d , den wir mittels

$$\begin{aligned} L_{k+1}^d : C^1(I_n, \mathcal{W}) &\rightarrow \mathbb{P}_{k+1}(I_n, \mathcal{W}), \\ L_{k+1}^d v(t_{n,\mu}) &= v(t_{n,\mu}), \quad \mu = 1, \dots, k+1, \\ \partial_t L_{k+1}^d v(t_n) &= \partial_t v(t_n) \end{aligned} \tag{4.78}$$

definieren.

Globaler Operator R_{k+1}^d :

Analog zum Operator R_{k+1}^c bei cGP-Verfahren, kann man zeigen, dass R_{k+1}^d , definiert durch

$$R_{k+1}^d|_{I_n} := R_{k+1}^n, \quad n = 1, \dots, N,$$

stetig ist und auf C^1 -Funktionen abbildet.

4.3.2 Eigenschaften der Interpolationsoperatoren

Wir definieren das Polynom $\hat{v} \in \mathbb{P}_{k+1}(\hat{I}, \mathcal{V}_h)$, welches durch die Bedingungen $\partial_{\hat{t}} \hat{v}(1) = 1$ und $\hat{v}(\hat{t}_j) = 0$ für alle $j = 1, \dots, k+1$ definiert ist. Dabei sind \hat{t}_j die Integrationspunkte der $(k+1)$ -punktigen rechtsseitigen Gauß–Radau-Formel auf \hat{I} . Weiterhin betrachten wir auf jedem Teilintervall I_n das Polynom $\vartheta_n \in \mathbb{P}_{k+1}(\hat{I}, \mathcal{V}_h)$ definiert durch $\vartheta_n(t) := \frac{\tau_n}{2} \hat{v}(t)$ mit $\hat{t} := T_n^{-1}(t)$. Dabei ist T_n die in 4.27 definierte affine Referenztransformation.

Wie im cGP-Fall lässt sich auch hier mithilfe der Referenztransformation T_n ein Zusammenhang zwischen den auf \hat{I} definierten Operatoren und den auf I_n definierten Operatoren herstellen. Es gilt $\hat{I}_k^d \hat{v} = I_k^d v$ und $\hat{L}_{k+1}^d \hat{v} = L_{k+1}^d v$ für jede hinreichend glatte Funktion v und \hat{v} mit $\hat{v}(\hat{t}) = v(t)$.

Das folgende Lemma kann ebenso wie im cGP-Fall bewiesen werden.

Lemma 4.22. Sei $w \in \mathbb{P}_{k+1}(I_n, \mathcal{V}_h)$. Es gilt

$$w(t) = I_k^d w(t) + l_{n-1} \vartheta_n(t) \quad \forall t \in I_n. \quad (4.79)$$

Weiterhin gilt für $w \in C^1(I_n, \mathcal{V}_h)$

$$L_{k+1}^d w(t) = I_k^d w(t) + \tilde{l}_{n-1} \vartheta_n(t) \quad \forall t \in I_n.$$

Die zweite Aussage bedeutet, wie wir bereits im cGP-Fall feststellen konnten, dass sich $L_{k+1}^d w$ und $I_k^d w$ auf I_n nur um ein Vielfaches von ϑ_n unterscheiden.

Sei nun v eine Funktion aus $C^{k+3}(I_n, \mathcal{V}_h)$ und $r \in \mathbb{N}$, dann erhalten wir für den oben definierten Interpolationsoperator J_{k+2}^d die Fehlerabschätzung

$$\sup_{t \in I_n} |J_{k+2}^d v(t) - v(t)|_r \leq C \tau_n^{k+3} \sup_{t \in I_n} |\partial_t^{k+3} v(t)|_r \quad (4.80)$$

und für die Zeitableitungen von J_{k+2}^d gilt

$$\sup_{t \in I_n} |\partial_t J_{k+2}^d v(t) - \partial_t v(t)|_r \leq C \tau_n^{k+2} \sup_{t \in I_n} |\partial_t^{k+3} v(t)|_r$$

und

$$\sup_{t \in I_n} |\partial_t^2 J_{k+2}^d v(t) - \partial_t^2 v(t)|_r \leq C \tau_n^{k+1} \sup_{t \in I_n} |\partial_t^{k+3} v(t)|_r.$$

Diese folgen unmittelbar aus den $(k+3)$ Bedingungen in der Definition von J_{k+1}^d .

Aus der Definition von R_{k+1}^d folgt, dass $\partial_t R_{k+1}^d$ und $\partial_t J_{k+2}^d$ die Funktion v an denselben Stützstellen $t_{n,\mu}$, $\mu = 1, \dots, k+1$, interpolieren. Darausfolgend erhalten wir die folgende Interpolationsfehlerabschätzung

$$\sup_{t \in I_n} \|\partial_t J_{k+2}^d v(t) - \partial_t R_{k+1}^d u(t)\|_0 \leq C \tau_n^{k+1} \sup_{t \in I_n} \|\partial_t^{k+2} v(t)\|_0.$$

Der Beweis der folgenden Fehlerabschätzungen bezüglich R_{k+1}^d folgt wie im cGP-Fall, vgl. dazu Lemma 4.6 und Korollar 4.7.

Lemma 4.23. [R_{k+1}^d -Fehlerabschätzung] Für alle $n = 1, \dots, N$ und alle $v \in C^{k+2}(I_n, H^{r+1}(\Omega))$ gilt

$$\sup_{t \in I_n} \|R_{k+1}^d v(t) - v(t)\|_{r+1} \leq C \tau_n^{k+2} \sup_{t \in I_n} \|\partial_t^{k+2} v(t)\|_{r+1}. \quad (4.81)$$

Außerdem gilt

$$\sup_{t \in I_n} \|R_{k+1}^d v(t)\|_{r+1} \leq C \|v\|_{C^1(I_n, H^{r+1}(\Omega))} \quad \forall v \in C^1(I_n, H^{r+1}(\Omega)). \quad (4.82)$$

Korollar 4.24. [$\partial_t R_{k+1}^d$ -Fehlerabschätzung] Für alle $n = 1, \dots, N$ und alle $v \in C^{k+2}(I_n, H^{r+1}(\Omega))$ gilt

$$\sup_{t \in I_n} \|\partial_t R_{k+1}^d v(t) - \partial_t v(t)\|_{r+1} \leq C \tau_n^{k+1} \sup_{t \in I_n} \|\partial_t^{k+2} v(t)\|_{r+1}. \quad (4.83)$$

Darüber hinaus gilt

$$\sup_{t \in I_n} \|\partial_t R_{k+1}^d v(t)\|_{r+1} \leq C \|v\|_{C^1(I_n, H^{r+1}(\Omega))} \quad \forall v \in C^1(I_n, H^{r+1}(\Omega)). \quad (4.84)$$

Bezüglich der Hermite/Lagrange-Interpolation L_{k+1}^d gilt die Fehlerabschätzung

$$\sup_{t \in I_n} \|L_{k+1}^d v(t) - v(t)\|_0 \leq C \tau_n^{k+2} \sup_{t \in I_n} \|\partial_t^{k+2} v(t)\|_0.$$

4.3.3 Normen

Satz 4.25. *Die Ausdrücke*

$$\left(\int_{I_n} \|v(t)\|_W^2 dt \right)^{1/2} \quad (4.85)$$

und

$$\left(\int_{I_n} \|I_k^d v(t)\|_W^2 dt + \frac{\tau_n}{2} \|v(t_{n-1})\|_W^2 \right)^{1/2} \quad (4.86)$$

sind äquivalente Normen auf $\mathbb{P}_{k+1}(I_n, W)$, wobei die Äquivalenzkonstanten nicht von τ_n abhängen und W einen beliebigen Banachraum darstellt.

Beweis. Unter Ausnutzung der Transformation von $(t_{n-1}, t_n]$ auf $(-1, 1]$ ist die Behauptung äquivalent zu:

Die Ausdrücke

$$\left(\int_{-1}^1 \|\hat{v}(\hat{t})\|_W^2 d\hat{t} \right)^{1/2} \quad (4.87)$$

und

$$\left(\int_{-1}^1 \|\hat{I}_k^d \hat{v}(\hat{t})\|_W^2 d\hat{t} + \|\hat{v}(-1)\|_W^2 \right)^{1/2} \quad (4.88)$$

sind äquivalente Normen auf dem endlich dimensionalen Raum $\mathbb{P}_{k+1}(\hat{I}, W)$. Der Ausdruck (4.88) erfüllt nach Konstruktion alle Eigenschaften einer Halbnorm. Es ist ausreichend die verbleibenden Normeigenschaften zu zeigen. Die Aussage folgt dann mittels der Normäquivalenz auf dem endlich dimensionalen Raum $\mathbb{P}_{k+1}(\hat{I}, W)$. Es bleibt daher zu zeigen, dass sich aus

$$0 = \int_{-1}^1 \|\hat{I}_k^d \hat{v}(\hat{t})\|_W^2 d\hat{t} + \|\hat{v}(-1)\|_W^2 \quad (4.89)$$

die Folgerung $\hat{v} \equiv 0$ ergibt. Aus (4.89) folgt, dass $\hat{I}_k^d \hat{v} = 0$. Da \hat{v} und $\hat{I}_k^d \hat{v}$ nach Definition von \hat{I}_k^d an den Gauß-Radau-Punkten $\hat{t}_\mu, \mu = 1, \dots, k+1$, übereinstimmen und $\hat{v}(-1) = 0$ folgt, dass $\hat{v} = 0$ an den $(k+2)$ Punkten gilt. Anschließend ergibt sich aus $\hat{v} \in \mathbb{P}_{k+1}(\hat{I}, W)$ die erwünschte Folgerung $\hat{v} \equiv 0$. Damit ist der Ausdruck (4.88) eine Norm. \square

4.3.4 Volldiskretes Problem mit Ortsstabilisierung

Das volldiskrete dG-Verfahren kombiniert mit der Ortsstabilisierung lautet:

Finde $u_{\tau h}|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V}_h)$ mit $u_{\tau h}(t_{n-1}^+) = u_{\tau h}^{n-1}$ und

$$\int_{I_n} \langle \partial_t u_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt + ([u_{\tau h}]_{n-1}, v_{\tau h}(t_{n-1}^+)) = Q_n[(f, v_{\tau h})] \quad \forall v_{\tau h} \in \mathbb{P}_k(I_n, \mathcal{V}_h). \quad (4.90)$$

4.3.5 Volldiskretes Problem und Postprocessing

Sei $u_{\tau h}(0) = \tilde{P}_h u_0$ eine Approximation an den Anfangswert u_0 . Mit der Anwendung des Wertes $u_{\tau h}(t_{n-1}^+)$ aus dem vorherigen Intervall I_{n-1} lässt sich das lokale volldiskrete Problem für $n = 1, \dots, N$ folgendermaßen formulieren:

Finde $u_{\tau h}|_{I_n} \in \mathbb{P}_k(I_n, \mathcal{V}_h)$ mit $u_{\tau h}(t_{n-1}^+) = u_{\tau h}^{n-1}$ und

$$\int_{I_n} \langle \partial_t u_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt + ([u_{\tau h}]_{n-1}, v_{\tau h}(t_{n-1}^+)) = Q_n[(f, v_{\tau h})] \quad \forall v_{\tau h} \in \mathbb{P}_k(I_n, \mathcal{V}_h). \quad (4.91)$$

Wir betrachten auch in diesem Abschnitt das Postprocessing aus Kapitel 3, jedoch für die dG-Verfahren. Auf dem Intervall I_n setzen wir

$$\tilde{u}_{\tau h}(t) = u_{\tau h}(t) + d_n \vartheta_n(t),$$

wobei $d_n \in \mathcal{V}_h$ eine ortsdiskrete Funktion darstellt und $\vartheta_n \in \mathbb{P}_{k+1}(I_n, \mathcal{V}_h)$, mit $\vartheta_n(t_{n,\mu}) = 0 \quad \forall \mu = 1, \dots, k+1$, und $\partial_t \vartheta_n(t_n) = 1$. Dabei sind $t_{n,\mu}$ die Gauß–Radau-Punkte.

Wie auch schon im cGP-Fall soll die nachbearbeitete Lösung $\tilde{u}_{\tau h}$ im L^2 -Ortssinne die zeitliche Differentialgleichung in $t = t_n$ erfüllen, d.h.,

$$(\partial_t \tilde{u}_{\tau h}(t_n), v) = \langle f(t_n) - A_h \tilde{u}_{\tau h}(t_n), v \rangle.$$

Infolge von $\vartheta_n(t_n) = 0$ ist $\tilde{u}_{\tau h}(t_n) = u_{\tau h}(t_n)$, somit ist die Differentialgleichung äquivalent zu

$$(\partial_t u_{\tau h}(t_n), v) + (d_n, v) \partial_t \vartheta_n(t_n) = \langle f(t_n) - A_h u_{\tau h}(t_n), v \rangle.$$

Aus $\partial_t \vartheta_n(t_n) = 1$ erhalten wir

$$(d_n, v) \partial_t \vartheta_n(t_n) = \langle f(t_n) - A_h u_{\tau h}(t_n) - \partial_t u_{\tau h}(t_n), v \rangle.$$

Dies ergibt

$$d_n = P_h f(t_n) - A_h u_{\tau h}(t_n) - \partial_t u_{\tau h}(t_n),$$

wobei P_h die L^2 -Projektion in den Finite-Elemente-Raum \mathcal{V}_h darstellt. Auch im dG-Fall stimmen $u_{\tau h}$ und $\tilde{u}_{\tau h}$ in den Gauß–Radau-Punkten überein. Weiterhin folgt aus Lemma 3.2, dass

$$-d_n \vartheta_n(t_{n-1}^+) = [u_{\tau h}]_{n-1}. \quad (4.92)$$

4.3.6 Fehleranalyse

Im Folgenden betrachten wir ein wichtiges Ergebnis bezüglich der durch Postprocessing nachbearbeiteten Lösung.

Lemma 4.26. *Für alle $n = 1, \dots, N$ gilt*

$$\int_{I_n} \langle \partial_t \tilde{u}_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt = Q_n[(f, v_{\tau h})] \quad \text{für alle } v_{\tau h} \in \mathbb{P}_k(I_n, \mathcal{V}_h).$$

Beweis. Aus der Definition vom Postprocessing folgt

$$\int_{I_n} (\partial_t \tilde{u}_{\tau h}(t), v_{\tau h}(t)) dt = \int_{I_n} (\partial_t u_{\tau h}(t), v_{\tau h}(t)) dt + \underbrace{\int_{I_n} (d_n \partial_t \vartheta_n(t), v_{\tau h}(t)) dt}_{:=T}. \quad (4.93)$$

Beim Betrachten des letzten Terms auf der rechten Seite liefert eine partielle Integration

$$T = - \int_{I_n} (d_n \vartheta_n(t), \partial_t v_{\tau h}(t)) dt + (d_n \vartheta_n(t), v_{\tau h}(t)) \Big|_{t_{n-1}^+}^{t_n}.$$

Aus $\vartheta_n \in \mathbb{P}_k(I_n)$ und $\partial_t v_{\tau h} \in \mathbb{P}_{k-1}(I_n)$ folgt $(d_n \vartheta_n, \partial_t v_{\tau h}) \in \mathbb{P}_{2k-1}(I_n)$. Dieser Term wird mittels der $(k+1)$ -punktigen Gauß–Radau-Quadraturformel exakt integriert. Es gilt

$$\begin{aligned} T &= -Q_n[(d_n \vartheta_n, \partial_t v_{\tau h})] + d_n \vartheta_n(t_n) v_{\tau h}(t_n) - d_n \vartheta_n(t_{n-1}^+) v_{\tau h}(t_{n-1}^+) \\ &= -d_n \vartheta_n(t_{n-1}^+) v_{\tau h}(t_{n-1}^+) \\ &= ([u_{\tau h}]_{n-1}, v_{\tau h}(t_{n-1}^+)). \end{aligned} \quad (4.94)$$

Der letzte Term folgt unmittelbar aus (4.92). Die Behauptung erhalten wir aus der Kombination von (4.90), (4.93) und (4.94). \square

Im Folgenden betrachten wir auf jedem Teilintervall I_n , $n = 1, \dots, N$, die Bilinearform B_n aus dem vorherigen Abschnitt definiert durch

$$B_n(w, v) := Q_n[(\partial_t w, v)] + Q_n[a_h(w, v)].$$

Lemma 4.27. *Die volldiskrete Lösung erfüllt für alle $n = 1, \dots, N$*

$$B_n(\tilde{u}_{\tau h}, v_{\tau h}) = Q_n[(f, v_{\tau h})] \quad \forall v_{\tau h} \in \mathbb{P}_k(I_n, \mathcal{V}_h).$$

Beweis. Aus der Definition der Bilinearform B_n folgt

$$B_n(\tilde{u}_{\tau h}, v_{\tau h}) = Q_n[(\partial_t \tilde{u}_{\tau h}, v_{\tau h})] + Q_n[a_h(\tilde{u}_{\tau h}, v_{\tau h})].$$

Im ersten Summanden auf der rechten Seite liegt das Produkt in $\mathbb{P}_{2k}(I_n)$, damit ist die Gauß–Radau-Formel exakt und wir erhalten

$$\begin{aligned} Q_n[(\partial_t \tilde{u}_{\tau h}, v_{\tau h})] &= \int_{I_n} (\partial_t \tilde{u}_{\tau h}(t), v_{\tau h}(t)) dt \\ &= \int_{I_n} (\partial_t u_{\tau h}(t), v_{\tau h}(t)) dt + ([u_{\tau h}]_{n-1}, v_{\tau h}(t_{n-1}^+)). \end{aligned}$$

Hier ist die letzte Gleichung aus dem Beweis von Lemma 4.26 zu entnehmen. Für den zweiten Summanden ergibt sich

$$Q_n[a_h(\tilde{u}_{\tau h}, v_{\tau h})] = Q_n[a_h(u_{\tau h}, v_{\tau h})] = \int_{I_n} a_h(u_{\tau h}(t), v_{\tau h}(t)) dt,$$

weil $u_{\tau h}$ und $\tilde{u}_{\tau h}$ in den Gauß–Radau-Punkten übereinstimmen. Der letzte Schritt ist möglich, da $a_h(u_{\tau h}, v_{\tau h}) \in \mathbb{P}_{2k}(I_n)$ gilt. Die Quadraturformel ist dementsprechend exakt und es folgt

$$B_n(\tilde{u}_{\tau h}, v_{\tau h}) = \int_{I_n} \langle \partial_t u_{\tau h}(t) + A_h u_{\tau h}(t), v_{\tau h}(t) \rangle dt + ([u_{\tau h}]_{n-1}, v_{\tau h}(t_{n-1}^+)) = Q_n[(f, v_{\tau h})]$$

für alle $v_{\tau h} \in \mathbb{P}_k(I_n, \mathcal{V}_h)$. □

Lemma 4.28. *[Stabilität] Sei $w \in \mathbb{P}_{k+1}^d(I, \mathcal{V}_h)$, dann gilt*

$$B_n(w, I_k^d w) \geq \frac{1}{2} \|w(t_n)\|_0^2 - \frac{1}{2} \|w(t_{n-1})\|_0^2 + \frac{1}{2} \|l_{n-1}\|_0^2 \|\vartheta_n(t_{n-1}^+)\|_0^2 + Q_n[\|I_k^d w\|^2],$$

wobei l_{n-1} in Lemma 4.22 definiert ist.

Beweis. Aus der Definition der Bilinearform B_n erhalten wir

$$B_n(w, I_k^d w) = Q_n[(\partial_t w, I_k^d w)] + Q_n[a_h(w, I_k^d w)].$$

Im ersten Summanden auf der rechten Seite liegt das Produkt in $\mathbb{P}_{2k}(I_n)$. Die Gauß–Radau-Formel ist somit exakt. Dies ergibt $Q_n[(\partial_t w, I_k^d w)] = \int_{I_n} (\partial_t w, I_k^d w) dt$. Weiterhin erhalten wir aus (4.79)

$$\begin{aligned} Q_n[(\partial_t w_{\tau h}, I_k^d w)] &= \int_{I_n} (\partial_t w(t), w(t) - l_{n-1} \vartheta_n(t)) dt \\ &= \int_{I_n} (\partial_t w(t), w(t)) dt - \int_{I_n} (\partial_t w(t), l_{n-1} \vartheta_n(t)) dt \\ &= \frac{1}{2} \int_{I_n} \frac{d}{dt} \|w(t)\|_0^2 dt - \int_{I_n} (\partial_t w(t), l_{n-1} \vartheta_n(t)) dt. \end{aligned}$$

Wir schreiben das letzte Integral folgendermaßen um:

$$\begin{aligned}
 \int_{I_n} (\partial_t w(t), l_{n-1} \vartheta_n(t)) dt &= \int_{I_n} (\partial_t (I_k^d w(t) + l_{n-1} \vartheta_n(t)), l_{n-1} \vartheta_n(t)) dt \\
 &= \int_{I_n} (\partial_t I_k^d w(t), l_{n-1} \vartheta_n(t)) dt + \frac{1}{2} \|l_{n-1}\|_0^2 \int_{I_n} \frac{d}{dt} \|\vartheta_n(t)\|_0^2 dt \\
 &= Q_n [(\partial_t I_k^d w, l_{n-1} \vartheta_n)] + \frac{1}{2} \|l_{n-1}\|_0^2 (\|\vartheta_n(t_n)\|_0^2 - \|\vartheta_n(t_{n-1}^+)\|_0^2) \\
 &= -\frac{1}{2} \|l_{n-1}\|_0^2 \|\vartheta_n(t_{n-1}^+)\|_0^2.
 \end{aligned}$$

Dabei haben wir zum einen ausgenutzt, dass $(\partial_t I_k^d w, l_{n-1} \vartheta_n) \in \mathbb{P}_{2k-1}(I_n)$ durch die Gauß–Radau-Formel exakt integriert wird und zum anderen, dass ϑ_n an allen Gauß–Radau-Punkten verschwindet. Es folgt

$$\begin{aligned}
 Q_n [(\partial_t w, I_k^d w)] &= \frac{1}{2} \int_{I_n} \frac{d}{dt} \|w(t)\|_0^2 dt + \frac{1}{2} \|l_{n-1}\|_0^2 \|\vartheta_n(t_{n-1}^+)\|_0^2 \\
 &= \frac{1}{2} \|w(t_n)\|_0^2 - \frac{1}{2} \|w(t_{n-1})\|_0^2 + \frac{1}{2} \|l_{n-1}\|_0^2 \|\vartheta_n(t_{n-1}^+)\|_0^2. \quad (4.95)
 \end{aligned}$$

Weiterhin haben wir

$$\begin{aligned}
 Q_n [a_h(w, I_k^d w)] &= Q_n [a_h(I_k^d w, I_k^d w)] + Q_n [a_h(l_{n-1} \vartheta_n, I_k^d w)] \\
 &= \int_{I_n} \langle A_h I_k^d w(t), I_k^d w(t) \rangle dt \\
 &\geq \int_{I_n} \|I_k^d w(t)\|^2 dt = Q_n [\|I_k^d w\|^2], \quad (4.96)
 \end{aligned}$$

wobei wir erneut die Eigenschaft ausgenutzt haben, dass ϑ_n an den Gauß–Radau-Punkten 0 ist. Weiterhin haben wir die Koerzivität (4.9) des Operators A_h verwendet.

Schließlich ergibt sich aus der Zusammenfassung von (4.95) und (4.96) die Behauptung. \square

Die folgenden Lemmata erhalten wir mit denselben Vorgehensweisen wie in den Beweisen von Lemmata 4.13 - 4.15, jedoch werden hier die im dG-Rahmen definierten Operatoren eingesetzt sowie die erhaltene Stabilitätsaussage in Lemma 4.28 verwendet. Wir betrachten die Fehlerzerlegung

$$\tilde{\epsilon}(t) = \eta(t) + \tilde{\epsilon}_{\tau h}(t), \quad (4.97)$$

wobei wir

$$\begin{aligned}
 \eta(t) &:= u(t) - \tilde{P}_h R_{k+1}^d u(t), \\
 \tilde{\epsilon}_{\tau h}(t) &:= \tilde{P}_h R_{k+1}^d u(t) - \tilde{u}_{\tau h}(t)
 \end{aligned}$$

setzen.

Lemma 4.29. *[Fehlerabschätzung] Es gilt*

$$|B_n(\tilde{\epsilon}_{\tau h}, v_{\tau h})|^2 \leq C \tau_n \left\{ E_1^n(u) + E_2^n(u) + E_3^n(u) \right\} \left\{ Q_n [\|v_{\tau h}\|^2] \right\}, \quad (4.98)$$

wobei die Fehler im Ort durch

$$\begin{aligned}
 E_1^n(u) &:= h^{2r+1} \sup_{t \in I_n} \|u(t)\|_{r+1}^2, \\
 E_2^n(u) &:= (\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2
 \end{aligned}$$

gegeben sind, wohingegen der Fehler in der Zeit durch

$$E_3^n(u) := \tau_n^{2(k+2)} \left\{ \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + \sup_{t \in I_n} \|\partial_t^{k+3} u(t)\|_0^2 \right\}$$

gegeben ist.

Lemma 4.30. [Fehlerabschätzung] Es gilt für $m = 1, \dots, N$,

$$\|\tilde{e}_{\tau h}(t_m)\|_0^2 + \sum_{n=1}^m Q_n [\|I_k^d \tilde{e}_{\tau h}\|^2] \leq C \sum_{n=1}^m \tau_n \left\{ E_1^n(u) + E_2^n(u) + E_3^n(u) \right\}, \quad (4.99)$$

wobei E_i^n , $i = 1, \dots, 3$, dem Lemma 4.29 zu entnehmen sind. Weiterhin folgt die Abschätzung

$$\|\tilde{e}_{\tau h}\|_{L^2(I, L^2(\Omega))}^2 \leq C \sum_{n=1}^N \tau_n \left\{ E_1^n(u) + E_2^n(u) + E_3^n(u) \right\}. \quad (4.100)$$

Beweis. Die erste Aussage lässt sich mit denselben Argumenten wie im Beweis von Lemma 4.14 zeigen. Das Einsetzen von $w = \tilde{e}_{\tau h}$ in der Stabilitätsungleichung in Lemma 4.28, das Einsetzen von $v_{\tau h} = I_k^d \tilde{e}_{\tau h}$ in der Fehlerabschätzung in Lemma 4.29 und die Kombination der Abschätzungen in beiden Lemmata führen zur folgenden Abschätzung:

$$\begin{aligned} & \frac{1}{2} \|\tilde{e}_{\tau h}(t_n)\|_0^2 - \frac{1}{2} \|\tilde{e}_{\tau h}(t_{n-1})\|_0^2 + Q_n [\|I_k^d \tilde{e}_{\tau h}\|^2] \\ & \leq C \tau_n \left\{ E_1^n(u) + E_2^n(u) + E_3^n(u) \right\}^{1/2} \left\{ Q_n [\|I_k^d \tilde{e}_{\tau h}\|^2] \right\}^{1/2}. \end{aligned}$$

Aus der Young-Ungleichung erhalten wir

$$\frac{1}{2} \|\tilde{e}_{\tau h}(t_n)\|_0^2 - \frac{1}{2} \|\tilde{e}_{\tau h}(t_{n-1})\|_0^2 + \frac{1}{2} Q_n [\|I_k^d \tilde{e}_{\tau h}\|^2] \leq C \tau_n \left\{ E_1^n(u) + E_2^n(u) + E_3^n(u) \right\}. \quad (4.101)$$

Weiterhin liefert das Aufsummieren dieser Abschätzung für $n = 1, \dots, m$ die Behauptung (4.99). Wir haben aus der Eigenschaft $\tilde{P}_h u_0 = u_h(0)$ Gebrauch gemacht.

Nun multiplizieren wir die Abschätzung (4.99) mit τ_m und vernachlässigen die nicht-negative Quadraturformel auf der linken Seite. Das Aufsummieren über $m = 1, \dots, N$ liefert

$$\begin{aligned} \sum_{m=1}^N \tau_m \|\hat{e}_{\tau h}(t_{m-1})\|_0^2 & \leq C \left(\sum_{m=1}^N \tau_m \right) \sum_{n=1}^{m-1} \tau_n \left\{ (E^n(u)) \right\} \\ & \leq CT \sum_{n=1}^N \tau_n \left\{ (E^n(u)) \right\}, \end{aligned} \quad (4.102)$$

wobei $E^n(u) := E_1^n(u) + E_2^n(u) + E_3^n(u)$.

Es ist offensichtlich, dass $e_{\tau h}$ eine stetige stückweise polynomiale Funktion in der Zeit vom Grad kleiner oder gleich k ist. Aus der Normäquivalenz in Satz 4.25 erhalten wir

$$\begin{aligned} \|e_{\tau h}\|_{L^2(I, L^2(\Omega))}^2 & = \sum_{n=1}^N \|e_{\tau h}\|_{L^2(I_n, L^2(\Omega))}^2 \\ & \leq C \sum_{n=1}^N \left(Q_n [\|e_{\tau h}\|_0^2] + \frac{\tau_n}{2} \|\hat{e}_{\tau h}(t_{n-1})\|_0^2 \right) \\ & \leq C \sum_{n=1}^N \tau_n \left\{ (E^n(u)) \right\}. \end{aligned}$$

□

Das folgende Lemma lässt sich genau wie im cGP-Fall beweisen.

Lemma 4.31. [*L²-Fehlerabschätzung*] *Es gilt für $m = 1, \dots, N$,*

$$\|\tilde{e}(t_m)\|_0^2 \leq C\tau_m \max_{1 \leq n \leq m} \left\{ C_{T,n}(u)\tau_n^{2(k+2)} + C_{S,n}(u)(\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \right\} + C'_{S,m}(u)h^{2r+2},$$

wobei

$$\begin{aligned} C_{T,n}(u) &:= \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + \sup_{t \in I_n} \|\partial_t^{k+3} u(t)\|_0^2, \\ C_{S,n}(u) &:= \sup_{t \in I_n} \|u(t)\|_{r+1}^2 + \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2, \\ C'_{S,m}(u) &:= \|u(t_m)\|_{r+1}^2. \end{aligned}$$

4.3.7 Fehlerabschätzung für die Zeitableitung

Im Gegensatz zum cGP-Fall, liefert das Postprocessing bei dG-Verfahren eine diskrete Lösung in $C^0(I, \mathcal{V}_h)$. Darausfolgend muss die Wohlgestellttheit der Zeitableitung der durch Postprocessing nachbereiteten diskreten Lösung sicher gestellt werden, um eine Fehlerabschätzung für die Zeitableitung zu erhalten. Wie in [14] erläutert wurde, ist die Zeitableitung von $\tilde{u}_{\tau h}$ im Inneren des Intervalls (t_{n-1}, t_n) wohldefiniert, da $\tilde{u}_{\tau h} \in \mathbb{P}_{k+1}^d(I, \mathcal{V}_h) \cap C^0(I, \mathcal{V}_h)$. Eine Erweiterung von $\tilde{u}_{\tau h}$ auf der linken Seite von t_n erhält man durch den Ansatz

$$\partial_t \tilde{u}_{\tau h}(t_n) := \lim_{t \nearrow t_n} \partial_t \tilde{u}_{\tau h}(t) \quad \forall n = 1, \dots, N. \quad (4.103)$$

Wir erhalten daher eine Definition von $\partial_t \tilde{u}_{\tau h}$ über das gesamte Intervall I . Die Funktion $\partial_t \tilde{u}_{\tau h} \in \mathbb{P}_k^d(I, \mathcal{V}_h)$ ist somit über I wohldefiniert.

Das folgende Lemma ist, bis auf die verwendeten Zeit-Interpolatoren und Quadraturformeln, identisch mit Lemma 4.16 beim cGP-Verfahren und lässt sich auch unter Verwendung der Gauß–Radau-Stützstellen auf die gleiche Weise zeigen.

Lemma 4.32. *Sei $u_{\tau h}$ die volldiskrete Lösung, dann gilt*

$$\partial_t \tilde{u}_{\tau h}(t) + A_h u_{\tau h}(t) = P_h I_k^d f(t) \quad \forall t \in I,$$

wobei I_k^d der in (4.76) definierte Gauß–Radau-Zeitinterpolator ist.

Nun führen wir die folgende modifizierte Bilinearform \tilde{B}_n ein:

$$\tilde{B}_n(v, w) := Q_n [\langle \partial_t v + A_h v, w \rangle] + ([v]_{n-1}, w_{n-1}^+).$$

Lemma 4.33. *Sei $w \in \mathbb{P}_k(I_n, \mathcal{V}_h)$. Es gilt*

$$\tilde{B}_n(w, w) \geq \frac{1}{2} \|w(t_n)\|_0^2 - \frac{1}{2} \|w_{n-1}\|_0^2 + \|[w]_{n-1}\|_0^2 + Q_n [\|w\|^2].$$

Beweis. Aus der Definition von \tilde{B}_n folgt

$$\begin{aligned} \tilde{B}_n(w, w) &= Q_n [(\partial_t w, w)] + Q_n [a_h(w, w)] + ([w]_{n-1}, w_{n-1}^+) \\ &= \int_{I_n} (\partial_t w(t), w(t)) dt + \int_{I_n} a_h(w(t), w(t)) dt + (w_{n-1}^+ - w_{n-1}, w_{n-1}^+). \end{aligned}$$

Dabei haben wir ausgenutzt, dass beide $(\partial_t w, w) \in \mathbb{P}_{2k-1}(I_n)$ und $a_h(w, w) \in \mathbb{P}_{2k}(I_n)$ mit der Gauß–Radau-Formel exakt integriert werden, d.h. in beiden Fällen lassen sich die Quadraturformeln als Integrale schreiben. Weiterhin folgt aus der Identität $(a-b, a) = \|a-b\|_0^2 + \frac{1}{2} \|a\|_0^2 - \frac{1}{2} \|b\|_0^2$

und der Koerzivitat von A_h , vgl. (4.9)

$$\begin{aligned}\tilde{B}_n(w, w) &\geq \frac{1}{2} \int_{I_n} \frac{d}{dt} \|w(t)\|_0^2 dt + \int_{I_n} \|w(t)\|^2 dt + \frac{1}{2} \|w_{n-1}^+ - w_{n-1}\|_0^2 + \frac{1}{2} \|w_{n-1}^+\|_0^2 - \frac{1}{2} \|w_{n-1}\|_0^2 \\ &= \frac{1}{2} \|w(t_n)\|_0^2 - \frac{1}{2} \|w(t_{n-1}^+)\|_0^2 + \int_{I_n} \|w(t)\|^2 dt + \|[w]_{n-1}\|_0^2 + \frac{1}{2} \|w_{n-1}^+\|_0^2 - \frac{1}{2} \|w_{n-1}\|_0^2 \\ &= \frac{1}{2} \|w(t_n)\|_0^2 - \frac{1}{2} \|w(t_{n-1})\|_0^2 + \|[w]_{n-1}\|_0^2 + Q_n[\|w\|^2].\end{aligned}$$

Dies ist was zu zeigen ist. \square

Herleitung einer Fehlergleichung fur die Zeitableitung:

Wir zerlegen den Fehler bei der Zeitableitung wie im cGP-Fall und setzen

$$\begin{aligned}\hat{e}(t) &:= \partial_t u(t) - \partial_t \tilde{u}_{\tau h}(t), \\ \hat{\eta}(t) &:= \partial_t u(t) - P_h \partial_t R_{k+1}^d u(t), \\ \hat{e}_{\tau h}(t) &:= P_h \partial_t R_{k+1}^d u(t) - \partial_t \tilde{u}_{\tau h}(t).\end{aligned}$$

Es gilt

$$\hat{e} = \hat{\eta} + \hat{e}_{\tau h}. \quad (4.104)$$

Aus der Definition der Bilinearform \tilde{B}_n folgt

$$\begin{aligned}\tilde{B}_n(\partial_t u, v_{\tau h}) &= Q_n[\langle \partial_t^2 u + A_h \partial_t u, v_{\tau h} \rangle] + ([\partial_t u]_{n-1}, v_{n-1}^+) \\ &= Q_n[\langle \partial_t(\partial_t u + A_h u), v_{\tau h} \rangle],\end{aligned}$$

wobei $[\partial_t u]_{n-1} = 0$, da die exakte Losung u glatt ist. Aus der L^2 -Projektion folgt $A_h u = Au + S_h u$, somit ist

$$\begin{aligned}\tilde{B}_n(\partial_t u, v_{\tau h}) &= Q_n[\langle \partial_t(\partial_t u + A \partial_t u + S_h u), v_{\tau h} \rangle] \\ &= Q_n[\langle \partial_t(f + S_h u), v_{\tau h} \rangle] \\ &= Q_n[\langle \partial_t f, v_{\tau h} \rangle] + Q_n[\langle S_h \partial_t u, v_{\tau h} \rangle].\end{aligned} \quad (4.105)$$

Hier haben wir ausgenutzt, dass A bzw. S_h von der Zeit unabhangig sind und andererseits haben wir

$$\begin{aligned}\tilde{B}_n(\partial_t \tilde{u}_{\tau h}, v_{\tau h}) &= Q_n[\langle \partial_t^2 \tilde{u}_{\tau h} + A_h \partial_t \tilde{u}_{\tau h}, v_{\tau h} \rangle] + ([\partial_t \tilde{u}_{\tau h}]_{n-1}, v_{n-1}^+) \\ &= Q_n[\langle \partial_t(\partial_t \tilde{u}_{\tau h} + A_h(u_{\tau h} + d_n \vartheta_n), v_{\tau h} \rangle] + ([\partial_t \tilde{u}_{\tau h}]_{n-1}, v_{n-1}^+) \\ &= Q_n[\langle \partial_t(\partial_t \tilde{u}_{\tau h} + A_h u_{\tau h}), v_{\tau h} \rangle] + Q_n[\langle A_h d_n \partial_t \vartheta_n, v_{\tau h} \rangle] + ([\partial_t \tilde{u}_{\tau h}]_{n-1}, v_{n-1}^+) \\ &= Q_n[\langle \partial_t P_h I_k^d f, v_{\tau h} \rangle] + Q_n[\langle A_h d_n \partial_t \vartheta_n, v_{\tau h} \rangle] + ([\partial_t \tilde{u}_{\tau h}]_{n-1}, v_{n-1}^+),\end{aligned}$$

wobei wir die Beziehung aus Lemma 4.32 nach der Zeit differenziert und verwendet haben. Aus der Definition der L^2 -Projektion P_h folgt schlielich

$$\tilde{B}_n(\partial_t \tilde{u}_{\tau h}, v_{\tau h}) = Q_n[\langle \partial_t I_k^d f, v_{\tau h} \rangle] + Q_n[\langle A_h d_n \partial_t \vartheta_n, v_{\tau h} \rangle] + ([\partial_t \tilde{u}_{\tau h}]_{n-1}, v_{n-1}^+). \quad (4.106)$$

Die beiden Faktoren $\partial_t \vartheta_n$ und $v_{\tau h}$ liegen in $\mathbb{P}_k(I_n, \mathcal{V}_h)$ und darausfolgend ist

$$\begin{aligned}Q_n[\langle A_h d_n \partial_t \vartheta_n, v_{\tau h} \rangle] &= \int_{I_n} \langle A_h d_n \partial_t \vartheta_n(t), v_{\tau h}(t) \rangle dt \\ &= - \int_{I_n} \langle A_h d_n \vartheta_n(t), \partial_t v_{\tau h}(t) \rangle dt + \langle A_h d_n \vartheta_n(t_n), v_{\tau h}(t_n) \rangle \\ &\quad - \langle A_h d_n \vartheta_n(t_{n-1}^+), v_{\tau h}(t_{n-1}^+) \rangle \\ &= Q_n[\langle A_h d_n \vartheta_n, \partial_t v_{\tau h} \rangle] - \langle A_h d_n \vartheta_n(t_{n-1}^+), v_{n-1}^+ \rangle \\ &= - \langle A_h d_n \vartheta_n(t_{n-1}^+), v_{n-1}^+ \rangle.\end{aligned}$$

Dabei haben wir sowohl eine partielle Integration verwendet, als auch den Fakt, dass ϑ_n an allen Gauß–Radau-Punkten verschwindet. Weiterhin folgt aus (4.92)

$$-A_h d_n \partial_t \vartheta_n = A_h [u_{\tau h}]_{n-1},$$

somit erhalten wir

$$Q_n[\langle A_h d_n \partial_t \vartheta_n, v_{\tau h} \rangle] = \langle A_h [u_{\tau h}]_{n-1}, v_{n-1}^+ \rangle. \quad (4.107)$$

Aus (4.105), (4.106) und (4.107) ergibt sich

$$\begin{aligned} \tilde{B}_n(\partial_t(u - \tilde{u}_{\tau h}), v_{\tau h}) &= Q_n[(\partial_t f, v_{\tau h})] - Q_n[(\partial_t I_k^d f, v_{\tau h})] + Q_n[\langle S_h \partial_t u, v_{\tau h} \rangle] \\ &\quad - \langle A_h [u_{\tau h}]_{n-1}, v_{n-1}^+ \rangle - ([\partial_t \tilde{u}_{\tau h}]_{n-1}, v_{n-1}^+) \\ &= Q_n[(\partial_t f, v_{\tau h})] - Q_n[(\partial_t I_k^d f, v_{\tau h})] + Q_n[\langle S_h \partial_t u, v_{\tau h} \rangle] \\ &\quad - ([P_h I_k^d f]_{n-1}, v_{n-1}^+). \end{aligned} \quad (4.108)$$

Dabei folgt der letzte Schritt aus Lemma 4.32. Weiterhin lässt sich der zweite Term auf der rechten Seite anhand einer partiellen Integration wie folgt formulieren:

$$\begin{aligned} Q_n[(\partial_t I_k^d f, v_{\tau h})] &= \int_{I_n} (\partial_t I_k^d f(t), v_{\tau h}(t)) dt \\ &= - \int_{I_n} (I_k^d f(t), \partial_t v_{\tau h}(t)) dt + (I_k^d f(t_n), v_{\tau h}(t_n)) - (I_k^d f(t_{n-1}^+), v_{\tau h}(t_{n-1}^+)) \\ &= -Q_n[(I_k^d f, \partial_t v_{\tau h})] + (I_k^d f(t_n), v_{\tau h}(t_n)) - (I_k^d f(t_{n-1}^+), v_{\tau h}(t_{n-1}^+)) \\ &= -Q_n[(L_{k+1}^d f, \partial_t v_{\tau h})] + (f(t_n), v_{\tau h}(t_n)) - (I_k^d f(t_{n-1}^+), v_{\tau h}(t_{n-1}^+)). \end{aligned} \quad (4.109)$$

Wir haben hier den Lagrange-Interpoplator L_{k+1}^d eingeführt, der mit I_k^d an allen Gauß–Radau-Punkten übereinstimmt. Eine weitere partielle Integration liefert

$$\begin{aligned} -Q_n[(L_{k+1}^d f, \partial_t v_{\tau h})] &= - \int_{I_n} (L_{k+1}^d f(t), \partial_t v_{\tau h}(t)) dt \\ &= \int_{I_n} (\partial_t L_{k+1}^d f(t), v_{\tau h}(t)) dt - (L_{k+1}^d f(t_n), v_{\tau h}(t_n)) + (L_{k+1}^d f(t_{n-1}^+), v_{\tau h}(t_{n-1}^+)) \\ &= Q_n[(\partial_t L_{k+1}^d f, v_{\tau h})] - (f(t_n), v_{\tau h}(t_n)) + (f(t_{n-1}), v_{\tau h}(t_{n-1}^+)). \end{aligned}$$

Zusammen mit (4.109) ergibt dies

$$\begin{aligned} Q_n[(\partial_t I_k^d f, v_{\tau h})] &= Q_n[(\partial_t L_{k+1}^d f, v_{\tau h})] + (f(t_{n-1}), v_{\tau h}(t_{n-1}^+)) - (I_k^d f(t_{n-1}^+), v_{\tau h}(t_{n-1}^+)) \\ &= Q_n[(\partial_t L_{k+1}^d f, v_{\tau h})] + (I_k^d f(t_{n-1}^-), v_{\tau h}(t_{n-1}^+)) - (I_k^d f(t_{n-1}^+), v_{\tau h}(t_{n-1}^+)) \\ &= Q_n[(\partial_t L_{k+1}^d f, v_{\tau h})] - ([I_k^d f]_{n-1}, v_{\tau h}(t_{n-1}^+)). \end{aligned} \quad (4.110)$$

Abschließend folgt aus der Kombination von (4.108) und (4.110)

$$\tilde{B}_n(\partial_t(u - \tilde{u}_{\tau h}), v_{\tau h}) = Q_n[(\partial_t f - \partial_t I_k^d f, v_{\tau h})] + Q_n[\langle S_h \partial_t u, v_{\tau h} \rangle]. \quad (4.111)$$

Diese Umformulierung von \tilde{B}_n ist bis auf die verwendeten Quadraturpunkte ähnlich zu der von B_n im cGP-Fall, siehe (4.63). Somit erhalten wir die gleichen Fehlerabschätzungen, obwohl die entstandenen volldiskreten Lösungen unterschiedlich sind. Mit denselben Vorgehensweisen wie bei den Beweisen von Lemmata 4.18 - 4.20 kann man die folgenden Fehlerabschätzungen mithilfe der entsprechenden Ergebnisse aus dem dG-Fall zeigen.

Lemma 4.34. [*Fehlerabschätzung für die Zeitableitung*] Es gilt

$$|B_n(\hat{e}_{\tau h}, v_{\tau h})|^2 \leq C\tau_n \left\{ \widehat{E}_1^n(u) + \widehat{E}_2^n(u) + \widehat{E}_3^n(u) + \widehat{E}_4^n(f) \right\} Q_n [\|v_{\tau h}\|^2],$$

wobei die Fehler im Ort durch

$$\begin{aligned} \widehat{E}_1^n(u) &:= h^{2r+1} \sup_{t \in I_n} \|\partial_t u\|_{r+1}^2 + h^{2r+1} \sup_{t \in I_n} \|\partial_t^2 u\|_{r+1}^2, \\ \widehat{E}_2^n(u) &:= (\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2 \end{aligned}$$

gegeben sind, wohingegen die Fehler in der Zeit durch

$$\begin{aligned} \widehat{E}_3^n(u) &:= \tau_n^{2(k+1)} \left\{ \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + \sup_{t \in I_n} \|\partial_t^{k+3} u(t)\|_1^2 \right\}, \\ \widehat{E}_4^n(f) &:= \tau_n^{2(k+1)} \sup_{t \in I_n} \|\partial_t^{k+2} f(t)\|_0^2 \end{aligned}$$

gegeben sind.

Lemma 4.35. [*$\hat{e}_{\tau h}$ -Fehlerabschätzung*] Es gilt für $m = 1, \dots, N$

$$\|\hat{e}_{\tau h}(t_m)\|_0^2 + \sum_{n=1}^m Q_n [\|\hat{e}_{\tau h}\|^2] \leq C \sum_{n=1}^m \tau_n \left\{ \widehat{E}_1^n(u) + \widehat{E}_2^n(u) + \widehat{E}_3^n(u) + \widehat{E}_4^n(f) \right\}, \quad (4.112)$$

wobei E_i^n , $i = 1, \dots, 4$, dem Lemma 4.34 zu entnehmen sind. Weiterhin erhalten wir die L^2 -Abschätzung

$$\|\hat{e}_{\tau h}\|_{L^2(I, L^2(\Omega))}^2 \leq C \sum_{n=1}^N \tau_n \left\{ \widehat{E}_1^n(u) + \widehat{E}_2^n(u) + \widehat{E}_3^n(u) + \widehat{E}_4^n(f) \right\}.$$

Beweis. Es folgt aus dem Einsetzen von $w = \hat{e}_{\tau h}$ in der Stabilitätsungleichung in Lemma 4.33, dem Einsetzen von $v_{\tau h} = \hat{e}_{\tau h}$ in der Fehlerabschätzung in Lemma 4.34 und die Kombination der Abschätzungen in beiden Lemmata

$$\begin{aligned} &\frac{1}{2} \|\hat{e}_{\tau h}(t_n)\|_0^2 - \frac{1}{2} \|\hat{e}_{\tau h}(t_{n-1})\|_0^2 + Q_n [\|\hat{e}_{\tau h}\|^2] \\ &\leq C\tau_n \left\{ \widehat{E}_1^n(u) + \widehat{E}_2^n(u) + \widehat{E}_3^n(u) + \widehat{E}_4^n(f) \right\}^{1/2} \left\{ Q_n [\|\hat{e}_{\tau h}\|^2] \right\}^{1/2}. \end{aligned}$$

Aus der Young-Ungleichung erhalten wir

$$\begin{aligned} &\frac{1}{2} \|\hat{e}_{\tau h}(t_n)\|_0^2 - \frac{1}{2} \|\hat{e}_{\tau h}(t_{n-1})\|_0^2 + Q_n [\|\hat{e}_{\tau h}\|^2] \\ &\leq C\tau_n \left\{ \widehat{E}_1^n(u) + \widehat{E}_2^n(u) + (\widehat{E}_3^n(u))^2 + \widehat{E}_4^n(f) \right\}. \end{aligned}$$

Weiterhin liefert das Aufsummieren der Abschätzung (4.101) für $n = 1, \dots, m$ die Behauptung (4.64), wobei wir von der Eigenschaft $\tilde{P}_h u_0 = u_h(0)$ Gebrauch gemacht haben.

Die zweite Aussage lässt sich analog zum Lemma 4.19 beweisen. \square

Die folgenden Lemmata lassen sich wie im cGP-Fall zeigen, siehe 4.20 und 4.21.

Lemma 4.36. [*L^2 -Fehlerabschätzung*] Es gilt für $m = 1, \dots, N$

$$\begin{aligned} \|\hat{e}(t_m)\|_0^2 &\leq C\tau_m \max_{1 \leq n \leq m} \left\{ \widehat{C}_{T,n}(u) \tau_n^{2(k+1)} + \widehat{C}_{T,n}(f) \tau_n^{2(k+1)} + \widehat{C}_{S,n}(u) (\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \right\} \\ &\quad + C'_{S,m} (\partial_t u) h^{2r+2}, \end{aligned}$$

wobei

$$\begin{aligned}\widehat{C}_{T,n}(u) &:= \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_1^2 + \sup_{t \in I_n} \|\partial_t^{k+3} u(t)\|_1^2, \\ \widehat{C}_{T,n}(f) &:= \sup_{t \in I_n} \|\partial_t^{k+2} f(t)\|_0^2, \\ \widehat{C}_{S,n}(u) &:= \|u\|_{C^1(I_n, H^{r+1}(\Omega))}^2 + \sup_{t \in I_n} \|\partial_t u(t)\|_{r+1}^2 + \sup_{t \in I_n} \|\partial_t^2 u(t)\|_{r+1}^2.\end{aligned}$$

$C'_{S,m}$ ist in Lemma 4.15 definiert. Weiterhin gilt

$$\|\hat{e}\|_{L^2(I, L^2(\Omega))}^2 \leq CT \max_{1 \leq n \leq N} \left\{ \widehat{C}_{T,n}(u) \tau_n^{2(k+1)} + \widehat{C}_{T,n}(f) \tau_n^{2(k+1)} + \widehat{C}_{S,n}(u) (\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \right\}.$$

Lemma 4.37. [$L^2(L^2)$ -Fehlerabschätzung] Es gilt

$$\begin{aligned}\|\tilde{e}_{\tau h}\|_{L^2(I, L^2(\Omega))}^2 &\leq CT \sum_{n=1}^N \tau_n \left\{ E_1^n(u) + E_2^n(u) + E_3^n(u) \right\} \\ &\quad + CT \tau^2 \sum_{n=1}^N \tau_n \left\{ \widehat{E}_1^n(u) + \widehat{E}_2^n(u) + \widehat{E}_3^n(u) + \widehat{E}_4^n(f) \right\},\end{aligned}\tag{4.113}$$

wobei E_i^n , $i = 1, \dots, 3$, und \widehat{E}_i^n , $i = 1, \dots, 4$, den Lemmata 4.13 und 4.18 zu entnehmen sind und $\tau := \max_{1 \leq n \leq N} \tau_n$. Weiterhin gilt

$$\begin{aligned}\|\tilde{e}\|_{L^2(I, L^2(\Omega))}^2 &\leq CT \tau^{2(k+2)} \max_{1 \leq n \leq N} \left\{ C_{T,n}(u) + \widehat{C}_{T,n}(u) + \widehat{C}_{T,n}(f) \right\} \\ &\quad + CT (\varepsilon^{1/2} + h^{1/2})^2 h^{2r} \max_{1 \leq n \leq N} \left\{ C_{S,n}(u) + \tau^2 \widehat{C}_{S,n}(u) \right\},\end{aligned}\tag{4.114}$$

wobei $C_{T,n}$, $C_{S,n}$, $\widehat{C}_{T,n}$ und $\widehat{C}_{S,n}$ in den Lemmata 4.15 und 4.20 definiert sind.

4.4 Numerische Ergebnisse

Wir präsentieren in diesem Abschnitt einige numerische Ergebnisse und diskutieren hierbei die von uns gewonnenen theoretischen Ergebnisse bezüglich der Stabilität und der Genauigkeit der betrachteten Galerkin-Verfahren in Kombination mit einer lokalen Stabilisierung im Ort (LPS).

Um den Fehler berechnen zu können, betrachten wir wissenschaftliche Testsituationen, bei denen wir eine ausgewählte exakte Lösung u in das Problem einsetzen werden. Die rechte Seite sowie die Rand- und Anfangsbedingung lassen sich dementsprechend aus der betrachteten exakten Lösung ableiten. Für die Implementierung der numerischen Ergebnisse dient MooNMD [26].

In unseren numerischen Berechnungen haben wir transformierte Finite-Elemente-Räumen verwendet [11], wobei die angereicherten Blasen-Räume auf der Referenzzelle \widehat{K} durch

$$\begin{aligned}\mathbb{P}_r^{\text{bl}}(\widehat{K}) &:= \mathbb{P}_r(\widehat{K}) + \hat{b}_\Delta \mathbb{P}_{r-1}(\widehat{K}), & r \geq 1, \\ \mathbb{Q}_r^{\text{bl}}(\widehat{K}) &:= \mathbb{Q}_r(\widehat{K}) + \text{span} \{ \hat{b}_\square \hat{x}_i^{r-1}, i = 1, 2 \}, & r \geq 1,\end{aligned}\quad (4.115)$$

gegeben sind. Dabei sind \hat{b}_Δ und \hat{b}_\square die kubische Blasen-Funktion auf dem Referenzdreieck und die biquadratische Blasen-Funktion auf dem Referenzviereck. Die auf $\mathbb{P}_r^{\text{bl}}(\widehat{K})$ und $\mathbb{Q}_r^{\text{bl}}(\widehat{K})$ basierenden transformierten Finite-Elemente-Räume werden wir mit \mathbb{P}_r^{bl} und \mathbb{Q}_r^{bl} bezeichnen.

Die Kombinationen $\mathcal{V}_h = \mathbb{P}_r^{\text{bl}}$, $\mathcal{D}_h(K) = \mathbb{P}_{r-1}(K)$, $r \geq 1$, auf Dreiecksgittern und die Kombinationen $\mathcal{V}_h = \mathbb{Q}_r^{\text{bl}}$, $\mathcal{D}_h(K) = \mathbb{P}_{r-1}(K)$, $r \geq 1$, auf Vierecksgittern erfüllen die Annahmen 4.1 und 4.2. Weitere Beispiele für Näherungsräume \mathcal{V}_h und Projektionsräume $\mathcal{D}_h(K)$, die die Annahmen 4.1 und 4.2 erfüllen, sind in [30, 35] angegeben. Als Stabilisierungsparameter betrachten wir hier

$$\beta_k = 0.1h_K \quad \forall K \in \mathcal{T}_h. \quad (4.116)$$

Mit

$$\begin{aligned}\|v\|_{L^2(L^2)} &:= \left(\int_0^T \|v(t)\|_{L^2(\Omega)}^2 dt \right)^{1/2}, \\ \|v\|_{H^1(L^2)} &:= \left(\int_0^T \|v'(t)\|_{L^2(\Omega)}^2 dt \right)^{1/2}, \\ \|v\|_{\ell^\infty(L^2)} &:= \max_{0 \leq n \leq N} \|v(t_n)\|_{L^2(\Omega)}\end{aligned}\quad (4.117)$$

definieren wir jeweils die $L^2(L^2)$ -Norm, die $H^1(L^2)$ -Norm und die diskrete $\ell^\infty(L^2)$ -Norm in der Zeit. Dazu betrachten wir die gitterabhängige Norm

$$\|v\|_{L^2(LPS)} := \left(\int_0^T \| |u(t)| \|^2 dt \right)^{1/2}.$$

Beispiel 1:

Als Modellproblem für unsere numerischen Tests wählen wir als reines Transportproblem in zwei Dimensionen das rotierende Gauß'sche Benchmark-Problem, [10, 2]. Daher betrachten wir das Problem (4.1) in

$$\Omega := \{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1 \}$$

mit den Daten

$$\varepsilon = 0, \quad b = (-y, x)^T, \quad c = f = 0, \quad T = 2\pi$$

und der in (0.3, 0.3) zentrierten Gauß'schen Anfangsbedingung

$$u_0(x, y) = \exp(-10(x - 0.3)^2 - 10(y - 0.3)^2).$$

Man beachte, dass die Lösung u zu allen Zeiten durch einen entsprechend gedrehten Ausgangszustand gegeben ist.

In diesem Beispiel sind wir an der Konvergenzordnung im Raum interessiert. Zu diesem Zweck halten wir die Fehler in der Zeit klein, indem wir das cGP(3)-Verfahren mit dem Zeitschritt $\tau = 2\pi \times 10^{-3}$ verwenden. Daher entsprechen 1000 Zeitschritte einer vollständigen Umdrehung des Gauß'schen Benchmarks. Die Berechnungen wurden auf Dreiecksgittern durchgeführt, wobei die Gitter auf den einzelnen Stufen durch sukzessive Verfeinerung mit Randanpassung aufgrund des krummen Gebietsrandes aus der Anfangstriangulierung erhalten werden. Das Ausgangsgitter (Stufe 0) und das Gitter nach dreifacher Verfeinerung (Stufe 3) sind in Abb. 4.1 dargestellt. Um der krummlinigen Berandung Rechnung zu tragen, wurden isoparametrische Finite-Elemente verwendet.

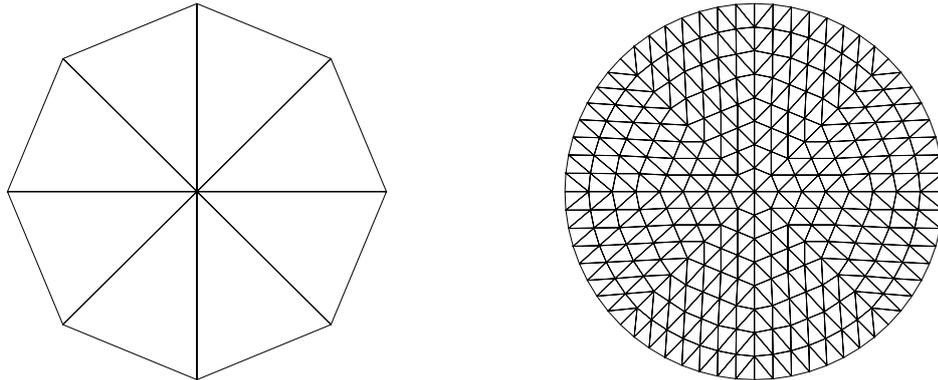


Abbildung 4.1: *Triangulierung für Beispiel 1: Ausgangsgitter (links) und Gitter nach dreifacher Verfeinerung (rechts).*

Bei den LPS-Diskretisierungen verwenden wir für das Paar $(V_h, \mathcal{D}_h(K))$ die drei folgenden Varianten:

$$(\mathbb{P}_1^{\text{bl}}, \mathbb{P}_0(K)), \quad (\mathbb{P}_2^{\text{bl}}, \mathbb{P}_1(K)) \quad \text{und} \quad (\mathbb{P}_3^{\text{bl}}, \mathbb{P}_2(K))$$

Die Tabelle 4.1 zeigt für die Zeitdiskretisierung cGP(3) die Fehler und die Konvergenzordnungen in der $L^2(LPS)$ -Norm. Es ist deutlich zu sehen, dass die in Lemmata 4.15 und 4.21 vorhergesagte Konvergenzordnung im Ort bestätigt wird.

Tabelle 4.1: *Beispiel 1: Fehler und Konvergenzordnungen für cGP(3) im Ort bezüglich der $L^2(LPS)$ -Norm.*

Stufe	$(\mathbb{P}_1^{\text{bl}}, \mathbb{P}_0(K))$		$(\mathbb{P}_2^{\text{bl}}, \mathbb{P}_1(K))$		$(\mathbb{P}_3^{\text{bl}}, \mathbb{P}_2(K))$	
	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung
2	2.525-01		4.404-02		8.091-03	
3	8.330-02	1.60	8.378-03	2.39	6.932-04	3.54
4	2.937-02	1.49	1.531-03	2.45	6.143-05	3.49
5	1.033-02	1.51	2.840-04	2.43	5.432-06	3.49
Theorie		1.5		2.5		3.5

Beispiel 2:

Wir betrachten das Problem (4.1) in $\Omega := (-1, 1) \times (-1, 1)$ mit $\varepsilon = 10^{-8}$, $b = (1+x, 2-y)^T$, $c = 1$ und $T = 1$. Die rechte Seite f und die Anfangsbedingung u_0 sind so gewählt, dass

$$u(x, y) = \sin(t) + \cos(t)x^2 + t^{100}xy$$

die Lösung des Problems (4.1) ist.

Um die Wirkung der angewendeten Zeitdiskretisierung zu beurteilen und die hierzu hergeleitete Theorie zu bestätigen, schließen wir in diesem Beispiel die räumlichen Fehler aus. Wir haben dazu die LPS-Diskretisierung mit $\mathcal{V}_h = \mathbb{Q}_1^{\text{bl}}$, $\mathcal{D}_h(K) = \mathbb{P}_0(K)$ gewählt. Für die Triangulierung wurde ein uniformes Gitter bestehend aus 8×8 Quadraten verwendet.

Als Zeitdiskretisierungen höherer Ordnung wurden cGP(2), cGP(3), cGP(4), dG(1), dG(2) und dG(3) angewendet.

Wir führen in den Tabellen 4.2–4.5 die Fehler $e = u_{\tau h} - u$ und Konvergenzordnungen der Verfahren in der $L^2(L^2)$ -, der $H^1(L^2)$ - und der $\ell^\infty(L^2)$ -Norm auf. Um einen schnellen Vergleich mit den theoretischen Ergebnissen zu ermöglichen, beinhalten alle Tabellen auch die theoretischen Konvergenzordnungen.

Die Zahlen in den Tabellen 4.2–4.4 zeigen, dass die cGP(k)-Verfahren für $k = 2, 3, 4$ von den Konvergenzordnungen $k + 1$ in der $L^2(L^2)$ -Norm und k in der $H^1(L^2)$ -Norm sind. Die dG(k)-Verfahren für $k = 1, 2, 3$ sind von den Konvergenzordnungen $k + 1$ in der $L^2(L^2)$ -Norm und k in der $H^1(L^2)$ -Norm. Aus der Tabelle 4.5 kann man entnehmen, dass die cGP(k)-Verfahren für $k = 3, 4$ in der $\ell^\infty(L^2)$ -Norm von den Konvergenzordnungen $2k$ sind. Die dG(k)-Verfahren für $k = 2, 3$ sind in der $\ell^\infty(L^2)$ -Norm von den Konvergenzordnungen $2k + 1$.

Die Tabellen 4.6–4.8 zeigen die Fehler $\tilde{e} = \tilde{u}_{\tau h} - u$ und Konvergenzordnungen der durch das Postprocessing nachbereiteten Lösung $\tilde{u}_{\tau h}$ in der $L^2(L^2)$ - und $H^1(L^2)$ -Norm. Man beachte, dass die nachbearbeiteten Lösungen sowohl für die cGP- als auch für die dG-Verfahren die Super-Konvergenzordnung $k + 2$ in der $L^2(L^2)$ -Norm und die Super-Konvergenzordnung $k + 1$ in der $H^1(L^2)$ -Norm besitzen.

Fazit:

Wir haben variationelle Zeitdiskretisierungen höherer Ordnung durch stetige Galerkin–Petrov-Verfahren (cGP) und unstetige Galerkin-Verfahren (dG) mit lokaler Projektionsstabilisierung (LPS) im Ort kombiniert. Außerdem haben wir die Fehlerabschätzungen für die cGP- bzw. dG-Verfahren angegeben und bewiesen. Es ist deutlich zu sehen, dass die in Lemmata 4.15 und 4.21 für die cGP(k)-Verfahren und die in Lemmata 4.34 und 4.37 für die dG(k)-Verfahren vorhergesagten Konvergenzordnungen durch die numerischen Beispiele bestätigt wurden.

Tabelle 4.2: *Beispiel 2: Fehler und Konvergenzordnungen für cGP(2) und dG(1) in Abhängigkeit der Anzahl der Zeitschritte N .*

N	$\ u - u_{\tau h}\ _{L^2(L^2)}$				$\ u - u_{\tau h}\ _{H^1(L^2)}$			
	cGP(2)		dG(1)		cGP(2)		dG(1)	
	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung
20	4.562-03		1.709-02		1.099+00		2.948+00	
40	7.718-04	2.56	4.580-03	1.89	4.184-01	1.39	1.784+00	0.72
80	1.062-04	2.86	1.095-03	2.06	1.251-01	1.74	9.695-01	0.87
160	1.360-05	2.96	2.576-04	2.08	3.305-02	1.92	4.937-01	0.97
320	1.711-06	2.99	6.180-05	2.05	8.386-03	1.97	2.467-01	1.00
640	2.142-07	2.99	1.509-05	2.03	2.104-03	1.99	1.230-01	1.00
1280	1.094-09	2.99	3.728-06	2.01	5.266-04	1.99	6.136-02	1.00
Theorie		3		2		2		1

Tabelle 4.3: *Beispiel 2: Fehler und Konvergenzordnungen für cGP(3) und dG(2) in Abhängigkeit der Anzahl der Zeitschritte N.*

N	$\ u - u_{\tau h}\ _{L^2(L^2)}$				$\ u - u_{\tau h}\ _{H^1(L^2)}$			
	cGP(3)		dG(2)		cGP(3)		dG(2)	
	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung
20	9.663-04		4.161-03		3.757-01		1.571+00	
40	9.597-05	3.33	7.131-04	2.54	8.185-02	2.19	5.984-01	1.39
80	7.178-06	3.74	9.699-05	2.87	1.272-02	2.68	1.749-01	1.77
160	4.734-07	3.92	1.211-05	3.00	1.697-03	2.90	4.537-02	1.94
320	3.001-08	3.97	1.492-06	3.02	2.157-04	2.97	1.138-02	1.99
640	1.882-09	3.99	1.844-07	3.01	2.709-05	2.99	2.840-03	2.00
1280	1.411-10	3.73	2.289-08	3.01	3.390-06	2.99	7.087-04	2.00
Theorie		4		3		3		2

Tabelle 4.4: *Beispiel 2: Fehler und Konvergenzordnungen für cGP(4) und dG(3) in Abhängigkeit der Anzahl der Zeitschritte N.*

N	$\ u - u_{\tau h}\ _{L^2(L^2)}$				$\ u - u_{\tau h}\ _{H^1(L^2)}$			
	cGP(4)		dG(3)		cGP(4)		dG(3)	
	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung
20	1.928-04		9.866-04		1.046-01		6.282-01	
40	1.060-05	4.18	9.623-05	3.35	1.217-02	3.10	1.327-01	2.24
80	4.114-07	4.68	6.943-06	3.79	9.620-04	3.66	2.001-02	2.72
160	1.370-08	4.90	4.446-07	3.96	6.440-05	3.90	2.620-03	2.93
320	4.357-10	4.97	2.767-08	4.00	4.098-06	3.97	3.298-04	2.98
640	1.490-11	4.87	1.719-09	4.00	2.573-07	3.99	4.118-05	3.00
Theorie		5		4		4		3

Tabelle 4.5: *Beispiel 2: Fehler und Konvergenzordnungen für cGP(3), cGP(4), dG(2) und dG(3) in Abhängigkeit der Anzahl der Zeitschritte N.*

N	$\ u - u_{\tau h}\ _{\ell^\infty(L^2)}$				$\ u - u_{\tau h}\ _{\ell^\infty(L^2)}$			
	cGP(3)		dG(2)		cGP(4)		dG(3)	
	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung
20	5.021-02		1.648-01		4.412-03		1.826-02	
40	2.240-03	4.48	1.404-02	3.55	5.472-05	6.33	4.528-04	5.33
80	5.206-05	5.42	6.624-04	4.40	3.327-07	7.36	5.583-06	6.34
160	9.154-07	5.82	2.368-05	4.80	1.505-09	7.78	5.056-08	6.78
320	1.475-08	5.95	7.715-07	4.93				
Theorie		6		5		8		7

Tabelle 4.6: *Beispiel 2: Fehler und Konvergenzordnungen für cGP(2) und dG(1) in Abhängigkeit der Anzahl der Zeitschritte N für die nachbereitete Lösung.*

N	$\ u - \tilde{u}_{\tau h}\ _{L^2(L^2)}$				$\ u - \tilde{u}_{\tau h}\ _{H^1(L^2)}$			
	cGP(2)		dG(1)		cGP(2)		dG(1)	
	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung
20	6.350-03		1.999-02		6.839-01		1.958+00	
40	7.869-04	3.01	3.797-03	2.39	1.499-01	2.18	7.500-01	1.38
80	6.470-05	3.60	5.557-04	2.77	2.327-02	2.68	2.208-01	1.76
160	4.385-06	3.88	7.137-05	2.96	3.101-03	2.90	5.761-02	1.93
320	2.799-07	3.96	8.872-06	3.00	3.942-04	2.97	1.450-02	1.99
640	1.758-08	3.99	1.100-06	3.01	4.947-05	2.99	3.625-03	2.00
1280	1.180-09	3.89	1.367-07	3.00	6.674-06	2.88	9.053-04	2.00
Theorie		4		3		3		2

Tabelle 4.7: *Beispiel 2: Fehler und Konvergenzordnungen für cGP(3) und dG(2) in Abhängigkeit der Anzahl der Zeitschritte N für die nachbereitete Lösung.*

N	$\ u - \tilde{u}_{\tau h}\ _{L^2(L^2)}$				$\ u - \tilde{u}_{\tau h}\ _{H^1(L^2)}$			
	cGP(3)		dG(2)		cGP(3)		dG(2)	
	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung
20	1.235-03		4.433-03		1.754-01		6.831-01	
40	7.512-05	4.03	4.792-04	3.20	2.057-02	3.09	1.470-01	2.21
80	2.993-06	4.64	3.616-05	3.72	1.627-03	3.66	2.251-02	2.70
160	1.003-07	4.89	2.361-06	3.93	1.088-04	3.90	2.972-03	2.92
320	3.193-09	4.97	1.483-07	3.99	6.926-06	3.97	3.758-04	2.98
640	1.005-10	4.98	9.249-09	4.00	4.347-07	3.99	4.704-05	2.99
1280			5.855-10	3.98	2.724-08	3.99	5.877-06	3.00
Theorie		5		4		4		3

Tabelle 4.8: *Beispiel 2: Fehler und Konvergenzordnungen für cGP(4) und dG(3) in Abhängigkeit der Anzahl der Zeitschritte N für die nachbereitete Lösung.*

N	$\ u - \tilde{u}_{h,\tau}\ _{L^2(L^2)}$				$\ u - \tilde{u}_{h,\tau}\ _{H^1(L^2)}$			
	cGP(4)		dG(3)		cGP(4)		dG(3)	
	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung	Fehler	Ordnung
20	2.007-04		8.742-04		3.884-02		2.001-01	
40	6.167-06	5.02	4.907-05	4.15	2.359-03	4.04	2.277-02	3.13
80	1.232-07	5.64	1.882-06	4.70	9.397-05	4.64	1.768-03	3.68
160	2.066-09	5.89	6.189-08	4.92	3.149-06	4.89	1.171-04	3.91
320	3.633-11	5.82	1.950-09	4.98	1.001-07	4.97	7.417-06	3.98
Theorie		6		5		5		4

5 Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten

In diesem Kapitel beginnen wir mit der Präsentation der Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten, wobei die Bewegung der Gebietsgrenze vorbeschrieben ist. Die zeitliche Änderung der Gebiete wird durch die ALE-Formulierung behandelt. Für die Zeitdiskretisierung werden wir die unstetigen Galerkin-Verfahren (dG) als variationelle Zeitdiskretisierungen höherer Ordnung anwenden. Dabei untersuchen wir weiterhin sowohl die Stabilität als auch eine a-priori Fehlerabschätzung.

5.1 Problemstellung

Als Motivation betrachten wir für u in einem beschränkten zeitabhängigen Gebiet $\Omega_t \subset \mathbb{R}^d$, $d = 2, 3$, wobei $t \in I := [0, T]$, die Konvektions-Diffusions-Reaktions-Gleichung:

$$\begin{aligned} \partial_t u - \varepsilon \Delta_{\mathbf{x}} u + \mathbf{b} \cdot \nabla_{\mathbf{x}} u + cu &= f, & \mathbf{x} \in \Omega_t, t \in I, \\ u &= 0, & \mathbf{x} \in \partial\Omega_t, t \in I, \\ u &= u_0, & \mathbf{x} \in \Omega_0, t = 0. \end{aligned} \quad (5.1)$$

$0 < \varepsilon \ll 1$ ist eine kleine positive Diffusionskonstante. \mathbf{b} und c sind zeitabhängige Funktionen, die von \mathbf{x} abhängen können. Diese bezeichnen jeweils die Konvektion und die Reaktion und erfüllen

$$c(t, \mathbf{x}) - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}(t, \mathbf{x}) \geq c_0 > 0 \quad \forall \mathbf{x} \in \Omega_t \quad \forall t \in I. \quad (5.2)$$

Hierbei ist c_0 eine Konstante. Die äußere Kraft f und der Anfangswert u_0 sind gegeben und erfüllen die folgenden Eigenschaften

$$f \in L^2(\mathcal{S}) \quad \text{und} \quad u_0 \in H_0^1(\Omega_0).$$

Dabei ist \mathcal{S} , der in (2.1) definierte Zeit-Raum-Zylinder. Der Einfachheit halber haben wir das Problem mit einer homogenen Dirichlet-Randbedingung ausgestattet. Mit $\nabla_{\mathbf{x}}$, $\nabla_{\mathbf{x}} \cdot$ und $\Delta_{\mathbf{x}}$ bezeichnen wir jeweils den Gradienten, die Divergenz und den Laplace-Operator bezüglich der Ortsvariable \mathbf{x} .

Die Bedingung (5.2) stellt, ebenso wie bei festen Gebieten, keine Einschränkung der Allgemeinheit dar. Eine ausführliche Erklärung diesbezüglich kann man dem Unterabschnitt 4.1 entnehmen. Diese lässt sich auch auf zeitabhängige Gebiete und ebenso auf zeitabhängige Konvektions- bzw. Reaktionsterme verallgemeinern.

Um die schwache Formulierung im ALE-Rahmen herzuleiten, betrachten wir den Raum $\mathcal{Y}(\widehat{\Omega}) = H_0^1(\widehat{\Omega})$ der zulässigen Testfunktionen $\hat{v} : \widehat{\Omega} \rightarrow \mathbb{R}$ auf dem Referenzgebiet $\widehat{\Omega}$. Wir transformieren die Testfunktionen aus $\mathcal{Y}(\widehat{\Omega})$ mit dem Homöomorphismus \mathcal{A}_t und erhalten die Testfunktionen $v = \hat{v} \circ \mathcal{A}_t^{-1} \in H^1(H_0^1(\Omega_t))$, mit deren Hilfe wir den Raum

$$\mathcal{X} := \left\{ v : \mathcal{S} \rightarrow \mathbb{R} : v = \hat{v} \circ \mathcal{A}_t^{-1}, \quad \hat{v} \in \mathcal{Y}(\widehat{\Omega}), \quad t \in I \right\}$$

der Testfunktionen auf \mathcal{S} konstruieren. Im Folgenden verwenden wir für die zeit- und ortsabhängigen Funktionen die vereinfachte Schreibweise $w(t)$ anstatt $w(t, \mathbf{x})$. Im Laufe dieses Kapitels werden wir eine Konstante C verwenden, die je nach Einsatz und auch von Schritt zu Schritt variieren kann.

5.1.1 Stabilitätsanalyse der Differentialgleichung im ALE-Rahmen

In diesem Unterabschnitt überprüfen wir, ob die im ALE-Rahmen geschriebene Differentialgleichung die Stabilitätseigenschaften einer Konvektions-Diffusions-Reaktions-Gleichung auf einem festen Gebiet beibehält. Bezüglich der Konvektionsgeschwindigkeit betrachten wir die Eigenschaft (5.2).

Bevor wir das Stabilitätsergebnis präsentieren, erläutern wir im Folgenden eine Abschätzung, die wir benötigen werden. Diese Abschätzung lässt sich anhand der Cauchy-Schwarz- und Young-Ungleichung auf $[t_i, t_j]$ für alle $v \in \mathcal{X}$ und $f \in L^2(\mathcal{S})$ in einfachen Schritten zeigen.

$$\int_{t_i}^{t_j} (f(s), v(s))_{\Omega_s} ds \leq \frac{1}{2c_0} \int_{t_i}^{t_j} \|f(s)\|_{L^2(\Omega_s)}^2 ds + \frac{c_0}{2} \int_{t_i}^{t_j} \|v(s)\|_{L^2(\Omega_s)}^2 ds \quad (5.3)$$

für alle $0 \leq t_i < t_j \leq T$.

Nach dem Einsetzen von (2.2) lässt sich die erste Gleichung in (5.1) im ALE-Rahmen folgendermaßen umschreiben:

$$D_t u - \varepsilon \Delta_{\mathbf{x}} u + (\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} u + cu = f, \quad \mathbf{x} \in \Omega_t, t \in I. \quad (5.4)$$

Weiterhin erhalten wir aus einer partiellen Integration im Ort die Identität

$$\begin{aligned} (cu, u)_{\Omega_t} + ((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} u, u)_{\Omega_t} &= (cu, u)_{\Omega_t} + \int_{\Omega_t} \frac{\mathbf{b} - \mathbf{w}}{2} \cdot \nabla_{\mathbf{x}}(u^2) d\mathbf{x} \\ &= (cu, u)_{\Omega_t} - \frac{1}{2} \int_{\Omega_t} (\nabla_{\mathbf{x}} \cdot (\mathbf{b} - \mathbf{w})) u^2 d\mathbf{x} \\ &\quad + \frac{1}{2} \int_{\partial\Omega_t} (\mathbf{b} - \mathbf{w}) \cdot \mathbf{n} u^2 d\Gamma \\ &= \left(c - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}, u^2 \right)_{\Omega_t} + \frac{1}{2} (\nabla_{\mathbf{x}} \cdot \mathbf{w}, u^2)_{\Omega_t}. \end{aligned} \quad (5.5)$$

Der letzte Schritt folgt aus $\int_{\partial\Omega_t} (\mathbf{b} - \mathbf{w}) \cdot \mathbf{n} u^2 d\Gamma = 0$, da $u|_{\partial\Omega_t} = 0$. Dabei ist \mathbf{n} das äußere Normalenvektorfeld an $\partial\Omega_t$.

Nun geben wir das Stabilitätsergebnis für das Problem (5.4) an.

Satz 5.1. *[Stabilität] Die Lösung u des Problems (5.4) erfüllt*

$$\begin{aligned} \|u(t)\|_{L^2(\Omega_t)}^2 + 2\varepsilon \int_{t_0}^t |u(s)|_{1, \Omega_s}^2 ds + c_0 \int_{t_0}^t \|u(s)\|_{L^2(\Omega_s)}^2 ds \\ \leq \frac{1}{c_0} \int_{t_0}^t \|f(s)\|_{L^2(\Omega_s)}^2 ds + \|u(t_0)\|_{L^2(\Omega_{t_0})}^2. \end{aligned}$$

Beweis. Wir multiplizieren die Gleichung (5.4) mit u , integrieren im Ort partiell und erhalten

$$(D_t u, u)_{\Omega_t} + \varepsilon (\nabla_{\mathbf{x}} u, \nabla_{\mathbf{x}} u)_{\Omega_t} + ((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} u, u)_{\Omega_t} + (cu, u)_{\Omega_t} = (f, u)_{\Omega_t}. \quad (5.6)$$

Aus der Reynolds-Identität (2.6) erhalten wir

$$(D_t u, u)_{\Omega_t} = \frac{1}{2} \int_{\Omega_t} D_t u^2 d\mathbf{x} = \frac{1}{2} \frac{d}{dt} \|u\|_{L^2(\Omega_t)}^2 - \frac{1}{2} \int_{\Omega_t} u^2 \nabla_{\mathbf{x}} \cdot \mathbf{w} d\mathbf{x}. \quad (5.7)$$

Mithilfe der Identitäten (5.7) und (5.5) lässt sich (5.6) wie folgt umschreiben:

$$\begin{aligned} (f, u)_{\Omega_t} &= \frac{1}{2} \frac{d}{dt} \|u\|_{L^2(\Omega_t)}^2 - \frac{1}{2} (\nabla_{\mathbf{x}} \cdot \mathbf{w}, u^2)_{\Omega_t} + \varepsilon |u|_{1, \Omega_t}^2 + \left(c - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}, u^2 \right)_{\Omega_t} + \frac{1}{2} (\nabla_{\mathbf{x}} \cdot \mathbf{w}, u^2)_{\Omega_t} \\ &= \frac{1}{2} \frac{d}{dt} \|u\|_{L^2(\Omega_t)}^2 + \varepsilon |u|_{1, \Omega_t}^2 + \left(c - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}, u^2 \right)_{\Omega_t} \\ &\geq \frac{1}{2} \frac{d}{dt} \|u\|_{L^2(\Omega_t)}^2 + \varepsilon |u|_{1, \Omega_t}^2 + c_0 \|u\|_{L^2(\Omega_t)}^2. \end{aligned}$$

Die Integration über $[t_0, t]$ liefert

$$\begin{aligned} \frac{1}{2}\|u(t)\|_{L^2(\Omega_t)}^2 - \frac{1}{2}\|u(t_0)\|_{L^2(\Omega_t)}^2 + \varepsilon \int_{t_0}^t |u(s)|_{1,\Omega_s}^2 ds + c_0 \int_{t_0}^t \|u(s)\|_{L^2(\Omega_s)}^2 ds \\ \leq \int_{t_0}^t (f(s), u(s))_{\Omega_s} ds \\ \leq \frac{1}{2c_0} \int_{t_0}^t \|f(s)\|_{L^2(\Omega_t)}^2 ds + \frac{c_0}{2} \int_{t_0}^t \|u(s)\|_{L^2(\Omega_s)}^2 ds, \end{aligned}$$

wobei der letzte Schritt aus (5.3) folgt. Dadurch erhalten wir unmittelbar eine Stabilitätsabschätzung, die anstatt des Terms $\frac{1}{\varepsilon}$, der für sehr kleine ε die Abschätzung zu grob werden lässt, nur die Konstante $\frac{1}{c_0}$ enthält, vgl. Bonito und Co. [7]. \square

5.2 Schwache Formulierung im ALE-Rahmen

Für den Übergang zur schwachen (variationellen) Formulierung multiplizieren wir die Differentialgleichung (5.1) mit einer zulässigen Testfunktion $v \in \mathcal{X}$. Schließlich integrieren wir wie üblich über Ω_t und erhalten

$$(\partial_t u, v)_{\Omega_t} + (\mathbf{b} \cdot \nabla_{\mathbf{x}} u, v)_{\Omega_t} + \varepsilon (\nabla_{\mathbf{x}} u, \nabla_{\mathbf{x}} v)_{\Omega_t} + (cu, v)_{\Omega_t} = (f, v)_{\Omega_t} \quad \forall v \in \mathcal{X}. \quad (5.8)$$

Diese Formulierung enthält Zeitableitungen im Euler-Rahmen. Die Testfunktionen v müssen zeitabhängig sein, da die wesentliche Randbedingung $v = 0$ auf $\partial\Omega_t$ verlangt wird. Es ist daher sinnvoll, dass alle Variablen der Entwicklung des Gebietes folgen. Dementsprechend wird (5.8) im ALE-Rahmen umgeschrieben. Die Wohldefiniertheit der Zeitableitungen der Funktionen in \mathcal{X} folgt aus der Annahme, dass \mathcal{A}_t ein Homöomorphismus ist. Daraus folgt die Glätte von u und v . Die Regularität der ALE-Abbildungen garantiert die Existenz einer eindeutigen Lösung $u \in \mathcal{X}$.

5.2.1 Nichtkonservative ALE-Formulierung

Die nichtkonservative Variante der schwachen Formulierung folgt unmittelbar aus der variationellen Formulierung. Dabei ersetzen wir die Eulersche partielle Zeitableitung ∂_t durch die ALE-Zeitableitung D_t . Nachdem wir (2.2) in (5.8) eingesetzt haben, lässt sich das Problem folgendermaßen formulieren:

Finde $u \in \mathcal{X}$ mit $u(0) = u_0$, sodass

$$(D_t u, v)_{\Omega_t} + \varepsilon (\nabla_{\mathbf{x}} u, \nabla_{\mathbf{x}} v)_{\Omega_t} + ((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} u, v)_{\Omega_t} + (cu, v)_{\Omega_t} = (f, v)_{\Omega_t} \quad \forall v \in \mathcal{X}. \quad (5.9)$$

Es ist offensichtlich, dass aus $u \in H^1(H_0^1(\Omega_t))$ die Stetigkeit von $u : I \rightarrow L^2(\Omega_t)$ folgt. Dies garantiert die Wohlgestelltheit der Anfangsbedingung $u(0) = u_0$.

5.2.2 Konservative ALE-Formulierung

Wir konstruieren hier eine konservative Formulierung, indem wir die verallgemeinerte Reynolds-Identität (2.6) verwenden. Durch das Einsetzen der Reynolds-Transportformel (2.6) in (5.9) erhalten wir die konservative Variante:

Finde $u \in \mathcal{X}$ mit $u(0) = u_0$, sodass

$$\begin{aligned} \frac{\partial}{\partial t} (u, v)_{\Omega_t} - ((\nabla_{\mathbf{x}} \cdot \mathbf{w})u, v)_{\Omega_t} + ((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} u, v)_{\Omega_t} + \varepsilon (\nabla_{\mathbf{x}} u, \nabla_{\mathbf{x}} v)_{\Omega_t} \\ + (cu, v)_{\Omega_t} - (u, D_t v)_{\Omega_t} = (f, v)_{\Omega_t} \quad \forall v \in \mathcal{X}. \end{aligned} \quad (5.10)$$

Man beachte, der einzige Unterschied zwischen den beiden Formulierungen ist, dass der Term $(D_t u, v)_{\Omega_t}$ in der nichtkonservativen Form durch $\frac{\partial}{\partial t}(u, v)_{\Omega_t} - ((\nabla_{\mathbf{x}} \cdot \mathbf{w})u, v)_{\Omega_t} - (u, D_t v)_{\Omega_t}$ in der konservativen Form ersetzt wird.

Der Einfachheit halber führen wir die folgende Notation ein:

$$A_n := \max_{\substack{r=0,1,2 \\ l=0,1}} \|\nabla_{\mathbf{Y}} \mathcal{A}_{t_n \rightarrow t}\|_{W^{r+1,\infty}(I_n, L^\infty(\Omega_n))}^l \|(\nabla_{\mathbf{Y}} \mathcal{A}_{t_n \rightarrow t})^{-1}\|_{W^{r,\infty}(I_n, L^\infty(\Omega_n))}^{1-l}. \quad (5.11)$$

Um die Bewegung des Gebietes zu behandeln, werden wir im Folgenden die dG-Zeitdiskretisierung vorstellen. Erwähnenswert ist vor allem, dass hier keine CFL-Restriktionen notwendig sind. Zur Vermeidung der Komplexität bei der Gebietsbehandlung verzichten wir vorerst auf die Ortsdiskretisierung. Diese werden wir in Kapitel 6 behandeln.

5.3 Zeitdiskrete ALE-Abbildung und Gittergeschwindigkeit

Bevor wir uns mit den dG-Verfahren in zeitabhängigen Gebieten befassen, diskretisieren wir zunächst einmal die ALE-Abbildung \mathcal{A}_t sowie die Gittergeschwindigkeit \mathbf{w} und approximieren den Raum Ω_t in der Zeit.

5.3.1 Exakte Gittergeschwindigkeit

Im Falle einer vorbeschriebenen stetigen Bewegung \mathcal{A}_t ist die Gitterposition an jedem Zeitpunkt bekannt. Die Gitter- bzw. die Gebietsgeschwindigkeit ist somit durch eine Ableitung in der Zeit und dementsprechend auch exakt zu erhalten. Es gilt

$$\mathbf{w}(t, \mathbf{x}) = \left. \frac{\partial \mathbf{x}}{\partial t} \right|_{\mathbf{Y}} (t, \mathcal{A}_t^{-1}(\mathbf{x})).$$

5.3.2 Approximative Gittergeschwindigkeit

Wie auch bei der ALE-Abbildung kann in der Praxis die Gittergeschwindigkeit nur an bestimmten Zeitpunkten berechnet werden, daher wird die Gittergeschwindigkeit beispielsweise durch ein Interpolationspolynom approximiert. In Bezug auf die dG-Verfahren definieren wir das Polynom $\tilde{\mathcal{A}}_t|_{I_n} \in \mathbb{P}_k$ durch

$$\tilde{\mathcal{A}}_t|_{I_n} := \sum_{i=1}^{k+1} \phi_{n,i}(t) \mathcal{A}_{t_{n,i}}, \quad (5.12)$$

wobei $t_{n,i}$ die Gauß–Radau-Punkte sind. $\phi_{n,i}$ sind hierbei die Lagrange-Basis-Polynome mit $\phi_{n,i}(t_{n,j}) = \delta_{i,j}$, $i, j = 1, \dots, k+1$.

Die Gittergeschwindigkeit ergibt sich durch

$$\mathbf{w}|_{I_n} = \partial_t \tilde{\mathcal{A}}_t|_{I_n} \in \mathbb{P}_{k-1}.$$

Beachtenswert ist, dass die Beziehung

$$\tilde{\mathcal{A}}_{t_n} = \tilde{\mathcal{A}}_{t_{n-1}} + \int_{t_{n-1}}^{t_n} \mathbf{w}(t) dt$$

nicht notwendig erfüllt wird. Dies ist eine Folgerung der möglichen Unstetigkeit von $\tilde{\mathcal{A}}_t$ an t_{n-1} .

5.3.3 Approximative Gittergeschwindigkeit mit Korrektur

Eine bessere Variante erhalten wir, indem wir den Sprung am Zeitpunkt t_{n-1} durch eine Stetigkeitsbedingung korrigieren. Nun definieren wir $\tilde{\mathcal{A}}_t|_{I_n} \in \mathbb{P}_{k+1}$ durch

$$\tilde{\mathcal{A}}_t|_{I_n} := \sum_{i=0}^{k+1} \phi_{n,i}(t) \mathcal{A}_{t_{n,i}}, \quad (5.13)$$

wobei $t_{n,i}$ die Gauß–Radau-Punkte sind und $t_{n,0} = t_{n-1}$ ist. $\phi_{n,i}$ sind die Lagrange-Basis-Polynome mit $\phi_{n,i}(t_{n,j}) = \delta_{i,j}$, $i, j = 0, \dots, k+1$.

Somit erhalten wir die Gittergeschwindigkeit durch

$$\mathbf{w}|_{I_n} = \partial_t \tilde{\mathcal{A}}_t|_{I_n} \in \mathbb{P}_k.$$

Es gilt die Eigenschaft von $\tilde{\mathcal{A}}_t$

$$\tilde{\mathcal{A}}_{t_n} = \tilde{\mathcal{A}}_{t_{n-1}} + \int_{t_{n-1}}^{t_n} \mathbf{w}(t) dt.$$

5.3.4 Approximatives Gebiet

Die Einführung einer approximativen ALE-Abbildung $\tilde{\mathcal{A}}_t$ führt unmittelbar zu einem approximativen Gebiet $\tilde{\Omega}_t$, welches das Gebiet Ω_t annähert. Da die approximative ALE-Abbildung $\tilde{\mathcal{A}}_t$ die ALE-Abbildung \mathcal{A}_t nur an den Gauß–Radau-Stützstellen abgreift, stimmen dementsprechend die Gebiete $\tilde{\Omega}_t$ und Ω_t nur an diesen Stützstellen überein. Das approximative Gebiet $\tilde{\Omega}_t$ lässt sich auf jedem Intervall I_n folgendermaßen definieren:

$$\tilde{\Omega}_t := \left\{ \mathbf{x} \in \mathbb{R}^d : \mathbf{x} = \sum_{i=1}^{k+1} \hat{\phi}_i(T_n^{-1}(t)) \tilde{\mathcal{A}}_{t_{n,i}}(\mathbf{Y}), \quad \mathbf{Y} \in \hat{\Omega} \right\}, \quad (5.14)$$

wobei $\hat{\phi}_i$, $i = 1, \dots, k+1$, die Lagrange-Basis-Polynome zu den Gauß–Radau-Punkten auf \hat{I} sind.

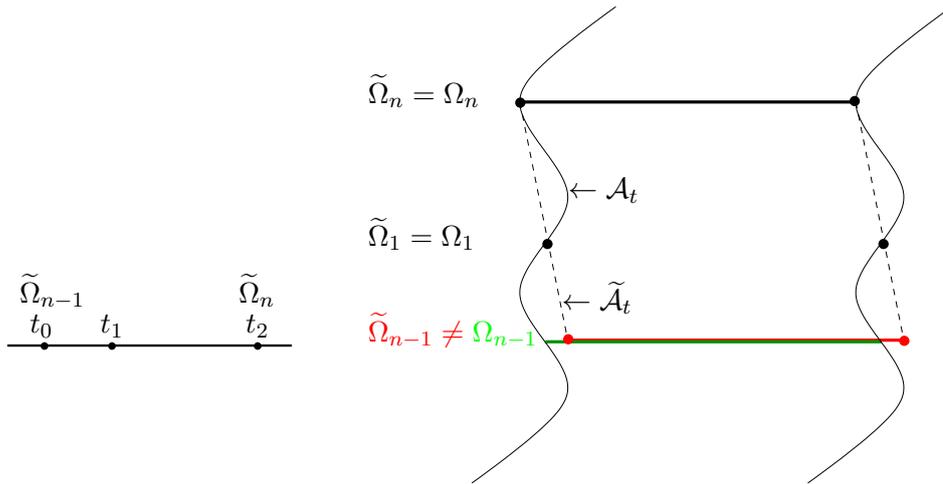


Abbildung 5.1: Zeitdiskretisierung des Gebietes Ω_t auf I_n mithilfe von $dG(1)$.

Um den Unterschied zwischen dem Gebiet Ω_t und dem approximativen Gebiet $\tilde{\Omega}_t$ zu illustrieren, betrachten wir den $dG(1)$ -Fall mit den Stützstellen $t_{n,1} = (t_{n-1} + t_n)/3$ und $t_{n,2} = t_n$ aus der

Gauß–Radau-Formel in der Abbildung 5.1. Aus dieser kann man entnehmen, dass sich die beiden Gebiete auch am Anfangspunkt jedes Teilintervalls I_n unterscheiden können, da dieser keine Interpolationsstützstelle darstellt.

Die korrigierte Approximation der ALE-Abbildung in (5.13) garantiert die Übereinstimmung der approximativen Gebiete $\tilde{\Omega}_{n-1}^-$ und $\tilde{\Omega}_{n-1}^+$, die die Definition der Sprungterme sinnvoll macht. Wir werden in Kapitel 6 durch numerische Beispiele sehen, welchen Einfluss diese Korrektur der ALE-Abbildung auf die Lösung hat. Im weiteren Verlauf dieses Kapitels ist $\tilde{\Omega}_t$ das in (5.14) definierte approximative Gebiet. Als ALE-Abbildung verwenden wir die in (5.13) definierte Approximation.

5.4 Unstetige Galerkin-Verfahren

Wir betrachten die Zeitdiskretisierung wie in Kapitel 2, jedoch im ALE-Rahmen, um das Problem (5.1) in der Zeit zu diskretisieren. Hier werden wir den approximativen Zeit-Raum-Zylinder

$$\tilde{\mathcal{S}} := \left\{ (t, \mathbf{x}) \in \mathbb{R}^{d+1} : t \in I, \mathbf{x} \in \tilde{\Omega}_t \right\} \quad (5.15)$$

und den in der Zeit diskreten Raum

$$\mathcal{X}_k := \left\{ V \in \tilde{\mathcal{X}} : V(\cdot, \mathbf{x})|_{I_n} = \hat{V}(\cdot, \tilde{\mathcal{A}}_t^{-1}(\mathbf{x}))|_{I_n} \in \mathbb{P}_k(I_n) \quad \forall n = 1, \dots, N \right\} \quad (5.16)$$

benötigen, wobei

$$\tilde{\mathcal{X}} := \left\{ v : \tilde{\mathcal{S}} \rightarrow \mathbb{R} : v = \hat{v} \circ \tilde{\mathcal{A}}_t^{-1}, \quad \hat{v} \in \mathcal{Y}(\hat{\Omega}), \quad t \in I \right\}$$

der Raum der Testfunktionen auf $\tilde{\mathcal{S}}$ ist. Nach dieser Definition brauchen die Funktionen aus \mathcal{X}_k von links an den Zeitpunkten $t_n, n = 1, \dots, N$, nicht stetig zu sein. Sei nun

$$\mathcal{S}_n := \left\{ (t, \mathbf{x}) \in \tilde{\mathcal{S}} : t \in I_n \right\},$$

wobei $\tilde{\mathcal{S}}$ in (5.15) definiert ist. Da wir eine Zeitdiskretisierung anstreben, definieren wir den Raum

$$\mathcal{X}_k(I_n) := \left\{ V : \mathcal{S}_n \rightarrow \mathbb{R} : V = W|_{\mathcal{S}_n}, W \in \mathcal{X}_k \right\}, \quad n = 1, \dots, N,$$

als Einschränkung des Raumes \mathcal{X}_k auf dem Intervall I_n . Man beachte, dass der Unterschied zwischen den Gebieten Ω_t und $\tilde{\Omega}_t$ zu einer numerischen Lösung U führt, die nicht die Lösung u_{ex} in Ω_t , sondern die Lösung u in $\tilde{\Omega}_t$ approximiert. Die exakte Lösung ergibt sich dann aus

$$u_{ex}(t, \mathbf{x}) := u(t, \tilde{\mathcal{A}}_t \circ \mathcal{A}_t^{-1}(\mathbf{x})) \quad \forall \mathbf{x} \in \Omega_t \quad \forall t \in I_n.$$

Für die gesuchte Funktion $u(t, \mathbf{x})$ setzt man als Ansatzfunktion auf dem Intervall I_n die zeitlich diskrete, aber im Ort kontinuierliche Näherung $U(\cdot, \mathbf{x}) : I_n \rightarrow \mathcal{X}_k(I_n)$ als eine Linearkombination von Basis-Funktionen aus $\mathcal{X}_k(I_n)$ an, d.h.

$$U(t, \mathbf{x}) := \sum_{j=1}^{k+1} U_n^j(\tilde{\mathcal{A}}_t(\mathbf{Y})) \phi_{n,j}(t, \tilde{\mathcal{A}}_t(\mathbf{Y})), \quad \tilde{\mathcal{A}}_t(\mathbf{Y}) = \mathbf{x}(t, \mathbf{Y}), \mathbf{x} \in \tilde{\Omega}_t. \quad (5.17)$$

Die Koeffizienten $U_n^j \in H_0^1(\Omega_{t_{n,j}})$ sind ortsabhängig und die Basis-Funktionen $\phi_{n,j}$ sind definiert durch

$$\phi_{n,j}(t, \mathbf{x}) = \hat{\phi}_{n,j}(t, \tilde{\mathcal{A}}_t^{-1}(\mathbf{x})), \quad t \in I_n, \mathbf{x} \in \tilde{\Omega}_t,$$

wobei $\hat{\phi}_{n,j} \in \mathbb{P}_k(I_n)$ die Lagrange-Basis-Funktionen bezüglich der $(k+1)$ Knoten der in Kapitel 2 betrachteten rechtsseitigen Gauß–Radau-Formel auf I_n sind. Die approximative Lösung auf Ω_t ergibt sich dann aus

$$U_{ex}(t, \mathbf{x}) := U(\tilde{\mathcal{A}}_t \circ \mathcal{A}_t^{-1}(\mathbf{x})) \quad \forall \mathbf{x} \in \Omega_t.$$

Der Einfachheit halber verwenden wir im Folgenden die Notation $\phi_{n,j}(t)$ anstatt $\phi_{n,j}(t, \mathbf{x})$.

5.4.1 Schwache Formulierung im ALE-Rahmen

Wir definieren die Bilinearform für die nichtkonservative Formulierung durch

$$a_{nc}(s; \varphi, \psi) := ((\mathbf{b}(s) - \mathbf{w}(s)) \cdot \nabla_{\mathbf{x}} \varphi, \psi)_{\tilde{\Omega}_s} + \varepsilon (\nabla_{\mathbf{x}} \varphi, \nabla_{\mathbf{x}} \psi)_{\tilde{\Omega}_s} + (c\varphi, \psi)_{\tilde{\Omega}_s}, \quad s \in I,$$

wohingegen die Bilinearform für die konservative Formulierung durch

$$\begin{aligned} a_c(s; \varphi, \psi) &:= ((\mathbf{b}(s) - \mathbf{w}(s)) \cdot \nabla_{\mathbf{x}} \varphi, \psi)_{\tilde{\Omega}_s} - ((\nabla_{\mathbf{x}} \cdot \mathbf{w}(s)) \varphi, \psi)_{\tilde{\Omega}_s} + \varepsilon (\nabla_{\mathbf{x}} \varphi, \nabla_{\mathbf{x}} \psi)_{\tilde{\Omega}_s} \\ &\quad + (c\varphi, \psi)_{\tilde{\Omega}_s} - (\varphi, D_t \psi)_{\tilde{\Omega}_s} \\ &= a_{nc}(s; \varphi, \psi) - ((\nabla_{\mathbf{x}} \cdot \mathbf{w}(s)) \varphi, \psi)_{\tilde{\Omega}_s} - (\varphi, D_t \psi)_{\tilde{\Omega}_s}, \quad s \in I, \end{aligned}$$

definiert wird. In Bezug auf den nichtkonservativen Fall lässt sich das globale dG-Problem folgendermaßen formulieren:

Finde $U \in \mathcal{X}_k$ mit $U(0^-) = u_0$ und

$$\sum_{n=1}^N \int_{I_n} (D_t U(t), V(t))_{\tilde{\Omega}_t} dt + \sum_{n=1}^N ([U]_{n-1}, V_{n-1}^+)_{\tilde{\Omega}_{n-1}} + \int_I a_{nc}(t; U, V) dt = \int_I (f(t), V(t))_{\tilde{\Omega}_t} dt \quad (5.18)$$

für alle $V \in \mathcal{X}_k$. Dabei sind der Sprung $[\varphi]_{n-1}$ sowie der rechtsseitige Wert φ_{n-1}^+ in (3.3) definiert. Während das globale dG-Problem bezüglich der konservativen Variante folgendermaßen lautet:

Finde $U \in \mathcal{X}_k$ mit $U(0^-) = u_0$ und

$$\sum_{n=1}^N \int_{I_n} \partial_t (U(t), V(t))_{\tilde{\Omega}_t} dt + \sum_{n=1}^N ([U]_{n-1}, V_{n-1}^+)_{\tilde{\Omega}_{n-1}} + \int_I a_c(t; U, V) dt = \int_I (f(t), V(t))_{\tilde{\Omega}_t} dt \quad (5.19)$$

für alle $V \in \mathcal{X}_k$. Wie man in Kapitel 3 sehen konnte, lässt sich das globale dG-Problem zeitlich aufspalten, indem man eine Testfunktion aus $\mathcal{X}_k(I_n)$ betrachtet und sie außerhalb \bar{I}_n durch 0 fortsetzt. Anhand der bekannten Werte $U_n^0 := U_{n-1}^-$ lässt sich die Lösung auf jedem Teilintervall I_n nacheinander bestimmen. Das lokale nichtkonservative dG(k)-Problem lautet:

Finde $U \in \mathcal{X}_k(I_n)$ mit $U_n^0 := U_{n-1}^-$ und

$$\int_{I_n} (D_t U(t), V(t))_{\tilde{\Omega}_t} dt + ([U]_{n-1}, V_{n-1}^+)_{\tilde{\Omega}_{n-1}} + \int_{I_n} a_{nc}(t; U, V) dt = \int_{I_n} (f(t), V(t))_{\tilde{\Omega}_t} dt \quad (5.20)$$

für alle $V \in \mathcal{X}_k(I_n)$. Das lokale konservative dG(k)-Problem lautet:

Finde $U \in \mathcal{X}_k(I_n)$ mit $U_n^0 := U_{n-1}^-$ und

$$\int_{I_n} \partial_t (U(t), V(t))_{\tilde{\Omega}_t} dt + ([U]_{n-1}, V_{n-1}^+)_{\tilde{\Omega}_{n-1}} + \int_{I_n} a_c(t; U, V) dt = \int_{I_n} (f(t), V(t))_{\tilde{\Omega}_t} dt \quad (5.21)$$

für alle $V \in \mathcal{X}_k(I_n)$. Die konservative und die nichtkonservative Formulierung sind vollkommen äquivalent. Dies garantiert die Reynolds-Identität (2.6).

Wir definieren die Energienorm anhand von

$$\|v\|_{\varepsilon, \tilde{\Omega}_t} := \left\{ \varepsilon |v|_{1, \tilde{\Omega}_t}^2 + c_0 \|v\|_{L^2(\tilde{\Omega}_t)}^2 \right\}^{1/2}.$$

Die Bilinearform a_{nc} ist bezüglich der Energienorm koerziv. Unter Ausnutzung der Eigenschaft (5.2) kann man in einfachen Schritten zeigen, dass a_{nc} die Abschätzung

$$a_{nc}(t; v, v) \geq \gamma \|v\|_{\varepsilon, \tilde{\Omega}_t}^2 \quad (5.22)$$

mit $\gamma > 0$ erfüllt.

Lemma 5.2. *[Existenz und Eindeutigkeit] Für alle $n = 1, \dots, N$ existiert eine eindeutige Lösung $U \in \mathcal{X}_k(I_n)$ von (5.20) bzw. (5.21), die die Anfangsbedingung $U(0^-) = u_0$ erfüllt.*

Beweis. Zunächst ist der Raum $\mathcal{X}_k(I_n), n = 1, \dots, N$, ausgestattet mit der Norm $\|\cdot\|_{\varepsilon, I_n} = \left\{ \int_{I_n} \|\cdot\|_{\varepsilon, \Omega_t}^2 dt \right\}^{1/2}$ ein Hilbert-Raum. Anhand der Äquivalenz von (5.20) und (5.21) werden wir die Behauptung nur für ersteres zeigen. Dabei betrachten wir die in diesem Fall definierte Bilinearform.

Desweiteren führen wir die folgende Bilinearform ein:

$$b^{dG}(\varphi, \psi) := \int_{I_n} a_{nc}(t; \varphi, \psi) + (D_t \varphi(t), \psi(t))_{\tilde{\Omega}_t} dt + ([\varphi]_n, \psi_n^+)_{\tilde{\Omega}_n}.$$

Die Linearform l^{dG} sei gegeben durch

$$l^{dG}(\psi) := \int_{I_n} (f(t), \psi(t))_{\tilde{\Omega}_t} dt.$$

Infolgedessen ist die diskrete nichtkonservative Form (5.20) äquivalent zu:

Finde $U \in \mathcal{X}_k(I_n)$ mit $U(0^-) = u_0$ und

$$b^{dG}(U, V) := l^{dG}(V) \quad \forall V \in \mathcal{X}_k(I_n).$$

Aufgrund von Normäquivalenz auf dem endlichdimensionalen Raum $\mathcal{X}_k(I_n)$ der Polynome vom Grad $\leq k$ folgen die Koerzivität sowie die Beschränktheit der Bilinearform b^{dG} aus der Koerzivität und der Beschränktheit der Bilinearform a_{nc} . Dies zusammen mit der Stetigkeit von $\int_{I_n} (f(t), V(t))_{\tilde{\Omega}_t} dt$ liefert nach dem Satz von Lax-Milgram die Behauptung. \square

Bevor wir uns mit der Stabilität der dG-Verfahren beschäftigen, erläutern wir zunächst die folgende Identität, die wir im Laufe des Kapitels benötigen werden:

Lemma 5.3. *[Diskrete Reynolds-Identität] Sei $V \in \mathcal{X}_k(I_n), n = 1, \dots, N$. Es gilt*

$$\int_{I_n} \left((D_t V(t), V(t))_{\tilde{\Omega}_t} - (\mathbf{w}(t) \cdot \nabla_{\mathbf{x}} V(t), V(t))_{\tilde{\Omega}_t} \right) dt = \frac{1}{2} \|V(t_n^-)\|_{L^2(\tilde{\Omega}_n)}^2 - \frac{1}{2} \|V(t_{n-1}^+)\|_{L^2(\tilde{\Omega}_{n-1})}^2. \quad (5.23)$$

Beweis. Aus (2.2) folgt

$$\begin{aligned} \int_{I_n} \left((D_t V(t), V(t))_{\tilde{\Omega}_t} - (\mathbf{w}(t) \cdot \nabla_{\mathbf{x}} V(t), V(t))_{\tilde{\Omega}_t} \right) dt &= \int_{I_n} (\partial_t V(t), V(t))_{\tilde{\Omega}_t} dt \\ &= \int_{I_n} \frac{1}{2} \partial_t \|V(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ &= \frac{1}{2} \|V(t_n^-)\|_{L^2(\tilde{\Omega}_n)}^2 - \frac{1}{2} \|V(t_{n-1}^+)\|_{L^2(\tilde{\Omega}_{n-1})}^2. \end{aligned}$$

\square

5.4.2 Stabilität

Dieser Unterabschnitt befasst sich mit der Untersuchung der numerischen Stabilität der variationellen Zeitdiskretisierung im Falle einer exakten Integration. Im Gegensatz zu den anderen Untersuchungen, wie beispielsweise von Nobile in [32], Bonito und Co. in [7], enthält unser Problem (5.1) den zusätzlichen Reaktionsterm cu , den wir ausnutzen, um eine bessere Abschätzung für die Stabilität erhalten zu können. Als Zugang zur Stabilität erläutern wir den folgenden Satz:

Satz 5.4. *Sei $V \in \mathcal{X}_k(I_n)$. Es gilt für alle $0 \leq l < m \leq N$:*

$$\begin{aligned}
& \frac{1}{2} \|V(t_m^-)\|_{L^2(\tilde{\Omega}_m)}^2 - \frac{1}{2} \|V(t_l^-)\|_{L^2(\tilde{\Omega}_l)}^2 + \frac{1}{2} \sum_{j=l}^m \| [V]_{j-1} \|_{L^2(\tilde{\Omega}_j)}^2 + \varepsilon \int_{t_l}^{t_m} |V(t)|_{1, \tilde{\Omega}_t}^2 dt \\
& \quad + c_0 \int_{t_l}^{t_m} \|V(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\
& \leq \int_{t_l}^{t_m} (D_t V(t), V(t))_{\tilde{\Omega}_t} dt + \sum_{j=l}^m ([V]_{j-1}, V_{j-1}^+)_{\tilde{\Omega}_{j-1}} + \int_{t_l}^{t_m} ((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} V(t), V(t))_{\tilde{\Omega}_t} dt \\
& \quad + \varepsilon \int_{t_l}^{t_m} (\nabla_{\mathbf{x}} V(t), \nabla_{\mathbf{x}} V(t))_{\tilde{\Omega}_t} dt + \int_{t_l}^{t_m} (c(t)V(t), V(t))_{\tilde{\Omega}_t} dt. \tag{5.24}
\end{aligned}$$

Beweis. Die Betrachtung und das Aufsummieren von l bis m der Identitäten

$$((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} V, V)_{\tilde{\Omega}_t} = -\frac{1}{2} (\nabla_{\mathbf{x}} \cdot (\mathbf{b} - \mathbf{w}), V^2)_{\tilde{\Omega}_t}, \tag{5.25}$$

$$\frac{1}{2} \int_{I_j} ((\nabla_{\mathbf{x}} \cdot \mathbf{w}(t))V(t), V(t))_{\tilde{\Omega}_t} dt = - \int_{I_j} (\mathbf{w}(t) \cdot \nabla_{\mathbf{x}} V(t), V(t))_{\tilde{\Omega}_t} dt, \tag{5.26}$$

$$\begin{aligned}
& \int_{I_j} (D_t V(t), V(t))_{\tilde{\Omega}_t} dt + \frac{1}{2} \int_{I_j} ((\nabla_{\mathbf{x}} \cdot \mathbf{w}(t))V(t), V(t))_{\tilde{\Omega}_t} dt \\
& \quad = \frac{1}{2} \int_{I_j} \partial_t \|V(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt = \frac{1}{2} \|V(t_j^-)\|_{L^2(\tilde{\Omega}_j)}^2 - \frac{1}{2} \|V(t_{j-1}^+)\|_{L^2(\tilde{\Omega}_{j-1})}^2
\end{aligned} \tag{5.27}$$

und

$$\begin{aligned}
([V]_{j-1}, V_{j-1}^+)_{\tilde{\Omega}_{j-1}} &= (V_{j-1}^+ - V_{j-1}^-, V_{j-1}^+)_{\tilde{\Omega}_{j-1}} \\
&= \frac{1}{2} \|V(t_{j-1}^+)\|_{L^2(\tilde{\Omega}_{j-1})}^2 - \frac{1}{2} \|V(t_{j-1}^-)\|_{L^2(\tilde{\Omega}_{j-1})}^2 + \frac{1}{2} \| [V]_{j-1} \|_{L^2(\tilde{\Omega}_{j-1})}^2
\end{aligned} \tag{5.28}$$

führt zu

$$\begin{aligned}
& \int_{t_l}^{t_m} (D_t V(t), V(t))_{\tilde{\Omega}_t} dt + \sum_{j=l}^m ([V]_{j-1}, V_{j-1}^+)_{\tilde{\Omega}_{j-1}} + \int_{t_l}^{t_m} ((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} V(t), V(t))_{\tilde{\Omega}_t} dt \\
& \quad + \varepsilon \int_{t_l}^{t_m} (\nabla_{\mathbf{x}} V(t), \nabla_{\mathbf{x}} V(t))_{\tilde{\Omega}_t} dt + \int_{t_l}^{t_m} (c(t)V(t), V(t))_{\tilde{\Omega}_t} dt \\
& = \frac{1}{2} \|V(t_m^-)\|_{L^2(\tilde{\Omega}_m)}^2 - \frac{1}{2} \|V(t_l^-)\|_{L^2(\tilde{\Omega}_l)}^2 + \frac{1}{2} \sum_{j=l}^m \| [V]_{j-1} \|_{L^2(\tilde{\Omega}_{j-1})}^2 + \varepsilon \int_{t_l}^{t_m} |V(t)|_{1, \tilde{\Omega}_t}^2 dt \\
& \quad + \int_{t_l}^{t_m} \left((c - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b})(t), V^2(t) \right)_{\tilde{\Omega}_t} dt.
\end{aligned}$$

Dies und das Einsetzen der Eigenschaft (5.2) liefern die Behauptung. \square

Da die Stabilität bei der Erlangung der Konvergenz bei dG-Verfahren eine wichtige Rolle spielt, präsentieren wir als Folgerung des Satzes 5.4 das folgende Ergebnis:

Satz 5.5. *[Stabilität bei exakter Integration] Sei U die Lösung von (5.20) bzw. von (5.21). Es gilt für $1 \leq l < m \leq N$:*

$$\begin{aligned} \|U(t_m^-)\|_{L^2(\tilde{\Omega}_m)}^2 + \sum_{j=l}^{m-1} \|[U]_j\|_{L^2(\tilde{\Omega}_j)}^2 + 2\varepsilon \int_{t_l}^{t_m} |U(t)|_{1,\tilde{\Omega}_t}^2 dt + c_0 \int_{t_l}^{t_m} \|U(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ \leq \|U(t_l^-)\|_{L^2(\tilde{\Omega}_l)}^2 + \frac{1}{c_0} \int_{t_l}^{t_m} \|f(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt. \end{aligned} \quad (5.29)$$

Beweis. Aufgrund der Äquivalenz der konservativen und der nichtkonservativen Form reicht es aus, wenn wir die Behauptung für eine der beiden Formulierungen zeigen. Zunächst betrachten wir $V = U$ in (5.20) und erhalten

$$\begin{aligned} \int_{I_n} (D_t U(t), U(t))_{\tilde{\Omega}_t} dt + ([U]_{n-1}, U_{n-1}^+)_{\tilde{\Omega}_{n-1}} + \int_{I_n} \left((\mathbf{b} - \mathbf{w})(t) \cdot \nabla_{\mathbf{x}} U(t), U(t) \right)_{\tilde{\Omega}_t} dt \\ + \varepsilon \int_{I_n} |U(t)|_{1,\tilde{\Omega}_t}^2 dt + \int_{I_n} (c(t)U(t), V(t))_{\tilde{\Omega}_t} dt = \int_{I_n} (f(t), U(t))_{\tilde{\Omega}_t} dt. \end{aligned}$$

Aus den Cauchy-Schwarz- und Young-Ungleichungen folgt weiterhin, dass

$$\int_{I_n} (h(t), V(t))_{\tilde{\Omega}_t} dt \leq \frac{c_0}{2} \int_{I_n} \|V(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt + \frac{1}{2c_0} \int_{I_n} \|h(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt.$$

Die Behauptung folgt aus Satz 5.4. □

Im Unterschied zu der ähnlichen Stabilitätsabschätzung in [7] enthält diese Abschätzung anstatt des Terms $\frac{1}{\varepsilon}$, der für sehr kleine ε die Abschätzung zu grob werden lässt, nur die Konstante $\frac{1}{c_0}$. In den folgenden Ausführungen wollen wir eine Fehlerabschätzung herleiten, für die wir die folgende ALE-Projektion benötigen werden:

Definition 5.6. *[ALE-Projektion] Wir definieren den ALE-Projektionsoperator $P : C(H_0^1(\mathcal{S})) \rightarrow \mathcal{X}_k$ wie folgt: Für alle $n = 1, \dots, N$ gilt*

$$Pu(t_n, \cdot) = u(t_n, \cdot) \text{ in } \tilde{\Omega}_n \quad (5.30)$$

und

$$\int_{I_n} (Pu(t) - u(t), v(t))_{\tilde{\Omega}_t} dt = 0 \quad \forall v \in \mathbb{P}_{k-1}(I_n). \quad (5.31)$$

Nun nehmen wir an, dass die ALE-Abbildung die zusätzliche Regularitätsbedingung

$$\mathcal{A}_{t_n \rightarrow t} \in L^\infty(I_n, W^{2,\infty}(\tilde{\Omega}_{t_n})), \quad n = 1, \dots, N, \quad (5.32)$$

erfüllt. Dies garantiert einerseits $Pu(t) \in H_0^1(\tilde{\Omega}_t)$ und andererseits die Existenz sowie die Eindeutigkeit der Funktion $Pu \in \mathcal{X}_k$, sodass (5.30) und (5.31) für alle $u \in C(I_n, H_0^1(\mathcal{S}))$ gelten. Bezüglich der Beweise der Existenz sowie der Wohldefiniertheit der Projektion P verweisen wir auf die Arbeit von Bonito, Kyza und Nocketto, vgl. [6].

5.4.3 A-priori Fehleranalyse

In diesem Abschnitt befassen wir uns mit der Untersuchung des Fehlers bei einer exakten Integration. Wir spalten den Fehler in einen ALE-Projektionsfehler $\boldsymbol{\delta} := u - Pu$ und in einen diskreten Fehleranteil $\boldsymbol{\theta} := U - Pu$ auf. Es gilt

$$U - u = \boldsymbol{\theta} - \boldsymbol{\delta}. \quad (5.33)$$

Als nächstes betrachten wir die folgende Eigenschaft des ALE-Projektionsfehlers, die für die Untersuchung des Fehlers besonders relevant ist.

Lemma 5.7. *Seien u und Pu jeweils die Lösung von (5.9) und ihre Projektion in den Raum \mathcal{X}_k . Der Fehler $\boldsymbol{\delta} = u - Pu$ erfüllt für alle $1 \leq n \leq N$*

$$\int_{I_n} (D_t \boldsymbol{\delta}(t), V(t))_{\tilde{\Omega}_t} dt + ([\boldsymbol{\delta}]_{n-1}, V(t_{n-1}^+))_{\tilde{\Omega}_{n-1}} + \int_{I_n} (\boldsymbol{\delta}(t) \nabla_{\mathbf{x}} \cdot \mathbf{w}(t), V(t))_{\tilde{\Omega}_t} dt = 0 \quad (5.34)$$

für alle $V \in \mathcal{X}_k(I_n)$.

Beweis. Aus der Reynolds-Identität (2.6) folgt

$$(D_t \boldsymbol{\delta}, V)_{\tilde{\Omega}_t} = -(\boldsymbol{\delta}, D_t V)_{\tilde{\Omega}_t} - (\boldsymbol{\delta}, V \nabla_{\mathbf{x}} \cdot \mathbf{w})_{\tilde{\Omega}_t} + \partial_t (\boldsymbol{\delta}, V)_{\tilde{\Omega}_t}.$$

Die Integration über I_n liefert

$$\begin{aligned} \int_{I_n} (D_t \boldsymbol{\delta}(t), V(t))_{\tilde{\Omega}_t} dt &= - \int_{I_n} (\boldsymbol{\delta}(t), D_t V(t))_{\tilde{\Omega}_t} dt - \int_{I_n} (\boldsymbol{\delta}(t), V(t) \nabla_{\mathbf{x}} \cdot \mathbf{w})_{\tilde{\Omega}_t} dt \\ &\quad + (\boldsymbol{\delta}(t_n^-), V(t_n^-))_{\tilde{\Omega}_n} - (\boldsymbol{\delta}(t_{n-1}^+), V(t_{n-1}^+))_{\tilde{\Omega}_{n-1}}. \end{aligned} \quad (5.35)$$

Aus (5.30) folgt $(\boldsymbol{\delta}(t_n^-), V(t_n^-))_{\tilde{\Omega}_n} = 0 = (\boldsymbol{\delta}(t_{n-1}^+), V(t_{n-1}^+))_{\tilde{\Omega}_{n-1}}$. Dies und (5.31) in (5.35) eingesetzt liefern die Behauptung. \square

Anschließend zitieren wir aus [6] für die ALE-Projektion P die folgenden Abschätzungen, die wir aufgrund der unterschiedlichen Definitionen von I_n und A_n in (5.11) in unserem Fall angepasst haben.

Lemma 5.8. *[ALE-Projektionsfehler] Für alle $j = 1, \dots, k$ gilt*

$$\begin{aligned} \sup_{t \in I_n} \|(u - Pu)(t)\|_{L^2(\tilde{\Omega}_t)}^2 &\leq C_n \tau_n^{2j+1} \int_{I_n} \|D_t^{j+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt, \\ \sup_{t \in I_n} \|\nabla_{\mathbf{x}}(u - Pu)(t)\|_{L^2(\tilde{\Omega}_t)}^2 &\leq D_n \tau_n^{2j+1} \int_{I_n} \left(\|D_t^{j+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + \|\nabla_{\mathbf{x}} D_t^{j+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 \right) dt, \end{aligned}$$

wobei C_n und D_n Konstanten sind, die von A_n abhängen. Dabei ist A_n in (5.11) definiert.

Beweis. Mit demselben Vorgehen wie in [6] führen wir zunächst die dG-Projektion aus der Arbeit von Akrivis und Makridakis ein, vgl. [3, 26]. Für $1 \leq n \leq N$ betrachten wir $\tilde{\Omega}_n$ als Referenzgebiet und die dG-Projektion von \hat{u} bezeichnen wir mit $\hat{\psi}_u \in \mathcal{Y}_k(I_n)$, wobei

$$\mathcal{Y}_k(I_n) := \left\{ \hat{v} : \hat{v} = v \circ \tilde{\mathcal{A}}_t, \quad v \in \mathcal{X}_k(I_n) \right\}$$

und definieren diese Projektion durch

$$\hat{\psi}_u(\mathbf{Y}, t_n) := \hat{u}(\mathbf{Y}, t_n), \quad \mathbf{Y} \in \Omega_{t_n}, \quad \int_{I_n} (\hat{u} - \hat{\psi}_u, \hat{V})_{\tilde{\Omega}_n} dt = 0 \quad \forall \hat{V} \in \mathcal{Y}_{k-1}(I_n).$$

Die letzte Bedingung gilt nur für $k \geq 1$. Die dG-Projektion $\widehat{\psi}_u$ ist wohldefiniert und erfüllt für alle $j = 1, \dots, k$

$$\sup_{t \in I_n} \|(\hat{u} - \widehat{\psi}_u)(t)\|_{L^2(\widetilde{\Omega}_n)}^2 \leq \widehat{C}_1 \tau_n^{2j+1} \int_{I_n} \|\partial_t^{j+1} \hat{u}(s)\|_{L^2(\widetilde{\Omega}_n)}^2 ds, \quad (5.37)$$

$$\sup_{t \in I_n} \|\nabla \mathbf{Y}(\hat{u} - \widehat{\psi}_u)(t)\|_{L^2(\widetilde{\Omega}_n)}^2 \leq \widehat{C}_2 \tau_n^{2j+1} \int_{I_n} \|\partial_t^{j+1} \hat{u}(s)\|_{L^2(\widetilde{\Omega}_n)}^2 ds, \quad (5.38)$$

vgl.[6, Prop. 3.2]. Es ist offensichtlich, dass

$$\psi_u = \widehat{\psi}_u \circ \widetilde{\mathcal{A}}_t^{-1} \in \mathcal{X}_k(I_n). \quad (5.39)$$

Aus (5.30) folgt weiterhin

$$\psi_u(t_n, \cdot) = Pu(t_n, \cdot). \quad (5.40)$$

Aufbauend auf den Eigenschaften (5.39) und (5.40) von ψ_u können wir $\psi_u - Pu = (t_n - t)W$ mit W aus $\mathcal{X}_{k-1}(I_n)$ setzen. Aus (5.31) folgt

$$\int_{I_n} (\psi_u(t) - Pu(t), V(t))_{\widetilde{\Omega}_t} dt = \int_{I_n} (\psi_u(t) - u(t), V(t))_{\widetilde{\Omega}_t} dt \quad \forall V \in \mathcal{X}_{k-1}(I_n).$$

Sei nun $V = W$. Es gilt einerseits

$$\begin{aligned} \int_{I_n} \|(\psi_u - Pu)(t)\|_{L^2(\widetilde{\Omega}_t)}^2 dt &= \int_{I_n} (\psi_u - Pu)(t), (t_n - t)W(t)_{\widetilde{\Omega}_t} dt \\ &\leq \tau_n \int_{I_n} (\psi_u - Pu)(t), W(t)_{\widetilde{\Omega}_t} dt \\ &= \tau_n \int_{I_n} |(\psi_u - u)(t), W(t)_{\widetilde{\Omega}_t}| dt \\ &\leq \tau_n \left(\int_{I_n} \|(\psi_u - u)(t)\|_{L^2(\widetilde{\Omega}_t)}^2 dt \right)^{1/2} \left(\int_{I_n} \|W(t)\|_{L^2(\widetilde{\Omega}_t)}^2 dt \right)^{1/2} \end{aligned} \quad (5.41)$$

und andererseits haben wir

$$\tau_n^2 \int_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)}^2 dt = \tau_n^2 \int_{I_n} \int_{\Omega_n} \widehat{V}^2(t) \det \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}} d\hat{\mathbf{x}} dt.$$

Aus der Normäquivalenz auf dem endlichdimensionalen Raum der Polynome vom Grad $\leq 2k - 2 + dq$, wobei q der Grad der ALE-Abbildung ist und d die Dimension des Raumes $\widetilde{\Omega}_t$ darstellt, folgt

$$\begin{aligned} \tau_n^2 \int_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)}^2 dt &\leq C \int_{I_n} (t_n - t)^2 \int_{\Omega_n} \widehat{V}^2(t) \det \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}} d\hat{\mathbf{x}} dt \\ &\leq C \int_{I_n} (t_n - t)^2 \|V(t)\|_{L^2(\widetilde{\Omega}_t)}^2 dt. \end{aligned}$$

Wir erhalten somit

$$\tau_n^2 \int_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)}^2 dt \leq C \int_{I_n} (t_n - t)^2 \|V(t)\|_{L^2(\widetilde{\Omega}_t)}^2 dt.$$

Dies ist äquivalent zu

$$\tau_n^2 \int_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)}^2 dt \leq C \int_{I_n} \|(\psi_u - Pu)(t)\|_{L^2(\widetilde{\Omega}_t)}^2 dt.$$

Daraufbauend folgt aus (5.41)

$$\int_{I_n} \|(\psi_u - Pu)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \leq C \left(\int_{I_n} \|(\psi_u - u)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \right)^{1/2} \left(\int_{I_n} \|(\psi_u - Pu)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \right)^{1/2}$$

bzw.

$$\int_{I_n} \|(\psi_u - Pu)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \leq C \int_{I_n} \|(\psi_u - u)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt.$$

Es gilt die Normäquivalenz

$$\begin{aligned} \|(\psi_u - Pu)(t)\|_{L^2(\tilde{\Omega}_t)}^2 &\leq \frac{C}{\tau_n} \int_{I_n} \|(\psi_u - Pu)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ &\leq \frac{C}{\tau_n} \int_{I_n} \|(\psi_u - u)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ &\leq \frac{CA_n}{\tau_n} \int_{I_n} \|(\hat{\psi}_u - \hat{u})(t)\|_{L^2(\tilde{\Omega}_n)}^2 dt. \end{aligned} \quad (5.42)$$

Dies kombiniert mit (5.37) unter Betrachtung von

$$\|u - Pu\|_{L^2(\tilde{\Omega}_t)}^2 \leq 2\|u - \psi_u\|_{L^2(\tilde{\Omega}_t)}^2 + 2\|\psi_u - Pu\|_{L^2(\tilde{\Omega}_t)}^2$$

liefert die erste Behauptung in Lemma 5.8.

Um die zweite Abschätzung zu zeigen, setzen wir $\hat{U}_n := \hat{P}u$, schreiben die Eigenschaft (5.31) auf $\tilde{\Omega}_n$ um und erhalten

$$\int_{I_n} ((\hat{u} - \hat{U}_n)(t) \det \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}}, \hat{V}(t))_{\tilde{\Omega}_n} dt = 0 \quad \forall \hat{V} \in \mathcal{Y}_{k-1}(I_n).$$

Dies ist äquivalent zu

$$\int_{\tilde{\Omega}_n} \int_{I_n} (\hat{u} - \hat{U}_n)(t) t^j \det \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}} dt d\mathbf{Y} = 0, \quad j = 0, \dots, k-1.$$

Aus der Differenzierung im Ort erhalten wir

$$\int_{\tilde{\Omega}_n} \int_{I_n} \nabla_{\mathbf{Y}}((\hat{u} - \hat{U}_n)(t) t^j \det \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}}) dt d\mathbf{Y} = 0, \quad j = 0, \dots, k-1.$$

Dies impliziert

$$\int_{I_n} (\det \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}} \nabla_{\mathbf{Y}}(\hat{u} - \hat{U}_n)(t) + (\hat{u} - \hat{U}_n)(t) \nabla_{\mathbf{Y}} \det \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}}, \nabla_{\mathbf{Y}} \hat{V}(t))_{\tilde{\Omega}_n} dt = 0 \quad \forall \hat{V} \in \mathcal{Y}_{k-1}(I_n),$$

jedoch gilt andererseits

$$\begin{aligned} \int_{I_n} (\det \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}} \nabla_{\mathbf{Y}}(\hat{\psi}_u - \hat{U}_n)(t), \nabla_{\mathbf{Y}} \hat{V}(t))_{\tilde{\Omega}_n} dt &= \int_{I_n} (\det \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}} \nabla_{\mathbf{Y}}(\hat{\psi}_u - \hat{u})(t), \nabla_{\mathbf{Y}} \hat{V}(t))_{\tilde{\Omega}_n} dt \\ &\quad + \int_{I_n} ((\hat{u} - \hat{U}_n)(t) \nabla_{\mathbf{Y}} \det \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}}, \nabla_{\mathbf{Y}} \hat{V}(t))_{\tilde{\Omega}_n} dt \end{aligned}$$

für alle $\hat{V} \in \mathcal{Y}_{k-1}(I_n)$.

Aus $\nabla_{\mathbf{Y}} \hat{v}(Y) = \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}}^T \nabla_{\mathbf{x}} v(\mathbf{x})$ für alle $v \in H_0^1(\Omega_t)$ und aus $\hat{\psi}_u - \hat{U}_n = (t_n - t) \widehat{W}$, vgl. (5.30), folgt weiterhin

$$\int_{I_n} \|\nabla_{\mathbf{x}}(\psi_u - U_n)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \leq \tau_n \int_{I_n} (\det \mathbf{J}_{\mathcal{A}_{t_n \rightarrow t}} (\nabla_{\mathbf{Y}} \mathcal{A}_{t_n \rightarrow t})^{-T} \nabla_{\mathbf{Y}}(\hat{\psi}_u - \hat{U}_n)(t), \nabla_{\mathbf{Y}} \widehat{W}(t))_{\tilde{\Omega}_n} dt.$$

Wir setzen nun $\widehat{V} = \widehat{W}$. Mit denselben Argumenten wie bereits im Beweis der ersten Aussage erhalten wir

$$\begin{aligned} \int_{I_n} \|\nabla_{\mathbf{x}}(\psi_u - U_n)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt &\leq CA_n^2 \int_{I_n} \|\nabla_{\mathbf{x}}(\psi_u - u)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ &\quad + CA_n^3 \int_{I_n} \|\nabla_{\mathbf{x}}(u - U_n)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt. \end{aligned}$$

Die in (5.42) stehende Normäquivalenz lässt sich auch auf den Ortsgradienten übertragen, somit erhalten wir

$$\|\nabla_{\mathbf{x}}(\psi_u - Pu)(t)\|_{L^2(\tilde{\Omega}_t)}^2 \leq C \frac{1}{\tau_n} \int_{I_n} \|\nabla_{\mathbf{x}}(\psi_u - Pu)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt.$$

Die Dreiecks- und Young-Ungleichung liefern die Behauptung. \square

Nun können wir eine Abschätzung optimaler Ordnung für den Fehler $U - u$ präsentieren. Es stellt sich heraus, dass keinerlei Einschränkung der Zeitschritte nötig ist, dementsprechend liegt eine unbedingte Fehlerabschätzung vor.

Satz 5.9. [*A-priori Fehlerabschätzung bei exakter Integration*] Seien u die Lösung von (5.1) und U die dG-Lösung von (5.20) bzw. von (5.21). Die ALE-Abbildung erfüllt (5.32). Es gilt für alle $j = 0, \dots, k$

$$\begin{aligned} &\max_{0 \leq n \leq N} \|(U - u)(t_n)\|_{L^2(\tilde{\Omega}_n)}^2 + \sum_{n=0}^{N-1} \|[U - u]_n\|_{L^2(\tilde{\Omega}_n)}^2 + \frac{\varepsilon}{2} \int_I |(U - u)(t)|_{1, \tilde{\Omega}_t}^2 dt \\ &\quad + c_0 \int_I \|(U - u)(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ &\leq \sum_{n=1}^N \tilde{C}_n \tau_n^{2j+2} \left(\|(c + \nabla_{\mathbf{x}} \cdot \mathbf{b})(t)\|_{L^\infty(I_n, \tilde{\Omega}_t)}^2 \right. \\ &\quad \left. + \|\nabla_{\mathbf{x}} \cdot (\mathbf{b} - \mathbf{w})(t)\|_{L^\infty(I_n, \tilde{\Omega}_t)}^2 \right) \int_{I_n} \|D_t^{j+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ &\quad + \sum_{n=1}^N \tilde{D}_n \tau_n^{2j+2} \|\mathbf{b} - \mathbf{w}(t)\|_{L^\infty(I_n, \tilde{\Omega}_t)}^2 \int_{I_n} \left(\|D_t^{j+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + \|\nabla_{\mathbf{x}} D_t^{j+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 \right) dt. \end{aligned}$$

Dabei sind \tilde{C}_n und \tilde{D}_n ALE-abhängige Konstanten.

Beweis. Aufgrund der Äquivalenz zwischen der nichtkonservativen Form (5.20) und der konservativen Form (5.21) werden wir die Aussage lediglich für die Gleichung (5.20) zeigen. Wir schreiben (5.9) in $\tilde{\Omega}_t$ und integrieren diese über I_n . Die erhaltene Gleichung subtrahieren wir von (5.20), so ergibt sich

$$\begin{aligned} &\int_{I_n} (D_t(U(t) - u(t)), V(t))_{\tilde{\Omega}_t} dt + \int_{I_n} ((\mathbf{b} - \mathbf{w})(t) \cdot \nabla_{\mathbf{x}}(U(t) - u(t)), V(t))_{\tilde{\Omega}_t} dt \\ &\quad + ([U]_{n-1}, V(t_{n-1}^+))_{\tilde{\Omega}_{n-1}} + \varepsilon \int_{I_n} (\nabla_{\mathbf{x}}(U(t) - u(t)), \nabla_{\mathbf{x}} V(t))_{\tilde{\Omega}_t} dt \\ &\quad + \int_{I_n} (c(t)(U(t) - u(t)), V(t))_{\tilde{\Omega}_t} dt = 0. \end{aligned}$$

Das Einsetzen von (5.33) und (5.34) in die letzte Gleichung liefert

$$\begin{aligned}
& \int_{I_n} (D_t \boldsymbol{\theta}(t), V(t))_{\tilde{\Omega}_t} dt + ([\boldsymbol{\theta}]_{n-1}, V(t_{n-1}^+))_{\tilde{\Omega}_{n-1}} + \int_{I_n} ((\mathbf{b} - \mathbf{w})(t) \cdot \nabla_{\mathbf{x}} \boldsymbol{\theta}(t), V(t))_{\tilde{\Omega}_t} dt \\
& + \varepsilon \int_{I_n} (\nabla_{\mathbf{x}} \boldsymbol{\theta}(t), \nabla_{\mathbf{x}} V(t))_{\tilde{\Omega}_t} dt + \int_{I_n} (c(t) \boldsymbol{\theta}(t), V(t))_{\tilde{\Omega}_t} dt - \int_{I_n} (c(t) \boldsymbol{\delta}(t), V(t))_{\tilde{\Omega}_t} dt \\
& - \varepsilon \int_{I_n} (\nabla_{\mathbf{x}} \boldsymbol{\delta}(t), \nabla_{\mathbf{x}} V(t))_{\tilde{\Omega}_t} dt - \int_{I_n} ((\mathbf{b} - \mathbf{w})(t) \cdot \nabla_{\mathbf{x}} \boldsymbol{\delta}(t), V(t))_{\tilde{\Omega}_t} dt \\
& + \int_{I_n} (\boldsymbol{\delta}(t) \nabla_{\mathbf{x}} \cdot \mathbf{w}(t), V(t))_{\tilde{\Omega}_t} dt = 0. \tag{5.43}
\end{aligned}$$

Aus einer partiellen Integration im Ort erhalten wir

$$\begin{aligned}
& \int_{I_n} ((\mathbf{b} - \mathbf{w})(t) \cdot \nabla_{\mathbf{x}} \boldsymbol{\delta}(t), V(t))_{\tilde{\Omega}_t} dt \\
& = \int_{I_n} (\boldsymbol{\delta}(t) (\mathbf{b} - \mathbf{w})(t), \nabla_{\mathbf{x}} V(t))_{\tilde{\Omega}_t} dt - \int_{I_n} (\boldsymbol{\delta}(t) \nabla_{\mathbf{x}} \cdot (\mathbf{b} - \mathbf{w})(t), V(t))_{\tilde{\Omega}_t} dt.
\end{aligned}$$

Schließlich folgt aus dem Einsetzen der letzten Gleichung in (5.43):

$$\begin{aligned}
& \int_{I_n} (D_t \boldsymbol{\theta}(t), V(t))_{\tilde{\Omega}_t} dt + ([\boldsymbol{\theta}]_{n-1}, V(t_{n-1}^+))_{\tilde{\Omega}_{n-1}} + \int_{I_n} ((\mathbf{b} - \mathbf{w})(t) \cdot \nabla_{\mathbf{x}} \boldsymbol{\theta}(t), V(t))_{\tilde{\Omega}_t} dt \\
& + \varepsilon \int_{I_n} (\nabla_{\mathbf{x}} \boldsymbol{\theta}(t), \nabla_{\mathbf{x}} V(t))_{\tilde{\Omega}_t} dt + \int_{I_n} (c(t) \boldsymbol{\theta}(t), V(t))_{\tilde{\Omega}_t} dt \\
& = \int_{I_n} (\varepsilon \nabla_{\mathbf{x}} \boldsymbol{\delta}(t) + \boldsymbol{\delta}(t) (\mathbf{b} - \mathbf{w})(t), \nabla_{\mathbf{x}} V(t))_{\tilde{\Omega}_t} dt + \int_{I_n} (\boldsymbol{\delta}(t) (c + \nabla_{\mathbf{x}} \cdot \mathbf{b})(t), V(t))_{\tilde{\Omega}_t} dt. \tag{5.44}
\end{aligned}$$

Nun setzen wir $V = \boldsymbol{\theta}$ in (5.44) und erhalten mithilfe des Satzes 5.4 die Abschätzung

$$\begin{aligned}
& \|\boldsymbol{\theta}(t_m)\|_{L^2(\tilde{\Omega}_m)}^2 + \sum_{j=l}^{m-1} \|[\boldsymbol{\theta}]_j\|_{L^2(\tilde{\Omega}_j)}^2 + \varepsilon \int_{t_l}^{t_m} |\boldsymbol{\theta}(t)|_{1, \tilde{\Omega}_t}^2 dt + c_0 \int_{t_l}^{t_m} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\
& \leq \|\boldsymbol{\theta}(t_l)\|_{L^2(\tilde{\Omega}_l)}^2 + \int_{I_n} (\varepsilon \nabla_{\mathbf{x}} \boldsymbol{\delta}(t) + \boldsymbol{\delta}(t) (\mathbf{b} - \mathbf{w})(t), \nabla_{\mathbf{x}} \boldsymbol{\theta}(t))_{\tilde{\Omega}_t} dt \\
& \quad + \int_{I_n} (\boldsymbol{\delta}(t) (c + \nabla_{\mathbf{x}} \cdot \mathbf{b})(t), \boldsymbol{\theta}(t))_{\tilde{\Omega}_t} dt. \tag{5.45}
\end{aligned}$$

Weiterhin erhalten wir aus einer partiellen Integration im Ort Folgendes:

$$\begin{aligned}
& |(\boldsymbol{\delta}(\mathbf{b} - \mathbf{w}), \nabla_{\mathbf{x}} \boldsymbol{\theta})_{\tilde{\Omega}_t}| \\
& = |-(\nabla_{\mathbf{x}} \boldsymbol{\delta}, (\mathbf{b} - \mathbf{w}) \boldsymbol{\theta})_{\tilde{\Omega}_t} - (\nabla_{\mathbf{x}} \cdot (\mathbf{b} - \mathbf{w}), \boldsymbol{\delta} \boldsymbol{\theta})_{\tilde{\Omega}_t}| \\
& \leq \|\nabla_{\mathbf{x}} \boldsymbol{\delta}\|_{L^2(\tilde{\Omega}_t)} \|\mathbf{b} - \mathbf{w}\|_{L^\infty(\tilde{\Omega}_t)} \|\boldsymbol{\theta}\|_{L^2(\tilde{\Omega}_t)} + \|\boldsymbol{\delta}\|_{L^2(\tilde{\Omega}_t)} \|\nabla_{\mathbf{x}} \cdot (\mathbf{b} - \mathbf{w})\|_{L^\infty(\tilde{\Omega}_t)} \|\boldsymbol{\theta}\|_{L^2(\tilde{\Omega}_t)}. \tag{5.46}
\end{aligned}$$

Außerdem haben wir

$$(\boldsymbol{\delta} (c + \nabla_{\mathbf{x}} \cdot \mathbf{b}), \boldsymbol{\theta})_{\tilde{\Omega}_t} \leq \|\boldsymbol{\delta}\|_{L^2(\tilde{\Omega}_t)} \|c + \nabla_{\mathbf{x}} \cdot \mathbf{b}\|_{L^\infty(\tilde{\Omega}_t)} \|\boldsymbol{\theta}\|_{L^2(\tilde{\Omega}_t)}. \tag{5.47}$$

Der verbleibende Term auf der rechten Seite von (5.45) lässt sich mithilfe der Young-Ungleichung folgendermaßen abschätzen:

$$\begin{aligned}
\varepsilon (\nabla_{\mathbf{x}} \boldsymbol{\delta}, \nabla_{\mathbf{x}} \boldsymbol{\theta})_{\tilde{\Omega}_t} & \leq \varepsilon^{1/2} |\boldsymbol{\delta}|_{1, \tilde{\Omega}_t} \varepsilon^{1/2} |\boldsymbol{\theta}|_{1, \tilde{\Omega}_t} \\
& \leq \frac{\varepsilon}{2} |\boldsymbol{\delta}|_{1, \tilde{\Omega}_t}^2 + \frac{\varepsilon}{2} |\boldsymbol{\theta}|_{1, \tilde{\Omega}_t}^2. \tag{5.48}
\end{aligned}$$

Das Einsetzen der Abschätzungen (5.44)–(5.48) in (5.45), $t_l = 0$ und $t_m = T$ liefert die Behauptung. Dabei haben wir $\|\boldsymbol{\theta}(0)\|_{L^2(\tilde{\Omega}_0)}^2 = 0$ ausgenutzt. □

Ähnliche Ergebnisse kann man bei Bonito, Kyza und Nochetto in [6] finden, jedoch ohne den Reaktionsterm cu .

5.4.4 Quadraturformel und Stabilität

In diesem Abschnitt wollen wir die unbedingte Stabilität unter Verwendung der in Kapitel 2 definierten $(k+1)$ -punktigen Gauß–Radau-Quadratur Q_n erzielen.

Nach dem Einsetzen der Quadraturformel ist die konservative Formulierung nicht mehr äquivalent zu der nichtkonservativen Formulierung. Dies ist darauf zurückzuführen, dass die Reynolds-Identität ungültig wird. Wir werden im Nachfolgenden dennoch sehen, dass sich trotz der fehlenden Äquivalenz, die Stabilität bei beiden Formulierungen mit den gleichen Argumenten zeigen lässt.

Um die Stabilität unter Verwendung der Gauß–Radau-Quadraturformel zu zeigen, betrachten wir das Problem in einer nichtkonservativen Form. Sei $u \in C^1(I, H_0^1(\tilde{\Omega}_t)) \cap \mathcal{X}$ die Lösung des stetigen nichtkonservativen Problems (5.9). Dann gilt

$$\begin{aligned} Q_n \left[(D_t u, v)_{\tilde{\Omega}_t} \right] + Q_n \left[((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} u, v)_{\tilde{\Omega}_t} \right] + \varepsilon Q_n \left[(\nabla_{\mathbf{x}} u, \nabla_{\mathbf{x}} v)_{\tilde{\Omega}_t} \right] + Q_n \left[(cu, v)_{\tilde{\Omega}_t} \right] \\ = Q_n \left[(f, v)_{\tilde{\Omega}_t} \right] \end{aligned} \quad (5.49)$$

für alle $\mathbf{v} \in \{v : \mathbf{v}(t_{n,i}) \in H_0^1(\tilde{\Omega}_{t_{n,i}})\}$, wobei Q_n die in (2.7) definierte Gauß–Radau-Quadraturformel ist.

Sei nun u die Lösung des stetigen konservativen Problems (5.21), dann erfüllt u die (5.49) entsprechende konservative Form

$$\begin{aligned} Q_n \left[\partial_t (u, v)_{\tilde{\Omega}_t} \right] - Q_n \left[((\nabla_{\mathbf{x}} \cdot \mathbf{w}) u, v)_{\tilde{\Omega}_t} \right] + Q_n \left[((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} u, v)_{\tilde{\Omega}_t} \right] + \varepsilon Q_n \left[(\nabla_{\mathbf{x}} u, \nabla_{\mathbf{x}} v)_{\tilde{\Omega}_t} \right] \\ + Q_n \left[(cu, v)_{\tilde{\Omega}_t} \right] - Q_n \left[(u, D_t v)_{\tilde{\Omega}_t} \right] = Q_n \left[(f, v)_{\tilde{\Omega}_t} \right] \end{aligned} \quad (5.50)$$

für alle $\mathbf{v} \in \{v : \mathbf{v}(t_{n,i}) \in H_0^1(\tilde{\Omega}_{t_{n,i}})\}$. Das dG-Verfahren im nichtkonservativen Fall lässt sich über $I_n, n = 1, \dots, N$, folgendermaßen formulieren:

Finde $U \in \mathcal{X}_k(I_n)$ mit $U(0^-) = u_0$ und

$$\begin{aligned} Q_n \left[(D_t U, V)_{\tilde{\Omega}_t} \right] + ([U]_{n-1}, V(t_{n-1}^+))_{\tilde{\Omega}_{n-1}} + Q_n \left[((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} U, V)_{\tilde{\Omega}_t} \right] \\ + \varepsilon Q_n \left[(\nabla_{\mathbf{x}} U, \nabla_{\mathbf{x}} V)_{\tilde{\Omega}_t} \right] + Q_n \left[(cU, V)_{\tilde{\Omega}_t} \right] = Q_n \left[(f, V)_{\tilde{\Omega}_t} \right] \quad \forall V \in \mathcal{X}_k(I_n). \end{aligned} \quad (5.51)$$

Das dG-Verfahren bezüglich der konservativen Variante lautet wie folgt:

Finde $U \in \mathcal{X}_k(I_n)$ mit $U(0^-) = u_0$ und

$$\begin{aligned} Q_n \left[\partial_t (U, V)_{\tilde{\Omega}_t} \right] - Q_n \left[((\nabla_{\mathbf{x}} \cdot \mathbf{w}) U, V)_{\tilde{\Omega}_t} \right] - Q_n \left[(U, D_t V)_{\tilde{\Omega}_t} \right] + ([U]_{n-1}, V(t_{n-1}^+))_{\tilde{\Omega}_{n-1}} \\ + Q_n \left[((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} U, V)_{\tilde{\Omega}_t} \right] + \varepsilon Q_n \left[(\nabla_{\mathbf{x}} U, \nabla_{\mathbf{x}} V)_{\tilde{\Omega}_t} \right] + Q_n \left[(cU, V)_{\tilde{\Omega}_t} \right] \\ = Q_n \left[(f, V)_{\tilde{\Omega}_t} \right] \quad \forall V \in \mathcal{X}_k(I_n). \end{aligned} \quad (5.52)$$

Nun betrachten wir den folgenden Quadraturfehler auf jedem Teilintervall I_n :

$$E_n(\varphi) := \int_{I_n} \varphi(t) dt - Q_n[\varphi]. \quad (5.53)$$

Im Folgenden werden wir sowohl mit der diskreten Norm $Q_n \left[\|V\|_{L^2(\tilde{\Omega}_t)}^2 \right]$ als auch mit der stetigen Norm $\int_{I_n} \|V(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt$ arbeiten. Dazu zitieren wir das folgende Lemma aus [7], das wir in unserem Fall angepasst haben:

Lemma 5.10. *[Diskrete und stetige Norm] Es gilt für alle $n = 1, \dots, N$ und für alle $V \in \mathcal{X}_k(I_n)$*

$$\frac{1}{A_n^2} Q_n \left[\|V\|_{L^2(\tilde{\Omega}_t)}^2 \right] \leq c_1 \int_{I_n} \|V\|_{L^2(\tilde{\Omega}_t)}^2 dt \leq c_2 A_n^2 Q_n \left[\|V\|_{L^2(\tilde{\Omega}_t)}^2 \right], \quad (5.54)$$

wobei A_n in (5.11) definiert ist.

Beweis. Wir gehen hier wie in [7] vor und betrachten $\tilde{\Omega}_n$ als Referenzgebiet. Aus einer Transformation auf $\tilde{\Omega}_n$ erhalten wir

$$\int_{I_n} \|V\|_{L^2(\tilde{\Omega}_t)}^2 dt = \int_{I_n} \int_{\tilde{\Omega}_n} |\widehat{V}(t)|^2 \det \mathbf{J}_{\tilde{\mathbf{A}}_{t_n \rightarrow t}} d\widehat{\mathbf{x}} dt \leq A_n \int_{I_n} \int_{\tilde{\Omega}_n} |\widehat{V}(t)|^2 d\widehat{\mathbf{x}} dt. \quad (5.55)$$

\widehat{V} ist ein Polynom in der Zeit vom Grad kleiner oder gleich k . Darausfolgend ist $\int_{\tilde{\Omega}_n} |\widehat{V}|^2 d\widehat{\mathbf{x}}$ ein Polynom in der Zeit vom Grad kleiner oder gleich $2k$. Somit gilt

$$\int_{I_n} \|\widehat{V}(t)\|_{L^2(\tilde{\Omega}_n)}^2 dt = \frac{\tau_n}{2} \sum_{i=1}^{k+1} \hat{\omega}_{n,i} \|\widehat{V}(t_{n,i})\|_{L^2(\tilde{\Omega}_n)}^2 \leq A_n Q_n \left[\|V\|_{L^2(\tilde{\Omega}_t)}^2 \right]. \quad (5.56)$$

Aus der Kombination von (5.55) und (5.56) folgt die erste Ungleichung in (5.54). Die zweite Ungleichung in (5.54) lässt sich analog dazu zeigen. \square

Wir zitieren ebenso das folgende Lemma aus [7], das wir an unseren Quadraturfehler angepasst haben und das sich trotzdem mit denselben Argumenten beweisen lässt, vgl. [7, Lemma 5.1] für $m = 1$.

Lemma 5.11. *[Gauß–Radau–Quadraturfehler] Der Fehler für die Gauß–Radau–Quadraturformel Q_n auf I_n erfüllt*

$$E_n[v\varphi] \leq C \|v\|_{L^1(I_n)} |\varphi|_{W^{2,\infty}(I_n)} \quad \forall v \in \mathbb{P}_{2k-1}(I_n), \forall \varphi \in W^{2,\infty}(I_n).$$

Im Folgenden wird die Bedingung $\tau_n A_n^2 \leq c_0$ auftreten, die keine Kopplung zwischen der Orts- und Zeitschrittweite aufweist. Diese ist jedoch keine CFL-Bedingung, sondern stellt nur eine Abhängigkeit der Dynamik des Systems dar.

Satz 5.12. *[Gauß–Radau–Quadraturformel und Stabilität] Seien $f \in L^2(S)$ und die ALE-Abbildung $\tilde{\mathbf{A}}_t$ erfüllt (5.32). U sei die Lösung des Problems (5.21). Sei $\tau_j A_j^2 \leq c_0$ für alle $j = m+1, \dots, n$ erfüllt. Es gilt*

$$\begin{aligned} \|U(t_n)\|_{L^2(\tilde{\Omega}_n)}^2 + \sum_{j=m}^{n-1} \|[U]_j\|_{L^2(\tilde{\Omega}_j)}^2 + \varepsilon \sum_{j=m+1}^n Q_j \left[\|\nabla_x U\|_{L^2(\tilde{\Omega}_t)}^2 \right] + c_0 \sum_{j=m+1}^n Q_j \left[\|U\|_{L^2(\tilde{\Omega}_t)}^2 \right] \\ \leq \|U(t_m)\|_{L^2(\tilde{\Omega}_m)}^2 + \frac{1}{c_0} \sum_{j=m+1}^n Q_j \left[\|f\|_{L^2(\tilde{\Omega}_t)}^2 \right], \quad 0 \leq m < n \leq N. \end{aligned}$$

Beweis. Wir zeigen die Behauptung zunächst für das konservative dG-Problem (5.52). Als nächstes setzen wir $V = U$ in (5.52) und erhalten

$$\begin{aligned} Q_n \left[\partial_t \|U\|_{L^2(\tilde{\Omega}_t)}^2 \right] - Q_n \left[\left((\nabla_{\mathbf{x}} \cdot \mathbf{w}) U, U \right)_{\tilde{\Omega}_t} \right] - Q_n \left[(U, D_t U)_{\tilde{\Omega}_t} \right] + ([U]_{n-1}, U(t_{n-1}^+))_{\tilde{\Omega}_{n-1}} \\ + Q_n \left[((\mathbf{b} - \mathbf{w}) \cdot \nabla_{\mathbf{x}} U, U)_{\tilde{\Omega}_t} \right] + \varepsilon Q_n \left[(\nabla_{\mathbf{x}} U, \nabla_{\mathbf{x}} U)_{\tilde{\Omega}_t} \right] + Q_n \left[(cU, U)_{\tilde{\Omega}_t} \right] \\ = Q_n \left[(f, U)_{\tilde{\Omega}_t} \right]. \end{aligned}$$

Unter Ausnutzung der Identität (5.25) ist dies äquivalent zu

$$\begin{aligned} Q_n \left[\partial_t \|U\|_{L^2(\tilde{\Omega}_t)}^2 \right] - \frac{1}{2} Q_n \left[(\nabla_{\mathbf{x}} \cdot \mathbf{w}) U, U \right]_{\tilde{\Omega}_t} - Q_n \left[(U, D_t U) \right]_{\tilde{\Omega}_t} + ([U]_{n-1}, U_{n-1}^+)_{\tilde{\Omega}_{n-1}} \\ + \varepsilon Q_n \left[(\nabla_{\mathbf{x}} U, \nabla_{\mathbf{x}} U) \right]_{\tilde{\Omega}_t} + Q_n \left[\left(c - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}, U^2 \right) \right]_{\tilde{\Omega}_t} = Q_n \left[(f, U) \right]_{\tilde{\Omega}_t}. \end{aligned}$$

Aus (5.53) und der Identität

$$E_n \left[(U, D_t U) \right]_{\tilde{\Omega}_t} + E_n \left[\frac{1}{2} (\nabla_{\mathbf{x}} \cdot \mathbf{w} U, U) \right]_{\tilde{\Omega}_t} = \frac{1}{2} E_n \left[\partial_t (U, U) \right]_{\tilde{\Omega}_t}$$

erhalten wir

$$\begin{aligned} \int_{I_n} \partial_t \|U(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt - \frac{1}{2} \int_{I_n} (\nabla_{\mathbf{x}} \cdot \mathbf{w}(t), U^2(t))_{\tilde{\Omega}_t} dt - \int_{I_n} (U(t), D_t U(t))_{\tilde{\Omega}_t} dt \\ + ([U]_{n-1}, U_{n-1}^+)_{\tilde{\Omega}_{n-1}} + \varepsilon Q_n \left[(\nabla_{\mathbf{x}} U, \nabla_{\mathbf{x}} U) \right]_{\tilde{\Omega}_t} + Q_n \left[\left(c - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}, U^2 \right) \right]_{\tilde{\Omega}_t} \\ = Q_n \left[(f, U) \right]_{\tilde{\Omega}_t} + \frac{1}{2} E_n \left[\partial_t \|U\|_{L^2(\tilde{\Omega}_t)}^2 \right]. \end{aligned}$$

Dies ist aufgrund von (5.27) und (5.28) äquivalent zu

$$\begin{aligned} \frac{1}{2} \|U(t_n^-)\|_{L^2(\tilde{\Omega}_n)}^2 - \frac{1}{2} \|U(t_{n-1}^-)\|_{L^2(\tilde{\Omega}_{n-1})}^2 + \frac{1}{2} \| [U]_{n-1} \|_{L^2(\tilde{\Omega}_{n-1})}^2 + \varepsilon Q_n \left[(\nabla_{\mathbf{x}} U, \nabla_{\mathbf{x}} U) \right]_{\tilde{\Omega}_t} \\ + Q_n \left[\left(c - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}, U^2 \right) \right]_{\tilde{\Omega}_t} = Q_n \left[(f, U) \right]_{\tilde{\Omega}_t} + \frac{1}{2} E_n \left[\partial_t \|U\|_{L^2(\tilde{\Omega}_t)}^2 \right]. \end{aligned} \quad (5.57)$$

Im Folgenden betrachten wir das Gebiet $\tilde{\Omega}_n$ als Referenzgebiet. Der Term in E_n lässt sich folgendermaßen umschreiben:

$$\begin{aligned} \partial_t \|U\|_{L^2(\tilde{\Omega}_t)}^2 &= \partial_t \int_{\tilde{\Omega}_t} U^2 d\mathbf{x} = \partial_t \int_{\tilde{\Omega}_n} \hat{U}^2 \det \mathbf{J}_{\tilde{\mathcal{A}}_{t_n \rightarrow t}} d\hat{\mathbf{x}} \\ &= \int_{\tilde{\Omega}_n} \left(\partial_t \hat{U}^2 \right) \det \mathbf{J}_{\tilde{\mathcal{A}}_{t_n \rightarrow t}} d\hat{\mathbf{x}} + \int_{\tilde{\Omega}_n} \hat{U}^2 \left(\partial_t \det \mathbf{J}_{\tilde{\mathcal{A}}_{t_n \rightarrow t}} \right) d\hat{\mathbf{x}} \\ &=: T_1 + T_2. \end{aligned}$$

Unter Ausnutzung des Lemmas 5.11 lässt sich der Quadraturfehler wie folgt abschätzen:

$$\begin{aligned} E_n[T_1] &\leq C\tau_n^2 \int_{\tilde{\Omega}_n} \left(\int_{I_n} \partial_t |\hat{U}(t)|^2 dt \right) d\hat{\mathbf{x}} |\det \mathbf{J}_{\tilde{\mathcal{A}}_{t_n \rightarrow t}}|_{W^{2,\infty}(I_n)} \\ &\leq C\tau_n A_n \int_{I_n} \|\hat{U}(t)\|_{L^2(\tilde{\Omega}_n)}^2 dt. \end{aligned} \quad (5.58)$$

Dabei folgt der letzte Schritt aus einer inversen Ungleichung und (5.11). Mit derselben Quadraturfehlerabschätzung erhalten wir für den zweiten Term

$$\begin{aligned} E_n[T_2] &\leq C\tau_n^2 \int_{\tilde{\Omega}_n} \left(\int_{I_n} |\hat{U}(t)|^2 dt \right) d\hat{\mathbf{x}} |\partial_t \det \mathbf{J}_{\tilde{\mathcal{A}}_{t_n \rightarrow t}}|_{W^{2,\infty}(I_n)} \\ &\leq C\tau_n A_n \int_{I_n} \|\hat{U}(t)\|_{L^2(\tilde{\Omega}_n)}^2 dt. \end{aligned}$$

Aufgrund von $\|\hat{U}\|_{L^2(\tilde{\Omega}_n)}^2 \in \mathbb{P}_{2k}$ ist die Gauß–Radau-Formel exakt. Anschließend erhalten wir mithilfe von (5.56) und Lemma 5.10 die Abschätzung

$$E_n \left[\partial_t (U, U) \right]_{\tilde{\Omega}_t} \leq C\tau_n A_n^2 Q_n \left[\|U\|_{L^2(\tilde{\Omega}_t)}^2 \right].$$

Bezüglich der äußeren Kraft f wenden wir sowohl die Cauchy–Schwarz- als auch die Young-Ungleichung auf die Quadratur Q_n an und erhalten

$$Q_n \left[(f, u)_{\tilde{\Omega}_t} \right] \leq \frac{c_0}{2} Q_n \left[\|u\|_{L^2(\tilde{\Omega}_t)}^2 \right] + \frac{1}{2c_0} Q_n \left[\|f\|_{L^2(\tilde{\Omega}_t)}^2 \right].$$

Dabei wurde die Positivität der Gewichte der Quadraturformel ausgenutzt. Nach der Kombination von (5.57), (5.2) und den vorherigen Überlegungen sowie $\tau_n A_n^2 \leq c_0$ erhalten wir nach dem Aufsummieren über $j = m + 1, \dots, n$ die Behauptung. \square

Mit derselben Vorgehensweise kann man die Behauptung für die nichtkonservative Form zeigen. Wir setzen $V = U$ in (5.51), nutzen die Identität (5.25) und erhalten

$$\begin{aligned} Q_n \left[(D_t U, U)_{\tilde{\Omega}_t} \right] + \frac{1}{2} Q_n \left[((\nabla_{\mathbf{x}} \cdot \mathbf{w}) U, U)_{\tilde{\Omega}_t} \right] + ([U]_{n-1}, U_{n-1}^+)_{\tilde{\Omega}_{n-1}} \\ + \varepsilon Q_n \left[(\nabla_{\mathbf{x}} U, \nabla_{\mathbf{x}} U)_{\tilde{\Omega}_t} \right] + Q_n \left[\left(c - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}, U^2 \right)_{\tilde{\Omega}_t} \right] = Q_n \left[(f, U)_{\tilde{\Omega}_t} \right]. \end{aligned}$$

Mithilfe der Definition (5.53) ergibt sich

$$\begin{aligned} \int_{I_n} (D_t U(t), U(t))_{\tilde{\Omega}_t} dt + \frac{1}{2} \int_{I_n} \left((\nabla_{\mathbf{x}} \cdot \mathbf{w}(t)) U(t), U(t) \right)_{\tilde{\Omega}_t} dt + ([U]_{n-1}, U_{n-1}^+)_{\tilde{\Omega}_{n-1}} \\ + \varepsilon Q_n \left[(\nabla_{\mathbf{x}} U, \nabla_{\mathbf{x}} U)_{\tilde{\Omega}_t} \right] + Q_n \left[\left(c - \frac{1}{2} \nabla_{\mathbf{x}} \cdot \mathbf{b}, U^2 \right)_{\tilde{\Omega}_t} \right] \\ = Q_n \left[(f, U)_{\tilde{\Omega}_t} \right] + E_n \left[(D_t U, U)_{\tilde{\Omega}_t} \right] + E_n \left[\left((\nabla_{\mathbf{x}} \cdot \mathbf{w}) U, U \right)_{\tilde{\Omega}_t} \right]. \end{aligned}$$

Mithilfe der Identitäten (5.27) und

$$E_n \left[(D_t U, U)_{\tilde{\Omega}_t} \right] + E_n \left[(\mathbf{w} \nabla_{\mathbf{x}} U, U)_{\tilde{\Omega}_t} \right] = E_n \left[\partial_t (U, U)_{\tilde{\Omega}_t} \right]$$

erhalten wir die Gleichung (5.57). Mit den gleichen Argumenten kann man, wie bereits für die konservative Variante dargestellt, weiter vorgehen.

Auch im diskreten Fall ist die Stabilitätsabschätzung frei von dem Term $\frac{1}{\varepsilon}$.

5.4.5 Quadraturformel und Fehlerabschätzung

Das Ziel hier ist eine unbedingte Fehlerabschätzung unter Verwendung der in Kapitel 2 definierten $(k + 1)$ -punktigen Gauß–Radau-Quadratur Q_n zu erhalten. Wie im vorherigen Unterabschnitt gesehen, sind die nichtkonservative Formulierung und die konservative Formulierung nicht mehr äquivalent. Dennoch lassen sich die Fehlerabschätzungen bei beiden Formulierungen mit den gleichen Argumenten zeigen.

Wir werden für das semidiskrete Problem (5.51) bzw. (5.52) eine Fehleranalyse präsentieren. Dabei werden wir Interpolationsoperatoren verwenden, die die Stützstellen der Gauß–Radau-Quadratur nutzen.

Bezüglich der nichtkonservativen Form liefert die Subtraktion der Gleichung (5.49) von (5.51) für alle $n = 1, \dots, N$:

$$Q_n \left[(D_t U - D_t u, V)_{\tilde{\Omega}_t} \right] + Q_n \left[a_{nc}(t; U - u, V) \right] + ([U - u]_{n-1}, V_{n-1}^+)_{\tilde{\Omega}_{n-1}} = 0 \quad \forall V \in \mathcal{X}_k(I_n). \quad (5.59)$$

Als nächsten Schritt betrachten wir den Lagrange-Interpolationsoperator in der Zeit

$$P_n[\mathbf{v}](t) := \sum_{i=1}^{k+1} \phi_{n,i}(t) \mathbf{v}(t_{n,i}),$$

wobei $t_{n,i}$ die Stützstellen der betrachteten Gauß–Radau-Quadraturformel und $\phi_{n,i}$ die Lagrange-Basis-Polynome mit $\phi_{n,i}(t_{n,j}) = \delta_{i,j}$, $i, j = 1, \dots, k+1$, sind.

Wie im vorherigen Abschnitt werden wir den Fehler in einen Interpolationsfehler $\boldsymbol{\delta} := u - P_n[u]$ und in einen diskreten Fehleranteil $\boldsymbol{\theta} := U - P_n[u]$ aufspalten. Es gilt

$$U - u = \boldsymbol{\theta} - \boldsymbol{\delta}. \quad (5.60)$$

Das Einsetzen dieser Fehlertrennung in (5.59) liefert

$$\begin{aligned} Q_n \left[(D_t \boldsymbol{\theta}, V)_{\tilde{\Omega}_t} \right] + Q_n \left[a_{nc}(t; \boldsymbol{\theta}, V) \right] + ([\boldsymbol{\theta}]_{n-1}, V_{n-1}^+)_{\tilde{\Omega}_{n-1}} \\ = Q_n \left[(D_t \boldsymbol{\delta}, V)_{\tilde{\Omega}_t} \right] + Q_n \left[a_{nc}(t; \boldsymbol{\delta}, V) \right] + ([\boldsymbol{\delta}]_{n-1}, V_{n-1}^+)_{\tilde{\Omega}_{n-1}}. \end{aligned} \quad (5.61)$$

Desweiteren geben wir die folgende Abschätzung für den Quadraturfehler an:

Lemma 5.13. *Sei Q_n die in (2.7) definierte Gauß–Radau-Quadraturformel. Seien $\tilde{\mathcal{A}}_{t_n \rightarrow t} \in W_\infty^{j+1}(I_n; W_\infty^1(\tilde{\Omega}_{t_n}))$ und $\hat{g} \in H^{j+1}(I_n; L^2(\tilde{\Omega}_{t_n}))$. Es gilt für $j = 0, \dots, k$ und $v \in \mathcal{X}_k(I_n)$*

$$|E_n[(g, v)_{\tilde{\Omega}_t}]| \leq C A_n B_{n,j+1} \tau_n^{j+1} \sum_{i=0}^{j+1} \int_{I_n} \|D_t^i g(t)\|_{L^2(\tilde{\Omega}_t)} \|v(t)\|_{L^2(\tilde{\Omega}_t)} dt, \quad (5.62)$$

wobei $B_{n,j+1} := \max_{r=0, \dots, j+1} \|\nabla_{\mathbf{Y}} \tilde{\mathcal{A}}_{t_n \rightarrow t}\|_{W^{r, \infty}(I_n, L^2(\tilde{\Omega}_{t_n}))}$ und C eine Konstante, die nur von k abhängt, sind. Weiter gilt für alle $\mu > 0$

$$|E_n[(g, v)_{\tilde{\Omega}_t}]| \leq \frac{A_n^2 B_{n,j+1}^2}{\mu c_0} \tau_n^{2(j+1)} \sum_{i=0}^{j+1} \int_{I_n} \|D_t^i g(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt + \mu c_0 \int_{I_n} \|v(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt. \quad (5.63)$$

Beweis. Mit \hat{I}_k bezeichnen wir die Lagrange-Interpolation an den Gauß–Radau-Punkten auf I_n . Zunächst schreiben wir den Term $Q_n[(g, v)]$ auf das Referenzgebiet $\tilde{\Omega}_n$ um. Es gilt

$$Q_n[(g, v)_{\tilde{\Omega}_t}] = \sum_{j=0}^k \omega_{n,j} (\hat{h}(t_{n,j}), \hat{v}(t_{n,j}))_{\tilde{\Omega}_{t_n}} = \int_{I_n} (\hat{I}_k \hat{h}(t), \hat{v}(t))_{\tilde{\Omega}_{t_n}},$$

wobei $\hat{h}(t) := \det \mathbf{J}_{\tilde{\mathcal{A}}_{t_n \rightarrow t}} \hat{g}(t)$. Weiterhin gilt

$$\begin{aligned} |E_n[(g, v)_{\tilde{\Omega}_t}]| &\leq \int_{I_n} |(\hat{h}(t) - \hat{I}_k \hat{h}(t), \hat{v}(t))_{\tilde{\Omega}_{t_n}}| dt \\ &\leq C \tau_n^{j+1} \int_{I_n} \|D_t^{j+1} \hat{h}(t)\|_{L^2(\tilde{\Omega}_{t_n})} \|\hat{v}(t)\|_{L^2(\tilde{\Omega}_{t_n})} dt \\ &\leq C A_n \tau_n^{j+1} \sum_{i=0}^{j+1} \int_{I_n} \|D_t^i \hat{g}(t)\|_{L^2(\tilde{\Omega}_t)} \|D_t^{j+1-i} \det \mathbf{J}_{\tilde{\mathcal{A}}_{t_n \rightarrow t}}\|_{L^2(\tilde{\Omega}_t)} \|\hat{v}(t)\|_{L^2(\tilde{\Omega}_t)} dt \\ &\leq C A_n B_{n,j+1} \tau_n^{j+1} \sum_{i=0}^{j+1} \int_{I_n} \|D_t^i g(t)\|_{L^2(\tilde{\Omega}_t)} \|v(t)\|_{L^2(\tilde{\Omega}_t)} dt. \end{aligned}$$

Die zweite Behauptung folgt aus der Young-Ungleichung, der Regularität von g , der Definition der ALE-Abbildung und (5.11). \square

Im Weiteren werden wir eine neue Projektion \hat{P}_n benötigen, die wir wie folgt definieren:

Definition 5.14. Sei \widehat{P}_n der Lagrange-Interpolationsoperator in der Zeit definiert durch

$$\widehat{P}_n[\mathbf{v}](t) := \sum_{i=0}^{k+1} \phi_{n,i}(t) \mathbf{v}(t_{n,i}),$$

wobei $t_{n,i}$ die Stützstellen der betrachteten Gauß–Radau-Quadraturformel sind und $t_{n,0} = t_{n-1}$ ist. Weiterhin sind $\phi_{n,i}$ die Lagrange-Basis-Polynome mit $\phi_{n,i}(t_{n,j}) = \delta_{i,j}$, $i, j = 0, \dots, k+1$.

Es gelten die folgenden Fehlerabschätzungen für die Interpolation \widehat{P}_n :

$$\int_{I_n} \|u(t) - \widehat{P}_n[u](t)\|_{L^2(\widetilde{\Omega}_t)} dt \leq C\tau_n^{k+3} \sup_{t \in I_n} \|\partial_t^{k+2} u(t)\|_{L^2(\widetilde{\Omega}_t)}, \quad (5.64)$$

$$\sup_{I_n} \|P_n[u](t) - \widehat{P}_n[u](t)\|_{L^2(\widetilde{\Omega}_t)} \leq C\tau_n^{k+1} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\widetilde{\Omega}_t)} \quad (5.65)$$

und

$$\sup_{I_n} \|D_t^r P_n[u](t) - D_t^r \widehat{P}_n[u](t)\|_{L^2(\widetilde{\Omega}_t)} \leq C\tau_n^{k+1-r} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\widetilde{\Omega}_t)}, \quad r = 1, \dots, k+1. \quad (5.66)$$

Als Zugang zur Fehlerabschätzung betrachten wir das folgende Lemma:

Lemma 5.15. Für alle $V \in \mathcal{X}_k(I_n)$ gilt

$$\begin{aligned} Q_n \left[(D_t \boldsymbol{\delta}, V)_{\widetilde{\Omega}_t} \right] + Q_n \left[a_{nc}(t; \boldsymbol{\delta}, V) \right] + ([\boldsymbol{\delta}]_{n-1}, V_{n-1}^+)_{\widetilde{\Omega}_{n-1}} \\ \leq C\tau_n^{k+2} \sup_{I_n} \|\partial_t^{k+2} u(t)\|_{L^2(\widetilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)} \\ + CA_n \tau_n^{k+2} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\widetilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)} \\ + CA_n B_{n,k+1} \tau_n^{k+2} \sum_{i=0}^k \tau_n^{k-i} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\widetilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)}. \end{aligned} \quad (5.67)$$

Beweis. Wir beginnen mit der Abschätzung der einzelnen Terme auf der linken Seite von (5.61) und erhalten einerseits

$$Q_n[a_{nc}(t; \boldsymbol{\delta}, V)] = \tau_n \sum_{i=0}^k \omega_i a_{nc}(t_{n,i}; u(t_{n,i}) - P_n[u](t_{n,i})V(t_{n,i})) = 0 \quad (5.68)$$

und andererseits

$$Q_n \left[(D_t \boldsymbol{\delta}, V)_{\widetilde{\Omega}_t} \right] = Q_n \left[(D_t u, V)_{\widetilde{\Omega}_t} \right] - Q_n \left[(D_t P_n[u], V)_{\widetilde{\Omega}_t} \right].$$

Wir können den zweiten Term wie folgt umschreiben:

$$-Q_n \left[(D_t P_n[u], V)_{\widetilde{\Omega}_t} \right] = - \int_{I_n} (D_t P_n[u](t), V(t))_{\widetilde{\Omega}_t} dt + E_n \left[(D_t P_n[u], V)_{\widetilde{\Omega}_t} \right]. \quad (5.69)$$

Aus einer partiellen Integration erhalten wir

$$\begin{aligned} & - \int_{I_n} (D_t P_n[u](t), V(t))_{\widetilde{\Omega}_t} dt \\ & = \int_{I_n} (P_n[u](t), D_t V(t))_{\widetilde{\Omega}_t} dt - \int_{I_n} \partial_t (P_n[u](t), V(t))_{\widetilde{\Omega}_t} dt + \int_{I_n} (P_n[u](t), V(t) \nabla_{\mathbf{x}} \cdot \mathbf{w}(t))_{\widetilde{\Omega}_t} dt \\ & = \int_{I_n} (P_n[u](t), D_t V(t))_{\widetilde{\Omega}_t} dt - (P_n[u](t_n^-), V_{n-1}^-)_{\widetilde{\Omega}_n} + (P_n[u](t_{n-1}^+), V_{n-1}^+)_{\widetilde{\Omega}_{n-1}} \\ & \quad + \int_{I_n} (P_n[u](t), V(t) \nabla_{\mathbf{x}} \cdot \mathbf{w}(t))_{\widetilde{\Omega}_t} dt. \end{aligned} \quad (5.70)$$

Mithilfe der Definition der Interpolation \widehat{P}_n und des Quadraturfehlers haben wir

$$\begin{aligned} \int_{I_n} (P_n[u](t), D_t V(t))_{\widetilde{\Omega}_t} dt &= Q_n \left[(P_n[u], D_t V)_{\widetilde{\Omega}_t} \right] + E_n \left[(P_n[u], D_t V)_{\widetilde{\Omega}_t} \right] \\ &= Q_n \left[(\widehat{P}_n[u], D_t V)_{\widetilde{\Omega}_t} \right] + E_n \left[(P_n[u], D_t V)_{\widetilde{\Omega}_t} \right] \\ &= \int_{I_n} (\widehat{P}_n[u](t), D_t V(t))_{\widetilde{\Omega}_t} dt + E_n \left[(P_n[u], D_t V)_{\widetilde{\Omega}_t} \right] \\ &\quad - E_n \left[(\widehat{P}_n[u], D_t V)_{\widetilde{\Omega}_t} \right]. \end{aligned}$$

Weiterhin liefert eine partielle Integration

$$\begin{aligned} \int_{I_n} (\widehat{P}_n[u](t), D_t V(t))_{\widetilde{\Omega}_t} dt &= - \int_{I_n} (D_t \widehat{P}_n[u](t), V(t))_{\widetilde{\Omega}_t} dt + (\widehat{P}_n[u](t_n^-), V_n^-)_{\widetilde{\Omega}_n} \\ &\quad - (\widehat{P}_n[u](t_{n-1}^+), V_{n-1}^+)_{\widetilde{\Omega}_{n-1}} - \int_{I_n} (\widehat{P}_n[u](t), V(t) \nabla_{\mathbf{x}} \cdot \mathbf{w}(t))_{\widetilde{\Omega}_t} dt. \end{aligned}$$

Durch die Transformation des ersten Integrals in eine Quadraturformel erhalten wir

$$\int_{I_n} (D_t \widehat{P}_n[u](t), V(t))_{\widetilde{\Omega}_t} dt = Q_n \left[(D_t \widehat{P}_n[u], V)_{\widetilde{\Omega}_t} \right] + E_n \left[(D_t \widehat{P}_n[u], V)_{\widetilde{\Omega}_t} \right].$$

Aus der Definition von \widehat{P}_n kann man $\widehat{P}_n[u](t_n^-) = P_n[u](t_n^-)$ und $\widehat{P}_n[u](t_{n-1}^+) = u(t_{n-1}) = u(t_{n-1}^-) = P_{n-1}[u](t_{n-1}^-)$ entnehmen. Zusammengefasst erhalten wir

$$\begin{aligned} &Q_n \left[(D_t \boldsymbol{\delta}, V)_{\widetilde{\Omega}_t} \right] \\ &= Q_n \left[(D_t(u - \widehat{P}_n[u]), V)_{\widetilde{\Omega}_t} \right] + (P_n[u](t_{n-1}^+), V_{n-1}^+)_{\widetilde{\Omega}_{n-1}} - (P_{n-1}[u](t_{n-1}^-), V_{n-1}^+)_{\widetilde{\Omega}_{n-1}} \\ &\quad + E_n \left[(D_t \widehat{P}_n[u] - D_t P_n[u], V)_{\widetilde{\Omega}_t} \right] + E_n \left[(P_n[u] - \widehat{P}_n[u], D_t V)_{\widetilde{\Omega}_t} \right] \\ &\quad + \int_{I_n} (P_n[u](t) - \widehat{P}_n[u](t), V(t) \nabla_{\mathbf{x}} \cdot \mathbf{w}(t))_{\widetilde{\Omega}_t} dt. \end{aligned}$$

Der Sprungterm in (5.67) lässt sich wie folgt umschreiben:

$$\begin{aligned} &([\boldsymbol{\delta}]_{n-1}, V_{n-1}^+)_{\widetilde{\Omega}_{n-1}} \\ &= \underbrace{([u]_{n-1}, V_{n-1}^+)_{\widetilde{\Omega}_{n-1}}}_{=0} - (P_n[u](t_{n-1}^+), V_{n-1}^+)_{\widetilde{\Omega}_{n-1}} + (P_{n-1}[u](t_{n-1}^-), V_{n-1}^+)_{\widetilde{\Omega}_{n-1}}. \end{aligned}$$

Die Kombination der letzten beiden Gleichungen und (5.68) liefert

$$\begin{aligned} &Q_n \left[(D_t \boldsymbol{\delta}, V)_{\widetilde{\Omega}_t} \right] + Q_n \left[a_{nc}(t; \boldsymbol{\delta}, V) \right] + ([\boldsymbol{\delta}]_{n-1}, V_{n-1}^+)_{\widetilde{\Omega}_{n-1}} \\ &= Q_n \left[(D_t(u - \widehat{P}_n[u]), V)_{\widetilde{\Omega}_t} \right] + \int_{I_n} (P_n[u](t) - \widehat{P}_n[u](t), V(t) \nabla_{\mathbf{x}} \cdot \mathbf{w}(t)) dt + F_n, \end{aligned} \tag{5.71}$$

wobei $F_n := E_n \left[(D_t \widehat{P}_n[u] - D_t P_n[u], V)_{\widetilde{\Omega}_t} \right] + E_n \left[(P_n[u] - \widehat{P}_n[u], D_t V)_{\widetilde{\Omega}_t} \right]$. Weiterhin folgt aus der Cauchy-Schwarz-Ungleichung einerseits, dass

$$\begin{aligned} Q_n \left[(D_t(u - \widehat{P}_n[u]), V)_{\widetilde{\Omega}_t} \right] &\leq \tau_n \sum_{i=0}^k \omega_i \|D_t(u(t_{n,i}) - \widehat{P}_n[u](t_{n,i}))\|_{L^2(\widetilde{\Omega}_{t_{n,i}})} \|V(t_{n,i})\|_{L^2(\widetilde{\Omega}_{t_{n,i}})} \\ &\leq C \tau_n \sup_{I_n} \|D_t(u(t) - \widehat{P}_n[u](t))\|_{L^2(\widetilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)} \\ &\leq C \tau_n^{k+2} \sup_{I_n} \|\partial_t^{k+2} u(t)\|_{L^2(\widetilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)} \end{aligned} \tag{5.72}$$

und andererseits, dass

$$\begin{aligned} \int_{I_n} (P_n[u](t) - \widehat{P}_n[u](t), V(t) \nabla_{\mathbf{x}} \cdot \mathbf{w}(t))_{\widetilde{\Omega}_t} dt \\ \leq A_n \tau_n \sup_{I_n} \|(P_n[u](t) - \widehat{P}_n[u](t))\|_{L^2(\widetilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)}. \end{aligned}$$

Hierbei haben wir die Regularität der ALE-Abbildung ausgenutzt. Daraufaufbauend erhalten wir mithilfe von (5.65) die Abschätzung

$$\begin{aligned} \int_{I_n} (P_n[u](t) - \widehat{P}_n[u](t), V(t) \nabla_{\mathbf{x}} \cdot \mathbf{w}(t)) dt \\ \leq C A_n \tau_n^{k+2} \sup_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\widetilde{\Omega}_t)}. \end{aligned} \quad (5.73)$$

Zuletzt lassen sich die Fehlerterme mithilfe von (5.62) und von (5.66) folgendermaßen abschätzen. Es gilt einerseits

$$\begin{aligned} \left| E_n \left[(D_t P_n[u] - D_t \widehat{P}_n[u], V)_{\widetilde{\Omega}_t} \right] \right| \\ \leq C A_n B_{n,k+1} \tau_n^{k+1} \sum_{i=0}^{k+1} \int_{I_n} \|D_t^i (D_t P_n[u](t) - D_t \widehat{P}_n[u](t))\|_{L^2(\widetilde{\Omega}_t)} \|V(t)\|_{L^2(\widetilde{\Omega}_t)} dt \\ \leq C A_n B_{n,k+1} \tau_n^{k+2} \sum_{i=0}^k \tau_n^{k-i} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\widetilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)}, \end{aligned} \quad (5.74)$$

wobei wir in Betracht gezogen haben, dass $D_t^{k+1} (D_t P_n[u] - D_t \widehat{P}_n[u]) = 0$. Andererseits gilt

$$\begin{aligned} \left| E_n \left[(P_n[u] - \widehat{P}_n[u], D_t V)_{\widetilde{\Omega}_t} \right] \right| \\ \leq C A_n B_{n,k+1} \tau_n^{k+1} \sum_{i=0}^k \int_{I_n} \|D_t^i (P_n[u](t) - \widehat{P}_n[u](t))\|_{L^2(\widetilde{\Omega}_t)} \|D_t V(t)\|_{L^2(\widetilde{\Omega}_t)} dt \\ \leq C A_n B_{n,k+1} \tau_n^{k+2} \sum_{i=0}^k \tau_n^{k+1-i} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\widetilde{\Omega}_t)} \sup_{I_n} \|D_t V(t)\|_{L^2(\widetilde{\Omega}_t)} \\ \leq C A_n B_{n,k+1} \tau_n^{k+2} \sum_{i=0}^k \tau_n^{k-i} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\widetilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\widetilde{\Omega}_t)}, \end{aligned} \quad (5.75)$$

wobei wir eine inverse Ungleichung verwendet haben. Die Behauptung folgt aus der Kombination von (5.71)–(5.75). \square

Nun können wir die folgende Fehlerabschätzung angeben:

Lemma 5.16. *Seien U die Lösung des diskreten Problems (5.49) und u die Lösung von (5.4). Es gilt für hinreichend kleine τ_n , $n = 1, \dots, N$, Es gilt*

$$\begin{aligned} \|(U - u)(T)\|_{L^2(\widetilde{\Omega}_t)}^2 + \frac{\gamma c_0}{2} \int_0^T \|(U - u)(t)\|_{L^2(\widetilde{\Omega}_t)}^2 dt \\ \leq \|(U - u)(0^-)\|_{L^2(\widetilde{\Omega}_0)}^2 \\ + C \tau_n^{2k+2} \left(\sup_{t \in [0, T]} \|\partial_t^{k+2} u(t)\|_{L^2(\widetilde{\Omega}_t)}^2 + (1 + A_n^2) \sup_{t \in [0, T]} \|\partial_t^{k+1} u(t)\|_{L^2(\widetilde{\Omega}_t)}^2 + B_{k+1} \right), \end{aligned}$$

wobei

$$B_{k+1} := \max_{n=1, \dots, N} \frac{A_n^2 B_{n,k+1}^2}{\gamma c_0} \sum_{i=0}^k \tau_n^{2(k-i)} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\widetilde{\Omega}_t)}^2.$$

Hierbei ist $\gamma > 0$ die Konstante aus der Koerzivität von a_{nc} .

Beweis. Aus Lemma 5.15 und (5.61) ergibt sich

$$\begin{aligned}
 Q_n \left[(D_t \boldsymbol{\theta}, V)_{\tilde{\Omega}_t} \right] + Q_n \left[a_{nc}(t; \boldsymbol{\theta}, V) \right] + ([\boldsymbol{\theta}]_{n-1}, V_{n-1}^+)_{\tilde{\Omega}_{n-1}} \\
 \leq C \tau_n \tau_n^{k+1} \sup_{I_n} \|\partial_t^{k+2} u(t)\|_{L^2(\tilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\tilde{\Omega}_t)} \\
 + C A_n \tau_n \tau_n^{k+1} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\tilde{\Omega}_t)} \\
 + C A_n B_{n,k+1} \tau_n \tau_n^{k+1} \sum_{i=0}^k \tau_n^{k-i} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\tilde{\Omega}_t)}. \quad (5.76)
 \end{aligned}$$

Als nächstes setzen wir $V = \boldsymbol{\theta}$ in (5.76) und formulieren die linke Seite davon um. Der erste Term in (5.76) lässt sich folgendermaßen umschreiben:

$$\begin{aligned}
 Q_n \left[(D_t \boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] &= \int_{I_n} (D_t \boldsymbol{\theta}(t), \boldsymbol{\theta}(t))_{\tilde{\Omega}_t} dt - E_n \left[(D_t \boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] \\
 &= \frac{1}{2} \int_{I_n} \partial_t \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt - E_n \left[(D_t \boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] - \frac{1}{2} \int_{I_n} (\nabla_{\mathbf{x}} \cdot \mathbf{w}, \boldsymbol{\theta}^2)_{\tilde{\Omega}_t} dt \\
 &= \frac{1}{2} \|\boldsymbol{\theta}_n^-\|_{L^2(\tilde{\Omega}_n)}^2 - \frac{1}{2} \|\boldsymbol{\theta}_{n-1}^+\|_{L^2(\tilde{\Omega}_{n-1})}^2 - E_n \left[(D_t \boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] - \frac{1}{2} \int_{I_n} (\nabla_{\mathbf{x}} \cdot \mathbf{w}, \boldsymbol{\theta}^2)_{\tilde{\Omega}_t} dt.
 \end{aligned}$$

Desweiteren gilt

$$([\boldsymbol{\theta}]_{n-1}, \boldsymbol{\theta}_{n-1}^+)_{\tilde{\Omega}_{n-1}} = \frac{1}{2} \|\boldsymbol{\theta}_{n-1}^+\|_{L^2(\tilde{\Omega}_{n-1})}^2 - \frac{1}{2} \|\boldsymbol{\theta}_{n-1}^-\|_{L^2(\tilde{\Omega}_{n-1})}^2 + \frac{1}{2} \|[\boldsymbol{\theta}]_{n-1}\|_{L^2(\tilde{\Omega}_{n-1})}^2.$$

Somit ergibt sich aus (5.76)

$$\begin{aligned}
 \|\boldsymbol{\theta}_n^-\|_{L^2(\tilde{\Omega}_n)}^2 - \|\boldsymbol{\theta}_{n-1}^-\|_{L^2(\tilde{\Omega}_{n-1})}^2 + \|[\boldsymbol{\theta}]_{n-1}\|_{L^2(\tilde{\Omega}_{n-1})}^2 + 2Q_n \left[a_{nc}(t; \boldsymbol{\theta}, \boldsymbol{\theta}) \right] \\
 - 2E_n \left[(D_t \boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] - 2E_n \left[(\boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] - \int_{I_n} (\nabla_{\mathbf{x}} \cdot \mathbf{w}(t), \boldsymbol{\theta}^2(t))_{\tilde{\Omega}_t} dt \\
 \leq C \tau_n \tau_n^{k+1} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)} \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)} \\
 + C A_n \tau_n \tau_n^{k+1} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)} \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)} \\
 + C A_n B_{n,k+1} \tau_n \tau_n^{k+1} \sum_{i=0}^k \tau_n^{k-i} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)} \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}. \quad (5.77)
 \end{aligned}$$

Den Fehler $\left| E_n \left[(D_t \boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] \right|$ wollen wir nach oben abschätzen. Wir schreiben den Term auf das Referenzgebiet $\tilde{\Omega}_n$ um und verwenden Lemma 5.11. Somit haben wir

$$\begin{aligned}
 2 \left| E_n \left[(D_t \boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] \right| &\leq C \tau_n^2 \int_{\tilde{\Omega}_n} \left(\int_{I_n} |\partial_t \hat{\boldsymbol{\theta}}(t)| |\hat{\boldsymbol{\theta}}(t)| dt \right) d\hat{\mathbf{x}} |\partial_t \det \mathbf{J}_{\tilde{\mathcal{A}}_{t_n \rightarrow t}}|_{W^{2,\infty}(I_n)} \\
 &\leq C \tau_n^2 A_n \left(\int_{I_n} \|\partial_t \hat{\boldsymbol{\theta}}(t)\|_{L^2(\tilde{\Omega}_n)}^2 dt \right)^{1/2} \left(\int_{I_n} \|\hat{\boldsymbol{\theta}}(t)\|_{L^2(\tilde{\Omega}_n)}^2 dt \right)^{1/2},
 \end{aligned}$$

wobei der letzte Schritt aus der Cauchy-Schwarz-Ungleichung folgt. Weiterhin liefert eine inverse Ungleichung

$$2 \left| E_n \left[(D_t \boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] \right| \leq C \tau_n A_n \int_{I_n} \|\hat{\boldsymbol{\theta}}(t)\|_{L^2(\tilde{\Omega}_n)}^2 dt$$

und mithilfe von (5.56) und von Lemma 5.10 ergibt sich die Abschätzung

$$2E_n \left[(D_t \boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] \leq C \tau_n A_n Q_n \left[\|\hat{\boldsymbol{\theta}}(t)\|_{L^2(\tilde{\Omega}_n)}^2 \right] \leq C \tau_n A_n^2 \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt. \quad (5.78)$$

Anhand der Koerzivitat der Bilinearform a_{nc} erhalten wir

$$Q_n \left[a_{nc}(t; \boldsymbol{\theta}, \boldsymbol{\theta}) \right] \geq \gamma \int_{I_n} \|\boldsymbol{\theta}(t)\|_{\varepsilon, \tilde{\Omega}_t}^2 dt - \gamma E_n \left[(\boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right].$$

Analog zu (5.58) gilt nach Lemma 5.11 die Fehlerabschatzung

$$\begin{aligned} \gamma E_n \left[(\boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] &\leq C \tau_n^2 \int_{\tilde{\Omega}_n} \left(\int_{I_n} |\hat{\boldsymbol{\theta}}(t)|^2 dt \right) d\hat{\mathbf{x}} \left| \partial_t \det \mathbf{J}_{\tilde{\mathcal{A}}_{t_n \rightarrow t}} \right|_{W^{2,\infty}(I_n)} \\ &\leq C \tau_n A_n \int_{I_n} \|\hat{\boldsymbol{\theta}}(t)\|_{L^2(\tilde{\Omega}_n)}^2 dt. \end{aligned}$$

Daraus ergibt sich wie in (5.78) die Abschatzung

$$(\gamma + 2) E_n \left[(\boldsymbol{\theta}, \boldsymbol{\theta})_{\tilde{\Omega}_t} \right] \leq C \tau_n A_n Q_n \left[\|\hat{\boldsymbol{\theta}}(t)\|_{L^2(\tilde{\Omega}_n)}^2 \right] \leq C_1 \tau_n A_n^2 \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt.$$

Weiterhin erhalten wir anhand der Regularitat der ALE-Abbildung die Abschatzung

$$\int_{I_n} (\nabla_{\mathbf{x}} \cdot \mathbf{w}(t), \boldsymbol{\theta}^2(t))_{\tilde{\Omega}_t} dt \leq A_n \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_n)}^2 dt.$$

Desweiteren ergibt sich aus (5.63) und einer inversen Ungleichung

$$\begin{aligned} C A_n B_{n,k+1} \tau_n \tau_n^{k+1} \sum_{i=0}^k \tau_n^{k-i} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)} \sup_{I_n} \|V(t)\|_{L^2(\tilde{\Omega}_t)} \\ \leq C \tau_n \tau_n^{2k+2} B_{k+1} + \tau_n \tilde{C}_2 \gamma c_0 \tau_n \sup_{I_n} \|V(t)\|_{L^2(\tilde{\Omega}_t)}^2 \\ \leq C \tau_n \tau_n^{2k+2} B_{k+1} + C_2 \gamma c_0 \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt, \end{aligned}$$

wobei B_{k+1} in Lemma 5.16 definiert ist. Zusammengefasst ergibt sich unter Anwendung der Young- und einer inversen Ungleichung auf die restlichen Terme auf der rechten Seite von (5.77) die Abschatzung

$$\begin{aligned} \|\boldsymbol{\theta}_n^-\|_{L^2(\tilde{\Omega}_n)}^2 - \|\boldsymbol{\theta}_{n-1}^-\|_{L^2(\tilde{\Omega}_{n-1})}^2 + \|[\boldsymbol{\theta}]_{n-1}\|_{L^2(\tilde{\Omega}_{n-1})}^2 + 2\gamma \int_{I_n} \|\boldsymbol{\theta}(t)\|_{\varepsilon, \tilde{\Omega}_t}^2 dt \\ - \bar{C} \tau_n A_n^2 \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_n)}^2 dt - A_n \int_{I_n} \|\boldsymbol{\theta}\|_{L^2(\tilde{\Omega}_n)}^2 dt \\ \leq C \tau_n \tau_n^{2k+2} \sup_{I_n} \|\partial_t^{k+2} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + C_3 \gamma c_0 \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ + C A_n^2 \tau_n \tau_n^{2k+2} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + C_4 \gamma c_0 \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ + C \tau_n \tau_n^{2k+2} B_{k+1} + C_2 \gamma c_0 \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt. \end{aligned}$$

Fur τ_n hinreichend klein gilt $C_1 \tau_n A_n^2 \leq \frac{c_0}{2}$. Nun wahlen wir die Konstanten C_2, C_3 und C_4 , sodass fur eine hinreichend groe Konstante c_0 die Beziehung $C_2 + C_3 + C_4 + \frac{A_n}{\gamma c_0} = 1$ gilt. Dann folgt aufgrund von $\|\boldsymbol{\theta}\|_{L^2(\tilde{\Omega}_t)}^2 \leq \frac{1}{c_0} \|\boldsymbol{\theta}\|_{\varepsilon, \tilde{\Omega}_t}^2$ die Abschatzung

$$\begin{aligned} \|\boldsymbol{\theta}_n^-\|_{L^2(\tilde{\Omega}_n)}^2 - \|\boldsymbol{\theta}_{n-1}^-\|_{L^2(\tilde{\Omega}_{n-1})}^2 + \|[\boldsymbol{\theta}]_{n-1}\|_{L^2(\tilde{\Omega}_{n-1})}^2 + \frac{\gamma c_0}{2} \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ \leq C \tau_n \tau_n^{2k+2} \left(\sup_{I_n} \|\partial_t^{k+2} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + A_n^2 \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + B_{k+1} \right). \quad (5.79) \end{aligned}$$

Wenn wir die Abschätzung (5.79) über n summieren, den positiven Sprungterm vernachlässigen und $\tau_n \leq \tau$ ausnutzen, ergibt sich

$$\begin{aligned} & \|(U - P_n[u])_N^- \|_{L^2(\tilde{\Omega}_N)}^2 + \frac{\gamma c_0}{2} \int_0^T \|U(t) - P_n[u](t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ & \leq C\tau^{2k+2} \left(\sup_I \|\partial_t^{k+2} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + A_n^2 \sup_I \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + B_{k+1} \right) + \|(U - P_n[u])_0^- \|_{L^2(\tilde{\Omega}_0)}^2. \end{aligned} \quad (5.80)$$

Bezüglich des Interpolationsfehlers haben wir

$$\int_{I_n} \|u(t) - P_n[u](t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \leq C\tau_n^{2k+3} \sup_{t \in I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2.$$

Das Aufsummieren über n und die Anwendung von $\tau_n \leq \tau$ liefern

$$\int_0^T \|u(t) - P_n[u](t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \leq C\tau^{2k+2} \sup_{t \in [0, T]} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2. \quad (5.81)$$

Die Kombination von (5.80) und (5.81) liefert die Behauptung. \square

Die Voraussetzung $C_1\tau_n A_n^2 \leq \frac{c_0}{2}$ für alle $n = 1, \dots, N$ lässt sich für kleine τ_n erfüllen und ist keine CFL-Bedingung.

Das folgende Lemma liefert eine Fehlerabschätzung in der Supremumnorm.

Lemma 5.17. *Seien U die Lösung des diskreten Problems (5.49) und u die Lösung von (5.4). Es gilt*

$$\begin{aligned} & \sup_I \|u(t) - U(t)\|_{L^2(\tilde{\Omega}_t)}^2 \leq \\ & C\tau^{2k+2} \left(\sup_I \|\partial_t^{k+2} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + (1 + A_n^2) \sup_I \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + B_{k+1} \right), \end{aligned}$$

wobei B_{k+1} in Lemma 5.16 definiert ist und C eine Konstante ist, die von A_n abhängt.

Beweis. Wir definieren die Hilfsfunktion

$$\tilde{\theta}(t) := P_n \left(\frac{\tau_n}{t - t_{n-1}} \theta(t) \right).$$

Es gilt laut Vlasak und Roos [39, Lemma 5]:

$$Q_n \left[(D_t \theta, 2\tilde{\theta})_{\tilde{\Omega}_n} \right] + (\theta_{n-1}^+, 2\tilde{\theta}_{n-1}^+)_{\tilde{\Omega}_{n-1}} = \|\theta_n^-\|_{L^2(\tilde{\Omega}_n)}^2 + \frac{1}{\tau_n} Q_n \left[\|\tilde{\theta}\|_{L^2(\tilde{\Omega}_n)}^2 \right]. \quad (5.82)$$

Desweiteren folgt aus Koerzivität von a_{nc} die Abschätzung

$$\begin{aligned} 0 \leq \gamma Q_n \left[\|\theta\|_{\varepsilon, \tilde{\Omega}_t}^2 \right] &= \gamma \frac{\tau_n}{2} \sum_{i=1}^{k+1} \hat{\omega}_i \|\theta(t_{n,i})\|_{\varepsilon, \tilde{\Omega}_t}^2 \\ &\leq \frac{\tau_n}{2} \sum_{i=1}^{k+1} \hat{\omega}_i a_{nc}((t_{n,i}), \theta(t_{n,i}), \theta(t_{n,i})) \\ &\leq \frac{\tau_n}{2} \sum_{i=1}^{k+1} \hat{\omega}_i \frac{2\tau_n}{t_{n,i} - t_{n-1}} a_{nc}((t_{n,i}), \theta(t_{n,i}), \theta(t_{n,i})) \\ &= \frac{\tau_n}{2} \sum_{i=1}^{k+1} \hat{\omega}_i a_{nc}((t_{n,i}), \theta(t_{n,i}), 2\tilde{\theta}(t_{n,i})) = Q_n \left[a_{nc}(t; \theta, 2\tilde{\theta}) \right]. \end{aligned}$$

Aus (5.55) folgt die Normäquivalenz für die diskrete Funktion $\boldsymbol{\theta}$:

$$\sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 \leq C \frac{1}{\tau_n} \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \leq C \frac{A_n}{\tau_n} Q_n \left[\|\boldsymbol{\theta}\|_{L^2(\tilde{\Omega}_n)}^2 \right] \leq C \frac{A_n}{\tau_n} Q_n \left[\|\tilde{\boldsymbol{\theta}}\|_{L^2(\tilde{\Omega}_n)}^2 \right].$$

Somit ergibt sich aus (5.82)

$$\begin{aligned} \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 &\leq CA_n \left(\frac{1}{\tau_n} Q_n \left[\|\tilde{\boldsymbol{\theta}}\|_{L^2(\tilde{\Omega}_n)}^2 \right] + \|\boldsymbol{\theta}_n^-\|_{L^2(\tilde{\Omega}_n)}^2 + Q_n \left[a(t; \boldsymbol{\theta}, 2\tilde{\boldsymbol{\theta}}) \right] \right) \\ &= CA_n \left(Q_n \left[(D_t \boldsymbol{\theta}, 2\tilde{\boldsymbol{\theta}})_{\tilde{\Omega}_n} \right] + Q_n \left[a(t; \boldsymbol{\theta}, 2\tilde{\boldsymbol{\theta}}) \right] + ([\boldsymbol{\theta}]_{n-1}, 2\tilde{\boldsymbol{\theta}}_{n-1}^+)_{\tilde{\Omega}_{n-1}} \right. \\ &\quad \left. + (\boldsymbol{\theta}_{n-1}^-, 2\tilde{\boldsymbol{\theta}}_{n-1}^+)_{\tilde{\Omega}_{n-1}} \right) \\ &\leq CA_n^2 \left(Q_n \left[(D_t \boldsymbol{\theta}, 2\tilde{\boldsymbol{\theta}})_{\tilde{\Omega}_t} \right] + Q_n \left[a(t; \boldsymbol{\theta}, 2\tilde{\boldsymbol{\theta}}) \right] + ([\boldsymbol{\theta}]_{n-1}, 2\tilde{\boldsymbol{\theta}}_{n-1}^+)_{\tilde{\Omega}_{n-1}} \right. \\ &\quad \left. + (\boldsymbol{\theta}_{n-1}^-, 2\tilde{\boldsymbol{\theta}}_{n-1}^+)_{\tilde{\Omega}_{n-1}} \right). \end{aligned}$$

Dabei folgt der letzte Schritt aus (5.10). Unter Ausnutzung der Fehlergleichung (5.61) erhalten wir

$$\begin{aligned} \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 &\leq CA_n^2 \left(Q_n \left[(D_t \boldsymbol{\delta}, 2\tilde{\boldsymbol{\theta}})_{\tilde{\Omega}_t} \right] + Q_n \left[a(t; \boldsymbol{\delta}, 2\tilde{\boldsymbol{\theta}}) \right] + ([\boldsymbol{\delta}]_{n-1}, 2\tilde{\boldsymbol{\theta}}_{n-1}^+)_{\tilde{\Omega}_{n-1}} \right. \\ &\quad \left. + (\boldsymbol{\theta}_{n-1}^-, 2\tilde{\boldsymbol{\theta}}_{n-1}^+)_{\tilde{\Omega}_{n-1}} \right). \end{aligned}$$

Nach der Anwendung von Lemma 5.15 haben wir

$$\begin{aligned} \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 &\leq C \tau_n^{k+2} A_n^2 \sup_{I_n} \|\partial_t^{k+2} u(t)\|_{L^2(\tilde{\Omega}_t)} \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)} \\ &\quad + CA_n^3 \tau_n^{k+2} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)} \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)} \\ &\quad + CA_n^2 B_{n,k+1} \tau_n^{k+2} \sum_{i=0}^k \tau_n^{k-i} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)} \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)} \\ &\quad + A_n^2 (\boldsymbol{\theta}_{n-1}^-, 2\tilde{\boldsymbol{\theta}}_{n-1}^+)_{\tilde{\Omega}_{n-1}}. \end{aligned}$$

Weiterhin liefern die Young-Ungleichung und $\|\boldsymbol{\theta}\|_{L^2(\tilde{\Omega}_t)}^2 \leq \frac{1}{c_0} \|\boldsymbol{\theta}\|_{\varepsilon, \tilde{\Omega}_t}^2$ Folgendes:

$$\begin{aligned} \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 &\leq C \tau_n A_n^4 \tau_n^{2k+2} \sup_{I_n} \|\partial_t^{k+2} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + C_1 \gamma c_0 \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ &\quad + CA_n^6 \tau_n \tau_n^{2k+2} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + C_2 \gamma c_0 \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ &\quad + CA_n^4 \tau_n \tau_n^{2k+2} B_{k+1} + C_3 \gamma c_0 \int_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ &\quad + C_4 \|\tilde{\boldsymbol{\theta}}_{n-1}^+\|_{L^2(\tilde{\Omega}_{n-1})}^2 + C_5 A_n^4 \|\boldsymbol{\theta}_{n-1}^-\|_{L^2(\tilde{\Omega}_{n-1})}^2. \end{aligned} \tag{5.83}$$

Desweiteren haben wir

$$\|\tilde{\boldsymbol{\theta}}_{n-1}^+\|_{L^2(\tilde{\Omega}_{n-1})}^2 \leq \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2.$$

Nun wählen wir die Konstanten $C_i, i = 1, \dots, 4$, mit $\sum_{i=1}^4 C_i = 1/2$. Daraufaufbauend erhalten wir aus (5.83) die Abschätzung

$$\begin{aligned} \sup_{I_n} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 &\leq C \tau_n \left(\tau_n^{2k+2} \sup_{I_n} \|\partial_t^{k+2} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + A_n^2 \tau_n^{2k+2} \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + B_{k+1} \right) \\ &\quad + C_5 A_n^4 \|\boldsymbol{\theta}_{n-1}^-\|_{L^2(\tilde{\Omega}_t)}^2. \end{aligned} \tag{5.84}$$

Weiterhin liefert das Aufsummieren der Abschätzung (5.79) für $n = 1, \dots, k$

$$\begin{aligned} \|\boldsymbol{\theta}_k^-\|_{L^2(\tilde{\Omega}_k)}^2 - \|\boldsymbol{\theta}_0^-\|_{L^2(\tilde{\Omega}_0)}^2 + \sum_{n=1}^k \|[\boldsymbol{\theta}]_{n-1}\|_{L^2(\tilde{\Omega}_k)}^2 + \frac{\gamma c_0}{2} \int_0^{t_k} \|\boldsymbol{\theta}(t)\|_{L^2(\tilde{\Omega}_t)}^2 dt \\ \leq C \sum_{n=1}^k \tau_n \tau_n^{2k+2} \left(\sup_{I_n} \|\partial_t^{k+2} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + A_n^2 \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + B_{k+1} \right). \end{aligned}$$

Wir nutzen $\|\boldsymbol{\theta}_0^-\|^2 = 0$ und $\tau_n \leq \tau$ aus, vernachlässigen die positiven Sprungterme und erhalten somit

$$\begin{aligned} \|\boldsymbol{\theta}_k^-\|_{L^2(\tilde{\Omega}_k)}^2 &\leq \|\boldsymbol{\theta}_k^-\|_{L^2(\tilde{\Omega}_k)}^2 + \frac{\gamma c_0}{2} \int_0^{t_k} \|\boldsymbol{\theta}(t)\|_{\varepsilon, \hat{\Omega}_t}^2 dt \\ &\leq C \sum_{n=1}^k \tau_n \tau_n^{2k+2} \left(\sup_{I_n} \|\partial_t^{k+2} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + A_n^2 \sup_{I_n} \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + B_{k+1} \right) \\ &\leq C \tau^{2k+2} \left(\sup_I \|\partial_t^{k+2} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + A_n^2 \sup_I \|\partial_t^{k+1} u(t)\|_{L^2(\tilde{\Omega}_t)}^2 + B_{k+1} \right). \quad (5.85) \end{aligned}$$

Die Aussage wird aus der Kombination von (5.84), (5.85) und des Interpolationsfehlers geliefert. \square

6 FE-Diskretisierung der Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten

In diesem Kapitel wird die Diskretisierung der Konvektions-Diffusions-Reaktions-Gleichungen in zeitabhängigen Gebieten $\Omega_t \subset \mathbb{R}^d$, $d = 2, 3$, mit vorgeschriebenem Rand $\Gamma_t := \partial\Omega_t$ im Rahmen der Finite-Elemente-Methode (FEM) durchgeführt. Das beschränkte Polygon ($d = 2$) bzw. Polyeder ($d = 3$) $\widehat{\Omega} \subset \mathbb{R}^d$ mit dem Rand $\widehat{\Gamma} := \partial\widehat{\Omega}$ wird als Referenzgebiet betrachtet. Das aktuelle Gebiet Ω_t wird durch eine Transformation (ALE-Abbildung) beschrieben. Die Diskretisierung dieser Transformation \mathcal{A}_t wird ein wichtiger Teil unserer Finite-Elemente-Formulierung sein. Für unsere Untersuchungen betrachten wir das Problem (5.1) und werden sämtliche Annahmen und Notationen aus Kapitel 5 übernehmen.

6.1 FE-Semidiskretisierung im ALE-Rahmen

6.1.1 FE-Semidiskretisierung auf dem Referenzgebiet

Wir bezeichnen mit $\widehat{\mathcal{T}}_h$ eine Triangulierung des Referenzgebietes $\widehat{\Omega}$, sodass $\widehat{\Omega} = \cup_{\widehat{K} \in \widehat{\mathcal{T}}_h} \widehat{K}$. Dabei sind \widehat{K} Dreiecke ($d = 2$) oder Tetraeder ($d = 3$). Sei $\mathcal{Y}_h(\widehat{\Omega}) \subset H^1(\widehat{\Omega})$ der Finite-Elemente-Ansatzraum auf $\widehat{\Omega}$ definiert durch

$$\mathcal{Y}_h(\widehat{\Omega}) := \left\{ \widehat{\psi}_h : \widehat{\Omega} \rightarrow \mathbb{R} : \widehat{\psi}_h \in C^0(\widehat{\Omega}), \widehat{\psi}_h|_{\widehat{K}} \in \mathbb{P}_l(\widehat{K}) \forall \widehat{K} \in \widehat{\mathcal{T}}_h \right\},$$

wobei $\mathbb{P}_l(\widehat{K})$ der Raum der Polynome vom Grad kleiner oder gleich l ist. Dieser ist auf dem Element \widehat{K} definiert. Sei \mathcal{N} die Menge der Freiheitsgraden des FE-Gitters und $\mathcal{N}_{int} \subset \mathcal{N}$ die Menge, die nur die inneren Freiheitsgraden enthält. Die Menge der Basisfunktionen auf dem Referenzgebiet $\widehat{\Omega}$ wird gegeben durch

$$\left\{ \widehat{\psi}_i : \widehat{\psi}_i \in \mathcal{Y}_h(\widehat{\Omega}), i \in \mathcal{N} \right\}.$$

Jede Funktion $\widehat{v}_h \in \mathcal{Y}_h(\widehat{\Omega})$ lässt sich als eine Linearkombination der nodalen FE-Basisfunktionen $\widehat{\psi}_i$ schreiben, d.h.

$$\widehat{v}_h(\mathbf{Y}) := \sum_{i \in \mathcal{N}} \widehat{v}_i \widehat{\psi}_i(\mathbf{Y}), \quad \mathbf{Y} \in \widehat{\Omega}, \quad \widehat{v}_i \in \mathbb{R}. \quad (6.1)$$

6.1.2 FE-Semidiskretisierung der ALE-Abbildung

In den meisten Fällen wird die ALE-Abbildung an sich mittels der FEM dargestellt und beispielsweise durch Lösen einer geeigneten Differentialgleichung berechnet, die die Bewegung des Randes beschreibt, siehe Abschnitt 2.1.3. Wie bereits erwähnt wird die Beschreibung des Randes Γ_t als bekannt vorausgesetzt.

Für alle $t \in I$ berichten wir die FE-Diskretisierung von \mathcal{A}_t mittels stückweiser Lagrange-Finite-Elemente, vgl. [19], die wir mit

$$\mathcal{A}_{h,t} : \widehat{\Omega} \rightarrow \Omega_t \quad (6.2)$$

bezeichnen. Somit lässt sich \mathbf{x} im Ort folgendermaßen diskretisieren:

$$\mathbf{x}(t, \mathbf{Y}) := \mathcal{A}_{h,t}(\mathbf{Y}) = \sum_{i \in \mathcal{N}} \mathbf{x}_i(t) \widehat{\psi}_i(\mathbf{Y}),$$

wobei $\mathbf{x}_i(t) := \mathcal{A}_{h,t}(\mathbf{Y}_i)$ die Position des i -ten Knotens zur Zeit t bezeichnet. Beachtenswert ist, dass die Wahl des Finite-Elemente-Raumes die Diskretisierung der verwendeten ALE-Abbildung beeinflusst. Mit K_t bezeichnen wir das Bild des Elements $\widehat{K} \in \widehat{\mathcal{T}}_h$, d.h., $K_t = \mathcal{A}_{h,t}(\widehat{K})$. Damit K_t ein Dreieck ($d = 2$) bzw. Tetraeder ($d = 3$) ist, reicht es aus, wenn $\mathcal{A}_{h,t}$ in $\mathbb{P}_1(\widehat{K})$ liegt.

Desweiteren setzen wir $\mathcal{T}_{h,t} := \{K_t = \mathcal{A}_{h,t}(\widehat{K}), \widehat{K} \in \widehat{\mathcal{T}}_h\}$ und $\overline{\Omega}_t = \cup_{K_t \in \mathcal{T}_{h,t}} \overline{K}_t$. Dann stellt $\mathcal{T}_{h,t}$ eine Triangulierung von Ω_t bestehend aus konformen Dreiecken bzw. Tetraeder dar.

6.1.3 FE-Räume und Semi-diskrete Formulierung

Wir betrachten nun den Zeit-Raum-Zylinder

$$\mathcal{S}_h := \{(t, \mathbf{x}) \in \mathbb{R}^{d+1} : \mathbf{x} \in \Omega_t, t \in I\}.$$

Die FE-Basisfunktionen auf der Triangulierung $\mathcal{T}_{h,t}$ von Ω_t lassen sich unter Anwendung der ALE-Abbildung wie folgt konstruieren:

$$\psi_i = \widehat{\psi}_i \circ \mathcal{A}_{h,t}^{-1}, \quad i \in \mathcal{N}.$$

Der FE-Raum auf \mathcal{S}_h kann mithilfe des diskreten Homöomorphismus (ALE-Abbildung) $\mathcal{A}_{h,t}$ durch

$$\mathcal{X}_h := \left\{ \psi_h : \mathcal{S}_h \rightarrow \mathbb{R} : \psi_h(t) \in \mathcal{X}_h(\Omega_t) \right\}$$

konstruiert werden, wobei

$$\mathcal{X}_h(\Omega_t) := \left\{ \psi_h = \widehat{\psi}_h \circ \mathcal{A}_{h,t}^{-1}, \quad \widehat{\psi}_h \in \mathcal{Y}_h(\widehat{\Omega}) \right\} \quad (6.3)$$

ist. Somit stellt die Menge $\{\psi_i, i \in \mathcal{N}\}$ bzw. $\{\psi_i, i \in \mathcal{N}_{int}\}$ eine Basis von \mathcal{X}_h bzw. von $\mathcal{X}_h^0 := \mathcal{X}_h \cap H_0^1(\Omega_t)$ dar.

Desweiteren lässt sich jede Funktion $v_h \in \mathcal{X}_h$ als eine Linearkombination der nodalen FE-Basisfunktionen ψ_i schreiben, d.h.

$$v_h(t, \mathbf{x}) := \sum_{i \in \mathcal{N}} v_i(t) \psi_i(t, \mathbf{x}), \quad \mathbf{x} \in \Omega_t, \quad (6.4)$$

wobei $v_h(t, \mathbf{x}) = \widehat{v}_h(t, \mathcal{A}_{h,t}^{-1}(\mathbf{x}))$.

Die Zeitableitung im diskreten ALE-Rahmen ist definiert durch

$$D_t v_h : \mathcal{S}_h \rightarrow \mathbb{R}^d, \quad D_t v_h(t, \mathbf{x}) = \partial_t \widehat{v}_h(t, \mathbf{Y}), \quad \mathbf{Y} = \mathcal{A}_{h,t}^{-1}(\mathbf{x}).$$

Die diskrete Gittergeschwindigkeit $\mathbf{w}_h(t, \mathbf{x}) = D_t \mathbf{x}(t, \mathbf{Y})$ definieren wir im FE-Raum durch

$$\mathbf{w}_h(t, \mathbf{x}) := \sum_{i \in \mathcal{N}} \mathbf{w}_i(t) \psi_i(t, \mathbf{x}).$$

Weiterhin erhalten wir, wie bereits im stetigen Fall, die Beziehung

$$D_t v_h = \partial_t v_h + \mathbf{w}_h \cdot \nabla_{\mathbf{x}} v_h.$$

Für die gesuchte Funktion $u(t, \mathbf{x})$ setzt man als Ansatzfunktion die örtlich diskrete, aber in der Zeit kontinuierliche Näherung $u_h : I \rightarrow \mathcal{X}_h^0$ als eine Linearkombination von nodalen FE-Basisfunktionen an, d.h.

$$u_h(t, \mathbf{x}) := \sum_{i \in \mathcal{N}_{int}} u_i(t) \psi_i(t, \mathbf{x}), \quad (6.5)$$

wobei die Koeffizienten $u_i, i \in \mathcal{N}$, zeitabhängig sind. Als Approximation der Anfangsbedingung u_0 wird $u_{h,0} = \tilde{P}_h u_0$ betrachtet. Dabei ist der Operator \tilde{P}_h in der Annahme 4.1 definiert. Der Einfachheit halber verwenden wir im Folgenden die Notation $v_h(t)$ anstatt $v_h(t, \mathbf{x})$.

Um die schwache Formulierung zu erhalten, multiplizieren wir die Differentialgleichung (5.1) mit einer zulässigen Testfunktion $v_h \in \mathcal{X}_h^0$, integrieren wie üblich über Ω_t und erhalten

$$(\partial_t u_h, v_h)_{\Omega_t} + (\mathbf{b}_h \cdot \nabla_{\mathbf{x}} u_h, v_h)_{\Omega_t} + \varepsilon (\nabla_{\mathbf{x}} u_h, \nabla_{\mathbf{x}} v_h)_{\Omega_t} + (c_h u_h, v_h)_{\Omega_t} = (f_h, v_h)_{\Omega_t} \quad (6.6)$$

für alle $v_h \in \mathcal{X}_h^0$. Dabei ist $(u_h, v_h)_{\Omega_t} := \int_{\Omega_t} u_h v_h \, d\mathbf{x}$.

6.1.4 Nichtkonservative ALE-Formulierung

Die nichtkonservative Variante der schwachen Formulierung folgt unmittelbar aus der variationellen Formulierung. Die Eulersche partielle Zeitableitung ∂_t wird dabei durch die ALE-Zeitableitung D_t ersetzt. Nachdem wir (2.2) in (6.6) eingesetzt haben, lässt sich das Problem wie folgt formulieren:

Finde $u_h \in \mathcal{X}_h^0$ mit $u_h(0) = u_{h,0}$, sodass

$$\begin{aligned} (D_t u_h, \psi_h)_{\Omega_t} + ((\mathbf{b}_h - \mathbf{w}_h) \cdot \nabla_{\mathbf{x}} u_h, \psi_h)_{\Omega_t} + \varepsilon (\nabla_{\mathbf{x}} u_h, \nabla_{\mathbf{x}} \psi_h)_{\Omega_t} \\ + (c_h u_h, \psi_h)_{\Omega_t} = (f_h, \psi_h)_{\Omega_t} \quad \forall \psi_h \in \mathcal{X}_h^0. \end{aligned} \quad (6.7)$$

6.1.5 Konservative ALE-Formulierung

Indem wir die verallgemeinerte Reynolds-Identität (2.6) in (6.7) einsetzen, konstruieren wir hier eine konservative Formulierung und erhalten dementsprechend die folgende konservative Variante:

Finde $u_h \in \mathcal{X}_h^0$ mit $u_h(0) = u_{h,0}$, sodass

$$\begin{aligned} \frac{\partial}{\partial t} (u_h, \psi_h)_{\Omega_t} + ((\mathbf{b}_h - \mathbf{w}_h) \cdot \nabla_{\mathbf{x}} u_h, \psi_h)_{\Omega_t} - ((\nabla_{\mathbf{x}} \cdot \mathbf{w}_h) u_h, \psi_h)_{\Omega_t} + \varepsilon (\nabla_{\mathbf{x}} u_h, \nabla_{\mathbf{x}} \psi_h)_{\Omega_t} \\ - (u_h, D_t \psi_h)_{\Omega_t} + (c_h u_h, \psi_h)_{\Omega_t} = (f_h, \psi_h)_{\Omega_t} \quad \forall \psi_h \in \mathcal{X}_h^0. \end{aligned} \quad (6.8)$$

Man beachte, dass hier alle von Kapitel 5 übernommenen Notationen ortsdiskret sind.

6.2 FE-Approximierung

6.2.1 Algebraische Formulierung der nichtkonservativen Form

Aus (2.2) folgt

$$(D_t u_h, \psi_h)_{\Omega_t} = (\partial_t u_h, \psi_h)_{\Omega_t} + (\mathbf{w}_h \cdot \nabla_{\mathbf{x}} u_h, \psi_h)_{\Omega_t}.$$

Wir setzen den Ansatz (6.5) in (6.7) ein und verwenden ψ_j aus der Basis von \mathcal{X}_h^0 als Testfunktionen. Für den ersten Term erhalten wir

$$\begin{aligned} (D_t u_h, \psi_j)_{\Omega_t} &= \sum_{i \in \mathcal{N}} \int_{\Omega_t} D_t (u_i \psi_i) \psi_j \, d\mathbf{x} \\ &= \sum_{i \in \mathcal{N}} \int_{\Omega_t} D_t u_i \psi_i \psi_j \, d\mathbf{x} + \sum_{i \in \mathcal{N}} \int_{\Omega_t} u_i D_t \psi_i \psi_j \, d\mathbf{x} \\ &= \sum_{i \in \mathcal{N}} \partial_t u_i \int_{\Omega_t} \psi_i \psi_j \, d\mathbf{x} + \sum_{i \in \mathcal{N}} u_i \int_{\Omega_t} D_t \psi_i \psi_j \, d\mathbf{x}. \end{aligned} \quad (6.9)$$

Mit der üblichen Schreibweise für die Masse- sowie die Steifigkeitsmatrix wird die Gleichung (6.7) in ein Anfangswertproblem für ein System linearer Gleichungen überführt. Dann lautet die nichtkonservative Finite-Elemente-Semidiskretisierung von (5.1):

$$M(t) \frac{dU(t)}{dt} + (H(t) - A(t, \mathbf{w}_h(t)))U(t) = F(t) \quad (6.10)$$

mit dem Startwert $U(0) = u_{h,0}$ und der rechten Seite F . Dabei sind

$$M(t) := \left\{ \int_{\Omega_t} \psi_i \psi_j \, d\mathbf{x} \right\}_{i,j \in \mathcal{N}_{int}},$$

$$H(t) :=$$

$$\left\{ \varepsilon \int_{\Omega_t} \nabla_{\mathbf{x}} \psi_i \nabla_{\mathbf{x}} \psi_j \, d\mathbf{x} + \int_{\Omega_t} \mathbf{b}_h \nabla_{\mathbf{x}} \psi_i \psi_j \, d\mathbf{x} + \int_{\Omega_t} c_h \psi_i \psi_j \, d\mathbf{x} + \int_{\Omega_t} D_t \psi_i \psi_j \, d\mathbf{x} \right\}_{i,j \in \mathcal{N}_{int}},$$

$$A(t, \mathbf{w}_h) := \left\{ \int_{\Omega_t} \mathbf{w}_h \cdot \nabla_{\mathbf{x}} \psi_i \psi_j \, d\mathbf{x} \right\}_{i,j \in \mathcal{N}_{int}},$$

$$F(t) := \left\{ \int_{\Omega_t} f_h \psi_j \, d\mathbf{x} \right\}_{i,j \in \mathcal{N}_{int}}$$

und $U := \{u_i\}, i \in \mathcal{N}_{int}$.

6.2.2 Algebraische Formulierung der konservativen Form

Analog zum nichtkonservativen Fall lässt sich (6.8) in das folgende algebraische Schema umschreiben:

$$\frac{d}{dt} (M(t)U(t)) + (G(t) - B(t, \mathbf{w}_h))U(t) = F(t) \quad (6.11)$$

mit dem Startwert $U(0) = u_{h,0}$. Dabei sind

$$B(t, \mathbf{w}_h) := \left\{ \int_{\Omega_t} \mathbf{w}_h \cdot \nabla_{\mathbf{x}} \psi_i \psi_j \, d\mathbf{x} - \int_{\Omega_t} (\nabla_{\mathbf{x}} \cdot \mathbf{w}_h) \psi_i \psi_j \, d\mathbf{x} \right\}_{i,j \in \mathcal{N}_{int}}$$

und

$$G(t) := \left\{ \varepsilon \int_{\Omega_t} \nabla_{\mathbf{x}} \psi_i \nabla_{\mathbf{x}} \psi_j \, d\mathbf{x} + \int_{\Omega_t} \mathbf{b}_h \nabla_{\mathbf{x}} \psi_i \psi_j \, d\mathbf{x} + \int_{\Omega_t} c_h \psi_i \psi_j \, d\mathbf{x} + \int_{\Omega_t} D_t \psi_i \psi_j \, d\mathbf{x} - \int_{\Omega_t} \psi_i D_t \psi_j \, d\mathbf{x} \right\}_{i,j \in \mathcal{N}_{int}}.$$

In Kapitel 3 haben wir variationelle Zeitdiskretisierungen höherer Ordnung entwickelt. Diese wollen wir nun auf unsere semidiskrete Konvektions-Diffusions-Reaktions-Gleichung anwenden.

6.3 Implementierung der konservativen Form

Wie bereits dargestellt formulieren wir das aus der Finite-Elemente-Methode erhaltene semidiskrete nichtkonservative Problem auf I_n folgendermaßen um:

$$\partial_t V(t) = \tilde{F}(t, V(t)). \quad (6.12)$$

Dabei sind $V(t) = M(t)U(t)$ und die rechte Seite $\tilde{F}(t, V) = F(t) - D(t)M^{-1}(t)V(t)$ mit $D(t) := G(t) - B(t, \mathbf{w}_h(t))$. Als Startwert betrachten wir die Approximation $v_{h,0} = M(0)u_{h,0}$.

6.3.1 Konservatives cGP-Problem

Das cGP-Verfahren angewendet auf (6.12) auf I_n lautet:

Finde die Koeffizienten $\mathbf{U}_n^j \in \mathbb{R}^{\mathcal{N}_{int}}$, $j = 1, \dots, k$, sodass

$$\sum_{j=0}^k \alpha_{i,j} \mathbf{V}_n^j = \frac{\tau_n}{2} \left\{ \tilde{\mathbf{F}}(t_{n,i}, \mathbf{V}(t_{n,i})) + \beta_i \tilde{\mathbf{F}}(t_{n,0}, \mathbf{V}(t_{n,0})) \right\}$$

für alle $i = 1, \dots, k$. Die Koeffizienten $\alpha_{i,j}$ bzw. β_i definieren sich wie auch in (3.12) durch

$$\alpha_{i,j} := \hat{\phi}'_j(\hat{t}_i) + \beta_i \hat{\phi}'_j(-1), \quad \beta_i := \hat{\omega}_0 \hat{\psi}_i(-1). \quad (6.13)$$

Dabei wurde der Startwert $\mathbf{V}(0) = v_{h,0}$ gewählt.

Die Lösung für das ursprüngliche Problem ergibt sich schließlich aus

$$\mathbf{U}_n^j = \mathbf{M}(t_{n,j})^{-1} \mathbf{V}_n^j, \quad (6.14)$$

somit lautet das „**konservative cGP-Problem**“:

Finde die Koeffizienten $\mathbf{U}_n^j \in \mathbb{R}^{\mathcal{N}_{int}}$, $j = 1, \dots, k$, sodass

$$\sum_{j=0}^k \alpha_{i,j} \mathbf{M}(t_{n,j}) \mathbf{U}_n^j = \frac{\tau_n}{2} \left\{ \mathbf{F}(t_{n,i}) - \mathbf{D}(t_{n,i}) \mathbf{U}_n^i + \beta_i \mathbf{F}(t_{n,0}) - \beta_i \mathbf{D}(t_{n,0}) \mathbf{U}_n^0 \right\}$$

für alle $i = 1, \dots, k$.

Im Folgenden präsentieren wir die cGP(k)-Verfahren für die Fälle $k = 1$ und $k = 2$.

6.3.2 cGP(1)-Verfahren

Aus der Anwendung der 2-punktigen Gauß–Lobatto-Formel mit den Stützstellen $t_{n,0} = t_{n-1}$, $t_{n,1} = t_n$ und den Referenzgewichten $\hat{\omega}_0 = \hat{\omega}_1 = 1$ ergibt sich die bekannte Trapezregel

$$\left[\mathbf{M}(t_n) + \frac{\tau_n}{2} \mathbf{D}(t_n) \right] \mathbf{U}_n^1 - \left[\mathbf{M}(t_{n-1}) - \frac{\tau_n}{2} \mathbf{D}(t_{n-1}) \right] \mathbf{U}_n^0 = \frac{\tau_n}{2} \left\{ \mathbf{F}(t_n) + \mathbf{F}(t_{n-1}) \right\}.$$

Die Lösung wird hier durch $u_\tau(t_n) = \mathbf{U}_n^1$ geliefert.

6.3.3 cGP(2)-Verfahren

Die Verwendung der 3-punktigen Gauß–Lobatto-Formel mit den Stützstellen $t_{n,0} = t_{n-1}$, $t_{n,1} = (t_{n-1} + t_n)/2$, $t_{n,2} = t_n$ und den Referenzgewichten $\hat{\omega}_0 = \hat{\omega}_2 = 1/3$, $\hat{\omega}_1 = 4/3$ liefert die bekannte Simpsonregel

$$\begin{aligned} & \left[\mathbf{M}(t_{n,1}) + \frac{\tau_n}{2} \mathbf{D}(t_{n,1}) \right] \mathbf{U}_n^1 + \frac{1}{4} \mathbf{M}(t_{n,2}) \mathbf{U}_n^2 \\ & \quad = \left[\frac{5}{4} \mathbf{M}(t_{n,0}) - \frac{\tau_n}{4} \mathbf{D}(t_{n,0}) \right] \mathbf{U}_n^0 + \frac{\tau_n}{2} \left\{ \mathbf{F}(t_{n,1}) + \frac{1}{2} \mathbf{F}(t_{n,0}) \right\}, \\ & - 4 \mathbf{M}(t_{n,1}) \mathbf{U}_n^1 + \left[2 \mathbf{M}(t_{n,2}) + \frac{\tau_n}{2} \mathbf{D}(t_{n,2}) \right] \mathbf{U}_n^2 \\ & \quad = \left[-2 \mathbf{M}(t_{n,0}) + \frac{\tau_n}{2} \mathbf{D}(t_{n,0}) \right] \mathbf{U}_n^0 + \frac{\tau_n}{2} \left\{ \mathbf{F}(t_{n,2}) - \mathbf{F}(t_{n,0}) \right\}, \end{aligned}$$

wobei $U(t_{n,j}) = \mathbf{U}_n^j$, $j = 1, 2$.

6.3.4 Konservatives dG-Problem

Mit der Approximation $\mathbf{V}(0) = v_{h,0}$ an den Anfangswert v_0 und der Anwendung des Wertes $\mathbf{V}(t_{n-1}^+)$ aus dem vorherigen Intervall I_{n-1} lässt sich das dG-Problem, angewendet auf (6.12) für $n = 1, \dots, N$, folgendermaßen formulieren:

Finde die Koeffizienten $\mathbf{V}_n^j \in \mathbb{R}^{\mathcal{N}_{int}}$, $j = 1, \dots, k+1$, sodass

$$\sum_{j=1}^{k+1} \alpha_{i,j} \mathbf{V}_n^j = \frac{\tau_n}{2} \tilde{\mathbf{F}}(t_{n,i}, \mathbf{V}(t_{n,i})) + d_i \mathbf{V}_n^0$$

für alle $i = 1, \dots, k$, wobei $\mathbf{V}_{n-1}^- := \mathbf{V}|_{I_{n-1}}(t_{n-1})$ für alle $n \geq 2$ und $\mathbf{V}_{n-1}^- := v_{h,0}$ für $n = 1$ der Startwert ist. Die Koeffizienten $\alpha_{i,j}$ bzw. β_i definieren sich wie in (3.19) durch

$$\alpha_{i,j} := \hat{\phi}'_j(t_j) + c_j d_i, \quad c_j := \hat{\phi}'_j(-1), \quad d_i := (\hat{\omega}_i)^{-1} \hat{\phi}'_i(-1). \quad (6.15)$$

Mithilfe von (6.14) ergibt sich das „**konservative dG-Problem**“ :

Finde die Koeffizienten $\mathbf{U}_n^j \in \mathbb{R}^{\mathcal{N}_{int}}$, $j = 1, \dots, k+1$, sodass

$$\sum_{j=1}^{k+1} \alpha_{i,j} \mathbf{M}(t_{n,j}) \mathbf{U}_n^j = \frac{\tau_n}{2} \mathbf{F}(t_{n,i}) - \frac{\tau_n}{2} \mathbf{D}(t_{n,i}) \mathbf{U}_n^i + d_i \mathbf{M}(t_{n,0}) \mathbf{U}_n^0$$

für alle $i = 1, \dots, k$. Dabei ist $\mathbf{U}_{n-1}^- := \mathbf{U}|_{I_{n-1}}(t_{n-1})$ für alle $n \geq 2$ und $\mathbf{U}_{n-1}^- := u_{h,0}$ für $n = 1$ der Startwert.

Nun werden wir die Schemata für die numerisch integrierten dG(0)- und dG(1)-Verfahren vorstellen.

6.3.5 dG(0)-Verfahren

Wir verwenden hier die 1-punktige Gauß–Radau-Formel mit der Stützstelle $t_{n,1} = t_n$ und dem Referenzgewicht $\hat{\omega}_1 = 2$ und erhalten das bekannte implizite Euler-Verfahren

$$[\mathbf{M}(t_n) + \tau_n \mathbf{D}(t_{n,1})] \mathbf{U}_n^1 - \mathbf{M}(t_{n-1}) \mathbf{U}_n^0 = \tau_n \mathbf{F}(t_n),$$

wobei die diskrete Lösung durch $\mathbf{U}_n^1 = U(t_n) \in \mathcal{V}$ geliefert wird.

6.3.6 dG(1)-Verfahren

Im dG(1)-Verfahren wird hier die 2-punktige Gauß–Radau-Formel mit den Stützstellen $t_{n,1} = (t_{n-1} + t_n)/3$, $t_{n,2} = t_n$ und den Referenzgewichten $\hat{\omega}_1 = 3/2$, $\hat{\omega}_2 = 1/2$ angewendet. Es ergibt sich das folgende (2×2) -Block-System:

$$\begin{aligned} \left[\frac{3}{4} \mathbf{M}(t_{n,1}) + \frac{\tau_n}{2} \mathbf{D}(t_{n,1}) \right] \mathbf{U}_n^1 + \frac{1}{4} \mathbf{M}(t_{n,2}) \mathbf{U}_n^2 &= \mathbf{M}(t_{n,0}) \mathbf{U}_n^0 + \frac{\tau_n}{2} \mathbf{F}(t_{n,1}), \\ -\frac{9}{4} \mathbf{M}(t_{n,1}) \mathbf{U}_n^1 + \left[\frac{5}{4} \mathbf{M}(t_{n,2}) + \frac{\tau_n}{2} \mathbf{D}(t_{n,2}) \right] \mathbf{U}_n^2 &= -\mathbf{M}(t_{n,0}) \mathbf{U}_n^0 + \frac{\tau_n}{2} \mathbf{F}(t_{n,2}), \end{aligned}$$

wobei $U(t_{n,j}) = \mathbf{U}_n^j$, $j = 1, 2$.

6.4 Implementierung der nichtkonservativen Form

Wie wir in Unterabschnitt 6.2 gesehen haben, lässt sich das aus der Finite-Elemente-Methode erhaltene semidiskrete Problem auf I_n wie folgt formulieren:

$$\partial_t V(t) = \tilde{F}(t, V(t)), \quad (6.16)$$

wobei $V(t) = M(t)U(t)$ und die rechte Seite $\tilde{F}(t, V) = F(t) - C(t)M^{-1}(t)V(t)$ mit $C(t) := H(t) - A(t, \mathbf{w}_h(t)) - M'(t)$. Als Startwert wird die Approximation $v_{h,0} = M(0)u_{h,0}$ betrachtet.

6.4.1 Nichtkonservatives cGP-Problem

Unter Verwendung von (6.16) lässt sich das „**nichtkonservative cGP-Problem**“, wie bereits bei dem konservativen Fall, folgendermaßen herleiten:

Finde die Koeffizienten $\mathbf{U}_n^j \in \mathbb{R}^{\mathcal{N}_{int}}$, $j = 1, \dots, k$, sodass

$$\sum_{j=0}^k \alpha_{i,j} \mathbf{M}(t_{n,j}) \mathbf{U}_n^j = \frac{\tau_n}{2} \{ \mathbf{F}(t_{n,i}) - \mathbf{C}(t_{n,i}) \mathbf{U}_n^i + \beta_i \mathbf{F}(t_{n,0}) - \beta_i \mathbf{C}(t_{n,0}) \mathbf{U}_n^0 \}$$

für alle $i = 1, \dots, k$. Dabei sind \mathbf{U}_n^0 , $\alpha_{i,j}$ und β_i wie im konservativen Fall.

Für eine bessere Verständlichkeit der erhaltenen speziellen Formulierung (3.13) präsentieren wir nun die cGP(k)-Verfahren für die Fälle $k = 1$ und $k = 2$.

6.4.2 cGP(1)-Verfahren

Die 2-punktige Gauß–Lobatto-Formel mit den Stützstellen $t_{n,0} = t_{n-1}$, $t_{n,1} = t_n$ und den Referenzgewichten $\hat{\omega}_0 = \hat{\omega}_1 = 1$ liefert die Trapezregel

$$[\mathbf{M}(t_n) + \frac{\tau_n}{2} \mathbf{C}(t_n)] \mathbf{U}_n^1 - [\mathbf{M}(t_{n-1}) - \frac{\tau_n}{2} \mathbf{C}(t_{n-1})] \mathbf{U}_n^0 = \frac{\tau_n}{2} \{ \mathbf{F}(t_n) + \mathbf{F}(t_{n-1}) \},$$

wobei sich die Lösung durch $u_\tau(t_n) = \mathbf{U}_n^1$ ergibt.

6.4.3 cGP(2)-Verfahren

Aus der 3-punktigen Gauß–Lobatto-Formel mit den Stützstellen $t_{n,0} = t_{n-1}$, $t_{n,1} = (t_{n-1} + t_n)/2$, $t_{n,2} = t_n$ und den Referenzgewichten $\hat{\omega}_0 = \hat{\omega}_2 = 1/3$, $\hat{\omega}_1 = 4/3$ erhalten wir die Simpsonregel

$$\begin{aligned} & [\mathbf{M}(t_{n,1}) + \frac{\tau_n}{2} \mathbf{C}(t_{n,1})] \mathbf{U}_n^1 + \frac{1}{4} \mathbf{M}(t_{n,2}) \mathbf{U}_n^2 \\ & \quad = [\frac{5}{4} \mathbf{M}(t_{n,0}) - \frac{\tau_n}{4} \mathbf{C}(t_{n,0})] \mathbf{U}_n^0 + \frac{\tau_n}{2} \{ \mathbf{F}(t_{n,1}) + \frac{1}{2} \mathbf{F}(t_{n,0}) \}, \\ & - 4 \mathbf{M}(t_{n,1}) \mathbf{U}_n^1 + [2 \mathbf{M}(t_{n,2}) + \frac{\tau_n}{2} \mathbf{C}(t_{n,2})] \mathbf{U}_n^2 \\ & \quad = [-2 \mathbf{M}(t_{n,0}) + \frac{\tau_n}{2} \mathbf{C}(t_{n,0})] \mathbf{U}_n^0 + \frac{\tau_n}{2} \{ \mathbf{F}(t_{n,2}) - \mathbf{F}(t_{n,0}) \}, \end{aligned}$$

wobei $U(t_{n,j}) = \mathbf{U}_n^j$, $j = 1, 2$.

6.4.4 Nichtkonservatives dG-Problem

Das „**nichtkonservative dG-Problem**“ lässt sich wie auch bei dem konservativen Fall, allerdings unter Verwendung von (6.16), herleiten.

Finde die Koeffizienten $\mathbf{U}_n^j \in \mathbb{R}^{\mathcal{N}_{int}}$, $j = 1, \dots, k + 1$, sodass

$$\sum_{j=1}^{k+1} \alpha_{i,j} \mathbf{M}(t_{n,j}) \mathbf{U}_n^j + \frac{\tau_n}{2} \mathbf{C}(t_{n,i}) \mathbf{U}_n^i = \frac{\tau_n}{2} \mathbf{F}(t_{n,i}) + d_i \mathbf{M}(t_{n,0}) \mathbf{U}_n^0$$

für alle $i = 1, \dots, k + 1$. Dabei sind \mathbf{U}_n^0 , $\alpha_{i,j}$ und d_i wie im konservativen Fall.

Im Folgenden werden wir die Schemata für die numerisch integrierten dG(0)- und dG(1)-Verfahren präsentieren.

6.4.5 dG(0)-Verfahren

Aus der Anwendung der 1-punktigen Gauß–Radau-Formel mit der Stützstelle $t_{n,1} = t_n$ und dem Referenzgewicht $\hat{\omega}_1 = 2$ ergibt sich das implizite Euler-Verfahren

$$[\mathbf{M}(t_n) + \tau_n \mathbf{C}(t_n)] \mathbf{U}_n^1 - \mathbf{M}(t_{n-1}) \mathbf{U}_n^0 = \tau_n \mathbf{F}(t_n),$$

wobei wir die diskrete Lösung durch $\mathbf{U}_n^1 = U(t_n) \in \mathcal{V}$ erhalten.

6.4.6 dG(1)-Verfahren

Bei dem dG(1)-Verfahren verwenden wir die 2-punktige Gauß–Radau-Formel mit den Stützstellen $t_{n,1} = (t_{n-1} + t_n)/3$, $t_{n,2} = t_n$ und den Referenzgewichten $\hat{\omega}_1 = 3/2$, $\hat{\omega}_2 = 1/2$. Schließlich erhalten wir das (2×2) -Block-System:

$$\begin{aligned} \left[\frac{3}{4}\mathbf{M}(t_{n,1}) + \frac{\tau_n}{2}\mathbf{C}(t_{n,1}) \right] \mathbf{U}_n^1 + \frac{1}{4}\mathbf{M}(t_{n,2})\mathbf{U}_n^2 &= \mathbf{M}(t_{n,0})\mathbf{U}_n^0 + \frac{\tau_n}{2}\mathbf{F}(t_{n,1}), \\ -\frac{9}{4}\mathbf{M}(t_{n,1})\mathbf{U}_n^1 + \left[\frac{5}{4}\mathbf{M}(t_{n,2}) + \frac{\tau_n}{2}\mathbf{C}(t_{n,2}) \right] \mathbf{U}_n^2 &= -\mathbf{M}(t_{n,0})\mathbf{U}_n^0 + \frac{\tau_n}{2}\mathbf{F}(t_{n,2}), \end{aligned}$$

wobei $U(t_{n,j}) = \mathbf{U}_n^j$, $j = 1, 2$.

6.5 Numerische Ergebnisse

Dieser Abschnitt dient zur Verifizierung der implementierten Algorithmen für cGP- sowie dG-Verfahren und zur Veranschaulichung der analytischen Ergebnisse bezüglich des Fehlerverhaltens sowie der Konvergenzordnungen der betrachteten dG-Verfahren. Dabei wird die in Kapitel 5 eingeführte diskrete Gittergeschwindigkeit betrachtet und numerisch untersucht. Im Anschluss daran wird auch numerisch überprüft, ob sich die erhaltenen Resultate auf cGP-Verfahren verallgemeinern lassen. Für die Implementierung der numerischen Ergebnisse dient eine auf zeitabhängige Gebiete verallgemeinerte Variante von MoonNMD [26].

Beispiel 3:

Um einen Vergleich unserer Ergebnisse mit denjenigen von Bonito, Kyza und Nocketto in [6] zu ermöglichen, betrachten wir dasselbe Problem wie in [6]:

$$\begin{aligned} \frac{\partial u}{\partial t} - \Delta_x u &= f, \quad x \in \Omega_t, t \in I, \\ u &= u_D, \quad x \in \partial\Omega_t, t \in I, \\ u &= u_0, \quad x \in \Omega_0, t = 0, \end{aligned}$$

wobei $I := [0, 0.99]$ und $\Omega_0 := (-1, 1) \times (-1, 1)$ sind. Wir betrachten Ω_0 als Referenzgebiet $\hat{\Omega}$. Die Entwicklung des Gebietes Ω_t wird durch die ALE-Abbildung

$$\mathcal{A}_t : y \rightarrow y \left(1 + \frac{1}{2} T_{11}(t) \right)$$

beschrieben, wobei $T_n(t) := \cos(n \arccos(t))$ das n -te Tschebycheff-Polynom erster Art ist. Das Gebiet Ω_t ergibt sich dann mithilfe der ALE-Abbildung durch $\Omega_t = \mathcal{A}_t(\Omega_0)$, $t \in [0, 0.99]$. Weiterhin sei

$$u(t, x) := \exp(x_1 t) \sin(x_2 t), \quad x := (x_1, x_2) \in \Omega_t,$$

die exakte Lösung des Problems. Die rechte Seite f sowie die Randbedingung u_D und Anfangsbedingung u_0 lassen sich aus der betrachteten exakten Lösung ableiten.

In diesem Beispiel sind wir an der Konvergenzordnung in der Zeit interessiert und betrachten n äquidistante Zeitschritte. Die Ortsdiskretisierung wurde so gewählt, dass sie den Fehler in der Zeit nicht beeinflusst. Für die Ortsdiskretisierung haben wir die in (4.115) definierten transformierten Finite-Elemente-Räumen verwendet. Wir haben dazu die LPS-Ortsdiskretisierung mit $\mathcal{V}_h = \mathbb{Q}_1^{\text{bl}}$, $\mathcal{D}_h(K) = \mathbb{P}_0(K)$ gewählt. Wir haben als Stabilitätsparameter $\beta_K = 0.1h_K$ verwendet und die exakte Gittergeschwindigkeit. Für die Triangulierung des Referenzgebietes $\hat{\Omega}$ wurde ein uniformes Gitter bestehend aus 2×2 Quadraten verwendet. An den Quadraturpunkten der Zeintegration wird die Randposition in das Innere des Gebietes fortgesetzt, indem das Problem 2.4 gelöst wird.

Die Fehlerberechnungen werden bezüglich der in (4.117) definierten $L^2(L^2)$ -Norm, $H^1(L^2)$ -Norm und diskreten $\ell^\infty(L^2)$ -Norm in der Zeit durchgeführt. Als Zeitdiskretisierungen höherer Ordnung werden cGP(2), cGP(3), dG(1) und dG(2) angewendet.

Die Fehler der dG(1) und dG(2) in Abhängigkeit der Anzahl der Zeitschritte N werden in den Abbildungen 6.1 und 6.2 gezeigt. Diese ermöglichen einen direkten Vergleich mit den Ergebnissen aus [8, Fig. 4]. Hier kann man sehen, dass die dG(k)-Verfahren bezüglich der $\ell^\infty(L^2)$ -Norm sogar die Super-Konvergenzordnung $2k + 1$ haben, während in [8] nur die Konvergenzordnung $k + 1$ erreicht wird. Ein weiterer Vergleich wird zwischen den konservativen und den nichtkonservativen dG-Verfahren vorgestellt. Dabei lässt sich feststellen, dass die beiden Verfahren in der $\ell^\infty(L^2)$ -Norm dieselbe Konvergenzordnung aufweisen.

Die Abbildungen 6.3 und 6.4 bieten einen Überblick der Fehler und Konvergenzordnungen von dG(1) und dG(2) in der $L^2(L^2)$ - und $H^1(L^2)$ -Norm. Es lässt sich feststellen, dass die in Lemma 5.17 gezeigte Fehlerabschätzung auch numerisch bestätigt wurde.

Beispiel 4:

Wir betrachten nun ein weiteres Beispiel, wobei die Bedingung (5.2) erfüllt wird:

$$\begin{aligned} \frac{\partial u}{\partial t} - \varepsilon \Delta_x u + \mathbf{b} \cdot \nabla_{\mathbf{x}} u + cu &= f, & x \in \Omega_t, t \in I, \\ u &= u_D, & x \in \partial\Omega_t, t \in I, \\ u &= u_0, & x \in \Omega_0, t = 0, \end{aligned}$$

wobei $I := [0, 0.99]$ und $\Omega_0 := (-1, 1) \times (-1, 1)$. Auch in diesem Beispiel betrachten wir Ω_0 als Referenzgebiet. Wir wählen hier die Diffusionskonstante $\varepsilon = 10^{-8}$. Für die Konvektion nehmen wir $\mathbf{b} = (x, 0)^T$ und als Reaktionsterm $c = 1$. Analog zu dem dritten Beispiel ist auch hier die ALE-Abbildung gegeben durch

$$\mathcal{A}_t : y \rightarrow y \left(1 + \frac{1}{2} T_{11}(t) \right),$$

wobei $T_n(t) := \cos(n \arccos(t))$ das n -te Tschebycheff-Polynom erster Art ist. Das Gebiet Ω_t ergibt sich dann mithilfe der ALE-Abbildung durch $\Omega_t = \mathcal{A}_t(\Omega_0)$, $t \in [0, 0.99]$. Die rechte Seite f wird so gewählt, dass

$$u(t, x) := \exp(x_1 t) \sin(x_2 t), \quad x := (x_1, x_2) \in \Omega_t,$$

die exakte Lösung des Problems ist. Die Randbedingung u_D und die Anfangsbedingung u_0 ergeben sich aus der exakten Lösung.

Auch in diesem Beispiel interessieren wir uns für die Konvergenzordnung in der Zeit und betrachten n äquidistante Zeitschritte. Daher wurde die Ortsdiskretisierung so gewählt, dass sie den Fehler in der Zeit nicht beeinflusst. Für die Ortsdiskretisierung haben wir wie im dritten Beispiel die in (4.115) definierten transformierten Finite-Elemente-Räumen verwendet und die LPS-Ortsdiskretisierung mit $\mathcal{V}_h = \mathbb{Q}_1^{\text{bl}}$, $\mathcal{D}_h(K) = \mathbb{P}_0(K)$ gewählt. Wir haben auch hier als Stabilitätsparameter $\beta_K = 0.1h_K$ verwendet. Es wurde auch in diesem Beispiel ein uniformes Gitter bestehend aus 2×2 Quadraten für die Triangulierung des Referenzgebietes $\hat{\Omega}$ verwendet. Die Randposition wird an den diskreten Zeitpunkten in das Innere des Gebietes fortgesetzt, indem das Problem 2.4 gelöst wird.

Die Fehlerberechnungen werden in Bezug auf die in (4.117) definierten $L^2(L^2)$ -Norm, $H^1(L^2)$ -Norm und diskreten $\ell^\infty(L^2)$ -Norm in der Zeit durchgeführt. Hier werden cGP(2), cGP(3), dG(1) und dG(2) als Zeitdiskretisierungen höherer Ordnung angewendet.

Wir illustrieren in den Abbildungen 6.5–6.14 die Fehler $U - u$ und Konvergenzordnungen der dG(k)- und cGP($k+1$)-Verfahren, $k = 1, 2$, im konservativen wie auch im nichtkonservativen Fall in der $L^2(L^2)$ -Norm, der $H^1(L^2)$ -Norm und der $\ell^\infty(L^2)$ -Norm. Ein direkter Vergleich der erhaltenen Ergebnisse zeigt, dass die konservative und die nichtkonservative Formulierung dieselben Konvergenzordnungen besitzen, während sich die Fehler beider Varianten nur um ein Vielfaches unterscheiden.

Die Abbildungen 6.5 und 6.8 bieten einen Überblick der Fehler und Konvergenzordnungen der dG(1) und der dG(2) in der $L^2(L^2)$ -, der $H^1(L^2)$ - und der $\ell^\infty(L^2)$ -Norm, wobei eine exakte Gittergeschwindigkeit verwendet wurde. Daraus kann man entnehmen, dass die in Lemma 5.17 gezeigte Fehlerabschätzung auch numerisch bestätigt wurde.

Nun werden wir sowohl für die konservativen als auch für die nichtkonservativen dG-Verfahren numerisch zeigen, inwiefern die verwendete Gittergeschwindigkeit \mathbf{w} die numerische Lösung bzw. den Fehler beeinflussen kann. Dabei haben wir die exakte Gittergeschwindigkeit, die in (5.12) definierte approximative Gittergeschwindigkeit und die in (5.13) definierte approximative Gittergeschwindigkeit mit Korrektur verwendet. Ein Vergleich der Ergebnisse aus diesen drei verschiedenen Varianten in den Abbildungen 6.9–6.12 zeigt, dass die approximative Gittergeschwindigkeit die Genauigkeit der dG-Verfahren stark beeinflussen kann. Die approximative Gittergeschwindigkeit mit Korrektur hingegen liefert dieselben Ergebnisse wie die exakte Gittergeschwindigkeit.

Die Abbildungen 6.13–6.14 zeigen, dass die cGP(k)-Verfahren für $k = 2, 3$ dieselben Ergebnisse bezüglich des Fehlers und der Konvergenzordnung wie in festen Gebieten erhalten werden, vgl. dazu Beispiel 2.

Aus den Abbildungen 6.5–6.14 kann man entnehmen, dass die dG(k)-Verfahren für $k = 1, 2$ in der $\ell^\infty(L^2)$ -Norm von den Konvergenzordnungen $2k + 1$ sind. Die cGP(k)-Verfahren für $k = 2, 3$ sind in der $\ell^\infty(L^2)$ -Norm von den Konvergenzordnungen $2k$. Die Abbildungen zeigen auch, dass die dG(k)-Verfahren für $k = 1, 2$ von den Konvergenzordnungen $k + 1$ in der $L^2(L^2)$ -Norm und k in der $H^1(L^2)$ -Norm sind. Die cGP(k)-Verfahren für $k = 2, 3$ sind von den Konvergenzordnungen $k + 1$ in der $L^2(L^2)$ -Norm und k in der $H^1(L^2)$ -Norm.

Fazit:

Zusammenfassend haben wir in den zeitabhängigen Gebieten variationelle Zeitdiskretisierungen höherer Ordnung durch unstetige Galerkin-Verfahren (dG) durchgeführt. Wir haben die Fehlerabschätzungen für die dG-Verfahren angegeben und bewiesen. Es ist deutlich zu sehen, dass die in Lemmata 5.16 und 5.17 für die dG(k)-Verfahren vorhergesagten Konvergenzordnungen durch die numerischen Beispiele bestätigt wurden. Trotz der fehlenden Theorie hat sich numerisch gezeigt, dass für die cGP-Verfahren die erwarteten Ergebnisse bezüglich des Fehlers und der Konvergenzordnung erhalten werden.

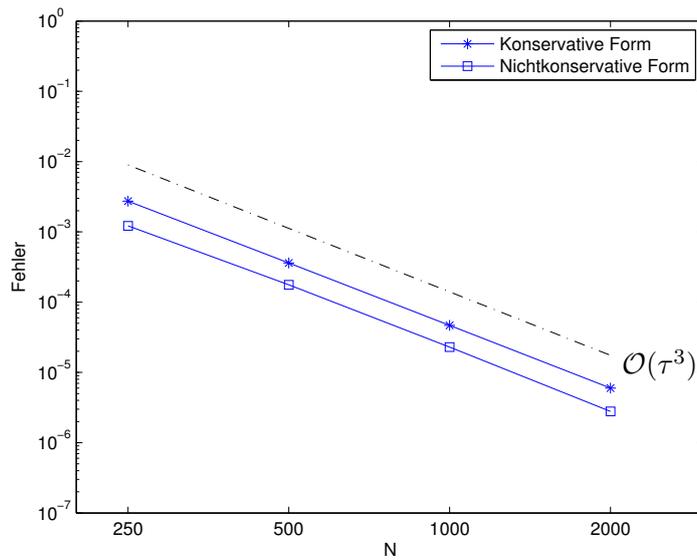


Abbildung 6.1: *Beispiel 3: Vergleich der Fehler des konservativen und des nichtkonservativen $dG(1)$ -Verfahrens bezüglich der $\ell^\infty(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .*

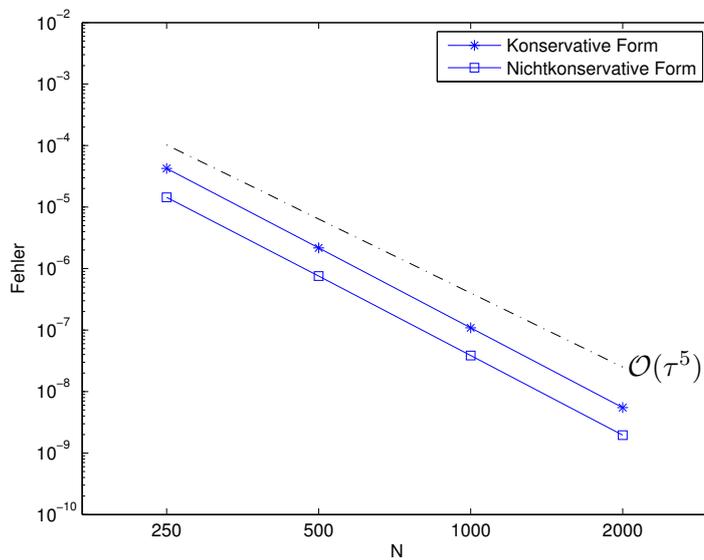


Abbildung 6.2: *Beispiel 3: Vergleich der Fehler des konservativen und des nichtkonservativen $dG(2)$ -Verfahrens bezüglich der $\ell^\infty(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .*

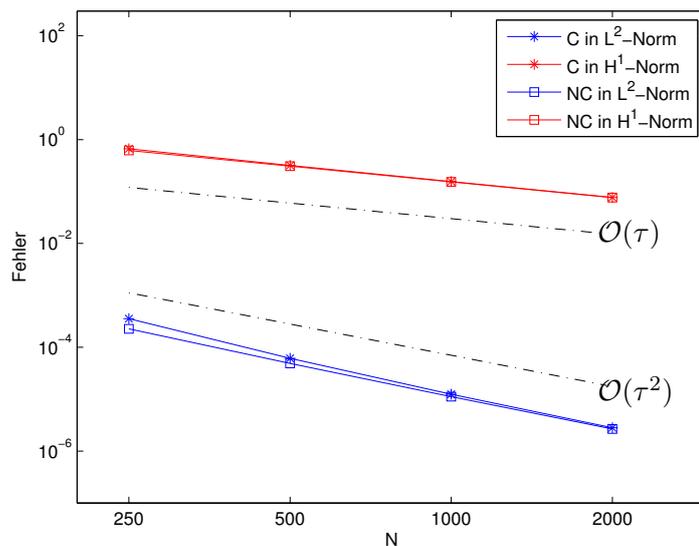


Abbildung 6.3: *Beispiel 3: Vergleich der Fehler des konservativen und des nichtkonservativen $dG(1)$ -Verfahrens bezüglich der $L^2(L^2)$ - und der $H^1(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .*

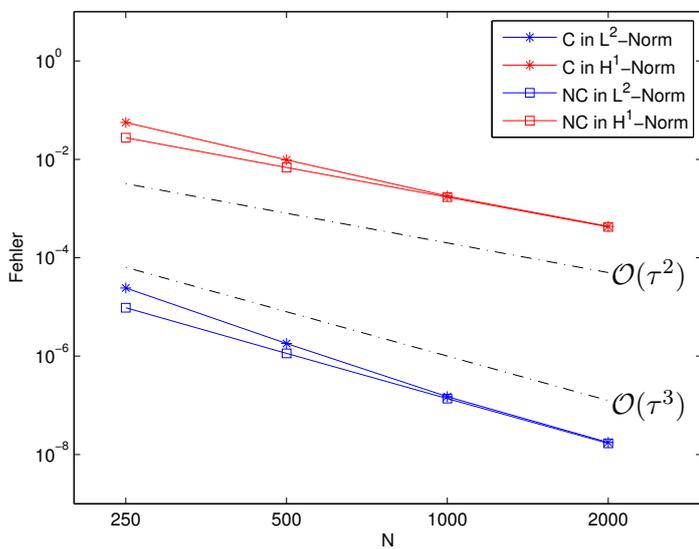


Abbildung 6.4: *Beispiel 3: Vergleich der Fehler des konservativen und des nichtkonservativen $dG(2)$ -Verfahrens bezüglich der $L^2(L^2)$ - und der $H^1(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .*

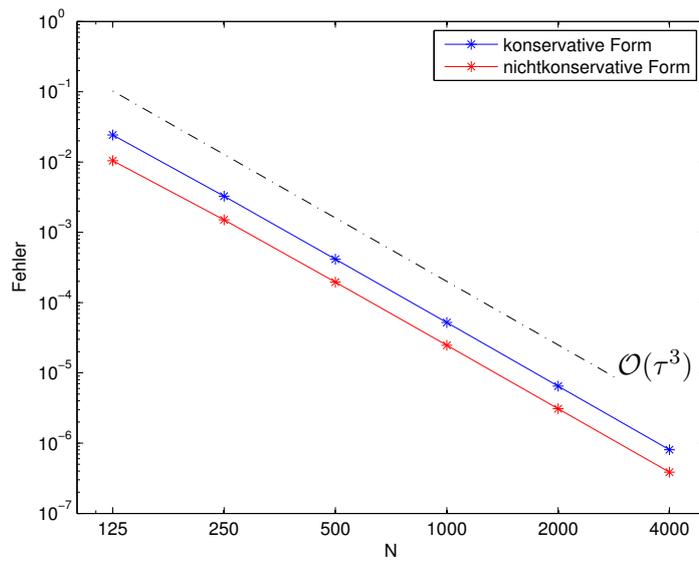


Abbildung 6.5: *Beispiel 4*: Vergleich der Fehler des konservativen und des nichtkonservativen $dG(1)$ -Verfahrens bezüglich der $\ell^\infty(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .

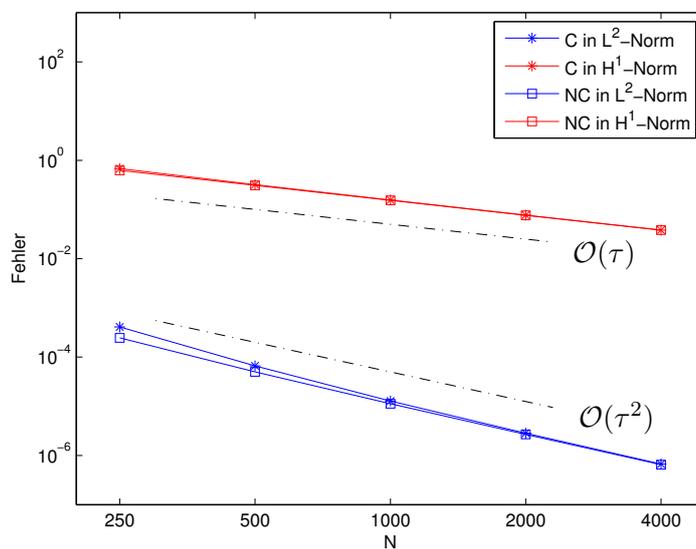


Abbildung 6.6: *Beispiel 4*: Vergleich der Fehler des konservativen (C) und des nichtkonservativen (NC) $dG(1)$ -Verfahrens bezüglich der $L^2(L^2)$ - und der $H^1(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .

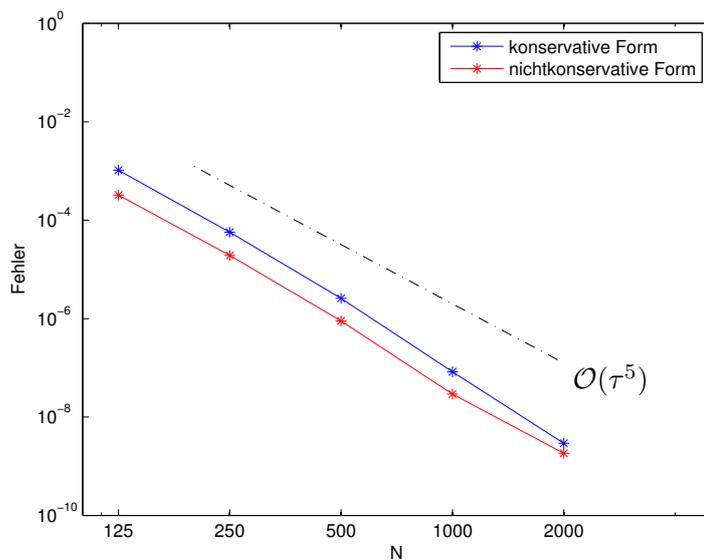


Abbildung 6.7: *Beispiel 4: Vergleich der Fehler des konservativen und des nichtkonservativen $dG(2)$ -Verfahrens bezüglich der $\ell^\infty(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .*

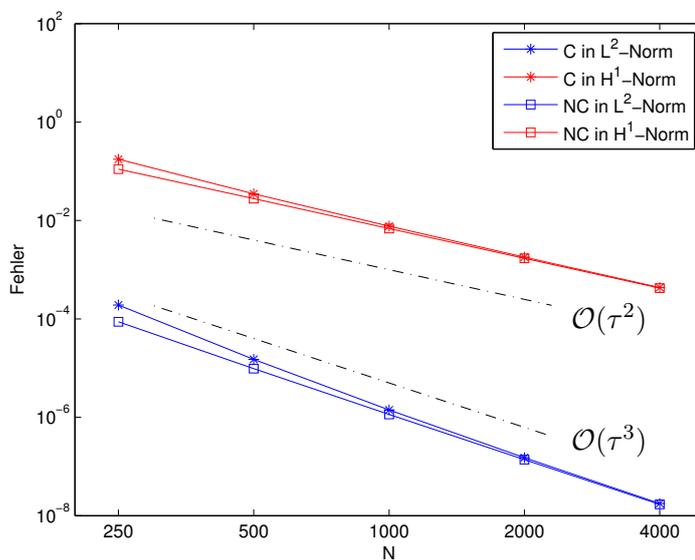


Abbildung 6.8: *Beispiel 4: Vergleich der Fehler des konservativen (C) und des nichtkonservativen (NC) $dG(2)$ -Verfahrens bezüglich der $L^2(L^2)$ - und der $H^1(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .*

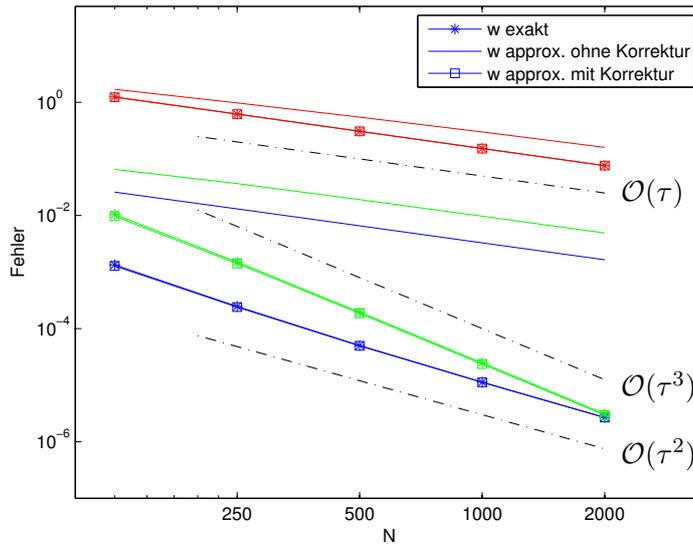


Abbildung 6.9: *Beispiel 4*: Vergleich der Fehler und Konvergenzordnung des nichtkonservativen $dG(1)$ -Verfahrens mit einer exakten, einer approximativen und einer approximativen Gittergeschwindigkeit w mit Korrektur bezüglich der $L^2(L^2)$ -, der $H^1(L^2)$ - und der $\ell^\infty(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .

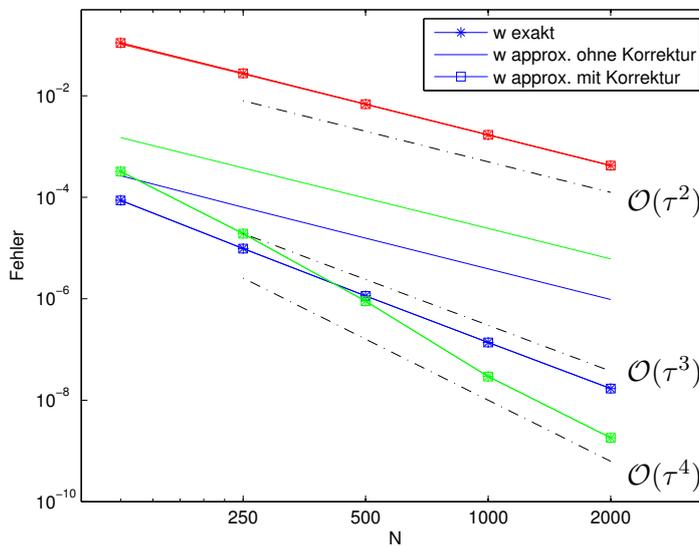


Abbildung 6.10: *Beispiel 4*: Vergleich der Fehler und Konvergenzordnung des nichtkonservativen $dG(2)$ -Verfahrens mit einer exakten, einer approximativen und einer approximativen Gittergeschwindigkeit w mit Korrektur bezüglich der $L^2(L^2)$ -, der $H^1(L^2)$ - und der $\ell^\infty(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .

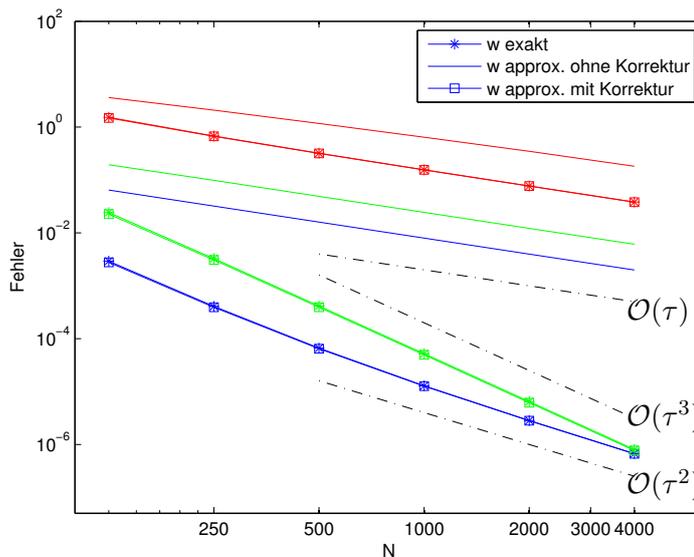


Abbildung 6.11: *Beispiel 4: Vergleich der Fehler und Konvergenzordnung des konservativen $dG(1)$ -Verfahrens mit einer exakten, einer approximativen und einer approximativen Gittergeschwindigkeit w mit Korrektur bezüglich der $L^2(L^2)$ -, der $H^1(L^2)$ - und der $\ell^\infty(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .*

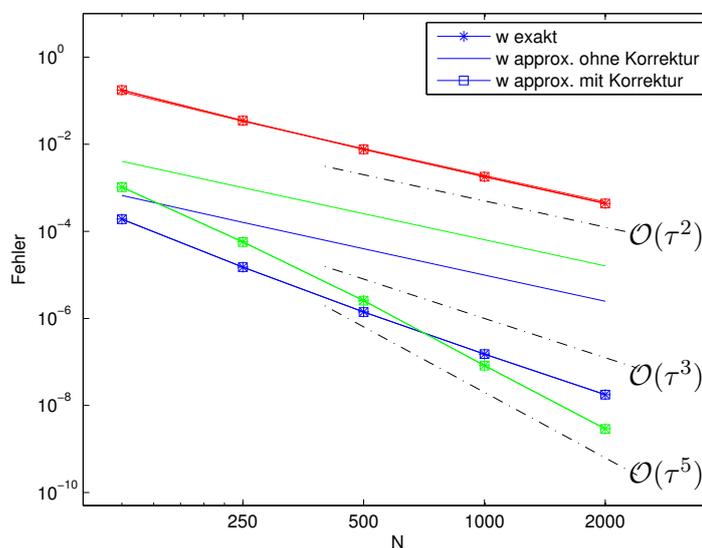


Abbildung 6.12: *Beispiel 4: Vergleich der Fehler und Konvergenzordnung des konservativen $dG(2)$ -Verfahrens mit einer exakten, einer approximativen und einer approximativen Gittergeschwindigkeit w mit Korrektur bezüglich der $L^2(L^2)$ -, der $H^1(L^2)$ - und der $\ell^\infty(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .*

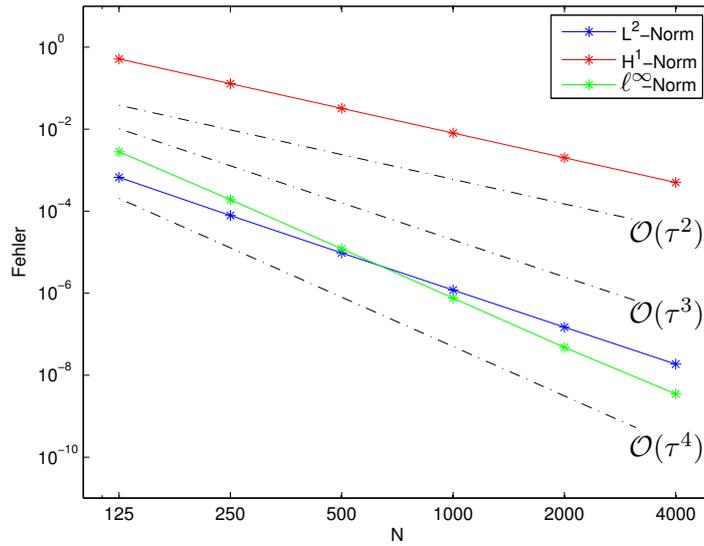


Abbildung 6.13: *Beispiel 4*: Fehler und Konvergenzordnung des konservativen $cGP(2)$ -Verfahrens bezüglich der $L^2(L^2)$ -, der $H^1(L^2)$ - und der $\ell^\infty(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .

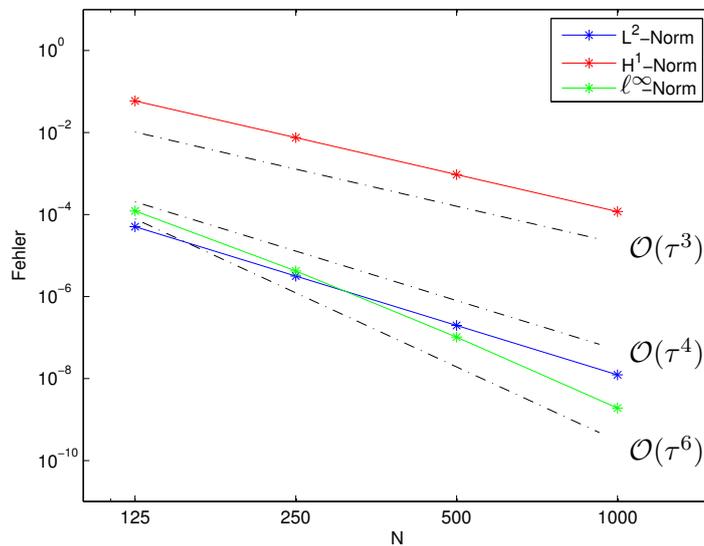


Abbildung 6.14: *Beispiel 4*: Fehler und Konvergenzordnung des konservativen $cGP(3)$ -Verfahrens bezüglich der $L^2(L^2)$ -, der $H^1(L^2)$ - und der $\ell^\infty(L^2)$ -Norm in Abhängigkeit der Anzahl der Zeitschritte N .

7 Zusammenfassung und Ausblick

Ein zentraler Aspekt dieser Arbeit ist die Erweiterung der Ausführungen von Matthies und Schieweck [29] mit dem Ziel variationelle Zeitdiskretisierungen höherer Ordnung in beweglichen Gebieten zu erhalten. Dabei wurden cGP- und dG-Verfahren entwickelt, die sich auf Probleme in zeitabhängigen Gebieten anwenden lassen. Diese Verfahren haben wir in Kapitel 6 auf unser im Ort semidiskretes Problem angewendet und implementiert. Zudem haben wir eine einheitliche Postprocessing-Methode für die Galerkin-Verfahren in zeitabhängigen Gebieten bewiesen, die sich auf andere Probleme, wie beispielsweise auf die Navier–Stokes-Gleichungen verallgemeinern lässt.

Inspiziert von der Arbeit von Ern und Schieweck [14] bezüglich dG-Verfahren haben wir theoretische Ergebnisse hinsichtlich der Stabilität sowie Fehlerabschätzung für die cGP-Verfahren in festen Gebieten erzielt. Dazu haben wir für cGP-Verfahren das in Kapitel 3 eingeführte einheitliche Postprocessing und einige modifizierte Interpolationsoperatoren verwendet. Mit einer Anpassung der Techniken des cGP-Falls haben wir Resultate für dG-Verfahren erreicht, die den Ergebnissen von Ern und Schieweck [14] entsprechen, wobei wir das Postprocessing in anderer Weise definiert haben.

Für ein durch dG-Verfahren diskretisiertes Problem haben wir wichtige Ergebnisse bezüglich der Stabilität und Fehlerabschätzung sowohl bei einer exakten Integration als auch bei der Verwendung einer Gauß–Radau-Quadraturformel gezeigt. Die analytischen Ergebnisse für dG-Verfahren, die erwartete Konvergenzordnung und die Stabilität bei cGP-Verfahren wurden durch numerische Beispiele bestätigt. Diesbezüglich eignet sich das hier vorgestellte ALE-Konzept, aufgrund der analytischen wie auch der numerischen Resultate, hervorragend für die betrachteten variationellen Zeitdiskretisierungen höherer Ordnung. Hinsichtlich der Gittergeschwindigkeit wurde durch numerische Beispiele gezeigt, inwiefern die verschiedenen Approximationen Einfluss auf die Genauigkeit der Zeitdiskretisierungsverfahren haben können. In diesem Rahmen wurden parallel zu der numerischen Analyse sämtliche Tools in MoonNMD [26] auf Probleme in zeitabhängigen Gebieten verallgemeinert.

Die Durchführung einer Vielzahl numerischer Tests hat gezeigt, dass die cGP-Verfahren die zu erwartenden Ergebnisse bezüglich der Konvergenzordnung liefern, demzufolge wäre eine Übertragung der gewonnenen Ergebnisse im dG-Fall auf den cGP-Fall möglich.

Mit Ausblick auf eine variationelle Zeitdiskretisierung höherer Ordnung der Navier–Stokes-Gleichungen in zeitabhängigen Gebieten lassen sich die in dieser Arbeit vorgestellten Techniken in Hinblick auf Implementierung, Stabilität- und Fehleruntersuchung für zukünftige Recherchen einsetzen.

Literaturverzeichnis

- [1] M. Abramowitz and A. Stegun. *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. Dover Publications, Incorporated, 1974.
- [2] N. Ahmed, G. Matthies, L. Tobiska, and H. Xie. Discontinuous galerkin time stepping with local projection stabilization for transient convection-diffusion-reaction problems. *Comput. Methods Appl. Mech. Engrg.*, 200(21–22):1747–1756, 2011.
- [3] G. Akrivis and C. Makridakis. Galerkin time-stepping methods for nonlinear parabolic equations. *ESAIM Math. Model. Numer. Anal.*, 38(2):261–289, 2004.
- [4] F. Armero and E. Love. An arbitrary lagrangian–eulerian finite element method for finite strain plasticity. *Int. J. Numer. Methods Eng.*, 57(4):471–508, 2003.
- [5] D. Boffi and L. Gastaldi. Stability and geometric conservation laws for {ALE} formulations. *Comput. Methods Appl. Mech. Engrg.*, 193(42–44):4717–4739, 2004.
- [6] A. Bonito, I. Kyza, and R. Nochetto. Time-discrete higher order ale formulations: a priori error analysis. *Numer. Math.*, 125(2):225–257, 2013.
- [7] A. Bonito, I. Kyza, and R. Nochetto. Time-discrete higher-order ale formulations: Stability. *SIAM J. Numer. Anal.*, 51(1):577–604, 2013.
- [8] A. Bonito, I. Kyza, and R. H. Nochetto. *A dG Approach to Higher Order ALE Formulations in Time*, pages 223–258. Springer International Publishing, Cham, 2014.
- [9] J.H. Bramble and S.R. Hilbert. Bounds for a class of linear functionals with applications to hermite interpolation. *Numer. Math.*, 16(4):362–369, 1971.
- [10] E. Burman and M. A. Fernández. Finite element methods with symmetric stabilization for the transient convection-diffusion-reaction equation. *Comput. Methods Appl. Mech. Engrg.*, 198(33):2508–2519, 2009.
- [11] P. G. Ciarlet. *The finite element method for elliptic problems*, volume 40. Society for Industrial and Applied Mathematics (SIAM), 2002.
- [12] Ph. Clément. Approximation by finite element functions using local regularization. *Revue française d’automatique, informatique, recherche opérationnelle. Analyse numérique*, 9(2):77–84, 1975.
- [13] J. Donea, S. Giuliani, and JP. Halleux. An arbitrary lagrangian-eulerian finite element method for transient dynamic fluid-structure interactions. *Comput. Methods Appl. Mech. Engrg.*, 33(1):689–723, 1982.
- [14] A. Ern and F. Schieweck. Discontinuous galerkin method in time combined with a stabilized finite element method in space for linear first–order pdes. *Math. Comp.*, electronically published on January 11, 2016, DOI:<http://dx.doi.org/10.1090/mcom/3073> (to appear in print).
- [15] C. Farhat and P. Geuzaine. Design and analysis of robust {ALE} time-integrators for the solution of unsteady flow problems on moving grids. *Comput. Methods Appl. Mech. Engrg.*, 193(39–41):4073–4095, 2004.

- [16] C. Farhat, P. Geuzaine, and C. Grandmont. The discrete geometric conservation law and the nonlinear stability of {ALE} schemes for the solution of flow problems on moving grids. *J. Comput. Phys.*, 174(2):669–694, 2001.
- [17] C. Farhat, M. Lesoinne, and N. Maman. Mixed explicit/implicit time integration of coupled aeroelastic problems: Three-field formulation, geometric conservation and distributed solution. *Internat. J. Numer. Methods Fluids*, 21:807–835, 1995.
- [18] M. Feistauer and J. Cesenek. On numerical simulation of airfoil vibrations induced by compressible flow. *Proceedings of ALGORITMY*, pages 22–31, 2012.
- [19] L. Formaggia and F. Nobile. A stability analysis for the arbitrary Lagrangian Eulerian formulation with finite elements. *East-West J. Numer. Math.*, 7(2):105–131, 1999.
- [20] L. Formaggia and F. Nobile. Stability analysis of second-order time accurate schemes for aleâfem. *Comput. Methods Appl. Mech. Engrg.*, 193(39–41):4097–4116, 2004.
- [21] L. Gastaldi. A priori error estimates for the arbitrary lagrangian eulerian formulation with finite elements. *East-West J. Numer. Math.*, 9, 2001.
- [22] P. Geuzaine, C. Grandmont, and C. Farhat. Design and analysis of {ALE} schemes with provable second-order time-accuracy for inviscid and viscous flow simulations. *J. Comput. Phys.*, 191(1):206–227, 2003.
- [23] H. Guillard and C. Farhat. On the significance of the geometric conservation law for flow computations on moving meshes. *Comput. Methods Appl. Mech. Engrg.*, 190:1467–1482, 2000.
- [24] CW. Hirt, A. Amsden, and JL. Cook. An arbitrary lagrangian–eulerian computing method for all flow speeds. *J. Comput. Phys.*, 14(3):227–253, 1974.
- [25] D. Jaubert and V. Moreau. Fluid-structure interaction: Applications in the context of a flow in arterial medium. 1997.
- [26] V. John and G. Matthies. MooNMD – a program package based on mapped finite element methods. *Comput. Vis. Sci.*, 6(2–3):163–169, 2004.
- [27] P. Knobloch. On the application of local projection methods to convection-diffusion-reaction problems. *BAIL 2008 - Boundary and Interior Layers*, 69:183–194, 2009.
- [28] P. Knobloch. A generalization of the local projection stabilization for convection-diffusion-reaction equations. *SIAM J. Numer. Anal.*, 48(2):659–680, 2010.
- [29] G. Matthies and F. Schieweck. Higher order variational time discretizations for nonlinear systems of ordinary differential equations. 2011.
- [30] G. Matthies, P. Skrzypacz, and L. Tobiska. A unified convergence analysis for local projection stabilisations applied to the oseen problem. *ESAIM: Math. Model. Numer. Anal.*, 41:713–742, 7 2007.
- [31] G. Matthies, P. Skrzypacz, and L. Tobiska. Stabilization of local projection type applied to convection-diffusion problems with mixed boundary conditions. *ETNA. Electron. Trans. Numer. Anal. [electronic only]*, 32:90–105, 2008.
- [32] F. Nobile. *Numerical approximation of fluid-structure interaction problems with application to haemodynamics*. EPFL, Lausanne, 2001.
- [33] A. Quarteroni, M. Tuveri, and A. Veneziani. Computational vascular fluid dynamics: problems, models and methods. *Comput. Vis. Sci.*, 2(4):163–197, 2000.

- [34] B. Rivière. *Discontinuous Galerkin methods for solving elliptic and parabolic equations*, volume 35 of *Frontiers in Applied Mathematics*. SIAM, Philadelphia, 2008.
- [35] H. G. Roos, M. Stynes, and L. Tobiska. *Robust numerical methods for singularly perturbed differential equations: convection-diffusion-reaction and flow problems*. Springer Ser. Comput. Math. Springer, 2008.
- [36] A. Rutherford. *Vectors, tensors, and the basic equations of fluid mechanics*. Dover Publications, New York, 1989.
- [37] L Ridgway Scott and Shangyou Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.
- [38] PD Thomas and CK Lombard. The geometric conservation law—a link between finite-difference and finite-volume methods of flow computation on moving grids. *11th Fluid and Plasma Dynamics Conference*, 1, 1978.
- [39] M. Vlasák and H.-G. Roos. An optimal uniform a priori error estimate for an unsteady singularly perturbed problem. *Int. J. Numer. Anal. Model.*, 11(1):24–33, 2014.

Erklärung

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.