Numerische Behandlung von Self-Consistent Field Theory-Modellen mittels Reduzierte-Basis-Methoden

Masterarbeit

am Institut für Mathematik,
Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität
in Mainz

Alexej Disterhoft
geboren in Leonidowka

Erstgutachter:	Prof. Dr. Thorsten Raasch
Zweitgutachter:	Prof. Dr. Martin Hanke-Bourgeois

Mainz, 10. September 2015

„I guess you could call it a ‚failure', but I prefer the term ‚learning experience'."

— Andy Weir, *The Martian*

Mein Dank gilt Herrn Prof. Dr. Raasch,
meiner Familie und meinen Freunden.

Inhaltsverzeichnis

1 Kapitel

Einleitung

Der Begriff *Polymer* bezeichnet einen aus Makromolekülen bestehenden Stoff. Diese Makromoleküle setzen sich wiederum aus vielen kleineren, sich wiederholenden Molekülen, in diesem Zusammenhang *Monomere* genannt, zusammen. Besteht ein Polymer aus nur einer Monomer-Gattung, dann spricht man von einem Homopolymer, sonst von einem Heteropolymer oder auch Copolymer. Obwohl Polymere in vielen verschiedenen Konfigurationen, beispielsweise ring- oder sternförmig, auftreten können, beschränken wir uns hier auf den Fall eines kettenförmigen Aufbaus. Weiter interessieren wir uns nur für Copolymere, vor allem die sogenannten Blockcopolymere, welche aus mehreren Monomer-Gattungen, die homogene zusammenhängende Blöcke bilden, aufgebaut sind (siehe Abbildung 1.1).

Von besonderem Interesse ist nun das Verhalten von Polymerschmelzen (engl. *polymer melt*), das heißt des flüssigen Aggregatzustands eines Polymers, sowie das Verhalten von Gemischen verschiedener polymerer Stoffe. So neigen die Gemische vieler Paare von Homopolymeren zu makroskopischer Phasenseparation, wie man es beispielsweise auch von Öl und Wasser kennt. Eine ähnliche Tendenz zur Separation findet man auch bei Schmelzen von Blockcopolymeren. Hierbei ist aber aufgrund der Verbindung zwischen den verschiedenen Monomer-Blöcken keine makroskopische Phasenseparation möglich, stattdessen kommt es zu einer periodischen mikroskopischen Separation (vergleiche Abbildung 1.2).

Da die experimentelle Bestimmung dieser Separationen ohne Vorwissen über mögliche stabile Anordnungen nur wenig erfolgversprechend ist, wird ein theoretisches Fundament benötigt, auf Basis dessen Vorhersagen getroffen werden können, die vorzugsweise auch zu experimentell belegbaren Anordnungen führen. Wir beschränken uns auf die Betrachtung des vergleichsweise einfachen Systems von Diblockcopolymeren, das heißt Blockcopolymeren, die aus zwei verschiedenen Monomer-Gattungen bestehen.

Als nützliches und relativ gut untersuchbares theoretisches Modell hat sich die sogenannte selbstkonsistente Feldtheorie (engl. *self-consistent field theory*, SCFT) herausgestellt,

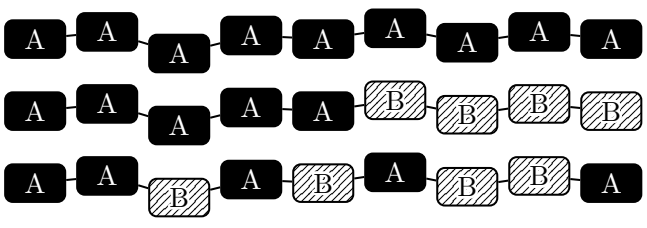


Abbildung 1.1. Skizzenhafte Darstellung verschiedener Polymerarten. Von oben nach unten: Homopolymer, ein AB-Diblockcopolymer und ein sogenanntes statistisches AB-Copolymer, bei dem die beiden Monomer-Arten zufällig verteilt sind.

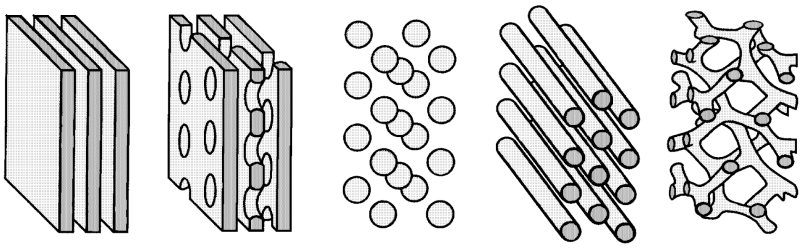


Abbildung 1.2. Verschiedene Phasen bei Diblockcopolymeren, welche experimentell beobachtet wurden, wobei hier nur eine der beiden Monomer-Gattungen dargestellt wird. Diese heißen von links nach rechts: lamellar, perforiert-lamellar, sphärisch, zylindrisch, gyroid. Diese Abbildung wurde [Mat07, Figure 1.18] entnommen.

die im Folgenden eingeführt und erläutert wird. Diese basiert auf einer makroskopischen Darstellung der Polymere und vernachlässigt die Vorgänge, welche auf atomarer Ebene stattfinden.

Mathematische Modellierung. Als Grundlage für die selbstkonsistente Feldtheorie dient eine Modellierung der Polymere als frei bewegliche Ketten (engl. *ideal chain*). Im Rahmen dieser Arbeit betrachten wir zwar nur die aus dem sogenannten stetigen Gaußschen Kettenmodell resultierende Theorie, erläutern aber zunächst ein diskretes Kettenmodell, welches eine Vorstufe des stetigen darstellt und die Idee dahinter verdeutlicht. Eine deutlich ausführlichere und vor allem mathematisch begründete Einführung in diese Thematik findet man bei Fredrickson [Fre05, Chapter 2] sowie Rubinstein und Colby [RC03].

Im Folgenden betrachten wir eine nicht näher spezifizierte Volumenzelle und geben mit dem Vektor $\mathbf{r}$ eine Position in diesem Volumen an.

Das diskrete Modell stellt die Polymerkette als eine diskrete Kette von Partikeln derart dar, dass aneinanderhängende Monomere ähnlich einem Scharnier frei beweglich sind. Dabei werden Wechselwirkungen zwischen benachbarten Monomeren berücksichtigt, während auf der Kette weit auseinanderliegende Partikel ignoriert werden. Diese Wechselwirkungen können in Abbildung 1.3 beispielsweise als Einschränkung des Winkels ϑ_9 durch die gegenseitige Beeinflussung der Partikel $8, 9$ und 10 auftreten.

Das stetige Gaußsche Kettenmodell, welches man unter anderem auch als stetigen Grenzfall des beschriebenen diskreten Modells erhält, hat sich sowohl bei analytischen als auch numerischen Betrachtungen als besonders nützlich erwiesen. Dabei wird die Polymerkette als stetige, linear elastische Faser aufgefasst und durch eine Kurve $\mathbf{r}(s)$ parametrisiert, wobei $s \in [0, 1]$ eine entlang der normalisierten Kontur der Kette verlaufende Variable ist.

Sowohl das diskrete als auch das stetige Kettenmodell haben einen starken Bezug zur Stochastik, da sie sich auch als Random Walks beziehungsweise stochastische Prozesse auffassen lassen. Dies stellt einen umfangreichen „Werkzeugkasten“ zur Untersuchung dieser zur Verfügung. Da die benötigten stochastischen Ausführungen und Herleitungen für diese Arbeit jedoch nebensächlich sind, belassen wir es bei diesen informalen Beschreibungen und widmen uns nun der darauf aufbauenden selbstkonsistenten Feldtheorie.

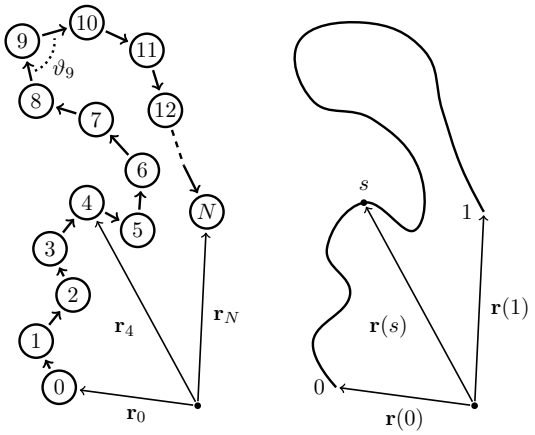


Abbildung 1.3. Schematische Darstellung einer Polymerkette im diskreten Kettenmodell (links) und im stetigen Gaußschen Kettenmodell (rechts). Abbildung reproduziert nach [Fre05, Figure 2.1 und 2.5].

Selbstkonsistente Feldtheorie. Bei der selbstkonsistenten Feldtheorie handelt es sich um ein weit verbreitetes theoretisches Modell der Physik, um das Verhalten von Teilchen unter Einwirkung von Kräften, die durch Wechselwirkungen mit weiteren Teilchen auftreten, zu studieren. Sie wird nicht nur im Zusammenhang mit Polymeren, sondern zum Beispiel auch in der Thermodynamik oder Informatik verwendet.

Die Grundidee ist folgende: in einem System vieler wechselwirkender Objekte kann die auf ein einzelnes Teilchen wirkende Gesamtkraft durch Mittelung aller Wechselwirkungen approximiert werden. Diese gemittelten Einwirkungen werden als externes Feld aufgefasst und ignorieren dabei Fluktuationen, das heißt Veränderungen der wirkenden Kräfte durch das lokale Verhalten, beispielsweise Bewegungen, der einzelnen Teilchen. Damit erreicht man effektiv die Reduktion eines Mehrkörperproblems auf ein Einkörperproblem und kann so das Verhalten eines solchen Systems untersuchen.

Dieses Prinzip lässt sich auch zur Untersuchung von Polymeren verwenden, da eine einzelne Polymerkette oftmals aus einer hohen vierstelligen Zahl von Atomen besteht und dadurch die Wechselwirkungen auf atomarem Level vernachlässigbar sind. Aufbauend auf den beschriebenen Modellen, in diesem Fall dem stetigen Gaußschen Kettenmodell, kann so die statistische räumliche Ausrichtung von Polymerketten bestimmt werden.

Wir beschränken uns im Folgenden auf die Beschreibung der SCFT für die inkompressible Schmelze eines AB-Diblockcopolymers und folgen dabei größtenteils den Ausführungen von Matsen und Schick [MS94] sowie Stasiak und Matsen [SM11].

Erneut wird eine einzelne Volumenzelle betrachtet, beispielsweise ein Würfel, welche selbst Teil eines größeren Systems sein kann. Diese Zelle enthalte n AB-Diblockcopolymere, welche jeweils aus einem A-Block und einem B-Block bestehen, wobei diese wiederum aus N_A Monomeren vom Typ A und aus N_B Monomeren vom Typ B zusammengesetzt sind. Der *Polymerisationsgrad*, das heißt die Gesamtanzahl an Monomeren in einem Polymer, ist durch $N = N_\mathrm{A} + N_\mathrm{B}$ gegeben. Weiter bezeichne $f = N_\mathrm{A}/N$ den Anteil an A-Monomeren im gesamten Polymer. Wie bei der Beschreibung des Gaußschen Modells ist $s \in [0, 1]$ eine normalisierte Distanz entlang der Kontur einer Polymerkette, wobei $s = 0$ und $s = 1$ den beiden Enden entspricht.

Als vereinfachende Annahmen sei die *statistische Länge* a eines Monomers, auch *Kuhn-Länge* genannt, der beiden Monomer-Gattungen gleich und ein Monomer beider Gattungen nehme das selbe Volumen $1/\rho_0$ ein. Das Gesamtvolumen der Schmelze in

dieser Zelle ist damit durch $V = nN/\rho_0$ gegeben.

Die wichtigsten Größen bei der SCFT sind nun die *Konzentrationen* $\phi_\mathrm{A}(\mathbf{r})$ und $\phi_\mathrm{B}(\mathbf{r})$ der A- und B-Monomere an einer Position $\mathbf{r}$ in der betrachteten Zelle und die *externen Felder* $\omega_\mathrm{A}(\mathbf{r})$ und $\omega_\mathrm{B}(\mathbf{r})$, welche auf die jeweiligen Monomer-Gattungen wirken.

Als Ausgangspunkt für die Bestimmungen möglicher stabiler Anordnungen in der Polymerschmelze dient das sogenannte *Freie-Energie-Funktional* des Systems, genauer ausgeführt in [Mat07; Fre05], welches für die freie Energie F eines einzelnen Polymers die Form

$$\frac{F}{nk_\mathrm{B}T} = -\ln\frac{Q}{V} + \frac{1}{V}\int\left[\chi N\phi_\mathrm{A}(\mathbf{r})\phi_\mathrm{B}(\mathbf{r}) - \omega_\mathrm{A}(\mathbf{r})\phi_\mathrm{A}(\mathbf{r}) - \omega_\mathrm{B}(\mathbf{r})\phi_\mathrm{B}(\mathbf{r})\right]\mathrm{d}\mathbf{r} \tag{1.1}$$

hat, wobei χ der sogenannte *Flory-Huggins-Wechselwirkungsparameter* für die Wechselwirkungen zwischen den Monomeren vom Typ A und B und $k_\mathrm{B}T$ die *thermische Energie* ist.

Stabile Anordnungen der Polymerketten entsprechen nun stationären Punkten des Freie-Energie-Funktionals bezüglich der Konzentrationen ϕ_A und ϕ_B sowie der Felder ω_A und ω_B. Betrachtet man die Funktionalableitungen von F bezüglich dieser Größen, dann bilden diese ein System von Gleichungen, kurz SCFT-Gleichungen genannt, anhand derer die gesuchten stationären Punkte bestimmt werden können. Diese Gleichungen bestehen aus der *Inkompressibilität* der Schmelze

$$\phi_\mathrm{A}(\mathbf{r}) + \phi_\mathrm{B}(\mathbf{r}) = 1 \tag{1.2}$$

sowie der Kopplung der Felder und der Konzentrationen durch

$$\omega_\mathrm{A}(\mathbf{r}) = \chi N\phi_\mathrm{B}(\mathbf{r}) + \xi(\mathbf{r}), \qquad \omega_\mathrm{B}(\mathbf{r}) = \chi N\phi_\mathrm{A}(\mathbf{r}) + \xi(\mathbf{r}), \tag{1.3}$$

wobei mit dem Lagrange-Multiplikator $\xi(\mathbf{r})$ die Inkompressibilität (1.2) erzwungen wird. Weiter erhält man eine Darstellung der Konzentrationen in Form von

$$\phi_\mathrm{A}(\mathbf{r}) = \frac{V}{Q}\int_0^f q(\mathbf{r},s)q^\dagger(\mathbf{r},s)\,\mathrm{d}s, \qquad \phi_\mathrm{B}(\mathbf{r}) = \frac{V}{Q}\int_f^1 q(\mathbf{r},s)q^\dagger(\mathbf{r},s)\,\mathrm{d}s, \tag{1.4}$$

wobei $Q = Q[\omega_\mathrm{A},\omega_\mathrm{B}]$ die *Partitionsfunktion* eines einzelnen Polymers ist und durch

$$Q = \int q(\mathbf{r},1)\,\mathrm{d}\mathbf{r} \tag{1.5}$$

bestimmt wird.

Die in den Gleichungen (1.4) und (1.5) auftretende Funktion $q(\mathbf{r},s)$ wird als *Vorwärts-Propagator* bezeichnet und erfüllt die parabolische partielle Differentialgleichung

$$\frac{\partial}{\partial s}q(\mathbf{r},s) = \frac{a^2N}{6}\Delta q(\mathbf{r},s) - \omega(\mathbf{r},s)q(\mathbf{r},s), \quad q(\mathbf{r},0) = 1. \tag{1.6}$$

Analog wird $q^\dagger(\mathbf{r},s)$ als *Rückwärts-Propagator* bezeichnet, da dieser eine ähnliche Differentialgleichung der Form

$$-\frac{\partial}{\partial s}q^\dagger(\mathbf{r},s) = \frac{a^2N}{6}\Delta q^\dagger(\mathbf{r},s) - \omega(\mathbf{r},s)q^\dagger(\mathbf{r},s), \quad q^\dagger(\mathbf{r},1) = 1 \tag{1.7}$$

erfüllt. Die Abbildung ω ist dabei definiert als

$$\omega(\mathbf{r}, s) = \begin{cases} \omega_{\mathrm{A}}(\mathbf{r}), & 0 \leq s < f, \\ \omega_{\mathrm{B}}(\mathbf{r}), & f \leq s \leq 1. \end{cases}$$

Je nachdem, welches Szenario betrachtet wird, entweder eine Volumenzelle innerhalb eines größeren Systems oder eine Zelle, welche durch feste Wände begrenzt wird, erhält man verschiedene Randbedingungen an die beiden Differentialgleichungen. Ersteres entspricht dabei periodischen Randbedingungen.

Ferner ist an dieser Stelle erwähnenswert, dass das Funktional der freien Energie (1.1) invariant ist bezüglich konstanter Verschiebungen der Felder ω_{A} und ω_{B}, wie beispielsweise [CF04] entnommen werden kann. Dies wird sich später sowohl in der theoretischen als auch numerischen Untersuchung als äußerst nützlich erweisen.

Einsatz numerischer Methoden. Es gibt verschiedene Ansätze, um die SCFT-Gleichungen in einem numerischen Verfahren zu verarbeiten. Oftmals führen diese zu einem dem nachfolgenden ähnlichen, iterativen Schema, welches Ähnlichkeiten mit einem Newton-Verfahren aufweist:

i. Zunächst wird $k = 0$ gesetzt und anschließend werden zufällig zwei externe Felder $\omega_{\mathrm{A}}^{(0)}$ und $\omega_{\mathrm{B}}^{(0)}$ generiert, um von vornherein auftretende Verzerrungen zu einem bestimmten stationären Punkt zu verhindern.

ii. Die Differentialgleichungen (1.6) und (1.7) werden für die Felder $\omega_{\mathrm{A}}^{(k)}$ und $\omega_{\mathrm{B}}^{(k)}$ gelöst.

iii. Die Konzentrationen $\phi_{\mathrm{A}}^{(k)}$ und $\phi_{\mathrm{B}}^{(k)}$ werden durch die Gleichungen (1.4) und (1.5) bestimmt.

iv. Diese Konzentrationen werden nun benutzt, um mit Hilfe der Gleichungen (1.3) die zugehörigen Felder zu bestimmen. Aus diesen Feldern werden mit einem Mixing-Verfahren die neuen Felder $\omega_{\mathrm{A}}^{(k+1)}$ und $\omega_{\mathrm{B}}^{(k+1)}$ für die nächste Iteration erzeugt. Das Mixing dient dazu, die Konvergenz des Verfahrens sicherzustellen beziehungsweise zu verbessern. Typischerweise gehen hier die Inkompressibilität (1.2) und zurückliegende Iterationen ein.

v. Wurde ein stationärer Punkt von (1.1) gefunden, so terminiert das Verfahren, ansonsten wird die Iteration ab Punkt *ii* fortgesetzt.

Bei dem beschriebenen Verfahren stellen sich die folgenden beiden Schritte als besonders wichtig heraus, da sie maßgeblich die Laufzeit des Iterationsverfahren beeinflussen:

i. das Lösungsverfahren für die Differentialgleichungen (1.6) und (1.7),

ii. das Mixing-Verfahren, mit dem iterativ neue Felder ω_{A} und ω_{B} bestimmt werden.

Auf den zweiten Punkt, das Mixing-Verfahren, werden wir im Verlauf dieser Arbeit nicht weiter eingehen und stattdessen hier einige der verwendeten Ansätze erwähnen. Da der Mixing-Schritt im Wesentlichen als Konvergenzbeschleuniger des Iterationsverfahrens

angedacht ist, lassen sich hierfür viele bekannte Verfahren der nichtlinearen Optimierung, aber auch aus anderen Bereichen, anwenden. Dies reicht von einem Quasi-Newton-Verfahren [MS94] bis zu Integrationsverfahren wie zum Beispiel Runge-Kutta-Verfahren oder Mehrschrittverfahren. Ein auf einem solchen Integrationsverfahren basierendes Mixing findet sich in [CF04], ferner findet man darin auch ein auf einem CG-Verfahren aufbauendes Mixing. Als besonders effektiver Ansatz hat sich das sogenannte Anderson-Mixing erwiesen, siehe die Arbeiten [TRL04; SM11]. Dabei werden neue Felder durch Kombination der Felder vieler zurückliegender Iterationen gewonnen. Weiter wurden in [DF99] Verfahren ähnlich einer Picard-Iteration betrachtet.

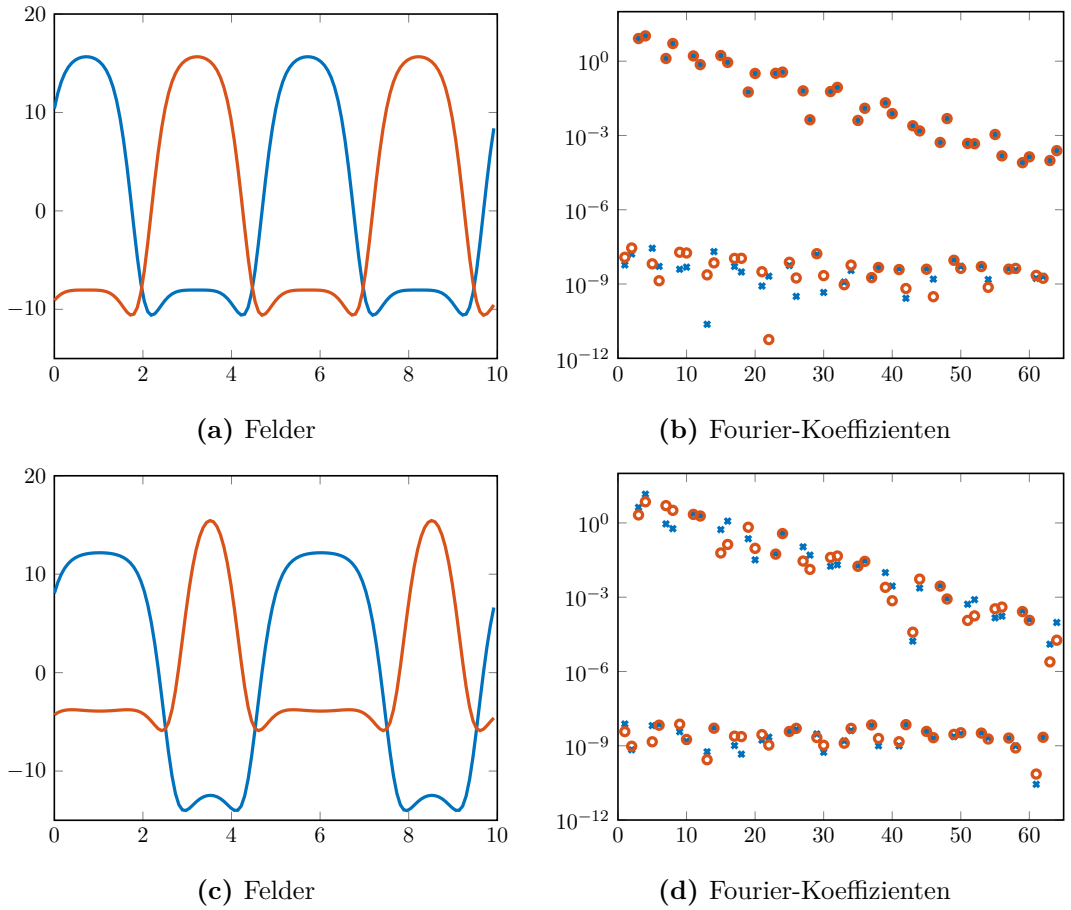
Unser Hauptaugenmerk in dieser Arbeit liegt auf dem ersten Problem, dem wiederholten Lösen der parabolischen partiellen Differentialgleichung (1.6). Da es abhängig vom gewählten Mixing-Verfahren oftmals eine Iterationsanzahl im mindestens dreistelligen Bereich benötigt, bis eine zufriedenstellende Genauigkeit bei den SCFT-Gleichungen erreicht ist, und damit insbesondere auch die partielle Differentialgleichung so oft gelöst werden muss, ist es wichtig, dass das Lösungsverfahren möglichst effizient ist. Weiter darf zu Gunsten der Laufzeit aber auch nicht die Genauigkeit des Lösers vernachlässigt werden, da sich dies im Iterationsverfahren durch Instabilität und zusätzliche Iterationen niederschlagen kann.

Ähnlich wie beim Mixing-Schritt wurden bereits viele verschiedene Ansätze mit mehr oder weniger zufriedenstellenden Ergebnissen verfolgt. Da es sich bei (1.6) im Grunde um eine Diffusionsgleichung handelt, lassen sich gut bekannte Methoden, zum Beispiel ein Finite-Differenzen-Verfahren, anwenden. So wird in [DF99] ein Crank-Nicolson-Verfahren eingesetzt, wobei hierbei explizit der Laufzeit Vorrang gegenüber der Genauigkeit gegeben wurde.

Als guter Kompromiss zwischen Laufzeit und Genauigkeit haben sich Spektral- und Pseudospektralverfahren etabliert. Erstere wurden von Matsen und Schick [MS94] erfolgreich eingesetzt, wobei hier erst das explizite Berücksichtigen der Symmetrien der zu erwartenden resultierenden Anordnung bei der Konstruktion des Spektralverfahrens zu annehmbaren Laufzeiten führt. Die damit verwandten Pseudospektralverfahren kommen zwar nicht an die Genauigkeit der Spektralverfahren heran, können aber unter Ausnutzung der Struktur der partiellen Differentialgleichung enorm Laufzeit einsparen. Dazu wird in [RK02] der Differentialoperator mittels Operator-Splitting so zerlegt, dass man das Lösen der Differentialgleichung mittels schneller Fourier-Transformation im Wesentlichen auf komponentenweise Vektor-Multiplikationen zurückführt. Das daraus resultierende Verfahren zweiter Ordnung wurde von [CGF06; RQM08] auf unterschiedliche Weisen zu Verfahren vierter Ordnung erweitert, ohne signifikant Laufzeit pro Iterationsschritt einzubüßen. Eine gute Übersicht über die meisten der hier genannten Methoden findet man bei Fredrickson [Fre05, Section 3.6] und Audus et al. [Aud+13].

Obwohl die Literatur zur SCFT für Polymere verschiedenste Verfahren für das Lösen der partiellen Differentialgleichung bietet, ist ein Galerkin-Ansatz basierend auf einer Raum-Zeit-Variationsformulierung unseres Wissens bisher nicht verfolgt worden.

Hier knüpfen wir an und betrachten das Problem, losgelöst vom eigentlichen SCFT-Iterationsverfahren eine Lösung für die Differentialgleichung zu finden, da der Berechnungsaufwand des im Folgenden hergeleiteten Verfahrens eine Integration in das Iterationsverfahren im Rahmen dieser Arbeit unmöglich macht. Die Grundidee ist die Verwendung eines Galerkin-Verfahrens für die Differentialgleichung mit anschließendem



(a) Felder **(b)** Fourier-Koeffizienten

(c) Felder **(d)** Fourier-Koeffizienten

Abbildung 1.4. Eindimensionale Beispiele einer stabilen Anordnung eines Diblockcopolymers, welche mittels SCFT und Pseudospektralverfahren bestimmt wurden. Das obere Beispiel wurde auf einem Intervall der Länge $L = 10$ mit den relevanten Größen $f = 1/2$, $\chi N = 25$ und $a^2 N/6 = 10/3$ simuliert. Für das untere Beispiel wurde lediglich f auf den Wert $1/3$ geändert. Monomer-Typ A entspricht den blauen und B dementsprechend den orangen Graphen. Die Fourier-Koeffizienten haben die Reihenfolge $\cos(2\pi x), \sin(2\pi x), \cos(4\pi x), \ldots$ und der konstante Anteil wurde vernachlässigt.

Aufsetzen eines Reduzierte-Basis-Ansatzes. Für diesen ist es notwendig, die Differentialgleichung zu parametrisieren, indem die in der Differentialgleichung auftretenden Felder in einem geeigneten Funktionensystem entwickelt werden. Dieses sollte so gewählt werden, dass die Entwicklungskoeffizienten möglichst schnell abfallen. Hierzu sei auf Abbildung 1.4 verwiesen, welche ein Felder-Paar von einem Iterationsdurchlauf und dessen Fourier-Koeffizienten zeigt.

Inhaltlicher Aufbau. In Kapitel 2 werden zunächst die funktionalanalytischen Grundlagen eingeführt beziehungsweise wiederholt, welche dann in Kapitel 3 verwendet werden, um die parabolische partielle Differentialgleichung, welche im Rahmen dieser Einleitung bereits vorgestellt wurde, unter passenden Rahmenbedingungen zu formalisieren. Weiter parametrisieren wir die Differentialgleichung und weisen unter bestimmten Bedingungen

eine Regularität der Lösungen der Differentialgleichung vom Parameter nach.

Nachfolgend werden in Kapitel 4 die ersten numerischen Grundlagen in Form des Petrov-Galerkin-Verfahrens gelegt und an Beispielen analysiert, um dann darauf aufbauend in Kapitel 5 die Reduzierte-Basis-Methode einzuführen und diese ebenfalls auf die in dieser Arbeit betrachtete Problemstellung anzuwenden.

Abschließend wird in Kapitel 6 ein Résumé der Arbeit und daran anknüpfend ein Ausblick, welcher mögliche Ansatzpunkte für Verbesserungen und Weiterentwicklungen erläutert, gegeben.

2 Kapitel

Funktionalanalytische Grundlagen

Um die in der Einleitung beschriebenen parabolischen partiellen Differentialgleichungen theoretisch und numerisch untersuchen zu können, müssen wir zunächst ein robustes Grundgerüst schaffen. Dies beginnen wir in diesem Kapitel mit der Einführung respektive Wiederholung der benötigten Grundlagen aus der Funktionalanalysis. Dabei orientieren wir uns maßgeblich an den Arbeiten von Dautray und Lions [DL92] und Schweizer [Sch13], in welchen die nachfolgenden Ausführungen weit detaillierter zu finden sind.

2.1. Bochner-Räume

Bevor wir uns an die Herleitung einer Raum-Zeit-Variationsformulierung parabolischer partieller Differentialgleichungen begeben, müssen wir zunächst die zugrundeliegenden Funktionenräume einführen. Hierbei konzentrieren wir uns auf die sogenannten *Bochner-Räume*, welche eine Verallgemeinerung der bekannten Lebesgue-Räume L_p auf Banachraum-wertige Funktionen darstellen, schränken uns dabei aber stets auf den für uns relevanten Fall eines endlichen Zeitintervalls ein. Weiter beschränken wir uns in dieser Arbeit auf die Betrachtung reeller Räume und Abbildungen, wobei ein Großteil der Aussagen auch für Strukturen über den komplexen Zahlen gilt.

Wir beginnen nun mit der folgenden Definition der Bochner-Räume nach [DL92, Definition XVIII.1.1].

Definition 2.1. Seien X ein Banachraum und $-\infty < a < b < \infty$. Als *Bochner-Raum* $L_2(a,b;X)$ bezeichnen wir die Menge (der Äquivalenzklassen) L_2-integrierbarer Funktionen $f\colon [a,b] \to X$, das heißt aller auf $[a,b]$ messbaren Funktionen mit

$$\|f\|_{L_2(a,b;X)} := \left(\int_a^b \|f(t)\|_X^2 \,\mathrm{d}t \right)^{1/2} < \infty.$$

Ferner ist der *Bochner-Raum* $L_\infty(a,b;X)$ definiert als die Menge (der Äquivalenzklassen) der für fast alle $t \in [a,b]$ wesentlich beschränkten Funktionen, also aller messbaren Funktionen $f\colon [a,b] \to X$ mit

$$\|f\|_{L_\infty(a,b;X)} := \operatorname*{ess\,sup}_{t\in[a,b]} \|f(t)\|_X < \infty.$$

Lemma 2.2. *Der Bochner-Raum $L_2(a,b;X)$ ist ein Banachraum. Ist ferner H ein Hilbertraum, so auch $L_2(a,b;H)$ mit dem Skalarprodukt*

$$\langle u, v\rangle_{L_2(a,b;H)} := \int_a^b \langle u(t), v(t)\rangle_H \,\mathrm{d}t \quad \textit{für } u, v \in L_2(a,b;H).$$

Beweis. Die erste Aussage findet sich in [DL92, Proposition XVIII.1.1], die zweite in [LM72, Abschnitt 1.1.3]. □

Weiter können wir für Funktionen aus einem Bochner-Raum eine Zeitableitung definieren, hier nach [DL92, S. 471; Sch13, Definition 10.6].

Definition 2.3. Seien X und Y Banachräume mit stetiger Einbettung $X \hookrightarrow Y$ und sei $u \in L_2(a,b;X)$. Existiert ein $v \in L_2(a,b;Y)$ mit

$$\int_a^b v(t)\varphi(t)\,\mathrm{d}t = -\int_a^b u(t)\varphi'(t)\,\mathrm{d}t \quad \text{für alle } \varphi \in \mathcal{C}_0^\infty((a,b),\mathbb{R}),$$

so bezeichnen wir diese distributionelle Ableitung $u_t := v$ als *Zeitableitung* von u.

Bemerkung 2.4. Der einfacheren Notation wegen werden wir die beiden Schreibweisen u_t und u' für die Zeitableitung verwenden.

Obige Definition einer Zeitableitung ermöglicht die Definition der Bochner-Sobolev-Räume. Wir beschränken uns hier auf den für uns relevanten Raum erster Ordnung, siehe [Eva10, Section 5.9.2].

Definition 2.5. Sei X ein Banachraum. Als *Bochner-Sobolev-Raum* erster Ordnung definieren wir den Raum

$$H^1(a,b;X) := \{u \in L_2(a,b;X) \,|\, u' \in L_2(a,b;X)\}.$$

Ferner gilt eine zu Lemma 2.2 analoge Aussage auch für Bochner-Sobolev-Räume. Da wir im Zuge dieser Arbeit ausschließlich mit Hilberträumen arbeiten werden, wird sich das folgende Konstrukt als grundlegend erweisen. Die nachfolgende Definition ist eine leichte Abwandlung von [Sch13, Abschnitt 10.2].

Definition 2.6. Seien V und H separable Hilberträume mit den Dualräumen V' und H'. Weiter sei die Einbettung $V \hookrightarrow H$ dicht und stetig. Durch die Identifikation $H \cong H'$ erhalten wir das sogenannte *Gelfand-Tripel* $V \hookrightarrow H \hookrightarrow V'$, wobei die zweite Inklusionen ebenfalls eine dichte stetige Einbettung ist.

Bemerkung 2.7. Es seien $\langle\cdot,\cdot\rangle_V$ und $\langle\cdot,\cdot\rangle_H$ die Skalarprodukte auf V respektive H. Weiter verwenden wir die Schreibweise $\langle\cdot,\cdot\rangle_{V'\times V}$ auch für die duale Paarung auf $V' \times V$, welche die eindeutige stetige Fortsetzung von $\langle\cdot,\cdot\rangle_H$ ist. Dadurch gilt insbesondere

$$\langle u, v\rangle_{V'\times V} = \langle u, v\rangle_H \quad \text{für alle } u \in H, v \in V.$$

Mit Hilfe dieser Gelfand-Tripel können wir nun die Räume definieren, welche wir letztendlich für die schwache Formulierung parabolischer partieller Differentialgleichungen benutzen werden. Dies geschieht analog zu [DL92, Definition XVIII.2.4].

Definition 2.8. Sei $V \hookrightarrow H \hookrightarrow V'$ ein Gelfand-Tripel. Wir definieren den Raum

$$W(a,b;V,V') := \{u \in L_2(a,b;V) \,|\, u' \in L_2(a,b;V')\},$$

wobei u' im Sinne von Definition 2.3 zu verstehen ist. Es gilt ferner

$$W(a,b;V,V') = L_2(a,b;V) \cap H^1(a,b;V').$$

Weiter können wir diesem Raum ein Skalarprodukt und damit auch eine Norm zuordnen. Dies liefert die folgende Aussage.

Lemma 2.9. *Versehen wir $W(a,b;V,V')$ mit dem Skalarprodukt*

$$\langle u,v\rangle_{W(a,b;V,V')} := \langle u,v\rangle_{L_2(a,b;V)} + \langle u',v'\rangle_{L_2(a,b;V')}$$

und der dadurch induzierten Norm

$$\|u\|_{W(a,b;V,V')} = \left(\|u\|^2_{L_2(a,b;V)} + \|u'\|^2_{L_2(a,b;V')}\right)^{1/2},$$

so ist $W(a,b;V,V')$ ein Hilbertraum.

Beweis. Ein Beweis findet sich in [DL92, Proposition XVIII.2.6]. □

Da die von uns betrachteten parabolischen partiellen Differentialgleichungen jeweils mit Anfangsbedingungen versehen sein werden, müssen wir an dieser Stelle klären, in welchem Sinne diese zu verstehen sind. Dies führt zum sogenannten Spursatz, welchen wir durch die folgende Einbettungsaussage erhalten.

Satz 2.10. *Seien $V \hookrightarrow H \hookrightarrow V'$ ein Gelfand-Tripel und $a,b \in \mathbb{R}$. Ferner sei $\mathcal{C}([a,b];H)$ die Menge aller stetigen Funktionen $f\colon [a,b] \to H$. Dann ist die Einbettung*

$$W(a,b;V,V') \hookrightarrow \mathcal{C}([a,b],H)$$

stetig. Insbesondere stimmt jedes $u \in W(a,b;V,V')$ fast überall mit einer stetigen Funktion aus $\mathcal{C}([a,b],H)$ überein.

Beweis. Siehe [DL92, Theorem XVIII.2.1; Sch13, Theorem 10.9]. □

Korollar 2.11 (Spursatz). *Seien $a,b \in \mathbb{R}$ und $u \in W(a,b;V,V')$. Dann sind die* Spuren *$u(a), u(b) \in H$ wohldefiniert.*

Beweis. Ergibt sich als direkte Folgerung aus Satz 2.10, siehe auch [DL92, Remark XVIII.2.4]. □

Ferner erhalten wir aus obiger Einbettungsaussage auch das folgende Ergebnis, welches im Rahmen der Betrachtung linearer Evolutionsgleichungen benötigt wird.

Korollar 2.12. *Seien $a,b \in \mathbb{R}$. Die Einbettungskonstante*

$$\gamma_e := \sup_{\substack{u \in W(a,b;V,V') \\ u \neq 0}} \frac{\|u(a)\|_H}{\|u\|_{W(a,b;V,V')}}$$

ist gleichmäßig beschränkt in der Wahl $V \hookrightarrow H$ und hängt nur im Fall $b \to a$ von b ab.

Beweis. Siehe [SS09, Section 5; DL92, Beweis zu Theorem XVIII.2.1]. □

Abschließen wollen wir diesen Abschnitt mit einer alternativen Charakterisierung der hier betrachteten Bochner-Räume als Tensor-Produkt, welche erst bei der numerischen Umsetzung in Kapitel 4 relevant wird.

Satz 2.13. *Seien V ein Hilbertraum und $a,b \in \mathbb{R}$ mit $a < b$. Dann ist der Bochner-Sobolev-Raum $H^m(a,b;V)$, $m \in \{0,1\}$, isometrisch isomorph zum Hilbertraum-Tensorprodukt $H^m([a,b]) \otimes V$, kurz*

$$H^m(a,b;V) \cong H^m([a,b]) \otimes V.$$

Beweis. Siehe [Aub00, Theorem 12.6.1, Theorem 12.7.1] für einen ausführlichen Nachweis. □

2.2. Lineare Evolutionsgleichungen

Nach der Einführung der benötigten Funktionenräume wenden wir uns nun den linearen Evolutionsgleichungen, einer bestimmten Unterart parabolischer partieller Differentialgleichungen, zu. Wir orientieren uns hierbei an Lions [Lio71], Schwab und Stevenson [SS09] sowie Urban und Patera [UP13], definieren den Begriff der linearen Evolutionsgleichungen, leiten eine schwache Formulierung her und weisen abschließend nach, dass diese korrekt gestellt ist.

Um den Begriff der linearen Evolutionsgleichungen definieren zu können, müssen wir zunächst die richtigen Rahmenbedingungen schaffen. Sei $V \hookrightarrow H \hookrightarrow V'$ ein Gelfand-Tripel im Sinne von Definition 2.6, das heißt, V und H seien separable Hilberträume und die Inklusionen seien jeweils dicht und stetig. Weiter seien $0 < T < \infty$ und $I := [0, T]$. Wir betrachten nun eine Familie $\{A(t) \in \mathcal{L}(V, V')\}_{t \in I}$ von stetigen linearen Operatoren. Nach dem Rieszschen Darstellungssatz, genauer siehe [Hal57, Theorem §22.1], können wir diesen Operatoren eine Familie von Bilinearformen $\{a(\cdot, \cdot; t) \in \mathcal{L}(V \times V, \mathbb{R})\}_{t \in I}$ zuordnen, wobei diese durch

$$\langle A(t)\eta, \zeta \rangle_{V' \times V} = a(\eta, \zeta; t) \quad \text{für alle } \eta, \zeta \in V$$

induziert werden.

Wie wir später sehen werden, stellt die nachfolgende Annahme sicher, dass die auf diesen Operatoren aufbauenden linearen Evolutionsgleichungen gewisse wünschenswerte Eigenschaften wie Lösbarkeit und Eindeutigkeit dieser Lösung besitzen.

Annahme 2.14. Die Familie $\{a(\cdot, \cdot; t) \in \mathcal{L}(V \times V, \mathbb{R})\}_{t \in I}$ von Bilinearformen erfülle die folgenden Eigenschaften:

(i) *Messbarkeit:* Die Abbildung $I \ni t \mapsto a(\eta, \zeta; t)$ sei messbar für alle $\eta, \zeta \in V$.

(ii) *Stetigkeit:* Es existiere ein $0 < \gamma_a < \infty$ mit

$$|a(\eta, \zeta; t)| \leq \gamma_a \|\eta\|_V \|\zeta\|_V \quad \text{für alle } \eta, \zeta \in V \text{ und fast alle } t \in I.$$

(iii) *Gårding-Ungleichung:* Es existieren $\alpha > 0$ und $\lambda \geq 0$ mit

$$a(\eta, \eta; t) + \lambda \|\eta\|_H^2 \geq \alpha \|\eta\|_V^2 \quad \text{für alle } \eta \in V \text{ und fast alle } t \in I.$$

Für den Rest dieses Abschnitts nehmen wir stets an, dass die obige Annahme erfüllt ist. Diese Vorarbeit erlaubt uns nun die nachfolgende Definition.

Definition 2.15. Seien $g \in L_2(I; V')$ ein *Quellterm* und $u_0 \in H$ ein *Anfangswert*. Als *lineare Evolutionsgleichung* bezeichnen wir die parabolische partielle Differentialgleichung

$$\begin{cases} u_t(t) + A(t)u(t) = g(t) & \text{in } V', \quad \text{für fast alle } t \in I, \\ u(0) = u_0 & \text{in } H. \end{cases} \tag{2.1}$$

Darauf aufbauend leiten wir als nächstes eine Raum-Zeit-Variationsformulierung für (2.1) her. Dazu benötigen wir geeignete Ansatz- und Testfunktionenräume, wofür die im vorherigen Abschnitt eingeführten Bochner-Räume verwendet werden.

Definition 2.16. Den *Ansatzfunktionenraum* $\mathcal{X}$ definieren wir als den Hilbertraum

$$\mathcal{X} := W(I;V,V') = L_2(I;V) \cap H^1(I;V') \tag{2.2}$$

mit dem Skalarprodukt

$$\langle u, v\rangle_{\mathcal{X}} := \langle u, v\rangle_{L_2(I;V)} + \langle u', v'\rangle_{L_2(I;V')}. \tag{2.3}$$

Der *Testfunktionenraum* $\mathcal{Y}$ sei als Produkt der beiden Hilberträume $\mathcal{Y}_1 := L_2(I;V)$ und $\mathcal{Y}_2 := H$ definiert als

$$\mathcal{Y} := \mathcal{Y}_1 \times \mathcal{Y}_2 = L_2(I;V) \times H \tag{2.4}$$

mit Skalarprodukt

$$\langle u, v\rangle_{\mathcal{Y}} := \langle u_1, v_1\rangle_{L_2(I;V)} + \langle u_2, v_2\rangle_H, \quad u = (u_1, u_2), v = (v_1, v_2) \in \mathcal{Y}. \tag{2.5}$$

Um nun aus (2.1) eine schwache Formulierung zu erhalten, „multiplizieren“ wir die lineare Evolutionsgleichung mit $v = (v_1, v_2) \in \mathcal{Y}$ und integrieren anschließend im Ort als auch über das Zeitintervall $I = [0, T]$.

Definition 2.17. Seien $\mathcal{X}$ und $\mathcal{Y}$ wie in (2.2) und (2.4), $g \in L_2(I;V')$ ein Quellterm und $u_0 \in H$ ein Anfangswert. Als *Raum-Zeit-Variationsformulierung* der linearen Evolutionsgleichung (2.1) bezeichnen wir das folgende Problem:

$$\text{Finde } u \in \mathcal{X} \text{ mit} \quad b(u, v) = f(v) \quad \text{für alle } v \in \mathcal{Y}. \tag{2.6}$$

Dabei sei die Bilinearform $b\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ durch

$$b(u, v) := \int_I [\langle u_t(t), v_1(t)\rangle_{V' \times V} + a(u(t), v_1(t); t)]\,\mathrm{d}t + \langle u(0), v_2\rangle_H \tag{2.7}$$

definiert und das Funktional $f\colon \mathcal{Y} \to \mathbb{R}$ durch

$$f(v) := \int_I \langle g(t), v_1(t)\rangle_{V' \times V}\,\mathrm{d}t + \langle u_0, v_2\rangle_H.$$

Bemerkung 2.18. Die Anfangswertbedingung $u(0) = u_0$ in H ist wegen Korollar 2.11 wohldefiniert.

Es bleibt nun zu zeigen, welche Bedingungen hinreichend sind, damit obige Raum-Zeit-Variationsformulierung *korrekt gestellt* ist. Dazu definieren wir zunächst, was wir darunter verstehen wollen und greifen dazu auf die Definition nach Hadamard [Had02] zurück.

Definition 2.19 (Hadamard). Seien X und Y zwei Hilberträume, $a \in \mathcal{L}(X \times Y, \mathbb{R})$ eine stetige Bilinearform und $f \in Y'$ ein stetiges lineares Funktional. Sei weiter ein abstraktes Variationsproblem gegeben durch:

$$\text{Finde } u \in X \text{ mit} \quad a(u, v) = f(v) \quad \text{für alle } v \in Y.$$

Wir nennen dieses *korrekt gestellt*, wenn eine eindeutige Lösung $u \in X$ existiert und diese eine a priori-Abschätzung der Form

$$\|u\|_X \le c\|f\|_{Y'} \quad \text{für alle } f \in Y'$$

mit einer von f unabhängigen Konstante $c > 0$ erfüllt.

Um dies für die Raum-Zeit-Variationsformulierung (2.6) nachzuweisen, werden wir indirekt auf den nachfolgenden wichtigen Satz zurückgreifen. Dieser findet sich in dieser oder ähnlicher Form bei Babuška [Bab71, Theorem 2.1], Babuška und Aziz [BA72, Theorem 5.2.1] und Braess [Bra13, Theorem §3.3.6].

Satz 2.20 (Banach-Nečas-Babuška-Theorem). *Seien X und Y zwei Hilberträume. Eine beschränkte lineare Abbildung $A\colon X \to Y'$ ist genau dann ein Isomorphismus, das heißt stetig invertierbar, wenn die zugehörige Bilinearform $a\colon X \times Y \to \mathbb{R}$ die folgenden Bedingungen erfüllt:*

(i) Stetigkeit: *Es existiert eine Konstante $0 < \gamma < \infty$ mit*

$$|a(u,v)| \leq \gamma \|u\|_X \|v\|_Y \quad \text{für alle } u \in X,\ v \in Y.$$

(ii) Inf-sup-Bedingung: *Es existiert eine Konstante $\beta > 0$ mit*

$$\inf_{u \in X} \sup_{v \in Y} \frac{a(u,v)}{\|u\|_X \|v\|_Y} \geq \beta.$$

(iii) Surjektivität: *Zu jedem $v \in Y$, $v \neq 0$, existiert ein $u \in X$ mit*

$$a(u,v) \neq 0.$$

Gelten die drei Bedingungen und ist weiter ein Funktional $f \in Y'$ gegeben, dann existiert eine eindeutige Lösung $\hat{u} \in X$ mit

$$a(\hat{u},v) = f(v) \quad \text{für alle } v \in Y$$

und es gilt

$$\|\hat{u}\|_X \leq \frac{1}{\beta} \|f\|_{Y'}.$$

Bemerkung 2.21. Nach [BA72, Remarks, S. 117] kann die Surjektivitätsbedingung (iii) durch eine weitere inf-sup-Bedingung ersetzt werden, denn existiert ein $\beta' > 0$ mit

$$\inf_{v \in Y} \sup_{u \in X} \frac{a(u,v)}{\|u\|_X \|v\|_Y} \geq \beta',$$

dann gilt insbesondere auch (iii).

Für das abstrakte Raum-Zeit-Variationsproblem (2.6) wurde für die hier vorliegenden Rahmenbedingungen bereits von Dautray und Lions [DL92, Section XVIII.3] nachgewiesen, dass es sich um ein korrekt gestelltes Problem handelt. Ein alternativer Beweis, welcher die Bedingungen des Banach-Nečas-Babuška-Theorems nachweist und weiter explizite Schranken für die Stetigkeitskonstante γ_b und die inf-sup-Konstante β der Bilinearform b aus (2.7) liefert, wurde von Schwab und Stevenson [SS09] geführt. Wir wiederholen die Kernaussage [SS09, Theorem 5.1] und verweisen für einen ausführlichen Beweis auf [SS09, Appendix A].

Satz 2.22. *Seien $\mathcal{X}$ und $\mathcal{Y}$ wie in* (2.2) *und* (2.4) *und sei $\{a(\cdot,\cdot;t) \in \mathcal{L}(V \times V, \mathbb{R})\}_{t\in I}$ eine Familie von Bilinearformen, welche Annahme 2.14 erfüllt. Dann ist das Raum-Zeit-Variationsproblem* (2.6) *korrekt gestellt, das heißt, für alle $f \in \mathcal{Y}'$ existiert eine eindeutige Lösung $u \in \mathcal{X}$ so, dass*

$$b(u,v) = f(v) \quad \text{für alle } v \in \mathcal{Y}$$

gilt. Ferner existiert eine von f unabhängige Konstante $\beta > 0$ mit

$$\|u\|_{\mathcal{X}} \leq \frac{1}{\beta}\|f\|_{\mathcal{Y}'}.$$

Weiter wollen wir im nachfolgenden Korollar die bereits angesprochenen Schranken für die Stetigkeitskonstante γ_b und die inf-sup-Konstante β der Bilinearform b angeben.

Korollar 2.23. *Unter den gleichen Voraussetzungen wie in Satz 2.22 und der Bedingung, dass die Bilinearformen $\{a(\cdot,\cdot;t)\}_{t\in I}$ die Gårding-Ungleichung mit $\lambda = 0$ erfüllen, gilt*

$$\gamma_b \leq \sqrt{2\max\{1,\gamma_a^2\} + \gamma_e^2}$$

und

$$\beta \geq \frac{\min\{\alpha\gamma_a^{-2}, \alpha\}}{\sqrt{2\max\{\alpha^{-2},1\} + \gamma_e^2}}.$$

Im Fall $\lambda \neq 0$ werden die Abschätzungen zu

$$\gamma_b \leq \frac{\sqrt{2\max\{1,\gamma_a^2\} + \gamma_e^2}}{\max\{\sqrt{1+2\lambda^2\rho^4}, \sqrt{2}\}}$$

und

$$\beta \geq \frac{\mathrm{e}^{-2\lambda T}}{\max\{\sqrt{1+2\lambda^2\rho^4}, \sqrt{2}\}} \cdot \frac{\min\{\alpha\gamma_a^{-2}, \alpha\}}{\sqrt{2\max\{\alpha^{-2},1\} + \gamma_e^2}}.$$

Die Größen γ_a, α und λ stammen aus Annahme 2.14, während die Konstanten ρ und γ_e, für letztere siehe auch Korollar 2.12, als die Einbettungskonstanten

$$\gamma_e := \sup_{0\neq u\in\mathcal{X}} \frac{\|u(0)\|_H}{\|u\|_{\mathcal{X}}}, \qquad \rho := \sup_{0\neq\eta\in V} \frac{\|\eta\|_H}{\|\eta\|_V}$$

definiert sind.

Beweis. Siehe [SS09, Appendix A]. □

Abschließen wollen wir dieses Kapitel mit einer Möglichkeit, das Raum-Zeit-Variationsproblem (3.6) zu einem äquivalenten Problem mit $\lambda = 0$ zu transformieren. Dazu seien erneut eine Familie von Bilinearformen $\{a(\cdot,\cdot;t)\}_{t\in I}$, welche Annahme 2.14 mit $\lambda \geq 0$ erfüllen, und weiter $u \in \mathcal{X}$, $v = (v_1, v_2) \in \mathcal{Y}$ sowie $g \in L_2(I; V')$ und $u_0 \in H$ gegeben. Wir definieren

$$\hat{u}(t) := u(t)\mathrm{e}^{-\lambda t}, \quad \check{v}_1(t) := v_1(t)\mathrm{e}^{\lambda t}, \quad \check{v} := (\check{v}_1, v_2), \quad \hat{g}(t) := g(t)\mathrm{e}^{-\lambda t}.$$

Weiter definieren wir das transformierte Variationsproblem

$$\text{Finde } \hat{u} \in \mathcal{X} \text{ mit } \quad \hat{b}(\hat{u},\check{v}) = \hat{f}(\check{v}) \quad \text{für alle } \check{v} = (\check{v}_1, v_2) \in \mathcal{Y}, \tag{2.8}$$

wobei die Bilinearform $\hat{b}\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ durch

$$\hat{b}(\hat{u}, \check{v}) := \int_I \left[\langle \hat{u}_t(t), \check{v}_1(t)\rangle_{V'\times V} + \lambda\langle \hat{u}(t), \check{v}_1(t)\rangle_H + a(\hat{u}(t), \check{v}_1(t); t)\right] \mathrm{d}t + \langle \hat{u}(0), v_2\rangle_H$$

und das Funktional $\hat{f}\colon \mathcal{Y} \to \mathbb{R}$ durch

$$\hat{f}(\check{v}) := \int_I \langle \hat{g}(t), \check{v}_1(t)\rangle_{V'\times V} \,\mathrm{d}t + \langle u_0, v_2\rangle_H$$

gegeben seien.

Proposition 2.24. *Sei $\{a(\cdot\,,\cdot\,; t) \in \mathcal{L}(V \times V, \mathbb{R})\}_{t\in I}$ eine Familie von Bilinearformen, die die Annahme 2.14 mit $\lambda \neq 0$ erfüllt. Dann ist u genau dann die Lösung des Raum-Zeit-Variationsproblems* (2.6)*, wenn $\hat{u}$ die Lösung von* (2.8) *ist. Insbesondere erfüllt die Familie $\{a(\cdot\,,\cdot\,; t) + \lambda\langle\cdot\,,\cdot\,\rangle_H \in \mathcal{L}(V \times V, \mathbb{R})\}_{t\in I}$ Annahme 2.14 mit $\lambda = 0$.*

Beweis. Wir beginnen mit der zweiten Aussage. Die Messbarkeit der neuen Familie von Bilinearformen ist direkt ersichtlich, ebenso die Stetigkeit, da $\lambda < \infty$ gilt. Auch die behauptete Gårding-Ungleichung mit $\lambda = 0$ folgt direkt aus der Konstruktion.

Die Äquivalenzaussage lässt sich unter Verwendung der Bilinearität der Skalarprodukte, der dualen Paarung und der Bilinearformen sowie Bemerkung 2.7 durch direktes nachrechnen nachweisen:

$$\begin{aligned}
\hat{b}(\hat{u}, \check{v}) &= \int_I \left[\langle \hat{u}_t(t), \check{v}_1(t)\rangle_{V'\times V} + \lambda\langle \hat{u}(t), \check{v}_1(t)\rangle_H + a(\hat{u}(t), \check{v}_1(t); t)\right] \mathrm{d}t + \langle \hat{u}(0), v_2\rangle_H \\
&= \int_I \Big[\langle (u(t)\mathrm{e}^{-\lambda t})_t, v_1(t)\mathrm{e}^{\lambda t}\rangle_{V'\times V} + \lambda\langle u(t)\mathrm{e}^{-\lambda t}, v_1(t)\mathrm{e}^{\lambda t}\rangle_H \\
&\qquad + a(u(t)\mathrm{e}^{-\lambda t}, v_1(t)\mathrm{e}^{\lambda t}; t)\Big] \,\mathrm{d}t + \langle u(0), v_2\rangle_H \\
&= \int_I \left[\langle u_t(t), v_1(t)\rangle_{V'\times V} + a(u(t), v_1(t); t)\right] \mathrm{d}t + \langle u(0), v_2\rangle_H = b(u, v).
\end{aligned}$$

Für das Funktional auf der rechten Seite ergibt sich analog

$$\begin{aligned}
\hat{f}(\check{v}) &= \int_I \langle \hat{g}(t), \check{v}_1(t)\rangle_{V'\times V} \,\mathrm{d}t + \langle u_0, v_2\rangle_H = \int_I \langle g(t)\mathrm{e}^{-\lambda t}, v_1(t)\mathrm{e}^{\lambda t}\rangle_{V'\times V} \,\mathrm{d}t + \langle u_0, v_2\rangle_H \\
&= \int_I \langle g(t), v_1(t)\rangle_{V'\times V} \,\mathrm{d}t + \langle u_0, v_2\rangle_H = f(v)
\end{aligned}$$

und damit insgesamt die Behauptung. □

3 Kapitel

Propagator-Differentialgleichung

Wir greifen nun die in Kapitel 1 eingeführten parabolischen Differentialgleichungen (1.6) und (1.7) auf, welche von den beiden Propagatoren q und $q^\dagger$ erfüllt werden. Im weiteren Verlauf der Arbeit verwenden wir den Begriff *Propagator-Differentialgleichung*, den wir noch genauer definieren werden, wenn wir uns auf die genannten partiellen Differentialgleichungen beziehen.

In diesem Kapitel konkretisieren wir zunächst diese Propagator-Differentialgleichungen, schaffen geeignete Rahmenbedingungen und leiten eine schwache Raum-Zeit-Formulierung her. Anschließend werden die in den Propagator-Differentialgleichungen auftretenden Felder ω_i parametrisiert und als Grundlage für eine parametrische schwache Formulierung verwendet. Für letztere weisen wir abschließend nach, dass sie korrekt gestellt ist und unter zusätzlichen Bedingungen eine gewisse Regularität bezüglich der Parameter aufweist.

3.1. Eine Raum-Zeit-Variationsformulierung

Wir wollen nun zunächst die aus der Einführung bekannten Propagator-Differentialgleichungen in einem allgemeineren Rahmen formulieren. Dabei halten wir an dem Fall zweier Felder fest, wobei diese Einschränkung nicht notwendig ist, da die nachfolgenden Ergebnisse in gleicher Weise auch für jede andere endliche Felderanzahl nachgewiesen werden können. Weiter sei an dieser Stelle angemerkt, dass es ausreicht sich auf die Betrachtung des Vorwärts-Propagators (1.6) zu beschränken, da der Rückwärts-Propagator (1.7) durch die einfache Transformation $s \mapsto 1 - s$ auf dieselbe Form, lediglich mit vertauschten Rollen bei den Feldern, gebracht werden kann.

Seien nun $0 < T_f < T < \infty$ reelle Konstanten und $I := [0, T]$ ein reelles Intervall, welches wir in die beiden disjunkten Teilintervalle $I_1 := [0, T_f)$ und $I_2 := [T_f, T]$ zerlegen. Wir interpretieren die Größe $t \in I$ als Zeit, wobei dies einen rein notationellen und keinen physikalischen Hintergrund hat. Weiter sei $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$, ein beschränktes Gebiet, das heißt offen, nichtleer, zusammenhängend und beschränkt, welches einen Lipschitz-Rand besitzt.

Zunächst wollen wir den Begriff der Felder und der Propagator-Differentialgleichung konkretisieren. Dies dient vor allem der einfacheren Notation und Benennung, weswegen wir den definierten Abbildungen an dieser Stelle möglichst wenige einschränkende Bedingungen auferlegen wollen.

Definition 3.1. Unter einem *Feld* verstehen wir eine $L_\infty(\Omega)$-Abbildung w. Seien w_1, w_2 Felder und χ_{I_1}, χ_{I_2} charakteristische Funktionen. Wir bezeichnen die Abbildung

$$\omega\colon I \times \Omega \to \mathbb{R}, \quad \omega(t, \mathbf{x}) := w_1(\mathbf{x})\chi_{I_1}(t) + w_2(\mathbf{x})\chi_{I_2}(t) = \begin{cases} w_1(\mathbf{x}), & t < T_f, \\ w_2(\mathbf{x}), & t \geq T_f, \end{cases} \tag{3.1}$$

als *Raum-Zeit-Feld*.

Definition 3.2. Seien ω ein Raum-Zeit-Feld wie in (3.1), $u_0\colon \Omega \to \mathbb{R}$ eine Anfangsbedingung, $g\colon I \times \Omega \to \mathbb{R}$ ein Quellterm sowie $c \in \mathbb{R}_+$ und $\mu \in \mathbb{R}$ Konstanten. Als *Propagator-Differentialgleichung* bezeichnen wir die parabolische partielle Differentialgleichung

$$\begin{cases} u_t(t, \mathbf{x}) - c\Delta u(t, \mathbf{x}) + \omega(t, \mathbf{x})u(t, \mathbf{x}) + \mu u(t, \mathbf{x}) = g(t, \mathbf{x}) & \text{auf } I \times \Omega, \\ u(0, \mathbf{x}) = u_0(\mathbf{x}) & \text{auf } \Omega, \\ u(t, \mathbf{x}) \text{ erfüllt eine vorgegebene Randbedingung} & \text{auf } I \times \partial\Omega. \end{cases} \tag{3.2}$$

Wie in der Einleitung erwähnt, hat der Mittelwert der Felder keinen Einfluss auf das Ergebnis des dort beschriebenen Iterationsverfahrens, weswegen wir den zusätzlichen Term $\mu u(t, \mathbf{x})$ einführen können. Dieser wird sich bei den folgenden theoretischen Untersuchungen und der numerischen Umsetzung als nützlich erweisen.

Unser Ziel ist es nun, eine Raum-Zeit-Variationsformulierung der Propagator-Differentialgleichung aus obiger Definition herzuleiten. Diese wird uns als Ausgangspunkt für unsere numerischen Verfahren dienen.

Zunächst schränken wir aber die möglichen Randbedingungen ein. Von größtem Interesse sind für uns, bedingt durch die Motivation der parabolischen Differentialgleichung in Kapitel 1, der Fall homogener Dirichlet-Randbedingungen, also $u(t, \mathbf{x}) = 0$ auf $I \times \partial\Omega$, sowie der Fall periodischer Randbedingungen. Letztere werden am Ende dieses Kapitels wieder aufgegriffen, während wir uns im Rest der Ausführungen auf den Fall homogener Dirichlet-Randbedingungen beschränken.

Bei der Herleitung der Raum-Zeit-Variationsformulierung der Propagator-Differentialgleichung werden wir schrittweise vorgehen und zunächst den stationären Fall betrachten, bevor wir darauf aufbauend die schwache Raum-Zeit-Variationsformulierung erhalten. Als Grundlage für die schwache Formulierung im stationären Fall verwenden wir die wohlbekannten Sobolev-Räume. Die folgende Bemerkung führt die notwendigen Notationen in diesem Zusammenhang ein.

Bemerkung 3.3. Wir schreiben kurz $V = H_0^1(\Omega)$ und $H = L_2(\Omega)$ für den bekannten Sobolev- respektive Lebesgue-Raum auf Ω. Dabei handelt es sich jeweils um einen Hilbertraum. Wir kennzeichnen durch den jeweiligen Index die entsprechenden Skalarprodukte und Normen. Als V-Norm wählen wir $\|\eta\|_V = (\|\eta\|_H^2 + \|\nabla\eta\|_H^2)^{1/2}$. Da V dicht in H eingebettet werden kann, definiert

$$V \hookrightarrow H \cong H' \hookrightarrow V' = (H_0^1(\Omega))' = H^{-1}(\Omega)$$

ein Gelfand-Tripel, vergleiche Definition 2.6. Motiviert durch Bemerkung 2.7 verwenden wir die Schreibweise $\langle \cdot, \cdot \rangle_{V' \times V}$ auch für die duale Paarung auf $V' \times V$.

Damit können wir nun die folgende Familie von Bilinearformen und die zugehörigen Operatoren definieren.

Definition 3.4. Seien ω ein Raum-Zeit-Feld sowie $c \in \mathbb{R}_+$ und $\mu \in \mathbb{R}$ Konstanten. Wir definieren für fast alle $t \in I$ eine Familie von Bilinearformen

$$a(\cdot\,,\cdot\,;t)\colon V \times V \to \mathbb{R}, \quad a(\eta,\zeta;t) := c\langle\nabla\eta,\nabla\zeta\rangle_H + \langle\omega(t,\,\cdot\,)\eta,\zeta\rangle_H + \mu\langle\eta,\zeta\rangle_H$$

und eine zugehörige Familie von Operatoren $A(t)\colon V \to V'$ via

$$\langle A(t)\eta,\zeta\rangle_{V'\times V} := a(\eta,\zeta;t), \quad \eta,\zeta \in V. \tag{3.3}$$

Bemerkung 3.5. Die Existenz des Operators $A(t)$ zur jeweiligen Bilinearform $a(\cdot\,,\cdot\,;t)$ lässt sich durch den Rieszschen Darstellungssatz begründen, siehe beispielsweise [Hal57, Theorem §22.1].

Wir wollen nun motivieren, warum die Bilinearformen so gewählt werden. Hierzu greifen wir auf die bekannte schwache Formulierung des Laplace-Operators $-\Delta\colon V \to V'$ zurück. Diese lässt sich für die vorliegende Wahl von V gerade durch die Bilinearform

$$V \times V \to \mathbb{R}, \quad (\eta,\zeta) \mapsto \langle\nabla\eta,\nabla\zeta\rangle_H$$

beschreiben. Unter Verwendung dieser ergibt sich, dass die Bilinearform $a(\cdot\,,\cdot\,;t)$ für fast alle $t \in I$ gerade die schwache Formulierung des Operators

$$A(t)\colon V \to V', \quad A(t)\eta = -\Delta\eta + \omega(t,\,\cdot\,)\eta + \mu\eta$$

ist, was gerade dem räumlichen Differentialoperator der Propagator-Differentialgleichung (3.2) entspricht.

Bemerkung 3.6. Nach Definition von ω in (3.1) gilt für fast alle $t \in I$ offenbar

$$\|\omega(t,\,\cdot\,)\|_{L_\infty(\Omega)} = \|w_1\|_{L_\infty(\Omega)}\chi_{I_1}(t) + \|w_2\|_{L_\infty(\Omega)}\chi_{I_2}(t)$$

und somit wegen der disjunkten Zerlegung $I = I_1 \cup I_2$ insbesondere

$$\|\omega\|_{L_\infty(I;L_\infty(\Omega))} = \operatorname*{ess\,sup}_{t\in I}\|\omega(t,\,\cdot\,)\|_{L_\infty(\Omega)} = \max\{\|w_1\|_{L_\infty(\Omega)},\|w_2\|_{L_\infty(\Omega)}\} < \infty.$$

Wir weisen nun zwei Eigenschaften für diesen Operator beziehungsweise die zugehörige Bilinearform nach, welche eine wichtige Rolle für die spätere Raum-Zeit-Variationsformulierung spielen werden.

Satz 3.7. *Sei $\{a(\cdot\,,\cdot\,;t)\}_{t\in I}$ die Familie von Bilinearformen aus Definition 3.4. Diese Bilinearformen erfüllen die folgenden Eigenschaften:*

(i) Messbarkeit*: Die Abbildung $I \ni t \mapsto a(\eta,\zeta;t)$ ist messbar für alle $\eta,\zeta \in V$.*

(ii) Stetigkeit: *Es gilt*

$$|a(\eta,\zeta;t)| \le \gamma_a\|\eta\|_V\|\zeta\|_V \quad \textit{für alle } \eta,\zeta \in V \textit{ und fast alle } t \in I$$

mit Stetigkeitskonstante $\gamma_a = \max\{c, \|\omega\|_{L_\infty(I;L_\infty(\Omega))} + |\mu|\} < \infty$.

(iii) Gårding-Ungleichung: *Es gilt*

$$a(\eta,\eta;t)+\lambda\|\eta\|_H^2 \geq \alpha\|\eta\|_V^2 \quad \textit{für alle } \eta\in V \textit{ und fast alle } t\in I \tag{3.4}$$

mit $\alpha = c\gamma_\Omega^2 > 0$ *und* $\lambda = \max\{\|\omega\|_{L_\infty(I;L_\infty(\Omega))}-\mu,0\}\geq 0$*, wobei* γ_Ω *die Poincaré-Friedrichs-Konstante ist.*

Beweis. Für den Nachweis der Messbarkeit stützen wir uns auf die Aussagen in [Fat05, S. 177] und [And12b, Theorem 2.7.9, Corollary 2.7.10, Lemma 8.1.1]. Demnach ist für $\xi\in L_\infty(I\times\Omega)$ und $\psi\in L_1(\Omega)$ die Abbildung $I\ni t\mapsto\langle\xi(t,\cdot),\psi\rangle_{L_2(\Omega)}$ messbar.

Wegen $\eta,\zeta\in V=H_0^1(\Omega)$ ist sowohl die Abbildung $\Omega\ni\mathbf{x}\mapsto\langle\nabla\eta(\mathbf{x}),\nabla\zeta(\mathbf{x})\rangle_H$ als auch $\Omega\ni\mathbf{x}\mapsto\eta(\mathbf{x})\zeta(\mathbf{x})$ in $L_1(\Omega)$. Es bleibt also lediglich $\omega\in L_\infty(I\times\Omega)$ zu zeigen. Dies ist aber aufgrund von $\chi_{I_1},\chi_{I_2}\in L_\infty(I)$ und $w_1,w_2\in L_\infty(\Omega)$ ebenfalls gegeben. Damit erfüllt jeder der Summanden von $a(\cdot,\cdot;t)$ und folglich auch die Bilinearform selbst die geforderte Messbarkeit.

Als nächstes zeigen wir die Stetigkeit. Seien dazu $\eta,\zeta\in V$ beliebig, dann erhalten wir unter Verwendung der Dreiecks- und der Cauchy-Schwarz-Ungleichung für beliebiges $t\in I$ die Abschätzung

$$\begin{aligned}
|a(\eta,\zeta;t)| &= |c\langle\nabla\eta,\nabla\zeta\rangle_H+\langle\omega(t,\cdot)\eta,\zeta\rangle_H+\mu\langle\eta,\zeta\rangle_H| \\
&\leq c|\langle\nabla\eta,\nabla\zeta\rangle_H|+|\langle\omega(t,\cdot)\eta,\zeta\rangle_H|+|\mu||\langle\eta,\zeta\rangle_H| \\
&\leq c\|\nabla\eta\|_H\|\nabla\zeta\|_H+\left(\|\omega\|_{L_\infty(I;L_\infty(\Omega))}+|\mu|\right)\|\eta\|_H\|\zeta\|_H \\
&\leq \max\{c,\|\omega\|_{L_\infty(I;L_\infty(\Omega))}+|\mu|\}\|\eta\|_V\|\zeta\|_V.
\end{aligned}$$

Für die Gårding-Ungleichung seien nun $\eta\in V$ und $\lambda\in\mathbb{R}$. Es gilt

$$\begin{aligned}
a(\eta,\eta;t)+\lambda\|\eta\|_H^2 &= c\|\nabla\eta\|_H^2+\langle\omega(t,\cdot)\eta,\eta\rangle_H+\mu\langle\eta,\eta\rangle_H+\lambda\langle\eta,\eta\rangle_H \\
&= c\|\nabla\eta\|_H^2+\langle(\omega(t,\cdot)+\mu+\lambda)\eta,\eta\rangle_H.
\end{aligned}$$

Wählen wir nun $\lambda:=\max\{\|\omega\|_{L_\infty(I;L_\infty(\Omega))}-\mu,0\}\geq 0$, dann gilt $\omega+\mu+\lambda\geq 0$ fast überall in Ω für fast alle $t\in I$ und wir erhalten die Abschätzung

$$a(\eta,\eta;t)+\lambda\|\eta\|_H^2\geq c\|\nabla\eta\|_H^2,$$

woraus wir durch Anwenden der Poincaré-Friedrichs-Ungleichung [Han09, Lemma 89.4]

$$a(\eta,\eta;t)+\lambda\|\eta\|_H^2\geq c\gamma_\Omega^2\|\eta\|_V^2$$

folgern können. □

Korollar 3.8. *Ist* $\lambda=0$*, insbesondere wenn* $\mu\geq\|\omega\|_{L_\infty(I;L_\infty(\Omega))}$ *gilt, dann sind die Bilinearformen* $\{a(\cdot,\cdot;t)\}_{t\in I}$ *für fast alle* $t\in I$ *elliptisch.*

Ferner wollen wir an dieser Stelle eine Aussage nachweisen, welche wir erst in Kapitel 4 benötigen werden.

Lemma 3.9. *Die Operatoren* $\{A(t)\}_{t\in I}$ *aus Definition 3.4 bilden eine Familie von selbstadjungierten Operatoren.*

Beweis. Dies kann aus der Selbstadjungiertheit des Laplace-Operators $-\Delta\colon V \to V'$ unter den gegebenen Bedingungen abgeleitet werden. Es ergibt sich

$$a(\eta, \zeta; t) = c\langle \nabla\eta, \nabla\zeta\rangle_H + \langle \omega(t,\cdot)\eta, \zeta\rangle_H + \mu\langle \eta, \zeta\rangle_H = \langle \eta, A(t)\zeta\rangle_{V\times V'},$$

insbesondere also

$$\langle A(t)\eta, \zeta\rangle_{V'\times V} = a(\eta, \zeta; t) = \langle \eta, A(t)\zeta\rangle_{V\times V'} \quad \text{für alle } \eta, \zeta \in V.$$

Dies zeigt gerade die Behauptung. □

Mit diesem Satz sind die notwendigen theoretischen Grundlagen abgeschlossen, sodass im Folgenden die Raum-Zeit-Variationsformulierung der Propagator-Differentialgleichung formuliert werden kann. Diese können wir informell durch Multiplikation der parabolischen Differentialgleichung (3.2) mit einer Raum-Zeit-Testfunktion v_1 und Integration über Ω und I sowie Addition der Anfangsbedingung, welche mit einer Raum-Testfunktion v_2 multipliziert und anschließend über Ω integriert wird, herleiten.

Um eine exakte Definition zu ermöglichen, benötigen wir zunächst zwei Raum-Zeit-Hilberträume, welche als Ansatz- respektive Testraum dienen werden. Es bietet sich an, die aus Definition 2.16 bereits bekannten Räume

$$\mathcal{X} = L_2(I; V) \cap H^1(I; V') \quad \text{und} \quad \mathcal{Y} = L_2(I; V) \times H, \tag{3.5}$$

zu verwenden.

Definition 3.10 (Schwache Formulierung). Seien $g \in L_2(I; V')$ ein Quellterm und $u_0 \in H$ eine Anfangsbedingung. Als *schwache Formulierung* oder *Raum-Zeit-Variationsformulierung* der Propagator-Differentialgleichung (3.2) bezeichnen wir das folgende Variationsproblem:

$$\text{Finde } u \in \mathcal{X} \text{ mit} \quad b(u, v) = f(v) \quad \text{für alle } v = (v_1, v_2) \in \mathcal{Y}. \tag{3.6}$$

Dabei sei die Bilinearform $b(\cdot,\cdot)\colon \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ durch

$$b(u, v) := \int_I \left[\langle u_t(t), v_1(t)\rangle_{V'\times V} + a(u(t), v_1(t); t)\right] \mathrm{d}t + \langle u(0), v_2\rangle_H$$

und das stetige lineare Funktional $f\colon \mathcal{Y} \to \mathbb{R}$ auf der rechten Seite durch

$$f(v) := \int_I \langle g(t), v_1(t)\rangle_{V'\times V} \,\mathrm{d}t + \langle u_0, v_2\rangle_H$$

gegeben.

Bemerkung 3.11. Die Linearität von f ist direkt ersichtlich; die Stetigkeit aber wollen wir an dieser Stelle nachweisen. Durch mehrfache Anwendung der Cauchy-Schwarz-Ungleichung erhalten wir

$$\begin{aligned}
f(v) &\le \int_I \|g(t)\|_{V'} \|v_1(t)\|_V \,\mathrm{d}t + \|u_0\|_H \|v_2\|_H \\
&\le \left(\int_I \|g(t)\|_{V'}^2 \,\mathrm{d}t\right)^{1/2} \left(\int_I \|v_1(t)\|_V^2 \,\mathrm{d}t\right)^{1/2} + \|u_0\|_H \|v_2\|_H \\
&= \|g\|_{L_2(I;V')} \|v_1\|_{L_2(I;V)} + \|u_0\|_H \|v_2\|_H \\
&\le \max\Big\{\|g\|_{L_2(I;V')}, \|u_0\|_H\Big\} \Big(\|v_1\|_{L_2(I;V)} + \|v_2\|_H\Big) \\
&\le \max\Big\{\|g\|_{L_2(I;V')}, \|u_0\|_H\Big\} \|v\|_{\mathcal{Y}}
\end{aligned}$$

und damit die Stetigkeit.

Der nächste Schritt ist nun, nachzuweisen, dass obige Raum-Zeit-Variationsformulierung im Sinne von Definition 2.19 korrekt gestellt ist, also eine eindeutige Lösung besitzt, welche stetig von dem Funktional $f \in \mathcal{Y}'$ abhängt. Hierzu werden wir mit Satz 2.22 ansetzen, welcher unter den gegebenen Rahmenbedingungen die zu prüfenden Eigenschaften auf die in Satz 3.7 bereits nachgewiesenen reduziert.

Korollar 3.12. *Seien Ansatz- und Testraum $\mathcal{X}$ und $\mathcal{Y}$ wie in* (3.5)*. Dann ist die Raum-Zeit-Variationsformulierung* (3.6) *korrekt gestellt.*

Beweis. Dies ist nach Definition der Raum-Zeit-Variationsformulierung eine unmittelbare Folgerung aus Satz 2.22 und Satz 3.7. □

Weiter erhalten wir als Nebenprodukt aus Korollar 2.23 auch Schranken für die Stetigkeitskonstante γ_b sowie die inf-sup-Konstante β der Bilinearform $b(\cdot, \cdot)$.

Korollar 3.13. *Seien die Voraussetzungen von Korollar 3.12 gegeben. Dann erfüllt die Bilinearform $b(\cdot, \cdot)$ die folgenden Eigenschaften:*

(i) Stetigkeit: *Es gilt*

$$\gamma_b := \sup_{u \in \mathcal{X}} \sup_{v \in \mathcal{Y}} \frac{b(u, v)}{\|u\|_{\mathcal{X}} \|v\|_{\mathcal{Y}}} < \infty.$$

(ii) inf-sup-Bedingung: *Es gilt*

$$\beta := \inf_{u \in \mathcal{X}} \sup_{v \in \mathcal{Y}} \frac{b(u, v)}{\|u\|_{\mathcal{X}} \|v\|_{\mathcal{Y}}} > 0.$$

Erfüllt die Bilinearform $a(\cdot, \cdot; t)$ die Gårding-Ungleichung (3.4) *mit $\lambda = 0$ für fast alle $t \in I$, dann gilt*

$$\gamma_b \leq \sqrt{2 \max\{1, c, (\|\omega\|_{L_\infty(I; L_\infty(\Omega))} + |\mu|)\}^2 + \gamma_e^2},$$
$$\beta \geq \frac{\gamma_\Omega^2 \min\{c, c^{-1}, c(\|\omega\|_{L_\infty(I; L_\infty(\Omega))} + |\mu|)^{-2}\}}{\sqrt{2 \max\{c^{-2} \gamma_\Omega^{-4}, 1\} + \gamma_e^2}}.$$

Ist dagegen $\lambda > 0$, dann erhalten wir stattdessen die erweiterten Abschätzungen

$$\gamma_b \leq \frac{\gamma_b'}{\max\{\sqrt{1 + 2(\|\omega\|_{L_\infty(I; L_\infty(\Omega))} - \mu)^2 \rho^4}, \sqrt{2}\}},$$
$$\beta \geq \frac{e^{-2(\|\omega\|_{L_\infty(I; L_\infty(\Omega))} - \mu)T}}{\max\{\sqrt{1 + 2(\|\omega\|_{L_\infty(I; L_\infty(\Omega))} - \mu)^2 \rho^4}, \sqrt{2}\}} \beta',$$

wobei γ_b' und β' den Größen γ_b und β des ersten Falles entsprechen. Die Größen γ_e und ρ entsprechen dabei denen aus Korollar 2.23.

Beweis. Die Schranken ergeben sich durch Einsetzen der Größen γ_a, λ und α aus Satz 3.7 in die Schranken von Korollar 2.23. □

3.2. Parametrische Formulierung

Nachdem nun eine erste schwache Formulierung der Propagator-Differentialgleichung eingeführt wurde, welche in dieser Form bereits als Grundlage für eine numerische Umsetzung verwendet werden kann, wollen wir nun als nächsten Schritt die darin auftretenden Felder möglichst niedrigdimensional parametrisieren. Dies wird vor allem durch das Iterationsverfahren aus Kapitel 1 motiviert, in dem die Propagator-Differentialgleichung immer wieder für leicht variierte Raum-Zeit-Felder ω berechnet wird.

Dazu kehren wir nun zunächst zum Operator $A(t)$ aus Definition 3.4 zurück und betrachten diesen zunächst unabhängig von der Zeit $t \in I$, aber in Abhängigkeit von einem Feld $w \in L_\infty(\Omega)$. Wir definieren für $w \in L_\infty(\Omega)$ eine Familie von Operatoren $A(w)$ als

$$A(w)\colon V \to V', \quad A(w)\eta := -c\Delta\eta + w\eta + \mu\eta. \tag{3.7}$$

Wie zuvor sei auch eine Familie von zugehörigen Bilinearformen $a(\cdot\,,\cdot\,;w)$ gegeben durch

$$a(\cdot\,,\cdot\,;w)\colon V \times V \to \mathbb{R}, \quad a(\eta,\zeta;w) := c\langle\nabla\eta,\nabla\zeta\rangle_H + \langle w\eta,\zeta\rangle_H + \mu\langle\eta,\zeta\rangle_H.$$

Um die Abhängigkeit der obigen Operatoren respektive Bilinearformen vom Feld auch für die nachfolgende numerische Umsetzung verwendbar zu machen, wollen wir diese Abhängigkeit von einer Abbildung $w \in L_\infty(\Omega)$ durch eine von einer diskreten Größe, beispielsweise einer Koeffizientenfolge aus $\ell_1(\mathbb{N})$ oder ähnlichen Folgenräumen, ersetzen. Dies erreichen wir durch folgende Einschränkung der verwendeten Felder $w \in L_\infty(\Omega)$ auf Funktionen, die wir als Reihenentwicklung von hier noch nicht näher spezifizierten Funktionen φ_i darstellen können. Auf die Wahl dieser φ_i werden wir bei der numerischen Untersuchung in Kapitel 4 erneut eingehen.

Definition 3.14. Sei $\{\varphi_i\}_{i\in\mathbb{N}} \subset L_\infty(\Omega)$ ein System von Funktionen und sei weiter ein Parameterraum $\mathcal{P} \subset \ell_\infty(\mathbb{N})$ gegeben. Wir nennen ein Feld $w \in L_\infty(\Omega)$ *darstellbar* durch $\{\varphi_i\}_{i\in\mathbb{N}}$, wenn ein $\boldsymbol{\sigma} \in \mathcal{P}$ existiert, so dass w mit

$$w(\boldsymbol{\sigma}) = \sum_{i=1}^{\infty} \sigma_i\varphi_i$$

im Sinne der gleichmäßigen Konvergenz übereinstimmt.

Wir können mit einem festen System $\{\varphi_i\}_{i\in\mathbb{N}}$ im Allgemeinen nicht alle möglichen $L_\infty(\Omega)$-Funktionen darstellen. Dies ist aber auch nicht nötig: Wie man in der Einführung bereits gesehen hat, weisen die während des Iterationsverfahrens tatsächlich auftretenden Felder gewisse Eigenschaften, beispielsweise Symmetrie oder Regularität, auf und wir können diese in die Wahl des Systems $\{\varphi_i\}_{i\in\mathbb{N}}$ einfließen lassen.

Bemerkung 3.15. Für den Rest dieses Kapitels beschränken wir uns bei der Wahl des Parameterraums auf $\mathcal{P} = [-1,1]^{\mathbb{N}}$. Dies dient hauptsächlich der Vereinfachung der Beweise und stellt keine Einschränkung dar, da die Funktionen $\{\varphi_i\}_{i\in\mathbb{N}}$ entsprechend umskaliert werden können.

Um sicherzustellen, dass derartige Felder $w(\boldsymbol{\sigma})$ wohldefinierte Operatoren $A(w(\boldsymbol{\sigma}))$ liefern, fordern wir die in der folgenden Annahme festgehaltene Eigenschaft von der Funktionenfolge $\{\varphi_i\}_{i\in\mathbb{N}}$.

Annahme 3.16. Das Funktionensystem $\{\varphi_i\}_{i\in\mathbb{N}} \subset L_\infty(\Omega)$ sei einfach summierbar in der L_∞-Norm, das heißt, es gelte $\{\|\varphi_i\|_{L_\infty(\Omega)}\}_{i\in\mathbb{N}} \in \ell_1(\mathbb{N})$.

Im Folgenden bezeichnen wir die obige $\ell_1(\mathbb{N})$-Norm der Kürze wegen als

$$c_\varphi := \sum_{i=1}^{\infty} \|\varphi_i\|_{L_\infty(\Omega)}. \tag{3.8}$$

Diese Annahme stellt insbesondere die gleichmäßige Konvergenz von $w(\boldsymbol{\sigma})$ für alle $\boldsymbol{\sigma} \in \mathcal{P}$ sicher, denn es gilt

$$\sup_{\boldsymbol{\sigma}\in\mathcal{P}} \|w(\boldsymbol{\sigma})\|_{L_\infty(\Omega)} \leq \sum_{i=1}^{\infty} \|\varphi_i\|_{L_\infty(\Omega)} = c_\varphi < \infty. \tag{3.9}$$

Legen wir uns auf ein konkretes Funktionensystem $\{\varphi_i\}_{i\in\mathbb{N}}$, welches die Annahme 3.16 erfüllt, fest, dann können wir die Operatoren $A(\omega)$ nun auch als Familie von Operatoren $A(\boldsymbol{\sigma})$ betrachten, denn durch Einsetzen von $w(\boldsymbol{\sigma})$ in (3.7) erhalten wir

$$A(\boldsymbol{\sigma})\colon V \to V', \quad A(\boldsymbol{\sigma})\eta = -c\Delta\eta + \sum_{i=1}^{\infty} \sigma_i\varphi_i\eta + \mu\eta. \tag{3.10}$$

Weiter können wir auch die zugehörige Bilinearform $a(\cdot\,,\cdot\,;\boldsymbol{\sigma})$ angeben als

$$a(\cdot\,,\cdot\,;\boldsymbol{\sigma})\colon V \times V \to \mathbb{R}, \quad a(\eta,\zeta;\boldsymbol{\sigma}) = c\langle\nabla\eta,\nabla\zeta\rangle_H + \sum_{i=1}^{\infty} \sigma_i\langle\varphi_i\eta,\zeta\rangle_H + \mu\langle\eta,\zeta\rangle_H.$$

Die bei der Bilinearform vorgenommene Vertauschung von Summe und H-Skalarprodukt ist durch Annahme 3.16 respektive Ungleichung (3.9) und den Satz von Lebesgue gerechtfertigt. Wie auch für den nicht-parametrischen Fall werden wir nachweisen, dass es sich hierbei um Operatoren beziehungsweise Bilinearformen handelt, welche die für uns wichtigen Eigenschaften der Stetigkeit und der Gültigkeit einer Gårding-Ungleichung besitzen. Zunächst wollen wir an dieser Stelle noch ein parametrisches Äquivalent der Raum-Zeit-Variationsformulierung aus Definition 3.10 formulieren.

Dies bedarf, wie zuvor, eines zeitlichen Wechsels zwischen mehreren Feldern w_j. Wir beschränken uns auf den bereits bekannten Fall zweier Felder und erweitern die obige Operator-Definition um die zeitliche Abhängigkeit. Zunächst definieren wir analog zu (3.1) ein parametrisches Raum-Zeit-Feld. Dies geschieht auf Basis der Darstellung von w_j aus Definition 3.14. Da diese den zeitlichen Wechsel noch nicht enthält, führen wir diesen durch die Aufteilung von $\boldsymbol{\sigma}$ in zwei Teilfolgen ein, um so die Notation möglichst einfach zu halten. Sei dazu $\boldsymbol{\sigma} \in \mathcal{P}$, dann definieren wir die folgenden Teilfolgen der ungeraden respektive geraden Indizes als $\boldsymbol{\sigma}_\mathrm{u} = (\sigma_{2i-1})_{i\in\mathbb{N}}$ und $\boldsymbol{\sigma}_\mathrm{g} = (\sigma_{2i})_{i\in\mathbb{N}}$.

Definition 3.17. Sei $\boldsymbol{\sigma} \in \mathcal{P}$. Unter einem *parametrischen Raum-Zeit-Feld* verstehen wir die Abbildung $\omega(\cdot\,,\cdot\,;\boldsymbol{\sigma})\colon I \times \Omega \to \mathbb{R}$, die durch

$$\begin{aligned} \omega(t,\mathbf{x};\boldsymbol{\sigma}) &:= w(\mathbf{x};\boldsymbol{\sigma}_\mathrm{u})\chi_{I_1}(t) + w(\mathbf{x};\boldsymbol{\sigma}_\mathrm{g})\chi_{I_2}(t) \\ &= \sum_{i=1}^{\infty} \left[\sigma_{2i-1}\chi_{I_1}(t) + \sigma_{2i}\chi_{I_2}(t)\right] \varphi_i(\mathbf{x}) \end{aligned} \tag{3.11}$$

gegeben ist.

Wegen der disjunkten Zerlegung $I = I_1 \cup I_2$ ist direkt ersichtlich, dass analog zu (3.9) die Abschätzung

$$\sup_{\boldsymbol{\sigma}\in\mathcal{P}} \|\omega(\boldsymbol{\sigma})\|_{L_\infty(I;L_\infty(\Omega))} = \sup_{\boldsymbol{\sigma}\in\mathcal{P}} \|w(\boldsymbol{\sigma})\|_{L_\infty(\Omega)} \leq \sum_{i=1}^{\infty} \|\varphi_i\|_{L_\infty(\Omega)} = c_\varphi \tag{3.12}$$

gilt.

Wir erweitern nun die Operator-Definition (3.10) um die Zeitabhängigkeit. Definieren wir also für $t \in I$ und $\boldsymbol{\sigma} \in \mathcal{P}$ die Operatorfamilie

$$A(t,\boldsymbol{\sigma})\colon V \to V', \quad A(t,\boldsymbol{\sigma})\eta := -c\Delta\eta + \omega(t,\,\cdot\,;\boldsymbol{\sigma})\eta + \mu\eta,$$

dann hat die zugehörige Familie von Bilinearformen $a(\,\cdot\,,\,\cdot\,;t,\boldsymbol{\sigma})\colon V \times V \to \mathbb{R}$ nach (3.11) die Form

$$a(\eta,\zeta;t,\boldsymbol{\sigma}) = c\langle\nabla\eta,\nabla\zeta\rangle_H + \sum_{i=1}^{\infty} \left[\sigma_{2i-1}\chi_{I_1}(t) + \sigma_{2i}\chi_{I_2}(t)\right] \langle\varphi_i\eta,\zeta\rangle_H + \mu\langle\eta,\zeta\rangle_H. \tag{3.13}$$

Mit dieser Vorarbeit können wir analog zu Definition 3.10 nun die folgende parametrische schwache Formulierung definieren.

Definition 3.18. Seien $g \in L_2(I;V')$ ein Quellterm und $u_0 \in H$ eine Anfangsbedingung. Als *parametrische schwache Formulierung* oder *parametrische Raum-Zeit-Variationsformulierung* der Propagator-Differentialgleichung (3.2) bezeichnen wir das folgende Variationsproblem:

$$\text{Sei } \boldsymbol{\sigma} \in \mathcal{P},\ \text{finde } u(\boldsymbol{\sigma}) \in \mathcal{X} \text{ mit} \quad b(u(\boldsymbol{\sigma}),v;\boldsymbol{\sigma}) = f(v) \quad \text{für alle } v \in \mathcal{Y}. \tag{3.14}$$

Dabei sei die Familie von Bilinearformen $b(\,\cdot\,,\,\cdot\,;\boldsymbol{\sigma})\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ gegeben durch

$$b(u,v;\boldsymbol{\sigma}) := \int_I \left[\langle u_t(t), v_1(t)\rangle_{V'\times V} + a(u(t),v_1(t);t,\boldsymbol{\sigma})\right] \mathrm{d}t + \langle u(0),v_2\rangle_H, \tag{3.15}$$

wobei $a(\,\cdot\,,\,\cdot\,;t,\boldsymbol{\sigma})$ wie in (3.13) definiert sei. Das stetige lineare Funktional $f\colon \mathcal{Y} \to \mathbb{R}$ sei wie zuvor

$$f(v) := \int_I \langle g(t), v_1(t)\rangle_{V'\times V}\,\mathrm{d}t + \langle u_0, v_2\rangle_H.$$

Hierfür weisen wir nun erneut nach, dass die schwache Formulierung korrekt gestellt ist. Wie zuvor wollen wir Satz 2.22 verwenden.

Satz 3.19. *Sei $a(\,\cdot\,,\,\cdot\,;t;\boldsymbol{\sigma})$, $t \in I$ und $\boldsymbol{\sigma} \in \mathcal{P}$, die Familie von Bilinearformen aus* (3.13)*. Dann gelten für alle $\boldsymbol{\sigma} \in \mathcal{P}$ die folgenden Eigenschaften:*

(i) Messbarkeit*: Die Abbildung $I \ni t \mapsto a(\eta,\zeta;t,\boldsymbol{\sigma})$ ist messbar für alle $\eta,\zeta \in V$.*

(ii) Stetigkeit: *Es gilt*

$$|a(\eta,\zeta;t;\boldsymbol{\sigma})| \leq \gamma_a \|\eta\|_V \|\zeta\|_V \quad \textit{für alle } \eta,\zeta \in V \textit{ und fast alle } t \in I$$

mit Stetigkeitskonstante $\gamma_a = \max\{c, c_\varphi + |\mu|\} < \infty$.

(iii) Gårding-Ungleichung: *Es gilt*

$$a(\eta, \eta; t; \boldsymbol{\sigma}) + \lambda \|\eta\|_H^2 \geq \alpha \|\eta\|_V^2 \quad \textit{für alle } \eta \in V \textit{ und fast alle } t \in I$$

mit $\alpha = c\gamma_\Omega^2 > 0$ *und* $\lambda = \max\{c_\varphi - \mu, 0\} \geq 0$.

Dabei ist c_φ *die Konstante aus* (3.8) *und die Konstanten* γ_a, λ *und* α *sind sowohl unabhängig von* $t \in I$ *als auch von* $\boldsymbol{\sigma} \in \mathcal{P}$.

Beweis. Der Nachweis erfolgt analog zum nicht-parametrischen Fall in Satz 3.7. Wir müssen zusätzlich lediglich die $L_\infty(\Omega)$-Norm von $\omega(t, \cdot; \boldsymbol{\sigma})$ mit Hilfe von (3.12) weiter durch c_φ abschätzen. □

Korollar 3.20. *Die parametrische schwache Formulierung* (3.14) *ist für alle* $\boldsymbol{\sigma} \in \mathcal{P}$ *korrekt gestellt. Ferner existieren analog zu Korollar 3.13 Schranken für die Stetigkeitskonstante* γ_b *und die inf-sup-Konstante* β, *welche unabhängig von* $\boldsymbol{\sigma} \in \mathcal{P}$ *sind.*

3.3. Regularität bezüglich der Parameter

Wir interessieren uns nun für die Regularität der Abhängigkeit der Lösung $u(\boldsymbol{\sigma})$ der parametrischen schwachen Formulierung vom Parameter $\boldsymbol{\sigma} \in \mathcal{P}$. Konkret werden wir nachweisen, dass die Lösung unter gewissen Annahmen an das Funktionensystem $\{\varphi_i\}_{i\in\mathbb{N}}$ analytisch vom Parameter $\boldsymbol{\sigma}$ abhängt. Diese Eigenschaft ist wünschenswert, da hierdurch die Anwendung der Reduzierte-Basis-Methode motiviert werden kann. Dies wird im Rahmen der Einführung dieser in Kapitel 5 etwas genauer beleuchtet.

In diesem Abschnitt orientieren wir uns an den Arbeiten von Cohen, DeVore und Schwab [CDS10; CDS11] sowie Kunoth und Schwab [KS13], weisen die Regularität aber direkt für die Raum-Zeit-Variationsformulierung nach, statt wie in den genannten Arbeiten den Umweg über den stationären Fall zu gehen.

Wir beginnen mit einigen notationellen Vorbemerkungen.

Bemerkung 3.21. Die Multiindexmenge $\mathcal{F}$ definieren wir als $\mathcal{F} := \{\boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}} \mid |\boldsymbol{\nu}| < \infty\}$, wobei

$$|\boldsymbol{\nu}| := \sum_{i=1}^{\infty} \nu_i$$

die $\ell_1(\mathbb{N})$-Norm sei. Anders formuliert, besteht $\mathcal{F}$ gerade aus denjenigen Folgen in $\mathbb{N}_0$, welche nur endliche viele Einträge ungleich Null aufweisen.

Seien $\boldsymbol{\nu} \in \mathcal{F}$ und $\mathbf{b} \in \ell_p(\mathbb{N})$, $p > 0$, dann schreiben wir

$$\mathbf{b}^{\boldsymbol{\nu}} := \prod_{i=1}^{\infty} b_i^{\nu_i}$$

mit der Konvention $0^0 = 1$. Wegen $|\boldsymbol{\nu}| < \infty$ ist dieses Produkt stets endlich.

Um die Notation für die nachfolgenden Beweise zu vereinfachen, ordnen wir die Darstellung der parametrischen Raum-Zeit-Felder um.

Bemerkung 3.22. Definiere neue charakteristische Funktionen und Entwicklungsfunktionen für $i \in \mathbb{N}$ durch

$$\tilde{\chi}_i := \begin{cases} \chi_{I_1}, & i \text{ ungerade}, \\ \chi_{I_2}, & i \text{ gerade}, \end{cases} \qquad \tilde{\varphi}_i := \begin{cases} \varphi_{(i+1)/2}, & i \text{ ungerade}, \\ \varphi_{i/2}, & i \text{ gerade}. \end{cases}$$

Damit können wir (3.11) auch schreiben als

$$w(t, \mathbf{x}; \boldsymbol{\sigma}) = \sum_{i=1}^{\infty} \sigma_i \tilde{\chi}_i(t) \tilde{\varphi}_i(\mathbf{x}).$$

Wir fixieren nun die rechte Seite $f \in \mathcal{Y}'$ der schwachen Formulierungen (3.6) und (3.14) und beginnen den Nachweis der Regularität mit einer Stabilitätsaussage, welche in den folgenden Beweisen nützlich sein wird.

Lemma 3.23. *Seien $\omega, \tilde{\omega}$ zwei Raum-Zeit-Felder wie in* (3.1) *und $u, \tilde{u}$ die zugehörigen Lösungen der schwachen Formulierung* (3.6)*. Dann gilt*

$$\|u - \tilde{u}\|_{\mathcal{X}} \leq \frac{\|f\|_{\mathcal{Y}'}}{\beta^2} \|\omega - \tilde{\omega}\|_{L_\infty(I; L_\infty(\Omega))},$$

wobei β eine feldunabhängige inf-sup-Konstante ist.

Beweis. Wir vernachlässigen der Kürze wegen im Folgenden die explizite Angabe der Zeitabhängigkeit der jeweiligen Funktionen. Weiter setzen wir $\vartheta = u - \tilde{u}$. Subtraktion der Variationsformulierung für die beiden Lösungen u und $\tilde{u}$ liefert für beliebige Testfunktionen $v = (v_1, v_2) \in \mathcal{Y}$ die Gleichung

$$\begin{aligned} 0 &= f(v) - f(v) = b(u, v; \omega) - b(\tilde{u}, v; \tilde{\omega}) \\ &= \int_I [\langle u_t - \tilde{u}_t, v_1 \rangle_{V' \times V} + a(u, v; \omega) - a(\tilde{u}, v; \tilde{\omega})] \, \mathrm{d}t + \langle u(0) - \tilde{u}(0), v_2 \rangle_H \\ &= \int_I [\langle \vartheta_t, v_1 \rangle_{V' \times V} + c\langle \nabla \vartheta, \nabla v_1 \rangle_H + \mu \langle \vartheta, v_1 \rangle_H + \langle \omega u - \tilde{\omega} \tilde{u}, v_1 \rangle_H] \, \mathrm{d}t + \langle \vartheta(0), v_2 \rangle_H \\ &= \int_I [\langle \vartheta_t, v_1 \rangle_{V' \times V} + a(\vartheta, v; \omega)] \, \mathrm{d}t + \langle \vartheta(0), v_2 \rangle_H + \int_I \langle (w - \tilde{w}) \tilde{u}, v_1 \rangle_H \, \mathrm{d}t \\ &= b(\vartheta, v; \omega) + \int_I \langle (w - \tilde{w}) \tilde{u}, v_1 \rangle_H \, \mathrm{d}t, \end{aligned}$$

welche wir auch als

$$b(\vartheta, v; \omega) = h(v) \tag{3.16}$$

mit der Abbildung

$$h \colon \mathcal{Y} \to \mathbb{R}, \quad h(v) := -\int_I \langle (w - \tilde{w}) \tilde{u}, v_1 \rangle_H \, \mathrm{d}t$$

auffassen können. Die Linearität von h ist klar. Wir weisen nun die Stetigkeit nach, betrachten also

$$\|h\|_{\mathcal{Y}'} = \sup_{\|v\|_{\mathcal{Y}} = 1} \Big| \int_I \langle (w - \tilde{w}) \tilde{u}, v_1 \rangle_H \, \mathrm{d}t \Big|$$

$$
\begin{aligned}
&\leq \|w - \tilde{w}\|_{L_\infty(I;L_\infty(\Omega))} \sup_{\|v\|_{\mathcal{Y}}=1} |\langle \tilde{u}, v_1 \rangle_{L_2(I;H)}| \\
&\leq \|w - \tilde{w}\|_{L_\infty(I;L_\infty(\Omega))} \sup_{\|v\|_{\mathcal{Y}}=1} \|\tilde{u}\|_{L_2(I;H)} \|v_1\|_{L_2(I;H)} \\
&\leq \|w - \tilde{w}\|_{L_\infty(I;L_\infty(\Omega))} \sup_{\|v\|_{\mathcal{Y}}=1} \|\tilde{u}\|_{\mathcal{X}} \|v\|_{\mathcal{Y}} \\
&= \|w - \tilde{w}\|_{L_\infty(I;L_\infty(\Omega))} \|\tilde{u}\|_{\mathcal{X}} < \infty.
\end{aligned}
$$

Damit ist $h \in \mathcal{Y}'$, wir können (3.16) also selbst als Raum-Zeit-Variationsproblem der Form (3.6) auffassen und erhalten damit durch Korollar 3.12 die Abschätzung

$$\|\vartheta\|_{\mathcal{X}} \leq \frac{1}{\beta} \|h\|_{\mathcal{Y}'}.$$

Wenden wir Korollar 3.12 weiter auf die schwache Formulierung zu $\tilde{\omega}$ an, dann gilt

$$\|\tilde{u}\|_{\mathcal{X}} \leq \frac{1}{\beta} \|f\|_{\mathcal{Y}'}.$$

Die Unabhängigkeit von β von den Feldern kann durch das Minimum der beiden Feld-abhängigen inf-sup-Konstanten sichergestellt werden.

Durch Zusammenfassen dieser drei Abschätzungen erhalten wir die Behauptung

$$\|u - \tilde{u}\|_{\mathcal{X}} = \|\vartheta\|_{\mathcal{X}} \leq \frac{\|f\|_{\mathcal{Y}'}}{\beta^2} \|\omega - \tilde{\omega}\|_{L_\infty(I;L_\infty(\Omega))}. \qquad \square$$

Als erster Schritt des Regularitätsnachweises wird zunächst die Existenz beliebiger partieller Ableitungen gezeigt, bevor dann nachfolgend schrittweise die Konvergenz der Taylorreihe der Lösung $u(\boldsymbol{\sigma}) \in \mathcal{X}$ nachgewiesen wird.

Satz 3.24. *Die Abbildung $\mathcal{P} \ni \boldsymbol{\sigma} \mapsto u(\boldsymbol{\sigma}) \in \mathcal{X}$, welche einem Parameter $\boldsymbol{\sigma}$ die eindeutige Lösung $u(\boldsymbol{\sigma})$ der parametrischen schwachen Formulierung* (3.14) *zuordnet, besitzt für alle $\boldsymbol{\nu} \in \mathcal{F}$ eine partielle Ableitung $\partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u(\boldsymbol{\sigma})$.*

Beweis. Erneut verzichten wir auf die explizite Notation der Zeitabhängigkeit. Wir beschränken uns darauf, die Behauptung exemplarisch für die partiellen Ableitungen erster Ordnung für ein festes $\boldsymbol{\sigma} \in \mathcal{P}$ nachzuweisen. Ohne Einschränkung sei nun $\boldsymbol{\nu} = \mathbf{e}_j \in \mathcal{F}$ für ein $j \in \mathbb{N}$ und ferner sei $h \in \mathbb{R}$. Wir definieren $\boldsymbol{\sigma}_h := \boldsymbol{\sigma} + h\boldsymbol{\nu} = \boldsymbol{\sigma} + h\mathbf{e}_j$ und

$$\vartheta_h := \frac{u(\boldsymbol{\sigma}_h) - u(\boldsymbol{\sigma})}{h},$$

wobei $u(\,\cdot\,)$ die Lösung der parametrischen schwachen Formulierung (3.14) für die entsprechenden Parameter ist. Ist $|h|$ klein genug, dann existieren diese auch im Fall $\boldsymbol{\sigma}_h \notin \mathcal{P}$, da analog zu (3.12) die schwache Formulierung nach wie vor korrekt gestellt ist. Dies ergibt sich durch die Abschätzung

$$\|\omega(\boldsymbol{\sigma}_h)\|_{L_\infty(I;L_\infty(\Omega))} \leq \sum_{i=1}^{\infty} \|\varphi_i\|_{L_\infty(\Omega)} + |h| \|\varphi_j\|_{L_\infty(\Omega)} \leq c_\varphi + |h| \|\varphi_j\|_{L_\infty(\Omega)} < \infty.$$

Zunächst schreiben wir die Differenz der parametrischen Raum-Zeit-Felder um zu

$$\omega(t,\mathbf{x};\boldsymbol{\sigma}_h)-\omega(t,\mathbf{x};\boldsymbol{\sigma})=\sum_{i=1}^{\infty}(\sigma_{h,i}-\sigma_i)\tilde{\chi}_i(t)\tilde{\varphi}_i(\mathbf{x})=h\tilde{\chi}_j(t)\tilde{\varphi}_j(\mathbf{x}).$$

Unter diesen Gegebenheiten betrachten wir nun die Differenz der zu $u(\boldsymbol{\sigma}_h)$ und $u(\boldsymbol{\sigma})$ zugehörigen Variationsprobleme. Für $v=(v_1,v_2)\in\mathcal{Y}$ gilt dann:

$$\begin{aligned}
0&=b(u(\boldsymbol{\sigma}_h),v;\boldsymbol{\sigma}_h)-b(u(\boldsymbol{\sigma}),v;\boldsymbol{\sigma})\\
&=\int_I\left[\langle u_t(\boldsymbol{\sigma}_h)-u_t(\boldsymbol{\sigma}),v_1\rangle_{V'\times V}+a(u(\boldsymbol{\sigma}_h),v_1;\boldsymbol{\sigma}_h)-a(u(\boldsymbol{\sigma}),v_1;\boldsymbol{\sigma})\right]\mathrm{d}t\\
&\quad+\langle u(0;\boldsymbol{\sigma}_h)-u(0;\boldsymbol{\sigma}),v_2\rangle_H\\
&=h\int_I\left[\langle(\vartheta_h)_t,v_1\rangle_{V'\times V}+c\langle\nabla\vartheta_h,\nabla v_1\rangle_H+\mu\langle\vartheta_h,v_1\rangle_H\right]\mathrm{d}t\\
&\quad+\int_I\left[\langle\omega(\boldsymbol{\sigma}_h)u(\boldsymbol{\sigma}_h),v_1\rangle_H-\langle\omega(\boldsymbol{\sigma})u(\boldsymbol{\sigma}),v_1\rangle_H\right]\mathrm{d}t+h\langle\vartheta_h(0),v_2\rangle_H\\
&=h\int_I\left[\langle(\vartheta_h)_t,v_1\rangle_{V'\times V}+a(\vartheta_h,v_1;\boldsymbol{\sigma})\right]\mathrm{d}t+h\langle\vartheta_h(0),v_2\rangle_H\\
&\quad+\int_I\langle(\omega(\boldsymbol{\sigma}_h)-\omega(\boldsymbol{\sigma}))u(\boldsymbol{\sigma}_h),v_1\rangle_H\,\mathrm{d}t\\
&=h\cdot b(\vartheta_h,v;\boldsymbol{\sigma})+h\int_I\tilde{\chi}_j\langle\tilde{\varphi}_ju(\boldsymbol{\sigma}_h),v_1\rangle_H\,\mathrm{d}t.
\end{aligned}$$

Dies schreiben wir erneut in Form der Gleichung

$$b(\vartheta_h,v;\boldsymbol{\sigma})=F_h(v)\tag{3.17}$$

mit der Abbildung

$$F_h\colon\mathcal{Y}\to\mathbb{R},\quad F_h(v):=-\int_I\tilde{\chi}_j\langle\tilde{\varphi}_ju(\boldsymbol{\sigma}_h),v_1\rangle_H\,\mathrm{d}t.$$

Vollkommen analog zum Beweis des vorherigen Lemmas kann gezeigt werden, dass F_h ein stetiges lineares Funktional auf $\mathcal{Y}$ definiert, das heißt, ϑ_h ist die eindeutige Lösung des Variationsproblems (3.17). Weiter ist $h\mapsto F_h(\,\cdot\,)$ stetig in $h=0$, denn für festes $v\in\mathcal{Y}$ gilt unter Verwendung der Cauchy-Schwarz-Ungleichung die Abschätzung

$$\begin{aligned}
|F_h(v)-F_0(v)|&=\Big|\int_I\tilde{\chi}_j\langle\tilde{\varphi}_j(u(\boldsymbol{\sigma}_h)-u(\boldsymbol{\sigma})),v_1\rangle_H\,\mathrm{d}t\Big|\\
&\le\|\tilde{\varphi}_j\|_{L_\infty(\Omega)}|\langle u(\boldsymbol{\sigma}_h)-u(\boldsymbol{\sigma}),v_1\rangle_{L_2(I;H)}|\\
&\le\|\tilde{\varphi}_j\|_{L_\infty(\Omega)}\|u(\boldsymbol{\sigma}_h)-u(\boldsymbol{\sigma})\|_{L_2(I;H)}\|v_1\|_{L_2(I;H)}\\
&\le\|\tilde{\varphi}_j\|_{L_\infty(\Omega)}\|u(\boldsymbol{\sigma}_h)-u(\boldsymbol{\sigma})\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}.
\end{aligned}$$

Hier setzen wir mit der Stabilitätsaussage aus Lemma 3.23 an, um $\|u(\boldsymbol{\sigma}_h)-u(\boldsymbol{\sigma})\|_{\mathcal{X}}$ weiter abzuschätzen und erhalten

$$\begin{aligned}
\|u(\boldsymbol{\sigma}_h)-u(\boldsymbol{\sigma})\|_{\mathcal{X}}&\le\frac{\|f\|_{\mathcal{Y}'}}{\beta^2}\|\omega(\boldsymbol{\sigma}_h)-\omega(\boldsymbol{\sigma})\|_{L_\infty(I;L_\infty(\Omega))}=\frac{\|f\|_{\mathcal{Y}'}}{\beta^2}\|h\tilde{\chi}_j\tilde{\varphi}_j\|_{L_\infty(I;L_\infty(\Omega))}\\
&\le\frac{\|f\|_{\mathcal{Y}'}}{\beta^2}|h|\|\tilde{\varphi}_j\|_{L_\infty(\Omega)}.
\end{aligned}$$

Zusammen mit obiger Ungleichung liefert dies

$$|F_h(v) - F_0(v)| \leq \|\tilde{\varphi}_j\|^2_{L_\infty(\Omega)} \|v\|_{\mathcal{Y}} \frac{\|f\|_{\mathcal{Y}'}}{\beta^2} |h| \to 0 \quad \text{für } h \to 0.$$

Das bedeutet, dass $F_h \to F_0$ in $\mathcal{Y}'$ für $h \to 0$ gilt, was insbesondere $\vartheta_h \to \vartheta_0$ in $\mathcal{X}$ für $h \to 0$ impliziert, da ϑ_h als Lösung des Variationsproblems (3.17) nach Korollar 3.20 stetig von F_h abhängt. Ferner ist durch $\partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u(\boldsymbol{\sigma}) = \vartheta_0$ als Lösung von

$$\text{Finde } \vartheta_0 \in \mathcal{X} \text{ mit } \quad b(\vartheta_0, v; \boldsymbol{\sigma}) = -\int_I \tilde{\chi}_j \langle \tilde{\varphi}_j u(\boldsymbol{\sigma}), v_1 \rangle_H \, \mathrm{d}t \quad \text{für alle } v \in \mathcal{Y}; \tag{3.18}$$

die Existenz und Wohldefiniertheit der gesuchten partiellen Ableitung gegeben.

Die Ableitungen höherer Ordnung lassen sich auf gleiche Weise durch Anwendung der beschriebenen Schritte auf die Variationsformulierung (3.18) et cetera konstruieren. □

Bemerkung 3.25. Sei erneut ohne Einschränkung $\boldsymbol{\nu} = \mathbf{e}_j \in \mathcal{F}$ für ein $j \in \mathbb{N}$. Alternativ erhält man das Variationsproblem (3.18) für die partielle Ableitung $\partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u(\boldsymbol{\sigma})$ auch durch formales Differenzieren der Variationsformulierung (3.14) nach σ_j, denn es gilt

$$\begin{aligned}
\partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} b(u(\boldsymbol{\sigma}), v; \boldsymbol{\sigma}) &= \partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} \Bigg(\int_I [\langle u_t(\boldsymbol{\sigma}), v_1 \rangle_{V' \times V} + c \langle \nabla u(\boldsymbol{\sigma}), \nabla v_1 \rangle_H + \mu \langle u(\boldsymbol{\sigma}), v_1 \rangle_H \\
&\qquad + \langle \omega(\boldsymbol{\sigma}) u(\boldsymbol{\sigma}), v_1 \rangle_H] \, \mathrm{d}t + \langle u(0; \boldsymbol{\sigma}), v_2 \rangle_H \Bigg) \\
&= \int_I [\langle \partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u_t(\boldsymbol{\sigma}), v_1 \rangle_{V' \times V} + \langle \nabla \partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u(\boldsymbol{\sigma}), \nabla v_1 \rangle_H + \mu \langle \partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u(\boldsymbol{\sigma}), v_1 \rangle_H \\
&\qquad + \langle \partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} \omega(\boldsymbol{\sigma}) u(\boldsymbol{\sigma}) + \omega(\boldsymbol{\sigma}) \partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u(\boldsymbol{\sigma}), v_1 \rangle_H] \, \mathrm{d}t + \langle \partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u(0; \boldsymbol{\sigma}), v_2 \rangle_H \\
&= b(\partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u(\boldsymbol{\sigma}), v; \boldsymbol{\sigma}) + \int_I \langle \partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} \omega(\boldsymbol{\sigma}) u(\boldsymbol{\sigma}), v_1 \rangle_H \, \mathrm{d}t \\
&= b(\partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u(\boldsymbol{\sigma}), v; \boldsymbol{\sigma}) + \int_I \tilde{\chi}_j \langle \tilde{\varphi}_j u(\boldsymbol{\sigma}), v_1 \rangle_H \, \mathrm{d}t
\end{aligned}$$

und ferner $\partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} f(v) = 0$. Daraus erhält man insgesamt erneut das Variationsproblem (3.18).

Satz 3.26. *Sei $\mathbf{b} = (b_i)_{i \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ die durch $b_i = \beta^{-1} \|\tilde{\varphi}_i\|_{L_\infty(\Omega)}$ gegebene Folge, wobei β die nach Korollar 3.12 existierende parameterunabhängige inf-sup-Konstante ist. Dann gilt*

$$\sup_{\boldsymbol{\sigma} \in \mathcal{P}} \|\partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u(\boldsymbol{\sigma})\|_{\mathcal{X}} \leq \frac{\|f\|_{\mathcal{Y}'}}{\beta} |\boldsymbol{\nu}|! \mathbf{b}^{\boldsymbol{\nu}} \quad \textit{für alle } \boldsymbol{\nu} \in \mathcal{F}. \tag{3.19}$$

Beweis. Wir beginnen damit, eine Darstellung der Variationsprobleme, welche von den partiellen Ableitungen erfüllt werden, herzuleiten. Diese lassen sich durch

$$b(\partial^{\boldsymbol{\nu}}_{\boldsymbol{\sigma}} u(\boldsymbol{\sigma}), v; \boldsymbol{\sigma}) = -\sum_{\{j \mid \nu_j \neq 0\}} \nu_j \int_I \tilde{\chi}_j \langle \tilde{\varphi}_j \partial^{\boldsymbol{\nu} - \mathbf{e}_j}_{\boldsymbol{\sigma}} u(\boldsymbol{\sigma}), v_1 \rangle_H \, \mathrm{d}t \tag{3.20}$$

rekursiv darstellen, was wir im Folgenden induktiv zeigen werden.

Den Fall $|\boldsymbol{\nu}| = 1$ haben wir in Bemerkung 3.25 bereits gezeigt. Sei nun also $|\boldsymbol{\nu}| > 1$. Sei weiter $k \in \mathbb{N}$ ein Index mit $\nu_k > 0$. Dann definieren wir $\tilde{\boldsymbol{\nu}} := \boldsymbol{\nu} - \mathbf{e}_k$ und es gilt offenbar $|\tilde{\boldsymbol{\nu}}| = |\boldsymbol{\nu}| - 1$. Nach Induktionsvoraussetzung gilt damit

$$b(\partial_{\boldsymbol{\sigma}}^{\tilde{\nu}} u(\boldsymbol{\sigma}), v; \boldsymbol{\sigma}) + \sum_{\{j|\tilde{\nu}_j \neq 0\}} \tilde{\nu}_j \int_I \tilde{\chi}_j \langle \tilde{\varphi}_j \partial_{\boldsymbol{\sigma}}^{\tilde{\boldsymbol{\nu}} - \mathbf{e}_j} u(\boldsymbol{\sigma}), v_1 \rangle_H \, \mathrm{d}t = 0,$$

wobei nach Definition $\nu_j = \tilde{\nu}_j$ für $j \neq k$ und $\tilde{\nu}_k = \nu_k - 1$ ist. Partielles Differenzieren dieser Gleichung nach σ_k analog zu Bemerkung 3.25 liefert dann die Gleichung

$$\begin{aligned} 0 = {} & b(\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}} u(\boldsymbol{\sigma}), v; \boldsymbol{\sigma}) \\ & + \int_I \tilde{\chi}_k \langle \tilde{\varphi}_k \partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu} - \mathbf{e}_k} u(\boldsymbol{\sigma}), v_1 \rangle_H \, \mathrm{d}t + (\nu_k - 1) \int_I \tilde{\chi}_k \langle \tilde{\varphi}_k \partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu} - \mathbf{e}_k} u(\boldsymbol{\sigma}), v_1 \rangle_H \, \mathrm{d}t \\ & + \sum_{\{j \neq k | \nu_j \neq 0\}} \nu_j \int_I \tilde{\chi}_j \langle \tilde{\varphi}_j \partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu} - \mathbf{e}_j} u(\boldsymbol{\sigma}), v_1 \rangle_H \, \mathrm{d}t, \end{aligned}$$

welche nach Zusammenfassen der Summanden Gleichung (3.20) entspricht.

Für die rechte Seite von (3.20) können wir wie zuvor nachweisen, dass es sich um ein stetiges lineares Funktional auf $\mathcal{Y}$ handelt. Wir können also Korollar 3.12 verwenden, um die Abschätzung

$$\|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}} u(\boldsymbol{\sigma})\|_{\mathcal{X}} \leq \frac{1}{\beta} \Big\| \sum_{\{j|\nu_j \neq 0\}} \nu_j \int_I \tilde{\chi}_j \langle \tilde{\varphi}_j \partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu} - \mathbf{e}_j} u(\boldsymbol{\sigma}), v_1 \rangle_H \, \mathrm{d}t \Big\|_{\mathcal{Y}'}$$

zu erhalten. Wir wollen nun die $\mathcal{Y}'$-Norm der rechten Seite weiter abschätzen. Dazu verwenden wir erneut die Cauchy-Schwarz-Ungleichung und erhalten wegen

$$\begin{aligned} & \Big| \sum_{\{j|\nu_j \neq 0\}} \nu_j \int_I \tilde{\chi}_j \langle \tilde{\varphi}_j \partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu} - \mathbf{e}_j} u(\boldsymbol{\sigma}), v_1 \rangle_H \, \mathrm{d}t \Big| \\ \leq {} & \sum_{\{j|\nu_j \neq 0\}} \nu_j \Big| \int_I \tilde{\chi}_j \langle \tilde{\varphi}_j \partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu} - \mathbf{e}_j} u(\boldsymbol{\sigma}), v_1 \rangle_H \, \mathrm{d}t \Big| \\ \leq {} & \sum_{\{j|\nu_j \neq 0\}} \nu_j \|\tilde{\varphi}_j\|_{L_\infty(\Omega)} \|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu} - \mathbf{e}_j} u(\boldsymbol{\sigma})\|_{L_2(I;H)} \|v_1\|_{L_2(I;H)} \\ \leq {} & \sum_{\{j|\nu_j \neq 0\}} \nu_j \|\tilde{\varphi}_j\|_{L_\infty(\Omega)} \|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu} - \mathbf{e}_j} u(\boldsymbol{\sigma})\|_{\mathcal{X}} \|v\|_{\mathcal{Y}}, \end{aligned}$$

weiter die Abschätzung

$$\|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}} u(\boldsymbol{\sigma})\|_{\mathcal{X}} \leq \sum_{\{j|\nu_j \neq 0\}} \nu_j \frac{\|\tilde{\varphi}_j\|_{L_\infty(\Omega)}}{\beta} \|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu} - \mathbf{e}_j} u(\boldsymbol{\sigma})\|_{\mathcal{X}}. \tag{3.21}$$

Um nun die eigentliche Behauptung zu beweisen, verfolgen wir erneut einen Induktionsansatz. Sei zunächst $|\boldsymbol{\nu}| = 0$. Dann entspricht

$$\sup_{\boldsymbol{\sigma} \in \mathcal{P}} \|u(\boldsymbol{\sigma})\|_{\mathcal{X}} \leq \frac{\|f\|_{\mathcal{Y}'}}{\beta}$$

Ungleichung (3.19) und ist nach Korollar 3.12 erfüllt. Sei also nun $|\boldsymbol{\nu}| > 0$. Dann gilt für die rekursive Darstellung (3.21) unter Verwendung der Induktionsvoraussetzung (3.19) für $\|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}-\mathbf{e}_j} u(\boldsymbol{\sigma})\|_{\mathcal{X}}$ die Abschätzung

$$\begin{aligned}
\|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}} u(\boldsymbol{\sigma})\|_{\mathcal{X}} &\leq \sum_{\{j|\nu_j \neq 0\}} \nu_j b_j \|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}-\mathbf{e}_j} u(\boldsymbol{\sigma})\|_{\mathcal{X}} \\
&\leq \sum_{\{j|\nu_j \neq 0\}} \nu_j b_j \frac{\|f\|_{\mathcal{Y}'}}{\beta} |\boldsymbol{\nu} - \mathbf{e}_j|! \mathbf{b}^{\boldsymbol{\nu}-\mathbf{e}_j} \\
&= \Big(\sum_{\{j|\nu_j \neq 0\}} \nu_j \Big) \Big(\frac{\|f\|_{\mathcal{Y}'}}{\beta} (|\boldsymbol{\nu}| - 1)! \mathbf{b}^{\boldsymbol{\nu}} \Big) \\
&= \frac{\|f\|_{\mathcal{Y}'}}{\beta} |\boldsymbol{\nu}|! \mathbf{b}^{\boldsymbol{\nu}}
\end{aligned}$$

und damit die Behauptung. □

Bevor zusätzliche Annahmen notwendig werden, um die benötigten Aussagen zu beweisen, wird an dieser Stelle zunächst erläutert, wie die obigen Aussagen zum Nachweis der analytischen Abhängigkeit der Lösung vom Parameter beitragen.

Definition 3.27. Wir nennen die Abbildung $\mathcal{P} \ni \boldsymbol{\sigma} \mapsto u(\boldsymbol{\sigma}) \in \mathcal{X}$ *analytisch*, wenn sie in jedem $\boldsymbol{\sigma}_0 \in \mathcal{P}$ als lokal gleichmäßig konvergente Potenzreihe

$$\sum_{\boldsymbol{\nu} \in \mathcal{F}} t_{\boldsymbol{\nu}} (\boldsymbol{\sigma} - \boldsymbol{\sigma}_0)^{\boldsymbol{\nu}}$$

mit Konstanten $t_{\boldsymbol{\nu}} \in \mathbb{R}$, $\boldsymbol{\nu} \in \mathcal{F}$, dargestellt werden kann.

Ist die Abbildung analytisch, dann entspricht die Potenzreihe gerade ihrer Taylorreihe. Diese Eigenschaft wollen wir ausnutzen, denn in diesem Fall sind die Koeffizienten $t_{\boldsymbol{\nu}}$ gerade durch

$$t_{\boldsymbol{\nu}} = \frac{1}{\boldsymbol{\nu}!} \partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}} u(\boldsymbol{\sigma}_0)$$

gegeben.

Betrachten wir beispielsweise die Taylorreihe von $u(\,\cdot\,)$ um den Nullpunkt $\mathbf{0} \in \mathcal{P}$, dann erhalten wir wegen $\mathcal{P} = [-1, 1]^{\mathbb{N}}$ die Abschätzung

$$\sup_{\boldsymbol{\sigma} \in \mathcal{P}} \Big\| \sum_{\boldsymbol{\nu} \in \mathcal{F}} t_{\boldsymbol{\nu}} \boldsymbol{\sigma}^{\boldsymbol{\nu}} \Big\|_{\mathcal{X}} \leq \sup_{\boldsymbol{\sigma} \in \mathcal{P}} \sum_{\boldsymbol{\nu} \in \mathcal{F}} \|t_{\boldsymbol{\nu}} \boldsymbol{\sigma}^{\boldsymbol{\nu}}\|_{\mathcal{X}} \leq \sum_{\boldsymbol{\nu} \in \mathcal{F}} \|t_{\boldsymbol{\nu}}\|_{\mathcal{X}} \leq \sum_{\boldsymbol{\nu} \in \mathcal{F}} \frac{1}{\boldsymbol{\nu}!} \|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}} u(\mathbf{0})\|_{\mathcal{X}},$$

beziehungsweise nach Satz 3.26 weiter

$$\sup_{\boldsymbol{\sigma} \in \mathcal{P}} \| \sum_{\boldsymbol{\nu} \in \mathcal{F}} t_{\boldsymbol{\nu}} \boldsymbol{\sigma}^{\boldsymbol{\nu}} \|_{\mathcal{X}} \leq \frac{\|f\|_{\mathcal{Y}'}}{\beta} \sum_{\boldsymbol{\nu} \in \mathcal{F}} \frac{|\boldsymbol{\nu}|!}{\boldsymbol{\nu}!} \mathbf{b}^{\boldsymbol{\nu}}.$$

Die Frage, ob die Abbildung $\mathcal{P} \ni \boldsymbol{\sigma} \mapsto u(\boldsymbol{\sigma}) \in \mathcal{X}$ analytisch ist, hat sich somit auf jene, unter welchen Bedingungen $\sum_{\boldsymbol{\nu} \in \mathcal{F}} (|\boldsymbol{\nu}|!)(\boldsymbol{\nu}!)^{-1} \mathbf{b}^{\boldsymbol{\nu}}$ konvergiert, reduziert. Wann dies der Fall ist, wurde beispielsweise in [CDS10, Theorem 7.2] untersucht. Wir geben die Aussage hier ohne Beweis wieder.

Satz 3.28. *Sei* $0 < p \leq 1$. *Die Folge* $(\frac{|\boldsymbol{\nu}|!}{\boldsymbol{\nu}!}\mathbf{b}^{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ *liegt genau dann in* $\ell_p(\mathcal{F})$, *wenn* $\|\mathbf{b}\|_{\ell_1(\mathbb{N})} < 1$ *und* $\mathbf{b} \in \ell_p(\mathbb{N})$ *gilt.*

Satz 3.29. *Das Funktionensystem* $\{\varphi_i\}_{i\in\mathbb{N}}$ *sei so gewählt, dass* $\mathbf{b} \in \ell_1(\mathbb{N})$, *definiert als* $b_i := \beta^{-1}\|\varphi_i\|_{L_\infty(\Omega)}$, *die Bedingung* $\|\mathbf{b}\|_{\ell_1(\mathbb{N})} < 1$ *erfüllt. Dann hängt die Lösung* $u(\boldsymbol{\sigma})$ *des parametrischen Raum-Zeit-Variationsproblems* (3.14) *analytisch vom Parameter* $\boldsymbol{\sigma} \in \mathcal{P}$ *ab.*

Beweis. Der vorherige Satz und vorangegangene Überlegungen liefern unter diesen Voraussetzungen die behauptete Aussage. □

Die Bedingung $\|\mathbf{b}\|_{\ell_1(\mathbb{N})} < 1$ kann dabei so interpretiert werden, dass der feldabhängige Teil der partiellen Differentialgleichung den feldunabhängigen Anteil nicht zu stark stören darf, um analytische Abhängigkeit garantieren zu können. Dies verträgt sich allerdings nur bedingt mit der Motivation der Propagator-Differentialgleichung, da bei dieser die Felder einen starken Einfluss haben. In Kapitel 5 werden wir auf diesen Punkt noch einmal eingehen und anhand der numerischen Ergebnisse veranschaulichen, inwiefern die Bedingung eingehalten wird beziehungsweise eingehalten werden kann.

3.4. Periodische Randbedingungen

Da wir in diesem Kapitel bisher ausschließlich mit homogenen Dirichlet-Randbedingungen gearbeitet haben, wollen wir an dieser Stelle auf den Fall periodischer Randbedingungen eingehen. Dabei werden wir feststellen, dass auf Grund der Struktur der Propagator-Differentialgleichung nur sehr geringe Unterschiede zum betrachteten homogenen Fall bestehen.

Zunächst müssen wir die Rahmenbedingungen für die Betrachtung periodischer Randbedingungen festlegen. Dazu beschränken wir uns auf den Fall, dass $\Omega = \bigtimes_{i=1}^{n}(0, l_i) \subset \mathbb{R}^n$ ein beschränkter offener Quader ist, wobei $l_i \in \mathbb{R}_+$ für $i = 1 \dots n$ sei. Weiter führen wir nun die Analoga des Lebesgue-Raums $L_2(\Omega)$ und des Sobolev-Raums $H^1(\Omega)$ für periodische Funktionen ein. Da dies für die diese Arbeit von untergeordneter Bedeutung ist, wird an dieser Stelle nur ein Überblick über die benötigten Ergebnisse gegeben. Genauere Ausführungen findet man beispielsweise bei Han und Atkinson [HA09].

Definition 3.30. Sei $\mathcal{C}^\infty_{\text{per}}(\Omega) \subset \mathcal{C}^\infty(\mathbb{R}^n)$ die Teilmenge der glatten Ω-periodischen Funktionen. Als den Lebesgue-Raum Ω-periodischer Funktionen $L_{2,\text{per}}(\Omega)$ definieren wir den Abschluss von $C^\infty_{\text{per}}(\Omega)$ bezüglich der L_2-Norm. Weiter definieren wir den Sobolev-Raum Ω-periodischer Funktionen $H^1_{\text{per}}(\Omega)$ als den Abschluss von $C^\infty_{\text{per}}(\Omega)$ bezüglich der H^1-Norm.

Diese Räume verwenden wir nun, um die Räume V und H (vergleiche Bemerkung 3.3) zu definieren. Nach Konstruktion handelt es sich bei $H^1_{\text{per}}(\Omega)$ und $L_{2,\text{per}}(\Omega)$ um Hilberträume und desweiteren ist $H^1_{\text{per}}(\Omega)$ ein dichter Unterraum von $L_{2,\text{per}}(\Omega)$. Wählen wir also $V := H^1_{\text{per}}(\Omega)$ und $H := L_{2,\text{per}}(\Omega)$, dann erhalten wir nach Definition 2.6 wie zuvor ein Gelfand-Tripel der Form

$$V \hookrightarrow H \cong H' \hookrightarrow V' = (H^1_{\text{per}}(\Omega))'.$$

Untersucht man die Ausführungen dieses Kapitels für den Fall homogener Dirichlet-Randbedingungen, dann stellt man fest, dass lediglich die aus Annahme 2.14 stammenden Bedingungen an die Bilinearformen $a(\cdot\,,\cdot\,;t)$, deren Darstellung durch den Wechsel zu periodischen Randbedingungen unverändert bleibt, nachgewiesen werden müssen.

Lemma 3.31. *Sei $\{a(\cdot\,,\cdot\,;t)\}_{t\in I}$ die Familie von Bilinearformen aus Definition 3.4, wobei die Hilberträume als $V = H^1_{\mathrm{per}}(\Omega)$ und $H = L_{2,\mathrm{per}}(\Omega)$ gegeben seien. Diese Bilinearformen erfüllen die folgenden Eigenschaften:*

(i) Messbarkeit: *Die Abbildung $I \ni t \mapsto a(\eta,\zeta;t)$ ist messbar für alle $\eta,\zeta \in V$.*

(ii) Stetigkeit: *Es gilt*

$$|a(\eta,\zeta;t)| \le \gamma_a \|\eta\|_V \|\zeta\|_V \quad \textit{für alle } \eta,\zeta \in V \textit{ und fast alle } t \in I$$

mit Stetigkeitskonstante $\gamma_a = \max\{c, \|\omega\|_{L_\infty(I;L_\infty(\Omega))} + |\mu|\} < \infty$.

(iii) Gårding-Ungleichung: *Es gilt*

$$a(\eta,\eta;t) + \lambda\|\eta\|_H^2 \ge \alpha\|\eta\|_V^2 \quad \textit{für alle } \eta \in V \textit{ und fast alle } t \in I$$

mit $\alpha = c > 0$ und $\lambda = \max\{\|\omega\|_{L_\infty(I;L_\infty(\Omega))} - \mu + c, c\} \ge c > 0$.

Beweis. Die Nachweise der Messbarkeit und Stetigkeit bleiben unverändert wie in Satz 3.7 und werden hier nicht wiederholt.

Für die Gårding-Ungleichung sei zunächst $\eta \in V$ beliebig und $\lambda \in \mathbb{R}$. Dann gilt

$$\begin{aligned} a(\eta,\eta;t) + \lambda\|\eta\|_H^2 &= c\|\nabla\eta\|_H^2 + \langle\omega(t,\cdot\,)\eta,\eta\rangle_H + \mu\,\langle\eta,\eta\rangle_H + \lambda\,\langle\eta,\eta\rangle_H \\ &= c\|\nabla\eta\|_H^2 + \langle\omega(t,\cdot\,) + \mu + \lambda)\eta,\eta\rangle_H\,. \end{aligned}$$

Wählen wir $\lambda := \max\{\|\omega\|_{L_\infty(I;L_\infty(\Omega))} - \mu + c, c\} \ge c > 0$, dann gilt $\omega(t,\cdot\,) + \mu + \lambda - c \ge 0$ fast überall in Ω und wir erhalten die Abschätzung

$$\begin{aligned} a(\eta,\eta;t) + \lambda\|\eta\|_H^2 &= c\|\nabla\eta\|_H^2 + \langle(\omega(t,\cdot\,) + \mu + \lambda - c)\eta,\eta\rangle_H + c\|\eta\|_H^2 \\ &\ge c\|\nabla\eta\|_H^2 + c\|\eta\|_H^2 \\ &= c\|\eta\|_V^2. \end{aligned}$$

□

Auf diesem Lemma aufbauend können die restlichen Ergebnisse des Kapitels analog auch auf den periodischen Fall übertragen werden.

4 Kapitel

Petrov-Galerkin-Verfahren

Mit diesem Kapitel beginnt der numerische Anteil dieser Arbeit. Wir führen zunächst das sogenannte Petrov-Galerkin-Verfahren ein, welches als Grundlage für die Reduzierte-Basis-Methode des nächsten Kapitels dienen wird. Wie bereits bei den funktionalanalytischen Grundlagen in Kapitel 2, wird dies zu Beginn allgemein gehalten und erst später auf die Problemstellung aus Kapitel 3 zugeschnitten. Als Quellen für dieses Kapitel wurden vor allem die Arbeiten von Nochetto, Siebert und Veeser [NSV09] sowie Braess [Bra13] verwendet.

Wir beginnen nun, indem wir die Rahmenbedingungen in Form eines abstrakten Variationsproblems festlegen. Seien $\mathcal{X}$ und $\mathcal{Y}$ zwei reelle Hilberträume und seien weiter $b\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ eine stetige Bilinearform und $f\colon \mathcal{Y} \to \mathbb{R}$ ein stetiges lineares Funktional. Das *abstrakte Variationsproblem* sei gegeben durch:

$$\text{Finde } u \in \mathcal{X} \text{ mit} \quad b(u,v) = f(v) \quad \text{für alle } v \in \mathcal{Y}. \tag{4.1}$$

Hinreichende Bedingungen für die Existenz und Eindeutigkeit einer Lösung haben wir bereits in Kapitel 2 gesehen, weswegen wir für den Rest dieses Kapitels annehmen, dass es sich dabei um ein korrekt gestelltes Problem handelt und insbesondere die inf-sup-Konstante β die Bedingung

$$\beta := \inf_{u\in\mathcal{X}} \sup_{v\in\mathcal{Y}} \frac{b(u,v)}{\|u\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} > 0 \tag{4.2}$$

erfüllt.

4.1. Grundlagen

Als *Petrov-Galerkin-Verfahren* werden diejenigen Galerkin-Verfahren bezeichnet, die auf Variationsprobleme mit nicht-übereinstimmendem Ansatz- und Testraum $\mathcal{X}$ und $\mathcal{Y}$ zugeschnitten sind.

Wie es bei Galerkin-Verfahren üblich ist, wird eine Diskretisierung des Variationsproblems (4.1) durch Approximation der im Allgemeinen unendlichdimensionalen Hilberträume mittels $\mathcal{N}$- beziehungsweise $\mathcal{M}$-dimensionaler Unterräume $\mathcal{X}_\mathcal{N} \subset \mathcal{X}$ und $\mathcal{Y}_\mathcal{M} \subset \mathcal{Y}$ erreicht. Zwar ist auch für den Fall $\mathcal{M} > \mathcal{N}$ eine sinnvolle Formulierung einer Diskretisierung möglich, diese erfolgt dann allerdings im Sinne einer Residuum-Minimierung, wie beispielsweise bei [And12b]. Da sich dies nicht ohne Weiteres mit der beabsichtigten Verwendung als Grundlage für eine Reduzierte-Basis-Methode verträgt, beschränken wir uns auf den Fall $\mathcal{N} = \mathcal{M}$.

Wir orientieren uns an [NSV09, Section 3.1], beschränken uns aber auf eine minimale Einführung.

Definition 4.1. Seien $\mathcal{X}_\mathcal{N} \subset \mathcal{X}$ und $\mathcal{Y}_\mathcal{N} \subset \mathcal{Y}$ Unterräume der Dimension $\mathcal{N} \in \mathbb{N}$. Als *Petrov-Galerkin-Lösung* von (4.1) bezeichnen wir eine Lösung $u_\mathcal{N} \in \mathcal{X}_\mathcal{N}$ des Variationsproblems:

$$\text{Finde } u_\mathcal{N} \in \mathcal{X}_\mathcal{N} \text{ mit} \quad b(u_\mathcal{N}, v) = f(v) \quad \text{für alle } v \in \mathcal{Y}_\mathcal{N}. \tag{4.3}$$

Zur einfacheren Unterscheidung bezeichnen wir (4.1) im Weiteren als stetiges und (4.3) als diskretes Variationsproblem.

Bemerkung 4.2. Anders als bei den Galerkin-Verfahren für elliptische Probleme führt eine solche Diskretisierung eines korrekt gestellten Problems nicht automatisch zu einem korrekt gestellten diskreten Problem. Wegen der Ungleichung

$$\sup_{v\in\mathcal{Y}} \frac{b(u,v)}{\|v\|_\mathcal{Y}} \geq \sup_{v\in\mathcal{Y}_\mathcal{N}} \frac{b(u,v)}{\|v\|_\mathcal{Y}} \quad \text{für alle } u \in \mathcal{X}$$

ist die stetige inf-sup-Bedingung (4.2) kein hinreichendes Kriterium für die Gültigkeit der diskreten inf-sup-Bedingung

$$\beta_\mathcal{N} := \inf_{u\in\mathcal{X}_\mathcal{N}} \sup_{v\in\mathcal{Y}_\mathcal{N}} \frac{b(u,v)}{\|u\|_\mathcal{X}\|v\|_\mathcal{Y}} > 0. \tag{4.4}$$

Allerdings gilt aufgrund der endlichen Dimension von $\mathcal{X}_\mathcal{N}$ und $\mathcal{Y}_\mathcal{N}$ stets die Gleichheit

$$\inf_{u\in\mathcal{X}_\mathcal{N}} \sup_{v\in\mathcal{Y}_\mathcal{N}} \frac{b(u,v)}{\|u\|_\mathcal{X}\|v\|_\mathcal{Y}} = \inf_{v\in\mathcal{Y}_\mathcal{N}} \sup_{u\in\mathcal{X}_\mathcal{N}} \frac{b(u,v)}{\|u\|_\mathcal{X}\|v\|_\mathcal{Y}}.$$

Die Stetigkeit dagegen muss nicht explizit nachgewiesen werden, da die diskreten Stetigkeitskonstanten stets durch die des stetigen Variationsproblems von oben beschränkt werden.

Dass die diskrete inf-sup-Konstante bei den Petrov-Galerkin-Verfahren eine wichtige Rolle spielt, wird unter anderem bei den folgenden beiden Aussagen deutlich.

Satz 4.3. *Gilt die diskrete inf-sup-Bedingung* (4.4)*, dann erfüllt die Petrov-Galerkin-Lösung $u_\mathcal{N} \in \mathcal{X}_\mathcal{N}$ die Abschätzung*

$$\|u_\mathcal{N}\|_\mathcal{X} \leq \frac{1}{\beta_\mathcal{N}} \|f\|_{\mathcal{Y}'_\mathcal{N}}.$$

Beweis. Folgt direkt aus dem Banach-Nečas-Babuška-Theorem, Satz 2.20. □

Satz 4.4 (Lemma von Céa). *Seien $u \in \mathcal{X}$ die Lösung des stetigen Variationsproblems* (4.1) *und $u_\mathcal{N} \in \mathcal{X}_\mathcal{N}$ die diskrete Lösung von* (4.3)*. Sei weiter γ die Stetigkeitskonstante der Bilinearform b aus* (4.1)*. Der Fehler $u - u_\mathcal{N} \in \mathcal{X}$ erfüllt die Ungleichung*

$$\|u - u_\mathcal{N}\|_\mathcal{X} \leq \frac{\gamma}{\beta_\mathcal{N}} \inf_{w\in\mathcal{X}_\mathcal{N}} \|u - w\|_\mathcal{X}.$$

Beweis. Siehe [NSV09, Theorem 3.2]. □

Durch Verfeinerung der Diskretisierung will man üblicherweise erreichen, dass die Abweichung zwischen stetiger und diskreter Lösung kleiner wird. Um dies aus dem Lemma von Céa zu erhalten, wird der folgende Stabilitätsbegriff notwendig.

Definition 4.5. Sei $\{(\mathcal{X}_\mathcal{N}, \mathcal{Y}_\mathcal{N})\}_{\mathcal{N}\geq 1}$ eine Folge von endlichdimensionalen Unterräumen mit zugehörigen diskreten inf-sup-Konstanten $\{\beta_\mathcal{N}\}_{\mathcal{N}\geq 1}$. Wir nennen diese Diskretisierungen *stabil*, wenn ein $\beta_{\mathrm{LB}} > 0$ mit

$$\inf_{\mathcal{N}\geq 1} \beta_\mathcal{N} \geq \beta_{\mathrm{LB}} > 0$$

existiert.

Abschließend führen wir eine alternative Darstellung der inf-sup-Konstante (4.2) ein, welche später insbesondere die numerische Berechnung von $\beta_\mathcal{N}$ ermöglichen wird.

Definition 4.6. Der *Supremizing-Operator* $T\colon \mathcal{X} \to \mathcal{Y}$ der Bilinearform b aus (4.1) sei definiert über die Gleichung

$$\langle Tu, v\rangle_\mathcal{Y} = b(u, v) \quad \text{für alle } u \in \mathcal{X} \text{ und } v \in \mathcal{Y}.$$

Lemma 4.7. *Der Supremizing-Operator T ist wohldefiniert, linear und stetig. Ferner gelten die Gleichungen*

$$Tu = \arg\sup_{v\in\mathcal{Y}} \frac{b(u,v)}{\|v\|_\mathcal{Y}} \qquad \textit{und} \qquad \beta = \inf_{u\in\mathcal{X}} \frac{\|Tu\|_\mathcal{Y}}{\|u\|_\mathcal{X}}.$$

Beweis. Wohldefiniertheit, Linearität und Stetigkeit ergeben sich aus der Definition und dem Rieszschen Darstellungssatz. Die erste Gleichung folgt mit der Anwendung der Cauchy-Schwarz-Ungleichung auf

$$\frac{b(u,v)}{\|v\|_\mathcal{Y}} = \frac{\langle Tu, v\rangle_\mathcal{Y}}{\|v\|_\mathcal{Y}} \leq \frac{\|Tu\|_\mathcal{Y}\|v\|_\mathcal{Y}}{\|v\|_\mathcal{Y}} = \|Tu\|_\mathcal{Y}$$

und der Gleichung

$$\frac{b(u,Tu)}{\|Tu\|_\mathcal{Y}} = \frac{\langle Tu, Tu\rangle_\mathcal{Y}}{\|Tu\|_\mathcal{Y}} = \|Tu\|_\mathcal{Y}.$$

Dies impliziert, dass das Supremum von b bezüglich des zweiten Arguments gerade von $v = Tu$ angenommen wird. Daraus folgt auch die zweite Gleichung, denn es gilt

$$\beta = \inf_{u\in\mathcal{X}} \sup_{v\in\mathcal{Y}} \frac{b(u,v)}{\|u\|_\mathcal{X}\|v\|_\mathcal{Y}} = \inf_{u\in\mathcal{X}} \frac{b(u,Tu)}{\|u\|_\mathcal{X}\|Tu\|_\mathcal{Y}} = \inf_{u\in\mathcal{X}} \frac{\langle Tu, Tu\rangle_\mathcal{Y}}{\|u\|_\mathcal{X}\|Tu\|_\mathcal{Y}} = \inf_{u\in\mathcal{X}} \frac{\|Tu\|_\mathcal{Y}}{\|u\|_\mathcal{X}}. \qquad \square$$

Ferner erhalten wir damit auch die nützliche Darstellung

$$\beta^2 = \inf_{u\in\mathcal{X}} \frac{\langle Tu, Tu\rangle_\mathcal{Y}}{\|u\|_\mathcal{X}^2}. \tag{4.5}$$

4.2. Raum-Zeit-Diskretisierung

Wir kehren nun zu der Raum-Zeit-Variationsformulierungen (3.6) und (3.14) der Propagator-Differentialgleichung aus Kapitel 3 zurück und wollen diese mit Hilfe eines Petrov-Galerkin-Verfahrens diskretisieren. Dies erfordert die Konstruktion endlichdimensionaler Unterräume $\mathcal{X}_\mathcal{N}$ und $\mathcal{Y}_\mathcal{N}$. Anschließend werden wir für diese eine hinreichende Bedingung für die Stabilität im Sinne von Definition 4.5 angeben.

Dieser Abschnitt orientiert sich an den Ausführungen von Andreev [And12a; And12b] und nutzt die Charakterisierung der Bochner-Sobolev-Räume als Hilbertraum-Tensorprodukte nach Satz 2.13, um die Konstruktion so in einen rein zeitlichen und einen rein räumlichen Anteil zu zerlegen.

Korollar 4.8. *Die Hilberträume $\mathcal{X}$ und $\mathcal{Y}$ aus* (2.2) *und* (2.4) *lassen sich schreiben als*

$$\mathcal{X} \cong (L_2(I) \otimes V) \cap (H^1(I) \otimes V'), \qquad \mathcal{Y} \cong (L_2(I) \otimes V) \times H. \tag{4.6}$$

Für den zeitlichen Anteil werden wir die beiden endlichdimensionalen Unterräume $E_\mathcal{K} \subset H^1(I) \subset L_2(I)$ und $F_\mathcal{K} \subset L_2(I)$ mit den Dimensionen $\dim E_\mathcal{K} = \mathcal{K} + 1$ und $\dim F_\mathcal{K} = \mathcal{K}$ verwenden. Die Räume V, H und V' der räumlichen Komponente können aufgrund der Gelfand-Tripel-Struktur durch den selben endlichdimensionalen Raum $V_\mathcal{J} \subset V$ diskretisiert werden. Dieser wird die Dimension $\dim V_\mathcal{J} = \mathcal{J}$ haben.

Diese Teilräume liefern dann zusammen mit der Tensorprodukt-Darstellung (4.6) die Diskretisierungen

$$\mathcal{X}_\mathcal{N} := E_\mathcal{K} \otimes V_\mathcal{J}, \qquad \mathcal{Y}_\mathcal{N} := (F_\mathcal{K} \otimes V_\mathcal{J}) \times V_\mathcal{J} \tag{4.7}$$

mit der Dimension

$$\mathcal{N} := \dim \mathcal{X}_\mathcal{N} = (\mathcal{K}+1)\mathcal{J} = \mathcal{K}\mathcal{J} + \mathcal{J} = \dim \mathcal{Y}_\mathcal{N}.$$

Wir merken an dieser Stelle an, dass wir stets die Skalarprodukte und die jeweiligen induzierten Normen von $\mathcal{X}$ und $\mathcal{Y}$ auf die jeweiligen Unterräume übertragen und für diese weiter die bekannten Bezeichnungen verwenden.

Zeitliche Komponente. Hierfür benötigen wir zunächst eine Diskretisierung des Zeitintervalls $I = [0, T]$ in Form eines nicht notwendigerweise äquidistanten Gitters

$$\mathcal{T}_\mathcal{K} := \{0 = t_0 < t_1 < \cdots < t_{\mathcal{K}-1} < t_\mathcal{K} = T\} \subset I.$$

Die Diskretisierung $E_\mathcal{K}$ des Ansatzraumes basiert auf stetigen, stückweise affinen Funktionen, genauer den klassischen Hutfunktionen ϑ_k auf den Gitterpunkten $t_k \in \mathcal{T}_\mathcal{K}$ für $k = 0, \ldots, \mathcal{K}$. Diese erfüllen $\vartheta_k(t_{\tilde{k}}) = \delta_{k\tilde{k}}$, wobei $\delta_{k\tilde{k}}$ das bekannte Kronecker-Delta sei. Wir fassen diese Hutfunktionen zu einer Basis $\mathcal{B}^E_\mathcal{K} := \{\vartheta_k \,|\, k = 0, \ldots, \mathcal{K}\}$ zusammen und definieren $E_\mathcal{K} := \operatorname{span} \mathcal{B}^E_\mathcal{K}$.

Für den Testraum-Anteil $F_\mathcal{K}$ verwenden wir stattdessen stückweise konstante Funktionen, die als charakteristische Funktionen $\xi_k := \chi_{(t_{k-1}, t_k)}$ der Teilintervalle $(t_{k-1}, t_k) \subset I$ mit $k = 1, \ldots, \mathcal{K}$ gegeben sind. Diese seien zu der Basismenge $\mathcal{B}^F_\mathcal{K} := \{\xi_k \,|\, k = 1, \ldots, \mathcal{K}\}$ zusammengefasst und weiter definieren wir $F_\mathcal{K} := \operatorname{span} \mathcal{B}^F_\mathcal{K}$.

Diese Wahl von Ansatz- und Testraum-Anteil führt nach [And12b] zu einem Crank-Nicolson-ähnlichen Verfahren, welches dementsprechend auch als Time-Stepping-Verfahren aufgefasst werden kann.

Räumliche Komponente. Hier wollen wir nur die verwendete Notation festlegen, da für den räumlichen Anteil die meisten von Galerkin-Verfahren bekannten Ansätze, beispielsweise Finite-Elemente oder globale Basisfunktionen, verwendet werden können.

Wir definieren erneut eine endliche Basismenge $\mathcal{B}_{\mathcal{J}}^{V} := \{\eta_j \,|\, j = 1, \dots, \mathcal{J}\}$ und darauf aufbauend den endlichdimensionalen Raum $V_{\mathcal{J}} := \operatorname{span} \mathcal{B}_{\mathcal{J}}^{V}$.

Raum-Zeit-Diskretisierung. Unter Verwendung der Tensorprodukt-Darstellung (4.6) können wir nun die beiden einzeln betrachteten Komponenten zu den endlichdimensionalen Raum-Zeit-Unterräumen zusammensetzen. Dazu definieren wir zunächst die Basen

$$\mathcal{B}_{\mathcal{N}}^{\mathcal{X}} := \{\vartheta \otimes \eta \,|\, \vartheta \in \mathcal{B}_{\mathcal{K}}^{E}, \eta \in \mathcal{B}_{\mathcal{J}}^{V}\}, \qquad \mathcal{B}_{\mathcal{K}\mathcal{J}}^{\mathcal{Y}_1} := \{\xi \otimes \eta \,|\, \xi \in \mathcal{B}_{\mathcal{K}}^{F}, \eta \in \mathcal{B}_{\mathcal{J}}^{V}\}.$$

und weiter

$$\mathcal{B}_{\mathcal{N}}^{\mathcal{Y}} := (\mathcal{B}_{\mathcal{K}\mathcal{J}}^{\mathcal{Y}_1} \times \{0\}) \cup (\{0\} \times \mathcal{B}_{\mathcal{J}}^{V}).$$

Nach Konstruktion und den Definitionen (4.7) gilt nun

$$\mathcal{X}_{\mathcal{N}} = E_{\mathcal{K}} \otimes V_{\mathcal{J}} = \operatorname{span} \mathcal{B}_{\mathcal{N}}^{\mathcal{X}}, \quad \mathcal{Y}_{\mathcal{N}} = (F_{\mathcal{K}} \otimes V_{\mathcal{J}}) \times V_{\mathcal{J}} = \operatorname{span} \mathcal{B}_{\mathcal{N}}^{\mathcal{Y}}. \tag{4.8}$$

Stabilität der Diskretisierung. Wie bereits erwähnt, wollen wir eine hinreichende Bedingung für die Stabilität der beschriebenen Diskretisierung im Sinne von Definition 4.5 angeben. Dies können wir an dieser Stelle nur für die nicht-parametrische Variationsformulierung (3.6) in Form einer knappen Übersicht ohne Beweise und tiefergehende Motivation abhandeln, weswegen auf die zugrundeliegende Arbeit [And12b, Section 5.2] verwiesen sei.

Die hier konstruierte Diskretisierung führt hauptsächlich zu einer Bedingung an die räumliche Komponente $V_{\mathcal{J}}$. Bevor wir diese angeben können, benötigen wir die nachfolgende Definition nach [And12b, S. 62], die in etwas anderer Form oft bei der Behandlung von hyperbolischen partiellen Differentialgleichungen auftritt.

Definition 4.9. Es seien $V_{\mathcal{J}} \subset V$ ein endlichdimensionaler Unterraum, $\mathcal{T}_{\mathcal{K}}$ ein Gitter des Zeitintervalls I und weiter durch $\max \Delta\mathcal{T}_{\mathcal{K}} := \max_{k=1,\dots,\mathcal{K}} |t_k - t_{k-1}|$ die maximale Schrittweite des Zeitgitters gegeben. Wir bezeichnen die Größe

$$\mathrm{CFL}_{\mathcal{N}} := \max \Delta\mathcal{T}_{\mathcal{K}} \sup_{\eta \in V_{\mathcal{J}} \setminus \{0\}} \frac{\|\eta\|_V}{\|\eta\|_{V'}}$$

als *Courant-Friedrichs-Levi-Zahl*, kurz *CFL-Zahl*, der Diskretisierung $(\mathcal{X}_{\mathcal{N}}, \mathcal{Y}_{\mathcal{N}})$ aus (4.8).

Bemerkung 4.10. Ist $V_{\mathcal{J}}$ endlichdimensional, so gilt insbesondere $\mathrm{CFL}_{\mathcal{N}} < \infty$.

Weiter führen wir eine gewisse inf-sup-Konstante ein, welche in [And12b, S. 57] aus der Konstruktion einer äquivalenten Norm auf $\mathcal{X}$ resultiert, die dann für den Stabilitätsnachweis [And12b, Theorem 5.2.6] verwendet wird. Dazu benötigen wir die Operatorfamilie $\{A(t)\}_{t \in I}$ aus (3.3), von welcher wir weiter fordern, dass sie selbstadjungiert ist und die Gårding-Ungleichung mit $\lambda = 0$ erfüllt. Ersteres gilt nach Lemma 3.9, während letzteres durch Proposition 2.24 erreicht werden kann. Unter diesen Voraussetzungen garantiert

der Satz von Lax-Milgram [Eva10, Section 6.2.1] die stetige Invertierbarkeit von $A(t)$. Ferner erlaubt es uns die Definition zweier Skalarprodukte beziehungsweise Normen durch

$$\langle v, \tilde{v}\rangle_+ := \int_I \langle A(t)v(t), \tilde{v}(t)\rangle_{V'\times V}\,\mathrm{d}t, \qquad \|v\|_+^2 := \langle v, v\rangle_+, \qquad v, \tilde{v} \in L_2(I;V),$$
$$\langle z, \tilde{z}\rangle_- := \int_I \langle A(t)^{-1}z(t), \tilde{z}(t)\rangle_{V\times V'}\,\mathrm{d}t, \qquad \|z\|_-^2 := \langle z, z\rangle_-, \qquad z, \tilde{z} \in L_2(I;V').$$

Mit dieser Vorarbeit können wir nun die angesprochene inf-sup-Konstante nach [And12b, S. 57] als

$$\beta_\pm(\mathcal{X}_\mathcal{N}, \mathcal{Y}_\mathcal{N}) := \inf_{u\in\mathcal{X}_\mathcal{N}} \sup_{v\in\mathcal{Y}_\mathcal{N}} \frac{\int_I \langle u_t(t), v_1(t)\rangle_{V'\times V}\,\mathrm{d}t}{\|u_t\|_-\|v_1\|_+}$$

definieren, wobei Infimum und Supremum bezüglich aller Elemente gebildet werden, für die der Nenner nicht Null wird.

Unter Verwendung der CFL-Zahl und dieser inf-sup-Konstante lässt sich letztendlich die folgende Stabilitätsaussage nachweisen.

Satz 4.11. *Seien* $\{(\mathcal{X}_\mathcal{N}, \mathcal{Y}_\mathcal{N})\}_{\mathcal{N}\geq 1}$ *Diskretisierungen der Form* (4.8)*. Gelten die Bedingungen*

$$\sup_{\mathcal{N}\geq 1} \mathrm{CFL}_\mathcal{N} < \infty \quad \textit{und} \quad \inf_{\mathcal{N}\geq 1} \beta_\pm(\mathcal{X}_\mathcal{N}, \mathcal{Y}_\mathcal{N}) > 0,$$

dann gilt für die diskrete inf-sup-Konstante $\beta_\mathcal{N}$ *aus* (4.4) *die Abschätzung*

$$\beta_\mathcal{N} \geq c_0 \min\{1, \beta_\pm(\mathcal{X}_\mathcal{N}, \mathcal{Y}_\mathcal{N})\} \min\{1, \mathrm{CFL}_\mathcal{N}^{-1}\} \quad \textit{für } \mathcal{N} \geq 1$$

mit einer von $\mathcal{X}_\mathcal{N}$ *und* $\mathcal{Y}_\mathcal{N}$ *unabhängigen Konstanten* $c_0 > 0$*.*

Beweis. Siehe [And12b, Subsection 5.2.2]. □

Da eine ausführlichere Bearbeitung der Stabilität respektive obiger Stabilitätsaussage den Rahmen dieser Arbeit übersteigt, werden wir erst am Ende dieses Kapitels dazu zurückkehren und die Stabilität anhand von Beispielen numerisch überprüfen.

4.3. Numerische Umsetzung

Nachdem das theoretische Fundament für die Diskretisierung gelegt ist, widmen wir uns nun der tatsächlichen Anwendung auf die Variationsformulierung der Propagator-Differentialgleichung. Um die in Abschnitt 3.2 durchgeführte Parametrisierung dieser numerisch umsetzen zu können, müssen wir zunächst den dort noch unendlichdimensionalen Parameterraum $\mathcal{P}$ auf einen endlichdimensionalen einschränken. Da wir in Einklang mit den Bedingungen aus Kapitel 3 bleiben wollen, erreichen wir dies durch die Wahl einer Dimension $N_\mathcal{P} \in \mathbb{N}$ und der Entwicklungsfunktionen $\varphi_i = 0$ für alle $i > N_\mathcal{P}$.

Wir wiederholen an dieser Stelle das parametrische Raum-Zeit-Variationsproblem noch einmal in aller Kürze. Seien $\mathcal{X}$ und $\mathcal{Y}$ die Ansatz- respektive Testräume aus (2.2) und (2.4). Weiter seien die Anzahl der Parameter durch $N_\mathcal{P} \in \mathbb{N}$, der Parameterraum $\mathcal{P} = [-1, 1]^{N_\mathcal{P}}$ und die Feld-Entwicklungsfunktionen durch die Menge

$$\mathcal{B}_{N_\mathcal{P}}^\omega := \{\varphi_i \in L_\infty(\Omega) \,|\, i = 1, \ldots, N_\mathcal{P}\}$$

gegeben. Das parametrische Raum-Zeit-Variationsproblem lautet damit:

$$\text{Sei } \boldsymbol{\sigma} \in \mathcal{P}, \text{ finde } u(\boldsymbol{\sigma}) \in \mathcal{X} \text{ mit } \quad b(u(\boldsymbol{\sigma}), v; \boldsymbol{\sigma}) = f(v) \quad \text{für alle } v \in \mathcal{Y}, \tag{4.9}$$

wobei b und f wie in Definition 3.18 gegeben seien.

Lemma 4.12. *Die parametrische Bilinearform aus* (3.15) *ist unter diesen Gegebenheiten affin vom Parameter $\boldsymbol{\sigma} \in \mathcal{P}$ abhängig. Genauer gilt*

$$b(\cdot, \cdot; \boldsymbol{\sigma}) = b_0(\cdot, \cdot) + \sum_{i=1}^{N_\mathcal{P}} \sigma_i b_i(\cdot, \cdot) \tag{4.10}$$

mit parameterunabhängigen stetigen Bilinearformen $b_i \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, $i = 0, \dots, N_\mathcal{P}$.

Beweis. Ohne Einschränkung schreiben wir das parametrische Raum-Zeit-Feld ω in der Darstellung

$$\omega \colon I \times \Omega \times \mathcal{P} \to \mathbb{R}, \quad \omega(t, \mathbf{x}; \boldsymbol{\sigma}) = \sum_{i=1}^{N_\mathcal{P}} \sigma_i \chi_i(t) \varphi_i(\mathbf{x})$$

mit charakteristischen Funktionen $\chi_i \in \{\chi_{I_1}, \chi_{I_2}\}$ für $i = 1, \dots, N_\mathcal{P}$. Wir definieren die Bilinearformen $b_i \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ durch

$$\begin{aligned} b_0 &:= \int_I \left[\langle u_t, v_1 \rangle_{V' \times V} + c \langle \nabla u, \nabla v_1 \rangle_H + \mu \langle u, v_1 \rangle_H \right] \mathrm{d}t + \langle u(0), v_2 \rangle_H, \\ b_i &:= \int_I \chi_i \langle \varphi_i u, v_1 \rangle_H \, \mathrm{d}t, \quad i = 1, \dots, N_\mathcal{P}, \end{aligned}$$

wobei die Stetigkeit durch einfache Abschätzungen folgt und hier nicht weiter ausgeführt wird. Unter Verwendung dieser gilt nun offenbar

$$b(u, v; \boldsymbol{\sigma}) = \int_I \left[\langle u_t, v_1 \rangle_{V' \times V} + a(u, v_1; \boldsymbol{\sigma}) \right] \mathrm{d}t + \langle u(0), v_2 \rangle_H = b_0(u, v) + \sum_{i=1}^{N_\mathcal{P}} \sigma_i b_i(u, v).$$

□

Wir beginnen die Herleitung der diskreten Darstellung des Variationsproblems (4.9) durch die Definition der Elemente des diskreten Ansatzraums $\mathcal{X}_\mathcal{N}$ und des Testraums $\mathcal{Y}_\mathcal{N}$.

Definition 4.13. Als *diskrete Ansatzfunktion* $u_\mathcal{N} \in \mathcal{X}_\mathcal{N}$ bezeichnen wir eine Linearkombination der Form

$$u_\mathcal{N} := \sum_{k=0}^{\mathcal{K}} \sum_{j=1}^{\mathcal{J}} u_j^k (\vartheta_k \otimes \eta_j)$$

mit dem durch

$$\mathbf{u}_\mathcal{N} := [\mathbf{u}_\bullet^k]_{k=0,\dots,\mathcal{K}} := [u_j^k]_{j=1,\dots,\mathcal{J};k=0,\dots,\mathcal{K}}$$

definierten zugehörigen Koeffizientenvektor $\mathbf{u}_\mathcal{N} \in \mathbb{R}^\mathcal{N}$.

Ebenso sei eine *diskrete Testfunktion* $v_\mathcal{N} = (y_{\mathcal{K}\mathcal{J}}, z_\mathcal{J}) \in \mathcal{Y}_\mathcal{N}$ als das Paar der Linearkombinationen

$$y_{\mathcal{K}\mathcal{J}} := \sum_{k=1}^{\mathcal{K}} \sum_{j=1}^{\mathcal{J}} y_j^k (\xi_k \otimes \eta_j), \qquad z_\mathcal{J} := \sum_{l=1}^{\mathcal{J}} z_l \eta_l$$

definiert, wobei die Koeffizientenvektoren $\mathbf{y}_{\mathcal{K}\mathcal{J}} \in \mathbb{R}^{\mathcal{K}\mathcal{J}}$ und $\mathbf{z}_{\mathcal{J}} \in \mathbb{R}^{\mathcal{J}}$ durch

$$\mathbf{y}_{\mathcal{K}\mathcal{J}} := [\mathbf{y}_\bullet^k]_{k=1,\dots,\mathcal{K}} := [y_j^k]_{j,=1,\dots,\mathcal{J};k=1,\dots,\mathcal{K}}, \qquad \mathbf{z}_{\mathcal{J}} := [z_l]_{l=1,\dots,\mathcal{J}}$$

gegeben sind.

Ferner definieren wir an dieser Stelle vorbereitend einige Gramsche Matrizen, welche sich aus dem rein zeitlichen beziehungsweise räumlichen Anteil der Diskretisierung gewinnen lassen. Diese werden uns später als Bausteine zur Berechnung der Raum-Zeit-Systeme dienen.

Definition 4.14. Die zeitlichen Gramschen Matrizen $\mathbf{M}_t^{FE}, \mathbf{C}_t^{FE}, \mathbf{M}_{t,i}^{FE} \in \mathbb{R}^{\mathcal{K}\times(\mathcal{K}+1)}$ des L_2-Skalarprodukts von Ansatz- und Testraumbasisfunktionen seien definiert als

$$\mathbf{M}_t^{FE} := \big[\langle \vartheta_k, \xi_l\rangle_{L_2(I)}\big]_{\substack{l=1,\dots,\mathcal{K}\\ k=0,\dots,\mathcal{K}}}, \qquad \mathbf{C}_t^{FE} := \big[\langle \vartheta_k', \xi_l\rangle_{L_2(I)}\big]_{\substack{l=1,\dots,\mathcal{K}\\ k=0,\dots,\mathcal{K}}},$$
$$\mathbf{M}_{t,i}^{FE} := \big[\langle \chi_i\vartheta_k, \xi_l\rangle_{L_2(I)}\big]_{\substack{l=1,\dots,\mathcal{K}\\ k=0,\dots,\mathcal{K}}} \quad \text{für } i = 1,\dots,N_{\mathcal{P}}.$$

Weiter seien durch $\mathbf{M}_t^E, \mathbf{A}_t^E \in \mathbb{R}^{(\mathcal{K}+1)\times(\mathcal{K}+1)}$ und $\mathbf{M}_t^F \in \mathbb{R}^{\mathcal{K}\times\mathcal{K}}$ die Gramschen Matrizen

$$\mathbf{M}_t^E := \big[\langle \vartheta_k, \vartheta_l\rangle_{L_2(I)}\big]_{k,l=0,\dots,\mathcal{K}}, \qquad \mathbf{A}_t^E := \big[\langle \vartheta_k', \vartheta_l'\rangle_{L_2(I)}\big]_{k,l=0,\dots,\mathcal{K}},$$
$$\mathbf{M}_t^F := \big[\langle \xi_k, \xi_l\rangle_{L_2(I)}\big]_{k,l=1,\dots,\mathcal{K}}$$

definiert sowie ein Zeilenvektor $\mathbf{e}_t^E \in \mathbb{R}^{1\times(\mathcal{K}+1)}$ durch $\mathbf{e}_t^E := \big[\vartheta_k(0)\big]_{k=0,\dots,\mathcal{K}}$.

Definition 4.15. Die räumlichen Gramschen Matrizen $\mathbf{H}_x, \mathbf{A}_x, \mathbf{V}_x, \mathbf{W}_{x,i} \in \mathbb{R}^{\mathcal{J}\times\mathcal{J}}$ seien als

$$\mathbf{H}_x := [\langle \eta_k, \eta_l\rangle_H]_{k,l=1,\dots,\mathcal{J}}, \qquad \mathbf{A}_x := [\langle \nabla\eta_k, \nabla\eta_l\rangle_H]_{k,l=1,\dots,\mathcal{J}},$$
$$\mathbf{V}_x := [\langle \eta_k, \eta_l\rangle_V]_{k,l=1,\dots,\mathcal{J}}, \qquad \mathbf{W}_{x,i} := [\langle \varphi_i\eta_k, \eta_l\rangle_H]_{k,l=1,\dots,\mathcal{J}} \quad \text{für } i = 1,\dots,N_{\mathcal{P}},$$

definiert.

Da nach Bemerkung 3.3 das V-Skalarprodukt durch $\langle\cdot,\cdot\rangle_V = \langle\cdot,\cdot\rangle_H + \langle\nabla\cdot,\nabla\cdot\rangle_H$ gegeben ist, gilt insbesondere $\mathbf{V}_x = \mathbf{H}_x + \mathbf{A}_x$.

Als letzten Teil der Vorbereitung beweisen wir das folgende allgemein gehalten Lemma, welches eine Möglichkeit liefert, die Rieszsche Darstellung eines Funktionals zu bestimmen.

Lemma 4.16. *Sei X ein endlichdimensionaler Hilbertraum mit Basis $\{\phi_i\}_{i=1,\dots,N}$. Weiter seien $g \in X'$ ein stetiges lineares Funktional und $v \in X$ dessen Rieszsche Darstellung, es gilt also $\langle g, w\rangle_{X'\times X} = \langle v, w\rangle_X$ für alle $w \in X$. Dann ist der Koeffizientenvektor $\mathbf{v} \in \mathbb{R}^N$ von v gegeben durch das lineare Gleichungssystem*

$$\mathbf{X}\mathbf{v} = \mathbf{g},$$

wobei $\mathbf{X} := [\langle\phi_i, \phi_j\rangle_X]_{i,j=1,\dots,N} \in \mathbb{R}^{N\times N}$ und $\mathbf{g} := [\langle g, \phi_i\rangle_{X'\times X}]_{i=1,\dots N}$ seien.

Beweis. Einsetzen einer beliebigen Testfunktion $w = \sum_{i=1}^N w_i\phi_i$ liefert

$$\langle g, w\rangle_{X'\times X} = \sum_{i=1}^N w_i\langle g, \phi_i\rangle_{X'\times X} = \mathbf{w}^\mathsf{T}\mathbf{g} = \mathbf{w}^\mathsf{T}\mathbf{X}\mathbf{v} = \langle\sum_{i=1}^N w_i\phi_i, \sum_{j=1}^N v_j\phi_j\rangle_X = \langle v, w\rangle_X. \quad \square$$

Mit Hilfe dieser Vorarbeit können nun alle benötigten Raum-Zeit-Objekte für die Diskretisierung hergeleitet werden. Wir beginnen mit dem Hauptanteil, der durch die Bilinearform erzeugten Systemmatrix, fahren danach mit dem Lastvektor fort und schließen mit der diskreten Darstellung der Normen und einer Möglichkeit zur Berechnung der diskreten inf-sup-Konstante ab.

Um die Notation möglichst knapp zu halten, verzichten wir sowohl auf die Argumente der Funktionen als auch auf das Symbol $\otimes$ für das Tensorprodukt.

Systemmatrix. Wir leiten die diskrete Darstellung für die einzelnen parameterunabhängigen Bausteine der affin-parametrischen Darstellung (4.10) her. Dazu werten wir die Bilinearformen für die Basisfunktionen $\vartheta_k \otimes \eta_j \in \mathcal{X}_\mathcal{N}$ sowie $(\xi_m \otimes \eta_l, 0), (0, \eta_n) \in \mathcal{Y}_\mathcal{N}$ aus und vereinfachen unter Verwendung der Tensorprodukt-Struktur soweit möglich. Im Fall von b_0 führt dies zu

$$\begin{aligned} &b_0(\vartheta_k \otimes \eta_j, (\xi_m \otimes \eta_l, 0)) \\ &= \int_I \left[\langle \vartheta_k' \eta_j, \xi_m \eta_l \rangle_{V' \times V} + c\langle \vartheta_k \nabla \eta_j, \xi_m \nabla \eta_l \rangle_H + \mu \langle \vartheta_k \eta_j, \xi_m \eta_l \rangle_H \right] \mathrm{d}t \\ &= \langle \vartheta_k', \xi_m \rangle_{L_2(I)} \langle \eta_j, \eta_l \rangle_H + c\langle \vartheta_k, \xi_m \rangle_{L_2(I)} \langle \nabla \eta_j, \nabla \eta_l \rangle_H + \mu \langle \vartheta_k, \xi_m \rangle_{L_2(I)} \langle \eta_j, \eta_l \rangle_H \end{aligned}$$

und

$$b_0(\vartheta_k \otimes \eta_j, (0, \eta_n)) = \langle \vartheta_k(0) \eta_j, \eta_n \rangle_H = \delta_{k,0} \langle \eta_j, \eta_n \rangle_H.$$

Für die restlichen b_i mit $i = 1, \dots, N_\mathcal{P}$ ergibt sich

$$b_i(\vartheta_k \otimes \eta_j, (\xi_m \otimes \eta_l, 0)) = \int_I \chi_i \langle \varphi_i \vartheta_k \eta_j, \xi_m \eta_l \rangle_H \, \mathrm{d}t = \langle \chi_i \vartheta_k, \xi_m \rangle_{L_2(I)} \langle \varphi_i \eta_j, \eta_l \rangle_H$$

sowie

$$b_i(\vartheta_k \otimes \eta_j, (0, \eta_n)) = 0.$$

Diese Darstellungen lassen sich nun mit Hilfe der zuvor definierten Gramschen Matrizen und des üblichen Kronecker-Produkts $\otimes$ zu Matrizen $\mathbf{B}_i \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$, $i = 0, \dots, N_\mathcal{P}$, der Form

$$\mathbf{B}_0 := \begin{bmatrix} \mathbf{C}_t^{FE} \otimes \mathbf{H}_x + c\,\mathbf{M}_t^{FE} \otimes \mathbf{A}_x + \mu \mathbf{M}_t^{FE} \otimes \mathbf{H}_x \\ \mathbf{e}_t^E \otimes \mathbf{H}_x \end{bmatrix}, \quad \mathbf{B}_i := \begin{bmatrix} \mathbf{M}_{t,i}^{FE} \otimes \mathbf{W}_{x,i} \\ \mathbf{0} \end{bmatrix}$$

zusammenfassen. Dadurch gilt nun für die diskreten Ansatz- und Testfunktionen aus Definition 4.13 offenbar die Gleichung

$$b_i(u_\mathcal{N}, (y_{\mathcal{K}\mathcal{J}}, z_\mathcal{J})) = (\mathbf{y}_{\mathcal{K}\mathcal{J}}; \mathbf{z}_\mathcal{J})^\mathsf{T} \mathbf{B}_i \mathbf{u}_\mathcal{N}, \quad \text{für } i = 0, \dots, N_\mathcal{P}.$$

Weiter gilt die affin-parametrische Darstellung (4.10) auch für diese Matrizen, sodass die Systemmatrix $\mathbf{B}(\boldsymbol{\sigma}) \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$ für ein $\boldsymbol{\sigma} \in \mathcal{P}$ durch

$$\mathbf{B}(\boldsymbol{\sigma}) := \mathbf{B}_0 + \sum_{i=1}^{N_\mathcal{P}} \sigma_i \mathbf{B}_i$$

ausgewertet werden kann.

Lastvektor. Als nächsten Schritt leiten wir eine diskrete Darstellung der rechten Seite des Variationsproblems (4.9) her. Diese hat nach Definition die Form

$$f(v) = \int_I \langle g, v_1 \rangle_{V' \times V} \, \mathrm{d}t + \langle u_0, v_2 \rangle_H.$$

Da der Quellterm $g \in L_2(I; V')$ im Allgemeinen nicht in rein zeit- und raumabhängige Funktionen zerlegt werden kann, wenden wir aufgrund der verwendeten stückweise konstanten Basisfunktionen für den Testraum in der zeitlichen Komponente eine Trapezformel als numerische Quadratur an. Dies führt für das Tupel $(\xi_k \otimes \eta_j, \eta_l) \in \mathcal{Y}_\mathcal{N}$ zu folgender Darstellung

$$\begin{aligned} f((\xi_k \otimes \eta_j, \eta_l)) &= \int_I \langle g, \xi_k \eta_j \rangle_{V' \times V} \, \mathrm{d}t + \langle u_0, \eta_l \rangle_H \\ &= \tfrac{1}{2} \Delta t_k \langle g(t_k) + g(t_{k-1}), \eta_j \rangle_{V' \times V} + \langle u_0, \eta_l \rangle_H, \end{aligned}$$

wobei $\Delta t_k := t_k - t_{k-1}$ die Schrittweite des Zeitgitters sei. Wir definieren die Vektoren $\mathbf{u}_0, \mathbf{g}^{k-1/2} \in \mathbb{R}^\mathcal{J}$ für $k = 1, \ldots, \mathcal{K}$ als

$$\mathbf{u}_0 := \left[\langle u_0, \eta_l \rangle_H \right]_{l=1,\ldots,\mathcal{J}}, \qquad \mathbf{g}^{k-1/2} := \left[\tfrac{1}{2} \langle g(t_k) + g(t_{k-1}), \eta_j \rangle_{V' \times V} \right]_{j=1,\ldots,\mathcal{J}}.$$

Die hier vorkommenden Skalarprodukte lassen sich nun mittels geeigneter Quadraturformeln auswerten, so dass wir den diskreten Lastvektor $\mathbf{f} \in \mathbb{R}^\mathcal{N}$ als

$$\mathbf{f} := \begin{bmatrix} \mathbf{g}^{1/2} \\ \vdots \\ \mathbf{g}^{\mathcal{K}-1/2} \\ \mathbf{u}_0 \end{bmatrix}$$

definieren können.

Bestimmung der diskreten Lösung. Die beschriebene Systemmatrix $\mathbf{B}(\boldsymbol{\sigma})$ und der Lastvektor $\mathbf{f}$ reichen bereits aus, um die Petrov-Galerkin-Lösung $u_\mathcal{N}(\boldsymbol{\sigma})$ des Variationsproblems (4.9) zu bestimmen. Dazu muss lediglich der Koeffizientenvektor $\mathbf{u}_\mathcal{N}(\boldsymbol{\sigma})$ der Lösung durch das Gleichungssystem

$$\mathbf{B}(\boldsymbol{\sigma}) \mathbf{u}_\mathcal{N}(\boldsymbol{\sigma}) = \mathbf{f}$$

bestimmt werden.

Normen. Weiter benötigen wir eine Möglichkeit, die Norm einer diskreten Ansatzfunktion beziehungsweise Testfunktion mit Hilfe des entsprechenden Koeffizientenvektors zu bestimmen. Dazu sei zunächst an das norminduzierende $\mathcal{X}$-Skalarprodukt

$$\langle u, v \rangle_\mathcal{X} = \langle u, v \rangle_{L_2(I;V)} + \langle u_t, v_t \rangle_{L_2(I;V')} \quad \text{für } u, v \in \mathcal{X} \tag{2.3}$$

erinnert. Einsetzen der Basisfunktionen des Ansatzraumes liefert

$$\langle \vartheta_k \otimes \eta_j, \vartheta_m \otimes \eta_l \rangle_\mathcal{X} = \int_I \langle \vartheta_k \eta_j, \vartheta_m \eta_l \rangle_V \, \mathrm{d}t + \int_I \langle \vartheta_k' \eta_j, \vartheta_m' \eta_l \rangle_{V'} \, \mathrm{d}t$$

$$= \langle \vartheta_k, \vartheta_m \rangle_{L_2(I)} \langle \eta_j, \eta_l \rangle_V + \langle \vartheta_k', \vartheta_m' \rangle_{L_2(I)} \langle \eta_j, \eta_l \rangle_{V'}$$

Bevor wir diese Darstellung zu einer handlichen Matrix-Schreibweise umformulieren können, müssen wir zunächst klären, wie das diskrete V'-Skalarprodukt bestimmt werden kann. Dazu verwenden wir Lemma 4.16 und erhalten für jedes $n = 1, \dots, \mathcal{J}$ für das Funktional $\langle \eta_n, \cdot \rangle_{V' \times V}$ den Koeffizientenvektor $\mathbf{h}_n = \mathbf{V}_x^{-1} \boldsymbol{\eta}_n$ mit $\boldsymbol{\eta}_n := [\langle \eta_n, \eta_j \rangle_{V' \times V}]_{j=1,\dots,\mathcal{J}}$. Nach dem Rieszschen Darstellungssatz gilt nun unter Verwendung der durch $\mathbf{h}_n \in \mathbb{R}^{\mathcal{J}}$, $n = 1, \dots, \mathcal{J}$, definierten $h_n \in V$ die Gleichung

$$\langle \eta_j, \eta_l \rangle_{V'} = \langle h_j, h_l \rangle_V = (\mathbf{V}_x^{-1} \boldsymbol{\eta}_j)^\mathsf{T} \mathbf{V}_x (\mathbf{V}_x^{-1} \boldsymbol{\eta}_l) = \boldsymbol{\eta}_j^\mathsf{T} \mathbf{V}_x^{-1} \boldsymbol{\eta}_l.$$

Die Gramsche Matrix $\mathbf{V}_{\text{dual}} \in \mathbb{R}^{\mathcal{J} \times \mathcal{J}}$ des V'-Skalarprodukts ist dementsprechend

$$\mathbf{V}_{\text{dual}} := [\boldsymbol{\eta}_j^\mathsf{T} \mathbf{V}_x^{-1} \boldsymbol{\eta}_l]_{j,l=1,\dots,\mathcal{J}} = \mathbf{H}_x^\mathsf{T} \mathbf{V}_x^{-1} \mathbf{H}_x.$$

Weiter können wir damit die Gramsche Matrix $\mathbf{X} \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$ des $\mathcal{X}$-Skalarprodukts auf $\mathcal{X}_\mathcal{N}$ durch

$$\mathbf{X} := \mathbf{M}_t^E \otimes \mathbf{V}_x + \mathbf{A}_t^E \otimes (\mathbf{H}_x^\mathsf{T} \mathbf{V}_x^{-1} \mathbf{H}_x)$$

definieren und können nun für alle $u \in \mathcal{X}_\mathcal{N}$ die $\mathcal{X}$-Norm über $\|u\|_\mathcal{X}^2 = \mathbf{u}^\mathsf{T} \mathbf{X} \mathbf{u}$ bestimmen.

Analog leiten wir auch eine diskrete Darstellung der Testraumnorm her. Auch hier sei an das norminduzierende $\mathcal{Y}$-Skalarprodukt

$$\langle u, v \rangle_\mathcal{Y} = \langle u_1, v_1 \rangle_{L_2(I;V)} + \langle u_2, v_2 \rangle_H \quad \text{für } u, v \in \mathcal{Y} \tag{2.5}$$

erinnert. Die Auswertung dieses Skalarprodukts für die Funktionen aus $\mathcal{Y}_\mathcal{N}$ liefert analog zur vorherigen Betrachtung

$$\langle (\xi_k \otimes \eta_j, \eta_l), (\xi_n \otimes \eta_m, \eta_o) \rangle_\mathcal{Y} = \langle \xi_k, \xi_n \rangle_{L_2(I)} \langle \eta_j, \eta_m \rangle_V + \langle \eta_l, \eta_o \rangle_H.$$

Zusammengefasst führt dies zu der Gramschen Matrix $\mathbf{Y} \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$, gegeben durch

$$\mathbf{Y} := \begin{bmatrix} \mathbf{M}_t^F \otimes \mathbf{V}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_x \end{bmatrix}.$$

Berechnung der inf-sup-Konstanten. Um die inf-sup-Konstante $\beta_\mathcal{N}(\boldsymbol{\sigma})$ zu berechnen, greifen wir auf den Supremizing-Operator aus Definition 4.6 zurück und nutzen die in Lemma 4.7 hergeleitete alternative Darstellung von $\beta_\mathcal{N}(\boldsymbol{\sigma})$.

Seien $\boldsymbol{\sigma} \in \mathcal{P}$ und $T_{\boldsymbol{\sigma}}$ der Supremizing-Operator von $b(\cdot, \cdot; \boldsymbol{\sigma})|_{\mathcal{X}_\mathcal{N} \times \mathcal{Y}_\mathcal{N}}$. Die Matrixdarstellung von $T_{\boldsymbol{\sigma}}$ ergibt sich nach dem Rieszschen Darstellungssatz durch

$$\mathbf{T}_{\boldsymbol{\sigma}} = \mathbf{Y}^{-1} \mathbf{B}(\boldsymbol{\sigma}) \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}.$$

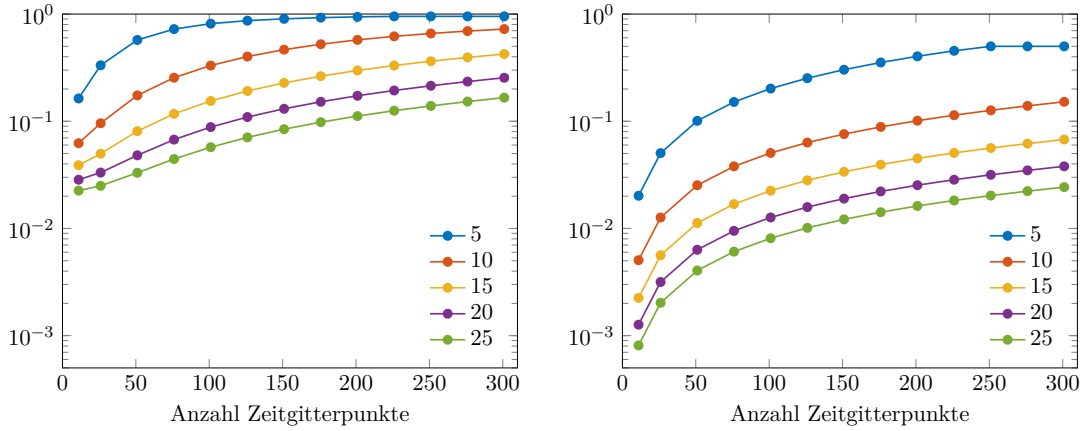
Dies führt weiter zu

$$\beta_\mathcal{N}^2(\boldsymbol{\sigma}) = \inf_{u \in \mathcal{X}_\mathcal{N}} \frac{\langle T_{\boldsymbol{\sigma}} u, T_{\boldsymbol{\sigma}} u \rangle_\mathcal{Y}}{\|u\|_\mathcal{X}^2} = \inf_{u \in \mathbb{R}^\mathcal{N}} \frac{\mathbf{u}^\mathsf{T} \mathbf{T}_{\boldsymbol{\sigma}}^\mathsf{T} \mathbf{Y} \mathbf{T}_{\boldsymbol{\sigma}} \mathbf{u}}{\mathbf{u}^\mathsf{T} \mathbf{X} \mathbf{u}} = \inf_{u \in \mathbb{R}^\mathcal{N}} \frac{\mathbf{u}^\mathsf{T} \mathbf{B}(\boldsymbol{\sigma})^\mathsf{T} \mathbf{Y}^{-1} \mathbf{B}(\boldsymbol{\sigma}) \mathbf{u}}{\mathbf{u}^\mathsf{T} \mathbf{X} \mathbf{u}}.$$

Es ist bekannt (vergleiche [PR06, Subsection 1.3.5]) dass der Ausdruck auf der rechten Seite mit Hilfe des verallgemeinerten Eigenwertproblems

$$\mathbf{B}(\boldsymbol{\sigma})^\mathsf{T} \mathbf{Y}^{-1} \mathbf{B}(\boldsymbol{\sigma}) \mathbf{v} = \lambda \mathbf{X} \mathbf{v}$$

berechnet werden kann, da das gesuchte Infimum gerade dem minimalen Eigenwert $\lambda_{\min}$ entspricht. Insbesondere gilt damit $\beta_\mathcal{N}(\boldsymbol{\sigma}) = \sqrt{\lambda_{\min}}$.



(a) Exakte inf-sup-Konstante $\beta_{\mathcal{N}}$.

(b) Untere Schranke nach Satz 4.11.

Abbildung 4.1. Zur Stabilität der Diskretisierung des ersten Beispiels. Die Farben repräsentieren dabei die verschiedenen Werte für die Dimension $\mathcal{J}$ der räumlichen Diskretisierung $V_{\mathcal{J}}$, während auf der horizontalen Achse die Anzahl $\mathcal{K}$ der Zeitgitterpunkte aufgetragen ist.

4.4. Beispiele

Wir widmen uns nun einigen Beispielen, wobei wir uns im Rahmen dieser Arbeit aus laufzeittechnischen Gründen auf den Fall einer Raumdimension beschränken wollen.

Zunächst erinnern wir an dieser Stelle an einige Modellgrößen. Das Zeitintervall sei durch $I := [0, T]$ für ein $0 < T < \infty$ gegeben. Ferner sei durch $T_f \in \mathbb{R}$ mit $0 < T_f < T$ der Zeitpunkt des Feldwechsels gegeben. Für den räumlichen Anteil sei ohne Einschränkung $\Omega := [0, L] \subset \mathbb{R}$ mit einem $L > 0$ gegeben.

Uns interessiert an dieser Stelle hauptsächlich die Stabilität der in diesem Kapitel beschriebenen Diskretisierung, weswegen wir den parametrischen Aspekt der betrachteten Raum-Zeit-Variationsformulierung erst im nächsten Kapitel weiterverfolgen werden.

Homogene Randbedingungen. In Abschnitt 4.2 wurde bisher nur die zeitliche Diskretisierung festgelegt. Es bleibt an dieser Stelle also zu klären, wie die räumliche Diskretisierung $V_{\mathcal{J}}$ gewählt wird. Wie in dieser Arbeit bereits mehrfach angeführt, weisen die Felder und die Lösungen der Propagator-Differentialgleichung aus Kapitel 1 oftmals Symmetrien und hohe Regularität auf, weswegen wir hierfür keinen Finite-Elemente-, sondern einen Fourier-Ansatz mit global definierten Funktionen verfolgen. Um die homogenen Randbedingungen direkt in die Basisfunktionen einzubauen, wählen wir

$$V_{\mathcal{J}} := \operatorname{span}\{\sin(\pi j \cdot / L) \,|\, j = 1, \dots, \mathcal{J}\}.$$

Als erstes einfaches Modell betrachten wir die Propagator-Differentialgleichung ohne zeitlichen Wechsel der Felder. Konkret wählen wir $I = [0, 1]$ sowie $\Omega = [0, 1]$ und ferner den Differentialoperator $A(t) \equiv A \colon V \to V'$ als

$$A\eta = -\Delta\eta + \eta + \sin(\pi \cdot)\eta.$$

Dies entspricht einer Verschiebung um $\mu = 1$ und einem zeitlich konstanten Feld $\omega \colon \Omega \to \mathbb{R}$, $\omega = \sin(\pi \cdot)$ mit $\|\omega\|_{L_\infty(I; L_\infty(\Omega))} = 1 = \mu$. Damit wird in Satz 3.7 die Gårding-Ungleichung

mit $\lambda = 0$ erfüllt und wir können die in Satz 4.11 beschriebene untere Schranke für die inf-sup-Konstante $\beta_{\mathcal{N}}$ auswerten.

Abbildung 4.1 zeigt sowohl die exakt bestimmte diskrete inf-sup-Konstante $\beta_{\mathcal{N}}$ als auch die untere Schranke für verschiedene Werte der Dimension $\mathcal{J}$ der räumlichen Diskretisierung sowie der Anzahl $\mathcal{K}$ an Zeitgitterpunkten. Zwar kann an dieser Stelle die Konstante c_0 aus Satz 4.11 nicht näher bestimmt werden, aber es ist auf den ersten Blick ersichtlich, dass die Schranke das tatsächliche Verhalten der inf-sup-Konstante qualitativ gut widerspiegelt. Ferner wird die CFL-Bedingung bekräftigt, da sich die inf-sup-Konstante für festes $\mathcal{J}$ und wachsendes $\mathcal{K}$ von unten dem Wert 1 annähert.

Das zweite, etwas komplexere Beispiel enthält nun einen zeitlichen Wechsel bei $T_f = 0.5$. Der Differentialoperator $A(t)\colon V \to V'$ sei gegeben durch

$$\begin{aligned} A(t)\eta = &- \Delta\eta + 3\eta + \chi_{[0,0.5)}(t)\,[\sin(\pi\,\cdot) - \sin(6\pi\,\cdot)]\,\eta \\ &+ \chi_{[0.5,1]}(t)\,[\sin(3\pi\,\cdot) + \sin(6\pi\,\cdot)]\,\eta. \end{aligned}$$

Erneut ist mit $\mu = 3$ sichergestellt, dass die Gårding-Ungleichung mit $\lambda = 0$ erfüllt wird. Abbildung 4.2 zeigt die Ergebnisse für diese Modelldaten. Auffallend ist, dass die resultierenden Werte nahezu identisch sind mit denen des einfachen Beispiels. Dies ist dadurch bedingt, dass in beiden Beispielen die Verschiebung μ so gewählt wurde, dass der Differentialoperator $A(t)$ elliptisch wird.

Periodische Randbedingungen. Auch auf die periodischen Randbedingungen wollen wir an dieser Stelle kurz eingehen und das zweite Beispiel nun in diesem Setting noch einmal betrachten. Dabei übernehmen wir die dortigen Gegebenheiten und passen lediglich die verwendete Diskretisierung $V_{\mathcal{J}}$ an. Diese wählen wir erneut in Form eines Fourier-Ansatzes, diesmal als

$$V_{\mathcal{J}} := \operatorname{span}\{v_j \,|\, j = 1, \dots, \mathcal{J}\}$$

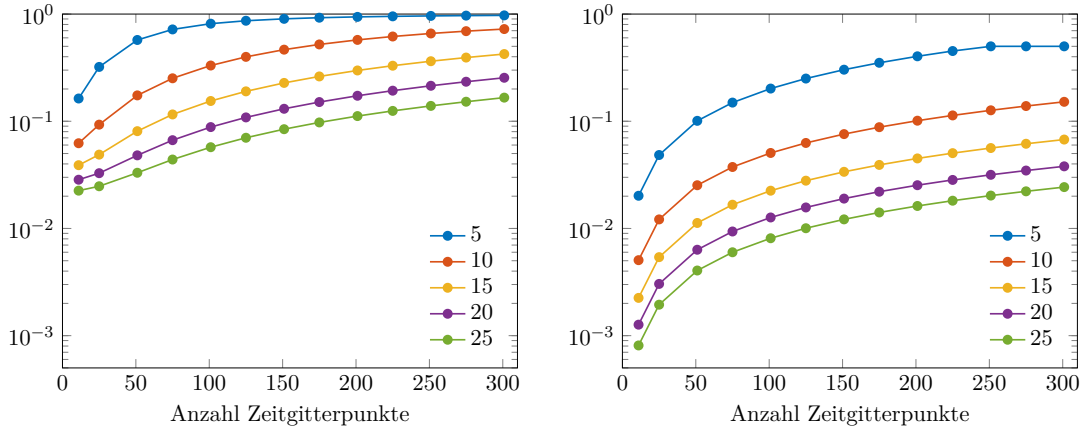
mit den Basisfunktionen

$$v_j := \begin{cases} \cos(\pi(j-1)\cdot/L), & \text{falls } j \text{ ungerade,} \\ \sin(\pi j\cdot/L), & \text{falls } j \text{ gerade.} \end{cases}$$

Nach den Aussagen in Abschnitt 3.4 können wir ohne Transformation im Sinne von Proposition 2.24 die Gårding-Ungleichung im Allgemeinen nicht mit $\lambda = 0$ erfüllen. Wir verzichten an dieser Stelle darauf, da diese im Wesentlichen nur für die Invertierbarkeit des Operators $A(t)$ benötigt wird und diese hier, wie bei der numerischen Umsetzung deutlich wurde, bereits gegeben ist.

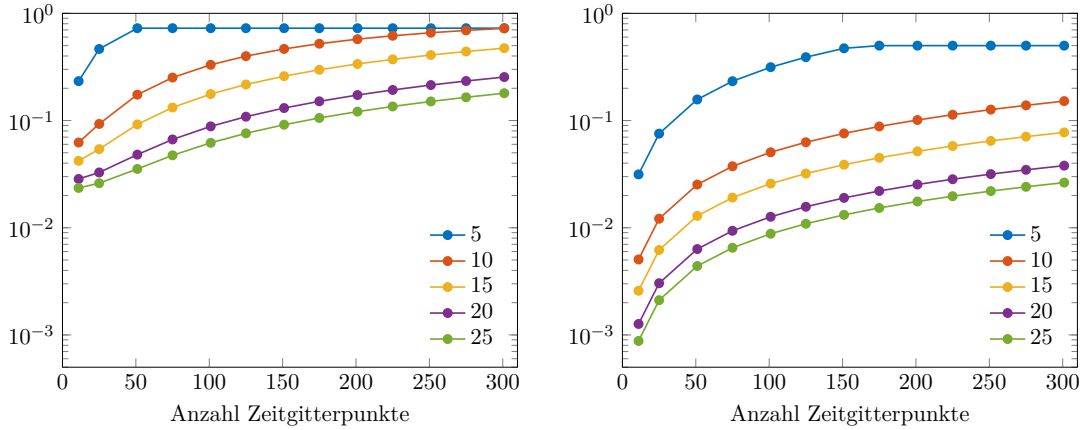
Die Ergebnisse sind in Abbildung 4.3 zu sehen. Diese weisen im Vergleich zum Fall homogener Randbedingungen nur kleine Differenzen auf und liefern sogar leicht bessere, das heißt größere Werte bei der exakt bestimmten diskreten inf-sup-Konstante.

Sowohl im Fall homogener als auch periodischer Randbedingungen bekräftigen die vorliegenden Ergebnisse die Wahl der Diskretisierung. Zwar handelt es sich dabei um relativ simple Modelle, allerdings legen diese bereits nahe, dass mit der Anpassung der Verschiebung μ der Felder ein Instrument zur Verfügung steht, mit dem auch in komplizierteren Fällen gute Stabilitätsergebnisse erzielt werden können.



(a) Exakte inf-sup-Konstante $\beta_{\mathcal{N}}$. **(b)** Untere Schranke nach Satz 4.11.

Abbildung 4.2. Ergebnisse des zweiten Beispiels mit homogenen Randbedingungen. Erneut repräsentieren die Farben die verschiedenen Werte für $\mathcal{J}$ und die horizontale Achse die Dimension $\mathcal{K}$.



(a) Exakte inf-sup-Konstante $\beta_{\mathcal{N}}$. **(b)** Untere Schranke nach Satz 4.11.

Abbildung 4.3. Ergebnisse des zweiten Beispiels im Falle periodischer Randbedingungen.

5 Kapitel

Reduzierte-Basis-Methode

Aufbauend auf dem Petrov-Galerkin-Verfahren des vorherigen Kapitels führen wir nun die Reduzierte-Basis-Methode (kurz RB-Methode) ein. Zunächst folgen einige theoretische Grundlagen und Ausführungen zur numerischen Umsetzung dieser Methode, bevor wir mit beispielhaften Anwendungen auf die betrachtete Problemstellung abschließen.

5.1. Grundlagen

Wir beginnen mit einer kurzen Motivation und orientieren uns in diesem Abschnitt an den Arbeiten von Rozza, Huynh und Patera [RHP08] sowie Patera und Rozza [PR06], welche einen weit tieferen Einblick bieten. Bei den Reduzierte-Basis-Methoden handelt es sich um Modellreduktionsverfahren, welche zwar bereits seit etwas mehr als 50 Jahren untersucht werden, aber vor allem in den letzten zehn Jahren viel Aufmerksamkeit und Weiterentwicklung erfahren haben.

Zunächst wollen wir erneut geeignete Rahmenbedingungen in Form eines abstrakten Variationsproblems schaffen. Seien dazu zwei Hilberträume $\mathcal{X}$ und $\mathcal{Y}$ und eine abgeschlossene, konvexe Parametermenge $\mathcal{P} \subset \mathbb{R}^p$ für ein $p \in \mathbb{N}$ gegeben. Weiter seien $b\colon \mathcal{X} \times \mathcal{Y} \times \mathcal{P} \to \mathbb{R}$ eine parametrische stetige Bilinearform und $f\colon \mathcal{Y} \to \mathbb{R}$ ein stetiges lineares Funktional. Wir betrachten das abstrakte parametrische Variationsproblem:

$$\text{Sei } \boldsymbol{\sigma} \in \mathcal{P},\ \text{finde } u(\boldsymbol{\sigma}) \in \mathcal{X} \text{ mit} \quad b(u(\boldsymbol{\sigma}), v; \boldsymbol{\sigma}) = f(v) \quad \text{für alle } v \in \mathcal{Y}. \tag{5.1}$$

Reduzierte-Basis-Methoden bieten sich dann an, wenn das Variationsproblem für einen gegebenen Parameter in Echtzeit oder immer wieder für viele verschiedene Parameter gelöst werden muss. Um dies mit zufriedenstellender Genauigkeit effizient bewerkstelligen zu können, konzentriert man sich auf die Lösungsmenge

$$\mathcal{M} := \{u(\boldsymbol{\sigma}) \in \mathcal{X} \,|\, \boldsymbol{\sigma} \in \mathcal{P}\}.$$

Erfüllt die Bilinearform aus (5.1) gewisse Regularitätseigenschaften bezüglich des Parameters, dann bildet $\mathcal{M}$ oftmals eine Mannigfaltigkeit mit vergleichsweise niedriger Dimension.

Um diese Tatsache für eine Modellreduktion nutzen zu können, müssen wir $\mathcal{M}$ zunächst durch ein diskretes Analogon ersetzen. Dazu werden bei den klassischen Reduzierte-Basis-Methoden Galerkin-Verfahren verwendet, weswegen wir an dieser Stelle das Petrov-Galerkin-Verfahren des vorherigen Kapitels nutzen, um eine abstrakte Diskretisierung durchzuführen. Seien $\mathcal{X}_\mathcal{N} \subset \mathcal{X}$ und $\mathcal{Y}_\mathcal{N} \subset \mathcal{Y}$ Unterräume der Dimension $\mathcal{N} \in \mathbb{N}$. Weiter

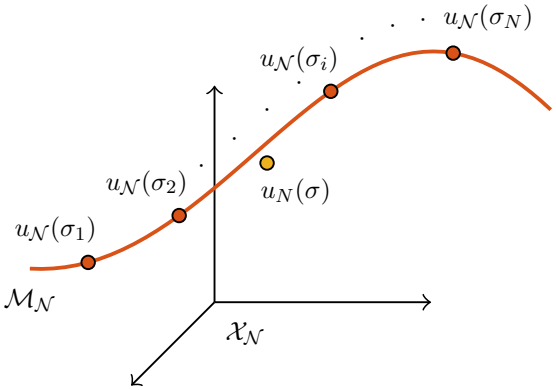


Abbildung 5.1. Skizze der Funktionsweise der Reduzierte-Basis-Methode im Falle eines eindimensionalen Parameters σ. Die Reduzierte-Basis-Lösung $u_N(\sigma)$ ergibt sich als Linearkombinationen der Truth-Lösungen $u_{\mathcal{N}}(\sigma_i)$ für gewisse $\sigma_i \in \mathcal{P}$.

nehmen wir an, dass diese Diskretisierung für alle $\boldsymbol{\sigma} \in \mathcal{P}$ zu einem korrekt gestellten Variationsproblem führt, welches gegeben sei durch:

$$\text{Sei } \boldsymbol{\sigma} \in \mathcal{P}, \text{ finde } u_{\mathcal{N}}(\boldsymbol{\sigma}) \in \mathcal{X}_{\mathcal{N}} \text{ mit } \quad b(u_{\mathcal{N}}(\boldsymbol{\sigma}), v; \boldsymbol{\sigma}) = f(v) \quad \text{für alle } v \in \mathcal{Y}_{\mathcal{N}}. \tag{5.2}$$

Da die RB-Methode maßgeblich auf dieser Petrov-Galerkin-Diskretisierung und nicht dem eigentlichen Variationsproblem (5.1) aufbaut, werden wir (5.2) mit dem Präfix *Truth* kenntlich machen. Die Truth-Lösungsmenge sei nun definiert als

$$\mathcal{M}_{\mathcal{N}} := \{u_{\mathcal{N}}(\boldsymbol{\sigma}) \in \mathcal{X}_{\mathcal{N}} \,|\, \boldsymbol{\sigma} \in \mathcal{P}\}.$$

Dies erlaubt die Konstruktion eines niedrigdimensionalen Unterraums $\mathcal{X}_N \subset \mathcal{M}_{\mathcal{N}} \subset \mathcal{X}_{\mathcal{N}}$ der Dimension N, welcher als Ansatzraum eines weiteren Galerkin-Verfahrens dienen wird. Dies ist die Grundidee der RB-Methoden, die in Abbildung 5.1 noch einmal skizzenhaft veranschaulicht wird.

Die Konstruktion des niedrigdimensionalen Ansatzraumes $\mathcal{X}_N$ erfolgt unter dem Aspekt, den durch das zweite Galerkin-Verfahren zusätzlich eingeführten Fehler für alle Parameter des Parameterraums $\mathcal{P}$ zu minimieren, was in einem hohen Rechenaufwand resultiert. Dies motiviert die bei RB-Methoden übliche Zerlegung in zwei Stufen. Die erste Stufe, die sogenannte *Offline-Phase*, dient dazu, den Ansatzraum $\mathcal{X}_N$ unter dem genannten Gesichtspunkt zu konstruieren und muss nur einmal ausgeführt werden. In der zweiten Stufe, der *Online-Phase*, kann dieses niedrigdimensionale System verwendet werden, um Lösungen für gegebene Parameter auszuwerten und eine garantierte Fehlerschranke zu berechnen. Da im Allgemeinen $N \ll \mathcal{N}$ gilt, wird hierbei als weiteres Ziel die Unabhängigkeit der Berechnungen in der Online-Phase von der Dimension $\mathcal{N}$ verfolgt.

Weiter führt diese Zerlegung zu einem wichtigen Kriterium für beziehungsweise gegen die Anwendung der RB-Methode, da der benötigte Aufwand für die Offline-Phase meist erst ab einer hohen Anzahl an zu bestimmenden Lösungen in der Online-Phase amortisiert wird.

Wir widmen uns nun den formalen Grundlagen der RB-Methode und beginnen mit einer grundlegenden Definition.

Definition 5.1. Es seien $N \in \mathbb{N}$ und $\mathcal{S}_N := \{\boldsymbol{\sigma}_n \,|\, n = 1, \dots, N\} \subset \mathcal{P}$ die sogenannte *Sample-Menge*. Wir bezeichnen die Truth-Lösungen $u_{\mathcal{N}}(\boldsymbol{\sigma}_n)$ als *Snapshots* und definieren den *Reduzierte-Basis-Ansatzraum* $\mathcal{X}_N$ als

$$\mathcal{X}_N := \operatorname{span}\{u_{\mathcal{N}}(\boldsymbol{\sigma}_n) \,|\, n = 1, \dots, N\} = \operatorname{span}\{\xi_n \,|\, n = 1, \dots, N\}, \tag{5.3}$$

wobei $\Xi_N := \{\xi_n \,|\, n = 1, \dots, N\}$ eine geeignete Orthonormalbasis ist. Außerdem definieren wir einen N-dimensionalen *Reduzierte-Basis-Testraum* $\mathcal{Y}_N \subset \mathcal{Y}_{\mathcal{N}}$. Als *Reduzierte-Basis-Variationsproblem* von (5.2) bezeichnen wir:

$$\text{Sei } \boldsymbol{\sigma} \in \mathcal{P}, \text{ finde } u_N(\boldsymbol{\sigma}) \in \mathcal{X}_N \text{ mit } \quad b(u_N(\boldsymbol{\sigma}), v; \boldsymbol{\sigma}) = f(v) \quad \text{für alle } v \in \mathcal{Y}_N. \tag{5.4}$$

Fernen nennen wir $u_N(\boldsymbol{\sigma})$ *Reduzierte-Basis-Lösung.*

RB-Ansatzräume der Art (5.3) werden in der Literatur zu RB-Methoden auch als *Lagrangeräume* bezeichnet, um sie damit von komplexeren Räumen, wie beispielsweise *Taylor-* und *Hermiteräumen* abzugrenzen. Bei diesen werden neben den Truth-Lösungen auch deren partielle Ableitungen bezüglich der Parameter berücksichtigt. Eine Übersicht darüber findet sich in [PR06, Chapter 3].

Wie zuvor bei den Petrov-Galerkin-Verfahren muss auch bei (5.4) sichergestellt werden, dass es sich dabei um ein korrekt gestelltes Problem handelt. Wir werden später sehen, dass diese Eigenschaft bei geschickter Wahl des Testraumes $\mathcal{Y}_N$ vom Truth-Variationsproblem vererbt wird, und nehmen deswegen an dieser Stelle an, dass dies stets erfüllt ist.

Da es sich bei der RB-Diskretisierung ebenfalls um ein Galerkin-Verfahren handelt, erhalten wir Analoga der Aussagen aus Satz 4.3 sowie Satz 4.4, welche hier nicht wiederholt werden, wobei die Truth-Räume $\mathcal{X}_{\mathcal{N}}$ und $\mathcal{Y}_{\mathcal{N}}$ die Rolle der Hilberträume $\mathcal{X}$ und $\mathcal{Y}$ einnehmen.

Weiter soll an dieser Stelle angemerkt werden, warum wir eine Galerkin-Diskretisierung (5.2) als Grundlage und nicht das eigentliche Variationsproblem (5.1) verwenden. Neben dem offensichtlichen Grund, der Berechenbarkeit der Lösung, ist die zugrundeliegende Prämisse die Folgende: Nach Satz 4.4 kann mit der Petrov-Galerkin-Lösung $u_{\mathcal{N}}(\boldsymbol{\sigma})$ die exakte Lösung $u(\boldsymbol{\sigma})$ unter gewissen Annahmen beliebig gut approximiert werden. Betrachten wir weiter die einfache Abschätzung

$$\|u(\boldsymbol{\sigma}) - u_N(\boldsymbol{\sigma})\|_{\mathcal{X}} \leq \|u(\boldsymbol{\sigma}) - u_{\mathcal{N}}(\boldsymbol{\sigma})\|_{\mathcal{X}} + \|u_{\mathcal{N}}(\boldsymbol{\sigma}) - u_N(\boldsymbol{\sigma})\|_{\mathcal{X}},$$

dann entspricht dies gerade der Tatsache, dass der erste Summand beliebig klein gehalten werden kann. Dies erlaubt uns, lediglich den Fehler zwischen Truth- und RB-Lösung zu berücksichtigen und führt weiter zu folgender Definition.

Definition 5.2. Als *Fehler* $e_N(\boldsymbol{\sigma}) \in \mathcal{X}_{\mathcal{N}}$ bezeichnen wir $e_N(\boldsymbol{\sigma}) := u_{\mathcal{N}}(\boldsymbol{\sigma}) - u_N(\boldsymbol{\sigma})$. Weiter definieren wir das *Residuum* $r_N(\cdot\,; \boldsymbol{\sigma}) \colon \mathcal{Y}_{\mathcal{N}} \to \mathbb{R}$ als

$$r_N(v; \boldsymbol{\sigma}) := b(e_N(\boldsymbol{\sigma}), v; \boldsymbol{\sigma}). \tag{5.5}$$

Lemma 5.3. *Das Residuum $r_N(\cdot\,; \boldsymbol{\sigma})$ ist für alle $\boldsymbol{\sigma} \in \mathcal{P}$ ein stetiges lineares Funktional, kurz also $r_N(\cdot\,; \boldsymbol{\sigma}) \in \mathcal{Y}'_{\mathcal{N}}$.*

Beweis. Sowohl Linearität als auch Stetigkeit sind direkt ersichtlich, denn nach Definition erhalten wir

$$\begin{aligned} r_N(v;\boldsymbol{\sigma}) &= b(e_N(\boldsymbol{\sigma}), v;\boldsymbol{\sigma}) = b(u_{\mathcal{N}}(\boldsymbol{\sigma}), v;\boldsymbol{\sigma}) - b(u_N(\boldsymbol{\sigma}), v;\boldsymbol{\sigma}) \\ &= f(v) - b(u_N(\boldsymbol{\sigma}), v;\boldsymbol{\sigma}) \end{aligned} \tag{5.6}$$

und diese Eigenschaften damit von f und b. □

Das Residuum kann nun verwendet werden, um eines der wichtigsten Konstrukte der RB-Methoden zu definieren.

Lemma 5.4. *Sei $\beta_{\mathrm{LB}}(\boldsymbol{\sigma}) > 0$ eine untere Schranke für $\beta_{\mathcal{N}}(\boldsymbol{\sigma})$. Dann gilt*

$$\|u_{\mathcal{N}}(\boldsymbol{\sigma}) - u_N(\boldsymbol{\sigma})\|_{\mathcal{X}} \leq \frac{\|r_N(\cdot;\boldsymbol{\sigma})\|_{\mathcal{Y}'_{\mathcal{N}}}}{\beta_{\mathrm{LB}}(\boldsymbol{\sigma})} =: \Delta_N(\boldsymbol{\sigma}),$$

wobei wir $\Delta_N(\boldsymbol{\sigma})$ als a posteriori-Fehlerschätzer *bezeichnen.*

Beweis. Wir können (5.5) als Variationsproblem (5.2) auffassen, welches unter den getroffenen Annahmen die eindeutige Lösung $e_N(\boldsymbol{\sigma})$ besitzt. Die Abschätzung folgt nun aus Satz 4.3. □

Dieser a posteriori-Fehlerschätzer bildet das Herzstück der RB-Methode, da mit ihm in der Offline-Phase eine möglichst optimale Wahl der Parameter-Samples erreicht werden kann, während er in der Online-Phase zur Berechnung einer garantierten Fehlerschranke zwischen Truth- und RB-Lösung verwendet wird. Dazu setzt man weiter woraus, dass die untere Schranke $\beta_{\mathrm{LB}}(\boldsymbol{\sigma})$ berechenbar ist. Wir werden bei der nachfolgenden numerischen Umsetzung darauf eingehen, wie dies bewerkstelligt werden kann.

Bevor wir uns dieser widmen, wollen wir noch ein Maß für die Güte des Fehlerschätzers einführen.

Lemma 5.5. *Sei $\gamma_{\mathcal{N}}(\boldsymbol{\sigma})$ die Stetigkeitskonstante der Bilinearform $b(\cdot,\cdot;\boldsymbol{\sigma})|_{\mathcal{X}_{\mathcal{N}}\times\mathcal{Y}_{\mathcal{N}}}$. Die* Effektivität *$\eta_N(\boldsymbol{\sigma})$ des a posteriori-Fehlerschätzers ist beschränkt durch*

$$1 \leq \eta_N(\boldsymbol{\sigma}) := \frac{\Delta_N(\boldsymbol{\sigma})}{\|u_{\mathcal{N}}(\boldsymbol{\sigma}) - u_N(\boldsymbol{\sigma})\|_{\mathcal{X}}} \leq \frac{\gamma_{\mathcal{N}}(\boldsymbol{\sigma})}{\beta_{\mathrm{LB}}(\boldsymbol{\sigma})}.$$

Beweis. Die Abschätzung nach unten ergibt sich direkt aus Lemma 5.4. Sei nun $\boldsymbol{\sigma} \in \mathcal{P}$ beliebig. Eine Anwendung des Rieszschen Darstellungssatzes liefert für $r_N(\cdot;\boldsymbol{\sigma})$ ein $\hat{e}_N \in \mathcal{Y}_{\mathcal{N}}$ mit $r_N(v;\boldsymbol{\sigma}) = \langle \hat{e}_N, v\rangle_{\mathcal{Y}}$ für alle $v \in \mathcal{Y}_{\mathcal{N}}$ und $\|r_N(\cdot;\boldsymbol{\sigma})\|_{\mathcal{Y}'_{\mathcal{N}}} = \|\hat{e}_N\|_{\mathcal{Y}}$. Mit der Stetigkeit der Bilinearform erhalten wir damit die Abschätzung

$$\|\hat{e}_N\|_{\mathcal{Y}}^2 = \langle \hat{e}_N, \hat{e}_N\rangle_{\mathcal{Y}} = r_N(\hat{e}_N;\boldsymbol{\sigma}) = b(e_N(\boldsymbol{\sigma}), \hat{e}_N;\boldsymbol{\sigma}) \leq \gamma_{\mathcal{N}}(\boldsymbol{\sigma})\|e_N(\boldsymbol{\sigma})\|_{\mathcal{X}}\|\hat{e}_N\|_{\mathcal{Y}},$$

oder kurz

$$\|\hat{e}_N\|_{\mathcal{Y}} \leq \gamma_{\mathcal{N}}(\boldsymbol{\sigma})\|e_N(\boldsymbol{\sigma})\|_{\mathcal{X}}.$$

Zusammen mit der Definition des a posteriori-Fehlerschätzers $\Delta_N(\boldsymbol{\sigma})$ liefert dies nun

$$\eta_N(\boldsymbol{\sigma}) = \frac{\Delta_N(\boldsymbol{\sigma})}{\|u_{\mathcal{N}}(\boldsymbol{\sigma}) - u_N(\boldsymbol{\sigma})\|_{\mathcal{X}}} \leq \frac{\gamma_{\mathcal{N}}(\boldsymbol{\sigma})\|e_N(\boldsymbol{\sigma})\|_{\mathcal{X}}}{\beta_{\mathrm{LB}}(\boldsymbol{\sigma})\|u_{\mathcal{N}}(\boldsymbol{\sigma}) - u_N(\boldsymbol{\sigma})\|_{\mathcal{X}}} = \frac{\gamma_{\mathcal{N}}(\boldsymbol{\sigma})}{\beta_{\mathrm{LB}}(\boldsymbol{\sigma})}.$$ □

Dieses Ergebnis ist wichtig, denn es garantiert, dass der a posteriori-Fehlerschätzer weder zu pessimistisch noch zu optimistisch ist. Insbesondere garantiert es, dass der tatsächliche Fehler durch den a posteriori-Fehlerschätzer stets korrekt nach oben abgeschätzt wird.

5.2. Numerische Umsetzung

Wir richten unsere Aufmerksamkeit nun auf die numerische Umsetzung der RB-Methode und orientieren uns dabei weiterhin an [RHP08; PR06]. Zunächst führen wir eine praktische Einschränkung der parametrischen Bilinearform des Variationsproblems (5.1) durch und abstrahieren die aus Lemma 4.12 bekannte affine Abhängigkeit.

Definition 5.6. Wir nennen eine parametrische Bilinearform $b\colon \mathcal{X} \times \mathcal{Y} \times \mathcal{P} \to \mathbb{R}$ *parametrisch affin*, wenn sie die Form

$$b(u, v; \boldsymbol{\sigma}) = \sum_{q=1}^{Q_b} \vartheta_q^b(\boldsymbol{\sigma}) b_q(u, v) \quad \text{für } u \in \mathcal{X},\ v \in \mathcal{Y},\ \boldsymbol{\sigma} \in \mathcal{P} \tag{5.7}$$

hat, wobei für $q = 1, \dots, Q_b$ durch $\vartheta_q^b\colon \mathcal{P} \to \mathbb{R}$ Funktionen und durch $b_q\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ parameterunabhängige Bilinearformen gegeben seien.

Für dieses Kapitel nehmen wir ferner an, dass die Funktionen ϑ_q^b effizient ausgewertet werden können. Weiter vereinfachen wir an dieser Stelle die Notation aus Abschnitt 4.3. Die endlichdimensionalen Räume $\mathcal{X}_\mathcal{N}$ und $\mathcal{Y}_\mathcal{N}$ der Truth-Diskretisierung fassen wir als Span entsprechender Basen auf und schreiben

$$\mathcal{X}_\mathcal{N} = \operatorname{span}\{\phi_i \,|\, i = 1, \dots, \mathcal{N}\}, \qquad \mathcal{Y}_\mathcal{N} = \operatorname{span}\{\psi_i \,|\, i = 1, \dots, \mathcal{N}\}.$$

Analog verfahren wir für die N-dimensionalen RB-Räume $\mathcal{X}_N$ und $\mathcal{Y}_N$ mit der Notation

$$\mathcal{X}_N = \operatorname{span}\{\xi_n \,|\, n = 1, \dots, N\}, \qquad \mathcal{Y}_N = \operatorname{span}\{\eta_n \,|\, n = 1, \dots, N\}.$$

Die bei der Petrov-Galerkin-Diskretisierung hergeleiteten Strukturen $\mathbf{X}, \mathbf{Y}, \mathbf{B}_q \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$ mit $q = 1, \dots, Q_b$ und $\mathbf{f} \in \mathbb{R}^\mathcal{N}$ lassen sich damit kurz schreiben als

$$\begin{aligned} \mathbf{X} &= \left[\langle \phi_i, \phi_j \rangle_\mathcal{X}\right]_{i,j=1,\dots,\mathcal{N}}, & \mathbf{Y} &= \left[\langle \psi_i, \psi_j \rangle_\mathcal{Y}\right]_{i,j=1,\dots,\mathcal{N}}, \\ \mathbf{f} &= \left[f(\psi_j)\right]_{j=1,\dots,\mathcal{N}}, & \mathbf{B}_q &= \left[b_q(\phi_i, \psi_j)\right]_{j,i=1,\dots,\mathcal{N}}. \end{aligned}$$

Nach diesen Vorbemerkungen geben wir nun als ersten Schritt eine Möglichkeit zur Bestimmung der Reduzierte-Basis-Äquivalente dieser Strukturen an.

Definition 5.7. Als *Reduzierte-Basis-Ansatzfunktion* $u_N \in \mathcal{X}_N$ respektive *Testfunktion* $v_N \in \mathcal{Y}_N$ bezeichnen wir die Linearkombinationen

$$u_N := \sum_{n=1}^{N} u_{n,N} \xi_n, \qquad v_N := \sum_{n=1}^{N} v_{n,N} \eta_n$$

mit den Koeffizientenvektoren $\mathbf{u}_N := [u_{n,N}]_{n=1,\dots,N}, \mathbf{v}_N := [v_{n,N}]_{n=1,\dots,N} \in \mathbb{R}^N$.

Nach Konstruktion können die Basisfunktionen ξ_n, η_n für alle $n = 1, \dots, N$ jeweils als Linearkombination von $\{\phi_i \,|\, i = 1, \dots, \mathcal{N}\}$ beziehungsweise $\{\psi_i \,|\, i = 1, \dots, \mathcal{N}\}$ geschrieben werden. Bezeichnen wir nun für jedes $n = 1, \dots, N$ die entsprechenden Koeffizientenvektoren mit $\boldsymbol{\xi}_n, \boldsymbol{\eta}_n \in \mathbb{R}^\mathcal{N}$, dann können wir damit weiter die Matrizen

$$\boldsymbol{\Xi} := [\boldsymbol{\xi}_n]_{n=1,\dots,N} \in \mathbb{R}^{\mathcal{N} \times N}, \qquad \mathbf{H} := [\boldsymbol{\eta}_n]_{n=1,\dots,N} \in \mathbb{R}^{\mathcal{N} \times N}$$

definieren. Diese ermöglichen nun die Berechnung der RB-Matrizen $\mathbf{B}_{q,N} \in \mathbb{R}^{N\times N}$ für $q = 1, \ldots, Q_b$ und des RB-Lastvektors $\mathbf{f}_N \in \mathbb{R}^N$ als

$$\begin{aligned} \mathbf{B}_{q,N} &:= [b_q(\xi_n, \eta_m)]_{m,n=1,\ldots,N} &&= \mathbf{H}^\mathsf{T}\mathbf{B}_q\boldsymbol{\Xi}, \quad q = 1, \ldots, Q_b, \\ \mathbf{f}_N &:= [f(\eta_m)]_{m=1,\ldots,N} &&= \mathbf{H}^\mathsf{T}\mathbf{f}. \end{aligned} \tag{5.8}$$

Analog können auch die norminduzierenden Matrizen $\mathbf{X}_N, \mathbf{Y}_N \in \mathbb{R}^{N\times N}$ definiert werden als

$$\begin{aligned} \mathbf{X}_N &:= [\langle \xi_n, \xi_m\rangle_{\mathcal{X}}]_{n,m=1,\ldots,N} &&= \boldsymbol{\Xi}^\mathsf{T}\mathbf{X}\boldsymbol{\Xi}, \\ \mathbf{Y}_N &:= [\langle \eta_n, \eta_m\rangle_{\mathcal{Y}}]_{n,m=1,\ldots,N} &&= \mathbf{H}^\mathsf{T}\mathbf{Y}\mathbf{H}. \end{aligned} \tag{5.9}$$

Die Bestimmung der RB-Lösung $u_N(\boldsymbol{\sigma})$ zu einem Parameter $\boldsymbol{\sigma} \in \mathcal{P}$ erfolgt, wie bereits vom Petrov-Galerkin-Verfahren bekannt, über das lineare Gleichungssystem

$$\mathbf{B}_N(\boldsymbol{\sigma})\mathbf{u}_N(\boldsymbol{\sigma}) = \mathbf{f}_N, \tag{5.10}$$

wobei $\mathbf{u}_N(\boldsymbol{\sigma})$ der Koeffizientenvektor der Lösung ist und die Systemmatrix $\mathbf{B}_N(\boldsymbol{\sigma})$ durch die affine Darstellung

$$\mathbf{B}_N(\boldsymbol{\sigma}) = \sum_{q=1}^{Q_b} \vartheta_q^b(\boldsymbol{\sigma})\mathbf{B}_{q,N} \tag{5.11}$$

ausgewertet werden kann. Damit wird die folgende Zerlegung möglich:

Offline: Berechne und speichere die Strukturen (5.8) und (5.9).

Online: Werte für gegebenes $\boldsymbol{\sigma}$ die Darstellung (5.11) aus und berechne die Lösung $u_N(\boldsymbol{\sigma})$ durch (5.10). Dies erfolgt unabhängig von der Größe $\mathcal{N}$.

Es bleibt zu klären, wie die RB-Räume $\mathcal{X}_N$ und $\mathcal{Y}_N$ konstruiert werden. Dieses Problem werden wir in mehrere Schritte aufteilen. Zunächst geben wir einen Algorithmus zur Konstruktion des Ansatzraums $\mathcal{X}_N$ an, fahren danach mit dem Testraum $\mathcal{Y}_N$ fort und klären anschließend die dabei auftretenden problematischen Punkte.

Konstruktion des RB-Ansatzraumes mittels Greedy-Algorithmus. Wie bereits angeführt wurde, wird der RB-Ansatzraum unter dem Aspekt, den Fehler zwischen Truth- und RB-Lösung für alle Parameter des Parameterraums $\mathcal{P}$ zu minimieren, konstruiert. Unter einem rein theoretischen Blick kann dies durch die sogenannten *Kolmogorow N-Weiten* optimal gelöst werden, wie beispielsweise [PR06, Section 3.4] entnommen werden kann. Da diese praktisch aber nicht umgesetzt werden können, wird hierfür oftmals auf einen *Greedy*-Ansatz zurückgegriffen. Dieser ist als Algorithmus 5.1 in Kurzform gegeben und wird im Folgenden erklärt.

Zunächst wird eine Diskretisierung $\mathcal{P}_{\text{train}} \subset \mathcal{P}$ des Parameterraums benötigt. Diese sollte „fein genug“ sein, um den Parameterraum mit dem beschriebenen Verfahren ausreichend auszuschöpfen. Es gibt verschiedene Strategien, wie diese gewählt werden kann, beispielsweise zufällig oder in Form eines äquidistanten Gitters.

Sei $\mathcal{X}_N$ ein N-dimensionaler RB-Ansatzraum mit Basis Ξ_N. Um daraus einen $N+1$-dimensionalen Raum $\mathcal{X}_{N+1}$ zu gewinnen, verfahren wir wie folgt:

Algorithmus 5.1. Greedy-Konstruktion der RB-Räume.

Eingabe : endliche Menge der Trainingsparameter $\mathcal{P}_{\text{train}} \subset \mathcal{P}$,
Fehlertoleranz $\varepsilon_{\text{tol}} > 0$
Ausgabe : RB-Ansatzraum $\mathcal{X}_N$ mit Basis Ξ_N,
Sample-Menge $\mathcal{S}_N$

Setze $N = 0$, $\mathcal{S}_0 = \emptyset$, $\Xi_0 = \emptyset$, $\mathcal{X}_0 = \{0\}$;
while $\max_{\boldsymbol{\sigma} \in \mathcal{P}_{\text{train}}} \Delta_N(\boldsymbol{\sigma}) > \varepsilon_{\text{tol}}$ **do**
  $\boldsymbol{\sigma}_{N+1} \leftarrow \arg\max_{\boldsymbol{\sigma} \in \mathcal{P}_{\text{train}}} \Delta_N(\boldsymbol{\sigma})$;
  $\mathcal{S}_{N+1} \leftarrow \mathcal{S}_N \cup \{\boldsymbol{\sigma}_{N+1}\}$;
  $\Xi_{N+1} \leftarrow$ `Orthonormalisiere`$(\Xi_N \cup \{u_{\mathcal{N}}(\boldsymbol{\sigma}_{N+1})\})$;
  $\mathcal{X}_{N+1} \leftarrow \operatorname{span} \Xi_{N+1}$;
  Berechne die neuen benötigten diskreten Objekte;
  $N \leftarrow N + 1$;

i. Bestimme den Parameter $\boldsymbol{\sigma}_{N+1} \in \mathcal{P}_{\text{train}}$, der den a posteriori-Fehlerschätzer Δ_N maximiert, also $\boldsymbol{\sigma}_{N+1} := \arg\max_{\boldsymbol{\sigma} \in \mathcal{P}_{\text{train}}} \Delta_N(\boldsymbol{\sigma})$,

ii. berechne die Truth-Lösung $u_{\mathcal{N}}(\boldsymbol{\sigma}_{N+1})$,

iii. definiere mit dieser nach einer Orthonormalisierung $\Xi_{N+1} := \Xi_N \cup \{u_{\mathcal{N}}(\boldsymbol{\sigma}_{N+1})\}$, $\mathcal{X}_{N+1} := \operatorname{span} \Xi_{N+1}$ und berechne die neuen benötigten Matrizen und Vektoren für das RB-System und den a posteriori-Fehlerschätzer.

Der beschriebene Ablauf wird nun iterativ ausgeführt, bis der maximale Fehler in Punkt *i* unter einer gewünschten Toleranz liegt. Um das Verfahren zu starten, kann das erste Parameter-Sample beliebig, beispielsweise zufällig, gewählt werden.

Die Orthonormalisierung wird durchgeführt, da a priori nicht sichergestellt werden kann, dass die Snapshots $u_{\mathcal{N}}(\boldsymbol{\sigma}_n)$ nicht nahezu linear abhängig sind. Dies würde in einer schlechten Kondition für die Systemmatrix (5.11) resultieren. Da diese Matrix nach Konstruktion ohnehin dicht besetzt ist, wird im Allgemeinen eine Gram-Schmidt-Orthonormalisierung durchgeführt, um so dem genannten Problem zu entgehen.

Diese Greedy-Heuristik hat zwar offensichtlich nur eine lokale Sicht auf die Wahl der optimalen Parameter, liefert dadurch aber ein Verfahren, welches nach [PR06, Subsection 3.4.3] zufriedenstellende Ergebnisse unter einem vertretbaren Aufwand liefert. Letzterer lässt sich wie folgt angeben: Sei $N_{\max}$ die Anzahl insgesamt durchgeführter Iterationen. Dann entspricht der Aufwand (i) $N_{\max} \cdot \#\mathcal{P}_{\text{train}}$ Auswertungen des a posteriori-Fehlerschätzers, (ii) $N_{\max}$ Aufrufen des Truth-Lösers und (iii) einem vollständigen Durchlauf der Gram-Schmidt-Orthonormalisierung.

Wie direkt ersichtlich ist, wird vor allem eine effiziente Möglichkeit zur Berechnung des a posteriori-Fehlerschätzers benötigt. Bevor wir uns dieser Aufgabe widmen, wollen wir auf die Konstruktion des zugehörigen RB-Testraumes eingehen.

Konstruktion des RB-Testraumes. Motiviert durch [MU14] erweist sich der bereits bekannte Supremizing-Operator aus Definition 4.6 als nützlich. Seien $\boldsymbol{\sigma} \in \mathcal{P}$ ein Parameter und $T_{\boldsymbol{\sigma}} \colon \mathcal{X}_{\mathcal{N}} \to \mathcal{Y}_{\mathcal{N}}$ der Supremizing-Operator der Bilinearform $b(\cdot, \cdot; \boldsymbol{\sigma})|_{\mathcal{X}_{\mathcal{N}} \times \mathcal{Y}_{\mathcal{N}}}$, dann

definieren wir den parameterabhängigen Testraum

$$\mathcal{Y}_N^{\boldsymbol{\sigma}} := \{T_{\boldsymbol{\sigma}} u_{\mathcal{N}}(\boldsymbol{\sigma}_i) \,|\, \boldsymbol{\sigma}_i \in \mathcal{S}_N,\ i = 1, \dots, N\}.$$

Satz 5.8. *Das RB-Variationsproblem* (5.4) *mit* $\mathcal{X}_N$ *und* $\mathcal{Y}_N^{\boldsymbol{\sigma}}$ *als Ansatz- beziehungsweise Testraum ist für alle* $\boldsymbol{\sigma} \in \mathcal{P}$ *korrekt gestellt.*

Beweis. Wir weisen die Bedingungen des Banach-Nečas-Babuška-Theorems nach, wobei wir uns auf die inf-sup-Bedingung beschränken, da die Stetigkeit vom Truth-Variationsproblem vererbt wird. Nach Annahme erfüllt die Truth-inf-sup-Konstante $\beta_{\mathcal{N}}(\boldsymbol{\sigma}) > 0$. Ferner gilt nach Definition für $u \in \mathcal{X}_N$ auch $T_{\boldsymbol{\sigma}} u \in \mathcal{Y}_N^{\boldsymbol{\sigma}}$, was zu der Abschätzung

$$\begin{aligned}
\beta_N(\boldsymbol{\sigma}) &= \inf_{u \in \mathcal{X}_N} \sup_{v \in \mathcal{Y}_N^{\boldsymbol{\sigma}}} \frac{b(u, v; \boldsymbol{\sigma})}{\|u\|_{\mathcal{X}} \|v\|_{\mathcal{Y}}} = \inf_{u \in \mathcal{X}_N} \frac{b(u, T_{\boldsymbol{\sigma}} u; \boldsymbol{\sigma})}{\|u\|_{\mathcal{X}} \|T_{\boldsymbol{\sigma}} u\|_{\mathcal{Y}}} \\
&= \inf_{u \in \mathcal{X}_N} \frac{\|T_{\boldsymbol{\sigma}} u\|_{\mathcal{Y}}}{\|u\|_{\mathcal{X}}} \geq \inf_{u \in \mathcal{X}_{\mathcal{N}}} \frac{\|T_{\boldsymbol{\sigma}} u\|_{\mathcal{Y}}}{\|u\|_{\mathcal{X}}} \\
&= \beta_{\mathcal{N}}(\boldsymbol{\sigma}) > 0
\end{aligned}$$

führt. □

Da der Testraum $\mathcal{Y}_N^{\boldsymbol{\sigma}}$ für jeden Parameter neu aufgebaut werden muss, ist eine sinnvolle Zerlegung in Offline- und Online-Phase wünschenswert. Dazu nutzen wir die parametrisch affine Darstellung (5.7) und definieren für jedes $b_q|_{\mathcal{X}_{\mathcal{N}} \times \mathcal{Y}_{\mathcal{N}}}$, $q = 1, \dots, Q_b$, den zugehörigen Supremizing-Operator T_q. Dann gilt

$$T_{\boldsymbol{\sigma}} = \sum_{q=1}^{Q_b} \vartheta_q^b(\boldsymbol{\sigma}) T_q.$$

Erneut kann die diskrete Darstellung $\mathbf{T}_q \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$ der Operatoren T_q nach dem Rieszschen Darstellungssatz durch $\mathbf{T}_q = \mathbf{Y}^{-1}\mathbf{B}_q$ bestimmt werden. Diese müssen nur einmal vorberechnet werden, während die restlichen Berechnungen folgendermaßen zerlegt werden können:

Offline: Während des Greedy-Verfahrens werden die Vektoren $\mathbf{v}_q^n := \mathbf{T}_q \mathbf{u}_{\mathcal{N}}(\boldsymbol{\sigma}_n)$ für jedes $q = 1, \dots, Q_b$ und $n = 1, \dots, N$ berechnet. Dies entspricht $N \cdot Q_b$ Matrix-Matrix-Multiplikationen der Größe $\mathcal{N} \times \mathcal{N}$.

Online: Bestimme $\mathcal{Y}_N^{\boldsymbol{\sigma}}$ durch $\mathbf{T}_{\boldsymbol{\sigma}} \mathbf{u}_{\mathcal{N}}(\boldsymbol{\sigma}_n) = \sum_{q=1}^{Q_b} \vartheta_q^b(\boldsymbol{\sigma}) \mathbf{v}_q^n$ für $n = 1, \dots, N$. Der Aufwand ist vernachlässigbar, da die Auswertung der $\vartheta_q^b(\boldsymbol{\sigma})$ nach Annahme effizient durchführbar ist.

Berechnung der Norm des Residuums. Diese stellt eine Hälfte des a posteriori-Fehlerschätzers dar und kann unter Verwendung des Rieszschen Darstellungssatzes effizient umgesetzt werden. Wir erhalten dadurch für jedes $\boldsymbol{\sigma} \in \mathcal{P}$ ein $\hat{e}_N(\boldsymbol{\sigma}) \in \mathcal{Y}_{\mathcal{N}}$ mit

$$\|\hat{e}_N(\boldsymbol{\sigma})\|_{\mathcal{Y}} = \|r_N(\cdot; \boldsymbol{\sigma})\|_{\mathcal{Y}_{\mathcal{N}}'}, \qquad r_N(v; \boldsymbol{\sigma}) = \langle \hat{e}_N(\boldsymbol{\sigma}), v \rangle_{\mathcal{Y}_{\mathcal{N}}} \quad \text{für alle } v \in \mathcal{Y}.$$

Aufgrund der Darstellung (5.6) des Residuums und der parametrisch affinen Darstellung (5.7) der Bilinearform gilt für alle $v \in \mathcal{Y}_\mathcal{N}$ die Gleichung

$$\langle \hat{e}_N(\boldsymbol{\sigma}), v\rangle_\mathcal{Y} = r_N(v;\boldsymbol{\sigma}) = f(v) - \sum_{q=1}^{Q_b}\sum_{n=1}^{N} \vartheta_q^b(\boldsymbol{\sigma}) u_{n,N}(\boldsymbol{\sigma}) b_q(\xi_n, v).$$

Weiter sind durch $b_q(\xi_n, \cdot)\colon \mathcal{Y}_\mathcal{N} \to \mathbb{R}$ für alle $n = 1,\dots,N$ und $q = 1,\dots,Q_b$ stetige lineare Funktionale gegeben, sodass eine erneute Anwendung des Rieszschen Darstellungssatzes die Existenz von $\hat{f}, \hat{b}_q^n \in \mathcal{Y}_\mathcal{N}$ mit

$$f(v) = \langle \hat{f}, v\rangle_\mathcal{Y}, \qquad b_q(\xi_n, v) = \langle \hat{b}_q^n, v\rangle_\mathcal{Y} \quad \text{für alle } v \in \mathcal{Y}_\mathcal{N}$$

liefert. Insgesamt ergibt dies die Darstellung

$$\hat{e}_N(\boldsymbol{\sigma}) = \hat{f} - \sum_{q=1}^{Q_b}\sum_{n=1}^{N} \vartheta_q^b(\boldsymbol{\sigma}) u_{n,N}(\boldsymbol{\sigma}) \hat{b}_q^n.$$

Darüber kann nun die Norm des Residuums als

$$\begin{aligned}
\|r_N(\cdot;\boldsymbol{\sigma})\|^2_{\mathcal{Y}'_\mathcal{N}} &= \|\hat{e}_N(\boldsymbol{\sigma})\|^2_\mathcal{Y} = \langle \hat{e}_N(\boldsymbol{\sigma}), \hat{e}_N(\boldsymbol{\sigma})\rangle_\mathcal{Y} \\
&= \langle \hat{f}, \hat{f}\rangle_\mathcal{Y} - 2\sum_{q=1}^{Q_b}\sum_{n=1}^{N} \vartheta_q^b(\boldsymbol{\sigma}) u_{n,N}(\boldsymbol{\sigma}) \langle \hat{b}_q^n, \hat{f}\rangle_\mathcal{Y} \\
&\quad + \sum_{q,p=1}^{Q_b}\sum_{n,m=1}^{N} \vartheta_q^b(\boldsymbol{\sigma})\vartheta_p^b(\boldsymbol{\sigma}) u_{n,N}(\boldsymbol{\sigma}) u_{m,N}(\boldsymbol{\sigma}) \langle \hat{b}_q^n, \hat{b}_p^m\rangle_\mathcal{Y}
\end{aligned}$$

bestimmt werden. Weiter bestimmen wir die Koeffizientenvektoren $\hat{\mathbf{f}}, \hat{\mathbf{b}}_q^n \in \mathbb{R}^\mathcal{N}$ der Rieszschen Darstellungen $\hat{f}$ und $\hat{b}_q^n$ über

$$\mathbf{Y}\hat{\mathbf{f}} = \mathbf{f}, \qquad \mathbf{Y}\hat{\mathbf{b}}_q^n = [b_q(\xi_n, \psi_j)]_{j=1,\dots,\mathcal{N}} = \mathbf{B}_q \boldsymbol{\xi}_n.$$

Verwenden wir diese, so wird obige Gleichung zur Berechnung der Norm des Residuums nach Auswerten der $\mathcal{Y}$-Skalarprodukte zu

$$\begin{aligned}
\|r_N(\cdot;\boldsymbol{\sigma})\|^2_{\mathcal{Y}'_\mathcal{N}} = \mathbf{f}^\mathsf{T}\mathbf{Y}^{-1}\mathbf{f} &- 2\sum_{q=1}^{Q_b}\sum_{n=1}^{N} \vartheta_q^b(\boldsymbol{\sigma}) u_{n,N}(\boldsymbol{\sigma}) \mathbf{f}^\mathsf{T}\mathbf{Y}^{-1}\mathbf{B}_q\boldsymbol{\xi}_n \\
&+ \sum_{q,p=1}^{Q_b}\sum_{n,m=1}^{N} \vartheta_q^b(\boldsymbol{\sigma})\vartheta_p^b(\boldsymbol{\sigma}) u_{n,N}(\boldsymbol{\sigma}) u_{m,N}(\boldsymbol{\sigma}) \boldsymbol{\xi}_n^\mathsf{T}\mathbf{B}_q^\mathsf{T}\mathbf{Y}^{-1}\mathbf{B}_p\boldsymbol{\xi}_m.
\end{aligned} \tag{5.12}$$

Mit Hilfe dieses Ausdrucks können wir nun wie folgt eine Zerlegung in Offline- und Online-Phase vornehmen:

Offline: Berechne die Größen $\mathbf{f}^\mathsf{T}\mathbf{Y}^{-1}\mathbf{f}$, $\mathbf{f}^\mathsf{T}\mathbf{Y}^{-1}\mathbf{B}_q\boldsymbol{\xi}_n$ und $\boldsymbol{\xi}_n^\mathsf{T}\mathbf{B}_q^\mathsf{T}\mathbf{Y}^{-1}\mathbf{B}_p\boldsymbol{\xi}_m$ für alle $n, m = 1,\dots,N$ und $q, p = 1,\dots,Q_b$. Ist $N_{\max}$ die Anzahl insgesamt ausgeführter Iterationen, dann müssen $N_{\max}^2 Q_b^2 + N_{\max} Q_b + 1$ der obigen Terme berechnet werden.

Online: Für einen gegebenen Parameter $\boldsymbol{\sigma} \in \mathcal{P}$ und dessen RB-Lösung $u_N(\boldsymbol{\sigma})$ muss nur die Formel (5.12) ausgewertet werden. Dies kann mit einer von $\mathcal{N}$ unabhängigen Laufzeit erreicht werden.

Bei der praktischen Umsetzung wird (5.12) nicht direkt in der gegebenen Form implementiert, sondern in einen parameterabhängigen und einen parameterunabhängigen Teil zerlegt, zu einem Vektor beziehungsweise einer Matrix zusammengefasst und mittels einer Matrix-Vektor-Multiplikation ausgewertet.

Berechnung einer unteren Schranke für die inf-sup-Konstante. Da die exakte Bestimmung der inf-sup-Konstante $\beta_{\mathcal{N}}(\boldsymbol{\sigma})$ für jeden Parameter $\boldsymbol{\sigma} \in \mathcal{P}_{\text{train}}$ mittels des in Abschnitt 4.3 beschriebenen verallgemeinerten Eigenwertproblems laufzeittechnisch nicht vertretbar ist, wird ein effizienteres Verfahren benötigt, welches eine möglichst genaue untere Schranke $\beta_{\text{LB}}(\boldsymbol{\sigma})$ liefert. Im Rahmen dieser Arbeit stützen wir uns hierfür auf die zu diesem Zweck entwickelte Successive Constraint Method (SCM) und orientieren uns an der Originalarbeit von Huynh et al. [Huy+07], wobei hier nur die Idee des Verfahrens wiederholt werden soll. Weiter wurden bei der Implementierung die Verbesserungen von Chen et al. [Che+09] berücksichtigt, die hier nicht weiter ausgeführt werden.

Da die SCM auf die Berechnung der Koerzivitätskonstanten einer symmetrischen parametrischen Bilinearform ausgelegt ist, müssen wir zunächst die nötigen Rahmenbedingungen schaffen. Wir werden später darauf eingehen, wie wir mit der SCM stattdessen die gewünschte inf-sup-Schranke bestimmen können.

Es sei also durch

$$a(u, v; \boldsymbol{\sigma}) = \sum_{q=1}^{Q_a} \vartheta_q^a a_q(u, v), \quad u, v \in \mathcal{X},$$

eine abstrakte symmetrische, parametrisch affine Bilinearform $a\colon \mathcal{X} \times \mathcal{X} \times \mathcal{P} \to \mathbb{R}$ gegeben, wobei $\vartheta_q^a\colon \mathcal{P} \to \mathbb{R}$ stetige Funktionen und $a_q\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ symmetrische, stetige und parameterunabhängige Bilinearformen seien. Wie zuvor sei $\mathcal{X}_{\mathcal{N}} \subset \mathcal{X}$ eine Diskretisierung eines Galerkin-Verfahrens. Wir wollen nun eine untere Schranke $\alpha_{\text{LB}}(\boldsymbol{\sigma})$ für die *Koerzivitätskonstante* $\alpha_{\mathcal{N}}(\boldsymbol{\sigma})$ bestimmen, welche als

$$\alpha_{\mathcal{N}}(\boldsymbol{\sigma}) := \inf_{u \in \mathcal{X}_{\mathcal{N}}} \frac{a(u, u; \boldsymbol{\sigma})}{\|u\|_{\mathcal{X}}^2}$$

definiert ist.

Das Kernstück der SCM ist es, die Bestimmung der Koerzivitätskonstanten als lineares Optimierungsproblem aufzufassen. Dazu sei zunächst die zulässige Menge $\mathcal{Z}$ definiert als

$$\mathcal{Z} := \left\{ \mathbf{z} = (z_1, \ldots, z_{Q_a}) \in \mathbb{R}^{Q_a} \;\middle|\; \exists\, u_z \in \mathcal{X}_{\mathcal{N}}\colon z_q = \frac{a_q(u_z, u_z)}{\|u_z\|_{\mathcal{X}}^2}, q = 1, \ldots, Q_a \right\}$$

und die lineare Zielfunktion als

$$\mathcal{J}\colon \mathbb{R}^{Q_a} \times \mathcal{P} \to \mathbb{R}, \quad \mathcal{J}(\mathbf{z}; \boldsymbol{\sigma}) := \sum_{q=1}^{Q_a} \vartheta_q^a(\boldsymbol{\sigma}) z_q.$$

Mit dieser Konstruktion gilt nun zwar $\alpha_{\mathcal{N}}(\boldsymbol{\sigma}) = \min_{\mathbf{z} \in \mathcal{Z}} \mathcal{J}(\mathbf{z}; \boldsymbol{\sigma})$, jedoch handelt es sich hierbei aufgrund der zulässigen Menge $\mathcal{Z}$ noch nicht um ein lineares Programm. Um dies

Algorithmus 5.2. Successive Constraint Method, Greedy-Konstruktion der $\mathcal{C}_K$.

Eingabe : endliche Menge der Trainingsparameter $\mathcal{P}_{\text{train}} \subset \mathcal{P}$,
Fehlertoleranz $\varepsilon_{\text{tol}} > 0$

Ausgabe : endliche Parametermenge $\mathcal{C}_K \subset \mathcal{P}_{\text{train}}$

Setze $K = 1$, $\mathcal{C}_1 = \{\boldsymbol{\sigma}_1\}$ mit zufälligem $\boldsymbol{\sigma}_1 \in \mathcal{P}_{\text{train}}$;
while $\max_{\boldsymbol{\sigma} \in \mathcal{P}_{\text{train}}} [(\alpha_{\text{UB}}(\boldsymbol{\sigma}; \mathcal{C}_K) - \alpha_{\text{LB}}(\boldsymbol{\sigma}; \mathcal{C}_K))/\alpha_{\text{UB}}(\boldsymbol{\sigma}; \mathcal{C}_K)] > \varepsilon_{\text{tol}}$ **do**
  $\boldsymbol{\sigma}_{K+1} \leftarrow \arg\max_{\boldsymbol{\sigma} \in \mathcal{P}_{\text{train}}} [(\alpha_{\text{UB}}(\boldsymbol{\sigma}; \mathcal{C}_K) - \alpha_{\text{LB}}(\boldsymbol{\sigma}; \mathcal{C}_K))/\alpha_{\text{UB}}(\boldsymbol{\sigma}; \mathcal{C}_K)]$;
  $\mathcal{C}_{K+1} \leftarrow \mathcal{C}_K \cup \{\boldsymbol{\sigma}_{K+1}\}$;
  $K \leftarrow K + 1$;

zu ändern, werden wir eine Unter- und eine Obermenge von $\mathcal{Z}$ konstruieren, die dann zu einem linearen Optimierungsproblem für eine untere Schranke von $\alpha_{\mathcal{N}}(\boldsymbol{\sigma})$ verarbeitet werden können.

Als ersten Schritt schränken wir die Komponenten z_q durch die Wahl von

$$\alpha_q^- := \inf_{u \in \mathcal{X}_{\mathcal{N}}} \frac{a_q(u,u)}{\|u\|_{\mathcal{X}}^2}, \quad \alpha_q^+ := \sup_{u \in \mathcal{X}_{\mathcal{N}}} \frac{a_q(u,u)}{\|u\|_{\mathcal{X}}^2} \qquad \text{für alle } q = 1, \dots, Q_a$$

und des Quaders $B_{Q_a} := \prod_{q=1}^{Q_a} [\alpha_q^-, \alpha_q^+] \subset \mathbb{R}^{Q_a}$ ein. Wir benötigen erneut eine endliche Teilmenge $\mathcal{P}_{\text{train}} \subset \mathcal{P}$ als Trainingsparameter. Ferner definieren wir hier die Notation $\mathcal{C}_K := \{\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_K\} \subset \mathcal{P}_{\text{train}}$ für eine weitere, noch zu konstruierende Parameterteilmenge. Weiter sei durch $P_M(\boldsymbol{\sigma}; E)$ eine Abbildung gegeben, welche für ein $M \in \mathbb{N}_0$ und eine Teilmenge $E \subset \mathcal{P}$ die bezüglich der euklidischen Norm M nächsten Punkte zu $\boldsymbol{\sigma}$ aus E liefert. Schließlich seien durch $M_\alpha, M_+ \in \mathbb{N}$ die Anzahl der Stabilitätsbedingungen respektive Positivitätsbedingungen festgelegt.

Mit dieser Vorarbeit können wir nun eine Untermenge $\mathcal{Z}_{\text{LB}}(\boldsymbol{\sigma}; \mathcal{C}_K) \subset \mathcal{Z}$ und eine Obermenge $\mathcal{Z}_{\text{UB}}(\mathcal{C}_K) \supset \mathcal{Z}$ definieren durch

$$\begin{aligned}
\mathcal{Z}_{\text{LB}}(\boldsymbol{\sigma}; \mathcal{C}_K) := &\left\{ \mathbf{z} \in B_{Q_a} \,\middle|\, \textstyle\sum_{q=1}^{Q_a} \vartheta_q^a(\tilde{\boldsymbol{\sigma}}) z_q \geq \alpha_{\mathcal{N}}(\boldsymbol{\sigma}) \text{ für alle } \tilde{\boldsymbol{\sigma}} \in P_{M_\alpha}(\boldsymbol{\sigma}; \mathcal{C}_K) \right\} \\
&\cap \left\{ \mathbf{z} \in B_{Q_a} \,\middle|\, \textstyle\sum_{q=1}^{Q_a} \vartheta_q^a(\tilde{\boldsymbol{\sigma}}) z_q \geq 0 \text{ für alle } \tilde{\boldsymbol{\sigma}} \in P_{M_+}(\boldsymbol{\sigma}; \mathcal{P}_{\text{train}}) \right\}, \\
\mathcal{Z}_{\text{UB}}(\mathcal{C}_K) := &\left\{ \mathbf{z}^*(\boldsymbol{\sigma}_k) := \arg\min_{\mathbf{z} \in \mathcal{Z}} \mathcal{J}(\mathbf{z}; \boldsymbol{\sigma}_k) \,\middle|\, k = 1, \dots, K \right\}.
\end{aligned}$$

Mit Hilfe dieser Mengen lassen sich nun die linearen Probleme

$$\alpha_{\text{LB}}(\boldsymbol{\sigma}; \mathcal{C}_K) := \min_{\mathbf{z} \in \mathcal{Z}_{\text{LB}}(\boldsymbol{\sigma}; \mathcal{C}_K)} \mathcal{J}(\mathbf{z}; \boldsymbol{\sigma}), \quad \alpha_{\text{UB}}(\boldsymbol{\sigma}; \mathcal{C}_K) := \min_{\mathbf{z} \in \mathcal{Z}_{\text{UB}}(\mathcal{C}_K)} \mathcal{J}(\mathbf{z}; \boldsymbol{\sigma})$$

auswerten, für die nach [Huy+07, Proposition 1] nachgewiesen werden kann, dass diese für beliebige $\mathcal{C}_K, M_\alpha$ und M_+ die Abschätzung

$$\alpha_{\text{LB}}(\boldsymbol{\sigma}; \mathcal{C}_K) \leq \alpha_{\mathcal{N}}(\boldsymbol{\sigma}) \leq \alpha_{\text{UB}}(\boldsymbol{\sigma}; \mathcal{C}_K) \quad \text{für alle } \boldsymbol{\sigma} \in \mathcal{P}$$

erfüllen. Algorithmus 5.2 zeigt auf, wie die noch fehlenden Teilmengen $\mathcal{C}_K \subset \mathcal{P}$ mittels eines Greedy-Ansatzes möglichst optimal konstruiert werden können.

Um mit diesem Vorgehen eine Schranke für die inf-sup-Konstante $\beta_{\mathcal{N}}(\boldsymbol{\sigma})$ zu bestimmen, müssen wir lediglich erneut auf den zugehörigen Supremizing-Operator aus Definition 4.6

zurückgreifen. Gleichung (4.5) erlaubt uns dann, $\beta^2_{\mathcal{N}}(\boldsymbol{\sigma})$ als Koerzivitätskonstante einer entsprechenden Bilinearform aufzufassen.

Abschließend wollen wir auf die Laufzeit dieses Verfahrens eingehen. Dazu erfolgt erneut eine Zerlegung in Offline- und Online-Phase.

Offline: Hier wird die in Algorithmus 5.2 beschriebene Konstruktion durchgeführt. Ist $K_{\max}$ die finale Anzahl durchgeführter Iterationen, dann wird der Aufwand maßgeblich durch (i) $2Q_a$ Eigenwertprobleme für B_{Q_a} sowie $K_{\max}$ Eigenwertprobleme für $\alpha_{\mathcal{N}}(\boldsymbol{\sigma}_k)$ und $\mathbf{z}^*(\boldsymbol{\sigma}_k)$ sowie (ii) $\#\mathcal{P}_{\text{train}} \cdot K_{\max}$ lineare Programme der Größe $\mathcal{O}(Q_a + M_\alpha + M_+)$ festgelegt.

Online: Es muss lediglich ein lineares Programm der Größe $\mathcal{O}(Q_a + M_\alpha + M_+)$ gelöst werden. Insbesondere ist die Berechnung von $\boldsymbol{\sigma} \mapsto \alpha_{\text{LB}}(\boldsymbol{\sigma})$ unabhängig von $\mathcal{N}$.

Zerlegung der Reduzierte-Basis-Methode in Offline- und Online-Phase. Um die RB-Methode als ganzes in eine Offline- und eine Online-Phase zu zerlegen, müssen wir lediglich die bereits behandelten Bausteine zusammensetzen. Der Aufwand der jeweiligen Berechnungen wurde an den entsprechenden Stellen angegeben.

Offline: Zunächst muss die Offline-Phase der Successive Constraint Method durchgeführt werden. Nach Beendigung dieser kann der Greedy-Algorithmus 5.2 zur Konstruktion der RB-Ansatz- und Testräume ausgeführt werden.

Online: Hier wird zu einem gegebenen Parameter $\boldsymbol{\sigma} \in \mathcal{P}$ die Reduzierte-Basis-Lösung $u_N(\boldsymbol{\sigma})$ bestimmt und anschließend der a posteriori-Fehlerschätzer $\Delta_N(\boldsymbol{\sigma})$ ausgewertet. Der Aufwand dieser Berechnungen ist unabhängig von der Dimension $\mathcal{N}$ der zugrundeliegenden Truth-Diskretisierung.

5.3. Beispiele

Erneut wollen wir die numerische Umsetzung durch Betrachtung verschiedener Beispiele abschließen. Wir konzentrieren uns dabei hauptsächlich auf ein einfaches Beispiel mit einem eindimensionalen Parameterraum, da dies die beste Möglichkeit bietet, das Verhalten der Reduzierte-Basis-Methode zu veranschaulichen. Weiter gehen wir auch auf ein komplexeres Beispiel ein, halten die Untersuchung dessen vor allem aus Laufzeitgründen jedoch sehr simpel.

Der Einfachheit halber nehmen wir bei beiden Beispielen homogene Randbedingungen an. Weiter wählen wir die Truth-Diskretisierung wie im ersten Beispiel in Abschnitt 4.4.

Um die Parametrisierung der Felder durchzuführen, verwenden wir Sinusfunktionen der Form $\sin(k\pi x)$, $k \in \mathbb{N}$, wobei bei den betrachteten Beispielen lediglich $\sin(\pi x)$ und $\sin(2\pi x)$ auftreten werden. Diese Wahl lässt sich durch die beispielhaften Ergebnisse aus der Einleitung rechtfertigen, da dort gut erkennbar ist, dass die Felder sowohl eine hohe Regularität als auch Symmetrien ausweisen.

Eindimensionaler Parameterraum. Um einen eindimensionalen Parameterraum zu erhalten, verzichten wir erneut auf den zeitlichen Wechsel der Felder. Es seien $I = [0, 1]$

K	max. rel. Spanne
1	1
2	1.553973×10^{-1}
3	9.525370×10^{-2}
4	3.911659×10^{-2}
5	3.834086×10^{-2}
6	2.385319×10^{-2}
7	2.378906×10^{-2}
8	9.824040×10^{-3}

Tabelle 5.1. Entwicklung der maximalen relativen Spanne (5.13) während der Greedy-Iteration der SCM beim ersten Beispiel.

und $\Omega = [0,1]$. Der parametrische Differentialoperator $A(t;\sigma) \equiv A(\sigma)\colon V \to V'$ sei ferner gegeben durch

$$A(\sigma)\eta = -\tfrac{1}{10}\Delta\eta + \eta + \sigma\sin(\pi\,\cdot)\eta$$

mit dem Parameter $\sigma \in \mathcal{P} := [-1,1]$. Wir verzichten auf einen Quellterm und verwenden als Anfangswert $u_0 = \frac{1}{2}\sin(\pi\,\cdot)$.

Als Dimensionen für die Truth-Diskretisierung nach Kapitel 4 verwenden wir für die zeitliche Komponente $\mathcal{K} = 250$ und für die räumliche $\mathcal{J} = 20$. Die Diskretisierung $\mathcal{P}_{\text{train}}$ des Parameterraums für die Greedy-Algorithmen wählen wir als äquidistantes Gitter auf $\mathcal{P}$ mit 1000 Punkten. Dies stellt eine relativ geringe Anzahl dar und ist hauptsächlich der Laufzeit der Analyse-Prozeduren geschuldet.

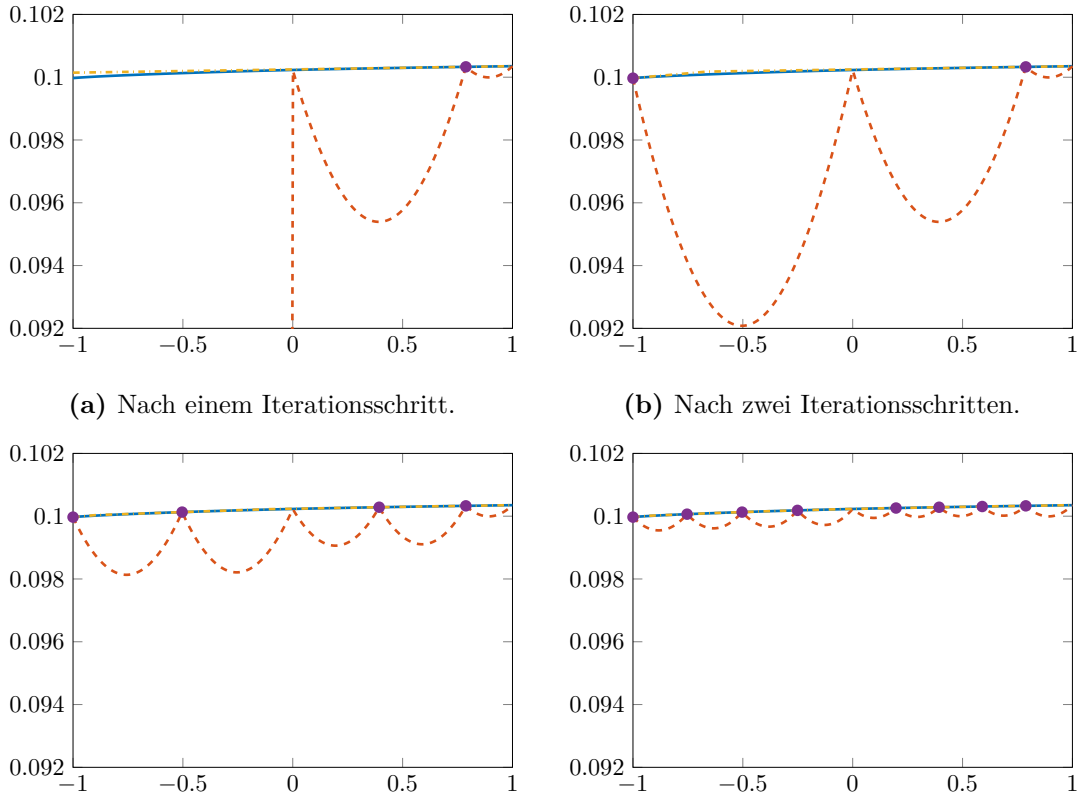
Wir wollen nun auf die verschiedenen numerischen Ergebnisse der Reduzierte-Basis-Methode und der Successive Constraint Method eingehen. Die benötigte Laufzeit für die Durchführung setzt sich zusammen aus 9 min 43 s für die Durchführung der Offline-Phase der SCM und 6 min 13 s für die RB-Offline-Phase. Vergleicht man dies mit den Laufzeiten für einen Aufruf des Truth- beziehungsweise RB-Lösers, die gemittelt die Werte 0.0817 s respektive 0.0637 s haben, dann wird deutlich, dass die Dauer der Offline-Phase unpraktikabel lange dauert, da circa 53000 Löser-Aufrufe nötig sind, um diese zu amortisieren. Zwar können diese Laufzeiten durch Optimierung der Implementierung zweifelsfrei deutlich verbessert werden, allerdings handelt es sich hierbei um das denkbar einfachste Beispiel. Dies lässt bereits vermuten, dass die Laufzeiten für realitätsnähere Beispiele unhandlich groß werden könnten.

Neben der Laufzeit sind natürlich auch die Ergebnisse der SCM und der RB-Methode interessant. Anhand von Abbildung 5.2 wird die Funktionsweise der Greedy-Iteration der SCM deutlich. Diese hat bereits nach acht Iterationsschritten die Fehlertoleranz von 10^{-2} unterschritten. Die Entwicklung der maximalen relativen Spanne (vergleiche Algorithmus 5.2)

$$\max_{\boldsymbol{\sigma}\in\mathcal{P}_{\text{train}}} \frac{\alpha_{\text{UB}}(\boldsymbol{\sigma};\mathcal{C}_K) - \alpha_{\text{LB}}(\boldsymbol{\sigma};\mathcal{C}_K)}{\alpha_{\text{UB}}(\boldsymbol{\sigma};\mathcal{C}_K)} \tag{5.13}$$

für $\alpha = \beta^2$ kann in Tabelle 5.1 abgelesen werden.

Da für diese Veranschaulichung die exakte diskrete inf-sup-Konstante $\beta_{\mathcal{N}}(\sigma)$ bereits berechnet wurde, können wir dies auch nutzen, um auf die hinreichende Bedingung für die parametrische Abhängigkeit aus Abschnitt 3.3 einzugehen. Verwenden wir die hier gegebenen Größen und nehmen wir weiter als grobe parameterunabhängige Näherung für $\beta_{\mathcal{N}}$ den Wert 0.1 an, dann reduziert sich die Bedingung zu $\|C\sin(\pi\,\cdot)\|_{L_\infty([0,1])} < 0.1$,



(a) Nach einem Iterationsschritt.

(b) Nach zwei Iterationsschritten.

(c) Nach vier Iterationsschritten.

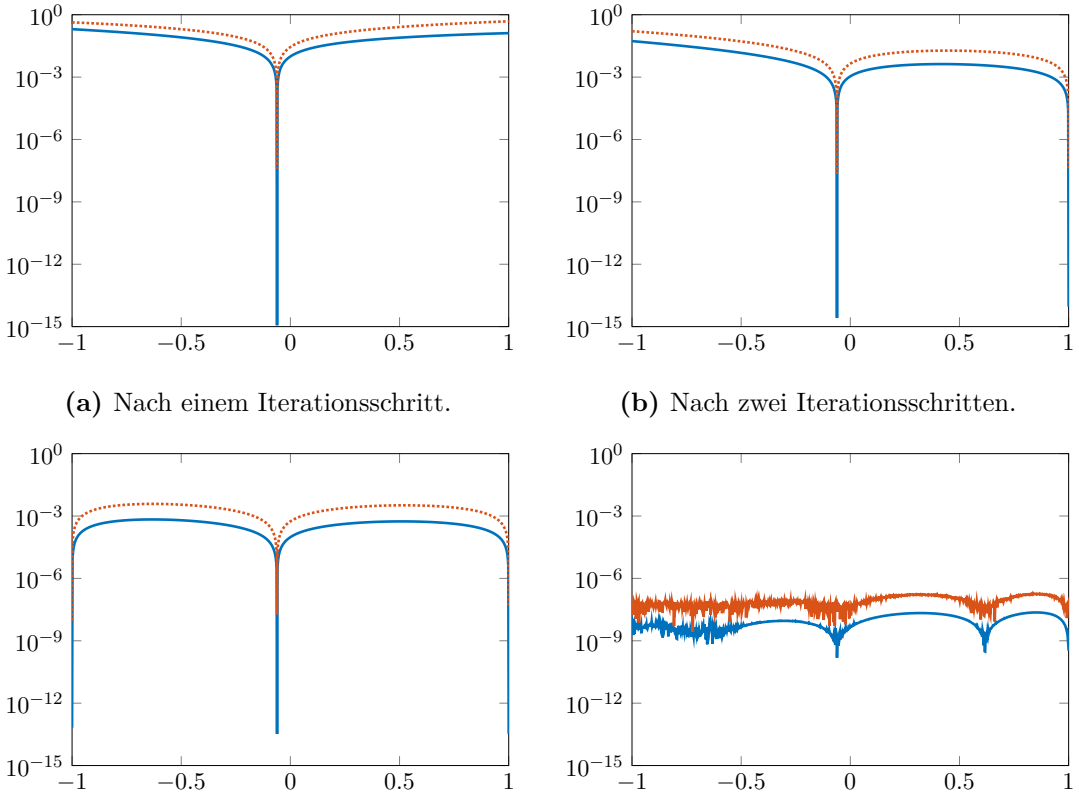
(d) Nach acht Iterationsschritten.

Abbildung 5.2. Vergleich zwischen exakt bestimmter inf-sup-Konstante $\beta_{\mathcal{N}}(\sigma)$ (blau), der unteren Schranke $\beta_{\mathrm{LB}}(\sigma)$ (orange gestrichelt) sowie der oberen Schranke $\beta_{\mathrm{UB}}(\sigma)$ (gelb gestrichelt) der SCM beim ersten Beispiel. Die horizontale Achse stellt den Parameterraum dar, wobei die Markierungen die vom Greedy-Verfahren ausgewählten Parameter anzeigen.

N	$\Delta_N^{\max}$	η_N^{ave}	η_N^{med}	$\eta_N^{\max}$
1	4.808368×10^{-1}	3.484555×10^{4}	2.848653	3.484268×10^{7}
2	1.606581×10^{-1}	1.271701×10^{4}	4.053004	8.790271×10^{6}
3	3.836088×10^{-3}	1.830208×10^{3}	5.742329	1.129403×10^{6}
4	2.843523×10^{-4}	3.783277×10^{2}	6.341930	1.679813×10^{5}
5	3.604637×10^{-6}	2.737704×10^{1}	7.006419	9.870312×10^{3}
6	1.892436×10^{-7}	1.136721×10^{1}	8.233159	2.421902×10^{2}

Tabelle 5.2. Entwicklung des Maximums $\Delta_N^{\max}$ des a posteriori-Fehlerschätzers sowie des arithmetischen Mittels η_N^{ave}, des Medians η_N^{med} und des Maximums $\eta_N^{\max}$ der Effektivität η_N über $\mathcal{P}_{\mathrm{train}}$ beim ersten Beispiel.

wobei $C > 0$ an dieser Stelle eine Skalierungsvariable sei. Konkret lässt sich dies so interpretieren, dass die hinreichende Bedingung nur für den Parameterteilraum $[-0.1, 0.1] \subset \mathcal{P}$ die analytische Regularität der Lösung garantiert. Dies ist offensichtlich ein äußerst kleiner Spielraum für den Parameter, welcher bei Hinzunahme weiterer Parameter noch weiter eingeschränkt wird.



(a) Nach einem Iterationsschritt.

(b) Nach zwei Iterationsschritten.

(c) Nach drei Iterationsschritten.

(d) Nach sechs Iterationsschritten.

Abbildung 5.3. Entwicklung des Fehlers zwischen Truth- und RB-Lösung für verschiedene Sample-Anzahlen N beim ersten Beispiel. Sowohl der exakt bestimmte Fehler (blau), als auch der a posteriori-Fehlerschätzer (orange) sind abgebildet.

In ähnlicher Weise zur SCM kann auch die Greedy-Iteration der RB-Methode dargestellt werden. Diese terminiert bereits nach sechs Iterationsschritten nach Unterschreiten der Fehlertoleranz 10^{-6}. Abbildung 5.3 zeigt, wie sich der a posteriori-Fehlerschätzer und der tatsächliche Fehler zwischen Truth- und RB-Lösung für wachsende Sample-Anzahl N verändern. Weiter ist in Tabelle 5.2 die Entwicklung des Maximums des a posteriori-Fehlerschätzers sowie des arithmetischen Mittels, des Medians und des Maximums der Effektivität η_N über $\mathcal{P}_{\text{train}}$ verzeichnet. Die für die Ausreißer der Effektivität verantwortlichen Spitzen treten bei den gewählten Parameter-Samples auf und sind vermutlich durch numerische Ungenauigkeiten bedingt, da sie für wachsendes N abgeschwächt werden.

Mehrdimensionaler Parameterraum. Das zweite und komplexere Beispiel beinhaltet einen zeitlichen Wechsel bei $T_f = 0.5$. Größen, die an dieser Stelle nicht explizit angegeben sind, entsprechen dem vorherigen Beispiel. Wir ändern die Dimension der zeitlichen Diskretisierung zu $\mathcal{K} = 200$ und belassen $\mathcal{J} = 20$. Der nun zeit- und parameterabhängige

N	$\Delta_N^{\max}$
1	5.335096×10^{-1}
5	8.204675×10^{-2}
10	8.472017×10^{-3}
15	1.249758×10^{-3}
20	1.637375×10^{-4}
25	4.508788×10^{-5}
30	1.337444×10^{-5}
35	4.012193×10^{-6}
40	1.417054×10^{-6}
41	9.452339×10^{-7}

Tabelle 5.3. Entwicklung des Maximums des a posteriori-Fehlerschätzers für steigendes N beim zweiten Beispiel. Siehe auch Abbildung 5.4.

Differentialoperator $A(t,\boldsymbol{\sigma})\colon V \to V'$ sei definiert als

$$\begin{aligned} A(t,\boldsymbol{\sigma})\eta = \tfrac{1}{10}\Delta\eta + 2\eta + \chi_{[0,0.5)}(t)[\sigma_1 \sin(\pi\,\cdot) + \sigma_2 \sin(2\pi\,\cdot)]\eta \\ + \chi_{[0.5,1]}(t)[\sigma_3 \sin(\pi\,\cdot) + \sigma_4 \sin(2\pi\,\cdot)]\eta \end{aligned}$$

mit dem Parameter $\boldsymbol{\sigma} \in \mathcal{P} := [-1,1]^4$. Die Diskretisierung $\mathcal{P}_{\text{train}}$ konstruieren wir mittels 5000 zufällig gleichverteilt gewählter Punkte aus $\mathcal{P}$.

Die benötigte Laufzeit für dieses Modell ist um ein Vielfaches höher als beim einfachen Beispiel. So benötigt die Offline-Phase der SCM 7 h 12 min, um nach der gewählten maximalen Iterationsanzahl von 500 zu terminieren, ohne die Fehlertoleranz von 10^{-2} zu unterschreiten. Weiter wird hier deutlich, dass die Trainingsparametermenge die tatsächliche Parametermenge nur unzureichend ausschöpft, da für viele Parameter die gelieferte untere Schranke α_{LB} einen negativen Wert annimmt und daher in einem imaginären Wert für $\beta_{\text{LB}} = \sqrt{\alpha_{\text{LB}}}$ resultiert. Dies verfälscht natürlich die Wahl der Parameter-Samples sowohl bei der SCM als auch bei der RB-Methode.

Auch die Offline-Phase der RB-Methode benötigt mit 5 h 58 min deutlich länger. Weiter können wir hier nicht angeben, ab wie vielen Auswertungen diese Laufzeit amortisiert werden würde. Bei dieser relativ niedrigen Dimension der Truth-Diskretisierung benötigt diese gemittelt 0.0849 s für eine Lösung, während die RB-Methode 0.1311 s für die Lösung inklusive der Auswertung des a posteriori-Fehlerschätzers benötigt. Letzterer sorgt für den Großteil dieser Laufzeit und würde bei einer höheren Dimension der Truth-Diskretisierung vermutlich weit weniger ins Gewicht fallen. Dies kann hier aber nicht weiter untersucht werden, da dies natürlich eine weitere Steigerung der Laufzeiten für die Offline-Phasen bedeutet.

Abbildung 5.4 und Tabelle 5.3 zeigen die Entwicklung des a posteriori-Fehlerschätzers. Es fällt auf, das bei lediglich vier Parametern bereits eine deutlich höhere Anzahl an Samples benötigt wird um eine Fehlertoleranz von 10^{-6} zu unterschreiten, der maximale a posteriori-Fehlerschätzer aber weiterhin monoton fällt.

Die Ergebnisse beider Beispiele sind unbefriedigend, zeigen sie doch, dass die in der Theorie auf diese Problemstellung durchaus gut anwendbare Reduzierte-Basis-Methode in der Praxis mit übermäßig hohen Laufzeiten einhergeht. Diese können zu Teilen auf die nicht optimierte Implementierung zurückgeführt werden. Vergleicht man aber beispielsweise die Laufzeit mit der in Kapitel 1 erwähnten Pseudospektralmethoden,

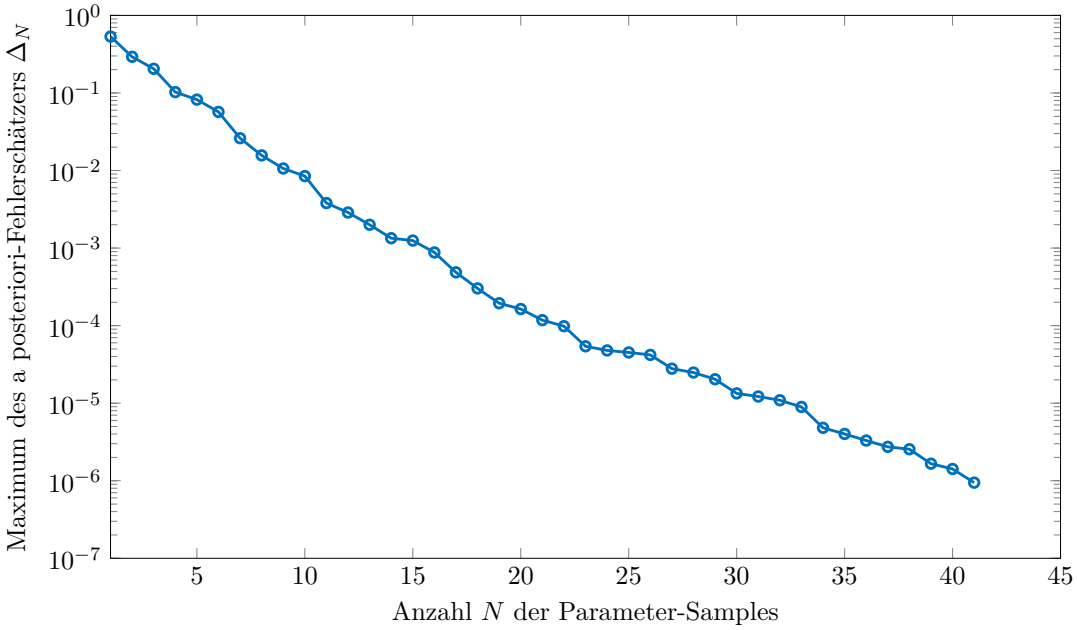


Abbildung 5.4. Entwicklung des Maximums des a posteriori-Fehlerschätzers für steigendes N beim zweiten Beispiel. Die geforderte Toleranz 10^{-6} wird beim 41. Iterationsschritt unterschritten. Siehe auch Tabelle 5.3.

welche bei ähnlichen Rahmenbedingungen und einer ähnlichen Genauigkeit im einstelligen Minutenbereich für das gesamte Iterationsverfahren liegt, so wird deutlich, dass der hier vorliegende Ansatz zur Modellreduktion weitere Untersuchungen benötigt, bevor daraus ein praktikables Verfahren gewonnen werden kann.

6 Kapitel

Fazit & Ausblick

In diesem abschließenden Kapitel wollen wir die Ergebnisse dieser Arbeit noch einmal diskutieren. Dabei betrachten wir diese hauptsächlich unter dem Aspekt der Anwendbarkeit auf die aus der Einleitung bekannte selbstkonsistente Feldtheorie.

Analog zum Aufbau der Arbeit beginnen wir mit den funktionalanalytischen Ergebnissen aus Kapitel 3, bevor wir weiter auf die zugrundeliegende Diskretisierung mittels Petrov-Galerkin-Verfahren und zuletzt auf die darauf aufbauende Reduzierte-Basis-Methode eingehen.

Funktionalanalytische Ergebnisse. Kapitel 3 lässt sich, neben der Herleitung der parametrischen Raum-Zeit-Variationsformulierung für die Propagator-Differentialgleichung, zu den folgenden beiden Hauptergebnissen zusammenfassen:

i. Der Nachweis, dass die hergeleitete Raum-Zeit-Variationsformulierung ein im Sinne von Hadamard korrekt gestelltes Problem darstellt.

ii. Die Untersuchung der Regularität der Lösungen der parametrischen Raum-Zeit-Variationsformulierung bezüglich der Parameter. Dabei wurde insbesondere eine hinreichende Bedingung dafür angegeben, wann diese Abhängigkeit analytisch ist.

Vor allem der zweite Punkt bietet noch Potenzial für weitere Untersuchungen. Die im Rahmen dieser Arbeit hergeleitete hinreichende Bedingung (siehe Satz 3.29) erweist sich als äußerst restriktiv, da sie maßgeblich festlegt, wie groß die Amplitude der verwendeten Felder sein darf und die zu erfüllende Schranke relativ niedrig ist.

Inwiefern dieses Ergebnis mit dem Ziel der analytischen Regularität verbessert werden kann, ist an dieser Stelle unklar. Im Rahmen der Forschungsphase für das erzielte Ergebnis wurden verschiedene Ansätze verfolgt, die im Endeffekt alle zu einer ähnlich einschränkenden hinreichenden Bedingung führten.

Alternativ kann der Anspruch an die Regularität verringert werden, da Analytizität für die Anwendung der Reduzierte-Basis-Methode zwar gut, aber nicht notwendig ist. Aufgrund der erzielten Ergebnisse liegt es nahe, dass schwächere Regularitätsergebnisse unter deutlich angenehmeren Bedingungen erreicht werden können.

Petrov-Galerkin-Diskretisierung. In Kapitel 4 wurde eine Diskretisierung durchgeführt, die als Grundlage für die Reduzierte-Basis-Methode verwendbar ist. Weiter wurde eine hinreichende CFL-Bedingung angegeben, unter welcher diese Diskretisierung stabil ist.

Auch hier bietet sich weiteres Verbesserungspotenzial. Als erste einfache Verbesserung bietet es sich an, bei der numerischen Umsetzung die Raum-Zeit-Struktur stärker zu nutzen. So müssen die vorkommenden Kronecker-Produkte nicht zu $\mathcal{N}$-dimensionalen Objekten ausgewertet werden, da viele Berechnungen, die auf diesen Objekten basieren, durch Äquivalenzaussagen zu Berechnungen in den Dimensionen $\mathcal{K}$ und $\mathcal{J}$ überführt werden können.

Eine deutlich weiter ausholende Option ist, ein bedingungslos stabiles Verfahren zu konstruieren. Andreev [And12b, Section 5.2] hat gezeigt, dass die CFL-Bedingung für das verwendete Verfahren möglicherweise verbessert, aber im Allgemeinen nicht entfernt werden kann. Eine einfache Variante, um aus dem vorliegenden Verfahren ein bedingungslos stabiles zu erhalten, ist es, für den Testraum eine verfeinerte Zeitdiskretisierung zu verwenden. Teilt man die Zeitgitterintervalle des Ansatzraumes für den Testraum jeweils in der Mitte auf, so führt dies zu bedingungsloser Stabilität. Als Problem erweist sich nun aber, dass das resultierende System in Ansatz- und Testraum nicht mehr die gleiche Dimension besitzt und dementsprechend nur noch als Residuum-minimierendes Variationsproblem aufgefasst werden kann. Dieses kann aber nicht mehr als Grundlage für die Reduzierte-Basis-Methode verwendet werden.

Es bleibt zu untersuchen, ob und wie das verwendete Verfahren optimiert oder möglicherweise durch ein „besseres“ Petrov-Galerkin-Verfahren ersetzt werden kann.

Reduzierte-Basis-Methode. Als nächster Schritt wurde in Kapitel 5 die Reduzierte-Basis-Methode eingeführt. Die theoretischen Grundlagen können zwar kurz gefasst werden, nichtsdestotrotz weist dieses Verfahren verschiedene verbesserungswürdige Punkte auf.

Wir beschränken uns auf den nach den Beispielen in Abschnitt 5.3 offensichtlichen Ansatzpunkt, die Berechnung einer unteren Schranke für die parameterabhängige inf-sup-Konstante. Dies stellt ein Kernstück der Reduzierte-Basis-Methode dar und wurde hier mit der Successive Constraint Method angegangen. Dabei hat sich herausgestellt, dass diese stark unter dem Fluch der Dimensionalität leidet. Ist die SCM bei einem Parameter noch in wenigen Minuten ausführbar, so benötigt sie bereits bei einer Parameterzahl im mittleren einstelligen Bereich mehrere Stunden.

Um dies zu verbessern, bietet sich zum einen die Verwendung einer optimierten SCM-Variante an wie beispielsweise [Huy+10]. Zum anderen können möglicherweise die Schranken nach [SS09] (siehe Korollar 2.23) als Ersatz zur SCM verwendet werden. Für Letzteres bleibt aber zunächst die Exaktheit dieser Schranken zu untersuchen. Spiegelt die Abschätzung das Verhalten der exakten inf-sup-Konstanten nur unzureichend wider, dann verfälscht dies insbesondere die Resultate des verwendeten Greedy-Verfahrens.

Auch das Greedy-Iterationsverfahren der Reduzierte-Basis-Methode scheint stark unter der Erhöhung der Parameterdimension zu leiden, liefert hier aber nur bedingt aussagekräftige Ergebnisse, da die Bestimmung der inf-sup-Konstanten die weitere Analyse für höhere Dimensionen erschwert.

Selbstkonsistente Feldtheorie. Als größter offener Punkt verbleibt die Frage, wie die in dieser Arbeit vorgestellte Modellreduktion verwendet werden kann, um die Berechnungen im Rahmen der selbstkonsistenten Feldtheorie zu beschleunigen.

Die vergleichsweise einfachste Option stellt die Substitution des in der Einleitung beschriebenen Differentialgleichung-Lösers im iterativen Verfahren durch die konstruierte

Reduzierte-Basis-Methode dar. Diese erfordert in der aktuellen Umsetzung a priori relativ tiefgehende Informationen, beispielsweise Symmetrien, über die resultierenden stabilen Felder, da sonst die Dimension des verwendeten Parameterraumes unpraktikabel groß wird. Weiter muss hierfür bei Änderung der Modellparameter vor dem iterativen Verfahren stets zunächst der Offline-Anteil der Reduzierte-Basis-Methode ausgeführt werden, was bei aktueller Umsetzung jegliche mögliche Zeiteinsparung durch die Modellreduktion zerstört.

Um auf solches a priori-Wissen verzichten zu können, ist eine dynamische Reduzierte-Basis-Methode gefragt, die nicht einfach nur als Löser-Ersatz für die Propagator-Differentialgleichungen verwendet wird. Diese sollte nach Möglichkeit direkt mit dem iterativen Verfahren gekoppelt sein und während der Ausführung „lernen“, welche Funktionen ein zur Modellreduktion der Felder geeignetes System darstellen.

Unseres Wissens wurden dynamische Reduzierte-Basis-Methoden dieser Art bisher nicht untersucht und stellen damit einen größeren, noch unerforschten Bereich dar.

A Anhang

Begleit-DVD

Dieser Arbeit liegt eine DVD bei, welche die Implementierungen der beschriebenen Verfahren sowie die in den Abschnitten 4.4 und 5.3 aufgeführten Beispiele enthält. Ebenso findet sich der Inhalt der DVD auch als Git-Repository unter `https://github.com/nobbs/master-thesis`.

Die Implementierungen und Skripte sind vollständig in MATLAB [MAT14] gehalten, bis mindestens Version 2013a abwärtskompatibel und plattformunabhängig ausführbar. Neben dem eigentlichen MATLAB-Softwarepaket wird die *Optimization Toolbox* für die linearen Programme der Successive Constraint Method benötigt.

Der Vollständigkeit halber sei hier auch die für die Simulationen verwendete und relevante Hardware genannt: zum Einsatz kam eine Intel Core 2 Duo CPU mit 3.0 GHz und 4 GB Arbeitsspeicher. Als Betriebssystem diente ferner openSUSE 11.3 Linux, Kernel 2.6.34.

Bevor der Aufbau der Implementierung genauer beleuchtet wird, soll ein kurzer Überblick über die wichtigsten Verzeichnisse geboten werden.

```
/
├── README.md
├── code ..... Enthält jeglichen Quellcode
│   ├── examples ..... Beispiele der einzelnen Kapitel
│   │   ├── chapter1
│   │   ├── chapter4
│   │   └── chapter5
│   ├── lib ..... Externe MatLab-Skripte und Libraries
│   ├── src ..... Hauptteil der Implementierung
│   │   ├── galerkin
│   │   └── rbm
│   └── test ..... Einige UnitTests für das Galerkin-Verfahren
├── doc ..... Automatisch generierte Dokumentation der Implementierung
│   └── index.html
└── tex ..... LaTeX-Dateien dieser Thesis
```

Die automatisch generierte Dokumentation im Verzeichnis `/doc` ist gut geeignet, um einen schnellen Überblick über die Implementierung zu bekommen, erspart den Blick in den Quellcode im Allgemeinen aber nicht. Um einen Einblick zu bekommen, wie die vorliegende Implementierung ausgeführt werden kann, empfiehlt sich ein Blick in die Beispiele in den Verzeichnissen `/code/examples/chapter{4,5}`. In diesen findet sich jeweils eine `README.md`, die diese Beispiele erklärt.

An dieser Stelle wollen wir in aller Kürze die wichtigen Dateien der Implementierung erwähnen.

- `/code/src/ProblemData.m` enthält eine Klasse, deren einziger Zweck ist, eine übersichtliche Definition der Modelldaten zu ermöglichen.
- `/code/src/rbm` enthält die beiden Dateien `RBM.m` und `SCM.m`, welche die Implementierung der Reduzierte-Basis-Methode beziehungsweise der Successive Constraint Method enthalten. Beide sind als Klassen implementiert und zu großen Teilen unabhängig von der zugrundeliegenden Petrov-Galerkin-Implementierung.
- `/code/src/galerkin` enthält die Bausteine des verwendeten Petrov-Galerkin-Verfahrens, die in folgende Verzeichnisse aufgeteilt sind:
  - `./spatial` enthält neben der abstrakten Klasse `SpatialAssemblyAbstract.m` die Implementierungen `SpatialAssemblyFourier.m` und `SpatialAssemblySine.m`. Letztere unterscheiden sich theoretisch zwar nur in der Wahl der Basisfunktionen, da die benötigten Gramschen Matrizen aber nicht durch numerische Quadratur sondern anhand äquivalenter Ausdrücke ausgewertet werden, ergeben sich größere Unterschiede im Quellcode.
  - `./temporal` umfasst die abstrakte Klasse `TemporalAssemblyAbstract.m` und die Implementierung `TemporalAssemblyNodal.m` sowie ferner eine weitere Implementierung `TemporalAssemblyLegendre.m`, welche ein Überbleibsel aus den Anfängen dieser Arbeit ist. Da letztere Legendre-Polynome für die zeitliche Diskretisierung verwendet und dementsprechend den zeitlichen Wechsel der Felder nur schlecht abbildet, wurde stattdessen zu dem aus Kapitel 4 bekannten Finite-Elemente-Ansatz in `TemporalAssemblyNodal.m` gewechselt.
  - `./solver` beinhaltet die abstrakte Klasse `SolverAbstract.m` und deren einzige Implementierung `SolverNodal.m`, welche die Klassen aus `./spatial` und `./temporal` zu dem Petrov-Galerkin-Verfahren kombiniert.
- `/code/test` enthält schließlich einige UnitTests zur Kontrolle der Diskretisierungen in `../src/galerkin/spatial`, da diese wie bereits erwähnt, nicht mittels numerischer Quadratur umgesetzt wurden.

Abbildungsverzeichnis

Tabellenverzeichnis

Literatur

[And12a] R. Andreev. „Space-time discretization of the heat equation. A concise Matlab implementation“. *ArXiv e-prints* (Dez. 2012). arXiv: 1212.6037 [math.NA].

[And12b] R. Andreev. „Stability of space-time Petrov-Galerkin discretizations for parabolic evolution equations“. Diss. ETH, Zürich, 2012. DOI: 10.3929/ethz-a-007563932.

[Aub00] J. P. Aubin. *Applied Functional Analysis*. 2. Aufl. Wiley-Blackwell, Jan. 2000. DOI: 10.1002/9781118032725.

[Aud+13] D. J. Audus et al. „Comparison of Pseudospectral Algorithms for Field-Theoretic Simulations of Polymers“. *Macromolecules* Bd. 46, Nr. 20 (Okt. 2013), S. 8383–8391. DOI: 10.1021/ma401804j.

[BA72] I. Babuška und A. K. Aziz. „Survey lectures on the mathematical foundations of the finite element method“. In: *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*. Hrsg. von A. K. Aziz. New York: Academic Press, 1972, S. 3–359.

[Bab71] I. Babuška. „Error-bounds for finite element method“. *Numer. Math.* Bd. 16, Nr. 4 (Jan. 1971), S. 322–333. DOI: 10.1007/bf02165003.

[Bra13] D. Braess. *Finite Elemente*. Springer Science + Business Media, 2013. DOI: 10.1007/978-3-642-34797-9.

[CDS10] A. Cohen, R. DeVore und C. Schwab. „Convergence Rates of Best N-term Galerkin Approximations for a Class of Elliptic sPDEs“. *Foundations of Computational Mathematics* Bd. 10, Nr. 6 (Juli 2010), S. 615–646. DOI: 10.1007/s10208-010-9072-2.

[CDS11] A. Cohen, R. DeVore und C. Schwab. „Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's“. *Anal. Appl.* Bd. 09, Nr. 01 (Jan. 2011), S. 11–47. DOI: 10.1142/s0219530511001728.

[CF04] H. D. Ceniceros und G. H. Fredrickson. „Numerical Solution of Polymer Self-Consistent Field Theory“. *Multiscale Modeling & Simulation* Bd. 2, Nr. 3 (Jan. 2004), S. 452–474. DOI: 10.1137/030601338.

[CGF06] E. W. Cochran, C. J. Garcia-Cervera und G. H. Fredrickson. „Stability of the Gyroid Phase in Diblock Copolymers at Strong Segregation“. *Macromolecules* Bd. 39, Nr. 7 (Apr. 2006), S. 2449–2451. DOI: 10.1021/ma0527707.

[Che+09] Y. Chen et al. „Improved successive constraint method based a posteriori error estimate for reduced basis approximation of 2D Maxwell's problem“. *ESAIM: M2AN* Bd. 43, Nr. 6 (Aug. 2009), S. 1099–1116. DOI: 10.1051/m2an/2009037.

[DF99] F. Drolet und G. H. Fredrickson. „Combinatorial Screening of Complex Block Copolymer Assembly with Self-Consistent Field Theory“. *Phys. Rev. Lett.* Bd. 83, Nr. 21 (Nov. 1999), S. 4317–4320. DOI: 10.1103/physrevlett.83.4317.

[DL92] R. Dautray und J.-L. Lions. *Mathematical analysis and numerical methods for science and technology. Volume 5 Evolution Problems I*. Berlin, Heidelberg: Springer-Verlag, Berlin, 1992.

[Eva10] L. C. Evans. *Partial Differential Equations*. 2. Aufl. Graduate studies in mathematics. American Mathematical Society, 2010.

[Fat05] H. O. Fattorini. *Infinite Dimensional Linear Control Systems: The Time Optimal and Norm Optimal Problems*. Bd. 201. North-Holland Mathematics Studies. Boston: Elsevier, 2005.

[Fre05] G. H. Fredrickson. *The Equilibrium Theory of Inhomogeneous Polymers*. Bd. 134. International Series of Monographs on Physics. Oxford University Press (OUP), Dez. 2005.

[HA09] W. Han und K. E. Atkinson. *Theoretical Numerical Analysis*. Springer New York, 2009. DOI: 10.1007/978-1-4419-0458-4.

[Had02] J. Hadamard. „Sur les problèmes aux dérivées partielles et leur signification physique“. *Princeton university bulletin* Bd. 13, Nr. 49-52 (1902), S. 28.

[Hal57] P. R. Halmos. *Introduction to Hilbert Space and the theory of spectral multiplicity*. 2. Aufl. New York: Chelsea, 1957, 114 S.

[Han09] M. Hanke-Bourgeois. *Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens*. Springer Science + Business Media, 2009. DOI: 10.1007/978-3-8348-9309-3.

[Huy+07] D. B. P. Huynh et al. „A successive constraint linear optimization method for lower bounds of parametric coercivity and inf–sup stability constants“. *Comptes Rendus Mathematique* Bd. 345, Nr. 8 (Okt. 2007), S. 473–478. DOI: 10.1016/j.crma.2007.09.019.

[Huy+10] D. B. P. Huynh et al. „A natural-norm Successive Constraint Method for inf-sup lower bounds“. *Computer Methods in Applied Mechanics and Engineering* Bd. 199, Nr. 29-32 (Juni 2010), S. 1963–1975. DOI: 10.1016/j.cma.2010.02.011.

[KS13] A. Kunoth und C. Schwab. „Analytic Regularity and GPC Approximation for Control Problems Constrained by Linear Parametric Elliptic and Parabolic PDEs“. *SIAM Journal on Control and Optimization* Bd. 51, Nr. 3 (Jan. 2013), S. 2442–2471. DOI: 10.1137/110847597.

[Lio71] J.-L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. Bd. 170. Grundlehren der mathematischen Wissenschaften. Springer-Verlag Berlin Heidelberg, 1971.

[LM72] J.-L. Lions und E. Magenes. *Non-Homogeneous Boundary Value Problems and Applications*. Bd. 181. Die Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 1972. DOI: 10.1007/978-3-642-65161-8.

[Mat07] M. W. Matsen. „Self-Consistent Field Theory and Its Applications“. In: *Soft Matter*. Wiley-VCH Verlag GmbH & Co. KGaA, 2007, S. 87–178. DOI: 10.1002/9783527617050.ch2.

[MAT14] MATLAB. *Release 2014b, Version 8.4.0.150421*, The MathWorks, Inc., Natick Massachusetts United States. Sep. 2014.

[MS94] M. W. Matsen und M. Schick. „Stable and unstable phases of a diblock copolymer melt“. *Phys. Rev. Lett.* Bd. 72, Nr. 16 (Apr. 1994), S. 2660–2663. DOI: 10.1103/physrevlett.72.2660.

[MU14] A. Mayerhofer und K. Urban. „A Reduced Basis Method for Parabolic Partial Differential Equations with Parameter Functions and Application to Option Pricing“. *ArXiv e-prints* (Aug. 2014). arXiv: 1408.2709 [math.NA].

[NSV09] R. H. Nochetto, K. G. Siebert und A. Veeser. „Theory of adaptive finite element methods: An introduction“. In: *Multiscale, Nonlinear and Adaptive Approximation*. Hrsg. von R. DeVore und A. Kunoth. Springer Berlin Heidelberg, 2009, S. 409–542. DOI: 10.1007/978-3-642-03413-8_12.

[PR06] A. T. Patera und G. Rozza. *Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations.* Version 1.0, Copyright MIT 2006, to appear in (tentative rubric) MIT Pappalardo Graduate Monographs in Mechanical Engineering. 2006.

[RC03] M. Rubinstein und R. H. Colby. *Polymer Physics*. Oxford University Press, 2003.

[RHP08] G. Rozza, D. B. P. Huynh und A. T. Patera. „Reduced Basis Approximation and a Posteriori Error Estimation for Affinely Parametrized Elliptic Coercive Partial Differential Equations“. *Archives of Computational Methods in Engineering* Bd. 15, Nr. 3 (Mai 2008), S. 229–275. DOI: 10.1007/s11831-008-9019-9.

[RK02] K. Ø. Rasmussen und G. Kalosakas. „Improved Numerical Algorithm for Exploring Block Copolymer Mesophases“. *Journal of Polymer Science Part B: Polymer Physics* Bd. 40, Nr. 16 (Juli 2002), S. 1777–1783. DOI: 10.1002/polb.10238.

[RQM08] A. Ranjan, J. Qin und D. C. Morse. „Linear Response and Stability of Ordered Phases of Block Copolymer Melts“. *Macromolecules* Bd. 41, Nr. 3 (Feb. 2008), S. 942–954. DOI: 10.1021/ma0714316.

[Sch13] B. Schweizer. *Partielle Differentialgleichungen*. Springer Science + Business Media, 2013. DOI: 10.1007/978-3-642-40638-6.

[SM11] P. Stasiak und M. W. Matsen. „Efficiency of pseudo-spectral algorithms with Anderson mixing for the SCFT of periodic block-copolymer phases“. *The European Physical Journal E* Bd. 34, Nr. 10 (Okt. 2011). DOI: 10.1140/epje/i2011-11110-0.

[SS09] C. Schwab und R. Stevenson. „Space-time adaptive wavelet methods for parabolic evolution problems“. *Mathematics of Computation* Bd. 78, Nr. 267 (Sep. 2009), S. 1293–1318. DOI: 10.1090/s0025-5718-08-02205-9.

[TRL04] R. B. Thompson, K. Ø. Rasmussen und T. Lookman. „Improved convergence in block copolymer self-consistent field theory by Anderson mixing“. *The Journal of Chemical Physics* Bd. 120, Nr. 1 (2004), S. 31. DOI: 10.1063/1.1629673.

[UP13] K. Urban und A. T. Patera. „An improved error bound for reduced basis approximation of linear parabolic problems“. *Mathematics of Computation* Bd. 83, Nr. 288 (Okt. 2013), S. 1599–1615. DOI: 10.1090/s0025-5718-2013-02782-2.

Eidesstattliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Masterarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, wobei ich alle wörtlichen und sinngemäßen Zitate als solche gekennzeichnet habe. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Mainz, den 10. September 2015