

FALTUNGSALGORITHMEN FÜR  
TRANSPARENTE  
RANDBEDINGUNGEN

Inaugural-Dissertation

zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

**Marina Elisabeth Fischer**

aus Düsseldorf

Düsseldorf, im Mai 2022

aus dem Institut für Mathematik  
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der  
Mathematisch-Naturwissenschaftlichen Fakultät der  
Heinrich-Heine-Universität Düsseldorf

Berichtersteller:

- (i) Prof. Dr. Achim Schädle  
Heinrich-Heine-Universität Düsseldorf
- (ii) PD Dr. Kersten Schmidt  
Technische Universität Darmstadt

Tag der mündlichen Prüfung: 15. August 2022

---

# ZUSAMMENFASSUNG

In dieser Arbeit betrachten wir zwei zeitabhängige Probleme – zum einen die fraktionale Schrödinger-Gleichung auf der reellen Achse, zum anderen die Wellengleichung auf einem zweidimensionalen periodischen Wellenleiter. Hierbei gehen wir von einem lokalen Defekt in einem beschränkten Gebiet  $\Omega_0$  des Ausbreitungsmediums aus. Für beide Probleme leiten wir transparente Randbedingungen in Form von Faltungsintegralen her, mit deren Hilfe die Lösung in  $\Omega_0$  bestimmt werden kann, sowie zwei Operatoren, anhand derer wir die Lösung in jeder beliebigen Periodizitätszelle des Außenraums ausgehend von den berechneten Daten auf dem Rand des Gebiets  $\Omega_0$  erneut unter Berechnung eines Faltungsintegrals erhalten. Hierbei beobachten wir, dass sowohl die Laplace-Transformierte des Faltungskerns für die Darstellung der transparenten Randbedingungen als auch die des Faltungskerns für die zellenweise Lösungsdarstellung im Fall der fraktionalen Schrödinger-Gleichung analytisch in einem Sektor  $\{\lambda \in \mathbb{C} \setminus \{0\} \mid |\arg(\lambda)| < \pi/2 + \theta_0\}$  mit  $0 < \theta_0 \leq \pi/2$  ist und dort die polynomielle Schranke  $\mathcal{O}(\lambda^{-\mu})$ ,  $\mu > 0$ , vorweist. Die Laplace-Transformierten im Fall der Wellengleichung dagegen sind analytisch und polynomiell beschränkt lediglich in der rechten Halbebene.

Im ersten Teil dieser Arbeit studieren wir daher Methoden zur Faltungsquadratur, die auf das unterschiedliche Verhalten der Laplace-Transformierten angepasst sind. Hier erweitern wir zum einen die, für den Fall  $\mu = 1$ , bereits bekannte Methode der exponentiellen Faltungsquadratur auf den Fall einer sektoriellen Laplace-Transformierten mit beliebigem Parameter  $\mu > 0$  und beweisen die Konvergenz des Verfahrens. Zum anderen geben wir die wichtigsten Grundzüge der Runge-Kutta basierten Faltungsquadratur im nicht-sektoriellen Fall an. Für beide Methoden beschreiben wir Faltungsalgorithmen, die jeweils durch verschiedene Merkmale als „schnell“ bezeichnet werden können.

Im zweiten Teil stellen wir unter Verwendung der inversen Laplace-Transformation einen Zusammenhang zwischen der zeitharmonischen Helmholtz-Gleichung und den zu untersuchenden zeitabhängigen Problemen her, wodurch wir auf eine bekannte Methode zur Bestimmung der benötigten Operatoren im Frequenzbereich zurückgreifen können, die uns letztendlich die oben genannten Darstellungen der transparenten Randbedingungen und der zellenweisen Lösung liefern. Die vollständigen Diskretisierungen beider Probleme sowie die Anwendungen der schnellen Faltungsalgorithmen zur Berechnung der Lösung in  $\Omega_0$  werden herausgearbeitet. Die numerische Berechnung der zellenweisen Lösung im Außenraum wird ebenfalls Bestandteil dieser Arbeit sein und zusammen mit den schnellen Algorithmen anhand von numerischen Experimenten untersucht.



---

# ABSTRACT

In this thesis, we focus on two time-dependent problems, namely the time-fractional Schrödinger equation on the real line and the wave equation in a two-dimensional periodic waveguide. We consider a local perturbation in a bounded domain  $\Omega_0$  of the propagation media. For either problem, we derive transparent boundary conditions in the form of convolution integrals which can be used to compute the solution in the domain  $\Omega_0$ . Additionally, we obtain two further operators that allow us to establish the solution on an arbitrary periodicity cell from the data on the boundary of  $\Omega_0$  by, once again, computing a convolution integral. By taking a closer look at the representation of the transparent boundary condition and of the cell-wise solution we notice that the Laplace transform of both convolution kernels in the case of the fractional Schrödinger equation are analytic and satisfy the polynomial bound  $\mathcal{O}(\lambda^{-\mu})$ ,  $\mu > 0$ , in a sector  $\{\lambda \in \mathbb{C} \setminus \{0\} \mid |\arg(\lambda)| < \pi/2 + \theta_0\}$  with  $0 < \theta_0 \leq \pi/2$ . When considering the wave equation the Laplace transforms prove to be analytic and polynomially bounded only in the right half plane.

Therefore, in the first part of this thesis, we study convolution quadrature methods that match the Laplace transforms' different behaviour. On the one hand, we expand the, in the case  $\mu = 1$ , known exponential quadrature rule to the case of a sectorial Laplace transform with an arbitrary parameter  $\mu > 0$  and prove the method's convergence. On the other hand, the most important basics of the Runge-Kutta based convolution quadrature for non-sectorial Laplace transforms are stated. For both methods, we describe convolution algorithms each of which can be denoted as "fast" due to different features.

On using the inverse Laplace transform in the second part, we establish a connection between the time-harmonic Helmholtz equation and the time-dependent problems that we focus on. Thus, a well-known approach for periodic media can be applied to derive operators in the frequency domain that finally yield the transparent boundary conditions and the cell-wise representation in the time domain that are mentioned above. The fully discretised versions of both problems as well as the application of the fast algorithms for computing the solution in  $\Omega_0$  are presented in detail. Furthermore, we discuss the computation of the cell-wise solution in the exterior space and, together with the fast algorithm, test it by conducting some numerical experiments.



---

# INHALTSVERZEICHNIS

<b>Einleitung</b>	<b>1</b>
<b>1 Exponentielle Faltungsquadratur für sektorielle Laplace-Transformierte</b>	<b>7</b>
1.1 Lagrange-Interpolation im Banachraum . . . . .	8
1.2 Exponentielle Quadratur . . . . .	9
1.2.1 Grundidee der Faltungsquadratur . . . . .	9
1.2.2 Eigenschaften der sektoriellen Laplace-Transformierten . . . . .	11
1.2.3 Faltungsquadratur durch Polynominterpolation . . . . .	15
1.3 Approximation der Faltungsgewichte . . . . .	20
1.4 Schneller Faltungsalgorithmus . . . . .	24
1.5 Numerischer Test . . . . .	27
<b>2 Runge-Kutta Faltungsquadratur für nicht-sektorielle Laplace-Transformierte</b>	<b>35</b>
2.1 Runge-Kutta Verfahren . . . . .	36
2.2 Runge-Kutta basierte Faltungsquadratur . . . . .	37
2.3 Implementierung mit Hilfe der schnellen Fourier-Transformation . . . . .	40
2.3.1 Berechnung der Faltungsgewichte . . . . .	40
2.3.2 Lösen einer Volterra Integralgleichung . . . . .	40
<b>3 Helmholtz-Gleichung im periodischen Wellenleiter mit lokalem Defekt</b>	<b>47</b>
3.1 Bestimmung der transparenten Randbedingungen . . . . .	48
3.1.1 Modellproblem . . . . .	48
3.1.2 Eigenschaften des Spektrums . . . . .	52
3.1.3 Existenz und Eindeutigkeit einer Lösung im Wellenleiter . . . . .	55
3.1.4 Helmholtz-Gleichung im Halb-Wellenleiter . . . . .	58
3.1.5 Lösen lokaler Probleme in $C_0^+$ und $\Omega_0$ . . . . .	61
3.2 Berechnungen im eindimensionalen Wellenleiter . . . . .	66
3.2.1 Übertragung der Resultate in 1D . . . . .	66
3.2.2 Ortskontinuierlicher Fall . . . . .	67
3.2.2.1 Homogener Wellenleiter . . . . .	68
3.2.2.2 Inhomogener Wellenleiter . . . . .	69
3.2.3 Ortsdiskreter Fall . . . . .	73
3.2.3.1 Homogener Wellenleiter . . . . .	73
3.2.3.2 Inhomogener Wellenleiter . . . . .	75

<b>4</b>	<b>Zeit-fractionale Schrödinger-Gleichung auf der reellen Achse</b>	<b>81</b>
4.1	Bestimmung der transparenten Randbedingungen . . . . .	82
4.1.1	Modellproblem . . . . .	82
4.1.2	Existenz und Eindeutigkeit einer Lösung . . . . .	83
4.1.3	Zusammenhang mit Helmholtz-Gleichung . . . . .	88
4.2	Schneller Faltungsalgorithmus für die fraktionale Schrödinger-Gleichung . . . . .	90
4.2.1	Anwendung der exponentiellen Faltungsquadratur . . . . .	90
4.2.2	Diskretisierung im Ort . . . . .	92
4.2.3	Vollständig diskretisiertes System und schneller Algorithmus . . . . .	94
4.3	Numerisches Experiment . . . . .	95
<b>5</b>	<b>Wellengleichung im periodischen Wellenleiter mit lokalem Defekt</b>	<b>99</b>
5.1	Bestimmung der transparenten Randbedingungen . . . . .	100
5.1.1	Modellproblem . . . . .	100
5.1.2	Existenz und Eindeutigkeit einer Lösung . . . . .	101
5.1.3	Zusammenhang mit Helmholtz-Gleichung . . . . .	103
5.2	Schneller Faltungsalgorithmus für die Wellengleichung . . . . .	105
5.2.1	Anwendung von Runge-Kutta Verfahren und Faltungsquadratur . . . . .	105
5.2.1.1	Semidiskrete Approximation in den Zellen . . . . .	106
5.2.1.2	Semidiskrete Approximation in $\Omega_0$ . . . . .	107
5.2.2	Diskretisierung im Ort . . . . .	108
5.2.2.1	Finite Differenzen . . . . .	109
5.2.2.2	Finite Elemente . . . . .	110
5.2.3	Vollständig diskretisiertes System . . . . .	112
5.2.4	Zusammenfassung des Algorithmus . . . . .	113
5.3	Numerische Experimente . . . . .	115
5.3.1	Wellengleichung in einer Raumdimension . . . . .	115
5.3.2	Wellengleichung in zwei Raumdimensionen . . . . .	118
<b>6</b>	<b>Ausblick</b>	<b>123</b>
	<b>Literatur</b>	<b>125</b>

---

# EINLEITUNG

In vielen Bereichen des Alltags begegnen uns Wellen, seien es Schallwellen, elektromagnetische Wellen oder einfach nur Wasserwellen, die sich in verschiedenen Medien ausbreiten können. Neben offensichtlichen Beispielen, wie Luft und Wasser, zählen Glasfaserkabel, (Halb)Leiter und natürlich vorkommende oder künstlich geschaffene photonische Kristalle zu weiteren Beispielen solcher Medien. Die Struktur des Mediums hat hierbei Einfluss auf die Ausbreitung der Welle: Ist das Medium homogen, so besitzt es überall dieselben physikalischen Eigenschaften. Damit bleibt beispielsweise die Ausbreitungsgeschwindigkeit der Welle konstant. Ein inhomogenes Medium, worunter auch periodische Medien zählen, kann dagegen unter anderem zu variierenden Geschwindigkeiten führen.

Für ein tiefergehendes Verständnis, wie sich eine Welle im jeweiligen Medium verhält, müssen daher mathematische Modelle gefunden werden, die die Ausbreitung möglichst genau in einen theoretischen Kontext übersetzen. Hierbei werden die Wellen in Form von partiellen Differentialgleichungen beschrieben; das Gebiet, in dem die Gleichung betrachtet wird, wird in der Regel als unbeschränkt in eine oder mehrere Richtungen angenommen. So kann es sich bei dem betrachteten Gebiet in einem zweidimensionalen Modell um den gesamten Raum handeln, oder beispielsweise um einen Wellenleiter, der in eine Richtung beschränkt, in die andere Richtung aber unbeschränkt ist.

Für die numerische Berechnung der Wellenausbreitung stellt sich die Frage, wie die theoretische Betrachtung eines unbeschränkten Gebiets auf ein endliches Problem reduziert werden kann. Hierbei kommt uns zu Gute, dass die Lösung in der Regel auch nur auf einem kleineren Bereich von Interesse ist.

Die Antwort auf die Frage findet sich hierbei im Stichwort der *transparenten* – oder auch nicht-reflektierenden beziehungsweise absorbierenden – *Randbedingungen*. Diese werden für das betrachtete Problem so konstruiert, dass dessen Lösung auf einem beschränkten Gebiet unter Verwendung des transparenten Randes und die Lösung, die zuerst auf dem unbeschränkten Gebiet berechnet und dann auf das beschränkte Gebiet eingeschränkt wird, übereinstimmen.

Nun könnte man sich die Frage stellen, welche Vorteile diese Herangehensweise für die numerische Berechnung bietet. Wäre es nicht auch möglich, ein etwas größeres beschränktes Gebiet unter Verwendung von Dirichlet-Randwerten zu betrachten, und es dabei so groß zu wählen, dass wir sicher sein können, die gesuchte Lösung in unserem kleineren Bereich ist noch unbeeinflusst von Reflexionen, die vom Rand zurückgeworfen werden?

Um den Vorteil der ersten Herangehensweise zu verdeutlichen und die Behandlung von transparenten Randbedingungen in dieser Arbeit zu motivieren, gehen wir kurz auf ein Beispiel unter vereinfachten Bedingungen ein:

Nehmen wir dazu an, wir betrachten eine Welle, die sich entlang der reellen Achse ausbreitet, unter der Annahme, dass der Träger des Startwerts zum Zeitpunkt  $t = 0$  im Intervall  $\Omega_0 = [-1/2, 1/2]$  liegt und die Ausbreitungsgeschwindigkeit im Außenraum, also außerhalb von  $\Omega_0$ , durch den Wert 1 beschränkt ist. Das heißt, zum Zeitpunkt  $t = 1/2$  befinden sich die Wellenberge höchstens an den Stellen  $x = \pm 1$ , zum Zeitpunkt  $t = 1$  höchstens bei  $x = \pm 3/2$  und so weiter. Um das Beispiel ein kleines bisschen komplizierter zu gestalten, nehmen wir weiter an, dass der eindimensionale Wellenleiter, durch den sich die Welle bewegt, in gleichbleibenden Abständen unterschiedliche Eigenschaften aufweist. Dadurch kann es zu Brechungen und Reflexionen kommen, die sich zwischen den beiden Wellenbergen befinden. Um numerisch die Lösung zum Zeitpunkt  $T = Nh$  in  $\Omega_0$  zu bestimmen, setzen wir voraus, dass das Intervall mit  $M$  Gitterpunkten diskretisiert ist, wobei Ortsgitterweite und Zeitschrittweite der Einfachheit halber als gleich groß angenommen werden, das heißt, es gilt

$$\frac{|\Omega_0|}{M-1} = \frac{1}{M-1} = \frac{T}{N}.$$

Berechnen wir die Lösung in  $\Omega_0$  mit Hilfe der transparenten Randbedingungen an  $x = \pm 1/2$ , so ergibt sich ein Aufwand von

$$c_1 NM + c_2 N \log(N) \quad \text{beziehungsweise} \quad c_1 NM + c_3 N (\log(N))^2,$$

wobei die Konstante  $c_1$  unter anderem durch das Verfahren zur Lösung der Gleichungssysteme, das wir als möglichst effizient annehmen, oder durch die Anwendung der Diskretisierung beeinflusst wird. Der zweite Summand und die entsprechende Konstante ergeben sich aus der Auswertung des transparenten Randes durch zwei unterschiedliche schnelle Algorithmen, auf die wir noch genauer eingehen werden.

Wollen wir stattdessen die Lösung auf einem größeren Intervall  $\Omega_{\text{Dir}} = [-b, b]$  mit Nullrandwerten links und rechts berechnen, müssen wir bedenken, dass die Wellenberge reflektiert werden, sobald sie an den Rand  $x = \pm b$  treffen. Um zum Zeitpunkt  $T$  damit garantieren zu können, dass noch keine Reflexion das Intervall  $\Omega_0$  erreicht hat, muss also  $b = (T+1)/2$  betrachtet werden. Aus dem Zusammenhang  $T/N = |\Omega_{\text{Dir}}|/(M_{\text{Dir}}-1)$  folgt somit, dass  $\Omega_{\text{Dir}}$  mit

$$M_{\text{Dir}} = \frac{N}{T} \cdot \frac{2(T+1)}{2} + 1 = N + \frac{N}{T} + 1 = N + M$$

äquidistanten Gitterpunkten diskretisiert werden muss. Für die Berechnung der Lösung in  $\Omega_{\text{Dir}}$  ergibt sich damit analog ein Aufwand von

$$c_1 N(N + M) = c_1 NM + c_1 N^2.$$

Damit lässt sich schon in diesem einfachen Beispiel ein geringerer Aufwand zur Berechnung der Lösung unter Verwendung von transparenten Randbedingungen erkennen, der mit steigenden Zeitschritten  $N$  besonders deutlich wird. Je höher die Ausbreitungsgeschwindigkeit

der Welle ist, desto größer muss auch das Intervall  $\Omega_{\text{Dir}}$  gewählt werden, was wiederum den Aufwand erhöht.

Bei dem Medium, das wir in dieser Arbeit zugrunde legen, gehen wir von der reellen Achse oder einem zweidimensionalen Wellenleiter aus, der in die  $x$ -Richtung unbeschränkt, in die  $y$ -Richtung aber beschränkt ist<sup>1</sup>. Zusätzlich berücksichtigt unser Modell eine periodische Struktur, die in einem beschränkten Gebiet  $\Omega_0$  einen Defekt aufweist. Wir werden sowohl die kontinuierlichen als auch die diskreten Darstellungen der transparenten Randbedingungen zur Berechnung der Lösung in  $\Omega_0$  für den Fall der Wellengleichung sowie der fraktionalen Schrödinger-Gleichung herleiten. Für beide Fälle werden wir schnelle Algorithmen zur Berechnung der Lösung in  $\Omega_0$  angeben, die jeweils an die unterschiedlichen Eigenschaften der betrachteten Probleme angepasst sind. Außerdem werden wir unter Ausnutzung der Periodizitätszellen und anhand der Daten auf den Rändern eine zellenweise Darstellung der Lösung im Außenraum angeben, deren numerische Berechnung ebenfalls Bestandteil dieser Arbeit ist.

Unter Verwendung der Laplace-Transformation in der Zeit, werden wir die beiden zu untersuchenden zeitabhängigen Probleme auf Helmholtz-Gleichungen im Frequenzbereich zurückführen. Zur Herleitung der transparenten Randbedingungen werden wir uns dabei des Vorgehens aus [20, 21, 40] bedienen, das wir auf den Fall der Neumann-zu-Dirichlet-Operatoren und auf die Verwendung von komplexen Frequenzen anpassen. Die Idee der Autoren besteht hierbei darin, die gesuchten Randbedingungen unter Betrachtung zweier lokaler Probleme in einer Periodizitätszelle zu konstruieren und bietet sich sowohl für die Untersuchung von ortskontinuierlichen als auch von ortsdiskreten Problemen an. Die Anwendung der inversen Laplace-Transformation, mit der wir die Theorie aus dem Frequenzbereich wieder in den Zeitbereich transferieren, liefert uns sowohl transparente Randbedingungen als auch die zellenweise Darstellung der Lösung in Form von Faltungsgleichungen, die vereinfacht als

$$\int_0^t k(t - \tau) g(\tau) d\tau, \quad t > 0$$

mit operatorwertigem Faltungskern  $k$  und vektorwertiger Funktion  $g$  angegeben werden können. Hierbei wird sich herausstellen, dass sowohl die Laplace-Transformierte  $K(\lambda)$  des Faltungskerns für die Darstellung der transparenten Randbedingungen als auch die des Faltungskerns für die zellenweise Lösungsdarstellung im Fall der fraktionalen Schrödinger-Gleichung analytisch in einem Sektor der Form  $\Sigma_{\pi/2+\theta_0} := \{\lambda \in \mathbb{C} \setminus \{0\} \mid |\arg(\lambda)| < \pi/2 + \theta_0\}$  mit einem Winkel  $0 < \theta_0 \leq \pi/2$  ist und dort ein polynomiell abfallendes Verhalten  $\mathcal{O}(\lambda^{-\mu})$ ,  $\mu > 0$ , vorweist. Die Laplace-Transformierten im Fall der Wellengleichung sind dagegen analytisch und polynomiell beschränkt lediglich in der rechten Halbebene.

Der Aufbau der vorliegenden Arbeit ist daher wie folgt: In den ersten beiden Kapiteln

<sup>1</sup>Da sich die Wellenausbreitung in einem solchen Wellenleiter lediglich auf die  $x$ -Achse bezieht, könnte man diesen auch als Spezialfall einer eindimensionalen Betrachtung auffassen. Zur besseren Unterscheidung zwischen der reellen Achse und dem beschriebenen Wellenleiter als Ausbreitungsmedium werden wir im Folgenden jedoch weiter von einem zweidimensionalen Wellenleiter sprechen. Die reelle Achse wird gelegentlich als eindimensionaler Wellenleiter bezeichnet.

widmen wir uns dem Thema der Faltungsquadratur. In Kapitel 1 werden wir ein exponentielles Faltungsquadratur-Verfahren herleiten, das auf der stückweisen Polynominterpolation der Funktion  $g$  basiert. Da wir hierbei von einer Laplace-Transformierten ausgehen, die im Sektor  $\Sigma_{\pi/2+\theta_0}$  die polynomielle Schranke  $\mathcal{O}(\lambda^{-\mu})$ ,  $\mu > 0$ , erfüllt, stellt das vorgestellte Verfahren eine Erweiterung der Theorie aus [36, 37] sowie aus [26, 27] dar. Unter Verwendung der Deformierbarkeit der Bromwich-Linie werden wir die Konvergenz des Verfahrens zeigen, die von der Anzahl der Stützstellen in der Polynominterpolation bestimmt wird. Anschließend werden wir mit Hilfe der Ideen aus [18, 60] einen schnellen Faltungsalgorithmus zur Berechnung der Lösung zu einem festen Zeitpunkt  $t_N = Nh$  entwickeln, der das Faltungsintegral mit einem Aufwand von nur  $\mathcal{O}(N \log(N))$  berechnet.

Im zweiten Kapitel werden wir Laplace-Transformierte betrachten, die ein polynomielles Abfallverhalten nur in der rechten Halbebene vorweisen. Hierfür werden wir die wichtigsten Bestandteile der Runge-Kutta basierten Faltungsquadratur angeben, die in [3, 4] auf die Behandlung dieses Falls angepasst wurde. Zusätzlich werden wir auf zwei Aspekte der numerischen Berechnung genauer eingehen – zum einen die Berechnung der Faltungsgewichte, zum anderen das Lösen einer Integralgleichung – und zeigen auf, wie auch hier schnelle Algorithmen verwendet werden können.

Nachdem wir die notwendigen Mittel zur Approximation von Faltungsintegralen aufgeführt haben, wird der Fokus in Kapitel 3 auf der zeitharmonischen Helmholtz-Gleichung im periodischen Wellenleiter mit lokalem Defekt liegen. Wir werden zunächst das zugehörige Spektrum untersuchen, um herauszufinden, ob die Betrachtung von komplexen Frequenzen möglich ist. Anschließend gehen wir auf die Konstruktion der transparenten Randbedingungen, mit denen wir die Lösung im lokal gestörten Teil des Wellenleiters bestimmen können, sowie auf die Darstellung der zellenweise Lösung ein und geben dazu die wichtigsten Resultate der Ideen aus den bereits genannten Papern [21, 40] wieder. Um eine Vorstellung des Abfallverhaltens der dabei vorkommenden Operatoren zu erhalten, werden wir im letzten Teil des dritten Kapitels zwei Beispiele im eindimensionalen Wellenleiter genauer untersuchen und sowohl die ortskontinuierlichen als auch -diskreten Operatoren herleiten.

In Kapitel 4 betrachten wir die fraktionale Schrödinger-Gleichung im eindimensionalen Wellenleiter. Durch den Zusammenhang, den wir mittels der Laplace-Transformation zwischen dem zeitabhängigen Problem und der zeitharmonischen Helmholtz-Gleichung ziehen, sind wir in der Lage, sowohl die transparenten Randbedingungen als auch die stückweise Berechnung der Lösung als Faltungsintegrale zu schreiben. Wir werden zusätzlich auf die Theorie aus [5, 18] zurückgreifen, mit der wir die Lösung der fraktionalen Gleichung auf der reellen Achse ebenfalls als Faltungsintegral darstellen können. Unter Angabe des diskretisierten Problems werden wir auf die Anwendung des schnellen Algorithmus aus Kapitel 1 eingehen.

Der Fokus in Kapitel 5 liegt auf der Behandlung der Wellengleichung im zweidimensionalen Wellenleiter. Erneut können wir hier die zeitabhängigen transparenten Randbedingungen, sowie die zellenweise Lösung als Faltungen darstellen, wobei wir im Gegensatz zu Kapitel 4 von dem gewünschten Abfallverhalten lediglich in der rechten Halbebene ausgehen werden. Hier werden wir die Faltungsquadratur aus Kapitel 2 anwenden und gehen dabei auch genauer auf die zellenweise Berechnung der Lösung ein. Dass sowohl deren Zeitdiskretisierung,

als auch die Zeitdiskretisierung der Lösung im Gebiet  $\Omega_0$  unter Verwendung der transparenten Randbedingungen konvergente Verfahren liefern, wird im Anschluss gezeigt. Um das Kapitel, und damit auch den inhaltlichen Teil der Arbeit, abzuschließen, werden wir die vollständige Diskretisierung des Problems angeben und die Resultate an numerischen Experimenten überprüfen.

Das letzte Kapitel dient dem Ausblick auf mögliche Aspekte, die als Gegenstand weiterer Untersuchungen von Interesse sein könnten.

Dies ist nun der passende Zeitpunkt, um denjenigen meinen Dank auszusprechen, die direkt oder auch indirekt zur Entstehung dieser Arbeit beigetragen haben.

Allen voran möchte ich meinem Betreuer Prof. Dr. Achim Schädle danken - natürlich für die Ermöglichung dieser Arbeit, für die vielen Stunden auf der Suche nach größeren und kleineren Fehlern in der Implementierung, für die vielen hilfreichen fachlichen Diskussionen, für die Hinweise und Tipps zur Verbesserung der Arbeit gerade in den letzten Wochen und für die Unterstützung während der gesamten Zeit.

Weiterer Dank gebührt auch allen ehemaligen und aktuellen Kolleginnen und Kollegen, insbesondere der Lehrstühle für Mathematische Optimierung und Angewandte Mathematik. Wir sind und waren schon immer ein toller Flur!

Zuletzt bedanke ich mich natürlich bei meinen besten Freundinnen und meiner Familie - meinen Eltern, meiner Schwester und meinem Freund. Vielen Dank für die Unterstützung, die Geduld, den Rückhalt und den notwendigen Ausgleich über die letzten Jahre.



## KAPITEL 1

---

# EXPONENTIELLE FALTUNGSQUADRATUR FÜR SEKTORIELLE LAPLACE-TRANSFORMIERTE

Der Ausgangspunkt für dieses und das folgende Kapitel ist die Betrachtung des Faltungsin-  
tegrals

$$\int_0^t k(t-\tau)g(\tau) d\tau, \quad t > 0, \quad (1.1)$$

für eine operatorwertige Funktion  $k : \mathbb{R}_0^+ \rightarrow \mathcal{L}(X)$  und eine vektorwertige Funktion  $g : \mathbb{R} \rightarrow X$ , die wir als genügend glatt annehmen. Hierbei ist  $(X, \|\cdot\|_X)$  ein Banachraum. Wir gehen davon aus, dass der Faltungskern  $k$  nur als Integral oder Reihe dargestellt werden kann oder schwer auswertbar ist, dessen Laplace-Transformierte  $K$  jedoch über eine geschlossene Formel verfügt, die gegebenenfalls von einfacherer Gestalt ist. An  $K$  stellen wir hierbei die Anforderung eines geeigneten Abfallverhaltens, das sich in diesem Kapitel auf einen Sektor der Form

$$\Sigma_{\pi/2+\theta_0} := \{\lambda \in \mathbb{C} \setminus \{0\} \mid |\arg(\lambda)| < \pi/2 + \theta_0\} \quad \text{mit} \quad 0 < \theta_0 \leq \pi/2$$

bezieht. Wir sprechen hierbei von einer *sektoriellen* Laplace-Transformierten  $K$ . Im darauffolgenden Kapitel gehen wir auch auf eine Klasse von Laplace-Transformierten ein, die zwar nicht sektoriell sind, dafür aber ein bestimmtes Abfallverhalten in der rechten Halbebene aufweisen.

Zur Approximation der Faltung (1.1) mit sektorieller Laplace-Transformierter sind unter anderem die Quadratur basierend auf Mehrschrittverfahren und die Runge-Kutta basierte Faltungsquadratur, siehe [46–48], zu nennen. Auf letztere werden wir im nächsten Kapitel genauer eingehen.

Anstatt jedoch diese bereits bekannten Verfahren zu verwenden, untersuchen wir in diesem Kapitel die Faltungsquadratur basierend auf der stückweise polynomiellen Interpolation der Funktion  $g$ . Das Verfahren, das wir auf diese Art erhalten, gehört damit zur Klasse der *exponentiellen Integriatoren* und ist auch als *exponentielles Quadraturverfahren* bekannt, siehe [36].

Hierbei wurde die exponentielle Quadratur unter anderem in [36, 37] bereits für den Fall untersucht, dass die Laplace-Transformierte des Faltungskerns durch  $K(\lambda) = (\lambda I - A)^{-1}$  mit einem dicht-definierten, abgeschlossenen und linearen Operator  $A : D(A) \rightarrow X$  gegeben ist und die Abschätzung in Annahme 1.2 mit  $\mu = 1$  erfüllt. Eine Anwendung auf die Betrachtung eines schwächeren Abfallverhaltens mit  $0 < \mu < 1$  ist beispielsweise in [26, 27] zu finden, allerdings wird dort zugrunde gelegt, dass der Faltungskern als konvergente Potenzreihe um den Entwicklungspunkt 0 geschrieben werden kann. Damit stellt die folgende Betrachtung sowohl eine Erweiterung der erstgenannten beiden Veröffentlichungen auf den allgemeineren Fall  $\mu > 0$  als auch eine Erweiterung der Theorie in den beiden letztgenannten auf den Fall dar, in dem wir nicht unbedingt von einer konvergenten Potenzreihendarstellung um den Punkt 0 ausgehen können.

Wir werden in diesem Kapitel zunächst kurz auf die Polynominterpolation in Banachräumen eingehen, bevor wir dann das exponentielle Quadraturverfahren für den Fall einer sektoriellen Laplace-Transformierten untersuchen. Hierbei werden wir den Integrationsweg in der inversen Laplace-Transformation geeignet deformieren, um eine Konvergenzaussage zu erhalten. Im zweiten Teil des Kapitels leiten wir einen Algorithmus her, der es uns ermöglicht, die Faltungsquadratur auf eine schnelle und parallele Art zu approximieren. Hierbei verwenden wir Ideen aus [18, 45].

## 1.1 Lagrange-Interpolation im Banachraum

Wir betrachten  $s$  paarweise verschiedene Knoten im Intervall  $[0, 1]$ , das heißt genauer haben wir die  $s$  Knoten

$$0 \leq c_1 < \dots < c_s \leq 1$$

gegeben und definieren  $t_n := nh$  für eine Zeitschrittweite  $h > 0$ .

**Bemerkung.** Die Betrachtung eines äquidistanten Zeitgitters ist hierbei für die folgenden Überlegungen nicht notwendig und wird nur der Übersichtlichkeit halber angenommen. Die theoretischen Ergebnisse sowie der schnelle Algorithmus lassen sich auf einfache Weise an variable Schrittweiten anpassen, siehe dazu beispielsweise [45, Bem. 2].

Für den Zeitpunkt  $T = Nh$  mit  $N \in \mathbb{N}$  wollen wir die Funktion  $g : [0, T] \rightarrow X$  auf jedem Teilintervall  $[t_n, t_{n+1}]$  durch ein Polynom  $p_n$  vom Grad  $s - 1$  approximieren, so dass die

Interpolationsbedingungen

$$p_n(t_n + c_i h) = g(t_n + c_i h) \quad \text{für } i = 1, \dots, s \quad \text{und } n = 0, \dots, N-1$$

erfüllt sind. Dazu wählen wir die bekannte Lagrange-Darstellung des Interpolationspolynoms und schreiben  $p_n$  in der Form

$$p_n(t_n + \vartheta h) = \sum_{i=1}^s \ell_i(\vartheta) g(t_n + c_i h), \quad (1.2)$$

wobei das  $i$ -te Lagrange-Fundamentalpolynom für  $\vartheta \in [0, 1]$  durch

$$\ell_i(\vartheta) := \prod_{\substack{k=1 \\ k \neq i}}^s \frac{\vartheta - c_k}{c_i - c_k}$$

gegeben ist, vergleiche beispielsweise [62]. Offensichtlich gilt  $\ell_i(c_j) = \delta_{ij}$ .

Für den standardmäßig betrachteten Fall  $X = \mathbb{R}$  lässt sich mit Hilfe des Satzes von Rolle schnell zeigen, dass es zu jedem  $\vartheta \in [0, 1]$  ein  $t \in [t_n, t_{n+1}]$  gibt, so dass der Interpolationsfehler die Gleichheit

$$g(t_n + \vartheta h) - p_n(t_n + \vartheta h) = \frac{h^s}{s!} \omega(\vartheta) g^{(s)}(t)$$

für das Knotenpolynom  $\omega(\vartheta) := \prod_{i=1}^s (\vartheta - c_i)$  erfüllt, siehe beispielsweise [62]. Für unseren allgemeinen Fall, in dem  $X$  ein Banachraum ist, stimmt dies nicht mehr. Hier gilt stattdessen der folgende Satz:

**Satz 1.1.** *Es sei  $g \in C^s([t_n, t_{n+1}], X)$ , dann erhalten wir den Interpolationsfehler*

$$\|g(t_n + \vartheta h) - p_n(t_n + \vartheta h)\|_X \leq \frac{h^s}{s!} |\omega(\vartheta)| \max_{t \in [t_n, t_{n+1}]} \|g^{(s)}(t)\|_X.$$

*Beweis.* Der Beweis lässt sich unter Betrachtung des Peano-Kerns führen und folgt direkt aus Beispiel 3.1 in [51].  $\square$

## 1.2 Exponentielle Quadratur

### 1.2.1 Grundidee der Faltungsquadratur

In diesem Kapitel betrachten wir, wie bereits erwähnt, die Laplace-Transformierte

$$K : D(K) \subset \mathbb{C} \rightarrow \mathcal{L}(X),$$

$$\lambda \mapsto K(\lambda) := \mathcal{L}\{k(t); \lambda\} = \int_0^\infty e^{-\lambda t} k(t) dt \quad \text{für } \operatorname{Re}(\lambda) > \sigma_0 > 0,$$

wobei mit  $\sigma_0 > 0$  die Konvergenzabszisse bezeichnet wird und nehmen an, dass die folgende Sektorialitätsbedingung erfüllt wird, vergleiche dazu [48]:

**Annahme 1.2.** Die Laplace-Transformierte  $K$  ist analytisch im Sektor  $\Sigma_{\pi/2+\theta_0}$  mit  $0 < \theta_0 \leq \pi/2$  und wird für einen Exponenten  $\mu > 0$  und die von  $\theta_0$  abhängige Konstante  $C(\theta_0) > 0$  beschränkt durch:

$$\|K(\lambda)\|_{\mathcal{L}(X)} \leq C(\theta_0) |\lambda|^{-\mu} \quad \text{für } \lambda \in \Sigma_{\pi/2+\theta_0}$$

Da wir, wie in der Einleitung des Kapitels erwähnt, von einer geschlossenen Formel der Funktion  $K$  ausgehen, stellen wir den Faltungskern  $k$  als deren inverse Laplace-Transformierte dar, das heißt, wir betrachten die Darstellung

$$k(t) = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} e^{\lambda t} K(\lambda) d\lambda \quad \text{für } t > 0 \quad \text{und } \sigma > \sigma_0, \quad (1.3)$$

deren Eindeutigkeit wir annehmen, siehe hierzu Bemerkung 1.3. Der Integrationsweg  $\sigma + i\mathbb{R}$  ist als *Bromwich*-Linie bekannt und liegt in der Konvergenzhalbebene  $\text{Re}(\lambda) > \sigma_0$ .

Setzen wir (1.3) in das Faltungsintegral  $u(t) = (k * g)(t)$  ein, erhalten wir durch die Vertauschung der Integrale

$$u(t) = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} K(\lambda) y_\lambda(t) d\lambda \quad \text{mit } y_\lambda(t) := \int_0^t e^{\lambda(t-\tau)} g(\tau) d\tau, \quad (1.4)$$

wobei  $y_\lambda(t)$  nach [55, Thm. 2.4] die Lösung des inhomogenen Anfangswertproblems

$$y'(t) = \lambda y(t) + g(t), \quad y(0) = 0 \quad (1.5)$$

beschreibt. Die Approximation der Funktion  $y_\lambda$  ist hierbei die zentrale Frage für die Herleitung der beiden verschiedenen Faltungsquadraturen, die wir in diesem sowie im nächsten Kapitel betrachten.

**Bemerkung 1.3.** Falls  $\mu > 1$  gilt, können wir direkt aus [15, Satz 28.2] folgern, dass die Laplace-Transformation eindeutig umkehrbar ist. Im Fall  $0 < \mu \leq 1$  können wir dagegen nicht als garantiert voraussetzen, dass uns das Integral (1.3) wieder  $k$  zurückliefert. Stattdessen bedienen wir uns der Überlegung in [3], in der  $\tilde{K}(\lambda) := \frac{K(\lambda)}{\lambda}$  betrachtet und  $u(t) := \frac{\partial}{\partial t} (\tilde{k} * g)(t)$  für die Transformation  $\tilde{k}(t) = \mathcal{L}^{-1}\{\tilde{K}(\lambda); t\}$  gesetzt. Für dieses  $u$  lässt sich wiederum zeigen, dass es in der Form (1.4) geschrieben werden kann.

Der Einfachheit halber nehmen wir in den folgenden theoretischen Betrachtungen an, dass die Laplace-Transformation für jedes  $\mu > 0$  eindeutig invertierbar ist.

### 1.2.2 Eigenschaften der sektoriellen Laplace-Transformierten

Bevor wir die exponentielle Quadratur herleiten, gehen wir zunächst genauer auf die Sektorialitätsannahme 1.2 ein. Diese ermöglicht uns den Integrationsweg in (1.3) zu deformieren.

**Satz 1.4.** *Erfüllt  $K$  die Sektorialitätsannahme 1.2 in  $\Sigma_{\pi/2+\theta_0}$ , so gilt für  $t > 0$*

$$\mathcal{L}^{-1}\{K(\lambda); t\} = \frac{1}{2\pi i} \int_{\Gamma} e^{\lambda t} K(\lambda) d\lambda$$

für jede Kurve  $\Gamma$  in  $\Sigma_{\pi/2+\theta}$  mit  $\theta < \theta_0$ , die in einem spitzen Winkel zur negativen reellen Achse gegen unendlich läuft und gegen den Uhrzeigersinn, das heißt von unten nach oben, orientiert ist.

*Beweis.* Der Beweis verwendet Ideen aus [17] und [61]. Es sei  $r > 0$ . Bezeichne mit  $B(r)$  die Teilmenge der Bromwich-Linie  $B$  aus (1.3), die von  $\sigma - ir$  bis  $\sigma + ir$  läuft. Ohne Einschränkung schneide  $\Gamma$  die reelle Achse in  $\tilde{\sigma}$  mit  $0 < \tilde{\sigma} < \sigma$ . Betrachte den Kreisbogen  $K_1(r) := re^{i\varphi}$  für

$$0 < \varphi_0(r) \leq \varphi \leq \varphi_1(r) < \pi,$$

wobei wir durch die Wahl  $\cos(\varphi_0(r)) = \frac{\sigma}{r}$  garantieren, dass  $K_1(r)$  die Bromwich-Linie  $B$  genau im Punkt  $\sigma + ir$  schneidet. Den Winkel  $\varphi_1(r)$  erhalten wir aus dem Schnittpunkt von  $K_1(r)$  mit der Kurve  $\Gamma$ , müssen diesen aber nicht genauer bestimmen. Hier genügt uns die Abschätzung  $\frac{\pi}{2} < \varphi_1(r) < \pi$ .

Analog definieren wir den Kreisbogen  $K_2(r)$  auf der unteren Halbebene, der die Bromwich-Linie im Punkt  $\sigma - ir$  schneidet.

Bezeichne nun mit  $\Gamma(r)$  denjenigen Teil von  $\Gamma$ , der auf der oberen Halbebene durch den Schnittpunkt mit  $K_1(r)$  und auf der unteren Halbebene durch den mit  $K_2(r)$  begrenzt wird. Definiere die Kurve  $\Upsilon(r) := B(r) \cup K_1(r) \cup \Gamma_-(r) \cup K_2(r)$ , wobei  $\Gamma_-(r)$  bedeutet, dass  $\Gamma(r)$  von oben nach unten abgelaufen wird. Wir erhalten somit eine stückweise stetige und geschlossene Kurve  $\Upsilon(r)$ , die vollständig im Holomorphiegebiet von  $K$  liegt und keine Singularitäten dieser Funktion umschließt. Nach dem Cauchyschen Integralsatz gilt daher

$$0 = \int_{B(r)} e^{\lambda t} K(\lambda) d\lambda + \int_{K_1(r)} e^{\lambda t} K(\lambda) d\lambda - \int_{\Gamma(r)} e^{\lambda t} K(\lambda) d\lambda + \int_{K_2(r)} e^{\lambda t} K(\lambda) d\lambda.$$

Wir betrachten nachfolgend das Integral über  $K_1(r)$  genauer. Mit der Abschätzung von  $K(\lambda)$  für  $\lambda \in \Sigma_{\pi/2+\theta}$  folgt

$$\begin{aligned} \left\| \int_{K_1(r)} e^{\lambda t} K(\lambda) d\lambda \right\|_{\mathcal{L}(X)} &\leq C \int_{K_1(r)} e^{\operatorname{Re}(\lambda)t} \frac{|d\lambda|}{|\lambda|^\mu} = C \int_{\varphi_0(r)}^{\varphi_1(r)} e^{\operatorname{Re}(re^{i\varphi})t} \frac{|d(re^{i\varphi})|}{|re^{i\varphi}|^\mu} \\ &\leq Cr^{1-\mu} \left( \int_{\varphi_0(r)}^{\pi/2} e^{rt \cos(\varphi)} d\varphi + \int_{\pi/2}^{\pi} e^{rt \cos(\varphi)} d\varphi \right). \end{aligned} \quad (*)$$

Mit der Substitution  $\varphi = \frac{\pi}{2} - \psi$  erhalten wir

$$\int_{\varphi_0(r)}^{\pi/2} e^{rt \cos(\varphi)} d\varphi = \int_0^{\psi_0(r)} e^{rt \sin(\psi)} d\psi \quad \text{mit} \quad \psi_0(r) = \frac{\pi}{2} - \varphi_0(r) = \arcsin\left(\frac{\sigma}{r}\right),$$

was wir wegen  $\sin(\psi) \leq \sin(\psi_0(r)) = \cos(\varphi_0(r)) = \frac{\sigma}{r}$  weiter durch

$$\int_0^{\psi_0(r)} e^{rt \frac{\sigma}{r}} d\psi = e^{t\sigma} \psi_0(r) = e^{t\sigma} \arcsin\left(\frac{\sigma}{r}\right)$$

abschätzen können. Hierbei gilt für große  $r$  die Näherung  $\arcsin\left(\frac{\sigma}{r}\right) \approx \frac{\sigma}{r}$ .

Für das zweite Integral verwenden wir die Substitution  $\varphi = \frac{\pi}{2} + \psi$  und die Abschätzung  $\frac{2}{\pi}\psi \leq \sin(\psi)$  für  $0 \leq \psi \leq \frac{\pi}{2}$  und erhalten

$$\int_{\pi/2}^{\pi} e^{rt \cos(\varphi)} d\varphi = \int_0^{\pi/2} e^{-rt \sin(\psi)} d\psi \leq \int_0^{\pi/2} e^{-rt \frac{2}{\pi}\psi} d\psi = \frac{\pi}{2rt} (1 - e^{-rt}).$$

Insgesamt können wir (\*) also für große  $r$  nach oben abschätzen durch

$$\tilde{C} r^{1-\mu} (r^{-1} + r^{-1} (1 - e^{-rt})) = \tilde{C} r^{-\mu} (2 - e^{-rt}),$$

wobei  $\tilde{C} > 0$  von  $\sigma$  und  $t$  abhängt. Eine analoge Abschätzung für das Integral über  $K_2(r)$  lässt sich ebenfalls zeigen. Da der Ausdruck für  $r \rightarrow \infty$  gegen 0 geht, folgt wie behauptet die Gleichheit

$$\frac{1}{2\pi i} \int_B e^{\lambda t} K(\lambda) d\lambda = \frac{1}{2\pi i} \int_{\Gamma} e^{\lambda t} K(\lambda) d\lambda.$$

□

Durch die Wahl eines geeigneten Integrationsweges lässt sich die folgende Aussage über den Faltungskern  $k$  treffen, die wir später benötigen.

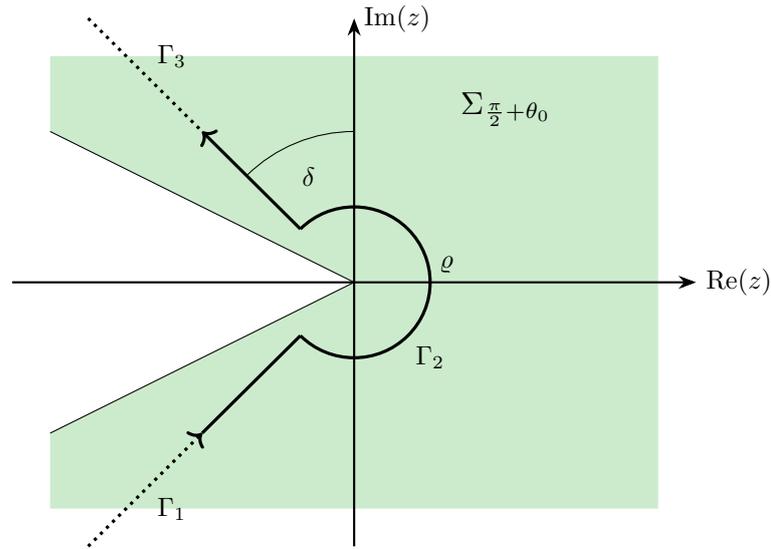
**Lemma 1.5.** *Es gelte für  $K$  die Annahme 1.2 im Sektor  $\Sigma_{\pi/2+\theta_0}$  und  $\Gamma$  erfülle die Voraussetzungen aus Satz 1.4. Dann existiert eine Konstante  $C > 0$ , so dass die Abschätzung*

$$\|k(t)\|_{\mathcal{L}(X)} \leq C t^{\mu-1} \quad \text{für} \quad t > 0$$

erfüllt ist.

*Beweis.* Vergleiche hierzu [9, Thm. 4.6] und [16, Prop. IV.4.3]. Es seien  $t > 0$ ,  $\varrho > 0$  und  $\delta \in (\theta, \theta_0)$  für  $\theta < \theta_0$ . Definiere die Kurve  $\Gamma := \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$ , deren Teilstücke durch

$$\begin{aligned} \Gamma_1 &:= \left\{ r e^{-i(\pi/2+\delta)} \mid \varrho \leq r < \infty \right\}, \\ \Gamma_2 &:= \left\{ \varrho e^{i\varphi} \mid |\varphi| \leq \pi/2 + \delta \right\} \quad \text{und} \\ \Gamma_3 &:= \left\{ r e^{i(\pi/2+\delta)} \mid \varrho \leq r < \infty \right\} \end{aligned}$$


 Abbildung 1.1: Die Kurve  $\Gamma$ , zusammengesetzt aus den Teilstücken  $\Gamma_1$ ,  $\Gamma_2$  und  $\Gamma_3$ .

gegeben sind und die von unten nach oben durchlaufen wird, siehe dazu Abbildung 1.1. Offensichtlich erfüllt  $\Gamma$  die Voraussetzungen aus Satz 1.4.

Aus der Sektorialitätsannahme 1.2 über  $K$  erhalten wir

$$\left\| \int_{\Gamma} e^{\lambda t} K(\lambda) d\lambda \right\|_{\mathcal{L}(X)} \leq C \int_{\Gamma} e^{t \operatorname{Re}(\lambda)} \frac{1}{|\lambda|^{\mu}} |d\lambda|.$$

Wir berechnen nun die Integrale über die Teilstücke  $\Gamma_1$ ,  $\Gamma_2$  und  $\Gamma_3$  einzeln. Sei dazu  $\rho = 1/t$  und  $a = \sin(\delta - \theta) \in (0, 1)$ . Für das Integral über  $\Gamma_1$  folgt damit

$$\begin{aligned} \int_{\Gamma_1} e^{t \operatorname{Re}(\lambda)} \frac{1}{|\lambda|^{\mu}} |d\lambda| &= \int_{\rho}^{\infty} e^{t \operatorname{Re}(re^{-i(\pi/2+\delta)})} \frac{|d(re^{-i(\pi/2+\delta)})|}{|re^{-i(\pi/2+\delta)}|^{\mu}} \\ &= \int_{\rho}^{\infty} e^{rt \cos(\pi/2+\delta)} \frac{1}{r^{\mu}} dr \\ &= \int_{\rho}^{\infty} e^{-r/e \sin(\delta)} \frac{1}{r^{\mu}} dr \\ &\stackrel{\text{subst.}}{=} \int_1^{\infty} e^{-\nu \sin(\delta)} \frac{1}{\nu^{\mu}} \underbrace{\rho^{1-\mu}}_{=t^{\mu-1}} d\nu. \end{aligned}$$

Analog bekommen wir für das Integral über  $\Gamma_3$

$$\int_{\Gamma_3} e^{t \operatorname{Re}(\lambda)} \frac{1}{|\lambda|^{\mu}} |d\lambda| = t^{\mu-1} \int_1^{\infty} e^{-\nu \sin(\delta)} \frac{1}{\nu^{\mu}} d\nu.$$

Da  $0 < \theta < \delta < \pi/2$  und damit  $0 < \delta - \theta < \delta$  folgt  $\sin(\delta) > \sin(\delta - \theta) = a$  und es ergibt sich

$$\int_1^{\infty} e^{-\nu \sin(\delta)} \frac{1}{\nu^{\mu}} d\nu < \int_1^{\infty} e^{-\nu a} \frac{1}{\nu^{\mu}} d\nu < \infty, \quad \text{da } a \in (0, 1) \quad \text{und } \mu > 0.$$

Für das Integral über das Teilstück  $\Gamma_2$  gilt

$$\begin{aligned} \int_{\Gamma_2} e^{t\operatorname{Re}(\lambda)} \frac{1}{|\lambda|^\mu} |d\lambda| &= \int_{-\pi/2-\delta}^{\pi/2+\delta} e^{t\operatorname{Re}(\varrho e^{i\varphi})} \frac{|d(\varrho e^{i\varphi})|}{|\varrho e^{i\varphi}|^\mu} \\ &= \int_{-\pi/2-\delta}^{\pi/2+\delta} e^{\varrho t \cos(\varphi)} \frac{\varrho}{\varrho^\mu} d\varphi \\ &= t^{\mu-1} \int_{-\pi/2-\delta}^{\pi/2+\delta} e^{\cos(\varphi)} d\varphi < t^{\mu-1} \int_{-\pi}^{\pi} e d\varphi. \end{aligned}$$

Insgesamt folgt die behauptete Abschätzung der inversen Laplace-Transformation von  $K$  aus

$$\left\| \int_{\Gamma} e^{\lambda t} K(\lambda) d\lambda \right\|_{\mathcal{L}(X)} \leq 2C t^{\mu-1} \left( \int_1^\infty e^{-\nu a} \frac{d\nu}{\nu^\mu} + \int_0^\pi e d\varphi \right) \leq \tilde{C} t^{\mu-1}.$$

□

Mit Hilfe der beiden vorigen Sätze lässt sich die folgende Aussage zeigen, die bei der numerischen Berechnung der inversen Laplace-Transformation von Bedeutung sein wird.

**Satz 1.6.** *Gelten für  $K$  die Sektorialitätsannahme 1.2 in  $\Sigma_{\pi/2+\theta_0}$  und die Kurve  $\Gamma$  erfülle die Annahmen aus Satz 1.4. Dann folgt*

$$\int_{\Gamma} \frac{1}{\lambda^{j+1}} K(\lambda) d\lambda = 0 \quad \text{für } j \in \mathbb{N}_0.$$

*Beweis.* Nach Voraussetzung ist  $K(\lambda) = \mathcal{L}\{k(t); \lambda\}$  in der Konvergenzhalbebene  $\operatorname{Re}(\lambda) > \sigma_0 > 0$  und es gilt mit Lemma 1.5 für jedes  $T > 0$

$$\int_0^T \|k(t)\|_{\mathcal{L}(X)} dt \leq C \int_0^T t^{\mu-1} dt = \frac{C}{\mu} T^\mu < \infty.$$

Damit folgt aus [15, Sätze 24.4 und 27.1] für  $j = 0$  und  $t \geq 0$  sofort die Aussage

$$\frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} e^{\lambda t} \frac{K(\lambda)}{\lambda} d\lambda = \int_0^t k(\tau) d\tau = (k * 1)(t) \quad \text{für } \sigma > \sigma_0.$$

Insbesondere ist hierbei die Faltung  $(k * 1)(t)$  nach [15, Satz 10.3] im Nullpunkt stetig, also gilt  $(k * 1)(0) = 0$ .

Für den Fall  $j \in \mathbb{N}$  setzen wir  $f_j(t) := \frac{1}{(j-1)!} t^{j-1}$  mit  $t > 0$ . Aus dem Faltungssatz, siehe beispielsweise [15, Satz 10.1], erhalten wir damit in der Konvergenzhalbebene  $\operatorname{Re}(\lambda) > \sigma_0 > 0$  die Gleichheit

$$\mathcal{L}\{(k * f_j)(t); \lambda\} = K(\lambda) F_j(\lambda) \quad \text{für } t > 0,$$

wobei  $F_j(\lambda) := \lambda^{-j} = \mathcal{L}\{f_j(t); \lambda\}$ . Wegen  $\|K(\lambda)\lambda^{-j}\|_{\mathcal{L}(X)} \leq C |\lambda|^{-\mu-j}$  für  $\lambda \in \Sigma_{\pi/2+\theta_0}$ , erfüllt die Funktion  $K(\lambda) F_j(\lambda)$  die Sektorialitätsannahme 1.2 und es gilt nach Lemma 1.5

für  $T > 0$ :

$$\int_0^T \|(k * f_j)(t)\|_{\mathcal{L}(X)} dt \leq C \int_0^T t^{\mu+j-1} dt \leq \frac{C}{\mu+j} T^{\mu+j} < \infty$$

Mit [15, Satz 27.1] erhalten wir daraus für  $j \in \mathbb{N}$  und  $t \geq 0$  den Zusammenhang

$$\frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} e^{\lambda t} \frac{1}{\lambda} \cdot \frac{K(\lambda)}{\lambda^j} d\lambda = \int_0^t (k * f_j)(t) dt = ((k * f_j) * 1)(t),$$

wobei erneut mit [15, Satz 10.3] die Aussage  $((k * f_j) * 1)(0) = 0$  folgt.

Es bleibt noch zu zeigen, dass die Deformation der Bromwich-Linie in die Kurve  $\Gamma$  auch im Fall  $t = 0$  zulässig ist: Dazu erinnern wir uns an den Kreisbogen  $K_1(r)$  für  $r > 0$  aus dem Beweis zu Satz 1.4 und betrachten  $j \in \mathbb{N}_0$ . Aus der Abschätzung  $\|K(\lambda)\lambda^{-j-1}\|_{\mathcal{L}(X)} \leq C|\lambda|^{-\mu-j-1}$  für  $\lambda \in \Sigma_{\pi/2+\theta_0}$  ergibt sich

$$\begin{aligned} \left\| \int_{K_1(r)} \frac{K(\lambda)}{\lambda^{j+1}} d\lambda \right\|_{\mathcal{L}(X)} &\leq C \int_{K_1(r)} \frac{1}{|\lambda|^{\mu+j+1}} |d\lambda| = C \int_{\varphi_0(r)}^{\varphi_1(r)} \frac{1}{|re^{i\varphi}|^{\mu+j+1}} |d(re^{i\varphi})| \\ &= C \int_{\varphi_0(r)}^{\varphi_1(r)} \frac{1}{r^{\mu+j}} d\varphi \leq Cr^{-\mu-j} \int_0^\pi d\varphi \\ &= C\pi r^{-\mu-j} \xrightarrow{r \rightarrow \infty} 0. \end{aligned}$$

Die Aussage folgt daher mit der Begründung aus dem Beweis von Satz 1.4.  $\square$

### 1.2.3 Faltungsquadratur durch Polynominterpolation

Damit können wir nun als Ausgangspunkt der Herleitung der exponentiellen Quadratur das Faltungsintegral  $u(t) = (k * g)(t)$  in Analogie zu (1.4) schreiben als

$$u(t) = \frac{1}{2\pi i} \int_{\Gamma} K(\lambda) y_{\lambda}(t) d\lambda \quad \text{mit} \quad y_{\lambda}(t) = \int_0^t e^{\lambda(t-\tau)} g(\tau) d\tau, \quad (1.6)$$

wobei die Kurve  $\Gamma$  die Annahmen aus Satz 1.4 erfüllt. Die Funktion  $y_{\lambda}$  approximieren wir mit Hilfe der stückweisen Polynominterpolation von  $g$ , siehe dazu (1.2). Wir erhalten damit zum Zeitpunkt  $t_n = nh$  die Approximation  $(y_{\lambda})_n$  von  $y_{\lambda}(t_n)$  durch

$$\begin{aligned} y_{\lambda}(t_n) &= \int_0^{t_n} e^{\lambda(t_n-\tau)} g(\tau) d\tau = \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} e^{\lambda(t_n-\tau)} g(\tau) d\tau \\ &= \sum_{j=0}^{n-1} \int_0^1 e^{\lambda(n-j-\vartheta)h} g(t_j + \vartheta h) h d\vartheta \\ &\approx h \sum_{j=0}^{n-1} e^{\lambda(n-j-1)h} \int_0^1 e^{\lambda(1-\vartheta)h} p_j(t_j + \vartheta h) d\vartheta \\ &= h \sum_{j=0}^{n-1} e^{\lambda(n-j-1)h} \int_0^1 e^{\lambda(1-\vartheta)h} \sum_{i=0}^s \ell_i(\vartheta) g(t_j + c_i h) d\vartheta \end{aligned}$$

$$= h \sum_{j=0}^{n-1} e^{\lambda(n-j-1)h} \sum_{i=0}^s b_i(h\lambda) g(t_j + c_i h) =: (y_\lambda)_n, \quad (1.7)$$

wobei die Funktionen

$$b_i(z) := \int_0^1 e^{z(1-\vartheta)} \ell_i(\vartheta) d\vartheta \quad \text{für } i = 1, \dots, s \quad (1.8)$$

als Linearkombinationen der  $\varphi$ -Funktionen

$$\varphi_k(z) := \int_0^1 e^{z(1-\vartheta)} \frac{\vartheta^{k-1}}{(k-1)!} d\vartheta, \quad k \geq 1$$

geschrieben werden können. Diese treten im Zusammenhang mit Verfahren aus der Klasse der exponentiellen Integratoren sehr häufig auf und lassen sich mit Hilfe der Rekursion

$$\varphi_{k+1}(z) = \frac{\varphi_k(z) - 1/k!}{z} \quad \text{mit } \varphi_0(z) = e^z$$

charakterisieren, vergleiche dazu [36].

**Lemma 1.7.** Für jedes  $h > 0$  und jedes  $\lambda \in \Gamma$ , wobei  $\Gamma$  die Voraussetzungen aus Satz 1.4 erfüllt, ist der Betrag der  $\varphi$ -Funktion  $\varphi_k(h\lambda)$ ,  $k \in \mathbb{N}$ , beschränkt.

*Beweis.* Da für  $\lambda \in \Gamma$  insbesondere  $\operatorname{Re}(\lambda) \leq c$  für eine Konstante  $c > 0$  gilt, folgt  $|e^{h\lambda}| \leq e^{hc}$ . Die Aussage ergibt sich damit direkt aus der Abschätzung

$$|\varphi_k(h\lambda)| \leq \int_0^1 |e^{h\lambda(1-\vartheta)}| \frac{\vartheta^{k-1}}{(k-1)!} d\vartheta.$$

□

Die Formulierung der exponentiellen Quadratur, die eine Verallgemeinerung der „exponentiellen Runge-Kutta Methode vom Kollokationstyp“ aus [36, 37] darstellt, lautet damit wie folgt:

**Satz 1.8.** Es seien die  $s$  paarweise verschiedenen Knoten  $0 \leq c_1 < \dots < c_s \leq 1$  gegeben und es erfülle die Laplace-Transformierte  $K(\lambda) := \mathcal{L}\{k(t); \lambda\}$  die Annahme 1.2. Dann erhalten wir zum Zeitpunkt  $t_n = nh$  mit Schrittweite  $h > 0$  eine Approximation von  $u(t) = (k * g)(t)$  als

$$u_n = h \sum_{j=0}^{n-1} \sum_{i=1}^s w_{i, n-1-j} g(t_j + c_i h),$$

wobei das Faltungsgewicht  $w_{i,j}$  gegeben ist durch

$$w_{i,j} = \frac{1}{2\pi i} \int_{\Gamma} K(\lambda) e^{\lambda(jh)} b_i(h\lambda) d\lambda \quad (1.9)$$

mit einer Kurve  $\Gamma$ , die die Voraussetzungen aus Satz 1.4 erfüllt.

*Beweis.* Die Approximation  $u_n$  von  $u(t_n)$  ergibt sich sofort durch das Einsetzen von  $y_\lambda(t_n)$  aus (1.7) in (1.6).  $\square$

Mit Hilfe der Fehlerabschätzung der Lagrange-Polynominterpolation und der Abschätzung des Faltungskerns  $k$  können wir die Konvergenz der exponentiellen Quadratur zeigen.

**Satz 1.9.** *Erfülle die Laplace-Transformierte  $K$  die Annahme 1.2 und seien  $s$  paarweise verschiedene Knoten  $c_i \in [0, 1]$  gegeben. Ist  $g \in C^s([0, T], X)$ , dann ist der Fehler der exponentiellen Quadratur aus Satz 1.8 für  $t_n = nh \leq T$  beschränkt durch*

$$\|u_n - u(t_n)\|_X \leq C \|\omega\|_\infty h^s \max_{\xi \in [t_0, t_n]} \|g^{(s)}(\xi)\|_X,$$

wobei  $\|\omega\|_\infty = \max_{x \in [0, 1]} |\omega(x)|$  und die Konstante  $C$  nur von  $T$ , der Anzahl der Interpolationspunkte und den Parametern in Annahme 1.2 abhängt, jedoch nicht von  $n$  und  $h$ .

*Beweis.* Für den Beweis greifen wir nicht auf die Darstellung von  $u$  durch die inverse Laplace-Transformation (1.6) zurück, sondern stellen  $u(t_n)$  und die Approximation  $u_n$  direkt als Faltung mit dem Faltungskern  $k$  dar, das heißt, wir betrachten

$$\begin{aligned} u(t_n) &= \int_0^{t_n} k(t_n - \tau h) g(\tau) d\tau = h \sum_{j=0}^{n-1} \int_0^1 k(t_n - t_j - \vartheta) g(t_j + \vartheta h) d\vartheta \quad \text{und} \\ u_n &= h \sum_{j=0}^{n-1} \int_0^1 k(t_n - t_j - \vartheta h) p_j(t_j + \vartheta h) d\vartheta. \end{aligned}$$

Damit erhalten wir unter Verwendung von Satz 1.1 und Lemma 1.5 die Abschätzung

$$\begin{aligned} \|u_n - u(t_n)\|_X &\leq h \sum_{j=0}^{n-1} \int_0^1 \|k(t_n - t_j - \vartheta h)(p_j(t_j + \vartheta h) - g(t_j + \vartheta h))\|_X d\vartheta \\ &\leq h \sum_{j=0}^{n-1} \int_0^1 \|k(t_n - t_j - \vartheta h)\|_{\mathcal{L}(X)} \|p_j(t_j + \vartheta h) - g(t_j + \vartheta h)\|_X d\vartheta \\ &\leq \frac{h^{s+1}}{s!} \sum_{j=0}^{n-1} \int_0^1 C(t_n - t_j - \vartheta h)^{\mu-1} |\omega(\vartheta)| \max_{\xi_j \in [t_j, t_{j+1}]} \|g^{(s)}(\xi_j)\|_X d\vartheta \\ &\leq C \frac{h^{s+1}}{s!} \max_{\zeta \in [0, 1]} |\omega(\zeta)| \max_{\xi \in [t_0, t_n]} \|g^{(s)}(\xi)\|_X \sum_{j=0}^{n-1} \int_0^1 (t_n - t_j - \vartheta h)^{\mu-1} d\vartheta. \end{aligned}$$

Für die verbleibende Summe gilt:

$$h \sum_{j=0}^{n-1} \int_0^1 (t_n - t_j - \vartheta h)^{\mu-1} d\vartheta = \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} (t_n - \tau)^{\mu-1} d\tau = \int_0^{t_n} (t_n - \tau)^{\mu-1} d\tau = \frac{1}{\mu} t_n^\mu$$

Mit  $nh \leq T$  ergibt sich so die behauptete Abschätzung

$$\|u_n - u(t_n)\|_X \leq C \frac{T^\mu}{\mu} \frac{h^s}{s!} \max_{\zeta \in [0,1]} |\omega(\zeta)| \max_{\xi \in [t_0, t_n]} \|g^{(s)}(\xi)\|_X.$$

□

**Bemerkung 1.10.** Unter der Annahme, dass der Faltungskern über die verallgemeinerte Mittag-Leffler Funktion, siehe (1.19), definiert ist, das heißt  $k(t) := \mathbf{e}_{\mu,\mu}(t; a)$  für  $\mu < 1$  und  $a \in \mathbb{R}$  gilt, wird in [28] eine detailliertere Konvergenzaussage gezeigt, die auch die Konstanten vor den gebrochenen Ordnungen  $h^{s+\nu\mu}$ ,  $\nu \in \mathbb{N}$ , enthält. In [26] wird die Aussage auf den Fall, dass  $a$  eine reelle Matrix ist, deren Eigenwerte große negative Realteile haben können, erweitert. Die hierbei von den Autoren hergeleiteten Konstanten hängen nur von den Knoten der Polynominterpolation ab, so dass sich Bedingungen herleiten lassen, mit denen höhere Ordnungen als  $\mathcal{O}(h^s)$  erreicht werden können. Im numerischen Experiment in Abschnitt 1.5 wird deutlich, dass die Konvergenzaussage für größere Werte von  $a$  jedoch nicht mehr zutrifft, was die Vermutung nahelegt, dass bestimmte Voraussetzungen für den Beweis vergessen oder nicht beachtet wurden. Die fehlenden Voraussetzungen werden in dieser Arbeit nicht weiter thematisiert.

Für die Untersuchung der Stabilität ist die Abschätzung der Faltungsgewichte (1.9) wichtig, die wir ebenfalls mit Hilfe der Überlegungen zum Faltungskern erhalten. Dafür benötigen wir das folgende Lemma:

**Lemma 1.11.** Für  $j \in \mathbb{N}_0$  und  $0 < \mu < 1$  gilt die Abschätzung

$$(j+1)^\mu - j^\mu \leq (j+1)^{\mu-1}.$$

*Beweis.* Für  $j = 0$  ist die Aussage klar. Sei also  $j \geq 1$ , dann erhalten wir mit Hilfe des binomischen Lehrsatzes

$$\begin{aligned} (j+1)^\mu - j^\mu &= (j+1)^\mu - (j+1)^\mu \left(1 - \frac{1}{j+1}\right)^\mu \\ &= (j+1)^\mu \left(1 - \sum_{k=0}^{\infty} \binom{\mu}{k} \left(\frac{-1}{j+1}\right)^k\right) \\ &= (j+1)^\mu \left(1 - \left(1 - \mu \frac{1}{j+1} + \sum_{k=2}^{\infty} \binom{\mu}{k} \left(\frac{-1}{j+1}\right)^k\right)\right) \\ &= (j+1)^{\mu-1} \left(\mu + \sum_{k=2}^{\infty} \binom{\mu}{k} \left(\frac{-1}{j+1}\right)^{k-1}\right). \end{aligned}$$

Hierbei gilt für  $k \geq 2$  insbesondere die Abschätzung

$$0 \leq (-1)^{k-1} \binom{\mu}{k} = (-1)^{k-1} \frac{\mu}{1} \cdot \frac{\mu-1}{2} \cdot \dots \cdot \frac{\mu-k+1}{k}$$

$$= \underbrace{\frac{\mu}{1}}_{<1} \cdot \underbrace{\frac{1-\mu}{2}}_{<1-\mu} \cdot \dots \cdot \underbrace{\frac{k-1-\mu}{k}}_{<1} < 1 - \mu$$

und damit unter Verwendung der geometrischen Reihe

$$\sum_{k=2}^{\infty} \binom{\mu}{k} \left(\frac{-1}{j+1}\right)^{k-1} \leq \sum_{k=2}^{\infty} \frac{1-\mu}{(j+1)^{k-1}} = (1-\mu) \left( \sum_{k=0}^{\infty} \frac{1}{(j+1)^k} - 1 \right) = \frac{1-\mu}{j} \leq 1 - \mu.$$

In die obige Gleichung eingesetzt, ergibt sich daraus

$$(j+1)^\mu - j^\mu \leq (j+1)^{\mu-1} (\mu + 1 - \mu) = (j+1)^{\mu-1}.$$

□

**Satz 1.12.** *Gilt für die Laplace-Transformierte  $K$  die Annahme 1.2, dann erfüllt das Faltungsgewicht  $w_{i,j}$  für  $j \in \mathbb{N}_0$  und  $i = 1, \dots, s$  aus Satz 1.8 die Abschätzung*

$$\|w_{i,j}\|_{\mathcal{L}(X)} \leq C \max_{\vartheta \in [0,1]} |\ell_i(\vartheta)| (t_{j+1})^{\mu-1},$$

wobei die Konstante  $C$  nur von den Parametern in Annahme 1.2 abhängt.

*Beweis.* Nehmen wir an, dass die Kurve  $\Gamma$  die Voraussetzungen aus Satz 1.4 erfüllt, so können wir das Faltungsgewicht nach Satz 1.8 schreiben als

$$\begin{aligned} w_{i,j} &= \frac{1}{2\pi i} \int_{\Gamma} K(\lambda) e^{\lambda(jh)} \int_0^1 e^{\lambda h(1-\vartheta)} \ell_i(\vartheta) d\vartheta d\lambda \\ &= \int_0^1 \frac{1}{2\pi i} \int_{\Gamma} K(\lambda) e^{\lambda h(j+1-\vartheta)} d\lambda \ell_i(\vartheta) d\vartheta \\ &= \int_0^1 k((j+1-\vartheta)h) \ell_i(\vartheta) d\vartheta. \end{aligned}$$

Mit Lemma 1.5 folgt daraus

$$\begin{aligned} \|w_{i,j}\|_{\mathcal{L}(X)} &\leq \int_0^1 \|k((j+1-\vartheta)h)\|_{\mathcal{L}(X)} |\ell_i(\vartheta)| d\vartheta \\ &\leq C \max_{\xi \in [0,1]} |\ell_i(\xi)| \int_0^1 ((j+1-\vartheta)h)^{\mu-1} d\vartheta. \end{aligned}$$

Wir erhalten für  $\mu \geq 1$

$$\int_0^1 ((j+1-\vartheta)h)^{\mu-1} d\vartheta \leq \max_{\vartheta \in [0,1]} ((j+1-\vartheta)h)^{\mu-1} = (t_{j+1})^{\mu-1}$$

und für  $\mu < 1$  mit Lemma 1.11

$$\int_0^1 ((j+1-\vartheta)h)^{\mu-1} d\vartheta = \frac{h^{\mu-1}}{\mu} ((j+1)^\mu - j^\mu) \leq \frac{1}{\mu} (t_{j+1})^{\mu-1}.$$

Damit folgt die behauptete Abschätzung

$$w_{i,j} \leq \tilde{C}(\mu) \max_{\xi \in [0,1]} |\ell_i(\xi)| (t_{j+1})^{\mu-1},$$

wobei für die Konstante  $\tilde{C}(\mu) = C \max(1, \mu^{-1})$  gilt.  $\square$

**Bemerkung 1.13.** Wenden wir die exponentielle Runge-Kutta Methode vom Kollokationstyp, die in [37] formuliert ist, direkt auf das Anfangswertproblem (1.5) an, erhalten wir die rekursive Darstellung

$$(y_\lambda)_n = e^{h\lambda} (y_\lambda)_{n-1} + h \sum_{i=1}^s b_i(h\lambda) g(t_{n-1} + c_i h) \quad \text{für } n \in \mathbb{N}.$$

Es lässt sich leicht nachrechnen, dass diese Rekursionsgleichung unter Verwendung von  $y_0 = 0$  äquivalent zur Darstellung (1.7) ist. Daher ließe sich das Verfahren aus Satz 1.8 auch herleiten, indem obige Darstellung zur Approximation von  $y_\lambda(t_n)$  in (1.6) eingesetzt würde. Die Motivation für unser Vorgehen, stattdessen (1.7) zu verwenden, liegt darin, dass wir so direkt die angegebene Darstellung der Quadratur aus Satz 1.8 erhalten.

### 1.3 Approximation der Faltungsgewichte

Um der Frage nachzugehen, wie die Faltungsgewichte  $w_{i,j}$  der exponentiellen Quadratur aus Satz 1.8 geeignet approximiert werden können und welche Fehler wir dadurch erhalten, folgen wir den Überlegungen in [43, 44] und [66]. Der Gedanke, der in den einzelnen Quellen verfolgt wird, ist der, die Integrale  $w_{i,j}$  zu parametrisieren und mit Hilfe der Trapezregel zu approximieren. Hierbei stellt sich die Frage, wie ein geeigneter Weg aussieht und welcher Diskretisierungsfehler dabei zu erwarten ist. Die Diskretisierung der Kurvenintegrale ist daher ein wichtiger Bestandteil für einen schnellen Algorithmus.

Da wir weder eine ausreichend gute Approximation erwarten können, wenn die Integrale  $w_{i,j}$  für alle  $j$  und alle  $i = 1, \dots, s$  unter Verwendung derselben Kurve berechnet werden, noch es effizient ist, für jedes Integral eine neue Kurve zu bestimmen, betrachten wir die Folge schnell wachsender Intervalle

$$I_0 := [0, h), \quad I_\ell := [B^{\ell-1}h, B^\ell h) \quad \text{für } \ell \geq 1 \tag{1.10}$$

mit ganzzahliger Basis  $B > 1$  und legen für jedes dieser Intervalle eine Kurve  $\Gamma_\ell$  fest. Ist dann  $jh \in I_\ell$ , integrieren wir (1.9) über  $\Gamma_\ell$ .

Bei der numerischen Berechnung der inversen Laplace-Transformation werden häufig drei verschiedene Wahlen der Integrationskontur untersucht. Während die Analyse der Talbot-Kontur, siehe beispielsweise [63, 65], vergleichsweise kompliziert ist, bietet sich eine Parabel

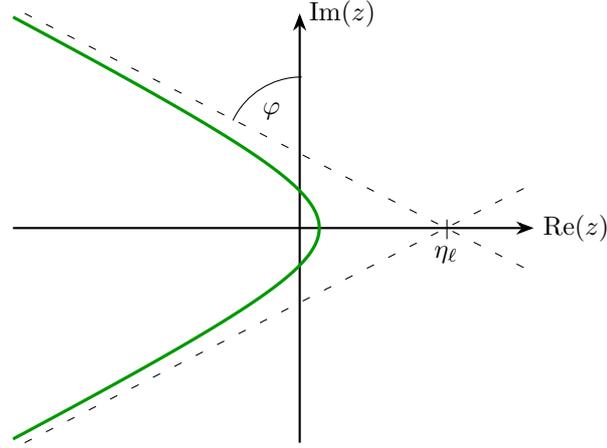


Abbildung 1.2: Darstellung der Hyperbel  $\Gamma_\ell$  und deren Asymptoten mit der Parametrisierung aus (1.11). Die Hyperbel wird hier gegen den Uhrzeigersinn abgelaufen.

im sektoriellen Fall nicht an, vergleiche dazu [66]. Wir wählen daher die Kurve  $\Gamma_\ell$  als Hyperbeln, die wir durch die stetig differenzierbaren Wege

$$\gamma_\ell : \mathbb{R} \rightarrow \Gamma_\ell, \quad x \mapsto \gamma_\ell(x) := \eta_\ell (1 + \sin(ix - \varphi)), \quad \ell \geq 0 \quad (1.11)$$

parametrisieren, siehe dazu Abbildung 1.2. Der Parameter  $\eta_\ell > 0$  hängt hierbei vom jeweiligen Intervall  $I_\ell$  ab und beschreibt die Stelle auf der reellen Achse, in der sich die Asymptoten der Hyperbel schneiden.

Mit  $\varphi$  wird der Winkel zwischen der oberen imaginären Achse und der danebenliegenden Asymptote angegeben. Setzen wir  $0 < \varphi < \theta$  für  $\theta < \theta_0$  voraus, so liegt die Hyperbel  $\Gamma_\ell$  im Sektor  $\Sigma_{\pi/2+\theta}$  und erfüllt damit die Annahme aus Satz 1.4.

Wegen der Annahme an das Abfallverhalten der Laplace-Transformierten  $K$  und der Beschränktheit der Funktionen  $b_i(h\lambda)$  für  $\lambda \in \Gamma$ , siehe Lemma 1.7, verschwindet der Integrand im Kurvenintegral (1.9) für  $\text{Re}(\lambda) \rightarrow -\infty$  schnell, wodurch sich die Diskretisierung mittels der Trapezregel anbietet, vergleiche dazu [44, 66]. Integrieren wir daher das Faltungsgewicht  $w_{i,j}$  für  $jh \in I_\ell$  und  $i = 1, \dots, s$  über die Parametrisierung  $\gamma_\ell$ , erhalten wir durch den Abbruch der Trapezregel nach  $2\kappa + 1$  Schritten die Approximation

$$\begin{aligned} w_{i,j} &= \frac{1}{2\pi i} \int_{-\infty}^{\infty} e^{\gamma_\ell(x)jh} K(\gamma_\ell(x)) b_i(h\gamma_\ell(x)) \gamma_\ell'(x) dx \\ &\approx \sum_{k=-\kappa}^{\kappa} \omega_k^{(\ell)} e^{jh\lambda_k^{(\ell)}} K(\lambda_k^{(\ell)}) b_i(h\lambda_k^{(\ell)}) =: w_{i,j}^{[\kappa]}, \end{aligned} \quad (1.12)$$

wobei die Gewichte  $\omega_k^{(\ell)}$  und Stützstellen  $\lambda_k^{(\ell)}$  gegeben sind durch

$$\omega_k^{(\ell)} = \frac{\tau}{2\pi i} \gamma_\ell'(x_k) \quad \text{und} \quad \lambda_k^{(\ell)} = \gamma_\ell(x_k) \quad (1.13)$$

mit  $x_k = k\tau$  und der Schrittweite  $\tau > 0$ .

Untersuchungen dazu, wie die Parameter der Hyperbeln und die Parameter in der Trapezregel gewählt werden können und welcher Diskretisierungsfehler dabei zu erwarten ist, sind beispielsweise in [29, 43, 66] zu finden. Wir folgen der Überlegung im Paper [44], in dem der Diskretisierungsfehler unter der Annahme, dass Intervalle der Form  $[y, \tilde{B}y]$  mit  $y > 0$  und  $\tilde{B} \geq 1$  betrachtet werden, angegeben wird. Damit bekommen wir für  $jh \in I_\ell$  und  $\ell \in \mathbb{N}$  die spektrale Abschätzung

$$\left\| w_{i,j} - w_{i,j}^{[\kappa]} \right\|_{\mathcal{L}(X)} = \mathcal{O}(e^{-c\kappa}) \quad \text{mit} \quad c = \mathcal{O}(1/\ln(B)). \quad (1.14)$$

Auf die ausführliche Darstellung des Fehlers und deren Beweis gehen wir hierbei nicht genauer ein, sondern verweisen auf das Theorem 1 und die Bemerkungen 1 und 2 in [44], in denen auch die folgende Wahl der Parameter begründet wird:

**Satz 1.14.** *Erfülle die Laplace-Transformierte  $K$  die Annahme 1.2 und seien der Winkel  $\varphi$  aus (1.11) und der Parameter  $d > 0$  so gesetzt, dass*

$$0 < \varphi - d < \varphi + d < \theta_0$$

*gilt. Seien außerdem  $\kappa \geq 1$  und  $\rho \in (0, 1)$ , dann erhalten wir für den Diskretisierungsfehler die Abschätzung (1.14) gleichmäßig für alle  $jh \in I_\ell$  mit  $\ell \in \mathbb{N}$ , wenn die Parameter*

$$\tau = \frac{1}{\kappa} a_s(\rho) \quad \text{und} \quad \eta_\ell = \frac{2\pi d \kappa (1 - \rho)}{B^\ell h a_s(\rho)}$$

*gewählt werden, wobei*

$$a_s(\rho) = \arccos \left( \frac{B}{s(1 - \rho) \sin(\varphi)} \right).$$

*Hierbei setzen wir  $s = 1$  für  $\mu \geq 1$  und  $s \in (0, 1)$  für  $\mu < 1$ .*

In der numerischen Berechnung ist es ausreichend,  $\varphi = d = \theta/2$  für  $\theta < \theta_0$  und  $s \equiv 1$  zu wählen, auch wenn dadurch die Voraussetzungen des Satzes verletzt werden, vergleiche dazu [18, 60]. Eine Strategie, wie der optimale Wert für den Parameter  $\rho$  gefunden werden kann, ist in [44] zu finden und besteht darin, den Ausdruck

$$\mathbf{eps} \epsilon_\kappa(\rho)^{\rho-1} + \epsilon_\kappa(\rho)^\rho \quad \text{mit} \quad \epsilon_\kappa(\rho) = \exp \left( \frac{-2\pi d \kappa}{a_s(\rho)} \right)$$

über  $\rho \in (0, 1)$  zu minimieren, wobei  $\mathbf{eps}$  die Maschinengenauigkeit ist.

**Bemerkung 1.15.** Wie in Satz 1.14 deutlich wird, hängt die Wahl des Winkels  $\varphi$  vom Winkel  $\theta_0$  aus der Sektorialitätsannahme 1.2 ab, jedoch nicht vom betrachteten Intervall  $I_\ell$ , weswegen es plausibel ist, ihn für alle Hyperbeln  $\Gamma_\ell$  gleich zu wählen. Auch die Schrittweite  $\tau$  der äquidistant liegenden  $2\kappa + 1$  Knotenpunkte ist damit unabhängig vom Intervall  $I_\ell$ .

Da es für die Fehlerabschätzung und die Wahl der Parameter aus Satz 1.14 notwendig ist, dass Intervalle der Form  $[y, \tilde{B}y]$  zugrunde liegen, was insbesondere auf die Intervalle  $I_\ell$ ,  $\ell \geq 1$  zutrifft, stellt sich die Frage, wie die optimalen Parameter im Fall von  $I_0$ , das heißt, für die numerische Berechnung des Faltungsgewichts  $w_{i,0}$ ,  $i = 1, \dots, s$  gewählt werden können. Wie das nächste Lemma zeigt, lässt sich das Faltungsgewicht  $w_{i,0}$  jedoch als eine inverse Laplace-Transformation zum Zeitpunkt  $t = h$  verstehen.

**Lemma 1.16.** *Erfülle  $K$  die Annahme 1.2 und die Kurve  $\Gamma$  die Annahmen aus Satz 1.4. Dann lässt sich das Faltungsgewicht  $w_{i,0}$  für  $i = 1, \dots, s$  aus Satz 1.8 schreiben als*

$$w_{i,0} = \frac{1}{2\pi i} \int_{\Gamma} K(\lambda) e^{h\lambda} \tilde{b}_i(h\lambda) d\lambda \quad \text{mit} \quad \tilde{b}_i(z) := \sum_{j=1}^s z^{-j} \ell_i^{(j-1)}(0).$$

*Beweis.* Da es sich bei den Lagrange-Fundamentalpolynomen  $\ell_i(z)$  um Polynome vom Grad  $s-1$  handelt, siehe Abschnitt 1.1, lässt sich mit Hilfe von partieller Integration leicht nachvollziehen, dass für  $z \in \mathbb{C}$  und  $i = 1, \dots, s$  die Gleichheit

$$b_i(z) = e^z \sum_{j=1}^s \frac{1}{z^j} \ell_i^{(j-1)}(0) - \sum_{j=1}^s \frac{1}{z^j} \ell_i^{(j-1)}(1)$$

gilt. Aus (1.9) erhalten wir damit

$$\begin{aligned} w_{i,0} &= \frac{1}{2\pi i} \int_{\Gamma} K(\lambda) b_i(h\lambda) d\lambda \\ &= \frac{1}{2\pi i} \int_{\Gamma} K(\lambda) \left( e^{h\lambda} \sum_{j=1}^s \frac{1}{(h\lambda)^j} \ell_i^{(j-1)}(0) - \sum_{j=1}^s \frac{1}{(h\lambda)^j} \ell_i^{(j-1)}(1) \right) d\lambda. \end{aligned}$$

Nun folgt mit Satz 1.6 sofort

$$\sum_{j=1}^s \frac{1}{h^j} \left( \frac{1}{2\pi i} \int_{\Gamma} K(\lambda) \frac{1}{\lambda^j} d\lambda \right) \ell_i^{(j-1)}(1) = 0$$

und dadurch die behauptete Darstellung des Faltungsgewichts.  $\square$

Die optimalen Parameter für die Berechnung des Faltungsgewichts  $w_{i,0}$  wählen wir daher bezogen auf den Zeitpunkt  $t = h$  beziehungsweise, um der oben genannten Form zu entsprechen, bezogen auf das Intervall  $[h, h]$ . Damit erhalten wir die Parameter aus Satz 1.14 unter Verwendung von  $B = 1$ .

Als einfache Folgerung aus Satz 1.9 und den Überlegungen in diesem Abschnitt 1.3 erhalten wir das folgende Korollar:

**Korollar 1.17.** *Es erfülle die Laplace-Transformierte  $K$  die Annahme 1.2 und es bezeichne  $U_n$  die Approximation von  $u_n$  aus Satz 1.8, die sich aus der Diskretisierung  $w_{i,j}^{[\kappa]}$  (1.12) der Faltungsgewichte  $w_{i,j}$  und unter Verwendung der Parameterwahlen aus Satz 1.14 ergibt.*

Ist  $g \in C^s([0, T], X)$  und  $t_n = nh \leq T$ , so erhalten wir den Fehler der vollständigen Diskretisierung als

$$\|U_n - u(t_n)\|_X = \mathcal{O}(e^{-c\kappa}) + \mathcal{O}(h^s) \quad \text{für } \kappa \rightarrow \infty \quad \text{und } h \rightarrow 0$$

mit einer Konstanten  $c > 0$ .

*Beweis.* Mit der Dreiecksungleichung und Satz 1.9 folgt sofort

$$\|U_n - u(t_n)\|_X \leq \|U_n - u_n\|_X + \|u_n - u(t_n)\|_X = \|U_n - u_n\|_X + \mathcal{O}(h^s).$$

Weiter erhalten wir mit (1.12) und (1.14) die Abschätzung

$$\begin{aligned} \|U_n - u_n\|_X &= \left\| h \sum_{j=0}^{n-1} \sum_{i=1}^s \left( w_{i, n-1-j}^{[\kappa]} - w_{i, n-1-j} \right) g(t_j + c_i h) \right\|_X \\ &\leq h \sum_{j=0}^{n-1} \sum_{i=1}^s \left\| w_{i, n-1-j}^{[\kappa]} - w_{i, n-1-j} \right\|_{\mathcal{L}(X)} \|g(t_j + c_i h)\|_X \\ &\leq C e^{-c\kappa} \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \|g(\tau)\|_X \, d\tau \\ &\leq C e^{-c\kappa} \int_0^T \|g(\tau)\|_X \, d\tau \end{aligned}$$

mit Konstanten  $c, C > 0$  und damit den behaupteten Fehler.  $\square$

## 1.4 Schneller Faltungsalgorithmus

Der im Folgenden angegebene schnelle Faltungsalgorithmus orientiert sich, wie bereits erwähnt, an den Verfahren aus [18, 45, 60]. Hierbei vermeiden wir, die Faltungsgewichte  $w_{i,j}$  einzeln zu berechnen, wodurch die Approximation der Lösung  $u(t) = (k * g)(t)$  zum Zeitpunkt  $t = Nh$  insgesamt  $\mathcal{O}(N)$  Auswertungen der Laplace-Transformierten  $K$  erfordern würde und alle  $\mathcal{O}(N)$  Faltungsgewichte gespeichert werden müssten. Stattdessen wird durch geschicktes Zusammenfassen und unter Ausnutzung der Intervalle  $I_\ell$  die Berechnung auf  $\mathcal{O}(\log(N))$  Auswertungen von  $K$  reduziert, die parallel durchgeführt werden können. Zusätzlich ist der Algorithmus vergesslich, was bedeutet, dass nur  $\mathcal{O}(\log(N))$  aktiver Speicherplatz zur Berechnung der Lösung benötigt wird. Hierauf gehen wir später genauer ein.

Der Algorithmus bietet sich damit zur schnellen Berechnung der Lösung auf einem kurzen Zeitintervall oder für eine kleine Anzahl an ausgewählten Zeitpunkten an. Er ist jedoch nicht sinnvoll, falls alle Approximationen  $u_1, \dots, u_N$  bestimmt werden sollen. Für Algorithmen mit diesem Zweck verweisen wir insbesondere auf [49, 60].

Zur Durchführung des schnellen Faltungsalgorithmus bestimmen wir zunächst die kleinste

ganze Zahl  $L$ , sodass  $N \leq B^L$  erfüllt ist und definieren

$$n_\ell := B^\ell \quad \text{für } \ell \geq 1 \quad \text{und} \quad n_L := N.$$

Damit ist  $jh \in I_\ell$  für  $\ell \geq 1$  genau dann, wenn  $n_{\ell-1} \leq j \leq n_\ell - 1$ . Der Übersicht halber definieren wir außerdem die Spaltenvektoren

$$b(h\lambda) = \left( b_i(h\lambda) \right)_{i=1}^s \quad \text{und} \quad \tilde{b}(h\lambda) = \left( \tilde{b}_i(h\lambda) \right)_{i=1}^s$$

sowie

$$g_j = \left( g(t_j + c_k h) \right)_{k=1}^s \quad \text{für } j = 0, \dots, N-1.$$

Damit können wir die Approximation  $u_N$  von  $u(t_N)$  mit  $t_N = Nh$ , die wir durch die Anwendung der exponentiellen Quadratur aus Satz 1.8 erhalten, in die  $L+1$  Summen

$$u_N = h \sum_{j=0}^{N-1} \sum_{i=1}^s w_{i,j} g(t_{N-1-j} + c_i h) = u_N^{(0)} + u_N^{(1)} + \dots + u_N^{(L)}$$

aufteilen, wobei wir wegen Lemma 1.16 den ersten Summanden als

$$\begin{aligned} u_N^{(0)} &:= h \sum_{i=1}^s w_{i,0} g(t_{N-1} + c_i h) \\ &= h \frac{1}{2\pi i} \int_{\Gamma_0} e^{h\lambda} \left( \tilde{b}(h\lambda)^T \otimes K(\lambda) \right) d\lambda g_{N-1} \end{aligned}$$

und die Summanden  $u_N^{(\ell)}$  für  $\ell = 1, \dots, L$  als

$$\begin{aligned} u_N^{(\ell)} &:= h \sum_{jh \in I_\ell} \sum_{i=1}^s w_{i,j} g(t_{N-1-j} + c_i h) \\ &= h \sum_{j=n_{\ell-1}}^{n_\ell-1} \frac{1}{2\pi i} \int_{\Gamma_\ell} e^{\lambda(jh)} \left( b(h\lambda)^T \otimes K(\lambda) \right) d\lambda g_{N-1-j} \end{aligned}$$

definieren.<sup>1</sup>

Mit der Diskretisierung  $w_{i,j}^{[\kappa]}$  (1.12) der Faltungsgewichte  $w_{i,j}$  sowie den Gewichten  $\omega_k^{(\ell)}$  und Stützstellen  $\lambda_k^{(\ell)}$  aus (1.13) erhalten wir die Approximation  $U_N^{(0)}$  von  $u_N^{(0)}$  als

$$U_N^{(0)} := h \sum_{k=-\kappa}^{\kappa} \omega_k^{(0)} e^{h\lambda_k^{(0)}} \left( \tilde{b}(h\lambda_k^{(0)})^T \otimes K(\lambda_k^{(0)}) \right) g_{N-1}. \quad (1.15)$$

<sup>1</sup>Die Schreibweise mit Hilfe des Tensor-Produkts verwenden wir hier der Übersicht halber auch für den allgemeinen Fall eines unendlich dimensionalen Raums  $X$ . Die Bedeutung erschließt sich hierbei aus der Anwendung auf einen endlich dimensionalen Raum.

Für  $\ell = 1, \dots, L$  können wir die Approximationen  $U_N^{(\ell)}$  von  $u_N^{(\ell)}$  entsprechend als

$$\begin{aligned} U_N^{(\ell)} &:= h \sum_{j=n_{\ell-1}}^{n_{\ell}-1} \sum_{k=-\kappa}^{\kappa} \omega_k^{(\ell)} e^{\lambda_k^{(\ell)}(jh)} \left( b(h\lambda_k^{(\ell)})^T \otimes K(\lambda_k^{(\ell)}) \right) g_{N-1-j} \\ &= h \sum_{k=-\kappa}^{\kappa} \omega_k^{(\ell)} \sum_{j=N-n_{\ell}}^{N-1-n_{\ell-1}} (e^{h\lambda_k^{(\ell)}})^{N-1-j} \left( b(h\lambda_k^{(\ell)})^T \otimes K(\lambda_k^{(\ell)}) \right) g_j \\ &= \sum_{k=-\kappa}^{\kappa} \omega_k^{(\ell)} (e^{h\lambda_k^{(\ell)}})^{n_{\ell}-1} K(\lambda_k^{(\ell)}) y_k^{(\ell)} \end{aligned}$$

schreiben, wobei

$$y_k^{(\ell)} = h \sum_{j=N-n_{\ell}}^{N-1-n_{\ell-1}} (e^{h\lambda_k^{(\ell)}})^{N-1-n_{\ell-1}-j} \left( b(h\lambda_k^{(\ell)})^T \otimes \text{Id} \right) g_j. \quad (1.16)$$

Verglichen mit der Rekursionsformel (1.7) lässt sich leicht einsehen, dass  $y_k^{(\ell)}$  genau der Approximation der Lösung  $y_{\lambda_k^{(\ell)}}(t)$  des lineares Anfangswertproblems

$$y'(t) = \lambda_k^{(\ell)} y(t) + g(t), \quad y((N - n_{\ell})h) = 0$$

zum Zeitpunkt  $t = (N - n_{\ell-1})h$  entspricht, die wir durch die stückweise Polynominterpolation der Funktion  $g$  erhalten. Diese Approximation bezeichnen wir im Folgenden als skalare exponentielle Quadratur. Im schnellen Algorithmus erhalten wir die Werte  $y_k^{(\ell)}$  daher für jedes der  $2\kappa + 1$  Anfangswertprobleme nach  $n_{\ell} - n_{\ell-1}$  Schritten der skalaren exponentiellen Quadratur.

Berechnen wir weiter  $x_k^{(\ell)}$  als

$$x_k^{(\ell)} = K(\lambda_k^{(\ell)}) y_k^{(\ell)}, \quad (1.17)$$

so können wir die Approximation  $U_N^{(\ell)}$  als die Linearkombination

$$U_N^{(\ell)} = \sum_{k=-\kappa}^{\kappa} c_k^{(\ell)} x_k^{(\ell)} \quad \text{mit} \quad c_k^{(\ell)} = \omega_k^{(\ell)} (e^{h\lambda_k^{(\ell)}})^{n_{\ell}-1} \quad (1.18)$$

schreiben. Hier wird deutlich, dass für die Berechnung von  $U_N^{(\ell)}$  insgesamt  $2\kappa + 1$  Auswertungen der Laplace-Transformierten notwendig sind. Wollen wir bei der Diskretisierung der Faltungsgewichte eine Genauigkeit von  $\mathcal{O}(\varepsilon)$  erreichen, so benötigen wir für die Berechnung von  $U_N^{(0)}, \dots, U_N^{(L)}$  daher insgesamt  $\mathcal{O}(\log(N) \log(1/\varepsilon))$  Auswertungen der Laplace-Transformierten, vergleiche dazu (1.14).

Des Weiteren ist für jedes  $U_N^{(\ell)}$  lediglich die Speicherung von  $\mathcal{O}(\log(1/\varepsilon))$  Auswertungen der Laplace-Transformierten sowie  $\mathcal{O}(\log(1/\varepsilon))$  Ergebnissen der skalaren exponentiellen Quadratur erforderlich. Damit beläuft sich der aktiv benötigte Speicherplatz für den schnellen Algorithmus auf  $\mathcal{O}(\log(N) \log(1/\varepsilon))$ .

Zur Berechnung aller erforderlichen Werte  $y_k^{(\ell)}$  werden insgesamt  $\mathcal{O}(N \log(1/\varepsilon))$  Schritte der skalaren exponentiellen Quadratur benötigt.

Sind sowohl der Faltungskern  $k$  als auch die Funktion  $g$  reellwertig, lässt sich schnell einsehen, dass sich bei der Berechnung von  $U_N^{(\ell)}$  die Anzahl der Auswertungen der Laplace-Transformierten auf  $\kappa + 1$  reduziert: Da die Stützstellen  $\lambda_k^{(\ell)}$  für  $k = -\kappa, \dots, \kappa$  symmetrisch bezüglich der reellen Achse liegen, sind auch die Werte  $y_k^{(\ell)}$  aus (1.16) sowie  $x_k^{(\ell)}$  aus (1.17) symmetrisch bezüglich der reellen Achse. Letzteres ergibt sich aus der Definition der Laplace-Transformierten  $K$ . Wegen  $\omega_k^{(\ell)} = \omega_{-k}^{(\ell)}$  löschen sich folglich die Imaginärteile in (1.18) aus. Damit reicht es, anstatt der Stützstellen  $\lambda_k^{(\ell)}$ ,  $k = -\kappa, \dots, \kappa$  lediglich die Stellen  $\text{Re}(\lambda_k^{(\ell)})$  für  $k = 0, \dots, \kappa$  zu betrachten.

Zusammengefasst besteht der schnelle Faltungsalgorithmus für die exponentielle Quadratur aus den folgenden Schritten:

1. *Schritt:* Berechne parallel die Summanden  $U_N^{(\ell)}$  für  $\ell = 0, \dots, L$ , die über (1.15) beziehungsweise (1.18) gegeben sind. Hierbei kann (1.16) zusätzlich über  $k$  und über die Einträge der Funktion  $g$  parallelisiert werden, während die Auswertungen (1.17) für jedes  $k$  parallel berechnet werden können.
2. *Schritt:* Erhalte die Approximation  $U_N$  von  $u_N$  durch die Summe

$$U_N = U_N^{(0)} + \dots + U_N^{(L)}.$$

## 1.5 Numerischer Test

Um die Resultate aus den vorherigen Abschnitten zu verdeutlichen, betrachten wir zunächst ein klassisches Testbeispiel  $u(t) = (k * g)(t)$ , in dem der Faltungskern bekannt und die Lösung  $u$  berechnet werden kann. In Kapitel 4 folgt die Anwendung auf ein komplizierteres Problem.

Wir benötigen hier die *Mittag-Leffler* Funktion mit zwei Parametern  $\alpha$  und  $\beta$ , die definiert ist als

$$E_{\alpha,\beta}(z) := \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + \beta)} \quad \text{für } \alpha, \beta > 0 \quad \text{und } z \in \mathbb{C}$$

und deren Verallgemeinerung

$$\mathbf{e}_{\alpha,\beta}(t; \lambda) := t^{\beta-1} E_{\alpha,\beta}(\lambda t^\alpha) \quad \text{mit } \lambda \in \mathbb{C} \quad \text{und } t > 0, \quad (1.19)$$

siehe beispielsweise [50, 56]. Damit definieren wir für eine Konstante  $a = r e^{i(\pi/2 + \varphi)}$  mit  $r > 0$ ,  $\varphi \in (0, \pi/2]$  und  $0 < \mu < 1$  den Faltungskern  $k_a$  als

$$k_a : \mathbb{R}^+ \rightarrow \mathbb{C}, \quad t \mapsto k_a(t) := \mathbf{e}_{\mu,\mu}(t; a) \quad (1.20)$$

sowie die Funktion  $g$  als

$$g : \mathbb{R} \rightarrow \mathbb{C}, \quad t \mapsto g(t) := \frac{1}{\Gamma(5 - \mu)} t^{4 - \mu},$$

vergleiche dazu [28]. Mit Hilfe von [56, Absch. 1.2.6] folgt

$$u(t) = (k_a * g)(t) = \mathbf{e}_{\mu,5}(t; a). \quad (1.21)$$

Ebenfalls bekannt ist die Laplace-Transformierte der verallgemeinerten Mittag-Leffler Funktion. Es gilt nach [50]

$$K_a(\lambda) := \mathcal{L}\{k_a(t); \lambda\} = (\lambda^\mu - a)^{-1} \quad \text{für } \operatorname{Re}(\lambda) > |a|^{1/\mu}.$$

Setzen wir den Winkel  $\theta_0$  abhängig vom Parameter  $\mu$  und vom Argument  $\arg(a) = \frac{\pi}{2} + \varphi$  als

$$\theta_0 := \min\left(\frac{\pi(1 - \mu) + 2\varphi}{2\mu}, \frac{\pi}{2}\right),$$

lässt sich leicht einsehen, dass  $|\arg(\lambda^\mu)| < \pi/2 + \varphi$  für  $\lambda \in \Sigma_{\pi/2 + \theta_0}$ . Damit erfüllt  $K_a(\lambda)$  die Sektorialitätsannahme 1.2 mit angegebenem Winkel  $\theta_0$  und Parameter  $\mu < 1$ .

Wie bereits in der Einleitung des Kapitels erwähnt, wird in [26, 27] die exponentielle Quadratur unter der Annahme der Darstellbarkeit des Faltungskerns als konvergente Potenzreihe betrachtet, konkreter wird hier der Faltungskern in der Form (1.20) angenommen und die Gewichte als Faltungen der verallgemeinerten Mittag-Leffler Funktion und der Lagrange-Polynome angegeben, vergleiche dazu die Darstellung der  $w_{i,j}$  im Beweis zu Satz 1.12. Insbesondere für den einfachen Fall einer Polynominterpolation mit einem Knoten werden die Gewichte in [26, Bsp. 8] mit Hilfe von partieller Integration vereinfacht. Dadurch ergibt sich die von der Wahl des Knoten unabhängige Darstellung

$$w_{1,j}^{\text{ML}} = h^{\mu-1} (\mathbf{e}_{\mu,\mu+1}(j+1; h^\mu a) - \mathbf{e}_{\mu,\mu+1}(j; h^\mu a)),$$

die wir dazu verwenden können, die Wahl der Parameter aus Satz 1.14 für die Diskretisierung  $w_{i,j}^{[\kappa]}$  mit  $s = 1$  zu überprüfen. Dazu stellen wir in Abbildung 1.3 den absoluten Fehler

$$\left| w_{1,j}^{\text{ML}} - w_{1,j}^{[\kappa]} \right|$$

in Abhängigkeit von  $j = 0, \dots, 10^4 - 1$  dar. Hierbei setzen wir  $h = 10^{-4}$ ,  $\mu = 0.3$  und  $a = -3$ . Nach obiger Überlegung erfüllt  $K_{-3}(\lambda)$  die Annahme 1.2 damit in  $\Sigma_\pi$ . Für die optimalen Parameter aus Satz 1.14 betrachten wir die Werte  $B = 5$  und  $B = 10$  sowie  $\kappa = 10$  und  $\kappa = 25$ . Die Mittag-Leffler Funktion in den Referenzgewichten  $w_{1,j}^{\text{ML}}$  werten wir mit Hilfe der MATLAB-Implementierung `mlf`, siehe [57], mit Maschinengenauigkeit aus. Wie in der Abbildung zu erkennen ist, werden zur Berechnung der Gewichte  $w_{1,j}$ ,  $jh \in [0, 10^4 h)$  für die Basis  $B = 5$  insgesamt 7 verschiedene Hyperbeln, also  $\Gamma_\ell$  für  $\ell = 0, \dots, 6$ , benötigt,

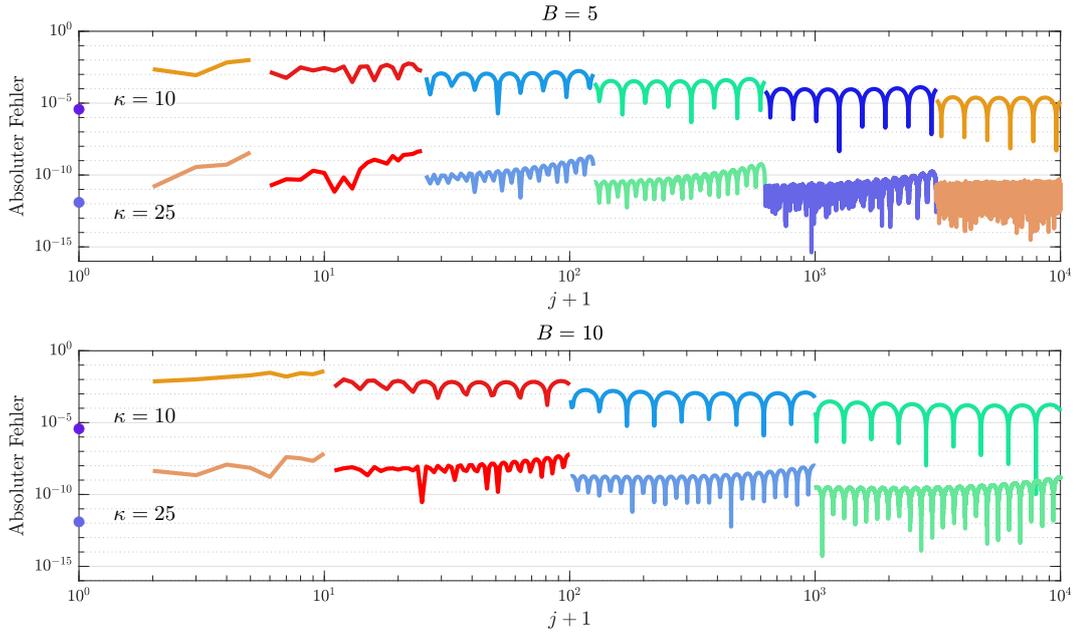


Abbildung 1.3: Absoluter Fehler in den Gewichten  $w_{1,j}^{[\kappa]}$  in Abhängigkeit des Index  $j + 1$  für die Werte  $B = 5$  und  $B = 10$  sowie  $\kappa = 10$  und  $\kappa = 25$ .

während die Basis  $B = 10$  zur Diskretisierung mit 5 verschiedenen Hyperbeln führt. Wie zu erwarten war, führt ein größerer Wert von  $\kappa$  zu einem kleineren Diskretisierungsfehler.

Um die Abschätzung der Faltungsgewichte aus Satz 1.12 zu bestätigen, berechnen wir die Gewichte  $w_{1,j}^{[\kappa]}$  für jeweils zwei Knotenwahlen  $c_i$ ,  $i = 1, \dots, s$ , wobei die beiden Fälle  $s = 2$  und  $s = 3$  betrachtet werden. Wir wählen hierbei wieder  $h = 10^{-4}$ ,  $\mu = 0.3$  und  $a = -3$ , und setzen die Parameter  $\kappa = 25$  und  $B = 5$  fest. In Abbildung 1.4 wird deutlich, dass die Schranke aus Satz 1.12 erfüllt ist, legt aber auch für größere  $j$  ein stärkeres Abfallverhalten nahe.

Das numerische Experiment in diesem Kapitel schließen wir mit einem Blick auf die Konvergenzaussagen aus Satz 1.9 und Bemerkung 1.10 ab. Wie in der Bemerkung erwähnt, werden in [26, 28] Bedingungen an die Knoten, die der Lagrange-Interpolation zugrunde liegen, hergeleitet, mit denen eine höhere Konvergenzordnung erreicht werden kann als nach Satz 1.9 zu erwarten ist. Erfüllen die Knoten die Bedingung

$$\int_0^1 \omega(\vartheta) d\vartheta = 0 \quad \text{mit} \quad \omega(\vartheta) := \prod_{i=1}^s (\vartheta - c_i),$$

so wird in den genannten Quellen die Konvergenzordnung  $s + \mu$  für  $0 < \mu < 1$  gefolgert. Gilt zudem die weitere Bedingung

$$\int_0^1 \omega(\vartheta) \zeta(1 - \mu, 1 - \vartheta) d\vartheta = 0$$

für die Hurwitz'sche Zeta-Funktion  $\zeta$ , so soll die Konvergenzordnung  $s + \min(1, 2\mu)$  erreicht

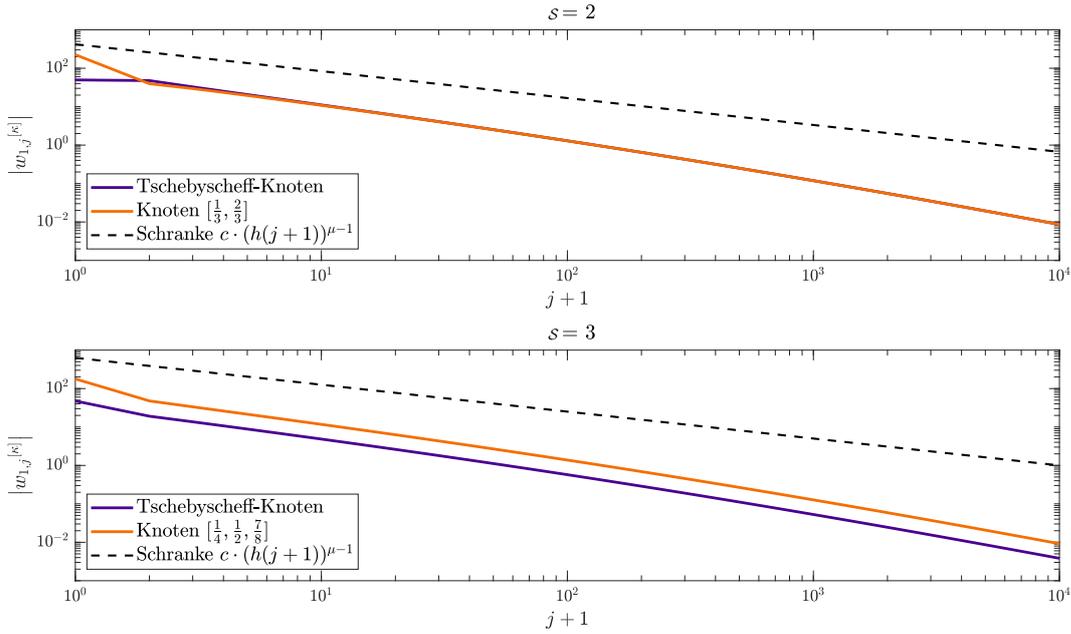


Abbildung 1.4: Absolutbetrag der Gewichte  $w_{1,j}^{[\kappa]}$  für  $s = 2, 3$  in Abhängigkeit des Index  $j + 1$  für die Parameter  $B = 5$  und  $\kappa = 25$ . Die Konstante  $c$  in der Schranke aus Satz 1.12 ist für den Plot entsprechend angepasst. Die Tschebyscheff-Knoten berechnen sich als  $c_i = \frac{1}{2} \left( 1 + \cos \left( \frac{(2j-1)\pi}{2s} \right) \right)$ ,  $i = 1, \dots, s$ .

werden. Für die Fälle  $s = 2$  und  $s = 3$  sowie bestimmte Werte für  $\mu$  werden in [28] Knoten angegeben, die die beiden obigen Bedingungen erfüllen. Damit die Knoten im Fall  $s = 3$  hierbei eindeutig sind, wird eine weitere Bedingung hinzugenommen, die die Konvergenzordnung  $s + \min(1, 3\mu)$  verspricht, vergleiche [28, Tabellen 1 und 2].

Um unsere Vermutung aus Bemerkung 1.10 zu bestätigen, die darin bestand, dass für die Konvergenzaussage aus [26, 28] Voraussetzungen fehlen, betrachten wir im folgenden numerischen Test Knoten, die die obigen Bedingungen erfüllen. Mit Hilfe des schnellen Algorithmus aus Abschnitt 1.4 berechnen wir dazu die Approximation  $u_N$  von (1.21) zum Zeitpunkt  $t_N = Nh = 1$ , wobei wir weiterhin  $\mu = 0.3$  wählen, jedoch verschiedene Werte für den Parameter  $a$  betrachten. Wir setzen hierbei die Werte  $B = 5$  und  $\kappa = 25$  fest. In Abbildung 1.5 liegt die stückweise lineare Interpolation zugrunde, das heißt, wir betrachten den Fall  $s = 2$ . Die hierbei verwendeten Tschebyscheff-Knoten erfüllen keine der obigen Bedingungen, das zweite Paar Knoten  $[0, 2/3]$  dagegen die erste Bedingung und die „optimalen Knoten“ sind in [28, Tabellen 1] zu finden. Nach [26, 28] erwarten wir daher die Ordnungen 2, 2.3 und 2.6, die durch die gestrichelten Linien dargestellt werden. Lassen sich für die Fälle  $a = -3$  sowie  $a = 3e^{i3\pi/4}$  die angegebenen Ordnungen noch annähernd beobachten, so wird für größere Werte, hier  $a = -32$  und  $a = 32e^{i3\pi/4}$ , deutlich, dass nur noch eine Konvergenz der Ordnung 2 erreicht wird.

Die gleiche Beobachtung lässt sich in Abbildung 1.6 für den Fall  $s = 3$  machen. Hierbei sind die Knoten der Interpolation so gewählt, dass in den genannten Quellen die Ordnung 3, 3.3 beziehungsweise 3.9 behauptet wird.

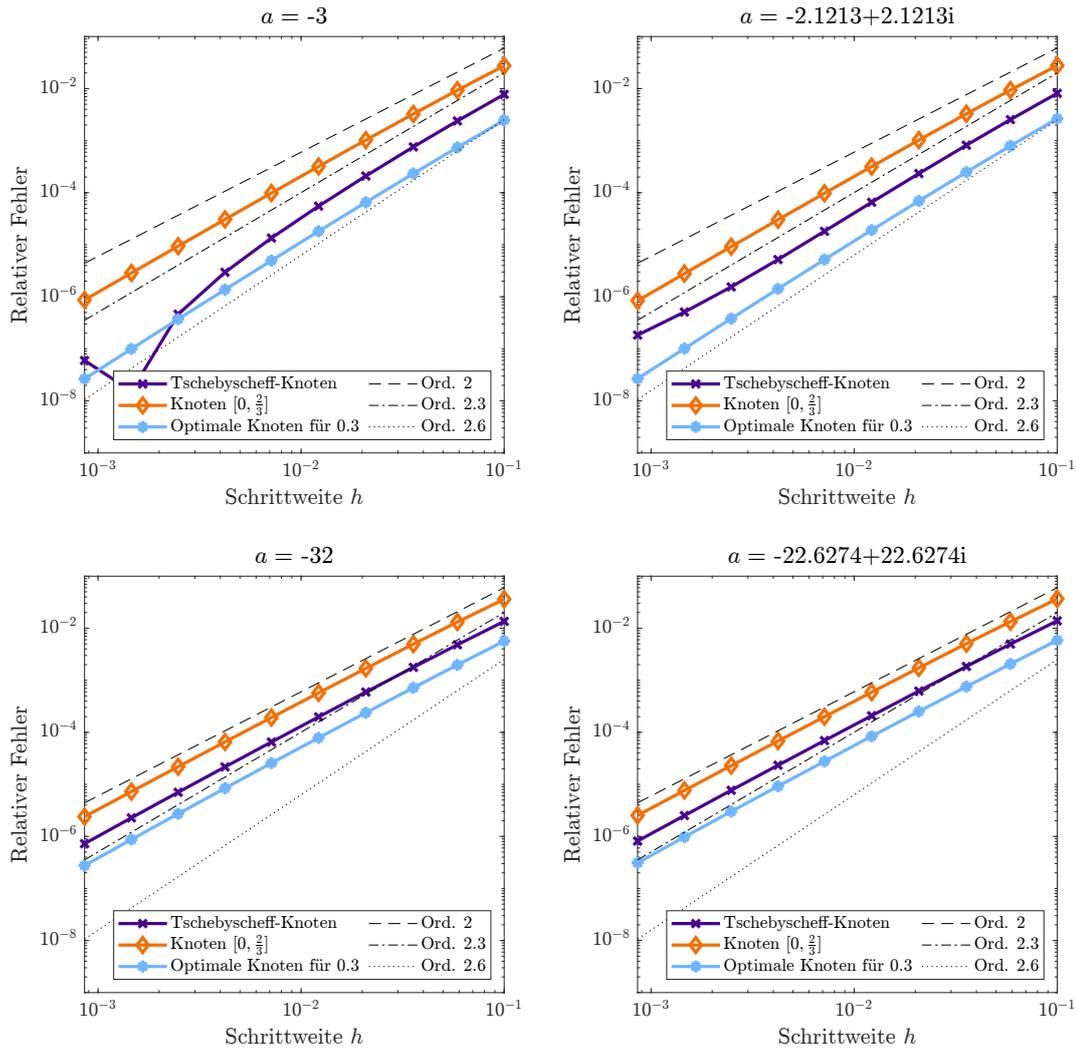


Abbildung 1.5: Relativer Fehler  $|u(t_N) - u_N| / |u(t_N)|$  mit  $t_N = 1$  und  $\mu = 0.3$  in Abhängigkeit der Zeitschrittweite  $h$  für die exponentielle Quadratur mit  $s = 2$ .

Betrachten wir den Wert  $\mu = 0.7$  lässt sich ebenfalls feststellen, dass für größere Werte von  $a$  die angegebenen Ordnungen von 3, 3.7 beziehungsweise 4 nicht mehr erreicht werden. Hierbei fällt jedoch auf, dass die Werte größer als im vorherigen Beispiel gewählt werden müssen. Für die beispielhaften Wahlen  $a = -3$  und  $a = 32e^{i3\pi/4}$ , bei denen im Fall  $\mu = 0.3$  die Beobachtung bereits deutlich wurde, sehen wir in Abbildung 1.7 noch die jeweils in [26, 28] behaupteten Ordnungen. Für die Werte  $a = -800$  und  $a = 800e^{i3\pi/4}$  sehen wir dagegen wieder die Ordnung 3, die wir nach Satz 1.9 erwarten können.

Da der schnelle Algorithmus für die exponentielle Quadratur auf dem schnellen Algorithmus aus [18, 60] basiert, verweisen wir für die Untersuchungen der Parallelisierbarkeit und des Aufwands auf die Ergebnisse aus [18, Bsp. 6.2].

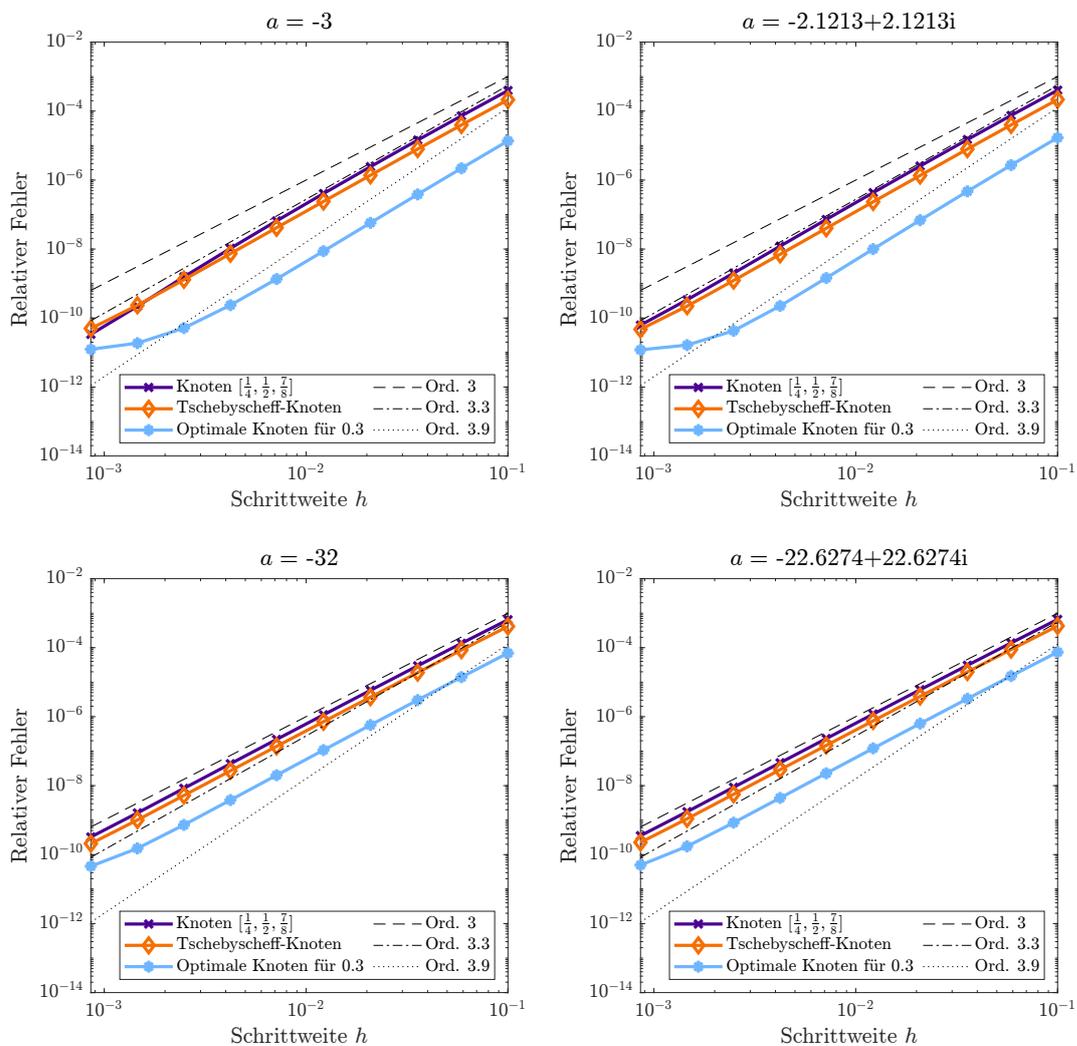


Abbildung 1.6: Relativer Fehler  $|u(t_N) - u_N| / |u(t_N)|$  mit  $t_N = 1$  und  $\mu = 0.3$  in Abhängigkeit der Zeitschrittweite  $h$  für die exponentielle Quadratur mit  $s = 3$ .

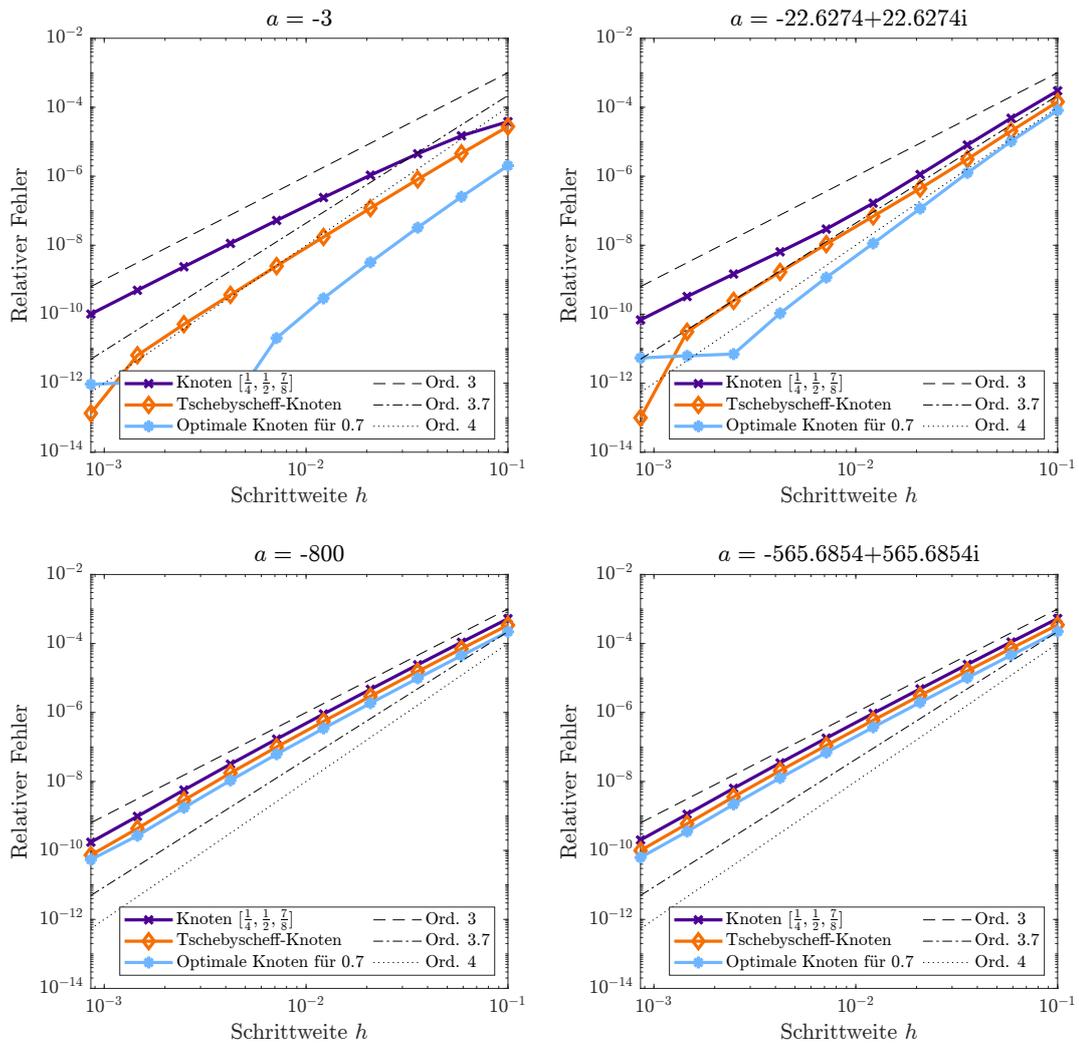


Abbildung 1.7: Relativer Fehler  $|u(t_N) - u_N|/|u(t_N)|$  mit  $t_N = 1$  und  $\mu = 0.7$  in Abhängigkeit der Zeitschrittweite  $h$  für die exponentielle Quadratur mit  $s = 3$ .



## KAPITEL 2

---

# RUNGE-KUTTA FALTUNGSQUADRATUR FÜR NICHT-SEKTORIELLE LAPLACE-TRANSFORMIERTE

Wie bereits angesprochen, betrachten wir in diesem Kapitel eine Klasse von Laplace-Transformierten, die die Sektorialitätsannahme 1.2 nicht erfüllen, jedoch über ein geeignetes Abfallverhalten auf der rechten Halbebene verfügen. Für die Forderung, dass die Laplace-Transformierte in der rechten Halbebene polynomiell beschränkt ist, wurde in [3] der Begriff *nicht-sektoriell* beziehungsweise *non-sectorial* eingeführt, den wir auch hier gelegentlich zur Unterscheidung zur sektoriellen Laplace-Transformierten verwenden. Wie wir in der Anwendung in Kapitel 3 sehen werden, benötigen wir zusätzlich zur polynomiellen Beschränkung in der Halbebene auch noch eine solche in Sektoren der Form  $|\arg(\lambda)| \leq \pi/2 - \theta$ ,  $0 < \theta \leq \pi/2$ .

In diesem Kapitel werden wir die wichtigsten Grundlagen der Runge-Kutta basierten Faltungsquadratur zur Diskretisierung des Faltungsintegrals im nicht-sektoriellen Fall wiedergeben. Das Verfahren wurde zunächst unter der Annahme einer sektoriellen Laplace-Transformierten von Lubich und Ostermann [48] vorgestellt und später von Banjai und Lubich in [3, 4] auf den nicht-sektoriellen beziehungsweise auf den oben beschriebenen Fall erweitert. Da hierbei insbesondere ein  $A$ -stabiles Verfahren zur Zeitdiskretisierung benötigt wird, lässt sich schnell der Vorteil einer Runge-Kutta basierten Faltungsquadratur gegenüber der ebenfalls häufig verwendeten Quadratur basierend auf Mehrschrittverfahren, siehe zum Beispiel [46, 47], einsehen: Während die Ordnung von  $A$ -stabilen Mehrschrittverfahren nach Dahlquist höchstens 2 sein kann, können  $A$ -stabile Runge-Kutta Verfahren beliebige Ordnungen erreichen, vergleiche [33].

Nachdem wir im ersten Teil des Kapitels einen kurzen Blick auf Runge-Kutta Verfahren geworfen haben, werden wir anschließend die Runge-Kutta Faltungsquadratur angeben und genauer auf das erwähnte Abfallverhalten eingehen. Im letzten Teil des Kapitels beschreiben

wir, an welchen Stellen in der Implementierung die schnelle Fourier-Transformation verwendet werden kann um Rechenaufwand zu sparen, wobei das Paper [32] von Hairer, Lubich und Schlichte zu nennen ist.

## 2.1 Runge-Kutta Verfahren

Im Folgenden betrachten wir ein implizites  $s$ -stufiges Runge-Kutta Verfahren mit Schrittweite  $h > 0$ . Angewendet auf das Anfangswertproblem

$$y'(t) = f(t, y(t)), \quad y(0) = y_0 \quad (2.1)$$

ist die Approximation  $y_{n+1}$  der Lösung  $y : \mathbb{R}_0^+ \rightarrow X$  zum Zeitpunkt  $t_{n+1} = (n+1)h$  rekursiv gegeben durch

$$\begin{aligned} Y_{ni} &= y_n + h \sum_{j=1}^s a_{ij} f(t_n + c_j h, Y_{nj}), \quad i = 1, \dots, s, \\ y_{n+1} &= y_n + h \sum_{j=1}^s b_j f(t_n + c_j h, Y_{nj}). \end{aligned} \quad (2.2)$$

Wir bezeichnen hierbei die Runge-Kutta Matrix mit  $\mathcal{O} = (a_{ij})_{i,j=1}^s$  und den Spaltenvektor der Gewichte mit  $b = (b_j)_{j=1}^s$ .

**Definition 2.1.** Das Verfahren (2.2) ist von der *Konsistenzordnung*  $p \geq 1$ , das heißt, der Fehler nach einem Schritt angewendet auf (2.1) mit genügend oft differenzierbarer rechter Seite, ist von der Größenordnung  $h^{p+1}$ , also

$$\|y_1 - y(h)\|_X = \mathcal{O}(h^{p+1}) \quad \text{für } h \rightarrow 0.$$

Das Verfahren hat zudem die *Stufenordnung*  $q \leq p$ , wenn für  $i = 1, \dots, s$

$$\|Y_{0i} - y(c_i h)\|_X = \mathcal{O}(h^{q+1}) \quad \text{für } h \rightarrow 0$$

erfüllt ist.

Die Anwendung des Runge-Kutta Verfahrens (2.2) auf die Testgleichung  $y' = \lambda y$ ,  $y(0) = y_0$  liefert die Approximation  $y_n = r(h\lambda)^n y_0$ , bei der die Stabilitätsfunktion  $r$  mit der obigen Notation geschrieben werden kann als

$$r(z) := 1 + z b^T (\text{Id} - z \mathcal{O})^{-1} \mathbb{1} \quad \text{mit} \quad \mathbb{1} := (1, \dots, 1)^T.$$

Für ein implizites Verfahren  $p$ -ter Ordnung ist die Stabilitätsfunktion eine rationale Approximation der Ordnung  $p$  an die Exponentialfunktion, siehe zum Beispiel [33].

Wie bereits in der Einleitung des Kapitels erwähnt, benötigen wir zur Faltungsquadratur  $A$ -stabile Verfahren. Dies ist für implizite Runge-Kutta Verfahren nach [33] genau dann gegeben, wenn für alle  $z \in \mathbb{C}$  mit  $\operatorname{Re}(z) \leq 0$  gilt

$$|r(z)| \leq 1 \quad \text{und} \quad \operatorname{Id} - z\mathcal{O} \text{ ist nicht-singulär.}$$

Zusätzlich zur  $A$ -Stabilität treffen wir die weiteren Annahmen:

(RK1)  $|r(i\alpha)| < 1$  für alle  $\alpha \in \mathbb{R} \setminus \{0\}$ ,

(RK2) die Runge-Kutta Matrix  $\mathcal{O}$  ist invertierbar und

(RK3) der Vektor der Gewichte  $b^T$  ist gleich der letzten Zeile der Matrix  $\mathcal{O}$ , das heißt, es gilt  $b_i = a_{si}$ ,  $i = 1, \dots, s$  und damit  $b^T \mathcal{O}^{-1} = e_s^T$ .

Hierbei folgt aus (RK2) und (RK3) direkt  $r(\infty) = 0$  und damit die  $L$ -Stabilität des Verfahrens, siehe [33, Prop. IV.3.8]. Weiter bedeutet (RK3), dass das Ergebnis der  $s$ -ten Stufe direkt die Approximation zum vollen Zeitschritt liefert, also  $Y_{ns} = y_{n+1}$  und daher entsprechend  $c_s = 1$  gilt.

Eine Familie von Runge-Kutta Verfahren, die die genannten Voraussetzungen erfüllen, sind die  $s$ -stufigen Radau IIA Verfahren. Diese sind nach [33]  $A$ -stabil, von der Ordnung  $2s - 1$  und haben die Stufenordnung  $s$ . In den numerischen Experimenten in den Kapiteln 4 und 5 werden wir diese Verfahren mit den Ordnungen 1, 3 und 5 zur Zeitdiskretisierung verwenden. Hierbei entspricht das 1-stufige Radau IIA Verfahren dem impliziten Euler Verfahren.

## 2.2 Runge-Kutta basierte Faltungsquadratur

Nehmen wir nun an, dass wir ein  $A$ -stabiles implizites Runge-Kutta Verfahren haben, das die Annahmen (RK1)-(RK3) erfüllt, so können wir die Runge-Kutta basierte Faltungsquadratur zur Diskretisierung des Faltungsintegrals (1.1) angeben. Wir gehen davon aus, dass die Laplace-Transformierte

$$K : D(K) \subset \mathbb{C} \rightarrow \mathcal{L}(X), \quad \lambda \mapsto K(\lambda) := \mathcal{L}\{k(t); \lambda\}$$

die folgenden Annahmen erfüllt:

**Annahme 2.2.** Die Laplace-Transformierte  $K$  ist analytisch in der Halbebene  $\operatorname{Re}(\lambda) > \sigma_0 > 0$  und es gibt Exponenten  $\mu_1 \geq 0$ ,  $\mu_2 \geq \mu_1$  und von  $\sigma_0$  und  $0 < \theta \leq \pi/2$  abhängige Konstanten  $C_1(\sigma_0)$ ,  $C_2(\sigma_0, \theta) > 0$ , so dass die folgenden Abschätzungen erfüllt sind:

(i)  $\|K(\lambda)\|_{\mathcal{L}(X)} \leq C_1(\sigma_0) |\lambda|^{-\mu_1}$  für  $\operatorname{Re}(\lambda) > \sigma_0$

(ii)  $\|K(\lambda)\|_{\mathcal{L}(X)} \leq C_2(\sigma_0, \theta) |\lambda|^{-\mu_2}$  für  $\operatorname{Re}(\lambda) > \sigma_0$  und  $|\arg(\lambda)| \leq \frac{\pi}{2} - \theta$

Wir führen hierbei die Notation  $\Sigma(\sigma_0) := \{\lambda \in \mathbb{C} \mid \operatorname{Re}(\lambda) > \sigma_0\}$  für  $\sigma_0 > 0$  ein.

Wie in [4] gezeigt wird, schließt Annahme 2.2 eine Klasse von Laplace-Transformierten mit ein, die die Abschätzung im nachfolgenden Lemma erfüllen.

**Lemma 2.3.** *Ist  $K$  analytisch in der rechten Halbebene  $\operatorname{Re}(\lambda) > \sigma_0 > 0$  und gibt es  $\mu, \nu \geq 0$  und eine Konstante  $C(\sigma_0) > 0$ , so dass*

$$\|K(\lambda)\|_{\mathcal{L}(X)} \leq C(\sigma_0) \frac{1}{|\lambda|^\mu \operatorname{Re}(\lambda)^\nu},$$

so erfüllt  $K$  die Annahme 2.2 mit  $\mu_1 = \mu$  und  $\mu_2 = \nu + \mu$ .

Die Idee hinter der Faltungsquadratur nach Lubich basiert nun wieder auf der Darstellung von  $u(t) = (k * g)(t)$  unter Verwendung der inversen Laplace-Transformation, die wir bereits in (1.4) angegeben haben. Anstatt hierbei die Funktion  $y_\lambda$  wie im vorigen Kapitel mit Hilfe der Polynominterpolation zu approximieren, nutzen wir aus, dass sie die Lösung des inhomogenen Anfangswertproblems (1.5) ist, die wir mit Hilfe des gegebenen Runge-Kutta Verfahrens diskretisieren können.

Dieser Ansatz liefert uns eine Approximation an  $y_\lambda$ , die wir in (1.4) einsetzen und mit der Cauchyschen Integralformel vereinfachen können. Wir erhalten damit die Runge-Kutta basierte Faltungsquadratur für nicht-sektorielle Laplace-Transformierte, die sich von der Formulierung in [48] für den sektoriellen Fall nur durch leicht geänderte Voraussetzungen an die Runge-Kutta Matrix unterscheidet, siehe [3].

**Satz 2.4.** *Betrachte ein  $s$ -stufiges implizites Runge-Kutta Verfahren der Form (2.2), das A-stabil ist und die Annahmen (RK1)-(RK3) erfüllt und gelte für die Laplace-Transformierte  $K(\lambda) := \mathcal{L}\{k(t); \lambda\}$  die Annahme 2.2. Mit der Schrittweite  $h > 0$  erhalten wir zum Zeitpunkt  $t_n = nh$  eine Approximation von  $u(t) = (k * g)(t)$  als*

$$u_n = h \sum_{j=0}^{n-1} w_{n-1-j} g_j,$$

wobei  $g_j = (g(t_j + c_k h))_{k=1}^s$  ein Spaltenvektor und  $w_j := (e_s^T \otimes \operatorname{Id}) W_j$  die letzte Zeile des  $s \times s$ -Faltungsgewichts  $W_j$  ist, das gegeben ist als der  $j$ -te Koeffizient der erzeugenden Funktion

$$h \sum_{j=0}^{\infty} W_j \zeta^j = K\left(\frac{\Delta(\zeta)}{h}\right) \quad \text{mit} \quad \Delta(\zeta) = \left(\alpha + \frac{\zeta}{1-\zeta} \mathbb{1} b^T\right)^{-1}. \quad (2.3)$$

**Bemerkung 2.5.** Wollen wir die inneren Stufen der Runge-Kutta basierten Faltungsquadratur bestimmen, das heißt die Approximationen von  $u(t_n + c_i h)$ ,  $i = 1, \dots, s$ , erhalten wir diese mit den Definitionen aus Satz 2.4 als

$$U_{ni} = (e_i^T \otimes \operatorname{Id}) h \sum_{j=0}^n W_j g_{n-j}, \quad i = 1, \dots, s.$$

Die vom Runge-Kutta Verfahren abhängige Funktion  $\Delta$  wird in [4] genauer untersucht. Darin wird das folgende Lemma angegeben:

**Lemma 2.6.** *Es sei ein  $A$ -stabiles Runge-Kutta Verfahren gegeben, das die Voraussetzungen (RK2) und (RK3) erfüllt. Dann ist das Spektrum  $\sigma(\Delta(\zeta))$  für  $|\zeta| < 1$  enthalten in der rechten Halbebene  $\operatorname{Re}(\lambda) > 0$ .*

*Beweis.* Unter den Voraussetzungen an das Runge-Kutta Verfahren wird in [48, Lemma 2.4] die Identität

$$(\Delta(\zeta) - z\operatorname{Id})^{-1} = \mathcal{O}(\operatorname{Id} - z\mathcal{O})^{-1} + \zeta \sum_{j=0}^{\infty} (r(z)\zeta)^j (\operatorname{Id} - z\mathcal{O})^{-1} \mathbb{1}b^T (\operatorname{Id} - z\mathcal{O})^{-1}$$

gezeigt, aus der sofort

$$\sigma(\Delta(\zeta)) \subset \sigma(\mathcal{O}^{-1}) \cup \{z \in \mathbb{C} \mid |r(z)\zeta| \geq 1\}$$

folgt. Aus der  $A$ -Stabilität erhalten wir hierbei  $|r(z)\zeta| < 1$  und die Invertierbarkeit von  $(\mathcal{O}^{-1} - z\operatorname{Id})$  für alle  $z$  mit  $\operatorname{Re}(z) \leq 0$ , so dass sowohl das Spektrum von  $\mathcal{O}^{-1}$  als auch  $\{z \in \mathbb{C} \mid |r(z)\zeta| \geq 1\}$  Teilmengen der rechten Halbebene sein müssen.  $\square$

**Bemerkung 2.7.** Wie auch in [4] begründet, können wir aus Lemma 2.6 folgern, dass  $\sigma(\Delta(\zeta)/h)$  für  $h \leq h_0$  mit genügend kleinem  $h_0$  in  $\Sigma(\sigma_0)$ ,  $\sigma_0 > 0$ , liegt und damit die Faltungsgewichte  $W_n$  wohldefiniert sind.

Der folgende Satz, für dessen Beweis wir auf [4] verweisen, zeigt, dass wir mit der Runge-Kutta basierten Faltungsquadratur aus Satz 2.4 unter bestimmten Voraussetzungen an die Funktion  $g$  und deren Ableitungen ein konvergentes Verfahren der Ordnung  $\min(p, q + 1 + \mu_2)$  erhalten.

**Satz 2.8.** *Die Laplace-Transformierte  $K$  erfülle die Annahme 2.2 und für das  $A$ -stabile Runge-Kutta Verfahren mit Ordnung  $p$  und Stufenordnung  $q$  gelten die Annahmen (RK1)-(RK3). Sei  $r > \max(p - \mu_1 + 1, p, q + 1)$  und  $g \in C^r([0, T], X)$ , wobei  $g^{(l)}(0) = 0$  für  $l = 0, \dots, r - 1$ . Dann gibt es ein  $h_0 > 0$ , so dass der Fehler der Faltungsquadratur aus Satz 2.4 für  $0 < h \leq h_0$  und  $nh \leq T$  durch*

$$\|u_n - u(t_n)\|_X \leq C h^{\min(p, q+1+\mu_2)} \left( \|g^{(r)}(0)\|_X + \int_0^{t_n} \|g^{(r+1)}(\tau)\|_X d\tau \right)$$

beschränkt ist. Hierbei hängen die Konstanten  $C$  und  $h_0$  nur von den Koeffizienten des Runge-Kutta Verfahrens, den Konstanten in der Abschätzung aus Annahme 2.2 und von der Länge des Zeitintervalls ab.

## 2.3 Implementierung mit Hilfe der schnellen Fourier-Transformation

### 2.3.1 Berechnung der Faltungsgewichte

Wie in [48] vorgeschlagen wird, können wir die Faltungsgewichte  $W_n$  aus (2.3) mit Hilfe der schnellen Fourier-Transformation (FFT) berechnen. Unter Verwendung der Cauchy'schen Integralformel und mit der Überlegung in Bemerkung 2.7 stellen wir dazu die Faltungsgewichte als Wegintegrale dar, wobei über den Rand eines Ursprungskreises mit Radius  $\vartheta < 1$  integriert wird, das heißt, wir betrachten

$$h W_n = \frac{1}{2\pi i} \int_{|\zeta|=\vartheta} \zeta^{-n-1} K\left(\frac{\Delta(\zeta)}{h}\right) d\zeta, \quad n = 0, 1, \dots$$

Approximieren wir diese Darstellung mit der Trapezregel erhalten wir die Näherung

$$h W_n \approx \frac{\vartheta^{-n}}{J} \sum_{j=0}^{J-1} K\left(\frac{\Delta(\zeta_j)}{h}\right) e^{-2\pi i n j/J} \quad \text{mit} \quad \zeta_j = \vartheta e^{2\pi i j/J}. \quad (2.4)$$

Hierbei wählen wir den Parameter  $J$  wie in [46, 47] beschrieben: Nehmen wir an, dass die Werte der Laplace-Transformierten  $K$  mit einer Genauigkeit von  $\varepsilon > 0$  berechnet wurden und wir die Faltungsgewichte  $W_0, \dots, W_N$  bestimmen wollen, erhalten wir in jedem  $W_n$  einen Fehler von  $\mathcal{O}(\sqrt{\varepsilon})$  für die Wahl  $J = N$  und  $\vartheta^J = \sqrt{\varepsilon}$ .

Die Faltungsgewichte können nun durch die FFT mit  $\mathcal{O}(J \log(J))$  Operationen berechnet werden.

### 2.3.2 Lösen einer Volterra Integralgleichung

In Kapitel 5 werden wir die Runge-Kutta basierte Faltungsquadratur auch dazu verwenden, eine Integralgleichung der Form

$$u(t) = (k * g(t, u))(t), \quad t > 0 \quad (2.5)$$

zu lösen. Während die Betrachtung des Faltungsintegrals (1.1) in Satz 2.4 für  $N \in \mathbb{N}$  eine Approximation  $u_{N+1}$  mit bekannter rechter Seite liefert, hängt diese nun bei der Anwendung auf die Integralgleichung von den inneren Stufen des Runge-Kutta Verfahrens ab. Daher machen wir uns zunächst klar, wie die Erweiterung der Faltungsquadratur aus Satz 2.4 auf die Gleichung (2.5) aussieht:

Betrachte dazu wieder ein  $A$ -stabiles Runge-Kutta Verfahren, das die Annahmen (RK1)-(RK3) erfüllt und definiere  $U_N$  als den Spaltenvektor der inneren Stufen, das heißt

$$U_N = (U_{Ni})_{i=1}^s,$$

vergleiche Bemerkung 2.5. Wegen (RK3) folgt, wie bereits gesehen,  $U_{Ns} = u_{N+1}$ . Damit erhalten wir aus Satz 2.4 die Approximation der Lösung  $u$  von (2.5) zum Zeitpunkt  $t_{N+1} = (N+1)h$  aus der Lösung  $U_N$  der impliziten Gleichung

$$U_N - hW_0 g_N = h \sum_{n=1}^N W_n g_{N-n}, \quad (2.6)$$

wobei hier  $g_N = (g(t_N + c_k h, U_{Nk}))_{k=1}^s$  und die Faltungsgewichte durch (2.3) gegeben und wie im vorigen Abschnitt berechnet werden können.

Im Gegensatz zur Berechnung in Satz 2.4, in der wir die Lösung von (1.1) zu einem festen Zeitpunkt  $t_{N+1}$  ohne Vorwissen über die Lösung zu vorherigen Zeitschritten in  $\mathcal{O}(N)$  Operationen bestimmen können, müssen wir nun zur Näherung an die Lösung  $u(t_{N+1})$  der Integralgleichung (2.5) zwangsläufig die Approximationen  $U_n$ ,  $n = 0, \dots, N$  berechnen und zwischenspeichern, auch wenn wir letztendlich nur an der Lösung zum letzten Zeitschritt interessiert wären. Eine naive Implementierung benötigt hierzu  $\mathcal{O}(N^2)$  Operationen.

Im Folgenden fassen wir ein Schema zusammen, das in [32] von Hairer, Lubich und Schlichte vorgestellt wurde und es uns ermöglicht, die Auswertung mit Hilfe von schnellen Fourier-Transformationen auf  $\mathcal{O}(N \log(N)^2)$  Operationen zu reduzieren. Der Speicheraufwand bleibt jedoch derselbe wie in der naiven Implementierung und wächst weiter linear an.

Zum Verständnis des Verfahrens aus [32] ist es hilfreich, zunächst die ersten paar Schritte zu beschreiben, bevor wir näher auf die Implementierung eingehen. Sei dazu  $r \in \mathbb{N} \setminus \{1\}$  fest gewählt.

1. *Schritt:* Durch das Lösen der impliziten Gleichung (2.6) bestimmen wir die Näherungen  $U_N$  für  $N = 0, \dots, r-1$  direkt.

2. *Schritt:* Definiere die Vektoren

$$\widetilde{W} = (W_1, \dots, W_{2r-1}, 0, \dots, 0) \quad \text{und} \quad \widetilde{g} = (g_0, \dots, g_{r-1}, 0, \dots, 0),$$

wobei jeweils so viele Nullen hinzugefügt werden, dass die Länge  $L$  der Vektoren der nächstgrößeren Zweierpotenz nach  $2r-1$  entspricht. Wir bekommen nun einen Teil der rechten Seiten von (2.6) für  $N = r, \dots, 2r-1$  aus den entsprechenden Einträgen der diskreten Faltung der Vektoren  $\widetilde{W}$  und  $\widetilde{g}$ . Genauer gilt

$$(\widetilde{W} * \widetilde{g})_k = \sum_{n=0}^{r-1} W_{k-n} g_n = \sum_{n=k-r+1}^k W_n g_{k-n} \quad \text{für} \quad k = r, \dots, 2r-1.$$

Die Faltung der Vektoren können wir hierbei mit Hilfe der FFT in  $\mathcal{O}(L \log(L))$  Operationen berechnen.

3. *Schritt:* Wir bestimmen die Werte  $U_N$ ,  $N = r, \dots, 2r-1$  induktiv durch das Lösen der

impliziten Gleichung (2.6), wobei wir die rechte Seite wie folgt erhalten:

$$\sum_{n=1}^N W_n g_{N-n} = \underbrace{\sum_{n=N-r+1}^N W_n g_{N-n}}_{\text{bekannt aus Schritt 2}} + \underbrace{\sum_{n=1}^{N-r} W_n g_{N-n}}_{\text{berechne induktiv}}$$

4. *Schritt*: Wiederhole Schritt 2, wobei nun statt  $r$  der Wert  $2r$  betrachtet wird, das heißt wir erhalten Teile der rechte Seite von (2.6) für  $N = 2r - 1, \dots, 4r - 1$ .

5. *Schritt*: Wiederhole Schritt 3 zur Berechnung von  $U_N$ ,  $N = 2r, \dots, 3r - 1$ .

6. *Schritt*: Wiederhole Schritt 2, wobei nun  $\tilde{g} = (g_{2r}, \dots, g_{3r-1}, 0, \dots, 0)$  gesetzt wird.

7. *Schritt*: Berechne die Werte  $U_N$ ,  $N = 3r, \dots, 4r - 1$ , wobei die rechte Seite in (2.6) zusammengesetzt wird aus

$$\sum_{n=1}^N W_n g_{N-n} = \underbrace{\sum_{n=N-2r+1}^N W_n g_{N-n}}_{\text{bekannt aus Schritt 4}} + \underbrace{\sum_{n=N-3r+1}^{N-2r} W_n g_{N-n}}_{\text{bekannt aus Schritt 6}} + \underbrace{\sum_{n=1}^{N-3r} W_n g_{N-n}}_{\text{berechne induktiv}}.$$

Die folgenden Schritte lassen sich nun leicht aus den angegebenen Schritten herleiten. Zur besseren Übersicht ist das Verfahren in Abbildung 2.1 für den Parameter  $r = 4$  dargestellt.

Wollen wir den Wert  $U_N$  für eine Zweierpotenz  $N$  bestimmen, berechnet sich der Aufwand des Algorithmus also wie bereits angegeben als

$$\mathcal{O}(N \log(N) + 2 \left(\frac{N}{2} \log\left(\frac{N}{2}\right)\right) + 4 \left(\frac{N}{4} \log\left(\frac{N}{4}\right)\right) + \dots) + \mathcal{O}(N) = \mathcal{O}(N \log(N)^2),$$

siehe [32].

Abschließend gehen wir kurz auf unsere Implementierung des Schemas ein. Dazu greifen wir auf die bereits in Abbildung 2.1 deutlich gewordene Visualisierung der Schritte durch Dreiecke und Quadrate zurück, das heißt, wir bezeichnen mit der Funktion  $\text{Dreieck}(k,1)$  die direkten Auswertung der Werte  $U_n$ ,  $n = k, \dots, l$  über die Gleichung (2.6). Mit  $\text{Quadrat}(j,k,1)$  ist die schnelle Fourier-Transformation der Vektoren

$$\widetilde{W} = (W_1, \dots, W_j, 0, \dots, 0) \quad \text{und} \quad \widetilde{g} = (g_k, \dots, g_l, 0, \dots, 0)$$

gemeint, vergleiche Schritt 2. Jedem Quadrat geben wir nun der Reihe seiner Ausführung nach eine Nummer, siehe Abbildung 2.2.

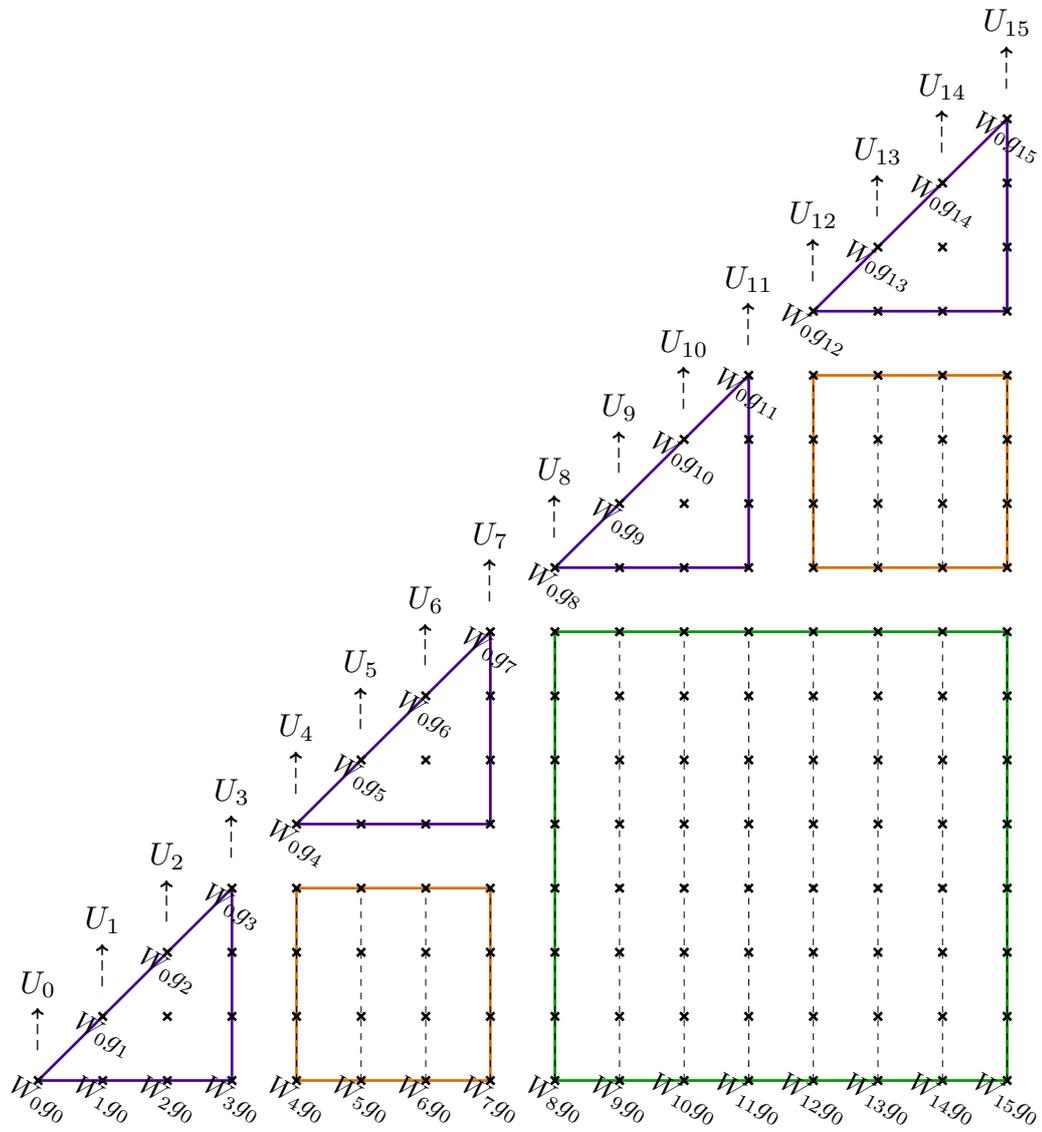


Abbildung 2.1: Darstellung des Verfahrens für den Parameter  $r = 4$ . Berechnet werden hier die Werte  $U_N$ ,  $N = 0, \dots, 15$  über die Gleichung (2.6). Die Schritte 1, 3, 5 und 7 werden durch die Dreiecke, die Schritte 2 und 6 durch die orangen Quadrate und Schritt 4 durch das grüne Quadrat dargestellt. Mit der Fourier-Transformation werden hier jeweils die spaltenweisen Summen berechnet, was durch die gestrichelten Linien angedeutet wird.

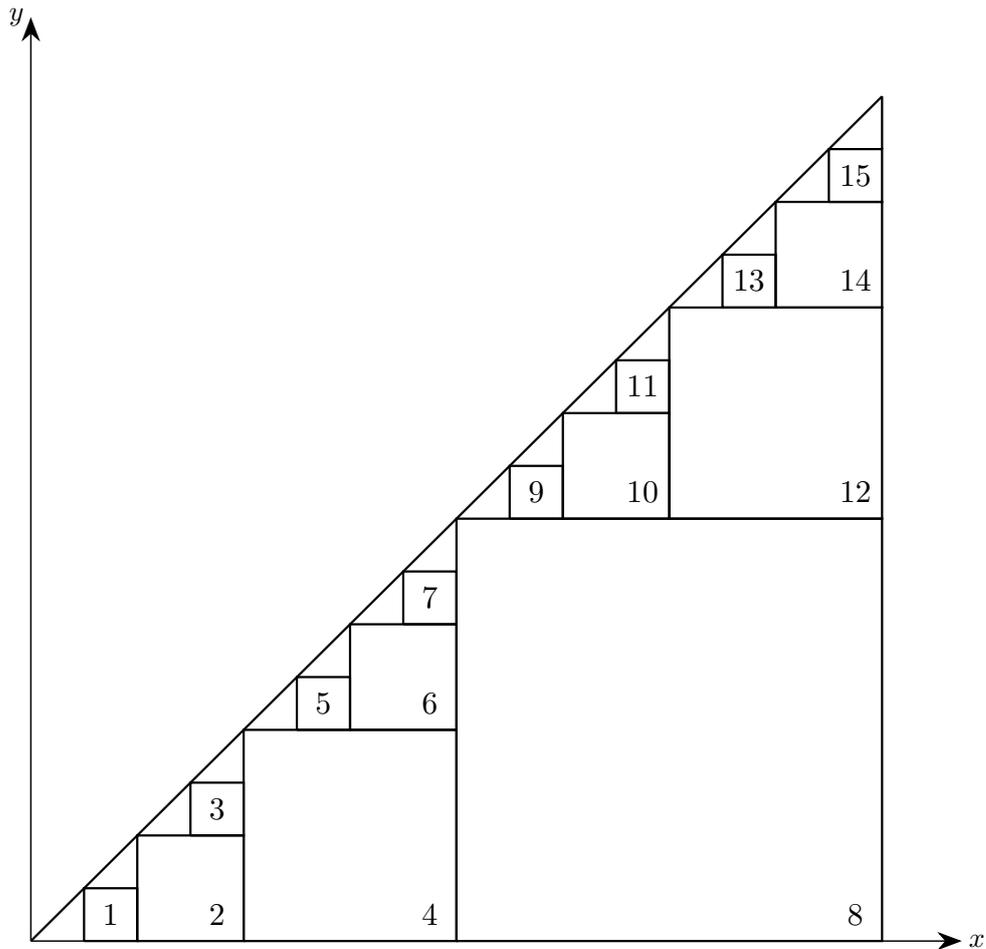


Abbildung 2.2: Reihenfolge der Ausführung der Funktionen `Quadrat`.

Mit Hilfe des folgenden Algorithmus 2.1 können wir aus dem Index  $i$  eines Quadrats seine Anordnung in  $x$ - und  $y$ -Richtung sowie seine Größe bestimmen.

---

**Algorithmus 2.1** : Anordnung des Quadrats mit Index  $i$  in  $x$ - und  $y$ -Richtung

---

**Input** : Index  $i$ ,

Parameter  $r \in \setminus\{1\}$

**Output** : Eckpunkte  $x = [x_1, x_2]$  des Quadrats in  $x$ -Richtung,

Eckpunkte  $y = [y_1, y_2]$  des Quadrats in  $y$ -Richtung,

Wert  $L$ , wobei die Größe des Quadrats  $2^{Lr} \times 2^{Lr}$  ist

Initialisiere  $L = 0$ ;

**while**  $\text{mod}(i, 2) = 0$  :

$i = i/2$ ;  
 $L = L + 1$ ;

$x = [ir, ir + 2^{Lr} - 1]$ ;

$y = [ir - 2^{Lr}, ir - 1]$ ;

---

Damit lässt sich das Verfahren zur Berechnung der Werte  $U_n$ ,  $n = 0, \dots, N$  wie folgt durchführen:

---

**Algorithmus 2.2 :** Schnelle Faltung nach [32] für die Gleichung (2.6)

---

**Input :** Faltungsgewichte  $W_n$ ,  $n = 0, \dots, N$ ,

Parameter  $r \in \mathbb{N} \setminus \{1\}$

**Output :** Werte  $U_n$ ,  $n = 0, \dots, N$

Initialisiere mit **Dreieck**(0,  $r-1$ );

**for**  $i = 1 : (2^{\lceil \log_2(N+1/r) \rceil} - 1)$  : // maximale Anzahl der benötigten Quadrate

    Bestimme die Eckpunkte  $x, y$  sowie den Wert  $L$  mit Algorithmus 2.1;

**if**  $x_2 \leq N$  : // volle Faltung im Quadrat

        |  $\nu = x_2 - x_1$ ;

**else if**  $x_1 \leq N$  und  $N < x_2$  : // benötige nur einen Teil der Faltung

        |  $\nu = N - x_1$

**else** // Quadrat wird nicht mehr durchlaufen

        | **break**;

    Berechne **Quadrat**( $2^L r + \nu$ ,  $y_1$ ,  $y_2$ ) und speichere die Einträge  $2^L r$  bis  $2^L r + \nu$ ;

    Berechne **Dreieck**( $x_1$ ,  $x_1 + \min(\nu, r - 1)$ );

---



## KAPITEL 3

---

# HELMHOLTZ-GLEICHUNG IM PERIODISCHEN WELLENLEITER MIT LOKALEM DEFECT

Nachdem wir nun Methoden zur Faltungsquadratur angegeben haben, werden wir in diesem Kapitel ein zeitharmonisches Problem in Form der Helmholtz-Gleichung in einem – bis auf einen lokalen Defekt – periodischen Wellenleiter betrachten und damit die Basis für die Anwendung der Faltungsalgorithmen legen, die wir in den darauffolgenden Kapiteln besprechen werden.

Die Motivation für die Betrachtung eines solchen periodischen Wellenleiters liegt in dessen besonderer Bedeutung für viele moderne Anwendungen, beispielsweise in Bereichen der Akustik oder der Mikro- und Nanooptik, siehe zum Beispiel [39, 41, 59]. Anders als in theoretischen Überlegungen sind periodische Medien in realen Anwendungen allerdings nicht perfekt periodisch, sondern beinhalten beschränkte Bereiche, in denen die periodische Struktur unterbrochen ist. Das kann beispielsweise bei photonischen Kristallen beobachtet werden, wobei die gestörten Bereiche im Vergleich zur Größe des Ausbreitungsgebiets jedoch sehr klein sind [59]. Es ist ebenfalls möglich, mit Hilfe von absichtlich eingebrachten Defekten, das Medium so zu verändern, dass sich bestimmte Wellenfrequenzen besonders gut oder schlecht ausbreiten.

Zur Herleitung numerischer Methoden für die Berechnung der Lösung einer linearen Differentialgleichung im Wellenleiter besteht ein bekannter Ansatz darin, die Berechnung auf das Gebiet zu reduzieren, das die lokale Störung enthält. Unter Ausnutzung der periodischen Struktur lassen sich transparente Randbedingungen in Form von Neumann-zu-Dirichlet Operatoren (kurz: NtD-Operatoren) auf dem Rand des beschränkten Gebiets bestimmen, anhand derer die Lösung im gestörten Bereich berechnet werden kann. Wir werden dazu einer Idee von Fliss und Joly folgen, die zuerst in [40] beschrieben und in [19–21, 23] weiterentwickelt wurde. In dieser genügt zur Konstruktion der NtD-Operatoren die Betrachtung lokaler Probleme auf einer Zelle.

Im ersten Teil dieses Kapitels werden wir die wichtigsten Resultate dieser Methode zusammenfassen, wobei wir die für die Erweiterung auf komplexe Frequenzen fehlenden Beweise ergänzen werden. Hierbei verfassen wir die Ergebnisse allgemein für den zweidimensionalen Wellenleiter.

Im zweiten Teil gehen wir kurz auf den eindimensionalen Fall ein und berechnen die hergeleiteten Operatoren für je ein Beispiel eines homogenen und eines inhomogenen Wellenleiters sowohl im ortskontinuierlichen als auch im ortsdiskreten Fall. Hierbei sind wir insbesondere am Abfallverhalten des NtD-Operators interessiert.

## 3.1 Bestimmung der transparenten Randbedingungen

### 3.1.1 Modellproblem

Unserem Modell zugrunde liegt ein zweidimensionaler Wellenleiter  $\Omega$ , der unbeschränkt in die Ausbreitungsrichtung  $x$ , jedoch beschränkt in  $y$ -Richtung ist, genauer  $\Omega = \mathbb{R} \times (0, b)$ . Auf dem Wellenleiter untersuchen wir zunächst<sup>1</sup> für  $s \in \mathbb{C}$  die Lösung  $u_s : \Omega \rightarrow \mathbb{C}$  der Helmholtz-Gleichung

$$\begin{cases} -\Delta u(x, y) + s \rho(x, y) u(x, y) = f(x, y) & \text{in } \Omega, \\ u(x, y) = 0 & \text{auf } \partial\Omega, \end{cases} \quad (\mathbf{H})$$

wobei wir über die Funktion  $\rho$  die folgenden Annahmen treffen:

(H1)  $\rho \in L^\infty(\Omega, \mathbb{R})$ , wobei es  $\rho_{\min}, \rho_{\max} \in \mathbb{R}^+$  gibt, so dass  $0 < \rho_{\min} \leq \rho \leq \rho_{\max}$ , und

(H2) es gibt Werte  $a^\pm \in \mathbb{R}$ ,  $L^\pm > 0$ , eine auf dem beschränkten Gebiet

$$\Omega_0 := \{(x, y) \in \Omega \mid a^- < x < a^+\}$$

definierte Funktion  $\rho_0$  sowie zwei in  $x$ -Richtung periodische Funktionen  $\rho^\pm$  mit Periodenlänge  $L^\pm$ , also

$$\begin{aligned} \rho^+(x, y) &= \rho^+(x + L^+, y) \quad \forall (x, y) \in \Omega^+ := \{(x, y) \in \Omega \mid a^+ < x\} \quad \text{und} \\ \rho^-(x, y) &= \rho^-(x - L^-, y) \quad \forall (x, y) \in \Omega^- := \{(x, y) \in \Omega \mid x < a^-\}, \end{aligned}$$

so dass  $\rho$  stückweise gegeben ist als

$$\rho(x, y) = \begin{cases} \rho^-(x, y) & \text{in } \Omega^-, \\ \rho_0(x, y) & \text{in } \Omega_0, \\ \rho^+(x, y) & \text{in } \Omega^+. \end{cases}$$

---

<sup>1</sup>Der Bereich wird in Abschnitt 3.1.3 weiter eingeschränkt.

Damit modelliert die Funktion  $\rho$  einen lokalen Defekt in der ansonsten periodischen Struktur, der auf das Gebiet  $\Omega_0$  beschränkt ist.

Von der Quellfunktion  $f$  nehmen wir außerdem an:

$$(H3) \quad f \in L^2(\Omega) \text{ und}$$

$$(H4) \quad \text{supp}(f) \subset \Omega_0.$$

Um die spezielle Struktur der Helmholtz-Gleichung unter den getroffenen Annahmen ausnutzen zu können, führen wir die folgenden Notationen ein:

Unter Verwendung der Periodenlängen  $L^\pm > 0$  können wir den Abschluss der „Halb-Wellenleiter“ in der Form  $\overline{\Omega^\pm} = \bigcup_{j=0}^{\infty} \overline{C_j^\pm}$  schreiben, wobei die Periodizitätszellen  $C_j^\pm$  gegeben sind durch

$$\begin{aligned} C_j^+ &:= \{(x, y) \in \Omega \mid a^+ + jL^+ < x < a^+ + (j+1)L^+\} \quad \text{und} \\ C_j^- &:= \{(x, y) \in \Omega \mid a^- - (j+1)L^- < x < a^- - jL^-\}. \end{aligned}$$

Weiterhin bezeichnen wir jeweils für  $j \in \mathbb{N}_0$  die linken Ränder der Zellen  $C_j^+$  beziehungsweise die rechten Ränder der Zellen  $C_j^-$  mit

$$\Gamma_j^\pm := \{(x, y) \in \Omega \mid x = a^\pm \pm jL^\pm\}.$$

Nun ist klar, dass alle Zellen  $C_j^\pm$ ,  $j \in \mathbb{N}$ , mit der Zelle  $C_0^\pm$  identifiziert werden können, genauso wie jeder Rand  $\Gamma_j^\pm$ ,  $j \in \mathbb{N}$ , mit  $\Gamma_0^\pm$ . Wie bereits erwähnt, werden wir uns in den folgenden Unterkapiteln diese Identifizierung zu Nutze machen. In Abbildung 3.1 stellen wir zur besseren Übersicht einen Wellenleiter mit der verwendeten Notation dar.

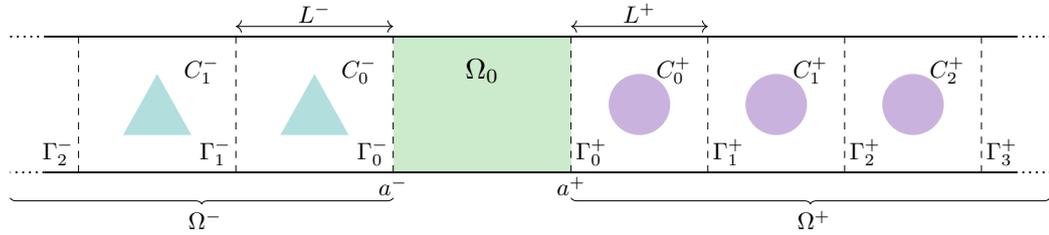


Abbildung 3.1: Zweidimensionaler Wellenleiter mit verwendeter Notation. Die periodischen Funktionen  $\rho^\pm$  auf den Halb-Wellenleitern werden hier durch die Dreiecke beziehungsweise Kreise symbolisiert.

**Bemerkung 3.1.** Im Fall  $\rho^+ = \rho^- \equiv c > 0$  sprechen wir auch von einem *homogenen*, andernfalls von einem *inhomogenen* Wellenleiter. Für die Betrachtung der zeitharmonischen Helmholtz-Gleichung in verschiedenen homogenen Medien in  $\mathbb{R}^2$  gibt es bereits zahlreiche Ansätze, die die Berechnung der Lösung auf den Bereich mit lokaler Störung zurückführt; für eine Zusammenfassung dieser Methoden siehe [21] beziehungsweise [6] für die konkrete Anwendung auf einen Wellenleiter. Da viele dieser Methoden die homogene Struktur des Außenraums verwenden, lassen sie sich nicht auf periodische Medien erweitern.

Das hier vorgestellte Verfahren für einen periodischen Wellenleiter ist dahingehend allgemeiner, da es uns auch die Betrachtung eines homogenen Wellenleiters erlaubt (die Annahmen (H1) und (H2) sind offensichtlich für  $\rho \equiv c > 0$  erfüllt). Des Weiteren liefert es eine Grundlage zur Betrachtung des gesamten zweidimensionalen Raums, siehe hierzu [21, 23], auf die wir jedoch nicht weiter eingehen.

**Definition 3.2.** Die komplexen Hilberträume, die wir im Folgenden benötigen werden, geben wir an dieser Stelle kurz an. Sei dazu  $G \subseteq \Omega$  offen.

- Im Raum  $L^2(G)$  definieren wir das von  $\rho$  abhängige Skalarprodukt als

$$\langle u, v \rangle_{0,\rho} := \int_G \rho u \bar{v} \, d(x, y)$$

und schreiben die dadurch induzierte Norm als  $\|\cdot\|_{0,\rho}$ , gegebenenfalls mit der entsprechenden Menge  $G$  im Index falls erforderlich. Wegen (H1) ist die Norm  $\|\cdot\|_{0,\rho}$  äquivalent zur üblicherweise auf  $L^2(G)$  betrachteten Norm  $\|\cdot\|_0 := \|\cdot\|_{0,1}$ .

- Den Sobolevraum  $H^1(G) := \left\{ u : G \rightarrow \mathbb{C} \mid u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y} \in L^2(G) \right\}$  versehen wir mit dem ebenfalls von  $\rho$  abhängigen Skalarprodukt

$$\langle u, v \rangle_{1,\rho} := \int_G \rho u \bar{v} \, d(x, y) + \int_G \nabla u \cdot \overline{\nabla v} \, d(x, y) \quad \forall u, v \in H^1(G).$$

Damit können wir die zugehörige Norm als  $\|u\|_{1,\rho}^2 = \|u\|_{0,\rho}^2 + |u|_1^2$  schreiben, wobei  $|u|_1 := \|\nabla u\|_0$  eine Seminorm auf  $H^1(G)$  ist.

- Den Sobolevraum mit verallgemeinerten Nullrandwerten  $H_0^1(G)$  definieren wir als Vervollständigung des Raums  $\{u \in C^\infty(\overline{G}) \mid u = 0 \text{ auf } \partial G\}$  bezüglich der  $\|\cdot\|_{1,\rho}$ -Norm.
- Wollen wir die verallgemeinerten Nullrandwerte auf die Menge  $\partial\Omega \cap \partial G$  einschränken, betrachten wir den  $H^1(G)$ -Unterraum  $H_{0,\Omega}^1(G)$ , den wir als Vervollständigung des Raums  $\{u \in C^\infty(\overline{G}) \mid u = 0 \text{ auf } \partial\Omega \cap \partial G\}$  bezüglich der  $\|\cdot\|_{1,\rho}$ -Norm verstehen.
- Mit  $H_0^1(\Delta, G)$  bezeichnen wir die Teilmenge von  $H_0^1(G)$ , für die zusätzlich  $\Delta u \in L^2(G)$  gilt. Analog ist  $H_{0,\Omega}^1(\Delta, G) := \{u \in H_{0,\Omega}^1(G) \mid \Delta u \in L^2(G)\}$ .
- Ist  $\Gamma$  ein Teil von  $\partial G$ , betrachten wir den Hilbertraum

$$H^{1/2}(\Gamma) := \left\{ \psi \in L^2(\Gamma) \mid \int_\Gamma \int_\Gamma \frac{|\psi(z_1) - \psi(z_2)|^2}{|z_1 - z_2|^2} \, d\sigma(z_1) \, d\sigma(z_2) < \infty \right\}$$

und definieren die zugehörige Norm durch

$$\|\psi\|_{1/2} := \left( \|\psi\|_{0,\Gamma}^2 + \int_\Gamma \int_\Gamma \frac{|\psi(z_1) - \psi(z_2)|^2}{|z_1 - z_2|^2} \, d\sigma(z_1) \, d\sigma(z_2) \right)^{\frac{1}{2}},$$

wobei  $\|\psi\|_{0,\Gamma}^2 = \int_\Gamma |\psi|^2 \, d\sigma$ , vergleiche [30, Def. 1.3.2.1].

- Die Vervollständigung von  $\{u \in C^\infty(\bar{\Gamma}) \mid u = 0 \text{ auf } \partial\Gamma\}$  bezüglich der  $\|\cdot\|_{1/2}$ -Norm bezeichnen wir als  $H_0^{1/2}(\Gamma)$ .
- Der Dualraum von  $H_0^{1/2}(\Gamma)$  wird mit  $H^{-1/2}(\Gamma)$  bezeichnet und mit der Norm

$$\|\varphi\|_{-1/2} = \sup_{\psi \neq 0} \frac{|\langle \varphi, \psi \rangle_\Gamma|}{\|\psi\|_{1/2}}$$

für  $\varphi \in H^{-1/2}(\Gamma)$  und  $\psi \in H_0^{1/2}(\Gamma)$  ausgestattet, wobei  $\langle \varphi, \psi \rangle_\Gamma = \varphi(\psi)$  die duale Paarung ist.

Unter bestimmten Voraussetzungen an die Teilmenge  $G$ , die sich aus der konkreten Form des betrachteten Wellenleiters ergeben, erhalten wir die folgenden Sätze über die Dirichlet- beziehungsweise über die Neumann-Spur:

**Satz 3.3.** *Es sei  $G = I \times (0, b) \subset \Omega$  mit endlichem Intervall  $I = (a_0, a_1)$  und bezeichne mit  $\Gamma_0$  den linken Rand beziehungsweise mit  $\Gamma_1$  den rechten Rand des Rechtecks  $G$ , das heißt  $\Gamma_0 = \{a_0\} \times (0, b)$  und  $\Gamma_1 = \{a_1\} \times (0, b)$ . Dann hat die Abbildung*

$$u \mapsto u|_{\Gamma_j},$$

die auf  $\{u \in C^\infty(\bar{G}) \mid u = 0 \text{ auf } \partial\Omega \cap \partial G\}$  definiert ist, für jedes  $j = 0, 1$  eine eindeutige stetige Fortsetzung

$$\gamma_{D,j} : H_{0,\Omega}^1(G) \rightarrow H_0^{1/2}(\Gamma_j).$$

*Beweis.* Der Satz ist eine direkte Anwendung von Theorem 1.5.2.1 aus [30]. □

**Satz 3.4.** *Es seien  $G$  und  $\Gamma_j$ ,  $j = 0, 1$ , wie in Satz 3.3. Dann haben die Abbildungen*

$$u \mapsto -\frac{\partial u}{\partial x}\Big|_{\Gamma_0} \quad \text{und} \quad u \mapsto \frac{\partial u}{\partial x}\Big|_{\Gamma_1},$$

die auf  $\{u \in C^\infty(\bar{G}) \mid u = 0 \text{ auf } \partial\Omega \cap \partial G\}$  definiert sind, eindeutige stetige Fortsetzungen

$$\gamma_{N,j} : H_{0,\Omega}^1(\Delta, G) \rightarrow H^{-1/2}(\Gamma_j), \quad j = 0, 1.$$

*Beweis.* Der Beweis des Satzes läuft analog zum Beweis von Theorem 1.5.3.10 aus [30]. Hier sei noch zu bemerken, dass der Raum  $V$ , der anfangs im genannten Beweis definiert wird, an unseren Fall mit verallgemeinerten Nullrandwerten angepasst werden muss. Die Wahl  $V = \{v \in H_{0,\Omega}^1(G) \mid \gamma_{D,k}(v) = 0 \text{ für } k \neq j\}$  liefert uns unsere Formulierung von Satz 3.4. □

**Bemerkung 3.5.** Der besseren Lesbarkeit halber werden wir im Folgenden bei der Betrachtung der Dirichlet-Spur von  $u \in H_{0,\Omega}^1(G)$  nur noch  $u|_{\Gamma_j}$  schreiben und meinen damit insbesondere die Abbildung  $\gamma_{D,j}(u)$ , auch wenn  $u \in H_{0,\Omega}^1(G)$  nicht stetig ist. Die gleiche Vereinfachung in der Notation verwenden wir auch für die Neumann-Spur  $\gamma_{N,j}$ .

Wie wir im nächsten Lemma sehen werden, gilt die Ungleichung von Poincaré unter den Annahmen (H1) und (H2) auch für die  $\rho$ -abhängigen Normen.

**Lemma 3.6.** *Sei  $G \subseteq \Omega$ , dann gelten für alle  $u \in H_0^1(G)$  die Ungleichungen*

$$\|u\|_{0,\rho} \leq C_1 |u|_1 \quad \text{und} \quad \|u\|_{1,\rho} \leq C_2 |u|_1$$

mit Konstanten  $C_1, C_2 > 0$ , die vom Gebiet  $\Omega$  und der Funktion  $\rho$  abhängt.

*Beweis.* Da  $G \subset \Omega$  und damit insbesondere in die  $y$ -Richtung beschränkt ist, kennen wir nach Proposition IV.7.1 aus [13] die Poincaré'sche Ungleichung bereits für die typischerweise betrachteten Normen in Sobolevräumen, das heißt mit unserer Notation gilt:

$$\exists C > 0 \text{ s.d. } \|u\|_{0,1} \leq C |u|_1 \quad \forall u \in H_0^1(G)$$

Mit Hilfe der Annahme (H1) folgt direkt

$$\|u\|_{0,\rho}^2 \leq \rho_{\max} \|u\|_{0,1}^2 \leq \rho_{\max} C^2 |u|_1^2$$

und daraus

$$\|u\|_{1,\rho}^2 = \|u\|_{0,\rho}^2 + |u|_1^2 \leq (\rho_{\max} C^2 + 1) |u|_1^2.$$

□

**Bemerkung 3.7.** Mit Lemma 3.6 lässt sich sofort einsehen, dass  $|\cdot|_1$  eine Norm auf  $H_0^1(G)$  beschreibt, die äquivalent zur  $\|\cdot\|_{1,\rho}$  ist, vergleiche dazu auch [13, Kor. IV.7.2].

### 3.1.2 Eigenschaften des Spektrums

Nachdem wir nun die notwendigen Räume definiert haben, können wir mit der Untersuchung des Spektrums der Helmholtz-Gleichung auf dem periodischen, lokal gestörten Wellenleiter unter den Annahmen (H1)-(H4) beginnen. Das Eigenwertproblem, das sich aus (H) ergibt, können wir nun ausdrücken als:

$$\text{Finde } \omega \in \mathbb{C} \text{ s.d. } \exists u \in H_0^1(\Delta, \Omega), u \neq 0 \text{ mit } -\Delta u = \omega \rho u$$

beziehungsweise äquivalent dazu

$$\text{Finde } \omega \in \mathbb{C} \text{ s.d. } \exists u \in H_0^1(\Delta, \Omega), u \neq 0 \text{ mit } Au = \omega u, \quad (3.1a)$$

wobei der Operator  $A$  definiert ist als

$$A : D(A) := H_0^1(\Delta, \Omega) \subset L^2(\Omega) \rightarrow L^2(\Omega), \quad u \mapsto Au := -\rho^{-1}\Delta u. \quad (3.1b)$$

**Satz 3.8.** *Der Operator  $A$  aus (3.1b) ist selbstadjungiert.*

*Beweis.* Wir verwenden hier [8, Prop. 7.6] und zeigen dazu, dass  $A$  symmetrisch und maximal monoton im Sinne der Definition aus [8, Kap. 7.1] ist.

Da  $C_0^\infty(\bar{\Omega}) \subset H_0^1(\Delta, \Omega)$ , ist  $D(A)$  dicht definiert in  $L^2(\Omega)$ . Mit Hilfe der Green'schen Formeln erhalten wir

$$\begin{aligned} \langle Au, v \rangle_{0,\rho} &= \int_{\Omega} \rho (-\rho^{-1}\Delta u) \bar{v} \, d(x, y) = \int_{\Omega} \nabla u \cdot \overline{\nabla v} \, d(x, y) \\ &= \int_{\Omega} -\rho u \overline{\rho^{-1}\Delta v} \, d(x, y) = \langle u, Av \rangle_{0,\rho} \quad \forall u, v \in D(A), \end{aligned}$$

das heißt  $A$  ist symmetrisch. Wegen

$$\langle Au, u \rangle_{0,\rho} = \int_{\Omega} |\nabla u|^2 \, d(x, y) \geq 0 \quad (3.2)$$

ist der Operator  $A$  für alle  $u \in D(A)$  positiv.

Um  $\text{ran}(A + \text{Id}) = L^2(\Omega)$  zu zeigen, betrachten wir die zu  $(A + \text{Id})u = \tilde{f}$  äquivalente Gleichung  $(M_\rho - \Delta)u = f$ , wobei  $M_\rho$  der von  $\rho$  induzierte Multiplikationsoperator ist, für beliebiges  $\tilde{f} \in L^2(\Omega)$  und damit nach (H1) für beliebiges  $f := \rho\tilde{f} \in L^2(\Omega)$ .

Zum Operator  $M_\rho - \Delta$  assoziieren wir die Sesquilinearform

$$a_1 : H_0^1(\Omega) \times H_0^1(\Omega) \rightarrow \mathbb{C}, \quad (u, v) \mapsto a_1(u, v) := \langle u, v \rangle_{1,\rho},$$

für die gilt:

- $a_1$  ist stetig, da mit der Cauchy-Schwarz'schen Ungleichung folgt

$$\begin{aligned} |a_1(u, v)| &\leq \int_{\Omega} |\nabla u| |\nabla v| + |\sqrt{\rho} u| |\sqrt{\rho} v| \, d(x, y) \\ &\leq \|u\|_1 \|v\|_1 + \|u\|_{0,\rho} \|v\|_{0,\rho} \leq 2 \|u\|_{1,\rho} \|v\|_{1,\rho}; \end{aligned}$$

- $a_1$  ist wegen  $a_1(u, u) = \overline{a_1(u, u)}$  hermitesch;
- $a_1$  ist  $H_0^1(\Omega)$ -elliptisch, da  $\text{Re}(a_1(u, u)) = a_1(u, u) = \|u\|_{1,\rho}^2$ .

Des Weiteren betrachten wir zu  $f \in L^2(\Omega)$  die Linearform

$$\ell_f : H_0^1(\Omega) \rightarrow \mathbb{C}, \quad v \mapsto \ell_f(v) := \int_{\Omega} f \bar{v} \, d(x, y),$$

die wegen

$$|\ell_f(v)| \leq \|f\|_0 \|v\|_0 \leq \rho_{\min}^{-1} \|f\|_{0,\rho} \|v\|_{0,\rho} \leq C \|v\|_{1,\rho}$$

stetig auf  $H_0^1(\Omega)$  ist. Nach dem Satz von Lax-Milgram, siehe beispielsweise [13], existiert nun für alle  $v \in H_0^1(\Omega)$  genau eine Lösung  $u \in H_0^1(\Omega)$  des Problems  $a_1(u, v) = \ell_f(v)$  beziehungsweise anders formuliert, es existiert genau eine schwache Lösung  $u \in H_0^1(\Omega)$  des äquivalenten Problems  $(A + \text{Id})u = \tilde{f}$ .

Betrachten wir die Testfunktion  $v \in C_0^\infty(\bar{\Omega})$  folgt mit dem Fundamentallemma der Variationsrechnung, dass für alle  $\tilde{f} \in L^2(\Omega)$  ein  $u \in H_0^1(\Omega)$  existiert, so dass  $Au + u = \tilde{f}$  fast überall gilt. Da sowohl  $\tilde{f} \in L^2(\Omega)$  als auch  $u \in L^2(\Omega)$ , erhalten wir  $Au \in L^2(\Omega)$  und damit unter Ausnutzung von  $\rho_{\min}^{-1} \geq \rho^{-1} > 0$  sofort  $\Delta u \in L^2(\Omega)$ . Nach [8, Prop. 7.6] ist damit die Selbstadjungiertheit des Operators  $A$  gezeigt.  $\square$

Mit Hilfe von Satz 3.8 bekommen wir nun das folgende Resultat, das im Falle eines lokal gestörten homogenen Wellenleiters unter Dirichlet-Randbedingungen bereits wohl bekannt ist, siehe [53].

**Korollar 3.9.** *Das Spektrum  $\sigma(A)$  des Operators  $A : D(A) \rightarrow L^2(\Omega)$  aus (3.1) ist eine Teilmenge von  $\mathbb{R}^+ = (0, \infty)$ .*

*Beweis.* Da  $A$  nach Satz 3.8 selbstadjungiert ist, folgt  $\sigma(A) \subset \mathbb{R}$  (siehe zum Beispiel [67, Satz VII.2.16]), was sich auf Grund der Abschätzung (3.2) direkt auf  $\sigma(A) \subset [0, \infty)$  reduzieren lässt. Dass 0 in der Resolventenmenge  $\rho(A)$  enthalten ist, begründen wir analog zum Beweis von Satz 3.8, wobei wir hier die zu  $Au = \tilde{f}$  äquivalente Gleichung  $-\Delta u = f$  für beliebiges  $f \in L^2(\Omega)$  betrachten:

Zum Operator  $-\Delta$  assoziieren wir die Sesquilinearform  $a_0 : H_0^1(\Omega) \times H_0^1(\Omega) \rightarrow \mathbb{C}$  mit

$$a_0(u, v) := \int_{\Omega} \nabla u \cdot \bar{\nabla} v \, d(x, y),$$

die offensichtlich wieder stetig und hermitesch ist. Die  $H_0^1(\Omega)$ -Elliptizität sehen wir mit der Poincaré'schen Ungleichung aus Lemma 3.6. Mit dieser gilt

$$\text{Re}(a_0(u, u)) = a_0(u, u) = |u|_1^2 \geq C \|u\|_{1,\rho}^2 \quad \forall u \in H_0^1(\Omega)$$

mit einer Konstante  $C > 0$ . Damit existiert nach dem Satz von Lax-Milgram für alle  $v \in H_0^1(\Omega)$  genau ein  $u \in H_0^1(\Omega)$  des Problems  $a_0(u, v) = \ell_f(v)$ , mit  $\ell_f$  aus dem vorherigen Beweis.

Wie zuvor folgt mit der Betrachtung der Testfunktionen aus  $C_0^\infty(\bar{\Omega})$ , dass  $Au = \tilde{f}$  fast überall in  $\Omega$  gilt, wobei  $Au$  und damit  $\Delta u$  wegen  $\tilde{f} \in L^2(\Omega)$  und (H1) selbst wieder in  $L^2(\Omega)$  liegt. Damit ist  $A$  bijektiv und folglich  $\sigma(A) \subset \mathbb{R}^+$ .  $\square$

Der Fall  $\omega \in \mathbb{R}^+$  in der Eigenwertgleichung (3.1) entspricht dem Fall  $s \in (-\infty, 0)$  im Helmholtz-Problem (H). Falls  $-s$  im Spektrum des Operators  $A$  liegt, ist es möglich, dass sich in  $\Omega$  Wellen ausbreiten, die im Unendlichen nicht verschwinden, vergleiche [22]. Für die Bestimmung einer Lösung des Problems (H), die auch in der „physikalischen“ Interpretation sinnvoll ist, werden daher zusätzliche Bedingungen benötigt, die deren Existenz und Eindeutigkeit garantieren.

Im Fall eines homogenen Wellenleiters bestehen physikalisch sinnvolle Lösungen aus „ausgehenden“ Wellen und können beispielsweise unter Hinzunahme der Sommerfeld’schen Abstrahlbedingungen eindeutig bestimmt werden [34]. Was jedoch im periodischen Fall als physikalisch sinnvolle Lösung bezeichnet werden kann, ist weniger intuitiv. Der Ansatz des *limiting absorption principles*, der in [40] und den nachfolgenden Papern der Autoren für den periodischen Fall verfolgt wird, stellt eine Möglichkeit dar, eine eindeutige und sinnvolle Lösung zu definieren. Angewendet auf den homogenen Wellenleiter führt dieser Ansatz wieder zu den beschriebenen „ausgehenden“ Wellen [53].

Für einen perfekt periodischen Wellenleiter, das heißt unter der Vereinfachung der Annahme (H2) dahingehend, dass  $\rho^+ = \rho^- \forall x \in \Omega$  und entsprechend  $L^+ = L^- := L$  gesetzt wird, sowie unter der Forderung, dass  $\Omega_0$  der Größe einer Periodizitätszelle entspricht, wird in [22] eine abzählbare Menge  $\sigma_0(A)$  an Eigenwerten bestimmt, für die es jedoch nicht möglich ist, die oben erwähnten zusätzlichen Bedingungen und damit eine physikalisch sinnvolle Lösung zu finden. Hierbei ist mindestens das Punktspektrum  $\sigma_p(A)$  des zugehörigen Operators  $A$  in der Menge  $\sigma_0(A)$  enthalten, das gegebenenfalls leer sein kann. Die Abhängigkeit des Punktspektrums  $\sigma_p(A)$  von der Funktion  $\rho$  beziehungsweise dem Defekt ist hierbei noch nicht vollständig untersucht. Nach [68] gibt es beispielsweise lokale Störungen eines homogenen Wellenleiters, für die das Punktspektrum  $\sigma_p(-\Delta)$  nicht leer ist. Wie daher die Menge  $\sigma_0(A)$  für den periodischen Wellenleiter mit lokaler Störung aussieht, ist derzeit eine noch offene Fragestellung. In [22] wird vermutet, dass es sich um eine abzählbare Teilmenge handelt, deren Häufungspunkte sich an den Rändern des essentiellen Spektrums befinden, vergleiche hierzu [19] für den Fall mit Blochperiodischen Randbedingungen auf  $\partial\Omega$  und [22] für Neumann-Randbedingungen.

### 3.1.3 Existenz und Eindeutigkeit einer Lösung im Wellenleiter

Da wir in den Kapiteln 4 und 5 das Helmholtz-Problem (H) über die Laplace-Transformation zweier zeitabhängiger Differentialgleichungen erhalten werden, ist es in unserer Betrachtung legitim, den Fall  $s \in (-\infty, 0)$  auszuschließen. Damit vermeiden wir auch das Bestimmen der zusätzlichen Bedingungen, das im vorherigen Abschnitt umrissen wurde. Wie wir in der Anwendung in den folgenden Kapiteln außerdem sehen werden, ist der Fall  $s = 0$  für uns nicht von besonderem Interesse. Wir werden daher ab jetzt die zusätzliche Annahme

$$(H5) \quad s \in \mathbb{C} \setminus \mathbb{R}_0^+ = \mathbb{C} \setminus (-\infty, 0]$$

treffen und unter dieser Annahme die Existenz und Eindeutigkeit einer Lösung von **(H)** in  $H_0^1(\Delta, \Omega)$  zeigen.

**Satz 3.10.** *Unter den Annahmen (H1)-(H5) gibt es für jedes  $s$  eine eindeutige Lösung  $u_s \in H_0^1(\Delta, \Omega)$  des Helmholtz-Problems **(H)**.*

*Beweis.* Die Aussage haben wir bereits in Satz 3.8 für den Fall  $s = 1$  gezeigt. Hier erweitern wir den Beweis auf die Werte  $s$ , die (H5) erfüllen<sup>2</sup>. Zum Operator  $sM_\rho - \Delta$  assoziieren wir die Sesquilinearform  $a_s : H_0^1(\Omega) \times H_0^1(\Omega) \rightarrow \mathbb{C}$  mit

$$a_s(u, v) := \int_{\Omega} \nabla u \cdot \overline{\nabla v} + s \rho u \bar{v} \, d(x, y).$$

Mit der Cauchy-Schwarz'schen Ungleichung erhalten wir

$$|a_s(u, v)| \leq (1 + |s|) \|u\|_{1,\rho} \|v\|_{1,\rho}$$

und damit die Stetigkeit für fest gewähltes  $s$ .

Im Fall  $\operatorname{Re}(s) \geq 0$  ergibt sich die  $H_0^1(\Omega)$ -Elliptizität mit Hilfe der Poincaré'schen Ungleichung, Lemma 3.6, durch

$$\operatorname{Re}(a_s(u, u)) = |u|_1^2 + \operatorname{Re}(s) \|u\|_{0,\rho}^2 \geq |u|_1^2 \geq C \|u\|_{1,\rho}^2.$$

Ist  $\operatorname{Re}(s) < 0$ , und insbesondere wegen  $s \in \mathbb{C} \setminus \mathbb{R}_0^+$  auch  $\operatorname{Im}(s) \neq 0$ , so zeigen wir anstelle der Elliptizität die Existenz eines  $\alpha > 0$ , so dass  $|a_s(u, u)| \geq \alpha \|u\|_{1,\rho}^2$  für alle  $u \in H_0^1(\Omega)$  gilt. Nach [13, Bem. VI.3.8] bleibt der Satz von Lax-Milgram damit weiterhin gültig und wir erhalten wieder die Existenz einer eindeutigen Lösung in  $H_0^1(\Omega)$ . Wir folgen hierbei der Argumentation aus [10, S.14]:

Setze  $s = -(a^2 + ib)$  und  $\kappa := \sqrt{-s} = \sqrt{a^2 + ib}$  mit  $a, b \in \mathbb{R} \setminus \{0\}$ , dann ist

$$\zeta := \operatorname{Re}(\kappa) = \left( \frac{a^2 + \sqrt{a^4 + b^2}}{2} \right)^{1/2} \quad \text{und} \quad \eta := \operatorname{Im}(\kappa) = \frac{b}{2} \left( \frac{2}{a^2 + \sqrt{a^4 + b^2}} \right)^{1/2}.$$

Wir erhalten damit für  $u \in H_0^1(\Omega)$  die Gleichung

$$\begin{aligned} \operatorname{Im} \left( \frac{1}{\kappa} a_s(u, u) \right) &= \operatorname{Im} \left( \frac{1}{\kappa} \left( |u|_1^2 - \kappa^2 \|u\|_{0,\rho}^2 \right) \right) = \operatorname{Im} \left( \frac{1}{\kappa} |u|_1^2 - \kappa \|u\|_{0,\rho}^2 \right) \\ &= \operatorname{Im} \left( \frac{\zeta - i\eta}{|\kappa|^2} |u|_1^2 - (\zeta + i\eta) \|u\|_{0,\rho}^2 \right) \\ &= -\frac{\eta}{|\kappa|^2} |u|_1^2 - \eta \|u\|_{0,\rho}^2. \end{aligned}$$

---

<sup>2</sup>Tatsächlich gilt die Aussage wie in Korollar 3.9 bewiesen auch für  $s = 0$ . Hierauf gehen wir jedoch nach obiger Begründung nicht weiter ein.

Wegen  $|z| \geq |\operatorname{Im}(z)| \quad \forall z \in \mathbb{C}$  folgt

$$\left| \frac{1}{\kappa} a_s(u, u) \right| \geq |\eta| \left( \frac{1}{|\kappa|^2} |u|_1^2 + \|u\|_{0,\rho}^2 \right) \geq |\eta| \frac{1}{|\kappa|^2} |u|_1^2$$

und daraus mit Lemma 3.6 die Abschätzung

$$|a_s(u, u)| \geq C \frac{|\eta|}{|\kappa|} \|u\|_{1,\rho}^2$$

mit einer Konstante  $C > 0$ .

Da die zu  $f \in L^2(\Omega)$  definierte Linearform  $\ell_f : H_0^1(\Omega) \rightarrow \mathbb{C}$ ,  $v \mapsto \ell_f(v) := \int_{\Omega} f \bar{v} \, d(x, y)$ , wie bereits im Beweis zu Satz 3.8 gezeigt, stetig in  $H_0^1(\Omega)$  ist, folgt die Existenz und Eindeutigkeit einer schwachen Lösung  $u \in H_0^1(\Omega)$  von **(H)** aus dem Satz von Lax-Milgram.

Analog zum Beweis von Satz 3.8 können wir damit zudem für  $s \in \mathbb{C} \setminus \mathbb{R}_0^+$  weiter folgern, dass die eindeutige Lösung von **(H)** in  $H_0^1(\Delta, \Omega)$  liegt.  $\square$

Das im nächsten Unterkapitel vorgestellte Verfahren von Fliss und Joly zur Berechnung der Lösung des Problems **(H)** basiert nun auf dessen Zerlegung in drei Teilprobleme, jeweils eines auf dem linken beziehungsweise rechten Halb-Wellenleiter und eines im Gebiet  $\Omega_0$  des lokalen Defekts. Diese Idee wird in der folgenden Proposition zusammengefasst, vergleiche hierzu [20]:

**Proposition 3.11.** *Unter den Annahmen (H1)-(H5) gilt:*

Für  $\varphi \in H^{-1/2}(\Gamma_0^\pm)$  sei  $u_s^\pm(\varphi) \in H_{0,\Omega}^1(\Delta, \Omega^\pm)$  die eindeutige Lösung des Helmholtz-Problems auf dem Halb-Wellenleiter

$$\left| \begin{array}{l} -\Delta u_s^\pm(\varphi) + s \rho u_s^\pm(\varphi) = 0 \quad \text{in} \quad \Omega^\pm, \\ u_s^\pm(\varphi) = 0 \quad \text{auf} \quad \partial\Omega \cap \partial\Omega^\pm, \\ \mp \frac{\partial u_s^\pm(\varphi)}{\partial x} = \varphi \quad \text{auf} \quad \Gamma_0^\pm. \end{array} \right. \quad (\mathbf{H}^\pm)$$

Definieren wir die beiden (linearen und stetigen) Operatoren  $\Lambda_s^\pm : H^{-1/2}(\Gamma_0^\pm) \rightarrow H_0^{1/2}(\Gamma_0^\pm)$  durch

$$\Lambda_s^\pm(\varphi) := u_s^\pm(\varphi)|_{\Gamma_0^\pm} \quad (3.3)$$

und bezeichnen mit  $u_s^0$  die eindeutige Lösung des Problems

$$\left| \begin{array}{l} -\Delta u_s^0 + s \rho u_s^0 = f \quad \text{in} \quad \Omega_0, \\ u_s^0 = 0 \quad \text{auf} \quad \partial\Omega \cap \partial\Omega_0, \\ u_s^0 - \Lambda_s^- \left( \frac{\partial u_s^0}{\partial x} \Big|_{\Gamma_0^-} \right) = 0 \quad \text{auf} \quad \Gamma_0^-, \\ u_s^0 - \Lambda_s^+ \left( -\frac{\partial u_s^0}{\partial x} \Big|_{\Gamma_0^+} \right) = 0 \quad \text{auf} \quad \Gamma_0^+ \end{array} \right. \quad (\mathbf{H}^0)$$

in  $H_0^1(\Delta, \Omega_0)$ , so können wir die Lösung  $u_s \in H_0^1(\Delta, \Omega)$  von **(H)** durch den Zusammenhang

$$u_s = \begin{cases} u_s^0 & \text{in } \Omega^0, \\ u_s^- \left( \frac{\partial u_s^0}{\partial x} \Big|_{\Gamma_0^-} \right) & \text{in } \Omega^-, \\ u_s^+ \left( -\frac{\partial u_s^0}{\partial x} \Big|_{\Gamma_0^+} \right) & \text{in } \Omega^+ \end{cases}$$

rekonstruieren.

*Beweis.* Haben wir eindeutige Lösungen  $u_s^\pm(\varphi) \in H_{0,\Omega}^1(\Delta, \Omega^\pm)$  und  $u_s^0 \in H_{0,\Omega}^1(\Delta, \Omega_0)$  der Helmholtz-Probleme gegeben, so ist klar, dass die angegebene stückweise definierte Funktion  $u_s$  die Helmholtz-Gleichung **(H)** auf dem gesamten Wellenleiter löst, eindeutig ist und in  $H_0^1(\Delta, \Omega)$  liegt. Ebenfalls ist die Linearität des Operators  $\Lambda_s^\pm$  klar.

Es bleibt daher noch zu zeigen, dass sowohl **(H $^\pm$ )** als auch **(H $^0$ )** eindeutige Lösungen in den angegebenen Unterräumen besitzen und die Definition (3.3) einen stetigen Operator liefert. Für diese Beweise verweisen wir auf die Sätze 3.12 und 3.17.  $\square$

Mit den in (3.3) definierten Operatoren gewährleisten wir, dass das Problem **(H $^0$ )** unter transparenten Randbedingungen gelöst wird. Wie auch in [20], betrachten wir hierbei  $\Lambda_s^\pm$  als *Neumann-zu-Dirichlet*-Operatoren, also als Abbildungen von  $H^{-1/2}(\Gamma_0^\pm)$  nach  $H_0^{1/2}(\Gamma_0^\pm)$ . Andere Möglichkeiten sind die Realisierung der transparenten Randbedingungen durch *Dirichlet-zu-Neumann*-Operatoren wie beispielsweise in [19, 21–23, 40] oder durch *Robin-zu-Robin*-Operatoren, siehe hierzu [24]. Die Begründung unserer Wahl findet sich in Abschnitt 3.2. Das vorgestellte Verfahren lässt sich mit leichten Modifikationen auf jeden dieser drei Typen anwenden.

### 3.1.4 Helmholtz-Gleichung im Halb-Wellenleiter

Da die NtD-Operatoren (3.3) basierend auf der Lösung der Helmholtz-Gleichung im Halb-Wellenleiter definiert sind, werden wir in diesem Abschnitt das Problem **(H $^\pm$ )** genauer untersuchen und, wie bereits zu Anfang des Kapitels erwähnt, die Periodizität im Halb-Wellenleiter zur Lösungsdarstellung ausnutzen.

Im Folgenden beschränken wir uns der besseren Lesbarkeit halber auf die Betrachtung des rechten Halb-Wellenleiters  $\Omega^+$  und erinnern an dessen Aufbau und die verwendete Notation, siehe Abbildung 3.1, insbesondere an die Identifizierung der Zellen  $C_j^+$ ,  $j \in \mathbb{N}$ , mit der Zelle  $C_0^+$  sowie der Ränder  $\Gamma_j^+$  mit dem Rand  $\Gamma_0^+$ . Sofern nicht anders markiert, beziehen sich die in diesem Abschnitt vorkommenden Normen also auf  $\Omega^+$ . Die Ergebnisse für den linken Halb-Wellenleiter ergeben sich analog.

**Satz 3.12.** *Es gelten die Annahmen (H1)-(H5). Für jedes  $s \in \mathbb{C} \setminus \mathbb{R}_0^+$  und jedes  $\varphi \in H^{-1/2}(\Gamma_0^+)$  gibt es eine eindeutige Lösung  $u_s^+(\varphi) \in H_{0,\Omega}^1(\Omega^+)$  des Helmholtz-Problems*

$$\left| \begin{array}{l} -\Delta u_s^+(\varphi) + s \rho u_s^+(\varphi) = 0 \quad \text{in } \Omega^+, \\ u_s^+(\varphi) = 0 \quad \text{auf } \partial\Omega \cap \partial\Omega^+, \\ -\frac{\partial u_s^+(\varphi)}{\partial x} = \varphi \quad \text{auf } \Gamma_0^+. \end{array} \right. \quad (\mathbf{H}^+)$$

Insbesondere ist damit der Operator  $\Lambda_s^+$  aus (3.3) wohldefiniert und es gilt

$$\Lambda_s^+ \in \mathcal{L}\left(H^{-1/2}(\Gamma_0^+), H_0^{1/2}(\Gamma_0^+)\right).$$

*Beweis.* Zum Operator  $sM_\rho - \Delta$  auf  $\Omega^+$  assoziieren wir die Sesquilinearform  $a_s^+ : H_{0,\Omega}^1(\Omega^+) \times H_{0,\Omega}^1(\Omega^+) \rightarrow \mathbb{C}$  mit

$$a_s^+(u, v) := \int_{\Omega^+} \nabla u \cdot \overline{\nabla v} + s \rho u \bar{v} \, d(x, y).$$

Deren Stetigkeit erhalten wir wie im Beweis von Satz 3.10. Für die  $H_{0,\Omega}^1(\Omega^+)$ -Elliptizität können wir uns am genannten Beweis orientieren, jedoch aufgrund der gemischten Randwerte nun nicht mehr Lemma 3.6 verwenden. Im Fall  $\operatorname{Re}(s) > 0$  erhalten wir hier

$$\operatorname{Re}(a_s^+(u, u)) = |u|_1^2 + \operatorname{Re}(s) \|u\|_{0,\rho}^2 \geq \min(\operatorname{Re}(s), 1) \|u\|_{1,\rho}^2$$

und damit die  $H_{0,\Omega}^1(\Omega^+)$ -Elliptizität, während wir nun im Fall  $\operatorname{Re}(s) \leq 0$  die Bedingung  $|a_s^+(u, u)| \geq \alpha \|u\|_{1,\rho}^2$  für  $\alpha > 0$  zeigen. Hier stellen wir fest, dass die Vorgehensweise aus dem Beweis von Satz 3.10, das heißt  $s = -(a^2 + ib) := -\kappa^2$  zu setzen, auch für den Fall  $a = 0$  möglich ist und damit  $\operatorname{Re}(s) = 0$  mit einschließt. Wegen (H5) ist dann  $\operatorname{Im}(s) \neq 0$  und damit  $\kappa \neq 0$ . Mit derselben Rechnung erhalten wir somit

$$|a_s^+(u, u)| \geq |\eta| |\kappa| \left( \frac{1}{|\kappa|^2} |u|_1^2 + \|u\|_{0,\rho}^2 \right) \geq |\eta| |\kappa| \min(|\kappa|^{-2}, 1) \|u\|_{1,\rho}^2.$$

Die Linearform  $\ell_\varphi : H_{0,\Omega}^1(\Omega^+) \rightarrow \mathbb{C}$  zum Problem  $(\mathbf{H}^+)$  ist gegeben durch

$$\ell_\varphi(v) := \int_{\Gamma_0^+} \varphi \bar{v} \, d\sigma(x, y).$$

Deren Stetigkeit in  $H_{0,\Omega}^1(\Omega^+)$  folgt aus dem Riesz'schen Darstellungssatz, siehe [13, Thm. VI.4.10], und der Stetigkeit der Dirichlet-Spur aus Satz 3.3:

$$\begin{aligned} |\ell_\varphi(v)| &= \left| \left\langle \varphi, v|_{\Gamma_0^+} \right\rangle_{\Gamma_0^+} \right| \leq \|\varphi\|_{-1/2} \left\| v|_{\Gamma_0^+} \right\|_{1/2} \\ &\leq C \|\varphi\|_{-1/2} \|v\|_{1,\rho,C_0^+} \leq C \|\varphi\|_{-1/2} \|v\|_{1,\rho} \end{aligned} \quad (3.4)$$

Damit erhalten wir für  $s \in \mathbb{C} \setminus \mathbb{R}_0^+$  und  $\varphi \in H^{-1/2}(\Omega^+)$  nach dem Satz von Lax-Milgram die Existenz und Eindeutigkeit einer Lösung  $u_s^+(\varphi) \in H_{0,\Omega}^1(\Omega^+)$  von  $(\mathbf{H}^+)$ . Mit derselben

Argumentation wie im Beweis von Satz 3.8 folgt aus der Betrachtung der glatten Testfunktionen mit kompaktem Träger, dass auch  $\Delta u \in L^2(\Omega^+)$  erfüllt ist, also  $u \in H_{0,\Omega}^1(\Delta, \Omega^+)$ . Insbesondere ist damit  $\Lambda_s^+(\varphi) \in H_0^{1/2}(\Gamma_0^+)$  wohldefiniert und wir erhalten aus  $|a(u_s^+(\varphi), u_s^+(\varphi))| = |\ell_\varphi(u_s^+(\varphi))|$  und den obigen Resultaten die Abschätzung

$$\|u_s^+(\varphi)\|_{1,\rho} \leq \tilde{C} \|\varphi\|_{-1/2},$$

das heißt, die Abbildung  $\varphi \mapsto u_s^+(\varphi)$  ist stetig. Zusammen mit  $\|u_s^+(\varphi)\|_{1,\rho,C_0^+} \leq \|u_s^+(\varphi)\|_{1,\rho}$  und Satz 3.3 folgt daraus die Stetigkeit des NtD-Operators  $\Lambda_s^+$ .  $\square$

Die Ergebnisse aus dem Beweis von Satz 3.12 und die Stetigkeit der Neumann-Spur, Satz 3.4, ermöglichen uns unter Ausnutzung der Einheitszelle  $C_0^+$  die Definition der beiden linearen und stetigen Operatoren

$$\begin{aligned} \mathcal{P}_s^+ &\in \mathcal{L}\left(H^{-1/2}(\Gamma_0^+)\right) \quad \text{durch} \quad \mathcal{P}_s^+(\varphi) := -\frac{\partial u_s^+(\varphi)}{\partial x} \Big|_{\Gamma_1^+} \quad \text{und} \\ \mathcal{S}_s^+ &\in \mathcal{L}\left(H^{-1/2}(\Gamma_0^+), H_{0,\Omega}^1(\Delta, C_0^+)\right) \quad \text{durch} \quad \mathcal{S}_s^+(\varphi) := u_s^+(\varphi) \Big|_{C_0^+}, \end{aligned} \quad (3.5)$$

mit denen wir die Lösung auf dem Halb-Wellenleiter  $\Omega^+$  ausgehend von  $\varphi \in H^{-1/2}(\Gamma_0^+)$  zellenweise bestimmen können. Der folgende Satz, der in [20, Lemma 2.2.3] bewiesen wird, fasst die wichtigsten Eigenschaften der Operatoren zusammen.

**Satz 3.13.** *Für die Operatoren aus (3.5) gilt:*

(i) *Der Propagationsoperator  $\mathcal{P}_s^+$  ist kompakt und injektiv. Sein Spektralradius erfüllt*

$$\rho(\mathcal{P}_s^+) < 1;$$

(ii) *die Einschränkung der Lösung  $u_s^+(\varphi)$  auf die Zelle  $C_j^+$ ,  $j \in \mathbb{N}_0$ , lässt sich durch*

$$u_s^+(\varphi) \Big|_{C_j^+} = \mathcal{S}_s^+ \circ (\mathcal{P}_s^+(\varphi))^j$$

bestimmen, vergleiche dazu Abbildung 3.2.

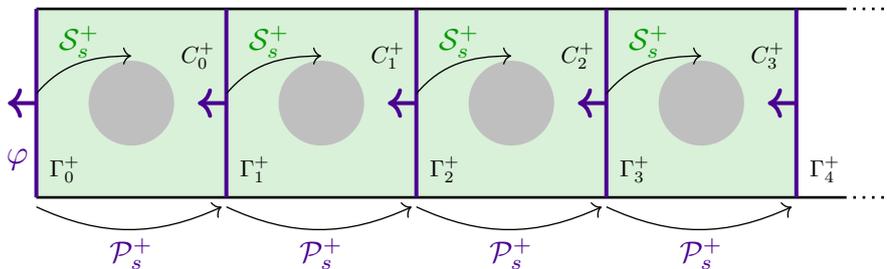


Abbildung 3.2: Darstellung der zellenweisen Berechnung der Lösung  $u_s^+$  von  $(\mathbf{H}^+)$  ausgehend von  $\varphi \in H^{-1/2}(\Gamma_0^+)$ .

Mit Hilfe des Operators  $\mathcal{S}_s^+$  erhalten wir direkt eine weitere Darstellung des NtD-Operators  $\Lambda_s^+$  aus (3.3) durch den Zusammenhang

$$\Lambda_s^+(\varphi) = \mathcal{S}_s^+(\varphi)|_{\Gamma_0^+} \quad \forall \varphi \in H^{-1/2}(\Gamma_0^+). \quad (3.6)$$

Die Berechnung des Operators  $\mathcal{S}_s^+$  können wir wiederum auf ein lokales Problem in der Zelle  $C_0^+$  reduzieren. Unter Verwendung des Propagationsoperators  $\mathcal{P}_s^+(\varphi)$  ist  $\mathcal{S}_s^+(\varphi)$  für  $\varphi \in H^{-1/2}(\Gamma_0^+)$  die Lösung des Helmholtz-Problems

$$\left\{ \begin{array}{ll} -\Delta \mathcal{S}_s^+(\varphi) + s \rho \mathcal{S}_s^+(\varphi) = 0 & \text{in } C_0^+, \\ \mathcal{S}_s^+(\varphi) = 0 & \text{auf } \partial\Omega \cap \partial C_0^+, \\ -\frac{\partial \mathcal{S}_s^+(\varphi)}{\partial x} = \varphi & \text{auf } \Gamma_0^+, \\ \frac{\partial \mathcal{S}_s^+(\varphi)}{\partial x} = -\mathcal{P}_s^+(\varphi) & \text{auf } \Gamma_1^+. \end{array} \right. \quad (3.7)$$

Haben wir also, bezogen auf das Helmholtz-Problem im gesamten Wellenleiter, die Propagationsoperatoren  $\mathcal{P}_s^\pm$  bestimmt, können wir die Operatoren  $\mathcal{S}_s^\pm$  sowie  $\Lambda_s^\pm$  berechnen und erhalten somit über die Lösung der Probleme  $(\mathbf{H}^0)$  und  $(\mathbf{H}^\pm)$  die vollständige Lösung auf dem Wellenleiter. Da die Definition des Operators  $\mathcal{P}_s^+$  in (3.5) noch auf einem unendlichen Gebiet basiert und daher für numerische Berechnungen nicht sinnvoll ist, beschäftigen wir uns im nächsten Abschnitt mit der Frage einer praktischen Charakterisierung von  $\mathcal{P}_s^+$ .

### 3.1.5 Lösen lokaler Probleme in $C_0^+$ und $\Omega_0$

Wie gerade in (3.7) für den Operator  $\mathcal{S}_s^+$  gesehen, können wir für die Berechnung von  $\mathcal{P}_s^+$  dank der Periodizität und der Identifizierung der Zellen auf ein lokales Problem in der Einheitszelle  $C_0^+$  zurückgreifen, vergleiche [20]. Wir bezeichnen dazu mit  $e_{s,k}^+(\varphi)$ ,  $k = 0, 1$  für  $s \in \mathbb{C} \setminus \mathbb{R}_0^+$  und  $\varphi \in H^{-1/2}(\Gamma_0^+)$  die Lösung des lokalen Problems

$$\left\{ \begin{array}{ll} -\Delta e_{s,k}^+(\varphi) + s \rho e_{s,k}^+(\varphi) = 0 & \text{in } C_0^+, \\ e_{s,k}^+(\varphi) = 0 & \text{auf } \partial\Omega \cap \partial C_0^+ \end{array} \right. \quad (3.8)$$

unter den zusätzlichen Randbedingungen

$$\left\{ \begin{array}{l} -\frac{\partial e_{s,0}^+(\varphi)}{\partial x} = \varphi \quad \text{auf } \Gamma_0^+, \\ \frac{\partial e_{s,0}^+(\varphi)}{\partial x} = 0 \quad \text{auf } \Gamma_1^+, \end{array} \right. \quad \text{bzw.} \quad \left\{ \begin{array}{l} -\frac{\partial e_{s,1}^+(\varphi)}{\partial x} = 0 \quad \text{auf } \Gamma_0^+, \\ \frac{\partial e_{s,1}^+(\varphi)}{\partial x} = \varphi \quad \text{auf } \Gamma_1^+. \end{array} \right.$$

**Satz 3.14.** *Unter den Voraussetzungen (H1)-(H4) gibt es für jedes  $s$ , das (H5) erfüllt, und jedes  $\varphi, \psi \in H^{-1/2}(\Gamma_0^+)$  eine eindeutige Lösung  $e_s(\varphi, \psi) \in H_{0,\Omega}^1(\Delta, C_0^+)$  des lokalen*

## Helmholtz-Problems

$$\left\{ \begin{array}{l} -\Delta e_s(\varphi, \psi) + s \rho e_s(\varphi, \psi) = 0 \quad \text{in } C_0^+, \\ e_s(\varphi, \psi) = 0 \quad \text{auf } \partial\Omega \cap \partial C_0^+, \\ -\frac{\partial e_s(\varphi, \psi)}{\partial x} = \varphi \quad \text{auf } \Gamma_0^+, \\ \frac{\partial e_s(\varphi, \psi)}{\partial x} = \psi \quad \text{auf } \Gamma_1^+. \end{array} \right.$$

*Beweis.* Wie bereits demonstriert, zeigen wir die Aussage unter Verwendung des Satzes von Lax-Milgram. Hier liefert uns die schwache Formulierung des Randwertproblems die Sesquilinearform  $a_{s,0}^+ : H_{0,\Omega}^1(C_0^+) \times H_{0,\Omega}^1(C_0^+) \rightarrow \mathbb{C}$  mit

$$a_{s,0}^+(u, v) := \int_{C_0^+} \nabla u \cdot \overline{\nabla v} + s \rho u \bar{v} \, d(x, y),$$

deren Stetigkeit und  $H_{0,\Omega}^1(C_0^+)$ -Elliptizität mit dem Vorgehen aus Satz 3.12 folgen. Die zum Problem assoziierte Linearform erhalten wir durch die Abbildung  $\ell_{\varphi,\psi} : H_{0,\Omega}^1(C_0^+) \rightarrow \mathbb{C}$ , wobei

$$\ell_{\varphi,\psi}(v) := \int_{\Gamma_0^+} \varphi \bar{v} \, d\sigma(x, y) + \int_{\Gamma_1^+} \psi \bar{v} \, d\sigma(x, y).$$

Die Stetigkeit der Linearform ergibt sich wie in Satz 3.12 aus dem Riesz'schen Darstellungssatz, der Stetigkeit der Dirichlet-Spur aus Satz 3.3 sowie der Identifizierung der Ränder  $\Gamma_0^+$  und  $\Gamma_1^+$ , vergleiche dazu (3.4).

Mit der Betrachtung von glatten Testfunktionen erhalten wir weiter  $\Delta u \in L^2(C_0^+)$  und damit eine eindeutige Lösung des Problems in  $H_{0,\Omega}^1(\Delta, C_0^+)$ . Ebenfalls wie im Beweis von Satz 3.12 können wir schließen, dass die Abbildung  $(\varphi, \psi) \mapsto e_s(\varphi, \psi)$  stetig ist.  $\square$

Insbesondere sind also die Lösungen  $e_{s,k}^+$ ,  $k = 0, 1$ , des Problems (3.8) eindeutig und liegen in  $H_{0,\Omega}^1(\Delta, C_0^+)$ . Damit können wir nun die Operatoren

$$\begin{aligned} \mathcal{T}_{s,00}^+(\varphi) &:= e_{s,0}^+(\varphi)|_{\Gamma_0^+}, & \mathcal{T}_{s,01}^+(\varphi) &:= e_{s,0}^+(\varphi)|_{\Gamma_1^+}, \\ \mathcal{T}_{s,10}^+(\varphi) &:= e_{s,1}^+(\varphi)|_{\Gamma_0^+}, & \mathcal{T}_{s,11}^+(\varphi) &:= e_{s,1}^+(\varphi)|_{\Gamma_1^+} \end{aligned} \quad (3.9)$$

definieren, die mit Satz 3.14 und Satz 3.3 als Abbildungen in  $\mathcal{L}\left(H^{-1/2}(\Gamma_0^+), H_0^{1/2}(\Gamma_0^+)\right)$  verstanden und daher auch als NtD-artige Operatoren angesehen werden können. In [20] werden die Operatoren genauer untersucht, worauf wir hier jedoch nicht weiter eingehen.

Was sich jedoch leicht einsehen lässt, ist die Möglichkeit, die Operatoren  $\mathcal{S}_s^+$  und  $\Lambda_s^+$  mit Hilfe der Lösung der lokalen Probleme auszudrücken. Aus (3.7) ergibt sich die Darstellung

$$\mathcal{S}_s^+(\varphi) = e_{s,0}^+(\varphi) - e_{s,1}^+(\mathcal{P}_s^+(\varphi)), \quad (3.10)$$

während (3.6) zum Ausdruck

$$\Lambda_s^+(\varphi) = \mathcal{T}_{s,00}^+(\varphi) - \mathcal{T}_{s,10}^+(\mathcal{P}_s^+(\varphi)) \quad (3.11)$$

führt. Auch der Operator  $\mathcal{P}_s^+$  kann mit Hilfe der NtD-artigen Operatoren bestimmt werden, wie der folgende Satz zeigt, dessen Beweis in [20, Thm. 2.2.5] zu finden ist.

**Satz 3.15.** *Der Operator  $\mathcal{P}_s^+ \in \mathcal{L}(H^{-1/2}(\Gamma_0^+))$  aus (3.5) ist die eindeutige Lösung des Riccati-Problems:*

*Finde ein kompaktes  $X \in \mathcal{L}(H^{-1/2}(\Gamma_0^+))$  mit einem Spektralradius kleiner als 1, so dass*

$$\mathcal{T}_{s,10}^+(X^2) - (\mathcal{T}_{s,00}^+ + \mathcal{T}_{s,11}^+)(X) + \mathcal{T}_{s,01}^+ = 0 \quad (3.12)$$

*erfüllt ist.*

Um die Zusammenfassung des von Fliss und Joly vorgestellten Verfahrens abzurunden, bleibt noch zu zeigen, dass das Helmholtz-Problem ( $\mathbf{H}^0$ ), das heißt, das Problem im Gebiet  $\Omega_0$  unter Verwendung der durch die NtD-Operatoren gegebenen transparenten Randbedingungen, eindeutig lösbar ist. Dazu benötigen wir das folgende Lemma, das auf die bereits erwähnten DtN-Operatoren eingeht und eine Erweiterung von [20, Prop. 2.2.9] darstellt.

**Lemma 3.16.** *Der Operator  $\Lambda_s^\pm$  aus (3.3) ist invertierbar. Für den inversen Operator  $\tilde{\Lambda}_s^\pm \in \mathcal{L}(H_0^{1/2}(\Gamma_0^\pm), H^{-1/2}(\Gamma_0^\pm))$  gelten insbesondere für alle  $\psi \in H_0^{1/2}(\Gamma_0^\pm)$ ,  $\psi \neq 0$ , die Abschätzungen*

$$\begin{cases} \operatorname{Re}\left(\left\langle \tilde{\Lambda}_s^\pm(\psi), \psi \right\rangle_{\Gamma_0^\pm}\right) > 0, & \text{falls } \operatorname{Re}(s) > 0 \text{ sowie} \\ \operatorname{Im}\left(\left\langle \tilde{\Lambda}_s^\pm(\psi), \psi \right\rangle_{\Gamma_0^\pm}\right) > 0, & \text{falls } \operatorname{Im}(s) > 0, \\ \operatorname{Im}\left(\left\langle \tilde{\Lambda}_s^\pm(\psi), \psi \right\rangle_{\Gamma_0^\pm}\right) < 0, & \text{falls } \operatorname{Im}(s) < 0, \end{cases}$$

wobei  $\langle \cdot, \cdot \rangle_{\Gamma_0^\pm}$  die duale Paarung in  $H_0^{1/2}(\Gamma_0^\pm)$  bezeichnet.

*Beweis.* Wir zeigen die Aussage der Übersichtlichkeit halber nur für den Operator  $\tilde{\Lambda}_s^+$ , der Beweis für den Operator auf dem linken Halb-Wellenleiter ist analog.

In gleicher Weise, wie wir es für den NtD-Operator in Proposition 3.11 gesehen haben, definieren wir nun für  $\psi \in H_0^{1/2}(\Gamma_0^+)$  den Operator  $\tilde{\Lambda}_s^+ : H_0^{1/2}(\Gamma_0^+) \rightarrow H^{-1/2}(\Gamma_0^+)$  durch

$$\tilde{\Lambda}_s^+(\psi) := -\frac{\partial v_s^+(\psi)}{\partial x} \Big|_{\Gamma_0^+},$$

wobei  $v_s^+(\psi) \in H_{0,\Omega}^1(\Delta, \Omega^+)$  die Lösung des Helmholtz-Problems

$$\left\{ \begin{array}{l} -\Delta v^+ + s \rho v^+ = 0 \quad \text{in } \Omega^+, \\ v^+ = 0 \quad \text{auf } \partial\Omega \cap \partial\Omega^+, \\ v^+ = \psi \quad \text{auf } \Gamma_0^+ \end{array} \right. \quad (3.13)$$

ist. Die Existenz und Eindeutigkeit der Lösung  $v \in H_{0,\Omega}^1(\Delta, \Omega^+)$  lässt sich erneut mit dem Satz von Lax-Milgram beweisen, wobei wir die bereits gezeigten Eigenschaften der Sesquilinearform  $a_s^+$  aus dem Beweis von Satz 3.12 verwenden können. Zusätzlich benötigen wir hierbei, dass es ein  $\tilde{v} \in H_{0,\Omega}^1(\Delta, \Omega^+)$  gibt, für das  $\tilde{v}|_{\Gamma_0^+} = \psi$  gilt. Dies ist auf Grund der Surjektivität der Dirichlet-Spur aus Satz 3.3, die in [13, Thm. IV.3.5] gezeigt wird, erfüllt. Wie in den vorherigen Beweisen können wir nun die Stetigkeit der Abbildungen  $\psi \mapsto v_s^+(\psi)$  beziehungsweise  $\psi \mapsto v_s^+(\psi)|_{C_0^+}$  folgern. Zusammen mit der Stetigkeit der Neumann-Spur, Satz 3.4, erhalten wir daraus die Stetigkeit der Abbildung  $\psi \mapsto \tilde{\Lambda}_s^+(\psi)$ .

Vergleichen wir nun das Problem mit inhomogener Neumann-Bedingung auf  $\Gamma_0^+$  ( $\mathbf{H}^+$ ) mit dem mit inhomogener Dirichlet-Bedingung auf  $\Gamma_0^+$  (3.13), lässt sich leicht einsehen, dass

$$\begin{aligned} \tilde{\Lambda}_s^+(\Lambda_s^+(\varphi)) &= \varphi \quad \forall \varphi \in H^{-1/2}(\Gamma_0^+) \quad \text{und} \\ \Lambda_s^+(\tilde{\Lambda}_s^+(\psi)) &= \psi \quad \forall \psi \in H_0^{1/2}(\Gamma_0^+) \end{aligned}$$

gilt, und damit durch  $\tilde{\Lambda}_s^+$  der inverse Operator zu  $\Lambda_s^+$  gegeben ist.

Die behaupteten Abschätzungen zeigen wir unter Verwendung der variationellen Formulierung des Problems ( $\mathbf{H}^+$ ), die wir im bereits erwähnten Satz 3.12 hergeleitet haben. Setzen wir in diese die Lösung  $u_s^+(\varphi)$  von ( $\mathbf{H}^+$ ) ein, erhalten wir die Gleichung

$$a_s^+(u_s^+(\varphi), u_s^+(\varphi)) = \ell_\varphi(u_s^+(\varphi)) \quad \forall \varphi \in H^{-1/2}(\Gamma_0^+)$$

beziehungsweise mit (3.3) und unter Verwendung des inversen Operators die äquivalente Formulierung

$$\begin{aligned} |u_s^+(\varphi)|_1^2 + s \|u_s^+(\varphi)\|_{0,\rho}^2 &= \int_{\Gamma_0^+} \varphi \overline{\Lambda_s^+(\varphi)} \, d\sigma(x, y) \\ &= \int_{\Gamma_0^+} \tilde{\Lambda}_s^+(\psi) \bar{\psi} \, d\sigma(x, y) = \langle \tilde{\Lambda}_s^+(\psi), \psi \rangle_{\Gamma_0^+} \end{aligned}$$

für  $\Lambda_s^+(\varphi) = \psi \in H_0^{1/2}(\Gamma_0^+)$ . Daraus ergeben sich sofort die behaupteten Abschätzungen.  $\square$

Mit Hilfe des Lemmas können wir nun den Beweis von Proposition 3.11 abschließen:

**Satz 3.17.** *Unter den Annahmen (H1)-(H5) hat das lokale Helmholtz-Problem ( $\mathbf{H}^0$ ) eine eindeutige Lösung in  $H_{0,\Omega}^1(\Omega_0)$ .*

*Beweis.* Anhand des inversen Operators aus Lemma 3.16 können wir die Nicht-Null-

Randbedingungen in  $(\mathbf{H}^0)$  auch schreiben als

$$\left| \mp \frac{\partial u^0}{\partial x} = \tilde{\Lambda}_s^\pm \left( u^0 \Big|_{\Gamma_0^\pm} \right) \right. \quad \text{auf } \Gamma_0^\pm,$$

wodurch sich die variationelle Formulierung, vergleiche Satz 3.12,

$$a_s^0(u, v) = \ell_f(v) := \int_{\Omega_0} f \bar{v} \, d(x, y) \quad \text{für } u, v \in H_{0,\Omega}^1(\Omega_0)$$

ergibt. Hierbei ist die Sesquilinearform  $a_s^0(u, v) : H_{0,\Omega}^1(\Omega_0) \times H_{0,\Omega}^1(\Omega_0) \rightarrow \mathbb{C}$  gegeben durch

$$\begin{aligned} a_s^0(u, v) &:= \int_{\Omega_0} \nabla u \cdot \overline{\nabla v} + s \rho u \bar{v} \, d(x, y) \\ &\quad + \int_{\Gamma_0^-} \tilde{\Lambda}_s^- \left( u \Big|_{\Gamma_0^-} \right) \bar{v} \, d\sigma(x, y) + \int_{\Gamma_0^+} \tilde{\Lambda}_s^+ \left( u \Big|_{\Gamma_0^+} \right) \bar{v} \, d\sigma(x, y). \end{aligned}$$

Unter Verwendung der Stetigkeit der DtN-Operatoren  $\tilde{\Lambda}_s^\pm$  aus Lemma 3.16 sowie der Stetigkeit der Dirichletspur, Satz 3.3, lässt sich die Stetigkeit der Sesquilinearform  $a_s^0$  leicht einsehen, vergleiche dazu (3.4). Die  $H_{0,\Omega}^1(\Omega_0)$ -Elliptizität zeigen wir wie in Satz 3.12:

Der besseren Lesbarkeit halber setzen wir hierbei

$$B^\pm := \left\langle \tilde{\Lambda}_s^\pm \left( u \Big|_{\Gamma_0^\pm} \right), v \right\rangle_{\Gamma_0^\pm} = \int_{\Gamma_0^\pm} \tilde{\Lambda}_s^\pm \left( u \Big|_{\Gamma_0^\pm} \right) \bar{v} \, d\sigma(x, y).$$

Ist  $\operatorname{Re}(s) > 0$  folgt mit Lemma 3.16

$$\begin{aligned} \operatorname{Re}(a_s^0(u, u)) &= |u|_1^2 + \operatorname{Re}(s) \|u\|_{0,\rho}^2 + \operatorname{Re}(B^-) + \operatorname{Re}(B^+) \\ &\geq \min(\operatorname{Re}(s), 1) \|u\|_{1,\rho}^2. \end{aligned}$$

Für den Fall  $\operatorname{Re}(s) \leq 0$  betrachten wir  $s = -(a^2 + ib)$  und setzen  $\kappa =: \sqrt{-s}$  mit  $a \in \mathbb{R}$ ,  $b \in \mathbb{R} \setminus \{0\}$  und erhalten so, wie bereits im Beweis von Satz 3.10 gesehen, den Zusammenhang

$$\zeta := \operatorname{Re}(\kappa) = \left( \frac{a^2 + \sqrt{a^4 + b^2}}{2} \right)^{1/2} \quad \text{und} \quad \eta := \operatorname{Im}(\kappa) = \frac{b}{2} \left( \frac{2}{a^2 + \sqrt{a^4 + b^2}} \right)^{1/2}.$$

Offensichtlich gilt  $\zeta > 0$  für alle  $\operatorname{Re}(s) \leq 0$ . Bei der Betrachtung des Imaginärteils von  $s$  müssen wir die folgende Unterscheidung machen, wobei wir die jeweils letzte Abschätzung aus Lemma 3.16 verwenden:

- Ist  $\operatorname{Im}(s) = -b > 0$ , also  $b < 0$ , so ist  $\eta < 0$  und  $\operatorname{Im}(B^\pm) > 0$ ;
- Ist  $\operatorname{Im}(s) = -b < 0$ , also  $b > 0$ , so ist  $\eta > 0$  und  $\operatorname{Im}(B^\pm) < 0$ .

In beiden Fällen können wir daher

$$\zeta \cdot \operatorname{Im}(B^\pm) = -\eta \cdot c^\pm \tag{*}$$

mit Konstanten  $c^\pm > 0$  schreiben. Damit erhalten wir für  $u \in H_{0,\Omega}^1(\Omega_0)$  mit dem Vorgehen aus dem Beweis zu Satz 3.10

$$\begin{aligned}
 \operatorname{Im}\left(\frac{1}{\kappa} a_s^0(u, u)\right) &= \operatorname{Im}\left(\frac{1}{\kappa} |u|_1^2 - \kappa \|u\|_{0,\rho}^2 + \frac{1}{\kappa} (B^- + B^+)\right) \\
 &= \operatorname{Im}\left(\frac{\zeta - i\eta}{|\kappa|^2} |u|_1^2 - (\zeta + i\eta) \|u\|_{0,\rho}^2 + \frac{\zeta - i\eta}{|\kappa|^2} (B^- + B^+)\right) \\
 &= \frac{-\eta}{|\kappa|^2} |u|_1^2 - \eta \|u\|_{0,\rho}^2 + \frac{1}{|\kappa|^2} \left(\zeta \operatorname{Im}(B^-) - \eta \operatorname{Re}(B^-) \dots \right. \\
 &\quad \left. + \zeta \operatorname{Im}(B^+) - \eta \operatorname{Re}(B^+)\right) \\
 &\stackrel{(*)}{=} -\eta \left(\frac{1}{|\kappa|^2} |u|_1^2 + \|u\|_{0,\rho}^2 + \frac{1}{|\kappa|^2} (c^- + \operatorname{Re}(B^-) + c^+ + \operatorname{Re}(B^+))\right)
 \end{aligned}$$

und weiter wegen  $|z| \geq |\operatorname{Im}(z)| \quad \forall z \in \mathbb{C}$  und da jeder Summand positiv ist

$$\begin{aligned}
 |a_s^0(u, u)| &\geq |\eta| |\kappa| \left(\frac{1}{|\kappa|^2} |u|_1^2 + \|u\|_{0,\rho}^2 + \frac{1}{|\kappa|^2} (c^- + \operatorname{Re}(B^-) + c^+ + \operatorname{Re}(B^+))\right) \\
 &\geq |\eta| |\kappa| \left(\frac{1}{|\kappa|^2} |u|_1^2 + \|u\|_{0,\rho}^2\right) \geq |\eta| |\kappa| \min(|\kappa|^{-2}, 1) \|u\|_{1,\rho}^2.
 \end{aligned}$$

Zusammen mit der Stetigkeit der Linearform  $\ell_f : H_{0,\Omega}^1(\Omega_0) \rightarrow \mathbb{C}$ , die wir aus den Annahmen (H3) und (H4) erhalten, folgt nun für jedes  $s \in \mathbb{C} \setminus \mathbb{R}_0^+$  die Existenz und Eindeutigkeit einer Lösung  $u \in H_{0,\Omega}^1(\Omega_0)$  von  $(\mathbf{H}^0)$ . Unter Betrachtung der glatten Testfunktionen können wir zusätzlich wieder  $\Delta u \in L^2(\Omega_0)$  folgern und erhalten  $u \in H_{0,\Omega}^1(\Delta, \Omega_0)$ .  $\square$

## 3.2 Berechnungen im eindimensionalen Wellenleiter

### 3.2.1 Übertragung der Resultate in 1D

Um die Operatoren  $\mathcal{P}_s^\pm$ ,  $\mathcal{S}_s^\pm$  und  $\Lambda_s^\pm$  besser verstehen zu können und insbesondere unsere Wahl des NtD- gegenüber der des DtN-Operators zu begründen, betrachten wir in diesem Unterkapitel den Spezialfall eines eindimensionalen Wellenleiters, das heißt, hier setzen wir  $\Omega = \mathbb{R}$ . Entsprechend erhalten wir die Vereinfachungen

$$\begin{aligned}
 \Omega_0 &= (a^-, a^+), \\
 C_j^+ &= (a^+ + jL^+, a^+ + (j+1)L^+), \\
 C_j^- &= (a^- - (j+1)L^-, a^- - jL^-) \quad \text{und} \\
 \Gamma_j^\pm &= a^\pm \pm jL^\pm,
 \end{aligned}$$

vergleiche dazu Abbildung 3.1. Da sich die Periodizität der Funktion  $\rho$  sowohl im Ein- wie auch im Zweidimensionalen nur auf die  $x$ -Richtung bezieht, lassen sich die Anforderungen (H1) und (H2) direkt auf den eindimensionalen Fall übertragen.

**Bemerkung 3.18.** Es lässt sich leicht einsehen, dass die für die Helmholtz-Probleme im zweidimensionalen Wellenleiter getroffenen Aussagen aus den vorherigen Abschnitten auch im eindimensionalen Fall ihre Gültigkeit behalten. Hier sei jedoch anzumerken, dass die Poincaré'sche Ungleichung aus Lemma 3.6 nicht für den Fall  $\Omega = \mathbb{R}$  gilt. Für den Operator  $A : D(A) = H^1(\Delta, \mathbb{R}) \rightarrow L^2(\mathbb{R})$  mit  $A := -\rho^{-1}\partial_x^2$  aus (3.1b) können wir dadurch, anstelle der Aussage in Korollar 3.9, aus Satz 3.8 nur folgern, dass  $\sigma(A) \subset \mathbb{R}_0^+$  erfüllt ist. Tatsächlich lässt sich leicht nachvollziehen, dass der Operator  $A$  nicht invertierbar ist. Dass die Operatoren  $\mathcal{P}_s^\pm$ ,  $\mathcal{S}_s^\pm$  und  $\Lambda_s^\pm$  im Fall  $s = 0$  nicht existieren, können wir auch in den nachfolgenden Beispielen beobachten.

Wir werden in den nächsten Abschnitten daher die eindimensionale Version des Helmholtz-Problems (**H**) sowohl im ortskontinuierlichen als auch -diskreten Fall, für je ein Beispiel eines homogenen und eines inhomogenen Wellenleiters betrachten und die Operatoren  $\mathcal{P}_s^+$ ,  $\mathcal{S}_s^+$  und  $\Lambda_s^+$  beziehungsweise deren Diskretisierungen bestimmen und untersuchen.

### 3.2.2 Ortskontinuierlicher Fall

Wie in Abschnitt 3.1.5 gezeigt, benötigen wir zur Bestimmung der Operatoren lediglich die Lösung der lokalen Helmholtz-Probleme auf der Zelle  $C_0^+ = (a^+, a^+ + L^+)$ , vergleiche Satz 3.15. Zur Vereinfachung und Übersichtlichkeit setzen wir hierbei

$$a^+ = 1 \quad \text{und} \quad L^+ = 2.$$

Damit lauten die lokalen Probleme, vergleiche (3.8): Finde für eine Konstante  $\varphi \in \mathbb{C} \setminus \{0\}$  die Lösung  $e_{s,k}^+(\varphi) \in H^1(\Delta, C_0^+)$ ,  $k = 0, 1$ , der eindimensionalen Helmholtz-Probleme

$$-\frac{\partial^2 e_{s,k}^+(\varphi)}{\partial x^2} + s\rho e_{s,k}^+(\varphi) = 0 \quad \text{für} \quad x \in C_0^+ := (1, 3) \quad (3.14a)$$

mit den Randbedingungen

$$\left| \begin{array}{l} -\frac{\partial e_{s,0}^+(\varphi)}{\partial x}(1) = \varphi, \\ \frac{\partial e_{s,0}^+(\varphi)}{\partial x}(3) = 0, \end{array} \right. \quad \text{bzw.} \quad \left| \begin{array}{l} -\frac{\partial e_{s,1}^+(\varphi)}{\partial x}(1) = 0, \\ \frac{\partial e_{s,1}^+(\varphi)}{\partial x}(3) = \varphi. \end{array} \right. \quad (3.14b)$$

Erneut aus Gründen der Übersicht setzen wir zudem im Folgenden

$$\sigma := \sqrt{s},$$

wobei wir zunächst wieder  $s \in \mathbb{C}$  zulassen werden. Es gilt  $\text{Re}(\sigma) \geq 0$ .

### 3.2.2.1 Homogener Wellenleiter

Für die Betrachtung eines homogenen Wellenleiters setzen wir  $\rho \equiv 1$ . Wie sich leicht nachrechnen lässt, erhalten wir hier für  $\varphi \in \mathbb{C} \setminus \{0\}$  und  $x \in (1, 3)$  die Lösungen

$$e_{s,0}^+(\varphi)(x) = \frac{(e^{(x-1)\sigma} + e^{(-x+5)\sigma})\varphi}{\sigma(e^{4\sigma} - 1)} \quad \text{und} \quad e_{s,1}^+(\varphi)(x) = \frac{(e^{(x+1)\sigma} + e^{(-x+3)\sigma})\varphi}{\sigma(e^{4\sigma} - 1)}.$$

Mit (3.9) ergeben sich daraus die Operatoren:

$$\mathcal{T}_{s,00}^+(\varphi) = \frac{(e^{4\sigma} + 1)\varphi}{\sigma(e^{4\sigma} - 1)} = \mathcal{T}_{s,11}^+(\varphi) \quad \text{und} \quad \mathcal{T}_{s,01}^+(\varphi) = \frac{2e^{2\sigma}\varphi}{\sigma(e^{4\sigma} - 1)} = \mathcal{T}_{s,10}^+(\varphi)$$

Damit vereinfacht sich die Ricatti-Gleichung (3.12) zu der quadratischen Gleichung

$$0 = X^2 - 2\frac{\mathcal{T}_{s,00}^+}{\mathcal{T}_{s,01}^+}X + 1 = X^2 - (e^{2\sigma} + e^{-2\sigma})X + 1,$$

die durch  $X_{s,+} = e^{2\sigma}$  und  $X_{s,-} = e^{-2\sigma}$  gelöst wird. Für  $s \in \mathbb{C} \setminus \mathbb{R}_0^+$  ist  $\operatorname{Re}(\sigma) > 0$  und damit  $X_{s,-}$  die Lösung, deren Spektralradius kleiner als 1 ist. Ist  $s \in \mathbb{R}_0^+$ , und daher  $\operatorname{Re}(\sigma) = 0$ , so hat die quadratische Gleichung keine Lösung, die die Anforderungen aus Satz 3.15 erfüllt, was mit der Überlegung in Bemerkung 3.18 übereinstimmt.

Mit  $\mathcal{P}_s^+ := X_{s,-}$  für  $s \in \mathbb{C} \setminus \mathbb{R}_0^+$  folgt weiter nach (3.10) und (3.11):

$$\begin{aligned} \mathcal{S}_s^+(\varphi)(x) &= e_{s,0}^+(\varphi)(x) - e_{s,1}^+(\mathcal{P}_s^+(\varphi))(x) = \frac{e^{(1-x)\sigma}\varphi}{\sigma} \quad \text{und} \\ \Lambda_s^+(\varphi) &= \mathcal{T}_{s,00}^+(\varphi) - \mathcal{T}_{s,10}^+(\mathcal{P}_s^+(\varphi)) = \frac{\varphi}{\sigma} \end{aligned} \quad (3.15)$$

Es lässt sich leicht einsehen, dass die Abbildung  $\Lambda_{(\cdot)}^+(\varphi) : \mathbb{C} \setminus \mathbb{R}_0^+ \rightarrow \mathbb{C}$ ,  $s \mapsto \Lambda_s^+(\varphi)$  die Abschätzung

$$|\Lambda_s^+(\varphi)| \leq C |s|^{-1/2} \quad \text{für} \quad s \in \mathbb{C} \setminus \mathbb{R}_0^+$$

erfüllt. Verändern wir den Definitionsbereich und verstehen diese stattdessen als Abbildung  $\Lambda_{(\cdot)}^+(\varphi) : \Sigma(0) \rightarrow \mathbb{C}$ ,  $\sigma \mapsto \Lambda_{\sigma^2}^+(\varphi)$ , so gilt offensichtlich wegen  $\sigma = \sqrt{s}$  die Abschätzung

$$|\Lambda_{\sigma^2}^+(\varphi)| \leq C |\sigma|^{-1} \quad \text{für} \quad \operatorname{Re}(\sigma) > 0. \quad (3.16)$$

Damit sind sowohl die Annahme 1.2 für  $s \in \Sigma_{\pi/2+\theta_0}$ ,  $0 < \theta_0 \leq \pi/2$ , als auch die Voraussetzungen aus Lemma 2.3 und damit die Annahme 2.2 für  $\operatorname{Re}(\sigma) > 0$  erfüllt. Gleiches gilt für die Abbildungen, die analog über die Komposition  $\mathcal{S}_s^+ \circ (\mathcal{P}_s^+)^j$ ,  $j \in \mathbb{N}_0$ , definiert sind.

**Bemerkung 3.19.** Wir haben bereits gesehen, dass die DtN- und NtD-Operatoren invers zueinander sind. Damit ist der DtN-Operator für dieses Beispiel gegeben durch  $\tilde{\Lambda}_s^+ = \sigma$  und erfüllt keine der Abschätzungen aus Annahme 2.2 und Annahme 1.2.

### 3.2.2.2 Inhomogener Wellenleiter

Als Beispiel eines inhomogenen Wellenleiters betrachten wir eine stückweise konstante Funktion  $\rho$ , genauer setzen wir

$$\rho^+(x)|_{C_0^+} := \begin{cases} 1, & \text{wenn } x \in (1, 2), \\ 4, & \text{wenn } x \in (2, 3). \end{cases}$$

Mit einer solchen Funktionen wird ein Wellenleiter modelliert, in dem sich Schichten mit verschiedenen Brechungsindizes abwechseln. In der Praxis wird diese Art von Medien als *Bragg-Spiegel* beziehungsweise *distributed Bragg reflector* bezeichnet, siehe beispielsweise [52, 69].

Wir bestimmen die Lösung des eindimensionalen lokalen Helmholtz-Problems auf  $C_0^+ = (1, 3)$  unter den Bedingungen (3.14b) mit Hilfe zweier lokaler Helmholtz-Probleme auf den Intervallen  $(1, 2)$  und  $(2, 3)$ , das heißt, wir lösen für  $\beta_k \in \mathbb{C}$  und  $k = 0, 1$  die Gleichungen

$$\begin{aligned} -\frac{\partial^2 v_{s,k}(\varphi, \beta_k)}{\partial x^2} + s v_{s,k}(\varphi, \beta_k) = 0 \quad \text{für } x \in (1, 2) \quad \text{mit} \quad (3.17) \\ \left| \begin{array}{l} -\frac{\partial v_{s,0}(\varphi, \beta_0)}{\partial x}(1) = \varphi, \\ \frac{\partial v_{s,0}(\varphi, \beta_0)}{\partial x}(2) = \beta_0, \end{array} \right. \quad \text{bzw.} \quad \left| \begin{array}{l} -\frac{\partial v_{s,1}(\varphi, \beta_1)}{\partial x}(1) = 0, \\ \frac{\partial v_{s,1}(\varphi, \beta_1)}{\partial x}(2) = \beta_1 \end{array} \right. \end{aligned}$$

sowie

$$\begin{aligned} -\frac{\partial^2 w_{s,k}(\varphi, \beta_k)}{\partial x^2} + 4s w_{s,k}(\varphi, \beta_k) = 0 \quad \text{für } x \in (2, 3) \quad \text{mit} \quad (3.18) \\ \left| \begin{array}{l} -\frac{\partial w_{s,0}(\varphi, \beta_0)}{\partial x}(2) = -\beta_0, \\ \frac{\partial w_{s,0}(\varphi, \beta_0)}{\partial x}(3) = 0, \end{array} \right. \quad \text{bzw.} \quad \left| \begin{array}{l} -\frac{\partial w_{s,1}(\varphi, \beta_1)}{\partial x}(2) = -\beta_1, \\ \frac{\partial w_{s,1}(\varphi, \beta_1)}{\partial x}(3) = \varphi. \end{array} \right. \end{aligned}$$

Nach Satz 3.14 und obiger Überlegung zur Gültigkeit im Eindimensionalen gilt  $v_{s,k} \in H^1(\Delta, (1, 2))$  und  $w_{s,k} \in H^1(\Delta, (2, 3))$ . Fordern wir zusätzlich noch, dass  $v_{s,k}(2) = w_{s,k}(2)$  erfüllt ist, erhalten wir die Lösung  $e_{s,k}^+ \in H^1(\Delta, (1, 3))$  durch  $e_{s,k}^+|_{(1,2)} := v_{s,k}$  und  $e_{s,k}^+|_{(2,3)} := w_{s,k}$ .

Die Rechnungen im Folgenden wurden mit MAPLE durchgeführt, zur Übersicht stellen wir jeweils nur die Ergebnisse vor. Wir erhalten für die lokalen Probleme (3.17) und (3.18) die Lösungsdarstellungen

$$v_{s,0}(\varphi, \beta_0)(x) = \frac{(e^{x\sigma} + e^{(2-x)\sigma})\beta_0 + (e^{(x-1)\sigma} + e^{(3-x)\sigma})\varphi}{\sigma(e^{2\sigma} - 1)} \quad \text{und}$$

$$w_{s,1}(\varphi, \beta_1)(x) = \frac{(e^{2(x-1)\sigma} + e^{2(3-x)\sigma})\varphi - (e^{2(x-2)\sigma} + e^{2(4-x)\sigma})\beta_1}{2\sigma(e^{4\sigma} - 1)} \quad \text{sowie}$$

$$v_{s,1}(\varphi, \beta_1)(x) = v_{s,0}(0, \beta_1)(x) \quad \text{und} \quad w_{s,0}(\varphi, \beta_0)(x) = w_{s,1}(0, \beta_0)(x).$$

Die Bedingungen  $v_{s,0}(2) = w_{s,0}(2)$  und  $v_{s,1}(2) = w_{s,1}(2)$  liefern

$$\beta_0 = -\frac{4e^\sigma(e^{2\sigma} + 1)\varphi}{3e^{4\sigma} + 4e^{2\sigma} + 3} \quad \text{und} \quad \beta_1 = \frac{2e^{2\sigma}\varphi}{3e^{4\sigma} + 4e^{2\sigma} + 3}$$

und damit

$$e_{s,k}^+(\varphi)|_{(1,2)} = v_{s,k}(\varphi, \beta_k) \quad \text{und} \quad e_{s,k}^+(\varphi)|_{(2,3)} = w_{s,k}(\varphi, \beta_k) \quad \text{für} \quad k = 0, 1.$$

Nach (3.9) erhalten wir daraus die Operatoren

$$\mathcal{T}_{s,00}^+(\varphi) = \frac{(e^{2\sigma} + 1)(3e^{4\sigma} - 4e^{2\sigma} + 3)}{(3e^{4\sigma} + 4e^{2\sigma} + 3)\sigma(e^{2\sigma} - 1)}\varphi,$$

$$\mathcal{T}_{s,01}^+(\varphi) = \mathcal{T}_{s,10}^+(\varphi) = \frac{4e^{3\sigma}}{(3e^{4\sigma} + 4e^{2\sigma} + 3)\sigma(e^{2\sigma} - 1)}\varphi \quad \text{und}$$

$$\mathcal{T}_{s,11}^+(\varphi) = \frac{(e^{2\sigma} + 1)(3e^{4\sigma} - 2e^{2\sigma} + 3)}{2(3e^{4\sigma} + 4e^{2\sigma} + 3)\sigma(e^{2\sigma} - 1)}\varphi,$$

mit denen wir die Ricatti-Gleichung (3.12) aufstellen können, deren Lösungen durch

$$X_{s,\pm} = \frac{1}{16e^{3\sigma}} ((e^{2\sigma} + 1)(9e^{4\sigma} - 10e^{2\sigma} + 9) \pm (e^{2\sigma} - 1)\chi_s) \quad (3.19)$$

$$\text{mit} \quad \chi_s := \sqrt{(3e^{2\sigma} + 4e^\sigma + 3)(3e^{2\sigma} - 4e^\sigma + 3)(3e^{2\sigma} + 2e^\sigma + 3)(3e^{2\sigma} - 2e^\sigma + 3)}$$

gegeben ist. Anders als im homogenen Fall ist der Spektralradius keiner der beiden Lösungen für alle  $\sigma \in \mathbb{C}$  mit  $\text{Re}(\sigma) > 0$  kleiner als 1, wie in Abbildung 3.3 dargestellt ist. Da sich die Ricatti-Gleichung wegen  $\mathcal{T}_{s,01}^+ = \mathcal{T}_{s,10}^+$  wie im homogenen ortskontinuierlichen Fall vereinfacht, gilt insbesondere  $X_{s,+} \cdot X_{s,-} = 1$ , wodurch der Propagationsoperator gegeben ist durch

$$\mathcal{P}_s^+ = \begin{cases} X_{s,+}, & \text{falls } |X_{s,+}| < 1, \\ X_{s,-}, & \text{falls } |X_{s,+}| > 1. \end{cases} \quad (3.20)$$

Für den NtD-Operator erhalten wir mit (3.11):

$$\Lambda_s^+ = \begin{cases} \frac{3e^{4\sigma} - 3 - \chi_s}{4(3e^{4\sigma} + 4e^{2\sigma} + 3)\sigma}, & \text{falls } |X_{s,+}| < 1, \\ \frac{3e^{4\sigma} - 3 + \chi_s}{4(3e^{4\sigma} + 4e^{2\sigma} + 3)\sigma}, & \text{falls } |X_{s,+}| > 1 \end{cases}$$

Wie im homogenen Beispiel hat  $\Lambda_s^+$  eine Singularität bei  $\sigma = 0$ . Betrachten wir die Wurzeln des Nennerpolynoms, das heißt genauer die Gleichung  $3e^{4\sigma} + 4e^{2\sigma} + 3 = 0$ , stellen wir fest,

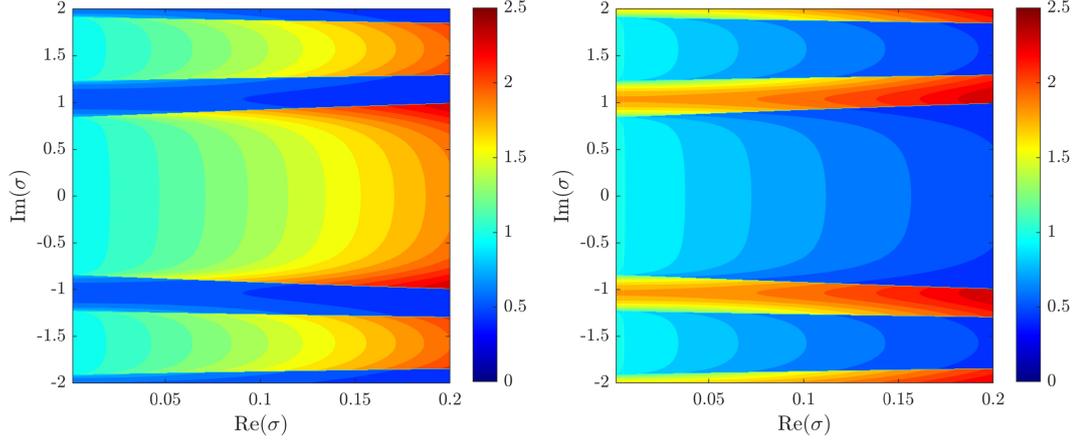


Abbildung 3.3: Absolutbeträge der Lösungen der Riccati-Gleichung im inhomogenen, kontinuierlichen Beispiel:  $|X_{s,+}|$  (links) und  $|X_{s,-}|$  (rechts), dargestellt für den Real- und Imaginärteil von  $\sigma = \sqrt{s}$ .

dass diese genau dann erfüllt ist, wenn

$$e^{2\sigma} = -\frac{2}{3} \pm \frac{i\sqrt{5}}{3} = \exp\left(\mp \arctan\left(\frac{\sqrt{5}}{2}\right) + \pi\right), \quad \text{also}$$

$$\sigma = \frac{i}{2} \left( \mp \arctan\left(\frac{\sqrt{5}}{2}\right) + (2k+1)\pi \right) \quad \forall k \in \mathbb{Z}.$$

Damit können wir folgern, dass Singularitäten von  $\Lambda_s^+$  nur auftreten, falls  $\text{Re}(\sigma) = 0$  gilt, was uns auch Abbildung 3.4 nahe legt.

Es lässt sich außerdem einsehen, dass es Konstanten  $C_1(\sigma_0)$  und  $C_2(\sigma_0, \theta)$  für  $\sigma_0 > 0$  und  $0 < \theta \leq \pi/2$  gibt, so dass für die Abbildung  $\sigma \mapsto \Lambda_{\sigma^2}^+(\varphi)$  die Abschätzungen

$$\begin{aligned} |\Lambda_{\sigma^2}^+(\varphi)| &\leq C_1(\sigma_0) \quad \text{für } \text{Re}(\sigma) > \sigma_0 \quad \text{sowie} \\ |\Lambda_{\sigma^2}^+(\varphi)| &\leq C_2(\sigma_0, \theta) |\sigma|^{-1} \quad \text{für } \text{Re}(\sigma) > \sigma_0 \quad \text{und} \quad |\arg(\sigma)| \leq \frac{\pi}{2} - \theta \end{aligned} \quad (3.21)$$

gelten und damit die Annahme 2.2 erfüllt ist.

**Bemerkung 3.20.** Betrachten wir, wie auch schon im homogenen Fall, die Abbildung  $s \mapsto \Lambda_s^+(\varphi)$ , so lässt sich wiederum einsehen, dass auch die Sektorialitätsannahme 1.2 für den Parameter  $\mu = 1/2$  in einem Sektor  $\Sigma_{\pi/2+\theta_0}$  mit  $0 < \theta_0 \leq \pi/2$  erfüllt ist. Für die Anwendung in Kapitel 5 sind wir in diesem Beispiel jedoch mehr daran interessiert, die Abbildung als  $\sigma \mapsto \Lambda_{\sigma^2}^+(\varphi)$  zu verstehen.

Dasselbe Verhalten können wir in Abbildung 3.5 für den Operator  $\mathcal{S}_s^+$  nachvollziehen. Um diesen angeben zu können, definieren wir folgende von  $\sigma = \sqrt{s}$  abhängige Werte:

$$\begin{aligned} \nu_s &:= 15e^{\sigma(5-x)} + 16e^{\sigma(3-x)} + 9e^{\sigma(1-x)} - 15e^{\sigma(x-1)} - 16e^{\sigma(x+1)} - 9e^{\sigma(x+3)}, \\ \tilde{\nu}_s &:= 9e^{4\sigma x} + 27e^{8\sigma} + 29e^{12\sigma} + 33e^{10\sigma} - 9e^{14\sigma} - 27e^{(4x+6)\sigma} - 29e^{(4x+2)\sigma} - 33e^{4\sigma(x+1)} \end{aligned}$$

Für den Fall  $|X_{s,+}| < 1$  erhalten wir damit

$$\mathcal{S}_s^+(x) = \begin{cases} \frac{-\chi_s (e^{\sigma(1-x)} + e^{\sigma(x-1)}) + \nu_s}{8\sigma (3e^{4\sigma} + 4e^{2\sigma} + 3)}, & \text{falls } x \in (1, 2), \\ \frac{(-\chi_s (e^{10\sigma} + 3e^{8\sigma} + 3e^{\sigma(4x+2)} + e^{4\sigma x}) + \tilde{\nu}_s) e^{\sigma(-2x-5)}}{32\sigma (3e^{4\sigma} + 4e^{2\sigma} + 3)}, & \text{falls } x \in (2, 3). \end{cases}$$

Ist  $|X_{s,+}| > 1$ , so ändern sich in  $\mathcal{S}_s^+(x)$  lediglich die Vorzeichen vor der Wurzel  $\chi_s$ .

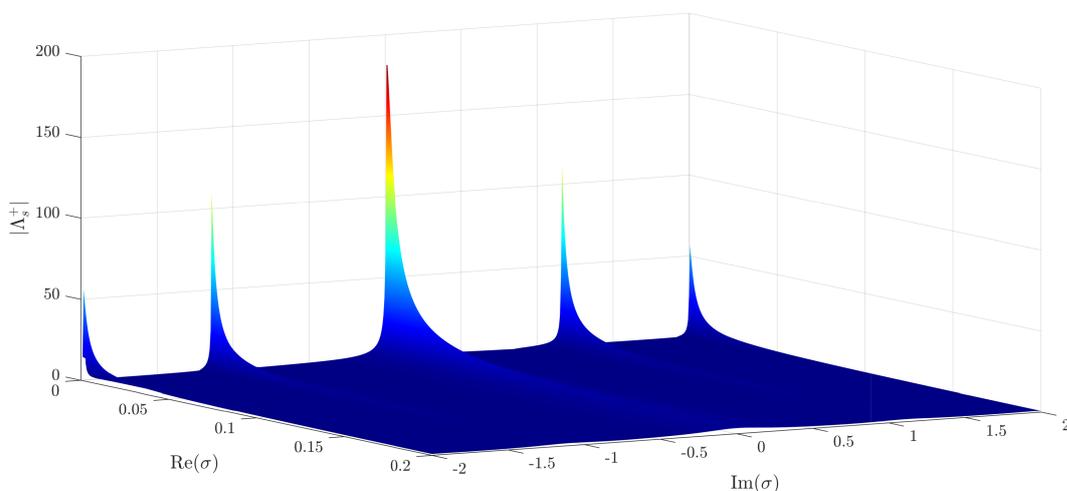


Abbildung 3.4: Absolutbetrag des NtD-Operators  $\Lambda_s^+$  im inhomogenen, kontinuierlichen Beispiel. Singularitäten sind hier zu erkennen bei  $0, \approx \pm 1.15i$  und  $\approx \pm 1.99i$ . Wir erinnern hier an den Zusammenhang  $\sigma = \sqrt{s}$ .

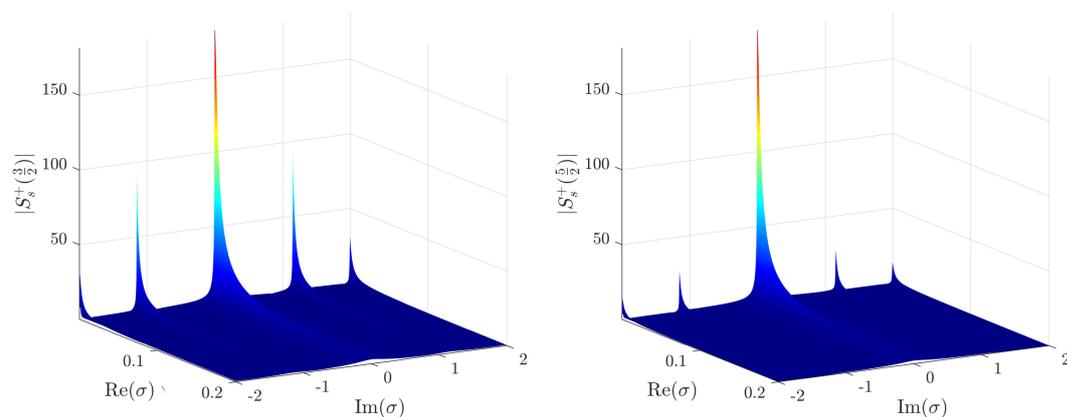


Abbildung 3.5: Absolutbetrag des Operators  $\mathcal{S}_s^+(x)$  im inhomogenen, kontinuierlichen Beispiel für  $x = \frac{3}{2}$  (links) und  $x = \frac{5}{2}$  (rechts). Auch hier sind Singularitäten bei  $0, \approx \pm 1.15i$  und  $\approx \pm 1.99i$  sichtbar.

### 3.2.3 Ortsdiskreter Fall

Für die Diskretisierung im Ort wählen wir ein äquidistantes Gitter auf dem Abschluss der Zelle  $C_0^+$  mit  $m$  Gitterpunkten und Gitterweite  $\Delta x = \frac{2}{m-1}$ . Die Gitterpunkte sind hierbei gegeben durch

$$x_j = (j-1)\Delta x + 1 \quad \text{für } j = 0, \dots, m,$$

das heißt, es gilt  $\overline{C_0^+} = [1, 3] = [x_1, x_m]$ . Weiter setzen wir für  $k = 0, 1$

$$(E_k)_j := e_{s,k}^+(\varphi)(x_j), \quad j = 0, \dots, m.$$

Um die ortsdiskrete Version der Helmholtz-Gleichung (3.14a) zu erhalten, verwenden wir zentrale Finite Differenzen zur Diskretisierung des Laplace-Operators. Da sich der NtD-Operator auf den linken Rand der Zelle  $C_0^+$  bezieht, greifen wir zur Diskretisierung derjenigen Randbedingungen (3.14b), die sich am linken Rand befinden, auf den außerhalb der Zelle liegenden Gitterpunkt  $x_0$  zu, während die Bedingungen am rechten Rand nur mit Hilfe der Gitterpunkte  $x_{m-1}$  und  $x_m$  realisiert werden.

Damit erhalten wir aus (3.14) das diskretisierte System

$$\frac{-(E_k)_{j+1} + 2(E_k)_j - (E_k)_{j-1}}{\Delta x^2} + s\rho(x_j)(E_k)_j = 0 \quad \text{für } j = 1, \dots, m-1, \quad (3.22a)$$

wobei die Randbedingungen durch

$$\left| \begin{array}{l} (E_0)_0 - (E_0)_1 = \Delta x \varphi, \\ (E_0)_m - (E_0)_{m-1} = 0, \end{array} \right. \quad \text{bzw.} \quad \left| \begin{array}{l} (E_1)_0 - (E_1)_1 = 0, \\ (E_1)_m - (E_1)_{m-1} = \Delta x \varphi \end{array} \right. \quad (3.22b)$$

gegeben sind.

#### 3.2.3.1 Homogener Wellenleiter

Wir setzen wie im ortskontinuierlichen Fall  $\rho \equiv 1$ . Damit lässt sich (3.22a) umschreiben zu

$$\lambda(E_k)_{j+1} + \mu_s(E_k)_j + \lambda(E_k)_{j-1} = 0 \quad \text{für } j = 1, \dots, m-1,$$

mit  $\lambda := -\frac{1}{\Delta x^2}$  und  $\mu_s := \frac{2}{\Delta x^2} + s$ .

Hierbei ist die Lösung der charakteristischen Gleichung  $\lambda\zeta^2 + \mu_s\zeta + \lambda = 0$  gegeben durch

$$\begin{aligned} \zeta_{s,\pm} &= \frac{-\mu_s \pm \sqrt{\mu_s^2 - 4\lambda^2}}{2\lambda} = \frac{-\mu_s \pm \sqrt{\mu_s - 2\lambda}\sqrt{\mu_s + 2\lambda}}{2\lambda} \\ &= 1 + \frac{s\Delta x^2}{2} \mp \frac{\sqrt{s}\sqrt{s\Delta x^2 + 4}\Delta x}{2}. \end{aligned} \quad (3.23)$$

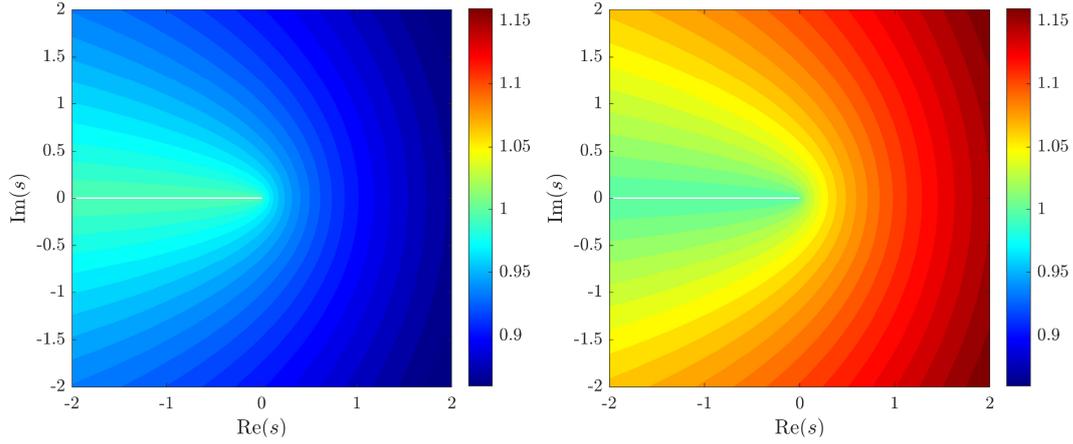


Abbildung 3.6: Absolutbeträge der Nullstellen der charakteristischen Gleichung (3.23) für  $\Delta x = \frac{1}{10}$ :  $|\zeta_{s,+}|$  (links) und  $|\zeta_{s,-}|$  (rechts).

Wie in Abbildung 3.6 dargestellt, gilt hier  $|\zeta_{s,+}| < 1$  für alle  $s \in \mathbb{C} \setminus \mathbb{R}_0^+$ . Setzen wir daher der Übersicht halber  $\zeta := \zeta_{s,+}$ , so lässt sich wegen  $\zeta_{s,+} \cdot \zeta_{s,-} = 1$  die Lösung  $(E_k)_j$  angeben als

$$(E_k)_j = c_{k,1} \zeta^j + c_{k,2} \zeta^{-j} \quad \text{mit} \quad c_{k,l} \in \mathbb{C}, \quad l = 1, 2.$$

Die Konstanten  $c_{k,1}$  und  $c_{k,2}$  können wir anhand der diskreten Randbedingungen (3.22b) für  $k = 0, 1$  eindeutig bestimmen und erhalten damit die Lösungsdarstellungen

$$(E_0)_j = \frac{\zeta (\zeta^{2j} \zeta + \zeta^{2m})}{\zeta^j (\zeta - 1) (\zeta^m - \zeta) (\zeta^m + \zeta)} \Delta x \varphi \quad \text{und}$$

$$(E_1)_j = \frac{\zeta^m \zeta (\zeta^{2j} + \zeta)}{\zeta^j (\zeta - 1) (\zeta^m + \zeta) (\zeta^m - \zeta)} \Delta x \varphi,$$

die uns wiederum die diskreten Operatoren

$$\mathcal{T}_{s,00}^+(\varphi) = (E_0)_1 = \frac{(\zeta^3 + \zeta^{2m}) \Delta x \varphi}{(\zeta - 1) (\zeta^m - \zeta) (\zeta^m + \zeta)},$$

$$\mathcal{T}_{s,01}^+(\varphi) = (E_0)_m = \frac{\zeta \zeta^m (\zeta + 1) \Delta x \varphi}{(\zeta - 1) (\zeta^m - \zeta) (\zeta^m + \zeta)} = (E_1)_1 = \mathcal{T}_{s,10}^+(\varphi) \quad \text{und}$$

$$\mathcal{T}_{s,11}^+(\varphi) = (E_1)_m = \frac{\zeta (\zeta^{2m} + \zeta) \Delta x \varphi}{(\zeta - 1) (\zeta^m - \zeta) (\zeta^m + \zeta)}$$

liefern. Die Lösungen der Ricatti-Gleichung lauten hier  $X_{s,+} = \zeta^{m-1}$  und  $X_{s,-} = \zeta^{1-m}$ . Da  $|\zeta| < 1$  gilt, erhalten wir mit  $\mathcal{P}_s^+ := X_{s,+}$  den Operator, der die Anforderungen aus Satz 3.15 erfüllt. Damit können wir den NtD- und den Lösungsoperator als

$$\Lambda_s^+(\varphi) = -\frac{\zeta}{\zeta - 1} \Delta x \varphi \quad \text{und} \quad (\mathcal{S}_s^+)_j(\varphi) = -\frac{\zeta^j}{\zeta - 1} \Delta x \varphi, \quad j \in \mathbb{N}$$

berechnen. Dass die beiden Abbildung  $s \mapsto \Lambda_s^+(\varphi)$  und  $s \mapsto (\mathcal{S}_s^+)_j(\varphi)$  wieder das gewünschte Abfallverhalten für  $s \in \mathbb{C} \setminus \mathbb{R}_0^+$  zeigen, sehen wir mit folgender Nebenrechnung unter

Verwendung von  $\zeta^{-1} = \zeta_{s,-}$ :

$$-\frac{\zeta \Delta x}{\zeta - 1} = \frac{\Delta x}{\zeta_{s,-} - 1} = \frac{2}{s \Delta x + \sqrt{s} \sqrt{s \Delta x^2 + 4}} = \frac{2}{\sqrt{s}(\sqrt{s \Delta x^2} + \sqrt{s \Delta x^2 + 4})} \quad (3.24)$$

Wegen  $|\sqrt{s \Delta x^2} + \sqrt{s \Delta x^2 + 4}| \geq 2$  folgt sofort  $|\Lambda_s^+| \leq \frac{1}{\sqrt{s}}$ . Unter Ausnutzung von  $j \geq 1$  und  $|\zeta| < 1$  erhalten wir dieselbe Abschätzung für  $|(\mathcal{S}_s^+)_j|$ .

**Bemerkung 3.21.** Wählen wir  $s$  fest, lässt sich mit obiger Rechnung leicht nachvollziehen, dass die diskreten Operatoren für  $\Delta x \rightarrow 0$  gegen die ortskontinuierlichen Operatoren aus Abschnitt 3.2.2.1 konvergieren, das heißt, es gilt

$$\begin{aligned} \mathcal{P}_s^+ &= \left( 1 + \frac{s \Delta x^2}{2} - \frac{\sqrt{s} \sqrt{s \Delta x^2 + 4} \Delta x}{2} \right)^{2/\Delta x} \xrightarrow{\Delta x \rightarrow 0} e^{-2\sqrt{s}}, \\ \Lambda_s^+ &= \frac{2}{\sqrt{s}(\sqrt{s \Delta x^2} + \sqrt{s \Delta x^2 + 4})} \xrightarrow{\Delta x \rightarrow 0} \frac{1}{\sqrt{s}} \end{aligned}$$

sowie

$$(\mathcal{S}_s^+)_j \xrightarrow{\Delta x \rightarrow 0} \frac{e^{\sqrt{s}(1-x)}}{\sqrt{s}} \quad \text{für } j = \frac{x-1}{\Delta x} + 1.$$

### 3.2.3.2 Inhomogener Wellenleiter

Wir betrachten dasselbe Beispiel mit stückweiser konstanter Funktion  $\rho$ , das wir bereits in Abschnitt 3.2.2.2 untersucht haben. Für die Ortsdiskretisierung gehen wir davon aus, dass die Anzahl  $m$  der Gitterpunkte in  $C_0^+ = [1, 3]$  ungerade ist und setzen  $n = \frac{m+1}{2}$ . Damit können wir derselben Idee wie im ortskontinuierlichen Fall folgen und das diskrete Helmholtz-Problem in  $C_0^+$  in zwei Probleme auf kleineren Intervallen zerlegen. Setze dazu

$$\begin{aligned} (V_k)_j &:= v_{s,k}(\varphi, \beta_k)(x_j), \quad j = 0, \dots, n \quad \text{und} \\ (W_k)_j &:= w_{s,k}(\varphi, \beta_k)(x_j), \quad j = n-1, \dots, m, \end{aligned}$$

dann lauten die diskreten Versionen der Helmholtz-Gleichungen (3.17) und (3.18):

$$\begin{aligned} \frac{-(V_k)_{j+1} + 2(V_k)_j - (V_k)_{j-1}}{\Delta x^2} + s(V_k)_j &= 0 \quad \text{für } j = 1, \dots, n-1, \\ \left. \begin{array}{l} (V_0)_0 - (V_0)_1 = \Delta x \varphi, \\ (V_0)_n - (V_0)_{n-1} = \beta_0, \end{array} \right\} & \text{bzw.} \quad \left. \begin{array}{l} (V_1)_0 - (V_1)_1 = 0, \\ (V_1)_n - (V_1)_{n-1} = \beta_1 \end{array} \right\} \end{aligned}$$

sowie

$$\frac{-(W_k)_{j+1} + 2(W_k)_j - (W_k)_{j-1}}{\Delta x^2} + 4s(W_k)_j = 0 \quad \text{für } j = n, \dots, m-1,$$

$$\left| \begin{array}{l} (W_0)_{n-1} - (W_0)_n = -\beta_0, \\ (W_0)_m - (W_0)_{m-1} = 0, \end{array} \right. \quad \text{bzw.} \quad \left| \begin{array}{l} (W_1)_{n-1} - (W_1)_n = -\beta_1, \\ (W_0)_m - (W_0)_{m-1} = \Delta x \varphi \end{array} \right.$$

Somit erhalten wir die Lösung von (3.22) als  $(E_k)_j = (V_k)_j$  für  $j = 0, \dots, n$  und  $(E_k)_j = (W_k)_j$  für  $j = n-1, \dots, m$ .

Wie im vorherigen Abschnitt drücken wir die Diskretisierungen  $(V_k)_j$  und  $(W_k)_j$  mit Hilfe der Wurzeln von quadratischen Gleichungen aus, das heißt, wir verwenden für  $c_{k,l}, d_{k,l} \in \mathbb{C}$ ,  $l = 1, 2$  den Ansatz

$$(V_k)_j = c_{k,1} \zeta^j + c_{k,2} \zeta^{-j} \quad \text{und} \quad (W_k)_j = d_{k,1} \eta^j + d_{k,2} \eta^{-j},$$

wobei mit  $\zeta$  eine der beiden Lösungen  $\zeta_{s,\pm}$  von (3.23) bezeichnet wird und  $\eta$  entsprechend als eine der beiden Lösungen der quadratischen Gleichung

$$-\frac{1}{\Delta x^2} \eta^2 + \left( \frac{2}{\Delta x^2} + 4s \right) \eta - \frac{1}{\Delta x^2} = 0$$

gewählt wird. Unter Verwendung der Randbedingungen und der Forderung  $(V_k)_n = (W_k)_n$  können wir die Konstanten  $c_{k,l}$ ,  $d_{k,l}$  und  $\beta_k$  bestimmen und erhalten als Lösung der Ricatti-Gleichung den Term

$$\begin{aligned} X_{s,\pm}(\zeta, \mu) = & \frac{1}{2\eta^{n+1}\zeta^{n+1}(\zeta^2 - 1)(\eta^2 - 1)} \left( \eta^4 \zeta^4 + \eta^{2n+2} \zeta^{2n+2} - \eta^4 \zeta^{2n} - 2\eta^3 \zeta^3 \right. \\ & + 2\eta^3 \zeta^{2n+1} - \eta^2 \zeta^{2n+2} - \eta^{2n+2} \zeta^2 + 2\eta^{2n+1} \zeta^3 - 2\eta^{2n+1} \zeta^{2n+1} \dots \\ & \left. - \eta^{2n} \zeta^4 + \eta^2 \zeta^2 + \eta^{2n} \zeta^{2n} \dots \right. \\ & \left. \pm \eta^2 \zeta^2 \sqrt{\eta^n \zeta^n - \eta^{n-1} \zeta^{n-1} - \eta^n + \eta^{n-1} \zeta + \eta \zeta^{n-1} - \zeta^n - \eta \zeta + 1} \dots \right. \\ & \sqrt{\eta^n \zeta^n - \eta^{n-1} \zeta^{n-1} - \eta^n + \eta^{n-1} \zeta - \eta \zeta^{n-1} + \zeta^n + \eta \zeta - 1} \dots \\ & \sqrt{\eta^n \zeta^n - \eta^{n-1} \zeta^{n-1} + \eta^n - \eta^{n-1} \zeta - \eta \zeta^{n-1} + \zeta^n - \eta \zeta + 1} \dots \\ & \left. \left. \sqrt{\eta^n \zeta^n - \eta^{n-1} \zeta^{n-1} + \eta^n - \eta^{n-1} \zeta + \eta \zeta^{n-1} - \zeta^n + \eta \zeta - 1} \right) \right). \end{aligned}$$

Im homogenen Beispiel aus dem vorherigen Abschnitt konnten wir noch klar erkennen, welchen Einfluss die Wahl der Wurzel des charakteristischen Polynoms auf den Propagationsoperator hat: Wählen wir dort unabhängig von  $s$  die Wurzel  $\zeta = \zeta_{s,+}$ , führt dies eindeutig zur betragsmäßig kleineren Lösung  $X_{s,+}$  der Ricatti-Gleichung. Hätten wir stattdessen fest  $\zeta = \zeta_{s,-}$  gewählt, wäre  $X_{s,-}$  die gesuchte Lösung gewesen.

Hier sehen wir diesen Zusammenhang dagegen nicht. Es kann beispielsweise sein, dass für ein festes  $s$  jede der vier Kombination  $X_{s,+}(\zeta_{s,\pm}, \eta_{s,\pm})$  einen Wert liefert, der betragsmäßig kleiner als 1 ist. Es stellt sich nun die Frage, welche der Kombinationen zur numerischen Berechnung verwendet werden sollte.

Ein erster Versuch, diese Frage zu beantworten, ist der direkte Vergleich mit den Wurzeln der ortskontinuierlichen Ricatti-Gleichung, die wir in (3.19) berechnet und in Abbildung 3.3 dargestellt haben. Hierzu betrachten wir in Abbildung 3.7 die Bereiche für  $\sigma = \sqrt{s}$  mit

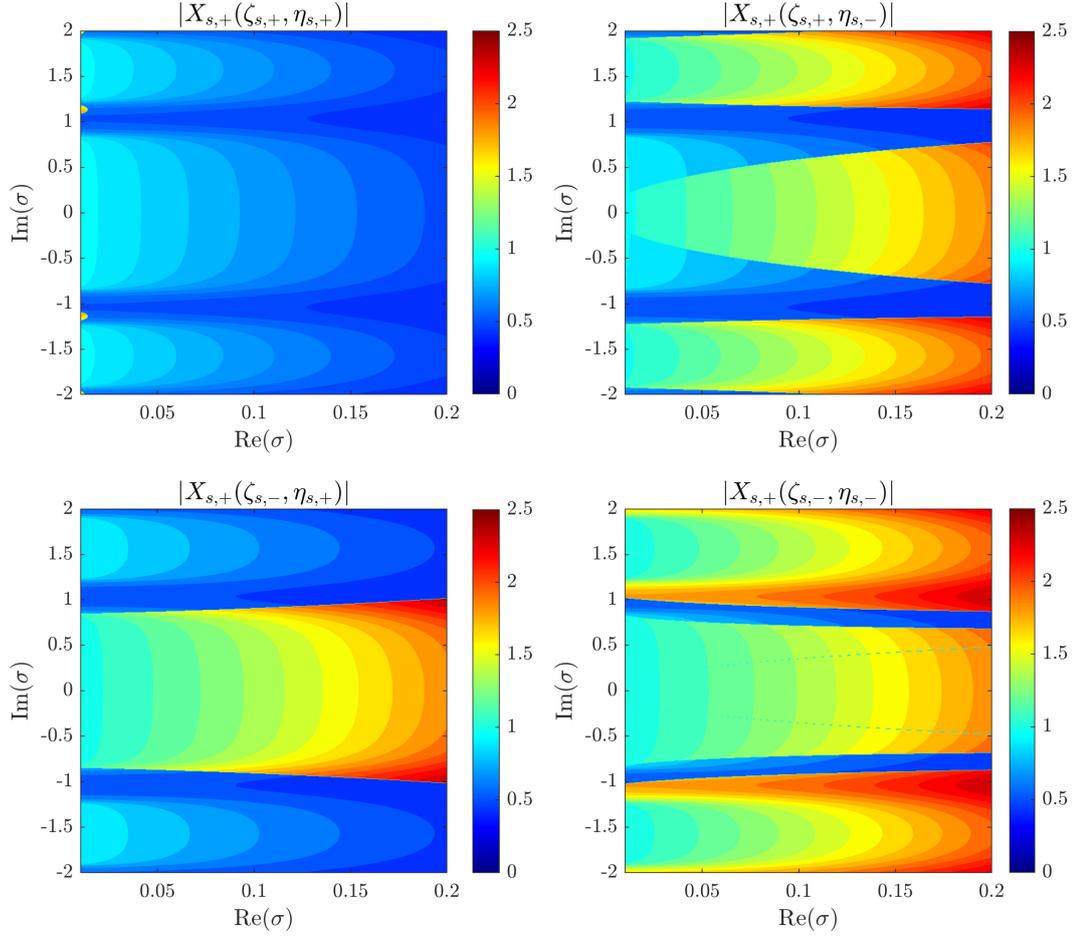


Abbildung 3.7: Absolutbeträge der Lösung  $X_{s,+}(\zeta_{s,\pm}, \eta_{s,\pm})$  der Ricatti-Gleichung im inhomogenen ortsdiskreten Beispiel für  $\Delta x = 10^{-4}$ , zur besseren Vergleichbarkeit mit Abbildung 3.3 dargestellt in Abhängigkeit vom Real- und Imaginärteil von  $\sigma = \sqrt{s}$ .

$\text{Re}(\sigma) \in (0, 0.2)$  und  $\text{Im}(\sigma) \in (-2, 2)$ , in denen die Lösung  $X_{s,+}(\zeta, \eta)$  für die verschiedenen Kombinationen aus  $\zeta = \zeta_{s,\pm}$  und  $\eta = \eta_{s,\pm}$  betragsmäßig kleiner als 1 ist. Vergleichen wir die Abbildungen 3.3 und 3.7, lässt sich hier vermuten, dass sich die Lösung  $X_{s,+}(\zeta, \eta)$  der diskreten Ricatti-Gleichung für immer kleiner werdendes  $\Delta x$  keiner der beiden Lösungen  $X_{s,\pm}$  der kontinuierlichen Ricatti-Gleichung nähert. Die gleiche Beobachtung lässt sich wegen  $X_{s,+}(\zeta, \mu) \cdot X_{s,-}(\zeta, \mu) = 1$  auch für  $X_{s,-}(\zeta, \mu)$  machen. Damit lässt sich anhand der Betrachtung der Wurzeln der Ricatti-Gleichung noch keine Entscheidung darüber treffen, welche der Kombination aus  $\zeta_{s,\pm}$  und  $\eta_{s,\pm}$  gewählt werden sollte.

Betrachten wir statt der einzelnen Wurzeln  $X_{s,\pm}(\zeta, \eta)$  den daraus zusammengesetzten Propagationsoperator

$$\mathcal{P}_s^+(\zeta, \eta) = \begin{cases} X_{s,+}(\zeta, \eta), & \text{falls } |X_{s,+}(\zeta, \eta)| < 1, \\ X_{s,-}(\zeta, \eta), & \text{falls } |X_{s,+}(\zeta, \eta)| > 1, \end{cases}$$

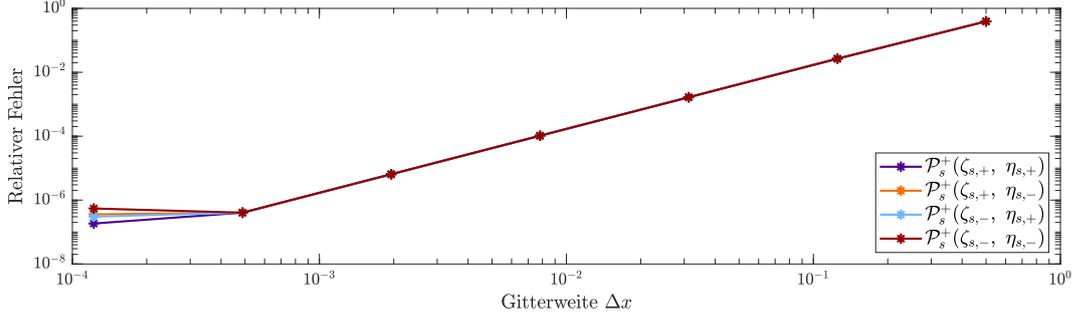


Abbildung 3.8: Relativer Fehler  $\|\mathcal{P}_s^+(\zeta, \eta) - \mathcal{P}_s^+\|_2 / \|\mathcal{P}_s^+\|_2$  in Abhängigkeit der Gitterweite  $\Delta x$  für  $s \in \mathbb{C}$  mit  $\text{Re}(\sigma) \in (0, 0.2)$  und  $\text{Im}(\sigma) \in (-2, 2)$ .

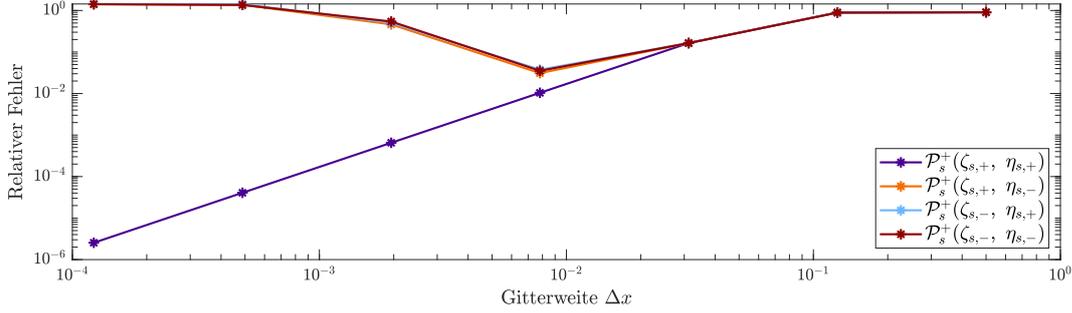


Abbildung 3.9: Relativer Fehler  $\|\mathcal{P}_s^+(\zeta, \eta) - \mathcal{P}_s^+\|_2 / \|\mathcal{P}_s^+\|_2$  in Abhängigkeit der Gitterweite  $\Delta x$  für  $s \in \mathbb{C}$  mit  $\text{Re}(\sigma) \in (0, 10)$  und  $\text{Im}(\sigma) \in (-10, 10)$ .

so legt die für das homogene Beispiel angestellte Beobachtung aus Bemerkung 3.21 nahe, dass wir auch im inhomogenen Fall die Konvergenz des Operators  $\mathcal{P}_s^+(\zeta, \eta)$  gegen dessen ortskontinuierliche Version  $\mathcal{P}_s^+$  aus (3.20) erwarten können. Tatsächlich zeigt Abbildung 3.8 für jede Kombination aus  $\zeta = \zeta_{s,\pm}$  und  $\eta = \eta_{s,\pm}$ , dass  $\mathcal{P}_s^+(\zeta, \eta)$  für  $\Delta x \rightarrow 0$  gegen  $\mathcal{P}_s^+$  konvergiert, wobei hier wieder  $\text{Re}(\sigma) \in (0, 0.2)$  und  $\text{Im}(\sigma) \in (-2, 2)$  betrachtet wird.

Erweitern wir den Bereich der betrachteten Werte auf  $\text{Re}(\sigma) \in (0, 10)$  sowie  $\text{Im}(\sigma) \in (-10, 10)$ , lässt sich beobachten, dass die Konvergenz nur noch für  $\zeta = \zeta_{s,+}$  und  $\eta = \eta_{s,+}$  besteht, vergleiche Abbildung 3.9, weswegen wir uns nun für diese Kombination entscheiden.

Daher bestimmen wir den NtD-Operator  $\Lambda_s^+(\zeta_{s,+}, \eta_{s,+})$  sowie den Lösungsoperator

$$(\mathcal{S}_s^+(\zeta_{s,+}, \eta_{s,+}))_j = \begin{cases} (\mathcal{S}_{s,1}^+(\zeta_{s,+}, \eta_{s,+}))_j, & \text{für } j = 1, \dots, n, \\ (\mathcal{S}_{s,2}^+(\zeta_{s,+}, \eta_{s,+}))_j, & \text{für } j = n, \dots, m \end{cases}$$

analog zum vorherigen Abschnitt, geben diese aufgrund ihrer Länge aber nicht weiter an. Wie in Abbildung 3.10 deutlich wird, lässt sich hier wieder die Konvergenz der diskreten Operatoren gegen ihre kontinuierlichen Versionen aus Abschnitt 3.2.2.2 beobachten. Es lässt sich schließen, dass die Operatoren wie im ortskontinuierlichen Fall das gewünschte Abfallverhalten für  $\text{Re}(\sigma) > 0$  aus Annahme 2.2 sowie Singularitäten für  $\text{Re}(\sigma) = 0$  vorweisen.

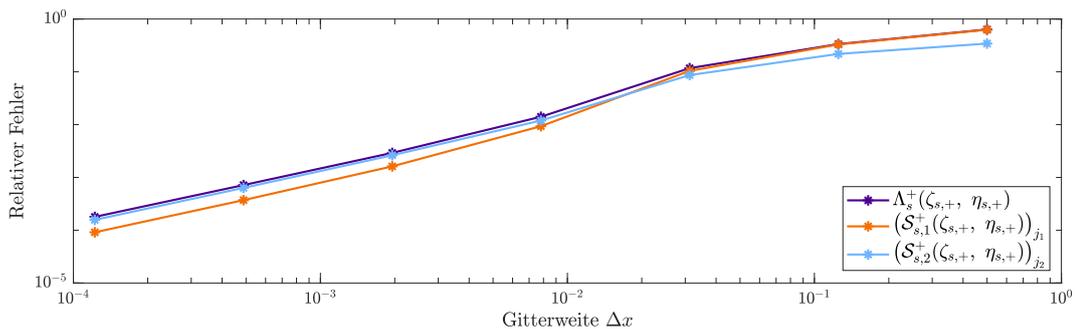


Abbildung 3.10: Relative Fehler  $\|\Lambda_s^+(\zeta_{s,+}, \eta_{s,+}) - \Lambda_s^+\|_2 / \|\Lambda_s^+\|_2$ ,  $\|(\mathcal{S}_{s,1}^+(\zeta_{s,+}, \eta_{s,+}))_{j_1} - \mathcal{S}_s^+(3/2)\|_2 / \|\mathcal{S}_s^+(3/2)\|_2$  und  $\|(\mathcal{S}_{s,2}^+(\zeta_{s,+}, \eta_{s,+}))_{j_2} - \mathcal{S}_s^+(5/2)\|_2 / \|\mathcal{S}_s^+(5/2)\|_2$  für  $j_1 = \frac{3/2-1}{\Delta x} + 1$  und  $j_2 = \frac{5/2-1}{\Delta x} + 1$  in Abhängigkeit der Gitterweite  $\Delta x$  für  $\sigma = \sqrt{s}$  mit  $\text{Re}(\sigma) \in (0, 10)$  und  $\text{Im}(\sigma) \in (-10, 10)$ .



## KAPITEL 4

---

# ZEIT-FRAKTIONALE SCHRÖDINGER-GLEICHUNG AUF DER REELLEN ACHSE

Für eine erste Anwendung, in der die beiden behandelten Themenblöcke „Faltungsquadratur“ und „transparente Randbedingungen“ zusammen gebracht werden, betrachten wir in diesem Kapitel die Zeit-fractionale Schrödinger-Gleichung im homogenen eindimensionalen Wellenleiter, ehe wir im nächsten Kapitel einen Blick auf die Wellengleichung im periodischen zweidimensionalen Wellenleiter werfen. Die fraktionale Ableitung in der Zeit verstehen wir hierbei als Differentiation nach *Caputo*. Wie wir sehen werden, führt die Anwendung der Resultate aus dem vorherigen Kapitel auf die fraktionale Schrödinger-Gleichung zu transparenten Randbedingungen im Zeitbereich und einer zellenweisen Lösung in den Halb-Wellenleitern, die in Form von Faltungen dargestellt werden können. Die Laplace-Transformierten der Faltungskerne erfüllen hierbei die Sektorialitätsannahme 1.2.

Als Motivation für die Betrachtung der fraktionalen Schrödinger-Gleichung in diesem Kapitel ist einerseits das in den letzten Jahrzehnten zunehmend gestiegene Interesse an fraktionalen Differentialgleichungen zu nennen. So werden Prozesse in der theoretischen Physik immer häufiger mit Hilfe von fraktionalen Gleichungen modelliert, besonders dann, wenn es um die Beschreibung von Phänomenen aus der Natur geht. Ein Stichwort ist hierbei die anomale Diffusion, bei der davon ausgegangen wird, dass sich eine Partikelwolke auf andere Weise verteilt als die klassische Diffusionsgleichung vorhersagen würde. Wir verweisen auf die Referenzen in [12, 25, 38] für eine Übersicht über Anwendungen und Beispiele.

Andererseits werden wir sehen, dass neben den zeitabhängigen transparenten Randbedingungen auch die Lösung der fraktionalen Schrödinger-Gleichung im Inneren als Faltung geschrieben werden kann, deren Faltungskern ebenfalls die Sektorialitätsannahme 1.2 erfüllt. Damit lässt sich anhand der fraktionalen Schrödinger-Gleichung und der zugehörigen transparenten Randbedingungen eine kompliziertere Anwendung der exponentiellen Quadratur beziehungsweise des schnellen Faltungsalgorithmus aus Kapitel 1 zeigen als dem numerischen Experiment in Abschnitt 1.5 zugrunde lag.

Untersuchungen von fraktionalem Differentialgleichungen zu theoretischen Aspekten, wie die Voraussetzungen zur Existenz und Eindeutigkeit einer Lösung, sind beispielsweise in [5, 9, 42] zu finden. Hier weisen wir im Besonderen auf die Dissertation [5] hin, die die Grundlage für die Lösungsdarstellung der fraktionalem Differentialgleichung in diesem Kapitel darstellt. Dort wird unter anderem für die Differentiation nach Caputo der Fall ohne Quellfunktion, also  $f \equiv 0$ , sowie unter Verwendung der Differentiation nach *Riemann-Liouville* der Fall einer beliebigen Quellfunktion betrachtet. Für die vollständige Lösungsdarstellung unter Annahme der Caputo-Ableitung, auf die wir im folgenden Kapitel kurz eingehen werden, verweisen wir zusätzlich auf [18].

Nachdem wir das Modellproblem und insbesondere die Differentiation nach Caputo vorgestellt haben, werden wir in diesem Kapitel kurz auf die zur Existenz- und Eindeutigkeitsaussage notwendige Theorie eingehen, aus der wir auch die Darstellung der Lösung der fraktionalem Schrödinger-Gleichung auf der reellen Achse erhalten. Durch den vorgestellten Zusammenhang mit der Helmholtz-Gleichung und den Ideen aus Kapitel 3 können wir die transparenten Randbedingungen sowie die zellenweise Lösungsdarstellung herleiten. Im Anschluss gehen wir auf die Anwendung des schnellen Faltungsalgorithmus ein und illustrieren die Ergebnisse anhand eines numerischen Experiments.

## 4.1 Bestimmung der transparenten Randbedingungen

### 4.1.1 Modellproblem

Wie bei den numerischen Experimenten in Abschnitt 3.2 setzen wir in diesem Kapitel  $\Omega = \mathbb{R}$  und  $\Omega_0 = (a^-, a^+)$ . Die Zellen  $C_j^\pm$  sowie die Ränder  $\Gamma_j^\pm$  sind entsprechend der Vereinfachung aus Abschnitt 3.2.1 charakterisiert.

Auf der reellen Achse, das heißt auf  $\Omega = \mathbb{R}$ , suchen wir für einen Parameter  $\alpha \in (0, 1)$  die Lösung  $u : \mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{C}$  der fraktionalem Schrödinger-Gleichung

$$\begin{cases} -i\Delta u(x, t) + D_t^\alpha u(x, t) = f(x, t) & \text{in } \mathbb{R} \times \mathbb{R}^+, \\ u(x, 0) = u_0(x) & \text{in } \mathbb{R}, \end{cases} \quad (\mathbf{S})$$

wobei die  $L^2$ -Realisierung des Schrödinger-Operators entsprechend zum Operator  $A$  aus Abschnitt 3.1.2 gegeben ist durch

$$i\Delta : D(i\Delta) := H^1(\Delta, \mathbb{R}) \rightarrow L^2(\mathbb{R}), \quad v \mapsto i\Delta v(x) := i \frac{\partial^2 v}{\partial x^2}(x).$$

Wie bereits erwähnt, verstehen wir den Operator  $D_t^\alpha$  als fraktionalem Differentiation nach Caputo der Ordnung  $\alpha$ , die definiert ist als

$$D_t^\alpha u(\cdot, t) := \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-\tau)^{-\alpha} \frac{\partial u}{\partial t}(\cdot, \tau) d\tau, \quad (4.1)$$

siehe dazu [50, 56]. Der Vorteil der Caputo-Differentiation gegenüber anderen Definitionen von fraktionaler Differentiation, wie beispielsweise der in der Einleitung des Kapitels erwähnten Definition nach Riemann-Liouville, liegt in den benötigten Anfangswerten. Während die physikalische Bedeutung der Anfangswerte für die Caputo-Differentiation klar erkennbar ist, werden für andere Definitionen meist Anfangswerte benötigt, die schwerer zu bestimmen sind und nicht immer in einen einfachen physikalischen Kontext gebracht werden können, vergleiche hierzu [35, 56].

Da wir nun ein zeitabhängiges Problem betrachten, müssen die Annahmen über die Quellfunktion aus Abschnitt 3.1.1 entsprechend angepasst werden. Anstelle der Annahmen (H3) und (H4) fordern wir nun

$$(S1) \quad f \in C^1(\mathbb{R}_0^+, L^2(\mathbb{R})) \text{ und}$$

$$(S2) \quad \text{supp}(f(\cdot, t)) \subset \Omega_0 \quad \forall t \in \mathbb{R}_0^+.$$

Zusätzlich nehmen wir über die Anfangsfunktion  $u_0$  an:

$$(S3) \quad u_0 \in D(i\Delta) \text{ sowie}$$

$$(S4) \quad \text{supp}(u_0) \subset \Omega_0$$

An dieser Stelle sei erwähnt, dass wir in Annahme (S3) eine stärkere Forderung als notwendig stellen, um im Faltungsalgorithmus aus Abschnitt 4.2 eine vereinfachte Lösungsdarstellung verwenden zu können. Für die Überlegungen im aktuellen Abschnitt 4.1 wäre die Annahme  $u_0 \in L^2(\mathbb{R})$  bereits ausreichend.

### 4.1.2 Existenz und Eindeutigkeit einer Lösung

Um die Existenz und Eindeutigkeit der Lösung der fraktionalen Schrödinger-Gleichung zu beweisen und eine Darstellung der Lösung als Faltungsintegral zu erhalten, geben wir im Folgenden kurz die wichtigsten theoretischen Ergebnisse aus [5] wieder, die ihrerseits auf der Theorie über Volterra'sche Integralgleichungen aus [58] basieren, und ergänzen diese durch die Überlegungen in [18].

Wir stellen hierzu die Funktion  $\Phi_\beta : \mathbb{R} \rightarrow \mathbb{R}_0^+$  vor, die für  $\beta \geq 0$  definiert ist durch

$$\Phi_\beta(t) := \begin{cases} \frac{1}{\Gamma(\beta)} t^{\beta-1}, & t > 0, \\ 0, & t \leq 0. \end{cases} \quad (4.2)$$

Es lässt sich leicht einsehen, dass diese eine Halbgruppen-Eigenschaft erfüllt, genauer gilt  $(\Phi_\alpha * \Phi_\beta)(t) = \Phi_{\alpha+\beta}(t)$  für  $\alpha, \beta \geq 0$ , vergleiche dazu [50, Kap. 1.1].

Mit dem im vorherigen Abschnitt angegebenen Definitionsbereich des Operators  $i\Delta$  als  $H^1(\Delta, \mathbb{R})$  definieren wir die (starke) Lösung der fraktionalen Schrödinger-Gleichung wie folgt, vergleiche dazu [5]:<sup>1</sup>

<sup>1</sup>In [5] wird in Anlehnung an [58] der Begriff der „starken“ Lösung verwendet. In anderen Quellen wird dagegen für diese Definition der Begriff der „klassischen“ Lösung vergeben, siehe zum Beispiel [54].

**Definition 4.1.** Eine Funktion  $u \in C(\mathbb{R}_0^+, L^2(\mathbb{R}))$  ist eine (starke) Lösung von **(S)**, wenn

- (i)  $u \in C(\mathbb{R}_0^+, H^1(\Delta, \mathbb{R}))$ ,
- (ii)  $\Phi_{1-\alpha} * (u - u_0) \in C^1(\mathbb{R}_0^+, L^2(\mathbb{R}))$  sowie
- (iii)  $u$  die fraktionale Gleichung **(S)** für  $t \in \mathbb{R}_0^+$  erfüllt.

Die Begründung für Bedingung (ii) ist hierbei im Zusammenhang zwischen der Differentiation nach Caputo und der nach Riemann-Liouville zu finden: Für, bezogen auf die Zeit, stetiges  $u$  gilt

$$D_t^\alpha u(t) = {}^{\text{RL}}D_t^\alpha (u(t) - u_0) = \frac{\partial}{\partial t} (\Phi_{1-\alpha} * (u - u_0))(t),$$

vergleiche dazu [50, Kapitel 1.2].

Ist  $u$  eine starke Lösung der fraktionalen Schrödinger-Gleichung, so können wir wie in [5] das fraktionale Riemann-Liouville-Integral auf **(S)** anwenden, was nichts anderes als eine Faltung beider Seiten der Differentialgleichung mit der Funktion  $\Phi_\alpha$  bedeutet. Aus (4.1) und der Halbgruppen-Eigenschaft von (4.2) erhalten wir so die Volterra-Gleichung

$$u(t) = u_0 + (\Phi_\alpha * i\Delta u)(t) + (\Phi_\alpha * f)(t). \quad (4.3)$$

Diese Umformung erlaubt es uns, auf das Konzept der Lösungsoperatoren beziehungsweise Resolventen zurückzugreifen, das grundlegend für die Theorie der linearen Volterra-Integralgleichungen ist, und die folgende Definition anzugeben, vergleiche [5, 58]:

**Definition 4.2.** Eine Familie  $\{S_\alpha(t)\}_{t \geq 0} \subset \mathcal{L}(L^2(\mathbb{R}))$ , die die Bedingungen

- (i)  $S_\alpha(t)$  ist stark stetig für  $t \in \mathbb{R}_0^+$  und  $S_\alpha(0) = \text{Id}$ ;
- (ii)  $S_\alpha(t)u_0 \in H^1(\Delta, \mathbb{R})$  und  $\Delta S_\alpha(t)u_0 = S_\alpha(t)\Delta u_0$  für alle  $u_0 \in H^1(\Delta, \mathbb{R})$  und  $t \in \mathbb{R}_0^+$ ;
- (iii) die Resolventengleichung

$$S_\alpha(t)u_0 = u_0 + (\Phi_\alpha * i\Delta S_\alpha u_0)(t)$$

gilt für alle  $u_0 \in H^1(\Delta, \mathbb{R})$  und  $t \in \mathbb{R}_0^+$

erfüllt, wird *Lösungsoperator* der Gleichung (4.3) genannt. Falls sich  $S_\alpha(t)$  zusätzlich analytisch in den Sektor  $\Sigma_{\rho_0}$  für  $\rho_0 \in (0, \pi/2]$  fortsetzen lässt und falls für jedes  $\rho < \rho_0$  und  $\omega > 0$  eine Konstante  $C(\rho, \omega) > 0$  existiert, so dass

$$\|S_\alpha(t)\| \leq C(\rho, \omega) e^{\omega \text{Re}(t)} \quad \text{für } t \in \Sigma_\rho$$

gilt, bezeichnet man den Operator als analytisch vom Typ  $(\rho_0, 0)$ .

Existiert der Lösungsoperator  $\{S_\alpha(t)\}_{t \geq 0}$  der Gleichung (4.3), so ist dieser nach [58, Kor. 1.1] eindeutig. Dass der Lösungsoperator tatsächlich existiert, lässt sich mit Hilfe von Theorem 2.14 aus [5] zeigen und wird im folgenden Satz angegeben:

**Satz 4.3.** *Der Lösungsoperator  $\{S_\alpha(t)\}_{t \geq 0}$  der fraktionalen Schrödinger-Gleichung (4.3) existiert und ist analytisch vom Typ  $(\theta_\alpha, 0)$ , wobei der Winkel  $\theta_\alpha \in (0, \pi/2]$  von der Ordnung  $\alpha$  der fraktionalen Differentiation abhängt.*

*Beweis.* Nach Satz 3.8 und der Überlegung in Bemerkung 3.18 ist der Laplace-Operator  $\Delta : H^1(\Delta, \mathbb{R}) \rightarrow L^2(\mathbb{R})$  selbstadjungiert, wodurch sofort folgt, dass der Schrödinger-Operator  $i\Delta : H^1(\Delta, \mathbb{R}) \rightarrow L^2(\mathbb{R})$  schiefselbstadjungiert ist. Mit dem Satz von Stone, [55, Thm. 1.10.8], generiert  $i\Delta$  damit eine unitäre  $C_0$ -Kontraktionsgruppe auf  $L^2(\mathbb{R})$ . Weiter erhalten wir aus dem Satz von Hille-Yosida, [55, Thm. 1.3.1], für die Resolvente  $R(\lambda, i\Delta) := (\lambda \text{Id} - i\Delta)^{-1}$  die Abschätzung

$$\|R(\lambda, i\Delta)\|_{\mathcal{L}(L^2(\mathbb{R}))} \leq \frac{1}{\lambda} \quad \text{für } \lambda > 0.$$

Insbesondere gilt hier  $\sigma(i\Delta) \subset i\mathbb{R}$ . Es lässt sich daraus nun mit Hilfe von Proposition 2.1.1 aus [31] folgern, dass die Resolvente  $R(\lambda, -i\Delta)$  für  $\lambda \in \mathbb{C} \setminus \overline{\Sigma_{\pi/2+\theta}}$  mit jedem Winkel  $0 < \theta < \pi/2$  analytisch ist und die Abschätzung

$$\|R(\lambda, -i\Delta)\|_{\mathcal{L}(L^2(\mathbb{R}))} \leq \frac{C(\theta)}{|\lambda|} \quad \text{für } \lambda \in \mathbb{C} \setminus \overline{\Sigma_{\pi/2+\theta}}$$

für eine von  $\theta$  abhängige Konstante  $C(\theta) \geq 1$  erfüllt. Damit ergibt sich sofort die äquivalente Abschätzung

$$\|R(\lambda, i\Delta)\|_{\mathcal{L}(L^2(\mathbb{R}))} \leq \frac{C(\theta)}{|\lambda|} \quad \text{für } \lambda \in \Sigma_{\pi/2-\theta}. \quad (4.4)$$

Um nun die Existenz des Lösungsoperators  $\{S_\alpha(t)\}_{t \geq 0}$  zu erhalten, betrachten wir die Resolvente  $R(\lambda^\alpha, i\Delta)$  für  $\lambda^\alpha \in \Sigma_{\pi/2-\theta}$ . Setzen wir wie in der Überlegung aus Abschnitt 1.5 den Winkel  $\theta_\alpha$  als

$$\theta_\alpha := \min \left( \frac{\pi(1-\alpha) - 2\theta}{2\alpha}, \frac{\pi}{2} \right), \quad (4.5)$$

so gilt  $|\arg(\lambda^\alpha)| < \pi/2 - \theta$  für  $\lambda \in \Sigma_{\pi/2+\theta_\alpha}$ . Damit ist  $R(\lambda^\alpha, i\Delta)$  analytisch im Sektor  $\Sigma_{\pi/2+\theta_\alpha}$  und wir erhalten aus (4.4) die Abschätzungen

$$\|R(\lambda^\alpha, i\Delta)\|_{\mathcal{L}(L^2(\mathbb{R}))} \leq \frac{C(\theta)}{|\lambda|^\alpha} \quad \text{und} \quad \|\lambda^{\alpha-1} R(\lambda^\alpha, i\Delta)\|_{\mathcal{L}(L^2(\mathbb{R}))} \leq \frac{C(\theta)}{|\lambda|}$$

für  $\lambda \in \Sigma_{\pi/2+\theta_\alpha}$ . Auf Grund der zweiten Abschätzung lässt sich [5, Thm. 2.14] anwenden und liefert die Existenz eines analytischen Lösungsoperators  $\{S_\alpha(t)\}_{t \geq 0}$  vom Typ  $(\theta_\alpha, 0)$ .  $\square$

Wie in [5] weiter gezeigt wird, kann der Lösungsoperator  $\{S_\alpha(t)\}_{t \geq 0}$  als die inverse

Laplace-Transformierte der operatorwertigen Funktion  $\lambda \mapsto \lambda^{\alpha-1}R(\lambda^\alpha, i\Delta)$  definiert werden. Da diese nach den gerade getroffenen Überlegungen die Sektorialitätsannahme 1.2 im Sektor  $\Sigma_{\pi/2+\theta_\alpha}$  erfüllt, können wir  $S_\alpha(t)$  für  $t \in \mathbb{R}^+$  nach Satz 1.4 schreiben als

$$\begin{aligned} S_\alpha(t) &= \mathcal{L}^{-1}\{\lambda^{\alpha-1}R(\lambda^\alpha, i\Delta); t\} \\ &= \frac{1}{2\pi i} \int_{\Gamma} e^{\lambda t} \lambda^{\alpha-1}R(\lambda^\alpha, i\Delta) d\lambda, \end{aligned} \quad (4.6)$$

wobei die Kurve  $\Gamma$  in  $\Sigma_{\pi/2+\theta}$  mit  $\theta < \theta_\alpha$  liegt, in einem spitzen Winkel zur negativen reellen Achse gegen unendlich läuft und gegen den Uhrzeigersinn orientiert ist.

Um die starke Lösung der fraktionalen Gleichung (S) mit beliebigem Quellterm, der (S1) erfüllt, angeben zu können, definieren wir analog zum Lösungsoperator  $S_\alpha(t)$  einen Operator  $P_\alpha(t)$  für  $t \in \mathbb{R}^+$  durch

$$\begin{aligned} P_\alpha(t) &:= \mathcal{L}^{-1}\{R(\lambda^\alpha, i\Delta); t\} \\ &= \frac{1}{2\pi i} \int_{\Gamma} e^{\lambda t} R(\lambda^\alpha, i\Delta) d\lambda, \end{aligned} \quad (4.7)$$

wobei dieselbe Kurve  $\Gamma$  wie in (4.6) gewählt werden kann. Der Operator  $P_\alpha(t)$  erfüllt einige nützliche Eigenschaften, die wir im Folgenden angeben:

**Lemma 4.4.** *Für den Operator  $P_\alpha(t)$  aus (4.7) gilt*

- (i)  $P_\alpha(t) \in \mathcal{L}(L^2(\mathbb{R}))$  und  $P_\alpha(t)u_0 \in H^1(\Delta, \mathbb{R})$  für  $t \in \mathbb{R}^+$  und  $u_0 \in L^2(\mathbb{R})$ ;
- (ii)  $(\Phi_{1-\alpha} * P_\alpha u_0)(t) = S_\alpha(t)u_0$  für  $t \in \mathbb{R}^+$  und  $u_0 \in L^2(\mathbb{R})$ .

*Beweis.* Der Beweis der Eigenschaft (i) ist in [5, Lemma 4.14] zu finden. Eigenschaft (ii) lässt sich leicht mit Hilfe der Laplace-Transformation und dem Faltungssatz zeigen, wobei wir verwenden, dass  $\mathcal{L}\{\Phi_\alpha(t); \lambda\} = \lambda^{-\alpha}$  für  $\lambda \in \Sigma_{\pi/2+\theta_\alpha}$ .  $\square$

Damit lässt sich nun die starke Lösung der fraktionalen Schrödinger-Gleichung wie folgt angeben:

**Satz 4.5.** *Erfülle der Quellterm  $f$  der Gleichung (S) die Voraussetzung (S1) und die Anfangsfunktion  $u_0$  die Annahme (S3). Dann ist die in Definition 4.1 definierte starke Lösung von (S) für  $t \in \mathbb{R}_0^+$  gegeben durch*

$$u(t) = S_\alpha(t)u_0 + (P_\alpha * f)(t). \quad (4.8)$$

*Beweis.* Dass  $S_\alpha(t)u_0$  für  $t \geq 0$  und jedes  $u_0 \in L^2(\mathbb{R})$  die eindeutige starke Lösung der Gleichung

$$D_t^\alpha u(x, t) = i \Delta u(x, t) \quad \text{in } \mathbb{R} \times \mathbb{R}^+ \quad \text{mit } u(x, 0) = u_0(x) \quad \text{in } \mathbb{R}$$

ist, wird in [5, Prop. 2.15] gezeigt und hier nicht weiter angegeben.  
Gibt es zudem eine starke Lösung  $v$  der Gleichung

$$D_t^\alpha v(x, t) = i \Delta v(x, t) + f(x, t) \quad \text{in } \mathbb{R} \times \mathbb{R}^+ \quad \text{mit } v(x, 0) = 0 \quad \text{in } \mathbb{R},$$

so lässt sich deren Eindeutigkeit einfach einsehen. Hierfür verweisen wir auf [18, Thm. 2.5].  
Für den Beweis der Existenz greifen wir erneut auf die Theorie über Volterra-Integralgleichungen aus [58] zurück. Nach [58, Prop. 1.2] liegt die Funktion

$$\tilde{v}(t) := (\Phi_\alpha * S_\alpha f(0))(t) + (\Phi_\alpha * S_\alpha * f')(t)$$

in  $C(\mathbb{R}_0^+, H^1(\Delta, \mathbb{R}))$  und löst die Volterra-Gleichung (4.3) für den Fall  $u_0 = 0$ . Weiter erhalten wir mit der starken Stetigkeit von  $S_\alpha(t)$  aus [1, Prop. 1.3.6] die Gleichung

$$(S_\alpha * f)'(t) = (S_\alpha * f')(t) + S_\alpha(t)f(0) \quad \text{für } t \in \mathbb{R}_0^+,$$

insbesondere ist  $(S_\alpha * f) \in C^1(\mathbb{R}_0^+, L^2(\mathbb{R}))$  und damit

$$(\Phi_\alpha * (S_\alpha * f)')(t) = (\Phi_\alpha * S_\alpha * f)'(t) \in C(\mathbb{R}_0^+, L^2(\mathbb{R})).$$

Nun folgt aus Lemma 4.4 und der Halbgruppen-Eigenschaft von  $\Phi_\alpha$  weiter für  $t \in \mathbb{R}_0^+$

$$(\Phi_\alpha * S_\alpha * f)(t) = (\Phi_\alpha * \Phi_{1-\alpha} * P_\alpha * f)(t) = (1 * P_\alpha * f)(t),$$

wodurch  $\tilde{v}(t) = (P_\alpha * f)(t)$  impliziert wird. Insbesondere ist damit  $(P_\alpha * f)(t) \in C(\mathbb{R}_0^+, H^1(\Delta, \mathbb{R}))$  und es gilt  $(P_\alpha * f)(0) = 0$ . Aus Lemma 4.4 erhalten wir zusätzlich  $(\Phi_{1-\alpha} * \tilde{v})(t) = (S_\alpha * f)(t) \in C(\mathbb{R}_0^+, L^2(\mathbb{R}))$ . Damit erfüllt  $\tilde{v}$  die Bedingungen (i) und (ii) aus Definition 4.1. Aus der Äquivalenz von (S) und (4.3) erhalten wir schließlich, dass die starke Lösung  $u$  von (S) gegeben ist durch

$$u(t) = S_\alpha(t)u_0 + \tilde{v}(t) = S_\alpha(t)u_0 + (P_\alpha * f)(t).$$

□

**Bemerkung 4.6.** Betrachten wir anstelle der fraktionalen Schödinger-Gleichung eine fraktionale Gleichung der Form

$$D_t^\alpha u(x, t) = au(x, t) + f(x, t) \quad \text{in } \mathbb{R} \times \mathbb{R}^+ \quad \text{mit } u(x, 0) = u_0 \quad \text{in } \mathbb{R},$$

wobei  $a \in \mathbb{C}$  gegeben ist, so lässt sich (4.8) auch über die verallgemeinerte Mittag-Leffler Funktion aus Abschnitt 1.5 als

$$u(t) = \mathbf{e}_{\alpha,1}(t; a)u_0 + \int_0^t \mathbf{e}_{\alpha,\alpha}(t - \tau; a)f(\tau) \, d\tau$$

darstellen, vergleiche dazu [56].

### 4.1.3 Zusammenhang mit Helmholtz-Gleichung

Um die Methode aus dem vorherigen Kapitel zur Bestimmung der transparenten Randbedingungen und zur Berechnung der Lösung außerhalb des Gebiets  $\Omega_0$  anwenden zu können, transformieren wir die fraktionale Schrödinger-Gleichung, die im Zeitbereich zu lösen ist, mit Hilfe der Laplace-Transformation in eine Gleichung im Frequenzbereich. Dazu nehmen wir an, dass das Laplace-Integral der Quellfunktion  $f$  für  $\lambda \in \Sigma_{\pi/2+\theta_\alpha}$  mit dem Winkel  $\theta_\alpha$  aus (4.5) konvergiert und daher

$$F(\cdot, \lambda) := \mathcal{L}\{f(\cdot, t); \lambda\} \quad \text{für } \lambda \in \Sigma_{\pi/2+\theta_\alpha}$$

wohldefiniert ist. Für die starke Lösung  $u$  von (S) können wir damit und aus der Darstellung aus Satz 4.5 folgern, dass

$$U(\cdot, t) := \mathcal{L}\{u(\cdot, t); \lambda\} \quad \text{für } \lambda \in \Sigma_{\pi/2+\theta_\alpha}$$

ebenfalls wohldefiniert ist.

Die Laplace-Transformation der Gleichung (S) liefert uns somit eine Helmholtz-Gleichung der Form

$$-i\Delta U(x, \lambda) + \lambda^\alpha U(x, \lambda) = F(x, \lambda) + \lambda^{\alpha-1} u_0(x) \quad \text{in } \mathbb{R} \times \Sigma_{\pi/2+\theta_\alpha}. \quad (4.9)$$

Multiplizieren wir (4.9) auf beiden Seiten mit  $-i$ , erhalten wir die Gleichung

$$-\Delta U(x, \lambda) - i\lambda^\alpha U(x, \lambda) = -iF(x, \lambda) - i\lambda^{\alpha-1} u_0 \quad \text{in } \mathbb{R} \times \Sigma_{\pi/2+\theta_\alpha}, \quad (4.10)$$

wodurch sich nun ein direkter Vergleich mit der Helmholtz-Gleichung (H) ziehen lässt. Anstelle der Frequenz  $s$  kommt nun die Frequenz  $-i\lambda^\alpha$  in der Gleichung vor. Da wir hierbei von  $\lambda \in \Sigma_{\pi/2+\theta_\alpha}$  ausgehen, ist  $\text{Re}(\lambda^\alpha) > 0$  nach der Überlegung aus dem Beweis von Satz 4.3. Damit folgt  $-i\lambda^\alpha \in \mathbb{C} \setminus \mathbb{R}_0^+$ , also erfüllt  $-i\lambda^\alpha$  die Annahme (H5).

Unter den Voraussetzungen (S1)-(S4) erhalten wir zusätzlich

$$-i\tilde{F} := -i(F + \lambda^{\alpha-1} u_0) \in L^2(\mathbb{R}) \quad \text{und} \quad \text{supp}(\tilde{F}) \subset \Omega_0 = (a^-, a^+),$$

wodurch der Quellterm  $-i\tilde{F}$  der Gleichung (4.10) die Annahmen aus Proposition 3.11 erfüllt. Damit erhalten wir die Lösung  $U$  der Helmholtz-Gleichung (4.9) beziehungsweise (4.10) eingeschränkt auf das Intervall  $\Omega_0 = (a^-, a^+)$  als die Lösung  $U^0$  des Problems

$$\begin{cases} -i\Delta U^0 + \lambda^\alpha U^0 = \tilde{F} & \text{in } \Omega_0, \\ U^0 - \Lambda_{(-i\lambda^\alpha)}^- \left( \frac{\partial U^0}{\partial x} \Big|_{x=a^-} \right) = 0, & \text{für } x = a^-, \\ U^0 - \Lambda_{(-i\lambda^\alpha)}^+ \left( -\frac{\partial U^0}{\partial x} \Big|_{x=a^+} \right) = 0 & \text{für } x = a^+. \end{cases} \quad (4.11)$$

**Bemerkung 4.7.** Da wir im Folgenden den Parameter  $s = -i\lambda^\alpha$  im Gegensatz zum vorher-

rigen Kapitel nicht mehr als fest gewählt annehmen, sondern den NtD-Operator  $\Lambda^\pm$  aus (3.3) als Abbildung von  $\Sigma_{\pi/2+\theta_\alpha}$  nach  $\mathcal{L}(H^{-1/2}(\{a^\pm\}), H^{1/2}(\{a^\pm\}))$  verstehen, schreiben wir nun  $\Lambda^\pm(-i\lambda^\alpha)$  anstelle von  $\Lambda^\pm_{(-i\lambda^\alpha)}$ . Die geänderte Notation übernehmen wir auch für  $\mathcal{P}^\pm_{(-i\lambda^\alpha)}$  und  $\mathcal{S}^\pm_{(-i\lambda^\alpha)}$ .

Eine Darstellung der transparenten Randbedingungen im Zeitbereich erhalten wir durch die Rücktransformation mit Hilfe der inversen Laplace-Transformation. Wenden wir diese auf das lokale Helmholtz-Problem (4.11) an, erhalten wir die Einschränkung der fraktionalen Schrödinger-Gleichung (S) auf das Intervall  $\Omega_0$  durch

$$\left\{ \begin{array}{ll} -i\Delta u + D_t^\alpha u = f & \text{in } \Omega_0 \times \mathbb{R}^+, \\ u(a^-, \cdot) = \lambda_\alpha^- \left( \frac{\partial u}{\partial x} \Big|_{x=a^-} \right) & \text{in } \mathbb{R}^+, \\ u(a^+, \cdot) = \lambda_\alpha^+ \left( -\frac{\partial u}{\partial x} \Big|_{x=a^+} \right) & \text{in } \mathbb{R}^+, \\ u(\cdot, 0) = u_0 & \text{in } \Omega_0, \end{array} \right. \quad (\mathbf{S}^0)$$

wobei wir den NtD-Operator der fraktionalen Gleichung im Zeitbereich als Abbildung

$$\lambda_\alpha^\pm : \left( \mathbb{R}^+ \rightarrow H^{-1/2}(\{a^\pm\}) \right) \rightarrow H^{1/2}(\{a^\pm\})$$

verstehen können und, wie in der folgenden Überlegung deutlich wird, über die inverse Laplace-Transformation des NtD-Operators (3.3) für die Helmholtz-Gleichung erhalten: Dazu seien

$$\Phi^\pm : \Sigma_{\pi/2+\theta_\alpha} \rightarrow H^{-1/2}(\{a^\pm\}) \quad \text{und} \quad \varphi^\pm : \mathbb{R}^+ \rightarrow H^{-1/2}(\{a^\pm\}),$$

wobei  $\Phi^\pm(\lambda) = \mathcal{L}\{\varphi^\pm(t); \lambda\}$  gelte. Unter Verwendung des inversen Faltungssatzes und Bemerkung 4.7 lässt sich der NtD-Operator  $\lambda_\alpha^\pm$  über

$$\begin{aligned} \lambda_\alpha^\pm(\varphi^\pm(t)) &:= \mathcal{L}^{-1}\{\Lambda^\pm(-i\lambda^\alpha)\Phi^\pm(\lambda); t\} \\ &= \left( \mathcal{L}^{-1}\{\Lambda^\pm(-i\lambda^\alpha); \cdot\} * \mathcal{L}^{-1}\{\Phi^\pm(\lambda); \cdot\} \right)(t) \\ &= \int_0^t \mathcal{L}^{-1}\{\Lambda^\pm(-i\lambda^\alpha); t-\tau\} \varphi^\pm(\tau) \, d\tau \end{aligned} \quad (4.12)$$

definieren.

**Bemerkung 4.8.** Mit demselben Vorgehen können wir auch eine Darstellung der Lösung in den Intervallen  $C_j^\pm$  des Außenraums, deren Länge  $L^\pm$  wir hier wegen des homogenen Wellenleiters als beliebig annehmen können, angeben: Die inverse Laplace-Transformation der zellenweisen Lösungsdarstellung im Frequenzbereich, vergleiche Satz 3.13, liefert

$$\begin{aligned} \mathcal{L}^{-1}\left\{ U(\Phi^\pm(\lambda)) \Big|_{C_j^\pm}; t \right\} &= \mathcal{L}^{-1}\left\{ \left( \mathcal{S}^\pm \circ (\mathcal{P}^\pm)^j \right) (-i\lambda^\alpha) \Phi^\pm(\lambda); t \right\} \\ &= \left( \mathcal{L}^{-1}\left\{ \left( \mathcal{S}^\pm \circ (\mathcal{P}^\pm)^j \right) (-i\lambda^\alpha); \cdot \right\} * \mathcal{L}^{-1}\{\Phi^\pm(\lambda); \cdot\} \right)(t) \end{aligned}$$

$$= \left( \mathcal{L}^{-1} \left\{ \left( \mathcal{S}^\pm \circ (\mathcal{P}^\pm)^j \right) (-i\lambda^\alpha); \cdot \right\} * \varphi^\pm(\cdot) \right) (t),$$

womit wir unter Verwendung der Lösung der lokalen fraktionalen Schrödinger-Gleichung ( $\mathbf{S}^0$ ) die zellenweise Lösung durch

$$u(t)|_{C_j^\pm} = \left( \mathcal{L}^{-1} \left\{ \left( \mathcal{S}^\pm \circ (\mathcal{P}^\pm)^j \right) (-i\lambda^\alpha); \cdot \right\} * \left( \mp \frac{\partial u(\cdot)}{\partial x} \Big|_{x=a^\pm} \right) \right) (t) \quad (4.13)$$

erhalten.

## 4.2 Schneller Faltungsalgorithmus für die fraktionale Schrödinger-Gleichung

### 4.2.1 Anwendung der exponentiellen Faltungsquadratur

Da die fraktionale Schrödinger-Gleichung eine Subdiffusion beschreibt, sind wir eher daran interessiert, die Lösung in einem Intervall  $\Omega_0$  zu erhalten, als diese in den Außenraum fortzusetzen. Die Vorgehensweise zur Berechnung der zellenweisen Lösung wird zudem im nächsten Kapitel für die Anwendung auf die Wellengleichung im zweidimensionalen Wellenleiter genauer betrachtet, siehe Abschnitt 5.2.1.1, und lässt sich mit leichten Modifikationen, wie beispielsweise die Verwendung der exponentiellen Quadratur anstelle der Runge-Kutta basierten Faltungsquadratur, einfach auf den Fall der fraktionalen Schrödinger-Gleichung anwenden, weswegen sie hier nicht weiter thematisiert wird. Wir werden uns daher in diesem Kapitel auf einen schnellen Algorithmus zur Berechnung der fraktionalen Gleichung ( $\mathbf{S}^0$ ) konzentrieren.

Wie bereits erwähnt, wollen wir in diesem ausnutzen, dass sowohl die Lösung der fraktionalen Schrödinger-Gleichung auf der reellen Achse als auch die NtD-Operatoren als Faltungsintegrale dargestellt werden können. Dazu vereinfachen wir die Darstellung aus Satz 4.5, indem wir feststellen, dass sich die Lösung  $u$  des Randwertproblems ( $\mathbf{S}$ ) schreiben lässt als  $u = v + u_0$ , wobei  $v$  gegeben ist als die Lösung des Problems

$$\begin{cases} -i\Delta v + D_t^\alpha v = f + i\Delta u_0 & \text{in } \mathbb{R} \times \mathbb{R}^+, \\ v(\cdot, 0) = 0 & \text{in } \mathbb{R}. \end{cases} \quad (4.14)$$

Hierbei können wir  $v$  nach Satz 4.5 schreiben als

$$v(x, t) = \int_0^t P_\alpha(t - \tau) (f(x, \tau) + i\Delta u_0(x)) d\tau. \quad (4.15)$$

Insbesondere erfüllt die Laplace-Transformierte  $R(\lambda^\alpha, i\Delta)$  des Faltungskerns  $P_\alpha(t)$ , wie im Beweis von Satz 4.5 gezeigt, die Sektorialitätsannahme 1.2 mit dem Parameter  $\mu = \alpha$  im Sektor  $\Sigma_{\pi/2+\theta_\alpha}$ .

Da wir uns somit im Ausgangspunkt von Kapitel 1 befinden, können wir die dort beschriebene exponentielle Quadratur zur Approximation der Faltung (4.15) verwenden. Für die Knoten  $0 \leq c_1 < \dots < c_s \leq 1$ ,  $N$  Zeitschritte und Schrittweite  $h$  erhalten wir nach Satz 1.8 zum Zeitpunkt  $t_N = Nh$  die Approximation an die Lösung  $v(x, t_N)$  durch

$$v_N(x) = h \sum_{j=0}^{N-1} \sum_{i=1}^s w_{i,N-1-j} (f(x, t_j + c_i h) + i\Delta u_0(x)), \quad (4.16a)$$

wobei die Faltungsgewichte durch

$$w_{i,j} = \frac{1}{2\pi i} \int_{\Gamma} R(\lambda^\alpha, i\Delta) e^{\lambda(jh)} b_i(h\lambda) d\lambda \quad (4.16b)$$

mit den Funktionen  $b_i(z)$  aus (1.8) gegeben sind. Von der Kurve  $\Gamma$  nehmen wir dabei an, dass sie die Voraussetzungen aus Satz 1.4 im Sektor  $\Sigma_{\pi/2+\theta_\alpha}$  erfüllt. Wie in Satz 1.9 gezeigt wurde, können wir hier eine Konvergenz der Ordnung  $s$  erwarten.

Wenden wir die obige Umformung  $u = v + u_0$  auf das lokale Problem  $(\mathbf{S}^0)$  an, erhalten wir unter der Annahme (S4) das Problem

$$\left\{ \begin{array}{ll} -i\Delta v + D_t^\alpha v = f + i\Delta u_0 & \text{in } \Omega_0 \times \mathbb{R}^+, \\ v(a^-, \cdot) = \lambda_\alpha^- \left( \frac{\partial v}{\partial x} \Big|_{x=a^-} \right) & \text{in } \mathbb{R}^+, \\ v(a^+, \cdot) = \lambda_\alpha^+ \left( -\frac{\partial v}{\partial x} \Big|_{x=a^+} \right) & \text{in } \mathbb{R}^+, \\ v(\cdot, 0) = 0 & \text{in } \Omega_0, \end{array} \right. \quad (4.17)$$

bei dem sich nun die Frage nach der Diskretisierung der NtD-Operatoren  $\lambda^\pm$  stellt. Diese sind gemäß (4.12) ebenfalls als Faltung definiert, wobei der Faltungskern durch die Funktion  $t \mapsto \mathcal{L}^{-1}\{\Lambda^\pm(-i\lambda^\alpha); t\}$  gegeben ist. Nun können wir auf die Resultate aus Abschnitt 3.2.2.1 zurückgreifen: Die Laplace-Transformierte  $\Lambda^+(-i\lambda^\alpha)$  des Faltungskerns lässt sich nach (3.15) wegen des Zusammenhangs  $s = -i\lambda^\alpha$  schreiben als die Abbildung

$$\lambda \mapsto \Lambda^+(-i\lambda^\alpha) = \frac{1}{\sqrt{-i\lambda^\alpha}}.$$

Wie sich leicht nachvollziehen lässt, gilt hier wegen  $\rho \equiv 1$  die Gleichheit  $\Lambda^+ \equiv \Lambda^-$ . Mit den Überlegungen in Abschnitt 3.2.2.1 folgt nun, dass die Laplace-Transformierten  $\Lambda^\pm(-i\lambda^\alpha)$  analytisch für  $\lambda \in \Sigma_{\pi/2+\theta_\alpha}$  sind und dort der Abschätzung

$$|\Lambda^\pm(-i\lambda^\alpha)| \leq \frac{C}{|\lambda|^{\alpha/2}}$$

genügen. Damit ist die Sektorialitätsannahme 1.2 im gleichen Sektor wie für die Laplace-Transformierte der Lösungsdarstellung erfüllt.

Wie gerade gesehen, können wir somit die exponentielle Quadratur zur Approximation der

transparenten Randbedingungen zum Zeitpunkt  $t_N$  verwenden, wodurch wir die Gleichung

$$v(a^\pm, t_N) = h \sum_{j=0}^{N-1} \sum_{i=1}^s \tilde{w}_{i,N-1-j} \left( \mp \frac{\partial}{\partial x} v(x, t_j + c_i h) \Big|_{x=a^\pm} \right) \quad (4.18a)$$

mit den Faltungsgewichten

$$\tilde{w}_{i,j} = \frac{1}{2\pi i} \int_{\Gamma} \Lambda^\pm(-i\lambda^\alpha) e^{\lambda(jh)} b_i(h\lambda) d\lambda \quad (4.18b)$$

erhalten. Hierbei kann dieselbe Kurve  $\Gamma$  wie für die Faltungsgewichte  $w_{i,j}$  gewählt werden.

Da wir dasselbe Faltungsquadratur-Verfahren sowohl für die Lösungsdarstellung (4.15) als auch für die NtD-Operatoren (4.12) verwenden können, ist nun der folgende Zusammenhang leicht nachvollziehbar:

**Korollar 4.9.** *Die Zeitdiskretisierung des lokalen Problems (4.17) unter Verwendung der exponentiellen Quadratur mit  $s$  Knoten zur Approximation von (4.12) und (4.15) liefert uns ein konvergentes Verfahren der Ordnung  $s$ .*

*Beweis.* Wir bezeichnen mit  $\tilde{v}_N(x)$  die Lösung der semidiskreten Version von (4.17) und (4.18). Dann gilt

$$\|\tilde{v}_N(x) - v(x, t_N)\|_{0,\Omega_0} \leq \|\tilde{v}_N(x) - v_N(x)\|_{0,\Omega_0} + \|v_N(x) - v(x, t_N)\|_{0,\Omega_0}.$$

Nach Konstruktion gilt  $\tilde{v}_N(x) = v_N(x)$  für alle  $x \in \Omega_0$ , wodurch der erste Summand in obiger Abschätzung verschwindet. Wegen

$$\|\tilde{v}_N(x) - v(x, t_N)\|_{0,\Omega_0} \leq \|v_N(x) - v(x, t_N)\|_{0,\Omega_0} \leq \|v_N(x) - v(x, t_N)\|_{0,\mathbb{R}}$$

und Satz 1.9 folgt damit aus der Konvergenz  $s$ -ter Ordnung der Diskretisierung von (4.14) direkt die Konvergenz der Diskretisierung von (4.17).  $\square$

## 4.2.2 Diskretisierung im Ort

Auch wenn wir gerade schon die zeitdiskrete, aber noch ortskontinuierliche, Version des lokalen Problems (4.17) untersucht und deren Konvergenz festgestellt haben, setzen wir für die Herleitung der Ortsdiskretisierung einen Schritt vorher an und betrachten stattdessen wie bereits in Abschnitt 4.1.3 die Laplace-Transformation von (4.17), die gegeben ist durch das lokale Helmholtz-Problem

$$\left\{ \begin{array}{ll} -i\Delta V + \lambda^\alpha V = F + i\Delta u_0 & \text{in } \Omega_0 := (-1, 1), \\ V - \Lambda^-(-i\lambda^\alpha) \left( \frac{\partial V}{\partial x} \Big|_{x=-1} \right) = 0 & \text{für } x = -1, \\ V - \Lambda^+(-i\lambda^\alpha) \left( -\frac{\partial V}{\partial x} \Big|_{x=1} \right) = 0 & \text{für } x = 1, \end{array} \right. \quad (4.19)$$

wobei  $V(x, \lambda)$  die Laplace-Transformierte der Lösung  $v(x, t)$  von (4.17) bezeichnet und wir  $\Omega_0$  zur Vereinfachung wie angegeben annehmen.

Da wir die Ergebnisse aus Abschnitt 3.2.3.1 verwenden wollen, diskretisieren wir den Ort hierbei mittels zentraler Finiten Differenzen. Dazu wählen wir auf dem Intervall  $\overline{\Omega_0} = [-1, 1]$  ein äquidistantes Gitter mit  $M$  Gitterpunkten und Gitterweite  $\Delta x = \frac{2}{M-1}$ , das heißt, die Gitterpunkte sind gegeben durch  $x_m = (m-1)\Delta x - 1$  für  $m = 1, \dots, M$ . Bezeichnen wir mit  $V_m(\lambda)$  die semidiskrete Näherung an  $V(x_m, \lambda)$ , so liefert uns die Anwendung der zentralen Finiten Differenzen auf das Helmholtz-Problem für  $m = 2, \dots, M-1$  die Gleichungen

$$\begin{aligned} & -i \left( \frac{V_{m+1}(\lambda) - 2V_m(\lambda) + V_{m-1}(\lambda)}{\Delta x^2} \right) + \lambda^\alpha V_m(\lambda) \\ & = F(x_m, \lambda) + i \left( \frac{u_0(x_{m+1}) - 2u_0(x_m) + u_0(x_{m-1}))}{\Delta x^2} \right). \end{aligned}$$

Nach (3.24) und da hier  $\Lambda^+ = \Lambda^-$  gilt, sind die ortsdiskreten NtD-Operatoren, die unter Verwendung der zentralen Finiten Differenzen berechnet wurden, für  $\varphi \in \mathbb{C}$  gegeben durch

$$\begin{aligned} \Lambda^\pm(-i\lambda^\alpha)(\varphi) &= \frac{\Delta x}{\zeta(\lambda) - 1} \varphi \quad \text{für } \lambda \in \Sigma_{\pi/2 + \theta_\alpha}, \quad \text{wobei} \\ \zeta(\lambda) &:= 1 + \frac{(-i\lambda^\alpha)\Delta x^2}{2} + \frac{\sqrt{-i\lambda^\alpha} \sqrt{-i\lambda^\alpha \Delta x^2 + 4\Delta x}}{2}. \end{aligned}$$

Bemerke dabei, dass die Nullstelle  $\zeta(\lambda)$  des zugrunde liegenden charakteristischen Polynoms (3.23) hierbei so gewählt ist, dass  $|\zeta(\lambda)| > 1$  für alle  $\lambda \in \Sigma_{\pi/2 + \theta_\alpha}$  gilt.

Passend zu den Randbedingungen der lokalen Probleme in den Zellen  $C_0^\pm$ , die in Abschnitt 3.2.3 zur Herleitung der ortsdiskreten Operatoren verwendet wurden, approximieren wir die Ableitungen an den Stellen  $x_1 = -1$  und  $x_M = 1$  mittels des Differenzenquotienten durch

$$\frac{\partial}{\partial x} V(x_1, \lambda) \approx \frac{V_2(\lambda) - V_1(\lambda)}{\Delta x} \quad \text{und} \quad -\frac{\partial}{\partial x} V(x_M, \lambda) \approx -\frac{V_M(\lambda) - V_{M-1}(\lambda)}{\Delta x}.$$

Damit erhalten wir die ortsdiskreten Versionen der transparenten Randbedingungen von (4.19) durch

$$\begin{aligned} V_1(\lambda) &= \Lambda^-(-i\lambda^\alpha) \left( \frac{V_2(\lambda) - V_1(\lambda)}{\Delta x} \right) = \frac{V_2(\lambda) - V_1(\lambda)}{\zeta(\lambda) - 1} \quad \text{sowie} \\ V_M(\lambda) &= \Lambda^+(-i\lambda^\alpha) \left( -\frac{V_M(\lambda) - V_{M-1}(\lambda)}{\Delta x} \right) = \frac{V_{M-1}(\lambda) - V_M(\lambda)}{\zeta(\lambda) - 1}. \end{aligned}$$

Für die semidiskreten Vektoren  $\mathbf{V}(\lambda) := (V_1(\lambda), \dots, V_M(\lambda))^T$ ,  $\mathbf{F}(\lambda) := (0, F(x_2, \lambda), \dots, F(x_{M-1}, \lambda), 0)^T$  und  $\mathbf{u}_0 := (u_0(x_1), \dots, u_0(x_M))^T$  lässt sich die ortsdiskrete Version von (4.19) damit schreiben als das System

$$\begin{aligned} \lambda^\alpha \mathbf{M}\mathbf{V}(\lambda) &= \mathbf{A}\mathbf{V}(\lambda) + \mathbf{B}(\lambda)\mathbf{V}(\lambda) + \Delta x \mathbf{F}(\lambda) + \mathbf{A}\mathbf{u}_0 \quad \text{bzw.} \\ (\lambda^\alpha \mathbf{M} - \mathbf{A} - \mathbf{B}(\lambda)) \mathbf{V}(\lambda) &= \Delta x \mathbf{F}(\lambda) + \mathbf{A}\mathbf{u}_0, \end{aligned}$$



wobei die Ortsdiskreten Faltungsgewichte gegeben sind durch

$$\mathbf{w}_{i,j} = \frac{1}{2\pi i} \int_{\Gamma} (\lambda^\alpha \mathbf{M} - \mathbf{A} - \mathbf{B}(\lambda))^{-1} e^{\lambda(jh)} b_i(h\lambda) d\lambda, \quad (4.21b)$$

vergleiche (4.16) und (4.18). Von der Kurve  $\Gamma$  nehmen wir weiterhin an, dass sie die Annahmen aus Satz 1.4 im Sektor  $\Sigma_{\pi/2+\theta_\alpha}$  erfüllt.

Parametrisieren wir die Kurve  $\Gamma$  gemäß der Überlegungen aus Abschnitt 1.3 mit unterschiedlichen Hyperbeln, so lässt sich leicht einsehen, dass der schnelle Faltungsalgorithmus zur Berechnung der Approximation an die Lösung  $v$  von (4.17) eine direkte Anwendung des schnellen Faltungsalgorithmus aus Abschnitt 1.4 ist. Für die Details den Algorithmus betreffend, verweisen wir auf das entsprechende Kapitel.

**Bemerkung 4.11.** Wie zu Beginn von Abschnitt 1.4 diskutiert, bietet sich der angegebene Algorithmus beispielsweise dann an, wenn die Lösung nur zum Endzeitpunkt  $t_N$  bestimmt werden soll, jedoch der zeitliche Verlauf nicht weiter von Interesse ist. Wollen wir die Lösung in  $\Omega_0$  zellenweise fortsetzen, werden jedoch die Auswertungen der Richtungsableitungen der Lösung an den Punkten  $(a^\pm, t_j + c_i h)$  für  $i = 0, \dots, s$  und  $j = 0, \dots, N - 1$  beziehungsweise Approximationen daran benötigt, vergleiche (4.13). Daher eignet sich der hier beschriebene Ansatz nur für die Behandlung der fraktionalen Gleichung im Intervall  $\Omega_0$ .

Zur vollständigen Beschreibung des Algorithmus, der die Approximation an die Lösung  $u$  des Problems  $(\mathbf{S}^0)$  berechnet, bleibt somit nur noch die Rücksubstitution durch den Zusammenhang  $u = v + u_0$  zu erwähnen.

## 4.3 Numerisches Experiment

Als Testproblem betrachten wir die fraktionale Schrödinger-Gleichung

$$\begin{cases} D_t^\alpha u(x, t) = i\Delta u(x, t) & \text{in } \mathbb{R} \times \mathbb{R}^+, \\ u(x, 0) = \chi_{[-0.99, 0.99]}(x) 10 \exp(-(8x)^2 + 10ix) & \text{in } \mathbb{R}, \end{cases}$$

wobei wir die Ordnung der fraktionalen Ableitung auf  $\alpha = 3/4$  festlegen. Nach (4.5) setzen wir daher  $\theta_\alpha = \pi/2$ . Wir berechnen die Approximation an die Lösung  $u$  von  $(\mathbf{S}^0)$  in  $\Omega_0 = (-1, 1)$  wie im vorherigen Abschnitt beschrieben, das heißt insbesondere verwenden wir die in Abschnitt 3.2.3.1 unter Verwendung von zentralen Finiten Differenzen hergeleitete Darstellung der NtD-Operatoren im entsprechenden Helmholtz-Problem und diskretisieren die Faltungsintegrale mit Hilfe der exponentiellen Quadratur. Ein analoges Beispiel wurde in [18] untersucht, jedoch wurden hier, wie bereits erwähnt, kompakte Finite Differenzen zur Ortsdiskretisierung sowie die Runge-Kutta basierte Faltungsquadratur für den Fall einer sektoriellen Laplace-Transformierten verwendet, siehe dazu auch [45].

Um die Genauigkeit der Diskretisierung der transparenten Randbedingungen  $\lambda_\alpha^\pm$  (4.12) zu testen, berechnen wir die Lösung  $u$  von  $(\mathbf{S}^0)$ , wobei wir das Intervall  $\overline{\Omega_0} = [-1, 1]$  mit

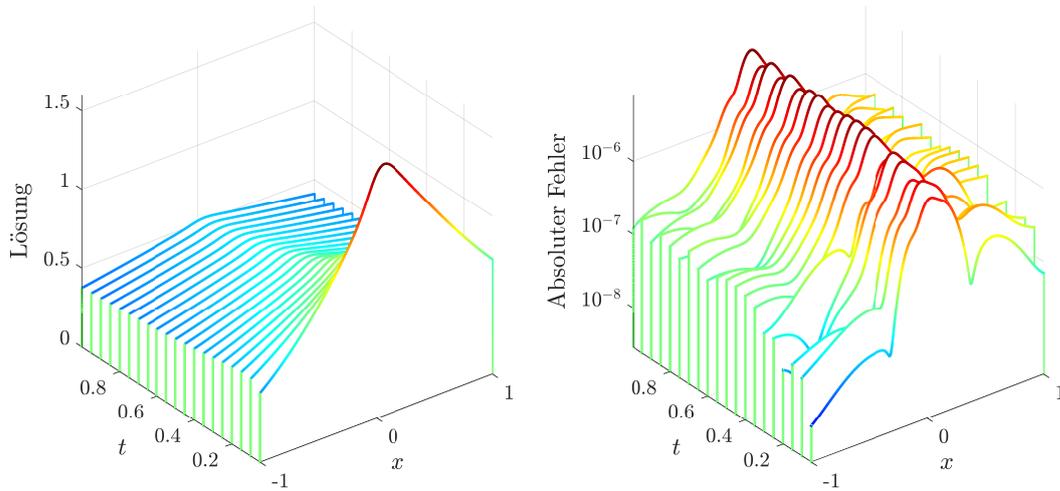


Abbildung 4.1: Absolutbetrag der Lösung und absoluter Fehler zu den Zeiten  $t = 0.05, 0.1, \dots, 1$  im Intervall  $\Omega_0 = (-1, 1)$  für die Parameter  $B = 5$  und  $\kappa = 50$ . Die  $z$ -Achse in der rechten Achse ist hierbei logarithmiert.

$M = 801$  Gitterpunkten diskretisieren und die Hyperbeln, die gemäß Abschnitt 1.3 zur Diskretisierung von (4.21b) gewählt werden, für den Wert  $\kappa = 50$  an  $2\kappa + 1$  Stützstellen auswerten. Hierbei setzen wir  $B = 5$  für die Aufteilung der Intervalle  $I_\ell$  aus (1.10). Für die Referenzlösung betrachten wir das Intervall  $[-8, 8]$  mit 6401 Gitterpunkten und setzen  $\kappa = 110$ . In beiden Fällen approximieren wir die Faltung (4.21) mit Hilfe der exponentiellen Quadratur der Ordnung  $s = 3$ , in der die Polynominterpolation unter Verwendung der Tschebyscheff-Knoten durchgeführt wird, und setzen die Zeitschrittweite auf  $h = 1/400$ . In Abbildung 4.1 stellen wir den Absolutbetrag der berechneten Lösung und des absoluten Fehlers in  $\Omega_0$  zu den Zeitpunkten  $t = 0.05, 0.1, \dots, 1$  dar. In der Abbildung lässt sich gut erkennen, dass der Fehler vom Inneren des Intervalls  $\Omega_0$  ausgeht, jedoch nicht von den transparenten Randbedingungen.

Haben wir in Abbildung 4.1 den festen Wert  $\kappa = 50$  verwendet, so wollen wir nun ein besseres Verständnis für den Zusammenhang zwischen der Anzahl der Stützstellen zur Diskretisierung der Hyperbeln und dem beobachteten Fehler erlangen. Dazu betrachten wir in Abbildung 4.2 den absoluten Fehler der berechneten Lösung in Abhängigkeit des Parameters  $\kappa$ . Die linke Seite der Abbildung zeigt dabei den Fehler zum Zeitpunkt  $t = 0.5$  mit Schrittweite  $h = 1/400$ , das heißt  $N = 200$  Zeitschritten, und mit den verschiedenen Werten  $M = 201, 401, 801$ . Für die rechte Seite der Abbildung berechnen wir den Fehler zum selben Zeitpunkt mit fester Wahl  $M = 401$  und betrachten verschiedene Werte für den Zeitschritt  $N$ . Die Referenzlösungen erhalten wir hierbei erneut mit der Wahl  $\kappa = 110$  und  $B = 5$  und unter Verwendung derselben Zeit- und Ortsschrittweite wie für die Approximation der jeweiligen Lösung in  $\Omega_0$ . Damit sehen wir in Abbildung 4.2 den Fehler, der durch die Approximation der Wegintegrale, siehe (4.21b) und (1.12), entsteht. Hier lässt sich die exponentielle Konvergenz, die wir in (1.14) angeben haben, klar erkennen.

An dieser Stelle lässt sich ein Vorteil der exponentiellen Faltungsquadratur gegenüber der Runge-Kutta basierten Faltungsquadratur, die in [18] betrachtet wurde, hervorheben: Dort

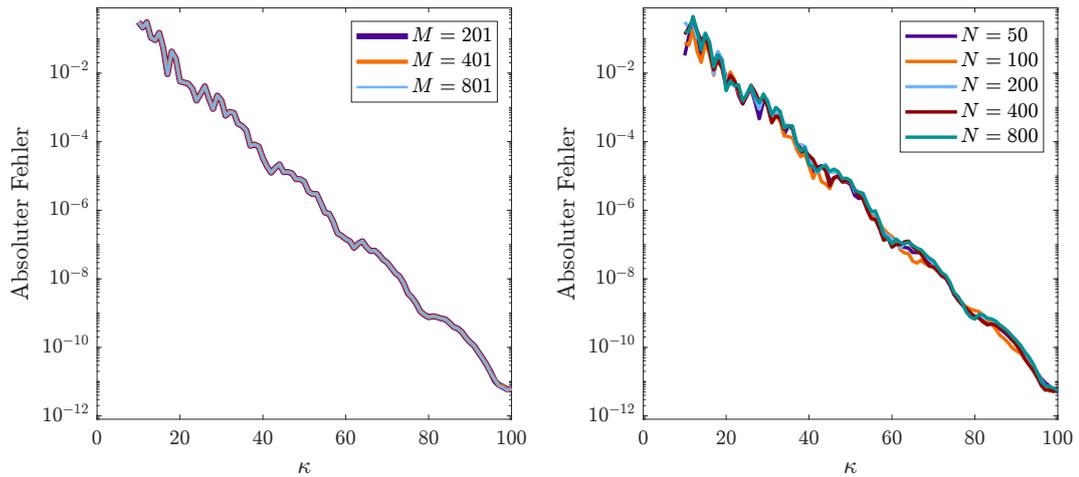


Abbildung 4.2: Absoluter Fehler zum Zeitpunkt  $t = 0.5$  in Abhängigkeit vom Parameter  $\kappa$  für verschiedene Werte des Zeitschritts  $N$  und verschiedene Anzahlen  $M$  an Punkten im Ortsgitter. Im linken Bild ist  $N = 200$  festgelegt, im rechten Bild betrachten wir  $M = 401$  Gitterpunkte.

erhalten wir für die Faltungsgewichte einen ähnlichen Diskretisierungsfehler wie in (1.14), jedoch gilt dieser nicht für die ersten paar Gewichte. Diese müssen stattdessen direkt über die Darstellung (2.4) berechnet werden, wodurch eine Fehlersättigung auftritt, die abhängig von der Anzahl der auf diese Weise berechneten Gewichte ist.



## KAPITEL 5

---

# WELLENGLEICHUNG IM PERIODISCHEN WELLENLEITER MIT LOKALEM DEFECT

In diesem Kapitel werden wir das in Kapitel 3 vorgestellte Verfahren nutzen, um die Wellengleichung in einem periodischen, lokal gestörten Wellenleiter zu untersuchen und die dort hergeleiteten Operatoren in den Zeitbereich zu übertragen. Hierbei erhalten wir diese in Form von Faltungen, die mit Hilfe der Runge-Kutta basierten Faltungsquadratur approximiert werden können. Verwenden wir dasselbe Runge-Kutta-Verfahren für die Diskretisierung der partiellen Differentialgleichung im Gebiet des lokalen Defekts, sind wir in der Lage einen Faltungsalgorithmus anzugeben, der es uns ermöglicht, die Lösung in diesem beschränkten Gebiet sowie davon ausgehend im Rest des Wellenleiters zu berechnen.

Unter der vereinfachenden Annahme, dass der Quellterm  $f$  der Wellengleichung harmonisch zeitabhängig ist, das heißt als Produkt einer nur vom Ort abhängigen Funktion und dem zeitabhängigen Faktor  $\exp(-i\omega t)$  mit Frequenz  $\omega > 0$  dargestellt werden kann, lässt sich die Wellengleichung durch Trennung der Variablen auf die Helmholtz-Gleichung zurückführen. Dies ist auch die Motivation zur Betrachtung der Helmholtz-Gleichung in [40] und den folgenden Veröffentlichungen, die sich daher hauptsächlich auf das zeitharmonische Problem konzentrieren und keine Möglichkeit aufzeigen, die transparenten Randbedingungen sowie die zellenweise Lösungsdarstellung im Zeitbereich zu erhalten.

Für die Betrachtung der Wellengleichung mit einer allgemeineren Quellfunktion in einem periodischen, lokal gestörten Wellenleiter zeigt Coatléven in seiner Dissertation [10] und dem daraus resultierenden Paper [11] eine Möglichkeit auf, die Ideen aus [40] zu nutzen. Hier wird die Wellengleichung zunächst anhand eines  $\theta$ -Schemas in der Zeit diskretisiert. Mit der Lösung des semidiskreten Systems im Halb-Wellenleiter im ersten Zeitschritt lassen sich rekursiv mit Hilfe des Vorgehens aus [40] diskrete DtN-Operatoren herleiten, anhand derer sich anschließend die transparenten Randbedingungen für das zeitdiskrete Problem im gesamten Halb-Wellenleiter als diskrete Faltung berechnen lassen.

Das hier vorgestellte Verfahren ermöglicht im Vergleich zur Idee aus [10] die Angabe einer zeitabhängigen transparenten Randbedingung, eine höhere Konvergenzordnung sowie eine schnelle und teils parallele Berechnung.

Dazu werden wir im ersten Teil des Kapitels das Modellproblem erläutern und wie im vorherigen Kapitel mit Hilfe der Ideen aus Kapitel 3 die transparenten Randbedingungen sowie die zellenweise Lösungsdarstellung herleiten. Der zweite Teil des Kapitels dient zur Angabe des schnellen Faltungsalgorithmus, wobei wir auch auf die Ortsdiskretisierung eingehen werden. Abschließend betrachten wir numerische Experimente, die die Anwendung und das Verhalten des Algorithmus illustrieren.

## 5.1 Bestimmung der transparenten Randbedingungen

### 5.1.1 Modellproblem

Wie in Abschnitt 3.1.1 betrachten wir einen zweidimensionalen periodischen Wellenleiter  $\Omega$ , der im Gebiet  $\Omega_0$  eine lokale Störung der Periodizität aufweist. Die eingeführten Notationen, siehe Abbildung 3.1, bleiben weiterhin bestehen.

Auf dem Wellenleiter suchen wir die Lösung  $u : \Omega \times \mathbb{R}_0^+ \rightarrow \mathbb{C}$  der Wellengleichung

$$\left\{ \begin{array}{ll} -\Delta_{x,y} u(x, y, t) + \rho(x, y) \partial_t^2 u(x, y, t) = f(x, y, t) & \text{in } \Omega \times \mathbb{R}^+, \\ u(x, y, t) = 0 & \text{auf } \partial\Omega \times \mathbb{R}^+, \\ u(x, y, 0) = u_0(x, y) & \text{in } \Omega, \\ \partial_t u(x, y, 0) = v_0(x, y) & \text{in } \Omega. \end{array} \right. \quad (\mathbf{W})$$

Während wir die Annahmen (H1) und (H2) über die Funktion  $\rho$  aus Kapitel 3 weiter beibehalten, müssen die Annahmen über den Quellterm  $f$  dem nun zeitabhängigen Problem angepasst beziehungsweise ergänzt werden, um weiterhin die spezielle Struktur des Wellenleiters ausnutzen zu können. Daher treffen wir anstelle von (H3) und (H4) die Annahmen

$$(W1) \quad f \in C^1(\mathbb{R}_0^+, L^2(\Omega)) \text{ und}$$

$$(W2) \quad \text{supp}(f(\cdot, t)) \subset \Omega_0 \quad \forall t \in \mathbb{R}_0^+$$

und fordern zusätzlich

$$(W3) \quad u_0 \in H_0^1(\Delta, \Omega) \text{ und } v_0 \in H_0^1(\Omega) \text{ sowie}$$

$$(W4) \quad \text{supp}(u_0) \cup \text{supp}(v_0) \subset \Omega_0.$$

### 5.1.2 Existenz und Eindeutigkeit einer Lösung

Unter Verwendung der soeben formulierten Annahmen können wir die Existenz und Eindeutigkeit einer bezüglich der Zeit klassischen Lösung der Wellengleichung **(W)** zeigen. Dazu erinnern wir uns an den in Abschnitt 3.1.2 definierten Operator

$$A : D(A) := H_0^1(\Delta, \Omega) \rightarrow L^2(\Omega), \quad u \mapsto Au := -\rho^{-1}\Delta,$$

mit dem wir das zu **(W)** äquivalente Problem

$$\begin{cases} Au + \partial_t^2 u = \rho^{-1} f & \text{in } \Omega \times \mathbb{R}^+, \\ u = 0 & \text{auf } \partial\Omega \times \mathbb{R}^+, \\ u(\cdot, 0) = u_0 & \text{in } \Omega, \\ \partial_t u(\cdot, 0) = v_0 & \text{in } \Omega \end{cases} \quad (5.1)$$

erhalten. Wie wir bereits in Satz 3.8 und Korollar 3.9 bewiesen haben, gelten folgende Resultate, die wir hier zur Übersichtlichkeit kurz wiedergeben:

- der Operator  $A$  ist unbeschränkt und selbstadjungiert in  $L^2(\Omega)$ ;
- die zu  $-\Delta$  assoziierte Sesquilinearform

$$a_0 : H_0^1(\Omega) \times H_0^1(\Omega) \rightarrow \mathbb{C}, \quad (u, v) \mapsto a_0(u, v) := \int_{\Omega} \nabla u \cdot \overline{\nabla v} \, d(x, y)$$

ist stetig auf  $H_0^1(\Omega) \times H_0^1(\Omega)$ , hermitesch sowie  $H_0^1(\Omega)$ -elliptisch;

- für  $u \in D(A)$  und  $v \in H_0^1(\Omega)$  gilt damit insbesondere  $a_0(u, v) = \langle Au, v \rangle_{0, \rho}$ .

Um den nächsten Satz formulieren zu können, wollen wir das Problem (5.1) in ein System erster Ordnung in der Zeit transformieren, das mit der Definition  $V = (V_1, V_2)^T := (u, \partial_t u)^T$  gegeben ist durch

$$\partial_t V = \begin{pmatrix} \partial_t u \\ \partial_t^2 u \end{pmatrix} = \begin{pmatrix} \partial_t u \\ -Au + \rho^{-1} f \end{pmatrix} = \begin{pmatrix} 0 & \text{Id} \\ -A & 0 \end{pmatrix} V + \begin{pmatrix} 0 \\ \rho^{-1} f \end{pmatrix}.$$

Wir definieren hierbei den Operator  $\mathcal{A} : D(\mathcal{A}) \rightarrow \mathcal{H}$  für  $V \in D(\mathcal{A})$  durch

$$\mathcal{A}V := \begin{pmatrix} 0 & \text{Id} \\ -A & 0 \end{pmatrix} V = \begin{pmatrix} 0 & \text{Id} \\ \rho^{-1}\Delta & 0 \end{pmatrix} V, \quad (5.2a)$$

wobei der Raum  $\mathcal{H} := H_0^1(\Omega) \times L^2(\Omega)$  ausgestattet mit dem Skalarprodukt

$$\langle U, V \rangle_{\mathcal{H}} := a_0(U_1, V_1) + \langle U_2, V_2 \rangle_{0, \rho} \quad (5.2b)$$

wieder ein Hilbertraum ist. Setzen wir

$$D(\mathcal{A}) := D(A) \times H_0^1(\Omega) \subset \mathcal{H}, \quad (5.2c)$$

so ist  $\mathcal{A}$  nach [13, Prop. VI.3.9] abgeschlossen in  $\mathcal{H}$ .

Mit diesen Überlegungen liefert [14, Thm. XVII.3.4] sofort das nächste Resultat:

**Satz 5.1.** *Der über (5.2) definierte Operator  $\mathcal{A}$  ist der infinitesimale Generator einer  $C_0$ -Kontraktionshalbgruppe auf  $\mathcal{H}$ .*

**Bemerkung 5.2.** Für den Beweis im genannten Theorem XVII.3.4 aus [14], dass tatsächlich eine Kontraktionshalbgruppe generiert wird, ist die  $H_0^1(\Omega)$ -Elliptizität, also

$$\operatorname{Re}(a_0(u, u)) \geq C \|u\|_{1,\rho}^2 \quad \forall u \in H_0^1(\Omega) \quad \text{mit} \quad C > 0$$

sowie die darüber definierte Norm auf  $\mathcal{H}$  von zentraler Bedeutung. Gilt hier nur

$$\operatorname{Re}(a_0(u, u)) + \lambda_0 \|u\|_{0,\rho}^2 \geq C \|u\|_{1,\rho}^2 \quad \text{für} \quad \lambda_0 \in \mathbb{R} \setminus \{0\},$$

so lässt sich weiterhin zeigen, dass der Operator eine  $C_0$ -Halbgruppe erzeugt. Diese muss jedoch keine Kontraktionshalbgruppe mehr sein, vergleiche [14, Bsp. XVIIB.3.6].

Mit Satz 5.1 folgen nun die Existenz und die Eindeutigkeit einer bezüglich der Zeit klassischen Lösung:

**Satz 5.3.** *Unter den Annahmen (H1), (W1) und (W3) besitzt die Wellengleichung (W) eine eindeutige Lösung*

$$u \in C^2(\mathbb{R}_0^+, L^2(\Omega)) \cap C^1(\mathbb{R}_0^+, H_0^1(\Omega)) \cap C(\mathbb{R}_0^+, D(\mathcal{A})).$$

*Beweis.* Da der Operator  $\mathcal{A}$  nach Satz 5.1 eine  $C_0$ -Kontraktionshalbgruppe erzeugt, der Vektor der Startwerte  $(u_0, v_0)^T$  nach (W3) in  $D(\mathcal{A})$  liegt und die Funktion  $F : \mathbb{R}_0^+ \rightarrow \mathcal{H}$ ,  $t \mapsto (0, \rho^{-1}f(\cdot, t))^T$  nach (W1) stetig differenzierbar auf  $\mathbb{R}_0^+$  ist, folgt die Existenz einer Lösung

$$V \in C^1(\mathbb{R}_0^+, \mathcal{H}) \cap C(\mathbb{R}_0^+, D(\mathcal{A})) \tag{5.3}$$

des inhomogenen Cauchy-Problems

$$\begin{cases} \partial_t V(t) = \mathcal{A}V(t) + F(t), & t \in \mathbb{R}^+ \\ V(0) = (u_0, v_0)^T \end{cases} \tag{5.4}$$

sofort mit [55, Kor. 4.2.5].

Um die Eindeutigkeit zu sehen, nehmen wir an, dass  $W_1$  und  $W_2$  zwei unterschiedliche Lösungen von (5.4) sind. Dann lässt sich leicht einsehen, dass  $W := W_1 - W_2$  eine (klassische) Lösung des homogenen Cauchy-Problems  $\partial_t W(t) = \mathcal{A}W(t)$ ,  $t \in \mathbb{R}^+$  mit Anfangswert  $W(0) = W_1(0) - W_2(0) = (0, 0)^T$  ist, vergleiche [55, S.100]. Da damit  $W(0) \in D(\mathcal{A})$  erfüllt ist, hat das homogene Cauchy-Problem nach [55, Thm. 1.2.4] die eindeutige Lösung  $W(t) \equiv 0$ , wodurch  $W_1 = W_2$  folgt.

Die angegebenen Regularitäten von  $u$  ergeben sich nun aus (5.3) unter Verwendung von (H1) und  $V = (u, \partial_t u)^T$ .  $\square$

### 5.1.3 Zusammenhang mit Helmholtz-Gleichung

Um, wie schon für die fraktionale Schrödinger-Gleichung gesehen, das Verfahren aus Kapitel 3 anwenden zu können, wollen wir statt eines Problems im Zeitbereich wieder ein Problem im Frequenzbereich betrachten. Unter Verwendung der Laplace-Transformation bezüglich der Zeit erhalten wir aus der Wellengleichung (**W**) eine Helmholtz-Gleichung der Form

$$\begin{aligned} -\Delta_{x,y} U(x, y, \lambda) + \rho(x, y) \lambda^2 U(x, y, \lambda) \\ = F(x, y, \lambda) + \rho(x, y) (\lambda u_0(x, y) + v_0(x, y)) \quad \text{in } \Omega \times \Sigma(\sigma_0) \end{aligned} \quad (5.5)$$

unter Dirichlet-Randbedingungen  $U(x, y, \lambda) = 0$  auf  $\partial\Omega \times \Sigma(\sigma_0)$ , wobei

$$U(\cdot, \lambda) := \mathcal{L}\{u(\cdot, t); \lambda\} \quad \text{und} \quad F(\cdot, \lambda) := \mathcal{L}\{f(\cdot, t); \lambda\}. \quad (5.6)$$

Hierbei ist  $\sigma_0 > 0$  die Konvergenzabszisse, das heißt,  $\sigma_0$  sei so gewählt, dass die vorkommenden Laplace-Integrale für  $\text{Re}(\lambda) > \sigma_0$  absolut konvergieren und damit (5.6) wohldefiniert ist, vergleiche [15, Satz 3.4].

Im Gegensatz zu der betrachteten Helmholtz-Gleichung (**H**) in Kapitel 3 kommt nun die Frequenz  $\lambda^2$  anstelle der Frequenz  $s$  in der Gleichung vor. Wegen  $\lambda \in \Sigma(\sigma_0)$ , erfüllt  $\lambda^2$  die Annahme (H5), das heißt insbesondere gilt  $\lambda^2 = s \in \mathbb{C} \setminus \mathbb{R}_0^+$ .

Des Weiteren lässt sich leicht einsehen, dass für festes  $\lambda \in \Sigma(\sigma_0)$  und unter den Voraussetzungen (H1) sowie (W1)-(W4) für die rechte Seite der Helmholtz-Gleichung gilt

$$\tilde{F} := F + \rho(\lambda u_0 + v_0) \in L^2(\Omega) \quad \text{und} \quad \text{supp}(\tilde{F}) \subset \Omega_0.$$

Damit werden die Voraussetzungen von Proposition 3.11 eingehalten und wir können die Lösung  $U$  des Problems (5.5) im gesamten Wellenleiter zusammensetzen aus der Lösung  $U^0$  des Problems im lokal gestörten Gebiet  $\Omega_0$ , also hier

$$\left\{ \begin{array}{l} -\Delta U^0 + \lambda^2 \rho U^0 = \tilde{F} \quad \text{in } \Omega_0, \\ U^0 = 0 \quad \text{auf } \partial\Omega \cap \partial\Omega_0, \\ U^0 - \Lambda_{\lambda^2}^- \left( \frac{\partial U^0}{\partial x} \Big|_{\Gamma_0^-} \right) = 0 \quad \text{auf } \Gamma_0^-, \\ U^0 - \Lambda_{\lambda^2}^+ \left( -\frac{\partial U^0}{\partial x} \Big|_{\Gamma_0^+} \right) = 0 \quad \text{auf } \Gamma_0^+, \end{array} \right. \quad (5.7)$$

vergleiche  $(\mathbf{H}^0)$ , sowie aus den Lösungen  $U^\pm(\varphi^\pm)$  der Halb-Wellenleiter-Probleme

$$\left| \begin{array}{ll} -\Delta U^\pm(\varphi^\pm) + \lambda^2 \rho U^\pm(\varphi^\pm) = 0 & \text{in } \Omega^\pm, \\ U^\pm(\varphi^\pm) = 0 & \text{auf } \partial\Omega \cap \partial\Omega^\pm, \\ \mp \frac{\partial U^\pm(\varphi^\pm)}{\partial x} = \varphi^\pm & \text{auf } \Gamma_0^\pm, \end{array} \right.$$

für die Wahl  $\varphi^\pm := \mp \frac{\partial U^0}{\partial x} \Big|_{\Gamma_0^\pm}$ . Nach Satz 3.13 ist damit die Lösung in der Zelle  $C_j^\pm$  gegeben als

$$U(\varphi^\pm) \Big|_{C_j^\pm} = \left( \mathcal{S}_{\lambda^2}^\pm \circ (\mathcal{P}_{\lambda^2}^\pm)^j \right) (\varphi^\pm), \quad j \in \mathbb{N}_0. \quad (5.8)$$

Um nun die transparenten Randbedingungen für die Wellengleichung  $(\mathbf{W})$  eingeschränkt auf das Gebiet  $\Omega_0$  zu bekommen, folgen wir demselben Vorgehen wie auch schon für die fraktionale Schrödinger-Gleichung, das heißt, wir wenden die inverse Laplace-Transformation auf das lokale Helmholtz-Problem (5.7) an. Dies liefert uns die lokale Wellengleichung

$$\left| \begin{array}{ll} -\Delta u + \rho \partial_t^2 u = f & \text{in } \Omega_0 \times \mathbb{R}^+, \\ u = 0 & \text{auf } \partial\Omega \cap \partial\Omega_0 \times \mathbb{R}^+, \\ u - \lambda^- \left( \frac{\partial u}{\partial x} \Big|_{\Gamma_0^-} \right) = 0 & \text{auf } \Gamma_0^- \times \mathbb{R}^+, \\ u - \lambda^+ \left( -\frac{\partial u}{\partial x} \Big|_{\Gamma_0^+} \right) = 0 & \text{auf } \Gamma_0^+ \times \mathbb{R}^+, \\ u(\cdot, 0) = u_0 & \text{in } \Omega_0, \\ \partial_t u(\cdot, 0) = v_0 & \text{in } \Omega_0. \end{array} \right. \quad (\mathbf{W}^0)$$

Verstehen wir den NtD-Operator (3.3) für die Helmholtz-Gleichung nun, wie bereits in Bemerkung 4.7 beschrieben, als Abbildung  $\lambda \rightarrow \Lambda^\pm(\lambda^2)$ , so können wir den NtD-Operator für die Wellengleichung

$$\lambda^\pm : \left( \mathbb{R}^+ \rightarrow H^{-1/2}(\Gamma_0^\pm) \right) \rightarrow H_0^{1/2}(\Gamma_0^\pm)$$

wieder mit Hilfe des Faltungssatzes über die inverse Laplace-Transformation definieren:

Es seien  $\Phi^\pm : \Sigma(\sigma_0) \rightarrow H^{-1/2}(\Gamma_0^\pm)$  und  $\varphi^\pm : \mathbb{R}^+ \rightarrow H^{-1/2}(\Gamma_0^\pm)$ , sodass  $\Phi^\pm(\lambda) = \mathcal{L}\{\varphi^\pm(t); \lambda\}$  erfüllt ist. Damit setzen wir

$$\begin{aligned} \lambda^\pm(\varphi^\pm(t)) &:= \mathcal{L}^{-1}\{\Lambda^\pm(\lambda^2) \Phi^\pm(\lambda); t\} \\ &= \left( \mathcal{L}^{-1}\{\Lambda^\pm(\lambda^2); \cdot\} * \mathcal{L}^{-1}\{\Phi^\pm(\lambda); \cdot\} \right)(t) \\ &= \int_0^t \mathcal{L}^{-1}\{\Lambda^\pm(\lambda^2); t - \tau\} \varphi^\pm(\tau) d\tau. \end{aligned} \quad (5.9)$$

Dasselbe Vorgehen ermöglicht uns auch wieder eine Darstellung der Lösung in der Zelle  $C_j^\pm$ ,  $j \in \mathbb{N}_0$ : Wenden wir die inverse Laplace-Transformation auf die zellenweise Lösung im

Frequenzbereich (5.8) an, erhalten wir

$$\begin{aligned} \mathcal{L}^{-1}\left\{U(\Phi^\pm(\lambda))\Big|_{C_j^\pm}; t\right\} &= \mathcal{L}^{-1}\left\{\left(\mathcal{S}^\pm \circ (\mathcal{P}^\pm)^j\right)(\lambda^2) \Phi^\pm(\lambda); t\right\} \\ &= \left(\mathcal{L}^{-1}\left\{\left(\mathcal{S}^\pm \circ (\mathcal{P}^\pm)^j\right)(\lambda^2); \cdot\right\} * \mathcal{L}^{-1}\{\Phi^\pm(\lambda); \cdot\}\right)(t) \\ &= \left(\mathcal{L}^{-1}\left\{\left(\mathcal{S}^\pm \circ (\mathcal{P}^\pm)^j\right)(\lambda^2); \cdot\right\} * \varphi^\pm(\cdot)\right)(t). \end{aligned}$$

Haben wir also die Lösung der lokalen Wellengleichung ( $\mathbf{W}^0$ ) in  $\Omega_0$  berechnet, so können wir sie durch

$$u(t)\Big|_{C_j^\pm} = \mathcal{L}^{-1}\left\{\left(\mathcal{S}^\pm \circ (\mathcal{P}^\pm)^j\right)(\lambda^2); \cdot\right\} * \left(\mp \frac{\partial u(\cdot)}{\partial x}\Big|_{\Gamma_0^\pm}\right)(t) \quad (5.10)$$

zellenweise fortsetzen.

## 5.2 Schneller Faltungsalgorithmus für die Wellengleichung

### 5.2.1 Anwendung von Runge-Kutta Verfahren und Faltungsquadratur

Da wir nun die NtD-Operatoren sowie die Lösung der Wellengleichung ( $\mathbf{W}$ ) in den Zellen als Faltungsintegrale dargestellt haben, können wir auf deren Diskretisierung in der Zeit unter Verwendung von Faltungsquadratur-Verfahren eingehen. Hierbei erinnern wir an die Beobachtungen aus den Abschnitten 3.2.2.1 und 3.2.2.2 für den eindimensionalen Fall: Sowohl im betrachteten homogenen als auch im inhomogenen Beispiel konnten wir feststellen, dass die Abbildungen  $\lambda \mapsto \Lambda^+(\lambda^2)$  und  $\lambda \mapsto \mathcal{S}^+(\lambda^2)$  in der Halbebene  $\Sigma(\sigma_0)$ ,  $\sigma_0 > 0$ , analytisch sind und dort die Abschätzungen aus Annahme 2.2, insbesondere mit  $\mu_2 = 1$ , erfüllen, vergleiche dazu (3.16) und (3.21). Da nach Konstruktion  $|\mathcal{P}^+| < 1$  gilt, zeigt auch die Abbildung  $\lambda \mapsto \mathcal{S}^+ \circ (\mathcal{P}^+)^j(\lambda^2)$  für  $j \in \mathbb{N}_0$  das gewünschte Abfallverhalten in der rechten Halbebene  $\Sigma(\sigma_0)$ .

Für beliebige Funktionen  $\rho^\pm$ , die nicht als periodische Treppenfunktionen gegeben sind, sowie allgemein für den zweidimensionalen Fall können wir weder die Operatoren analytisch berechnen noch deren Abfallverhalten angeben. Wir gehen jedoch auf Grund der Beobachtungen im Eindimensionalen davon aus, dass die Abbildungen  $\lambda \mapsto \Lambda^\pm(\lambda^2)$  und  $\lambda \mapsto \mathcal{S}^\pm \circ (\mathcal{P}^\pm)^j(\lambda^2)$  weiterhin die Abschätzung aus Annahme 2.2 mit  $\mu_2 = 1$  erfüllen. Das numerische Experiment im Abschnitt 5.3.2 untermauert diese Annahme.

Damit können sowohl der zeitabhängige NtD-Operator (5.9) der Wellengleichung als auch das Faltungsintegral zur zellenweisen Berechnung der Lösung (5.10) mit Hilfe der Runge-Kutta Faltungsquadratur aus Satz 2.4 approximiert werden. Im Folgenden betrachten wir dazu, wie im ersten Kapitel, ein  $A$ -stabiles Runge-Kutta Verfahren der Form (2.2) mit Ordnung  $p$  und Stufenordnung  $q$ , das die Bedingungen (RK1)-(RK3) erfüllt.

### 5.2.1.1 Semidiskrete Approximation in den Zellen

Obwohl im Algorithmus als erstes die Lösung im beschränkten Gebiet  $\Omega_0$  berechnet wird bevor sie daraus zellenweise fortgesetzt werden kann, gehen wir zunächst kurz auf diese zellenweise Berechnung ein, da sie lediglich aus der Approximation eines Faltungsintegrals besteht.

Dazu treffen wir die theoretische Annahme, dass wir bereits die Lösung  $u : \Omega_0 \times \mathbb{R}_0^+ \rightarrow \mathbb{C}$  der lokalen Wellengleichung ( $\mathbf{W}^0$ ) kennen und damit auch die Funktion

$$g^\pm : \mathbb{R}^+ \rightarrow H^{-1/2}(\Gamma_0^\pm), \quad t \mapsto g^\pm(t) := \mp \frac{\partial u(\cdot, t)}{\partial x} \Big|_{\Gamma_0^\pm}$$

für jedes  $t \in \mathbb{R}^+$  angeben können. Die Anwendung von Satz 2.4 auf das Faltungsintegral (5.10) liefert uns unter dieser Annahme zum Zeitpunkt  $t_{N+1} = (N+1)h$  mit Schrittweite  $h$  und  $N$  Zeitschritten die Approximation

$$u(\cdot, t_{N+1}) \Big|_{C_j^\pm} \approx h \sum_{n=0}^N \widetilde{w}_n^{\pm, j} G_{N-n}^\pm = (b^T \mathcal{O} \otimes \text{Id}) h \sum_{n=0}^N \widetilde{W}_n^{\pm, j} G_{N-n}^\pm, \quad (5.11)$$

wobei hier

$$G_n^\pm(\cdot) = (g^\pm(t_n + c_k h))_{k=1}^s = \left( \mp \frac{\partial}{\partial x} u(\cdot, t_n + c_k h) \Big|_{\Gamma_0^\pm} \right)_{k=1}^s$$

gesetzt wird und die Faltungsgewichte  $\widetilde{W}_n^{\pm, j}$  über die erzeugende Funktion

$$h \sum_{n=0}^{\infty} \widetilde{W}_n^{\pm, j} \zeta_n = \left( \mathcal{S}^\pm \circ (\mathcal{P}^\pm)^j \right) \left( \left( \frac{\Delta(\zeta)}{h} \right)^2 \right) \quad (5.12)$$

gebildet werden.

Nehmen wir nun weiter an, dass die Funktion  $g^\pm$  glatt genug ist und  $(g^\pm)^{(l)}(0) = 0$  für  $l = 0, \dots, r-1$  mit  $r > \max(p - \mu_1 + 1, p, q + 1)$  erfüllt, so können wir nach Satz 2.8 in der Approximation zum Zeitpunkt  $t_{N+1}$  einen Fehler in der Größenordnung  $\mathcal{O}(h^{\min(p, q+1+\mu_2)})$  erwarten.

In der praktischen Anwendung können wir die oben getroffene theoretische Annahme natürlich nicht erfüllen. Statt der Funktion  $g^\pm$  kennen wir lediglich die Approximationen  $U_{nk}(\cdot) \Big|_{\Gamma_0^\pm} \approx u(\cdot, t_n + c_k h) \Big|_{\Gamma_0^\pm}$  für  $n = 0, \dots, N$  und  $k = 1, \dots, s$ , wodurch wir anstelle von  $G_n^\pm(\cdot)$  in (5.11) die zellenweise Lösung mit Hilfe von

$$G_n^\pm(\cdot) = \left( \mp \frac{\partial}{\partial x} U_{nk}(\cdot) \Big|_{\Gamma_0^\pm} \right)_{k=1}^s$$

erhalten. Im numerischen Experiment in Abschnitt 5.3.2 lässt sich dennoch die erhoffte Konvergenz beobachten.

### 5.2.1.2 Semidiskrete Approximation in $\Omega_0$

Im lokalen Problem  $(\mathbf{W}^0)$  geht es nicht nur um die Approximation des Faltungsintegrals, mit dem wir den NtD-Operator (5.9) für die transparenten Randbedingung dargestellt haben, sondern auch um die passende Zeitdiskretisierung der Wellengleichung  $-\Delta u + \rho \partial_t^2 u = f$ . Wir gehen daher zunächst auf Letztere ein und betrachten dazu weiterhin ein  $A$ -stabiles implizites Runge-Kutta Verfahren, das die Annahmen (RK1)-(RK3) erfüllt.

Zur Diskretisierung der Wellengleichung  $(\mathbf{W})$  in  $\Omega$ , deren eindeutige Lösung in Satz 5.3 charakterisiert ist, transformieren wir diese in ein System erster Ordnung in der Zeit, was wir bereits in Abschnitt 5.1.2 gesehen haben und hier kurz mit für diesen Abschnitt passenderer Notation wiedergeben. Setzen wir  $v := \partial_t u$  und  $y := (u, v)^T$ , so ist  $(\mathbf{W})$  äquivalent zum System

$$\partial_t y = \begin{pmatrix} \partial_t u \\ \partial_t v \end{pmatrix} = \begin{pmatrix} v \\ -Au + \rho^{-1} f \end{pmatrix} = \mathcal{A}y + \begin{pmatrix} 0 \\ \rho^{-1} f \end{pmatrix},$$

wobei der Operator  $\mathcal{A}$  durch (5.2) definiert ist. Durch die Anwendung des gegebenen Runge-Kutta Verfahrens auf dieses System erster Ordnung erhalten wir die folgende Darstellung der inneren Stufe  $Y_{ni}(\cdot)$ ,  $i = 1, \dots, s$  als

$$Y_{ni} = \begin{pmatrix} U_{ni} \\ V_{ni} \end{pmatrix} = y_n + h \sum_{j=1}^s a_{ij} \begin{pmatrix} V_{nj} \\ -AU_{nj} + \rho^{-1} f(\cdot, t_n + c_j h) \end{pmatrix}, \quad (5.13)$$

wobei  $y_0 = (u_0, v_0)^T$  und wir die Approximation  $y_{N+1}$  an  $y(\cdot, t_{N+1})$  wegen (RK3) wieder direkt aus der letzten Stufe  $Y_{Ns}$  bekommen.

Da der Operator  $\mathcal{A}$  nach Satz 5.1 eine Kontraktionshalbgruppe generiert, können wir die Theoreme 1 und 3 aus [7] anwenden, die uns die Konvergenz von  $p$ -ter Ordnung des Verfahrens (5.13) liefern. Für den Beweis werden zusätzliche Anforderungen an die Regularität der Lösung  $y$  beziehungsweise an die Regularität des Quellterms gestellt, auf die wir nicht genauer eingehen.

Es stellt sich nun die Frage, wie die Diskretisierung der transparenten Randbedingungen gewählt werden sollte, um die Konvergenz auch bezüglich des Problems  $(\mathbf{W}^0)$  zu erhalten. Im Gegensatz zur Gleichung (5.10) aus dem vorherigen Abschnitt, bei deren Approximation wir auf eine bis dahin bereits bekannte rechte Seite zurückgreifen können, sind die transparenten Randbedingungen als Integralgleichungen

$$u(\cdot, t)|_{\Gamma_0^\pm} = \mp \int_0^t \mathcal{L}^{-1}\{\Lambda^\pm(\lambda^2); t - \tau\} \frac{\partial}{\partial x} u(\cdot, \tau)|_{\Gamma_0^\pm} d\tau$$

gegeben, vergleiche  $(\mathbf{W}^0)$  und (5.9). Wie wir bereits in Abschnitt 2.3.2 für eine Volterra Integralgleichung gesehen haben, führt die Anwendung der Runge-Kutta Faltungsquadratur hierbei auf eine implizite Gleichung zur Berechnung der inneren Stufen.

Wir bezeichnen daher wie im ersten Kapitel mit  $U_N$  den Spaltenvektor der inneren Stufen, also  $U_N = (U_{Ni})_{i=1}^s$ . Die Approximation der transparenten Randbedingungen erhalten wir somit als

$$U_N|_{\Gamma_0^\pm} = h \sum_{n=0}^N W_n^\pm \left( \mp \frac{\partial}{\partial x} U_{N-n}|_{\Gamma_0^\pm} \right), \quad (5.14a)$$

wobei hier die Faltungsgewichte  $W_n^\pm$  über die erzeugende Funktion

$$h \sum_{n=0}^{\infty} W_n^\pm \zeta^n = \Lambda^\pm \left( \left( \frac{\Delta(\zeta)}{h} \right)^2 \right) \quad (5.14b)$$

bestimmt werden.

Verwenden wir in der Runge-Kutta Faltungsquadratur dasselbe Runge-Kutta Verfahren, das auch der Diskretisierung (5.13) zugrunde liegt, so sind die Werte  $U_{Ni}|_{\Gamma_0^\pm}$ , die wir durch die semidiskreten transparenten Randbedingungen (5.14) erhalten, nach Konstruktion genau dieselben wie die, die durch die Einschränkung der  $U_{Ni}$  aus (5.13) auf  $\Gamma_0^\pm$  entstehen. Damit erhalten wir wie bereits in Korollar 4.9 gesehen aus der Konvergenz der Diskretisierung von  $(\mathbf{W})$  direkt die Konvergenz  $p$ -ter Ordnung des lokalen Problems  $(\mathbf{W}^0)$ , die wir auch anhand des numerischen Experiments in Abschnitt 5.3.2 nachvollziehen können.

## 5.2.2 Diskretisierung im Ort

Wir betrachten hier die beiden Methoden zur Diskretisierung des Ortes, die den numerischen Experimenten zugrunde liegen. Zum einen gehen wir kurz auf die zentralen Finiten Differenzen ein, unter deren Verwendung wir die benötigten ortsdiskreten Operatoren in den eindimensionalen Beispielen in Abschnitt 3.2.3 exakt ausrechnen konnten. Zum anderen beschäftigen wir uns mit der Diskretisierung mittels der Methode der Finiten Elemente, die wir im zweidimensionalen Experiment anwenden.

Hierbei gehen wir nicht genauer auf die Ortsdiskretisierung der lokalen zeitharmonischen Probleme (3.8) ein, anhand derer wir die Operatoren  $\mathcal{P}^\pm$ ,  $\mathcal{S}^\pm$  und  $\Lambda^\pm$  bestimmen, da diese für den schnellen Faltungsalgorithmus nicht von großem Interesse sind und sich leicht realisieren lassen. Des Weiteren lässt sich die zellenweise Berechnung der Lösung auf eine einfache Art vornehmen, sobald wir die benötigten Daten an den Rändern  $\Gamma_0^\pm$  durch die Lösung des Problems in  $\Omega_0$  bestimmt haben. Im Folgenden werden wir uns daher auf die Ortsdiskretisierung des semidiskreten Systems bestehend aus (5.13) und (5.14) konzentrieren und damit insbesondere nicht genauer auf die Berechnung aus Abschnitt 5.2.1.1 eingehen.

Der Übersichtlichkeit halber betrachten wir in diesem Abschnitt nicht das vollständige System (5.13), sondern lediglich die zweite Zeile des Vektors  $Y_{ni}$  unter der Annahme<sup>1</sup>  $s = 1$ ,

---

<sup>1</sup>Nach (RK3) ist das einstufige Runge-Kutta Verfahren damit durch die Werte  $a_{11} = b_1 = c_1 = 1$  charakterisiert.

das heißt die Gleichung

$$\rho V_{n1} = \rho v_n + h (\Delta U_{n1} + f(t_n + h)) \quad \text{in } \Omega_0. \quad (5.15)$$

Für das impliziten Euler-Verfahren gilt  $V_{n1} = v_{n+1}$  und  $U_{n1} = u_{n+1}$ . Da wir hier die Verallgemeinerung auf beliebige Stufenzahlen, die in Abschnitt 5.2.3 zu finden ist, möglichst nachvollziehbar machen wollen, bleiben wir bei der Notation über Großbuchstaben, die auf die inneren Stufen hinweist. Lediglich auf den zweiten Index  $i = 1$  in der Angabe der Vektoren  $U_{n1}$  und  $V_{n1}$  verzichten wir in den folgenden beiden Abschnitten und schreiben stattdessen  $U_n$  und  $V_n$ . Zum Verständnis der Ortsdiskretisierung ist die Betrachtung der Gleichung (5.15) ausreichend.

### 5.2.2.1 Finite Differenzen

Da uns die Finiten Differenzen nur für die Betrachtung der eindimensionalen Beispiele interessieren, setzen wir in diesem kurzen Abschnitt  $\Omega_0 = (a^-, a^+)$ ,  $a^- < a^+$ . Wir wählen ein äquidistantes Gitter auf  $\bar{\Omega}_0$  mit  $M$  Gitterpunkten und Gitterweite  $\Delta x = \frac{a^+ - a^-}{M-1}$ , wodurch die Gitterpunkte gegeben sind durch  $x_m = (m-1)\Delta x + a^-$  für  $m = 1, \dots, M$ .

Für die Näherung  $(U_n)_m \approx U_n(x_m)$  erhalten wir aus (5.15) für  $m = 2, \dots, M-1$  die ortsdiskretisierte Gleichung

$$\begin{aligned} \rho(x_m) (V_n)_m &= \rho(x_m) v_n(x_m) \\ &+ h \left( \frac{(U_n)_{m+1} - 2(U_n)_m + (U_n)_{m-1}}{\Delta x^2} + f(x_m, t_n + h) \right). \end{aligned} \quad (5.16)$$

Unter Verwendung des Differenzenquotienten zur Approximation der Normalenableitung, vergleiche dazu die Überlegung in Abschnitt 3.2.3, setzen wir

$$\begin{aligned} (G_n)_1 &:= \frac{(U_n)_2 - (U_n)_1}{\Delta x} \approx \frac{\partial}{\partial x} U_n(x_1) \quad \text{und} \\ (G_n)_M &:= -\frac{(U_n)_M - (U_n)_{M-1}}{\Delta x} \approx -\frac{\partial}{\partial x} U_n(x_M). \end{aligned}$$

Definieren wir die Vektoren

$$\begin{aligned} \mathbf{U}_n &:= ((U_n)_1, \dots, (U_n)_M)^T, \\ \mathbf{V}_n &:= ((V_n)_1, \dots, (V_n)_M)^T, \\ \mathbf{G}_n &:= ((G_n)_1, 0, \dots, 0, (G_n)_M)^T, \\ \mathbf{v}_n &:= (v_n(x_1), \dots, v_n(x_M))^T \quad \text{und} \\ \mathbf{f}_n &:= (0, f(x_2, t_n + h), \dots, f(x_{M-1}, t_n + h), 0)^T \in \mathbb{C}^M, \end{aligned}$$



Durch die Beschränkung auf den endlich-dimensionalen Raum  $\Phi_M$  erhalten wir die Galerkin-Approximation als

$$\begin{aligned} \langle V_{n(\Phi_M)}, w_M \rangle_{0,\rho,\Omega_0} &= \langle v_{n(\Phi_M)}, w_M \rangle_{0,\rho,\Omega_0} + h \left( -a(U_{n(\Phi_M)}, w_M) - g^+(U_{n(\Phi_M)}, w_M) \right. \\ &\quad \left. - g^-(U_{n(\Phi_M)}, w_M) + \langle f(t_n + h), w_M \rangle_{0,\Omega_0} \right) \quad \forall w_M \in V_M. \end{aligned}$$

Mit der Massematrix  $\mathbf{M}_\rho = \left( \langle \varphi_i, \varphi_j \rangle_{0,\rho,\Omega_0} \right)_{i,j=1}^M$ , der Steifigkeitsmatrix  $\mathbf{A} = \left( a(\varphi_i, \varphi_j) \right)_{i,j=1}^M$ , dem Lastvektor  $\mathbf{f}_n = \left( \langle f(t_n + h), \varphi_i \rangle_{0,\Omega_0} \right)_{i=1}^M$  und den Vektoren  $\mathbf{G}_n^\pm := \left( g^\pm(U_{n(\Phi_M)}, \varphi_i) \right)_{i=1}^M$  ergibt sich daraus das äquivalente  $M$ -dimensionale System

$$\mathbf{M}_\rho \mathbf{V}_n = \mathbf{M}_\rho \mathbf{v}_n + h(-\mathbf{A} \mathbf{U}_n - \mathbf{G}_n^+ - \mathbf{G}_n^- + \mathbf{f}_n).$$

Wir nehmen nun an, dass es  $m^\pm \in \{1, \dots, M\}$  gibt, so dass für die Ansatzfunktionen gilt:

$$\begin{aligned} \varphi_i|_{\Gamma_0^-} \neq 0 \quad \text{für } i = 1, \dots, m^- \quad \text{und} \quad \varphi_i|_{\Gamma_0^-} = 0 \quad \text{für } i = m^- + 1, \dots, M \quad \text{bzw.} \\ \varphi_i|_{\Gamma_0^+} = 0 \quad \text{für } i = 1, \dots, m^+ - 1 \quad \text{und} \quad \varphi_i|_{\Gamma_0^+} \neq 0 \quad \text{für } i = m^+, \dots, M \end{aligned}$$

Damit lassen sich die beiden Rand-Massematrizen

$$\mathbf{M}^- = \left( \langle \varphi_i, \varphi_j \rangle_{0,\Gamma_0^-} \right)_{i,j=1}^{m^-} \quad \text{und} \quad \mathbf{M}^+ = \left( \langle \varphi_i, \varphi_j \rangle_{0,\Gamma_0^+} \right)_{i,j=m^+}^M$$

definieren und die beiden Unterräume

$$\Phi_M^- := \text{span} \{ \varphi_1, \dots, \varphi_{m^-} \} \quad \text{und} \quad \Phi_M^+ := \text{span} \{ \varphi_{m^+}, \dots, \varphi_M \}$$

von  $\Phi_M$  angeben. Außerdem können wir die Einschränkung von  $U_{n(\Phi_M)}$  auf  $\Gamma_0^\pm$  auf die Darstellungen

$$\begin{aligned} U_{n(\Phi_M)}(x, y)|_{\Gamma_0^-} &= \sum_{m=1}^{m^-} (U_n)_m \varphi_m(x, y) \quad \text{und} \\ U_{n(\Phi_M)}(x, y)|_{\Gamma_0^+} &= \sum_{m=m^+}^M (U_n)_m \varphi_m(x, y) \end{aligned}$$

reduzieren, wobei wir hier die Koeffizientenvektoren mit  $\mathbf{Y}_n^- = ((U_n)_1, \dots, (U_n)_{m^-})^T$  und  $\mathbf{Y}_n^+ = ((U_n)_{m^+}, \dots, (U_n)_M)^T$  bezeichnen.

Unter Verwendung der Basiselemente der Räume  $\Phi_M^\pm$  können wir genauer auf die Vektoren  $\mathbf{G}_n^\pm$  eingehen. Wegen

$$\mathbf{G}_n^- = \left( g^-(U_{n(\Phi_M)}, \varphi_i) \right)_{i=1}^M = \left( g^-(\varphi_m, \varphi_i) \right)_{m,i=1}^M \cdot \mathbf{U}_n$$

und den Überlegungen über die Ansatzfunktionen am Rand  $\Gamma_0^-$  lässt sich leicht einsehen,

dass sich  $\mathbf{G}_n^-$  schreiben lässt als

$$\mathbf{G}_n^- = (\Theta_n^-, 0, \dots, 0)^T \quad \text{mit} \quad \Theta_n^- := \left( g^-(\varphi_m, \varphi_i) \right)_{m,i=1}^{m^-} \cdot \Upsilon_n^-.$$

Analog erhalten wir

$$\mathbf{G}_n^+ = (0, \dots, 0, \Theta_n^+)^T \quad \text{mit} \quad \Theta_n^+ := \left( g^+(\varphi_m, \varphi_i) \right)_{m,i=m^+}^M \cdot \Upsilon_n^+.$$

Zuletzt betrachten wir die schwache Formulierung der semidiskreten transparenten Randbedingungen (5.14). Hier lässt sich unter Ausnutzung der soeben definierten Vektoren einfach nachvollziehen, dass das  $m^-$ - beziehungsweise  $M - m^+ - 1$ -dimensionale Gleichungssystem, das sich durch die Finite-Elemente Diskretisierung ergibt, gegeben ist durch

$$\mathbf{M}^\pm \Upsilon_n^\pm = h \sum_{j=0}^n W_j^\pm \Theta_{n-j}^\pm.$$

### 5.2.3 Vollständig diskretisiertes System

Nachdem wir uns die Ortsdiskretisierung der vereinfachten Gleichung (5.15) mittels Finiten Elemente und, im eindimensionalen Fall, mittels Finiten Differenzen angesehen haben, können wir uns nun der vollständigen Diskretisierung der Gleichungen (5.13) und (5.14) widmen.

Anhand der Ergebnisse aus dem vorherigen Abschnitt lässt sich schnell einsehen, dass uns die vollständige Diskretisierung von (5.13) zwei  $M$ -dimensionale Gleichungen mit den Unbekannten  $\mathbf{U}_{ni}$ ,  $\mathbf{V}_{ni} \in \mathbb{C}^M$  für  $n = 0, \dots, N$  und  $i = 1, \dots, s$  liefert, die in der Form

$$\begin{cases} \mathbf{U}_{ni} = \mathbf{u}_n + h \sum_{j=1}^s a_{ij} \mathbf{V}_{nj}, \\ \mathbf{M}_\rho \mathbf{V}_{ni} = \mathbf{M}_\rho \mathbf{v}_n + h \sum_{j=1}^s a_{ij} (-\mathbf{A} \mathbf{U}_{nj} - \mathbf{G}_{nj} + \mathbf{f}_{ni}) \end{cases}$$

geschrieben werden können. Bezeichnen wir den Vektor der inneren Stufen mit  $\mathbf{U}_n = (\mathbf{U}_{ni})_{i=1}^s$ , genauso die Vektoren  $\mathbf{V}_n$  und  $\mathbf{G}_n$  sowie  $\mathbf{f}_n = (\mathbf{f}(t_n + c_i h))_{i=1}^s$ , so erhalten wir daraus das  $2sM$ -dimensionale Gleichungssystem

$$\begin{cases} \mathbf{U}_n = \mathbb{1}_s \otimes \mathbf{u}_n + h (\mathcal{O} \otimes \text{Id}_M) \mathbf{V}_n, \\ (\text{Id}_s \otimes \mathbf{M}_\rho) \mathbf{V}_n = \mathbb{1}_s \otimes \mathbf{M}_\rho \mathbf{v}_n + h (-(\mathcal{O} \otimes \mathbf{A}) \mathbf{U}_n + (\mathcal{O} \otimes \text{Id}_M) (-\mathbf{G}_n + \mathbf{f}_n)). \end{cases} \quad (5.17)$$

Ohne Einschränkung nehmen wir weiter an, dass es  $d^\pm$  gibt<sup>2</sup>, so dass der Vektor  $\mathbf{G}_{ni} \in \mathbb{C}^M$  gegeben ist als

$$\mathbf{G}_{ni} = (\Theta_{ni}^-, 0, \dots, 0, \Theta_{ni}^+)^T \quad \text{mit} \quad \Theta_{ni}^\pm \in \mathbb{C}^{d^\pm}.$$

<sup>2</sup>Für die Finiten Differenzen in Abschnitt 5.2.2.1 gilt  $d^\pm = 1$ . Bei der Betrachtung der Finiten Elemente in Abschnitt 5.2.2.2 setzen wir  $d^- := m^-$  und  $d^+ := M - m^+ + 1$ .

Bezeichnen wir nun noch die ersten  $d^-$  Einträge von  $\mathbf{U}_{ni}$  mit  $\mathbf{r}_{ni}^-$  sowie die letzten  $d^+$  Einträge von  $\mathbf{U}_{ni}$  mit  $\mathbf{r}_{ni}^+$ , so ist die vollständige Diskretisierung der transparenten Randbedingungen (5.14) geben als die diskreten Faltungen

$$(\text{Id}_s \otimes \mathbf{M}^\pm) \mathbf{r}_n^\pm = h \sum_{k=0}^n W_k^\pm \Theta_{n-k}^\pm$$

mit  $W_n^\pm \in \mathbb{C}^{(sd^\pm) \times (sd^\pm)}$ ,  $\Theta_n^\pm = (\Theta_{ni}^\pm)_{i=1}^s \in \mathbb{C}^{sd^\pm}$  und  $\mathbf{r}_n^\pm = (\mathbf{r}_{ni}^\pm)_{i=1}^s \in \mathbb{C}^{sd^\pm}$ . Setzen wir hierbei die umgeformte Gleichung

$$\Theta_n^\pm = (hW_0^\pm)^{-1} \left( -(\text{Id}_s \otimes \mathbf{M}^\pm) \mathbf{r}_n^\pm + h \sum_{k=1}^n W_k^\pm \Theta_{n-k}^\pm \right) \quad (5.18)$$

in Gleichung (5.17) ein, erhalten wir so ein System, mit dem wir die Vektoren  $\mathbf{U}_n$  und  $\mathbf{V}_n$  bestimmen können. Haben wir diese berechnet, bekommen wir wiederum aus (5.18) die Werte  $\Theta_n^\pm$ , die für die Berechnung des nächsten Zeitschritts notwendig sind.

An dieser Stelle wollen wir nicht zu weit in die Details der Implementierung gehen, jedoch deutlich machen, wie das Schema von Hairer, Lubich und Schlöcher, das in Abschnitt 2.3.2 beschrieben ist, verwendet wird um einen schnellen Algorithmus zu erhalten:

Die direkte Berechnung der  $\mathbf{U}_n$  mit  $n = kr, \dots, (k+1)r - 1$ ,  $k \in \mathbb{N}_0$ ,  $r \in \mathbb{N} \setminus \{1\}$ , die wir in Abschnitt 2.3.2 als die Funktion `Dreieck` bezeichnet haben, wird hier, wie gerade beschrieben, über das System (5.17), in das (5.18) eingesetzt wurde, durchgeführt. Daher wird in `Dreieck` ein System der Größe  $2sM$  gelöst. Ebenfalls schließt `Dreieck` das Lösen der Gleichung (5.18) mit ein. Damit liefert uns die direkte Berechnung in `Dreieck` neben  $\mathbf{U}_n$  und  $\mathbf{V}_n$  auch noch die Vektoren  $\Theta_n^\pm$ .

Die Funktion `Quadrat`, in der die FFT ihre Anwendung findet, bezieht sich auf die Summe in der rechten Seite von (5.18), das heißt auf  $sd^\pm$ -dimensionale Vektoren und Matrizen. Die jeweils dazu notwendigen Vektoren  $\Theta_n^\pm$  sind aus den vorherigen `Dreieck`-Funktionen bekannt.

## 5.2.4 Zusammenfassung des Algorithmus

Bevor wir zu den numerischen Experimenten kommen, geben wir abschließend eine Übersicht über die wichtigsten Schritte des vollständig diskretisierten Algorithmus zur Approximation an die Lösung  $u$  der Wellengleichung (**W**) an.

Sei dazu  $t_n = nh$  mit Zeitschrittweite  $h$  und  $n \leq N$ . Die Ortsdiskretisierung von  $u(\cdot, t_n)|_{\Omega_0} \in H_{0,\Omega}^1(\Omega_0)$  sei realisiert durch einen Vektor der Länge  $M$ , wobei sich jeweils  $d^\pm$  Freiheitsgrade auf die Ränder  $\Gamma_0^\pm$  beziehen, vergleiche dazu die Überlegung aus dem vorherigen Abschnitt. Des Weiteren sei die Ortsdiskretisierung von  $u(\cdot, t_n)|_{C_0^\pm} \in H_{0,\Omega}^1(C_0^\pm)$  durch einen Vektor der Länge  $M^\pm$  bestimmt, wobei wichtig ist, dass wir dieselben  $d^\pm$  Freiheitsgrade bezüglich der Ränder  $\Gamma_0^\pm$  haben.

Wollen wir nun die Approximation von  $u|_{\tilde{\Omega}}$  bezüglich Zeit und Ort für

$$\tilde{\Omega} = \Omega_L^- \cup \Omega_0 \cup \Omega_L^+$$

mit  $\Omega_L^\pm := C_0^\pm \cup \dots \cup C_L^\pm$ ,  $L \in \mathbb{N}_0$ , bestimmen, gehen wir dazu wie folgt vor:

1. *Schritt:* Führe Algorithmus 5.1 aus, das heißt, berechne die  $(sd^\pm \times sd^\pm)$ -Matrizen  $\mathcal{P}^\pm(Z_j^2)$  und  $\Lambda^\pm(Z_j^2)$  sowie die  $(sM^\pm \times sd^\pm)$ -Matrix  $\mathcal{S}^\pm(Z_j^2)$  für jede  $(s \times s)$ -Matrix  $Z_j = \Delta(\zeta_j)/h$  mit  $j = 0, \dots, J-1$ ,  $J > N$  und  $\zeta_j$  aus (2.4). Da die dazu betrachteten Helmholtz-Probleme für jedes  $j$  unabhängig voneinander sind, lassen sich diese Berechnungen, das heißt die Schleife über  $j$ , parallel durchführen. Ebenso können der linke und rechte Halb-Wellenleiter separat betrachtet und parallel berechnet werden.
2. *Schritt:* Berechne die Faltungsgewichte  $W_n^\pm \in \mathbb{C}^{sd^\pm \times sd^\pm}$ ,  $n = 0, \dots, N-1$ , die durch (5.14b) gegeben sind, mit Hilfe der FFT wie in Abschnitt 2.3.1 beschrieben.
3. *Schritt:* Löse das vollständig diskretisierte System (5.17) unter Verwendung von (5.18) und dem Verfahren von Hairer, Lubich und Schlichte wie in Abschnitt 5.2.3 beschrieben. Wir erhalten so die Näherungen  $\mathbf{U}_{nk} \in \mathbb{C}^M \approx u(x, y, t_n + c_k h)|_{\Omega_0}$  für  $n = 0, \dots, N-1$ ,  $k = 1, \dots, s$ .
4. *Schritt:* Berechne für  $\ell = 0, \dots, L$  die Faltungsgewichte  $\tilde{W}_n^{\pm, \ell} \in \mathbb{C}^{sM^\pm \times sd^\pm}$ ,  $n = 0, \dots, N-1$ , die durch (5.12) definiert sind, wieder mit Hilfe der FFT. Auch hier lassen sich die Berechnungen für jedes  $\ell$  parallel durchführen.
5. *Schritt:* Berechne für  $\ell = 0, \dots, L$  aus (5.11) die zellenweisen Approximationen  $\mathbf{U}_{n\ell}^\pm \in \mathbb{C}^{M^\pm} \approx u(x, y, t_{n+1})|_{C_\ell^\pm}$  für  $n = 0, \dots, N-1$ . Hierbei benötigen wir für die rechte Seite in (5.11) neben den Gewichten aus Schritt 4, nur noch die Vektoren  $\Theta_n^\pm \in \mathbb{C}^{sd^\pm}$ , die wir in Schritt 3 erhalten.

---

**Algorithmus 5.1** : Berechnung der Operatoren aus Schritt 1
 

---

**Input** : Schrittzahl  $N$ , Schrittweite  $h$ , Matrix  $\Delta$  des Runge-Kutta Verfahrens

Genauigkeit  $\varepsilon > 0$  aus Abschnitt 2.3.1

**Output** : Operatoren  $\mathcal{P}^\pm(Z_j^2)$ ,  $\mathcal{S}^\pm(Z_j^2)$ ,  $\Lambda^\pm(Z_j^2)$  für  $j = 0, \dots, J$

Setze  $J$  als die nächstgrößte Zweierpotenz nach  $N$ ;

**for**  $j = 0 : J - 1$  :

Berechne die Matrix  $Z_j = \frac{\Delta(\zeta_j)}{h}$  für  $\zeta_j = \varepsilon^{1/(2J)} e^{2\pi i j / J}$ , vergleiche (2.4);

Bestimme die Eigenwertzerlegung  $Z_j = U_j D_j U_j^{-1}$  mit  $D_j = \text{diag}((D_j)_1, \dots, (D_j)_s)$ ;

**for**  $i = 1 : s$  :

Erhalte die Operatoren  $\mathcal{T}_{kl}^\pm((D_j)_i^2)$  mit  $k, l \in \{0, 1\}$  über die Lösung der lokalen Helmholtz-Probleme (3.8) ;

Löse die Ricatti-Gleichung (3.12), siehe Bemerkung 5.4, und erhalte  $\mathcal{P}^\pm((D_j)_i^2)$ ;

Berechne  $\Lambda^\pm((D_j)_i^2)$  aus (3.11) und  $\mathcal{S}^\pm((D_j)_i^2)$  aus (3.10);

Bestimme die Operatoren  $\mathcal{P}^\pm(Z_j^2) = U_j \text{diag}\left(\left(\mathcal{P}^\pm\left((D_j)_i^2\right)\right)_{i=1}^s\right) U_j^{-1}$  sowie analog  $\Lambda^\pm(Z_j^2)$  und  $\mathcal{S}^\pm(Z_j^2)$

---

**Bemerkung 5.4.** Die Ricatti-Gleichung (3.12) kann mit Hilfe eines modifizierten Newton-Verfahrens gelöst werden, das in [23] beschrieben wird und die Anforderung an den Spektralradius durch einen Projektionsschritt in jeder Iteration berücksichtigt. Der Newton-Schritt ist hierbei für einen Startwert  $X_0$  gegeben durch  $X_{k+1} = X_k - Q_k$  für  $k = 0, 1, \dots$ , wobei sich  $Q_k$  als Lösung der Lyapunov-Gleichung

$$\mathcal{T}_{s,10}^+ (X_k Q_k + Q_k X_k) + (\mathcal{T}_{s,00}^+ + \mathcal{T}_{s,11}^+) Q_k = \mathcal{T}_{s,10}^+ X_k^2 - (\mathcal{T}_{s,00}^+ + \mathcal{T}_{s,11}^+) X_k + \mathcal{T}_{s,01}^+$$

mit Hilfe des Bartels-Stewart-Algorithmus berechnen lässt, siehe dazu die Implementierung [64].

Sind wir nur an der Lösung im Gebiet  $\Omega_0$  interessiert, so können wir sowohl in Algorithmus 5.1 auf die Berechnung der Operatoren  $\mathcal{S}^\pm(Z_j^2)$  verzichten als auch die Schritte 4 und 5 auslassen. Des Weiteren ist es nicht notwendig, links und rechts von  $\Omega_0$  dieselbe Anzahl an Zellen zu betrachten. Dies wurde im obigen Algorithmus lediglich der besseren Übersicht halber angenommen.

## 5.3 Numerische Experimente

### 5.3.1 Wellengleichung in einer Raumdimension

Im Gegensatz zur Subdiffusionsgleichung aus dem vorherigen Kapitel ist bei der Betrachtung der Wellenausbreitung besonders der inhomogene Fall interessant, also der, in dem die Funktion  $\rho$  keinen konstanten Wert annimmt. Daher untersuchen wir den schnellen Faltungsalgorithmus zunächst unter Verwendung der Überlegungen zur eindimensionalen inhomogenen Helmholtz-Gleichung aus Abschnitt 3.2.2.2. Für das Testproblem in einer Raumdimension setzen wir dazu

$$\left| \begin{array}{ll} \rho(x) \partial_t^2 u(x, t) = \partial_x^2 u(x, t) & \text{in } \mathbb{R} \times \mathbb{R}^+, \\ u(x, 0) = \chi_{[-1/2, 1/2]}(x) \cos^4(x\pi) & \text{in } \mathbb{R}, \end{array} \right. \quad (5.19)$$

wobei die Funktion  $\rho$  entsprechend (H2) stückweise definiert ist als

$$\rho(x) = \begin{cases} \rho^-(x) & \text{in } (-\infty, -1], \\ 1 & \text{in } (-1, 1), \\ \rho^+(x) & \text{in } [1, \infty). \end{cases}$$

Hierbei wählen wir die Periodenlängen  $L^\pm = 2$ , wodurch die Zellen als  $C_j^+ = (1 + 2j, 3 + 2j)$  beziehungsweise  $C_j^- = (-3 - 2j, -1 - 2j)$  für  $j \in \mathbb{N}_0$  gegeben sind und definieren die periodischen Funktionen  $\rho^\pm$  als

$$\rho^+(x)|_{C_0^+} := \begin{cases} 1, & x \in (1, 2), \\ 4, & x \in (2, 3) \end{cases} \quad \text{bzw.} \quad \rho^-(x)|_{C_0^-} := \begin{cases} 1, & x \in (-2, -1), \\ 4, & x \in (-3, -2). \end{cases}$$

Insbesondere haben wir für obige Funktion  $\rho^+$  in Abschnitt 3.2.3.2 die für die Ortsdiskretisierung des Laplace-Operators mittels zentraler Finiter Differenzen exakten NtD-Operatoren  $\Lambda_s^+(\zeta_{s,+}, \eta_{s,+})$  hergeleitet. Wie sich leicht nachvollziehen lässt, gilt hier nach Konstruktion außerdem  $\Lambda_s^+(\zeta_{s,+}, \eta_{s,+}) = \Lambda_s^-(\zeta_{s,+}, \eta_{s,+})$ .

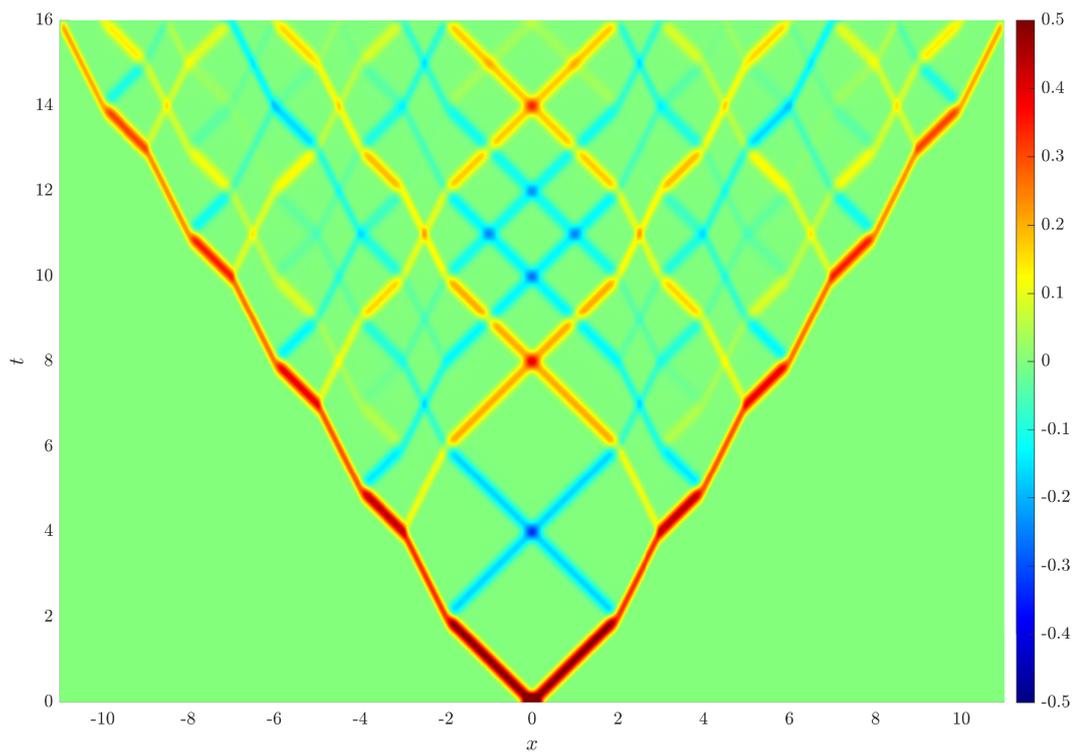
Anstelle der in Abschnitt 3.2.3.2 hergeleiteten Darstellung der ortsdiskreten Randbedingungen  $\Lambda_s^+$  können wir diese auch über die Lösung lokaler Helmholtz-Probleme, wie in Algorithmus 5.1 beschrieben, ausrechnen. Vergleichen wir die NtD-Operatoren, die über die hergeleitete Darstellung berechnet werden, mit denen, die über den angegebenen Algorithmus bestimmt werden, ergibt sich ein absoluter Fehler im Bereich von  $10^{-13}$  bis  $10^{-16}$ , was zeigt, dass die Berechnung auf beide Arten sinnvolle Ergebnisse liefert. Wir können in diesem numerischen Experiment daher die Darstellung aus Abschnitt 3.2.3.2 verwenden.

Die Approximation der transparenten Randbedingungen  $\lambda^\pm$  für das Problem (5.19) eingeschränkt auf  $\Omega_0 := (-1, 1)$ , die über (5.9) definiert sind, lässt sich damit wie in (5.14) unter Verwendung der FFT berechnen.

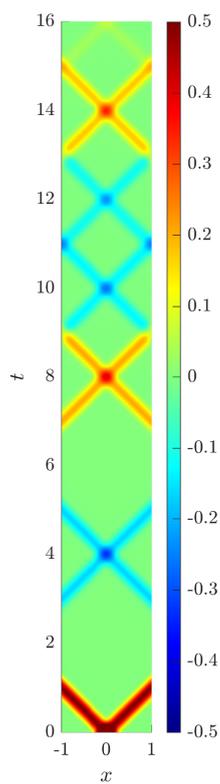
Wie schon im numerischen Experiment zur fraktionalen Schrödinger-Gleichung, wollen wir nun die Genauigkeit der diskreten transparenten Randbedingungen  $\lambda^\pm$  testen. Dazu berechnen wir die Lösung der Wellengleichung (5.19) im Intervall  $\Omega_0$  unter Verwendung des vorgestellten Verfahrens zu den Zeitpunkten  $t = 0, h, \dots, Nh = 16$ , wobei wir die Zeitschrittweite auf  $h = 1/400$  festlegen. Die Intervalle  $\overline{\Omega_0}$  und  $\overline{C_0^\pm}$  diskretisieren wir jeweils mit 801 Gitterpunkten. Für die Berechnung der Faltungsgewichte, siehe Schritt 2 in Abschnitt 5.2.4, setzen wir hierbei  $J = 2^{1+\log_2(N)}$  und  $\varepsilon = 10^{-13}$ , was sich in den Tests als gute Wahl gezeigt hat.

Um eine Referenzlösung zu erhalten, bestimmen wir die Lösung der Wellengleichung auf dem Intervall  $[-11, 11]$  mit Dirichlet-Randwerten zu denselben Zeitpunkten und gleicher Zeitschrittweite. Hierbei diskretisieren wir das Intervall mit 8801 Gitterpunkten. In beiden Rechnungen verwenden wir das dreistufige Radau IIA Verfahren der Ordnung 5 zur Zeitdiskretisierung.

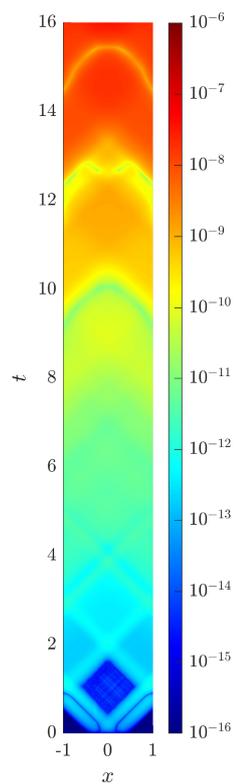
In Abbildung 5.1 stellen wir den Betrag der berechneten Referenzlösung sowie den Betrag der Lösung auf dem Intervall  $\Omega_0$  und den absoluten Fehler dar, der sich durch den Vergleich beider Lösungen ergibt. Im Zeitbereich  $t \in (0, 1]$  lässt sich gut erkennen, dass Fehler im Innern von  $\Omega_0$  entstehen, die NtD-Operatoren jedoch im Bereich der Maschinengenauigkeit exakt ausgerechnet wurden. Der mit fortschreitender Zeit wachsende Fehler lässt sich unter anderem durch Rundungsfehler beim Lösen der Gleichungssysteme erklären. Zusätzlich haben die Parameter  $\varepsilon$  und  $J$  Einfluss auf die Approximation der Faltungsgewichte für die transparenten Randbedingungen.



(a) Referenzlösung



(b) Lösung



(c) Fehler

Abbildung 5.1: Beträge der Referenzlösung im Gebiet  $[-11, 11]$  mit Dirichlet-Randwerten, der Lösung in  $\Omega_0$  mit transparenten Randbedingungen sowie des absoluten Fehlers in  $\Omega_0$  zu den Zeiten  $t = 0, h, \dots, 16$ .

### 5.3.2 Wellengleichung in zwei Raumdimensionen

Für unser Testproblem in zwei Raumdimensionen setzen wir  $\Omega_0 = (-1, 1)^2$  sowie  $\Omega^- = (-\infty, -1) \times (-1, 1)$  und  $\Omega^+ = (1, \infty) \times (-1, 1)$  und betrachten die Wellengleichung **(W)** unter den Anfangsbedingungen

$$u_0(x, y) = \chi_{[-1/2, 1/2]^2}(x, y) \cos^4(x) \cos^4(y) \quad \text{und} \quad v_0 \equiv 0$$

sowie der zeitabhängigen Quellfunktion

$$f(x, y, t) = \chi_{\Omega_0}(x, y) e^{-30(x^2+y^2)} \cos(\pi t).$$

Im Gegensatz zum vorherigen Abschnitt wählen wir eine kompliziertere Funktion  $\rho$ , so dass wir die benötigten Operatoren  $\mathcal{P}^\pm$ ,  $\mathcal{S}^\pm$  und  $\Lambda^\pm$  nicht mehr in einer geschlossenen Form darstellen können. Die Periodenlängen der Funktionen  $\rho^\pm$  wählen wir hierbei erneut als  $L^\pm = 2$ , wodurch die Zellen  $C_j^\pm$  jeweils in  $x$ - und  $y$ -Richtung von der Länge 2 sind. Wir definieren die periodischen Funktionen

$$\begin{aligned} \rho^+(x, y) &:= 1 + 16e^{-10(x-2(j+1))^2} e^{-10y^2} \quad \text{für } x \in C_j^+, j \in \mathbb{N}_0 \quad \text{sowie} \\ \rho^-(x, y) &:= 1 + \min(16 \sin(\pi x)^2, 4) e^{-10y^2} \quad \text{für } x \in \Omega^- \end{aligned}$$

und setzen  $\rho_0 \equiv 1$ . Abbildung 5.2 zeigt die gemäß (H2) stückweise zusammengesetzte Funktion  $\rho$ .

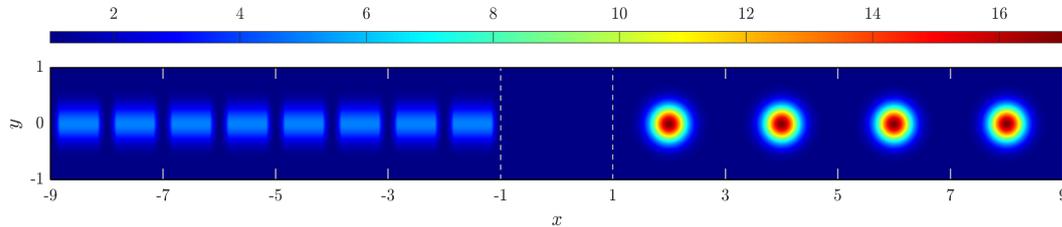


Abbildung 5.2: Funktion  $\rho$  auf einem Ausschnitt des Wellenleiters  $\Omega = \mathbb{R} \times (-1, 1)$ . Der lokale Defekt ist auf das Gebiet  $\Omega_0 = (-1, 1)^2$  beschränkt.

Wir überprüfen zunächst die Annahme aus Abschnitt 5.2.1, dass sowohl der NtD-Operator als auch der Operator  $\mathcal{S}^+ \circ (\mathcal{P}^+)^j$  das richtige Abfallverhalten in der rechten Halbebene erfüllen. Hierfür diskretisieren wir sowohl das Gebiet  $\overline{\Omega_0}$  als auch die Zellen  $\overline{C_0^\pm}$  mit einem äquidistanten Gitter mit  $M_x$  Gitterpunkten in  $x$ -Richtung und  $M_y$  Gitterpunkten in  $y$ -Richtung und verwenden bilineare Viereckselemente der Ordnung 2.

Hier sei erwähnt, dass wir die Anzahl  $M_x$  der Einfachheit halber gleich für  $\overline{C_0^\pm}$  und  $\overline{\Omega_0}$  setzen, es jedoch auch möglich ist, die Gebiete mit unterschiedlich vielen Gitterpunkten in  $x$ -Richtung zu diskretisieren. Wichtig ist an dieser Stelle, dass die Freiheitsgrade bezüglich des Randes  $\Gamma_0^+$  in  $\Omega_0$  und  $C_0^+$  identisch sind, genauso wie die Freiheitsgrade bezüglich  $\Gamma_0^-$  in  $\Omega_0$  und  $C_0^-$ .

In Abbildung 5.3 stellen wir die Spektralnorm der Operatoren  $\Lambda^-(\lambda^2)$  und  $\mathcal{S}^-(\lambda^2)$  sowie

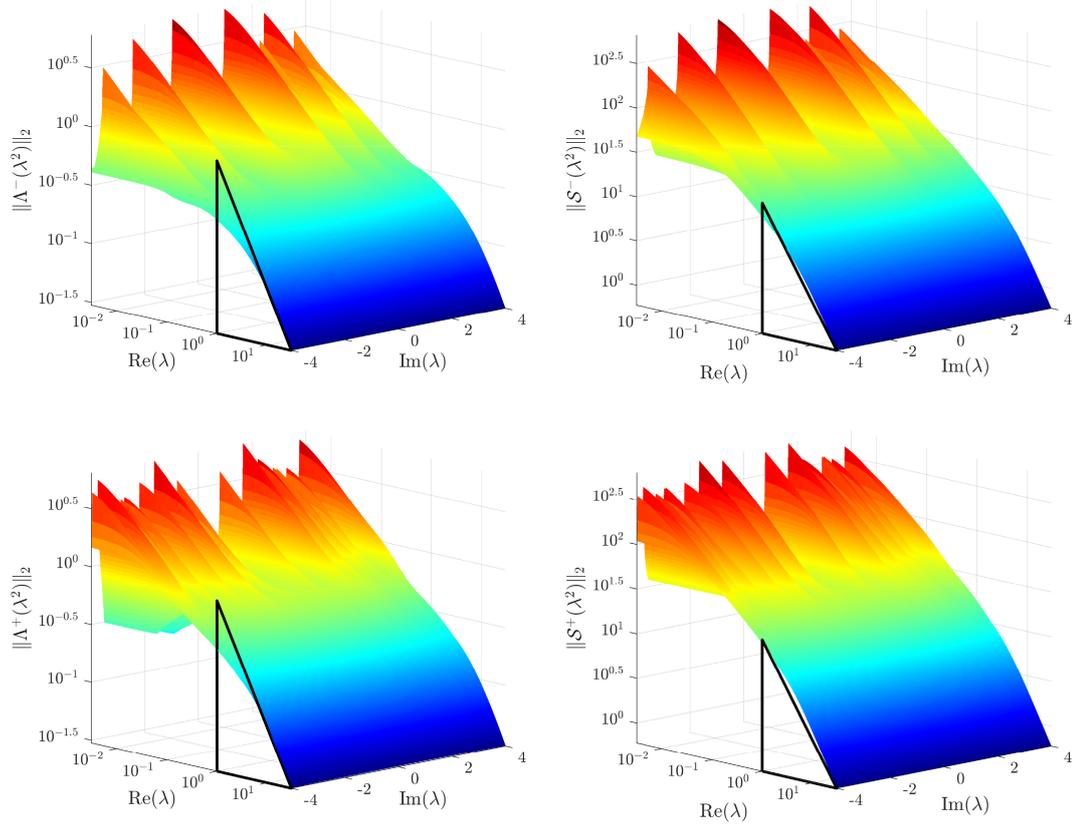


Abbildung 5.3: Spektralnorm der Operatoren  $\Lambda^\pm(\lambda^2)$  und  $\mathcal{S}^\pm(\lambda^2)$  für  $\operatorname{Re}(\lambda) \in (0.0032, 30)$  und  $\operatorname{Im}(\lambda) \in (-4, 4)$ . Um das Abfallverhalten besser ablesen zu können, sind hier die  $z$ - und  $x$ -Achse logarithmiert, das Dreieck stellt eine Steigung von  $-1$  dar.

$\Lambda^+(\lambda^2)$  und  $\mathcal{S}^+(\lambda^2)$  dar. Hierbei setzen wir  $M_x = M_y = 40$ . In der Abbildung ist zu erkennen, dass die Operatoren Singularitäten auf der imaginären Achse aufweisen, jedoch in der rechten Halbebene über das gewünschte Abfallverhalten verfügen. Wie aus den Überlegungen aus Kapitel 3 hervorgeht, sind für die zweidimensionale Wellengleichung mit Dirichlet-Randwerten keine Singularitäten an  $\lambda = 0$  zu erwarten.

Um eine Vorstellung der Wellenausbreitung in diesem Testproblem zu erhalten, bestimmen wir unter Verwendung der gerade beschriebenen Ortsdiskretisierung sowie des Radau IIA Verfahrens der Ordnung 3 die Approximation an die Lösung  $u(t_N)|_{\Omega_0}$  zu verschiedenen Zeitpunkten  $t_N = Nh$  und setzen die Lösung wie in Abschnitt 5.2.1.1 beschrieben in die Zellen  $C_j^\pm$  für  $j = 0, \dots, 4$  fort. In Abbildung 5.4 ist die Lösung im Gebiet  $[-11, 11] \times [-1, 1]$  für Zeitpunkte zwischen 0.12 und 20 dargestellt. Für die Berechnung der Faltungsgewichte, siehe Schritt 2 in Abschnitt 5.2.4, wählen wir hierbei wie im eindimensionalen Beispiel  $\varepsilon = 10^{-13}$  und  $J = 2^{1+\log_2(N)}$ . Das Gebiet  $\overline{\Omega}_0 = [-1, 1]^2$  sowie die Zellen  $\overline{C}_j^\pm$  diskretisieren wir wieder mit  $M_x = M_y = 40$  Gitterpunkten, das heißt, in dem Gebiet  $[-11, 11] \times [-1, 1]$  betrachten wir insgesamt 430 Gitterpunkte in  $x$ -Richtung. Die Zeitschrittweite setzen wir fest auf  $h = 1/40$ .

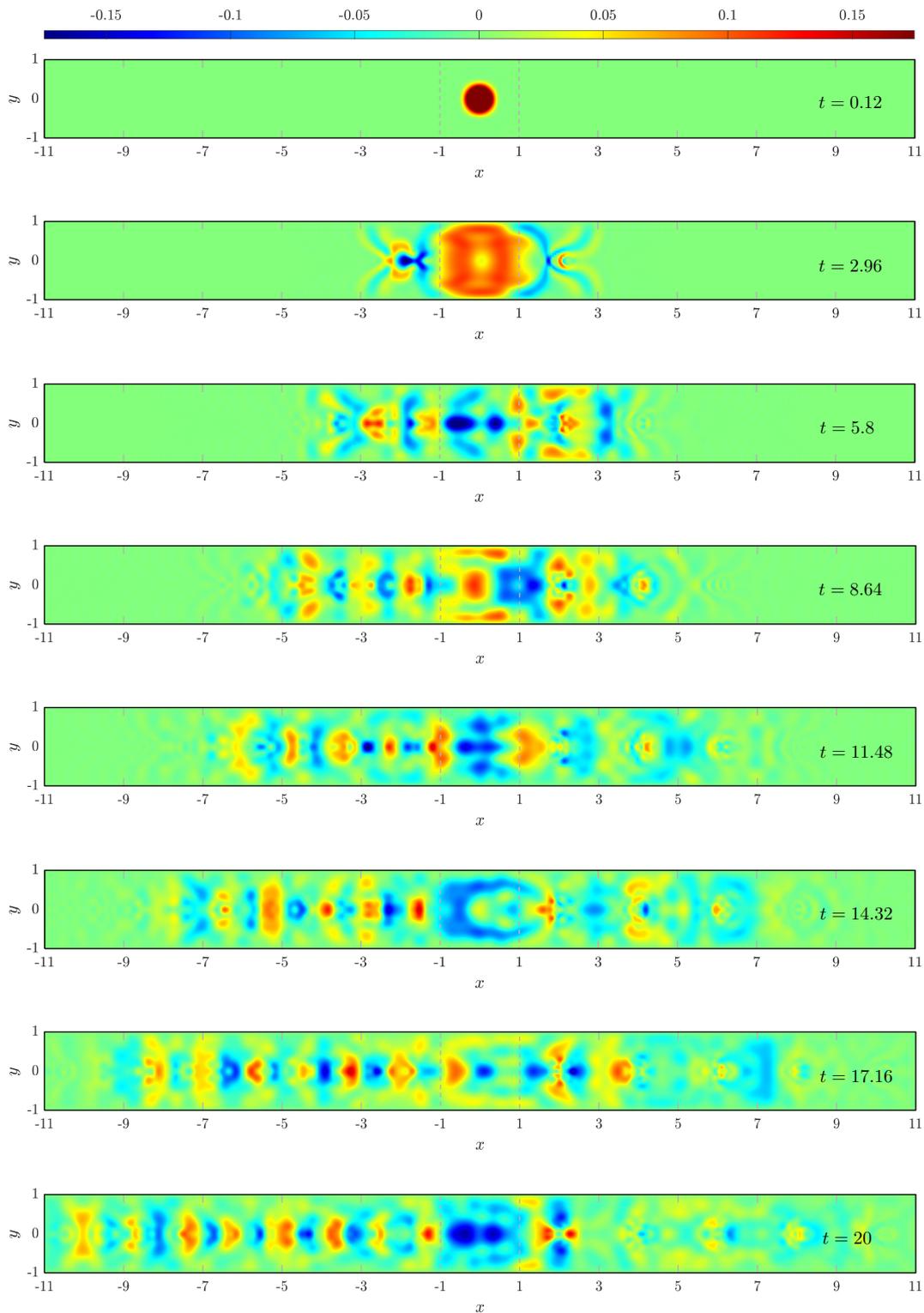


Abbildung 5.4: Absolutbetrag der Lösung des zweidimensionalen Testproblems im Gebiet  $[-11, 11] \times [-1, 1]$  zu den verschiedenen angegebenen Zeitpunkten  $t$ .

Zuletzt werfen wir noch einen Blick auf die Konvergenz des Verfahrens und betrachten dazu zwei unterschiedliche Zeitpunkte. Zum Zeitpunkt  $t_N = 1.23$  berechnen wir die Lösung  $u_N$  des Testbeispiels im Gebiet  $[-3, 3] \times [-1, 1]$ , indem wir das lokale Problem in  $\Omega_0$  lösen und die Lösung jeweils in einer Zelle links und einer rechts fortsetzen. Für die Lösung zum Zeitpunkt  $t_N = 4.56$  berechnen wir analog die Lösung in  $\Omega_0$  und setzen diese in die Zellen  $C_j^\pm$  für  $j = 0, 1, 2$  fort. In beiden Fällen setzen wir  $M_x = M_y = 60$  sowie erneut  $\varepsilon = 10^{-13}$  und  $J = 2^{1+\log_2(N)}$ .

Die Referenzlösungen erhalten wir durch das Lösen des Testproblems mit Dirichlet-Randwerten: Im Fall  $t_N = 1.34$  berechnen wir die Referenzlösung im Gebiet  $[-3, 3] \times [-1, 1]$  und setzen Dirichlet-Werte an die Ränder  $\{\pm 3\} \times [-1, 1]$ , im Fall  $t_N = 4.56$  vergrößern wir das Gebiet auf  $[-7, 7] \times [-1, 1]$  und betrachten Dirichlet-Werte bei  $\{\pm 7\} \times [-1, 1]$ . Für die Referenzlösung wählen wir dieselbe Ortsschrittweiten wie in den zu vergleichenden Lösungen, das heißt, in  $x$ -Richtung diskretisieren wir mit 178 beziehungsweise 414 Punkten. Während wir hier das Radau IIA Verfahren der Ordnung 5 mit einer Zeitschrittweite von etwa  $8.234 \cdot 10^{-5}$  verwenden, testen wir in der Berechnung der Lösung  $u_N$  die Approximation basierend auf den Radau IIA Verfahren der Ordnungen 1, 3 und 5 und stellen in den Abbildungen 5.5 und 5.6 den Fehler in der  $L^2$ -Norm in Abhängigkeit von verschiedenen Zeitschrittweiten dar.

Die linken Seiten der Abbildungen zeigen hierbei den Fehler eingeschränkt auf das Gebiet  $\Omega_0$ , das heißt, den Fehler, der durch die Lösung des Systems in Abschnitt 5.2.3 entsteht. Wie wir hierbei für den ortskontinuierlichen, aber zeitdiskreten Fall begründet haben, erwarten wir eine Konvergenzordnung, die der Ordnung des zugrunde liegenden Runge-Kutta Verfahrens entspricht, das heißt, hier erwarten wir die Ordnungen 1, 3 und 5. Diese Beobachtung wird in beiden Abbildungen bestätigt.

Die rechte Seite zeigt den Fehler in der Zelle  $C_0^+ = (1, 3) \times (-1, 1)$ . Hier ist eine Konvergenzordnung zu erwarten, die von der Runge-Kutta basierten Faltungsquadratur bestimmt wird, siehe dazu die Überlegungen in Abschnitt 5.2.1.1. Auch hier können die erhofften Ordnungen in der Abbildung beobachtet werden.

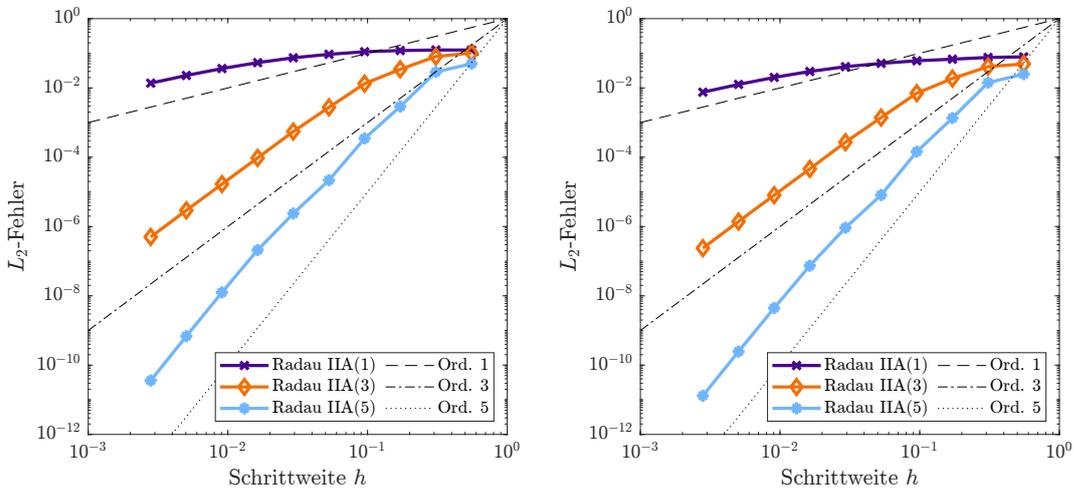


Abbildung 5.5: Fehler im Gebiet  $\Omega_0 = (-1, 1)^2$  (links) und im der Zelle  $C_0^+ = (1, 3) \times (-1, 1)$  (rechts) zum Zeitpunkt  $t = 1.23$  in der  $L_2$ -Norm in Abhängigkeit der Zeitschrittweite  $h$  für die Radau IIA Verfahren der Ordnungen 1, 3 und 5.

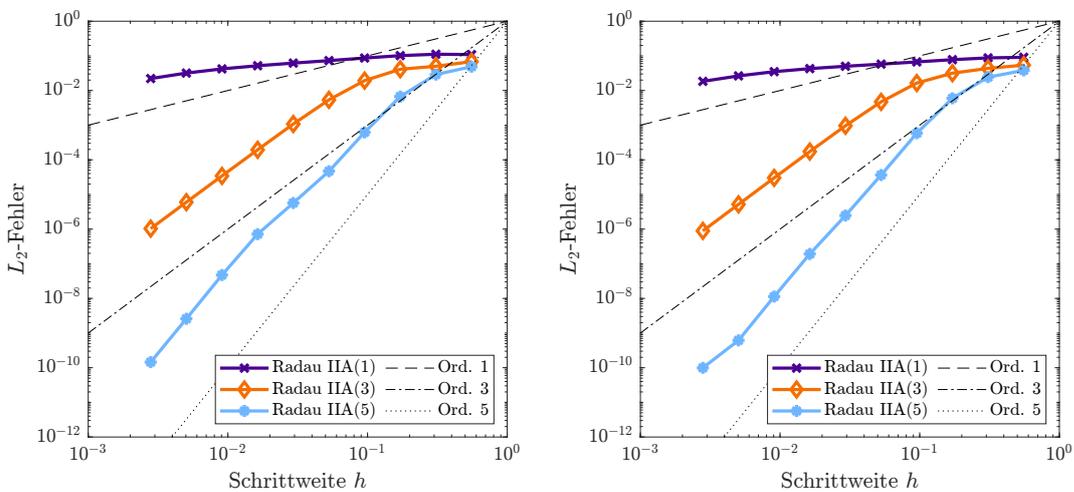


Abbildung 5.6: Fehler im Gebiet  $\Omega_0 = (-1, 1)^2$  (links) und im der Zelle  $C_0^+ = (1, 3) \times (-1, 1)$  (rechts) zum Zeitpunkt  $t = 4.56$  in der  $L_2$ -Norm in Abhängigkeit der Zeitschrittweite  $h$  für die Radau IIA Verfahren der Ordnungen 1, 3 und 5.

## KAPITEL 6

---

# AUSBLICK

Abschließend wollen wir einen kurzen Blick auf mögliche Aspekte werfen, die als Gegenstand weiterer Untersuchungen von Interesse sein könnten.

Im ersten Kapitel haben wir die auf der Polynom-Interpolation der Funktion  $g : \mathbb{R} \rightarrow X$ , vergleiche (1.1), basierende exponentielle Faltungsquadratur für den Fall einer sektoriellen Laplace-Transformierten hergeleitet, bei der wir von einer polynomiellen Schranke  $\mathcal{O}(\lambda^\mu)$  unter der Annahme  $\mu > 0$  ausgegangen sind. Wie bereits erwähnt, kann das Verfahren auch als Spezialfall des exponentiellen Runge-Kutta Verfahren vom Kollokationstyp verstanden werden und stellt eine Erweiterung der Überlegungen aus den Papern [36, 37] dar, in denen ausschließlich der Fall  $\mu = 1$  untersucht wurde. Dort konstruieren die Autoren neben der exponentiellen Quadratur auch allgemeine exponentielle Runge-Kutta Verfahren sowie unter anderem exponentielle Mehrschritt-Verfahren. Hier könnte eine Erweiterung auf den Fall  $\mu > 0$  und insbesondere die Untersuchung, ob und unter welchen Umständen der Parameter  $\mu$  Einfluss auf die Konvergenzordnungen hat, von Interesse sein.

Das Verfahren zur Bestimmung der zeitharmonischen Operatoren  $\Lambda^\pm(s)$ ,  $\mathcal{P}^\pm(s)$  und  $\Lambda^\pm(s)$ , das wir im dritten Kapitel vorgestellt und bezüglich des Vorkommens von komplexen Frequenzen untersucht haben, wurde in dieser Arbeit, wie auch in [40], ausschließlich für Helmholtz-Probleme in ein- oder zweidimensionalen Wellenleitern verwendet. In anschließenden Veröffentlichungen der AutorInnen, siehe unter anderem [20, 21, 23], ist eine Erweiterung des Verfahrens auf den gesamten Raum  $\mathbb{R}^2$  zu finden. Hierbei wird die Lösung mit Hilfe von Floquet-Bloch-Transformationen auf die Behandlung eines zweidimensionalen Wellenleiters und damit auf das beschriebene Verfahren zurückgeführt. Damit können die Überlegungen in dieser Arbeit dazu genutzt werden, zeitabhängige Probleme, die den geforderten Zusammenhang zur Helmholtz-Gleichung aufweisen, auch in Medien zu lösen, die in beide Richtungen unbeschränkt sind.

Um die transparenten Randbedingungen sowie die zellenweise Lösungsdarstellung in Kapitel 5 für die Wellengleichung im periodischen, lokal gestörten Wellenleiter mit Hilfe der Runge-Kutta basierten Faltungsquadratur berechnen zu können, haben wir angenommen, dass die NtD-Operatoren  $\Lambda^\pm(\lambda^2)$  sowie die Operatoren  $\mathcal{S}^\pm(\lambda^2)$ , deren Herleitung wir in Kapitel 3 angegeben haben, die Annahme 2.2 über die polynomielle Beschränkung in der

rechten Halbebene  $\operatorname{Re}(\lambda)$  für den Parameter  $\mu_2 = 1$  erfüllen. Diese Behauptung konnten wir in Abschnitt 3.2 an zwei eindimensionalen Beispielen überprüfen und in Abschnitt 5.3.2 für den zweidimensionalen Fall in einem numerischen Experimenten beobachten. Ein Beweis der Abschätzungen für den eindimensionalen Wellenleiter mit beliebiger Funktion  $\rho$  sowie für den zweidimensionalen Wellenleiter bleibt an dieser Stelle jedoch noch offen.

Auf Grund des höheren Aufwands bei der Verwendung eines impliziten Verfahrens wird bei der Berechnung der Wellengleichung in der Regel ein explizites Verfahren zur Zeitdiskretisierung vorgezogen. Wie wir jedoch in den Überlegungen in Abschnitt 5.2.1 gesehen haben, können wir die Konvergenz der Approximation der Lösung im beschränkten Gebiet  $\Omega_0$  nur folgern, wenn dasselbe implizite Runge-Kutta Verfahren sowohl der Diskretisierung der Gleichung im Inneren als auch der verwendeten Faltungsquadratur zur Berechnung der transparenten Randbedingungen zugrunde liegt. Hierzu gibt es Ansätze, wie beispielsweise im kürzlich erschienenen Paper [2], in denen explizite Verfahren im Inneren mit auf impliziten Verfahren basierenden Faltungsquadraturen kombiniert werden können mit der Absicht, weiterhin ein stabiles und konvergentes Verfahren zu erhalten. Eine Anwendung dieser Verfahren auf das betrachtete Problem stellt ein weiteres interessantes Ziel für nachfolgende Betrachtungen dar.

---

# LITERATUR

- [1] W. Arendt u. a. *Vector-valued Laplace Transforms and Cauchy Problems*. Springer Basel AG, 2001.
- [2] K. Baker und L. Banjai. „Numerical analysis of a wave equation for lossy media obeying a frequency power law“. In: *IMA J. Numer. Anal.* 00 (2021), S. 1–35.
- [3] L. Banjai und C. Lubich. „An error analysis of Runge-Kutta convolution quadrature“. In: *BIT Numer. Math.* 51 (2011), S. 483–496.
- [4] L. Banjai, C. Lubich und J. M. Melenk. „Runge-Kutta convolution quadrature for operators arising in wave propagation“. In: *Numer. Math.* 119 (2011), S. 1–20.
- [5] E. Bazhlekova. „Fractional evolution equations in Banach spaces“. Diss. Technische Universität Eindhoven, 2001.
- [6] E. Bécache, A.-S. Bonnet-Ben Dhia und G. Legendre. „Perfectly matched layers for the convected Helmholtz equation“. In: *SIAM J. Numer. Anal.* 42.1 (2004), S. 409–433.
- [7] P. Brenner, M. Crouzeix und V. Thomée. „Single step methods for inhomogeneous linear differential equations in Banach space“. In: *RAIRO-Anal Numer-Num* 16.1 (1982), S. 5–26.
- [8] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer New York, 2011.
- [9] C. Chen und M. Li. „On fractional resolvent operator functions“. In: *Semigroup Forum* 80.1 (2010), S. 121–142.
- [10] J. Coatléven. „Analyse mathématique et numérique de quelques problèmes d’ondes en milieu périodique“. Diss. Ecole Polytechnique X, 2011.
- [11] J. Coatléven. „Transparent boundary conditions for evolution equations in infinite periodic strips“. In: *SIAM J. Sci. Comput.* 34.3 (2012), A1563–A1583.
- [12] E. Cuesta, C. Lubich und C. Palencia. „Convolution quadrature time discretization of fractional diffusion-wave equations“. In: *Math. Comp.* 75.254 (2006), S. 673–696.
- [13] R. Dautray und J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology*. Bd. 2. Functional and Variational Methods. Springer-Verlag Berlin Heidelberg, 2000.

- [14] R. Dautray und J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology*. Bd. 5. Evolution Problems I. Springer-Verlag Berlin Heidelberg, 2000.
- [15] G. Doetsch. *Einführung in Theorie und Anwendung der Laplace-Transformation*. 2. Aufl. Birkhäuser Verlag, 1970.
- [16] K.-J. Engel und R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer-Verlag New York, 2000.
- [17] G. A. Evans und K. C. Chung. „Laplace transform inversions using optimal contours in the complex plane“. In: *Int. J. Comput. Math.* 73 (2000), S. 531–543.
- [18] M. Fischer. „Fast and parallel Runge-Kutta approximation of fractional evolution equations“. In: *SIAM J. Sci. Comput.* 41.2 (2019), A927–A947.
- [19] S. Fliss. „A Dirichlet-to-Neumann approach for the exact computation of guided modes in photonic crystal waveguides“. In: *SIAM J. Sci. Comput.* 35.2 (2013), B438–B461.
- [20] S. Fliss. „Etude mathématique et numérique de la propagation des ondes dans des milieux périodiques localement perturbés.“ Diss. Français: Ecole Polytechnique X, 2009.
- [21] S. Fliss und P. Joly. „Exact boundary conditions for time-harmonic wave propagation in locally perturbed periodic media“. In: *Appl. Numer. Math.* 59 (2009), S. 2155–2178.
- [22] S. Fliss und P. Joly. „Solutions of the time-harmonic wave equation in periodic waveguides: Asymptotic behaviour and radiation condition“. In: *Arch. Ration. Mech. Anal.* 219 (2016), S. 349–386.
- [23] S. Fliss und P. Joly. „Wave propagation in locally perturbed periodic media (case with absorption): Numerical aspects“. In: *J. Comput. Phys.* 231 (2012), S. 1244–1271.
- [24] S. Fliss, D. Klindworth und K. Schmidt. „Robin-to-Robin transparent boundary conditions for the computation of guided modes in photonic crystal waveguides“. In: *BIT Numer Math* 55 (2015), S. 81–115.
- [25] R. Garrappa. „Trapezoidal methods for fractional differential equations: Theoretical and computational aspects“. In: *Math. Comput. Simulation* 110 (2015), S. 96–112.
- [26] R. Garrappa und M. Popolizio. „Exponential integrators for fractional differential equations“. In: *Fractional Calculus: Theory*. 2014.
- [27] R. Garrappa und M. Popolizio. „Exponential quadrature rules for linear fractional differential equations“. In: *Mediterr. J. Math.* 12 (2015), S. 219–244.
- [28] R. Garrappa und M. Popolizio. „On accurate product integration rules for linear fractional differential equations“. In: *J. Comput. Appl. Math.* 235 (2011), S. 1085–1097.
- [29] I. P. Gavriljuk, W. Hackbusch und B. N. Khoromskij. „Data-sparse approximation to a class of operator-valued functions“. In: *Math. Comp.* 74.250 (2005), S. 681–708.
- [30] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. SIAM ed. Philadelphia: Society for Industrial und Applied Mathematics, 2011.
- [31] M. Haase. *The Functional Calculus for Sectorial Operators*. 1. Aufl. Birkhäuser Verlag, 2006.

- 
- [32] E. Hairer, C. Lubich und M. Schlichte. „Fast numerical solution of nonlinear Volterra convolution equations“. In: *SIAM J. Sci. Stat. Comput.* 6.3 (1985), S. 532–541.
- [33] E. Hairer und G. Wanner. *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*. 2. Aufl. Springer-Verlag, 2002.
- [34] I. Harari, I. Patlashenko und D. Givoli. „Dirichlet-to-Neumann maps for unbounded wave guides“. In: *J. Comput. Phys.* 143 (1998), S. 200–223.
- [35] N. Heymans und I. Podlubny. „Physical interpretation of initial conditions for fractional differential equations with Riemann-Liouville fractional derivatives“. In: *Rheol Acta* 45.5 (2006), S. 765–771.
- [36] M. Hochbruck und A. Ostermann. „Exponential integrators“. In: *Acta Numer.* 19 (2010), S. 209–286.
- [37] M. Hochbruck und A. Ostermann. „Exponential Runge-Kutta methods for parabolic problems“. In: *Appl. Numer. Math.* 53.2-4 (2005), S. 323–339.
- [38] S. Jiang u. a. „Fast evaluation of the Caputo fractional derivative and its applications to fractional diffusion equations“. In: *Commun. Comput. Phys.* 21.3 (2017), S. 650–678.
- [39] S. G. Johnson und J. D. Joannopoulos. *Photonic Crystals - The Road from Theory to Practice*. Bd. 1. Springer US, 2002.
- [40] P. Joly, J.-R. Li und S. Fliss. „Exact boundary conditions for periodic waveguides containing a local perturbation“. In: *Commun. Comput. Phys.* 1.6 (2006), S. 945–973.
- [41] A. K. Kaw. *Mechanics of Composite Materials*. Bd. 2. USA: CRC Press, 2006.
- [42] K. Li, J. Peng und J. Jia. „Cauchy problems for fractional differential equations with Riemann-Liouville fractional derivatives“. In: *J. Funct. Anal.* 263.2 (2012), S. 476–510.
- [43] M. López-Fernández und C. Palencia. „On the numerical inversion of the Laplace transform of certain holomorphic mappings“. In: *Appl. Numer. Math.* 51.2 (2004), S. 289–303.
- [44] M. López-Fernández, C. Palencia und A. Schädle. „A spectral order method for inverting sectorial Laplace transforms“. In: *SIAM J. Numer. Anal.* 44.3 (2006), S. 1332–1350.
- [45] M. López-Fernández u. a. „Fast Runge-Kutta approximation of inhomogeneous parabolic equations“. In: *Numer. Math.* 102.2 (2005), S. 277–291.
- [46] C. Lubich. „Convolution quadrature and discretized operational calculus. I“. In: *Numer. Math.* 52.2 (1988), S. 129–145.
- [47] C. Lubich. „Convolution quadrature and discretized operational calculus. II“. In: *Numer. Math.* 52.4 (1988), S. 413–425.
- [48] C. Lubich und A. Ostermann. „Runge-Kutta methods for parabolic equations and convolution quadrature“. In: *Math. Comp.* 60.201 (1993), S. 105–131.
- [49] C. Lubich und A. Schädle. „Fast convolution for nonreflecting boundary conditions“. In: *SIAM J. Sci. Comput.* 24.1 (2002), S. 161–182.

- [50] F. Mainardi. *Fractional Calculus and Waves in Linear Viscoelasticity. An Introduction to Mathematical Models*. London: Imperial College Press, 2012.
- [51] S. De Marchi und M. Vianello. „Peano’s kernel theorem for vector-valued functions and some applications“. In: *Numer. Funct. Anal. Optim.* 17.1+2 (1996), S. 57–64.
- [52] A. Mizrahi und L. Schächter. „Bragg reflection waveguides with a matching layer“. In: *Opt. Express* 12.14 (2004), S. 3156–3170.
- [53] K. Morgenröther und P. Werner. „On the principles of limiting absorption and limiting amplitude for a class of locally perturbed waveguides. Part 1: Time-independent theory“. In: *Math. Methods Appl. Sci.* 10 (1988), S. 125–144.
- [54] A. Al-Omari und H. Al-Saadi. „Existence of the classical and strong solutions for fractional semilinear initial value problems“. In: *Bound. Value Probl.* 157 (2018), S. 1–13.
- [55] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Applied Mathematical Sciences 44. Springer-Verlag New York, 1983.
- [56] I. Podlubny. *Fractional Differential Equations*. San Diego: Academic Press, 1999.
- [57] I. Podlubny. *Mittag-Leffler function*. 2012. URL: <https://de.mathworks.com/matlabcentral/fileexchange/8738-mittag-leffler-function/content/mlf.m>.
- [58] J. Prüss. *Evolutionary Integral Equations and Applications*. Springer Basel AG, 1993.
- [59] K. Sakoda. *Optical Properties of Photonic Crystals*. Bd. 2. Springer-Verlag Berlin Heidelberg, 2005.
- [60] A. Schädle, M. López-Fernández und C. Lubich. „Fast and oblivious convolution quadrature“. In: *SIAM J. Sci. Comput.* 28.2 (2006), S. 421–438.
- [61] M. R. Spiegel. *Laplace-Transformationen. Theorie und Anwendung*. Übers. von R. Michel. Düsseldorf: McGraw-Hill, 1977.
- [62] J. Stoer. *Numerische Mathematik 1*. 9. Aufl. Springer-Verlag Berlin Heidelberg, 2005.
- [63] A. Talbot. „The accurate numerical inversion of Laplace transforms“. In: *IMA J. Appl. Math.* 23 (1979), S. 97–120.
- [64] A. Townsend. *Bartels-Stewart algorithm*. 2012. URL: <https://github.com/ajt60gaibb/freelyap/blob/master/bartelsStewart.m>.
- [65] J. A. C. Weideman. „Optimizing Talbot’s contours for the inversion of the Laplace transform“. In: *SIAM J. Numer. Anal.* 44.6 (2006), S. 2342–2362.
- [66] J. A. C. Weideman und L. N. Trefethen. „Parabolic and hyperbolic contours for computing the Bromwich integral“. In: *Math. Comp.* 76.259 (2007), S. 1341–1356.
- [67] D. Werner. *Funktionalanalysis*. 8. Aufl. Springer Spektrum, 2018.
- [68] K. J. Witsch. „Examples of embedded eigenvalues for the Dirichlet Laplacian in perturbed waveguides“. In: *Math. Methods Appl. Sci.* 12 (1990), S. 91–93.
- [69] P. Yeh und A. Yariv. „Bragg reflection waveguides“. In: *Opt. Commun.* 19.3 (1976), S. 427–430.

---

# ERKLÄRUNG

Ich versichere an Eides statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Düsseldorf, den 24. Mai 2022

(Marina Fischer)