

Ein Beitrag zur Behandlung von Optimalsteuerungsaufgaben mit unendlichem Zeithorizont

- Theorie und Spektralverfahren -

Von der Fakultät für MINT - Mathematik, Informatik, Physik,
Elektro- und Informationstechnik
der Brandenburgischen Technischen Universität Cottbus-Senftenberg

genehmigte Dissertation

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

vorgelegt von

Master of Science Angie Burtchen
geboren am 18.03.1988 in W.-P. Stadt Guben, Deutschland

Vorsitzender: Prof. Dr. Ralf Wunderlich
Gutachterin: Prof. Dr. Sabine Pickenhain
Gutachter: Prof. Dr. Helmut Maurer
Gutachter: Prof. Dr. Matthias Gerdts

Tag der mündlichen Prüfung: 29.11.2019

Danksagung

Mein besonderer Dank gilt Prof. Sabine Pickenhain für die Möglichkeit, meine Dissertationsschrift unter ihrer Leitung anfertigen zu dürfen, für die Betreuung in den letzten Jahren und für die zahlreichen fachlichen Diskussionen und Hinweise, die zum Entstehen dieser Arbeit beigetragen haben.

Weiterhin möchte ich mich in besonderem Maße bei Prof. Helmut Maurer und Prof. Matthias Gerdts für die Begutachtung dieser Dissertation bedanken.

Für den Austausch, die interessanten Gespräche und das angenehme Arbeitsklima danke ich allen Kollegen und Kolleginnen des Lehrstuhls Mathematik insbesondere Optimierung sowie Martin Bähr, dessen wertvollen Impulse und Anregungen die Weiterentwicklung dieser Arbeit angetrieben haben.

Mein größter Dank gilt meinen Freunden, meiner Familie und insbesondere meinen Eltern. Bei ihnen möchte ich mich ganz besonders herzlich bedanken, da sie mich in jeder Phase des Studiums und während der Erstellung dieser Arbeit uneingeschränkt und unermüdlich unterstützt haben.

Zusammenfassung

Schwerpunkte dieser Arbeit sind das Studium linear-quadratischer Optimalsteuerungsaufgaben über dem a priori gegebenen unendlichen Zeithorizont $[0, \infty)$, die Untersuchung einer indirekten Spektralmethode zur Lösung dieser Aufgabenklasse sowie die Demonstration ihrer Anwendbarkeit mit Hilfe numerischer Experimente. Linear-quadratische Steuerungsaufgaben präsentieren eine bedeutende Aufgabenklasse in der Optimalsteuerung. Sie kommen sowohl für die Approximation nichtlinearer Aufgaben als auch bei Verfahren der sequentiellen quadratischen Optimierung zum Einsatz. Die Verwendung eines unendlichen Integrationsbereiches lässt sich dabei mit Hilfe verschiedener Anwendungsgebiete motivieren: bei dynamischen Systemen in der Ökonomie und der Biologie, wo aktuelle Entscheidungen Konsequenzen für zukünftige Generationen mit sich bringen; bei der Transformation von Aufgaben mit endlichem Horizont, um Singularitäten im Zielfunktional umgehen zu können (z.B. das Variationsproblem der Brachistochrone) und bei Anwendungen technischer und physikalischer Natur, bei denen sich das unendliche Intervall aus der Aufgabenstellung selbst ergibt (z.B. der quantenmechanische harmonische Oszillator). Darüber hinaus entstehen linear-quadratische Steuerungsaufgaben mit einem unendlichen Zeithorizont auch im Kontext der asymptotischen Stabilisierung dynamischer Systeme. Motiviert durch die zahlreichen Anwendungen wird diese Aufgabenklasse genauer studiert und somit der Ausgangspunkt für ein indirektes Lösungsverfahren geschaffen. Die Optimalsteuerungsaufgabe über $[0, \infty)$ wird in gewichteten Funktionenräumen formuliert, wodurch eine Einbettung in eine konvexe Hilbertraumtheorie möglich ist. Als Zustandsraum wird ein gewichteter Sobolev-Raum gewählt. Die Steuerungen werden in einem gewichteten Lebesgue-Raum betrachtet. Die Gewichtsfunktion tritt zusätzlich auch im Zielfunktional auf, wobei hier das Lebesgue-Integral zugrunde gelegt wird. Neben dem Nachweis der Existenz und Eindeutigkeit der Lösung beinhaltet die Arbeit den Beweis von Optimalitätsbedingungen in Form eines Pontryaginschen Maximumprinzips. Dieses liefert neben den kanonischen Gleichungen und der Maximumbedingung auch die Existenz einer Adjungierten, welche einer natürlichen Transversalitätsbedingung genügt. Für die Konstruktion der Spektralmethode spielt Letztere eine entscheidende Rolle. Durch eine sinnvolle Kombination der kanonischen Differentialgleichung mit dem Hauptsatz der Differential- und Integralrechnung gelingt es, aus der Transversalitätsbedingung eine Bestimmungsgleichung für den Anfangswert der adjungierten Funktion in Integralform abzuleiten. Die gewählten Approximationsansätze basieren auf den verallgemeinerten Laguerre-Polynomen, welche ein vollständiges Funktionensystem in einem gewichteten Lebesgue-Raum über $(0, \infty)$ repräsentieren. In Kombination mit einer Kollokation des kanonischen Differentialgleichungssystems und der integrierten Transversalitätsbedingung lässt sich damit das Lösen der Optimalsteuerungsaufgabe auf das Lösen eines linearen Gleichungssystems reduzieren. Der erfolgreiche Einsatz der Methode wird anhand verschiedener Beispiele der optimalen Steuerung und auch im Kontext der asymptotischen Stabilisierung eines dynamischen Systems veranschaulicht.

Abstract

This work focuses on linear-quadratic optimal control problems with the a priori given infinite horizon $[0, \infty)$, presents an indirect spectral method for the solution of this class of problems and shows its applicability in several numerical experiments. Linear-quadratic problems represent an important class in the theory of optimal control. They are used for the approximation of nonlinear problems as well as for methods of sequential quadratic programming. The formulation with an infinite horizon can be motivated by different applications: by dynamical systems in economics and biology, where recent decisions affect future generations; by problems with finite horizon, where a suitable transformation to an infinite horizon can be used to handle several singularities of the objective (e.g. the brachistochrone problem in calculus of variation) and by technical or physical applications, where the infinite horizon results naturally from the problem setting (e.g. the one-dimensional harmonic oscillator in quantum mechanics). Moreover linear-quadratic control problems with infinite horizon arise for stabilizing a dynamic system asymptotically. Motivated by this wide range of applications this class of problems is considered in detail to establish a basis for an indirect solution method. The problem formulation on $[0, \infty)$ implies a weighted space setting and can be treated as a linear-quadratic problem in Hilbert spaces with classical tools of convex analysis. For the state space a weighted Sobolev space is chosen and for the control space a weighted Lebesgue space is considered. The weight function is additionally involved in the objective, where the integral is understood in Lebesgue sense. The question of existence and uniqueness of an optimal solution is studied and a Pontryagin type maximum principle is established, which provides in addition to the canonical equation and the maximum condition, the existence of an adjoint satisfying a natural transversality condition. This plays a crucial role for the construction of the spectral method. Using a combination of the canonical equation and the fundamental theorem of calculus for Lebesgue integration an integral equation is established to estimate the initial value of the adjoint. The unknown functions are approximated by finite sums of generalized Laguerre polynomials, which form a complete orthogonal system in a weighted Lebesgue space over $(0, \infty)$. In combination with the integrated transversality condition the differential equations are collocated at a set of preselected points and the optimal control problem can be transformed into a linear equation system. Finally, the successful application of the proposed method is illustrated with some numerical examples of optimal control problems including a problem to find an asymptotically stabilizing control for a linearized dynamic system.

Inhaltsverzeichnis

Abbildungsverzeichnis	IX
Tabellenverzeichnis	XI
1 Einleitung	1
2 Räume und benötigte Grundlagen	7
2.1 Notationen	7
2.2 Klassische Lebesgue- und Sobolev-Räume	10
2.2.1 Die Räume $L_p(\Omega)$ und $W_p^k(\Omega)$	10
2.2.2 Die Räume $L_p^n(\Omega)$, $W_p^{k,n}(\Omega)$ und der Raum $L_\infty^{n \times m}(\Omega)$	15
2.2.3 Exkurs: Matrixnormen im $L_\infty^{n \times m}((0, \infty))$	23
2.3 Gewichtete Lebesgue- und Sobolev-Räume	29
2.3.1 Die Räume $L_p(\Omega, \omega)$, $W_p^k(\Omega, S_k)$ und $W_p^k(\Omega, \omega)$	29
2.3.2 Die Räume $L_p^n(\Omega, \omega)$ und $W_p^{k,n}(\Omega, \omega)$	34
2.4 Die Räume $L_2((0, \infty), \omega_{\alpha,\beta})$, $W_2^k((0, \infty), \omega_\beta)$ und $W_2^k((0, \infty), \mathbf{w}_{k,\beta})$	36
2.4.1 Die Gewichte $\omega_{\alpha,\beta}(t) := t^\alpha e^{-\beta t}$	36
2.4.2 Definition und Eigenschaften	37
2.5 Asymptotische Steuerbarkeit dynamischer Systeme	40
3 Ergebnisse aus der Approximationstheorie in Hilberträumen	45
3.1 Beste Approximation und Orthogonalprojektion	45
3.2 Orthonormalsysteme und Fourier-Entwicklung	49
3.3 Polynomiale Approximation in $L_2((a, b), \omega)$	60
3.3.1 Aufgabenstellung	61
3.3.2 Voraussetzungen und Vorbetrachtungen	63
3.3.3 Orthogonale Polynome und einige grundlegende Eigenschaften	66
4 Linear-quadratische Steuerungsaufgaben mit unendlichem Zeithorizont	71
4.1 Formulierung der Aufgabe $(LQ)_\infty$	71
4.2 Voraussetzungen	74
4.3 Differentialgleichungen in Sobolev-Räumen	77
4.4 Formulierung als Optimierungsproblem in einem Hilbertraum	82
4.5 Existenz und Eindeutigkeit einer Lösung	88
4.6 Optimalitätsbedingungen	97

5	Indirekte Spektralmethode zur Lösung linear-quadratischer Steuerungsaufgaben mit unendlichem Zeithorizont	115
5.1	Verallgemeinerte Laguerre-Polynome	119
5.1.1	Grundeigenschaften	120
5.1.2	Fourier-Laguerre-Approximation	125
5.2	Formulierung der Methode für $(LQ)_\infty$	148
5.2.1	Anfangswert der Adjungierten	148
5.2.2	Kollokationsgleichungen	149
5.2.3	Aufstellen der Koeffizientenmatrix	152
6	Beispiele	157
6.1	Beispiel von Garg, Hager und Rao	157
6.1.1	Diskussion der notwendigen Bedingungen	158
6.1.2	Optimale Lösung	160
6.1.3	Anwenden der Spektralmethode	160
6.1.4	Auswertung der numerischen Ergebnisse	161
6.2	Modifiziertes Beispiel	164
6.2.1	Diskussion der notwendigen Bedingungen	165
6.2.2	Optimale Lösung	167
6.2.3	Anwenden der Spektralmethode	168
6.2.4	Auswertung der numerischen Ergebnisse für $\beta = 1$ und $\beta = 2.4$	168
6.2.5	Auswertung der numerischen Ergebnisse für $\beta = 2.5$ und $\beta = 25$	174
6.3	Mathematisches Pendel	178
6.3.1	Diskussion der notwendigen Bedingungen	181
6.3.2	Anwenden der Spektralmethode	184
6.3.3	Auswertung der numerischen Ergebnisse für $\delta = 0.5$ und $\delta = 3$	185
6.4	Beispiel mit Steuerungsbeschränkungen	193
6.4.1	Diskussion der notwendigen Bedingungen	194
6.4.2	Anwenden der Spektralmethode	198
6.4.3	Auswertung der numerischen Ergebnisse für $x_0 = 1$ und $\bar{u} = 11/5$	199
7	Abschließende Bemerkungen	205
A	Quellcode für das Beispiel aus Abschnitt 6.1	207
A.1	LQR.m	207
A.2	Kollokation.m	207
A.3	LaguerreGenBeta.m	211
	Literaturverzeichnis	XIII

Abbildungsverzeichnis

6.1	$\beta = 2$: Näherungslösung $x_N(t)$ des Zustandes für $N = 8$ (—), $N = 10$ (—), $N = 12$ (—), $N = 14$ (—) im Vergleich mit der exakten Lösung $x^*(t)$ (—) . .	162
6.2	$\beta = 2$: Näherungslösung $u_N(t)$ der Steuerung für $N = 8$ (—), $N = 10$ (—), $N = 12$ (—), $N = 14$ (—) im Vergleich mit der exakten Lösung $u^*(t)$ (—) . .	163
6.3	Lösungsstruktur in Abhängigkeit von $\lambda_2(\beta)$ bzw. β	167
6.4	$\beta = 1$: Näherungslösung $x_N(t)$ des Zustandes für $N = 12$ (—), $N = 14$ (—), $N = 16$ (—), $N = 18$ (—) im Vergleich mit der exakten Lösung $x^*(t)$ (—) . .	172
6.5	$\beta = 1$: Näherungslösung $u_N(t)$ der Steuerung für $N = 12$ (—), $N = 14$ (—), $N = 16$ (—), $N = 18$ (—) im Vergleich mit der exakten Lösung $u^*(t)$ (—) . .	172
6.6	$\beta = 2.4$: Näherungslösung $x_N(t)$ des Zustandes für $N = 10$ (—), $N = 12$ (—), $N = 14$ (—), $N = 16$ (—) im Vergleich mit der exakten Lösung $x^*(t)$ (—)	173
6.7	$\beta = 2.4$: Näherungslösung $u_N(t)$ der Steuerung für $N = 10$ (—), $N = 12$ (—), $N = 14$ (—), $N = 16$ (—) im Vergleich mit der exakten Lösung $u^*(t)$ (—)	173
6.8	$\beta = 2.5$: Näherungslösung $x_N(t)$ des Zustandes für $N = 14$ (—), $N = 16$ (—), $N = 18$ (—), $N = 20$ (—) im Vergleich mit der exakten Lösung $x^*(t)$ (—)	176
6.9	$\beta = 2.5$: Näherungslösung $u_N(t)$ der Steuerung für $N = 14$ (—), $N = 16$ (—), $N = 18$ (—), $N = 20$ (—) im Vergleich mit der exakten Lösung $u^*(t)$ (—)	177
6.10	$\beta = 25$: Näherungslösung $x_N(t)$ des Zustandes für $N = 10$ (—), $N = 12$ (—), $N = 14$ (—), $N = 16$ (—) im Vergleich mit der exakten Lösung $x^*(t)$ (—) . .	177
6.11	$\beta = 25$: Näherungslösung $u_N(t)$ der Steuerung für $N = 10$ (—), $N = 12$ (—), $N = 14$ (—), $N = 16$ (—) im Vergleich mit der exakten Lösung $u^*(t)$ (—) . .	178
6.12	Mathematisches Pendel	178
6.13	$\delta = 0.5$ und $\beta = 20$: Näherungslösungen $u_N^{(1)}(t), u_N^{(2)}(t)$ der Steuerungen für $N = 18$ (—) im Vergleich mit den exakten Lösungen $u^{*(1)}(t), u^{*(2)}(t)$ (—) . .	189
6.14	$\delta = 0.5$ und $\beta = 20$: Näherungslösungen $x_N^{(1)}(t), x_N^{(2)}(t)$ der Zustände für $N = 18$ (—) im Vergleich mit den exakten Lösungen $x^{*(1)}(t), x^{*(2)}(t)$ (—) . .	190
6.15	$\delta = 3$ und $\beta = 50$: Näherungslösungen $u_N^{(1)}(t), u_N^{(2)}(t)$ der Steuerungen für $N = 24$ (—) im Vergleich mit den exakten Lösungen $u^{*(1)}(t), u^{*(2)}(t)$ (—) . .	191
6.16	$\delta = 3$ und $\beta = 50$: Näherungslösungen $x_N^{(1)}(t), x_N^{(2)}(t)$ der Zustände für $N = 24$ (—) im Vergleich mit den exakten Lösungen $x^{*(1)}(t), x^{*(2)}(t)$ (—)	192
6.17	Ansatz für die Steuerung bei Box-Restriktionen $[-\bar{u}, \bar{u}]$, $\bar{u} > 0$ mit $\bar{u}(\sqrt{2}-1) < x_0$	195
6.18	Darstellung der Schaltbedingung für $x_0 = 1$ und $\bar{u} = \frac{11}{5}$	197

6.19	$\beta = 4.01$: Näherungslösung $u_N(t)$ der Steuerung für $N = 9$ (—) im Vergleich mit der optimalen Steuerung $u^*(t)$ (—) auf $[0, 20]$ und in der Nähe des Umschaltzeitpunktes \bar{t}	200
6.20	Näherungslösung $u_N(t)$ der Steuerung für $\beta = 2$ und $N = 9$ und für $\beta = 5$ und $N = 7$	202

Tabellenverzeichnis

6.1	$\beta = 2$: L_{2,ω_β} -Fehler des Zustandes und der Steuerung	164
6.2	$\beta = 2$: Approximation des Zielfunktionswertes und des Anfangswertes der Adjungierten	164
6.3	Kollokationspunkte $\tau_N^{(N)}$ für $\beta = 1$ und $\beta = 2.4$	169
6.4	$\beta = 1$ und $\beta = 2.4$: L_{2,ω_β} -Fehler des Zustandes und der Steuerung	170
6.5	$\beta = 1$ und $\beta = 2.4$: Approximation des Zielfunktionswertes	170
6.6	$\beta = 1$ und $\beta = 2.4$: Approximation des Anfangswertes der Adjungierten . . .	171
6.7	Kollokationspunkte $\tau_N^{(N)}$ für $\beta = 2.5$ und $\beta = 25$	174
6.8	$\beta = 2.5$ und $\beta = 25$: Approximation des Zielfunktionswertes	175
6.9	$\beta = 2.5$ und $\beta = 25$: L_{2,ω_β} -Fehler des Zustandes und der Steuerung	175
6.10	$\beta = 2.5$ und $\beta = 25$: Approximation des Anfangswertes der Adjungierten . . .	176
6.11	L_{2,ω_β} -Fehler der Steuerung für $\delta = 0.5$ und $\beta = 4, 20, 50, 100$	186
6.12	L_{2,ω_β} -Fehler der Steuerung für $\delta = 3$ und $\beta = 4, 20, 50, 100$	187
6.13	Approximierter Zielfunktionswert $J(x_N, u_N)$ und Umschaltzeitpunkt \bar{t}_N für $\beta = 4.01, 4.41, 4.81$	199
6.14	Approximierter Zielfunktionswert $J(x_N, u_N)$ für $\beta = 2, 5$	201

1 Einleitung

Das Gebiet der optimalen Steuerung ist für zahlreiche Prozesse in den Natur-, Ingenieurs- und Wirtschaftswissenschaften von großem Interesse und hat sich durch zahlreiche Anwendungen zu einem wichtigen Gebiet der angewandten Mathematik etabliert. Beispiele für die vielfältige Anwendbarkeit von Optimalsteuerungsaufgaben sind in der Robotik, bei der Simulation von Testfahrten, bei der Flugbahn- und Straßenbahnoptimierung, bei der Modellierung biologischer und chemischer Prozesse und bei dynamischen Produktions- und Lagerhaltungsproblemen zu finden. In dieser Arbeit wird sich auf Optimalsteuerungsaufgaben mit einem a priori gegebenen unendlichem Zeithorizont konzentriert. Es gibt verschiedene Anwendungen, bei denen die Betrachtung eines unendlichen Integrationsintervalles erforderlich ist. So ist es zum Beispiel bei ökonomischen oder auch biologischen Systemen sinnvoll, einen unendlichen Planungshorizont zu wählen, da die derzeit getroffenen Entscheidungen nicht nur Auswirkungen mit kurzer Dauer implizieren, sondern oftmals auch Konsequenzen für zukünftige Generationen mit sich bringen (vgl. z.B. [AK70, Hau03]). An Beispielen des Drogenkonsums, der Korruption und der Terrorbekämpfung werden weitere interessante Anwendungen von Steuerungsproblemen über einem unendlichen Zeithorizont in [GCF⁺08] diskutiert. Dabei entsteht der unendliche Betrachtungshorizont in vielen Modellen als Grenzfall von Steuerungsaufgaben mit einem beliebig großen endlichen Zeithorizont $T > 0$. Es gibt jedoch auch Beispiele, bei denen der unendliche Zeithorizont auf andere Weise motiviert wird. Mit dem Ziel auftauchende Singularitäten im Zielfunktional zu behandeln, werden zum Beispiel in [Pic19] bewusst verschiedene Transformationen verwendet, um den unendlichen Zeithorizont zu erzeugen. An dem Beispiel des berühmten Variationsproblems der Brachistochrone wird demonstriert, wie eine geeignete Transformation auf ein Steuerungsproblem mit einem a priori gegebenen unendlichen Zeithorizont führt. Weiterhin gibt es viele Anwendungen, vor allem technischer und physikalischer Natur, welche Zielstellungen hervorbringen, bei denen sich ein unendliches Integrationsintervall in natürlicher Weise aus der Aufgabenstellung selbst ergibt. Als Beispiel sei die Bestimmung der Wellenfunktionen des quantenmechanischen harmonischen Oszillators angeführt, vgl. [PB18].

Der eindimensionale harmonische Oszillator ist eines der grundlegenden Modellsysteme der Quantenmechanik, mit welchem das Verhalten eines Teilchens in einem Potential der Form

$$V(x) = \frac{1}{2}kx^2 = \frac{1}{2}m\omega^2x^2$$

beschrieben wird. Dabei bezeichnen m die Teilchenmasse, $k > 0$ die Kraftkonstante und $\omega = \sqrt{k/m}$ die Kreisfrequenz des Oszillators. Der Hamilton-Operator beschreibt die Gesamtenergie eines quantenmechanischen Systems und enthält im Fall des harmonischen Oszillators einen Term für die kinetische Energie und einen Term, der dessen quadratisches

Potential beschreibt:

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2.$$

Dabei sind $\hat{p} = -i\hbar\frac{\partial}{\partial x}$ (\hbar ist das Plancksche Wirkungsquantum) und \hat{x} Operatoren zur Beschreibung des Impulses und des Ortes des Teilchens. Gesucht sind nun Funktionen ψ , die sogenannten Wellenfunktionen, mit denen sich das Verhalten des betrachteten physikalischen Systems beschreiben lässt. Um diese zu bestimmen, setzt man den Hamilton-Operator in die stationäre Schrödinger-Gleichung $\hat{H}\psi(x) = E\psi(x)$ ein und erhält im Fall des quantenmechanischen Oszillators die folgende lineare Differentialgleichung 2. Ordnung:

$$-\frac{\hbar^2}{2m}\psi''(x) + \frac{1}{2}m\omega^2x^2\psi(x) = E\psi(x). \quad (1.1)$$

Die Wellenfunktionen ergeben sich durch Lösen der Schrödinger-Gleichung, können aber auch mit Hilfe eines Variationsansatzes gewonnen werden. Dieser Variationszugang im Kontext eines komplexen Hilbertraumes geht auf Schrödinger selbst zurück, vgl. [LL79]. Im Fall des harmonischen Oszillators führt dieser Ansatz auf die Betrachtung der folgenden Variationsaufgabe

$$J(\psi) = \int_{-\infty}^{\infty} \left\{ \frac{\hbar^2\psi'(x)^2}{4m} + \frac{1}{4}(m\omega^2x^2 - E)\psi^2(x) \right\} dx \rightarrow \text{Stat!}$$

Zu bestimmen sind die stationären Zustände ψ und nicht nur die Minima des Zielfunktional. Wir stellen fest, dass die zugehörige Euler-Lagrange-Gleichung gerade mit der Schrödinger-Gleichung in (1.1) übereinstimmt. Unter der Annahme, dass die Funktionen ψ für $x \rightarrow \pm\infty$ gegen Null konvergieren, ergeben sich die Hermite-Funktionen als gesuchte Wellenfunktionen für den harmonischen Oszillator. An diesem Beispiel wird deutlich, dass die Verwendung eines endlichen Integrationsgebietes nicht in Frage kommt, da das unendliche Intervall in natürlicher Weise aus dem physikalischen Kontext der Aufgabe selbst entsteht. Insbesondere ergibt sich hierbei das unendliche Integrationsintervall aus der Annahme, dass das Universum als unendlich vorausgesetzt wird.

Ein weiteres Anwendungsgebiet für Optimalsteuerungsaufgaben, bei denen ein unendlicher Zeithorizont auftritt, sind Aufgaben der asymptotischen Stabilisierung dynamischer Systeme. Bereits in dem Jahr 1960 untersuchte Kalman den Zusammenhang zwischen der asymptotischen Steuerbarkeit eines dynamischen Systems und Aufgaben der optimalen Steuerung, vgl. [Kal60]. Dieser Grundgedanke wird in dieser Arbeit aufgegriffen und eine Philosophie präsentiert, bei der eine geeignete Wahl gewichteter Funktionenräume die asymptotische Stabilität des dynamischen Systems garantiert.

Die Bedeutung von Steuerungsaufgaben mit unendlichem Zeithorizont spiegelt sich in zahlreichen Veröffentlichungen und Arbeiten zu diesem Thema wider, vgl. beispielsweise [AC79, AK08, Ase17, AV15, Car95, CHL91, Hal74, LPW08, Lyk10, Lyk16, Mag80, Mic82, Pic09, Pic14, Pic15, PL17, Sei99, Tau10].

Den Schwerpunkt dieser Arbeit bilden linear-quadratische Optimalsteuerungsaufgaben mit

dem a priori gegebenen unendlichen Zeithorizont $[0, \infty)$. Diese Aufgaben repräsentieren eine wichtige Aufgabenklasse in der optimalen Steuerung, da sie einerseits für die Approximation nichtlinearer Aufgaben genutzt werden, andererseits aber auch bei dem klassischen Verfahren der sequentiellen quadratischen Optimierung (SQP-Verfahren, engl.: sequential quadratic programming) zum Einsatz kommen. Darüber hinaus spielen sie eine entscheidende Rolle im Rahmen der Stabilitätstheorie, bei der man durch Linearisierung um einen Gleichgewichtspunkt oder entlang einer Referenztrajektorie eine lineare Systemdynamik erhält. Das quadratische Zielfunktional entsteht dabei aus der Zielstellung, den Steuerungsaufwand und die Abweichung des Zustandes von der Referenztrajektorie zu minimieren.

Bei der Formulierung von Steuerungsproblemen über dem unendlichen Zeithorizont $[0, \infty)$ treten verschiedene Fragestellungen auf. Dazu gehören neben der Frage einer geeigneten Raumwahl auch die Frage nach der Wahl des Integraltypes im Zielfunktional, die Frage nach der Existenz und Eindeutigkeit einer Lösung und die Frage nach geltenden Optimalitätsbedingungen. Da das analytische Lösen von Problemen der optimalen Steuerung oftmals nicht zu bewältigen ist, studieren wir in dieser Arbeit ein indirektes Näherungsverfahren. Aufbauend auf den aus einem Pontryaginschen Maximumprinzip resultierenden Optimalitätsbedingungen leiten wir ein kanonisches Differentialgleichungssystem her, durch dessen Lösung wir die gesuchte optimale Steuerung mit Hilfe der Maximumbedingung berechnen können. Eine besondere Rolle nimmt dabei die ermittelte Transversalitätsbedingung für die adjungierte Funktion ein. Durch diese wird nicht nur die Raumzugehörigkeit der Adjungierten in einem gewichteten Sobolev-Raum bestimmt, sondern auch gleichzeitig ihr Anfangswert festgelegt. Die Idee dabei ist es, eine Kombination der kanonischen Differentialgleichung mit dem Hauptsatz der Differential- und Integralrechnung zu verwenden, um somit den Wert der Adjungierten in $t_0 = 0$ mit Hilfe einer Integralbedingung zu berechnen. Es resultiert ein Differentialgleichungssystem für den Systemzustand und die Adjungierte, wobei der Systemzustand in $t_0 = 0$ gegeben und der Anfangswert der Adjungierten mit Hilfe der Transversalitätsbedingung bestimmt werden kann. Das entstehende Anfangswertproblem lösen wir mit Hilfe einer Spektralmethode, da diese auf globalen Ansatzfunktionen basiert und daher gute Approximationen für den unendlichen Horizont liefert.

Neben Methoden aus der (Funktional-) Analysis, der Optimierung und der Numerik erfordert das numerische Verfahren eine theoretische Vorarbeit aus der Theorie der Optimalsteuerung über $[0, \infty)$. Daher werden in dem ersten Teil dieser Arbeit zunächst alle Grundlagen präsentiert, die für eine wohldefinierte Formulierung eines Optimalsteuerungsproblem mit einem unendlichen Zeithorizont benötigt werden. Im Anschluss wird die Existenz und Eindeutigkeit einer Lösung diskutiert und notwendige und hinreichende Optimalitätsbedingungen bewiesen. Aufbauend auf diesen theoretischen Vorbetrachtungen wird die Herleitung und numerische Umsetzung der von uns verwendeten indirekten Spektralmethode erläutert und ihr erfolgreicher Einsatz anhand verschiedener Beispiele demonstriert.

Die in dieser Arbeit vorgestellten Ergebnisse bauen auf unseren Forschungsarbeiten der letzten Jahre auf, wobei ein Teil dieser Arbeit bereits in den Publikationen [PB18, PB19, PBK16] veröffentlicht wurde. Unser Ziel ist es nun die einzelnen Ergebnisse gesamtheitlich zu betrachten, diese weiter auszubauen und Möglichkeiten für die Weiterentwicklung der bisherigen Resultate aufzuzeigen. Aus dem bisher Gesagten lassen sich die wesentlichen Beiträge dieser Arbeit wie folgt zusammenfassen:

- Der Zusammenhang zwischen dem Gebiet der Optimalsteuerungsaufgaben mit dem a priori gegebenen unendlichen Zeithorizont $[0, \infty)$ und der asymptotischen Steuerbarkeit dynamischer Systeme wird im Kontext gewichteter Sobolev-Räume beleuchtet.
- Es wird ein Maximumprinzip vom Pontryagin Typ für eine Klasse linear-quadratischer Optimalsteuerungsaufgaben über $[0, \infty)$ bewiesen. Die betrachteten Aufgaben beinhalten vektorwertige Zustände und Steuerungen in gewichteten Sobolev- und Lebesgue-Räumen und legen das Lebesgue-Integral zugrunde. Durch die Einbettung der Aufgabe in eine gewichtete Hilbertraumtheorie lässt sich diese mit Methoden der konvexen Analysis behandeln. Neben den kanonischen Gleichungen und der Maximumbedingung liefert das Maximumprinzip die Existenz einer Adjungierten, welche eine natürliche Transversalitätsbedingung erfüllt.
- Aufbauend auf den aus dem Maximumprinzip resultierenden Optimalitätskriterien wird eine indirekte Spektralmethode zur Lösung linear-quadratischer Aufgaben der optimalen Steuerung über $[0, \infty)$ vorgestellt. Im Unterschied zu denen in [FR08, GPF⁺11, GHR11] präsentierten Spektralverfahren für Steuerungsaufgaben mit einem unendlichen Zeithorizont wird in dieser Arbeit darauf verzichtet, das Intervall $[0, \infty)$ auf ein endliches Intervall zu transformieren. Insbesondere wird der Ansatz „First optimize, then discretize“ verfolgt, um die Optimalsteuerungsaufgabe durch Diskretisierung der Optimalitätsbedingungen in ein lineares Gleichungssystem zu überführen. Ein wesentlicher Bestandteil dabei ist die Wahl globaler Approximationsansätze und das Ausnutzen einer integrierten Transversalitätsbedingung, um den Anfangswert der Adjungierten numerisch zu bestimmen.
- Mit Resultaten zur punktweisen Konvergenz von Laguerre-Fourier-Reihen in gewichteten Sobolev-Räumen wird ein Beitrag zu bisherigen in der Literatur vorhandenen Konvergenzresultaten geleistet.
- Numerische Experimente demonstrieren die Anwendbarkeit der vorgestellten numerischen Lösungsmethode. Neben Aufgaben der optimalen Steuerung wird auch die asymptotische Stabilisierbarkeit eines linearisierten dynamischen Systems behandelt.

Im Detail ist die Arbeit wie folgt aufgebaut. In dem Kapitel 2 konzentrieren wir uns im Wesentlichen darauf, die Grundlagen aus der Theorie der Lebesgue- und Sobolev-Räume zusammenzutragen. Zunächst führen wir die klassischen Lebesgue- und Sobolev-Räume ein, um auf diese aufbauend die gewichteten Lebesgue- und Sobolev-Räume zu definieren und deren wichtigsten Eigenschaften zu untersuchen. Mit einem Einblick in das Gebiet der Stabilisierung und Steuerbarkeit dynamischer Systeme zeigen wir eine Anwendung linear-quadratischer Steuerungsaufgaben über $[0, \infty)$ in gewichteten Funktionenräumen auf und schließen damit das zweite Kapitel ab. In dem Kapitel 3 werden zentrale Resultate der Approximationstheorie in Hilberträumen vorgestellt, Begriffe wie Orthogonalprojektion und Fourier-Reihe definiert und speziell die polynomiale Approximation in dem gewichteten Lebesgue-Raum $L_2((a, b), \omega)$ diskutiert. Gegenstand des vierten Kapitels ist die Untersuchung linear-quadratischer Optimalsteuerungsaufgaben über dem gegebenen unendlichen Zeithorizont $[0, \infty)$. Die Formulierung dieser Aufgabenklasse erfolgt in gewichteten Funktionenräumen und bietet daher die Möglichkeit, diese Aufgaben in Hilberträumen zu studieren und aufgrund ihrer linear-quadratischen Struktur zusätzlich Methoden der konvexen Analysis heranzuziehen. Nachdem

grundlegende Voraussetzungen an diese Aufgaben vorgestellt und Differentialgleichungen in gewichteten Sobolev-Räumen studiert werden, werden die Existenz und Eindeutigkeit der Lösung, sowie notwendige und hinreichende Optimalitätskriterien bewiesen. Hauptgegenstand ist die Formulierung eines Maximumprinzips vom Pontryagin Typ. Dieses garantiert neben der Gültigkeit der kanonischen Gleichungen und der Maximumbedingung auch die Existenz einer Adjungierten, welche einer natürlichen Transversalitätsbedingung genügt. Die wesentliche Beweistechnik des Maximumprinzips basiert auf der Trennung konvexer Mengen. Schlussendlich legt dieses Resultat den Grundstein für die im nächsten Kapitel vorgestellte indirekte Spektralmethode. Zu Beginn des fünften Kapitels geben wir eine Übersicht zu verschiedenen numerischen Lösungsverfahren für Optimalsteuerungsaufgaben. Es werden die Unterschiede zwischen direkten und indirekten Verfahren erläutert und der Grundgedanke von Spektralmethoden beschrieben. Danach lässt sich das fünfte Kapitel in zwei Teile gliedern. In dem ersten Teil widmen wir uns der Untersuchung der verallgemeinerten Laguerre-Polynome. Diese bilden ein vollständiges Orthogonalsystem in dem gewichteten Lebesgue-Raum $L_2((0, \infty), \omega_{\alpha, \beta})$ und werden von uns für die Approximation der gesuchten Funktionen verwendet. Neben grundlegenden Eigenschaften dieses Polynomsystems präsentieren wir neue Ergebnisse zur punktweisen Konvergenz von Laguerre-Fourier-Reihen in gewichteten Sobolev-Räumen. Der zweite Teil beinhaltet die Formulierung der indirekten Spektralmethode zur numerischen Lösung der linear-quadratischen Optimalsteuerungsaufgabe über dem unendlichen Zeithorizont $[0, \infty)$. Durch die Auswertung der notwendigen Optimalitätsbedingungen überführen wir das Lösen der Steuerungsaufgabe auf das Lösen eines Differentialgleichungssystems, welches wir mit Hilfe der globalen Polynomansätze und eines Kollokationsansatzes approximieren. Die gesuchten Funktionen sind dabei der Zustand und die Adjungierte; die optimale Steuerung berechnen wir im Anschluss durch direktes Auswerten der Maximumbedingung. Ein besonderes Augenmerk liegt auf der Herleitung einer Integralbedingung, mit welcher der Anfangswert der Adjungierten numerisch bestimmt werden kann. Die Möglichkeit dazu bietet die aus dem Pontryaginschen Maximumprinzip resultierende Transversalitätsbedingung. Letztlich führt die indirekte Diskretisierung der Optimalsteuerungsaufgabe auf ein lineares Gleichungssystem, dessen Koeffizientenmatrix wir in dem letzten Abschnitt dieses Kapitels vorstellen. Die erfolgreiche Anwendbarkeit der vorgestellten indirekten Lösungsmethode wird in dem Kapitel 6 an ausgewählten Beispielen demonstriert. Neben Beispielen der optimalen Steuerung wird auch der Einsatz der Methode zur asymptotischen Stabilisierung eines linearisierten dynamischen Systems vorgestellt. Eine Zusammenfassung der Ergebnisse dieser Arbeit und ein Ausblick auf zukünftige Fragestellungen bilden den Abschluss dieser Arbeit.

2 Räume und benötigte Grundlagen

Für das Verständnis dieser Arbeit setzen wir voraus, dass der Leser mit den Grundlagen der Analysis, der linearen Algebra und der numerischen Mathematik vertraut ist. Für eine geschlossene Darstellung werden wir die wichtigsten Begriffsbildungen in Erinnerung rufen und entsprechende Zusammenhänge aufzeigen. Dabei legen wir insbesondere den Schwerpunkt auf eine für die nachfolgenden Kapitel vollständige Zusammenstellung wichtiger Notationen, Definitionen und benötigter Hilfsmittel. Wenn für das weitere Verständnis notwendig, skizzieren wir Beweisideen, verzichten aber ansonsten auf eine umfassende Angabe von Beweisen in diesem Kapitel und verweisen den interessierten Leser für eine umfassende Einführung an die jeweils angegebene Literatur.

Die Inhalte dieses Kapitels sind in Zusammenarbeit mit Torsten Ziemann entstanden und wurden während der Entstehung dieser Arbeit bereits unter dem Titel „Analytische Hilfsmittel der Steuerungstheorie“ veröffentlicht, vgl. [BZ18].

2.1 Notationen

Ziel dieses Abschnittes ist es, die von uns verwendeten Bezeichnungen und Begriffe einzuführen und kurz zu erläutern. Ausführlich sind die hier angesprochenen Inhalte in [AF03] zu finden.

Wir beginnen damit, erste hier verwendete Notationen vorzustellen:

- Die Menge der natürlichen Zahlen \mathbb{N} einschließlich der Null bezeichnen wir mit $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.
- Wenn nicht ausdrücklich anders erwähnt, sei im Folgenden stets $n, m \in \mathbb{N}$.
- Das Symbol $\mathbb{R}_+ := [0, \infty)$ verwenden wir für die Menge der positiven reellen Zahlen einschließlich der Null.
- Den Betrag einer Zahl $v \in \mathbb{R}$ bezeichnen wir mit $|v|$.
- Für $1 \leq p < \infty$ bezeichnen wir mit

$$|v|_p := \left(\sum_{i=1}^n |v_i|^p \right)^{1/p}$$

die p - Norm eines Vektors $v = (v_1, v_2, \dots, v_n)^T \in \mathbb{R}^n$. Ist $p = 2$ so erhalten wir die euklidische Norm und schreiben

$$|v|_2 := \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}.$$

- Für eine Matrix $A \in \mathbb{R}^{n \times m}$ bezeichnen wir die von der euklidischen Vektornorm induzierte Matrixnorm entsprechend auch mit $|A|_2$.

Des Weiteren möchten wir an Definitionen spezieller Funktionenräume erinnern. Dabei werden wir durchweg nur reelle Räume betrachten. Weiterhin verwenden wir

- das Symbol Ω oder die Bezeichnung *Gebiet* und meinen damit eine nichtleere offene Menge des s -dimensionalen euklidischen Vektorraumes \mathbb{R}^s ($s \in \mathbb{N}$)

und bezeichnen mit

- $\bar{\Omega}$ den Abschluss von Ω und mit $\partial\Omega = \bar{\Omega} \setminus \Omega$ den Rand von Ω .

Wir nutzen insbesondere die folgenden Räume stetiger Funktionen:

- Für $k \in \mathbb{N}_0$ bezeichnen wir mit $C^k(\Omega, \mathbb{R}^n)$ die Menge aller Funktionen $x : \Omega \rightarrow \mathbb{R}^n$, die zusammen mit ihren partiellen Ableitungen bis zur Ordnung k stetig auf Ω sind. Für $C^0(\Omega, \mathbb{R}^n)$ schreiben wir $C(\Omega, \mathbb{R}^n)$.
- $C^\infty(\Omega, \mathbb{R}^n)$ wird als $\bigcap_{k=0}^\infty C^k(\Omega, \mathbb{R}^n)$ definiert.
- Im Fall $n = 1$ setzen wir $C^k(\Omega) := C^k(\Omega, \mathbb{R})$, $C(\Omega) := C(\Omega, \mathbb{R})$ und korrespondierend $C^\infty(\Omega) := C^\infty(\Omega, \mathbb{R})$.

Für die Einführung der Sobolev-Räume benötigen wir das Konzept der schwachen Ableitungen. Dieses und die dafür notwendigen Begriffe stellen wir nun kurz zusammen.

Sei $x : \Omega \rightarrow \mathbb{R}^n$ eine Funktion, die *fast überall* (kurz: f.ü.) auf Ω definiert ist. Dann bezeichnen wir mit $\text{supp}(x)$ den *Träger* der Funktion x , welcher als Abschluss (bezüglich der Norm des \mathbb{R}^s) der Nichtnullstellenmenge von x definiert ist:

$$\text{supp}(x) := \overline{\{t \in \Omega : x(t) \neq 0\}}.$$

Die Menge der *Testfunktionen* beinhaltet all die Funktionen $x \in C^\infty(\Omega, \mathbb{R}^n)$, welche einen kompakten Träger in Ω besitzen:

$$C_0^\infty(\Omega, \mathbb{R}^n) := \{x \in C^\infty(\Omega, \mathbb{R}^n) : \text{supp}(x) \text{ ist kompakte Teilmenge von } \Omega\}.$$

Insbesondere gibt es demnach zu jeder Testfunktion $x \in C_0^\infty(\Omega, \mathbb{R}^n)$ ein von dieser abhängiges, also individuell verschiedenes Gebiet Ω_0 mit $\bar{\Omega}_0 \subset \Omega$, sodass x in einem Randstreifen von Ω verschwindet, also $x \equiv 0$ außerhalb von $\bar{\Omega}_0$ gilt. Per Definition ist jede Testfunktion beliebig oft differenzierbar und offensichtlich ist jede Ableitung selbst wieder eine Testfunktion. Daher haben Testfunktionen die Eigenschaft, dass sowohl sie selbst als auch alle vorhandenen Ableitungen die Randwerte Null haben. Auch hier schreiben wir in dem Fall $n = 1$ wieder $C_0^\infty(\Omega) := C_0^\infty(\Omega, \mathbb{R})$.

Weiter nennen wir eine Funktion $x : \Omega \rightarrow \mathbb{R}$ *lokal auf Ω Lebesgue-integrierbar*, falls sie auf *jeder* offenen Menge U mit den Eigenschaften

- (i) $\bar{U} \subset \Omega$ und
- (ii) \bar{U} ist eine kompakte Teilmenge des \mathbb{R}^s

Lebesgue-integrierbar ist. Die Menge dieser Funktionen bezeichnen wir mit $L_{1,\text{loc}}(\Omega)$. Für vektorwertige Funktionen $x = (x_1, \dots, x_n) : \Omega \rightarrow \mathbb{R}^n$ definieren wir

$$L_{1,\text{loc}}^n(\Omega) := L_{1,\text{loc}}(\Omega, \mathbb{R}^n) := \{x : \Omega \rightarrow \mathbb{R}^n \mid x_i \in L_{1,\text{loc}}(\Omega) \text{ für alle } i = 1, \dots, n\}.$$

Im Weiteren ist $\alpha \in \mathbb{N}_0^s$ ein *Multiindex*, d.h. ein Tupel $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s)$ von Indizes $\alpha_i \in \mathbb{N}_0$, der Ordnung $|\alpha| := \sum_{i=1}^s \alpha_i$. Für eine bessere Übersicht verwenden wir kurz das Symbol \mathcal{D}^α und meinen damit

$$\mathcal{D}_1^{\alpha_1} \dots \mathcal{D}_s^{\alpha_s} := \frac{\partial^{|\alpha|}}{\partial t_1^{\alpha_1} \dots \partial t_s^{\alpha_s}}.$$

Die Komponenten α_i des Multiindex geben also an, wie oft nach t_i differenziert wird. Insbesondere wird $D^{(0,\dots,0)}x := x$ gesetzt.

An dieser Stelle können wir nun den Begriff der schwachen Ableitung aufgreifen. Sei dafür $x \in L_{1,\text{loc}}(\Omega)$. Existiert eine Funktion $y \in L_{1,\text{loc}}(\Omega)$, sodass

$$\int_{\Omega} x(t) \mathcal{D}^\alpha \phi(t) dt = (-1)^{|\alpha|} \int_{\Omega} y(t) \phi(t) dt, \quad t \in \Omega$$

für alle Testfunktionen $\phi \in C_0^\infty(\Omega)$ erfüllt ist, dann heißt y die *schwache partielle Ableitung* von x und wird ebenfalls mit $\mathcal{D}^\alpha x = y$ bezeichnet.

Bemerkung 2.1.

1. Im Falle der Existenz ist die schwache Ableitung (bis auf eine Menge vom Maß Null) eindeutig bestimmt.
2. Ist eine Funktion im klassischen Sinne differenzierbar und ist die Ableitung Element des Raumes $L_{1,\text{loc}}(\Omega)$, existiert die schwache Ableitung und stimmt fast überall mit der klassischen Ableitung überein.
3. Nicht jede Funktion aus dem Raum $L_{1,\text{loc}}(\Omega)$ muss eine schwache Ableitung im $L_{1,\text{loc}}(\Omega)$ besitzen. Andererseits kann eine schwache Ableitung aber auch eine „bessere“ Regularität besitzen, also auch „besseren“ Räumen als nur dem $L_{1,\text{loc}}(\Omega)$ angehören.

Die Notation $AC([a, b])$ verwenden wir für den Raum der absolut stetigen¹ Funktionen $x : [a, b] \rightarrow \mathbb{R}$ mit $-\infty < a < b < \infty$. Für möglicherweise unbeschränkte Intervalle $I \subseteq \mathbb{R}$ definieren wir die Gesamtheit derjenigen Funktionen $x : I \rightarrow \mathbb{R}$, für welche die Einschränkungen $x|_{[a,b]}$ für alle $[a, b] \subset I$ Elemente des Raumes $AC([a, b])$ sind, als den Raum der lokal

¹ Eine Funktion $x : [a, b] \rightarrow \mathbb{R}$ mit $-\infty < a < b < \infty$ heißt absolut stetig, falls für jedes $\varepsilon > 0$ ein $\delta > 0$ existiert, sodass für alle $k \in \mathbb{N}$ und alle Folgen von Intervallen $\{(\alpha_i, \beta_i)\}_{i=1,\dots,k}$ mit $a \leq \alpha_1 < \beta_1 \leq \alpha_2 < \beta_2 \leq \dots \leq \alpha_k < \beta_k \leq b$ und $\sum_{i=1}^k (\beta_i - \alpha_i) < \delta$

$$\sum_{i=1}^k |x(\beta_i) - x(\alpha_i)| < \varepsilon$$

gilt, vgl. [Els11], S. 303, 4.11 Definition (G. VITALI 1905).

absolut stetigen Funktionen und versehen ihn mit der Bezeichnung $AC_{\text{loc}}(I)$, vgl. [Els11], S. 307, 4.18 Definition. Für Funktionen $x = (x_1, \dots, x_n): I \rightarrow \mathbb{R}^n$ definieren wir

$$AC^n([a, b]) := AC([a, b], \mathbb{R}^n) := \{x: I \rightarrow \mathbb{R}^n \mid x_i \in AC([a, b]) \text{ für alle } i = 1, \dots, n\}$$

und

$$AC_{\text{loc}}^n(I) := AC_{\text{loc}}(I, \mathbb{R}^n) := \{x: I \rightarrow \mathbb{R}^n \mid x_i \in AC_{\text{loc}}(I) \text{ für alle } i = 1, \dots, n\}.$$

2.2 Klassische Lebesgue- und Sobolev-Räume

2.2.1 Die Räume $L_p(\Omega)$ und $W_p^k(\Omega)$

Wir nennen $\mathcal{M}(\Omega)$ die Menge aller Lebesgue-messbaren Funktionen $x: \Omega \rightarrow \mathbb{R}$ und führen zunächst die Lebesgue-Räume ein.

Definition 2.1 ([AF03], S. 23 f., 2.1)

Sei $1 \leq p < \infty$. Dann bezeichnen wir mit $L_p(\Omega)$ den Raum der messbaren und im Sinne von Lebesgue p -fach integrierbaren Funktionen:

$$L_p(\Omega) := \left\{ x \in \mathcal{M}(\Omega) \mid \int_{\Omega} |x(t)|^p dt < \infty \right\}$$

Durch

$$\|x\|_p := \left(\int_{\Omega} |x(t)|^p dt \right)^{1/p}$$

wird eine Norm auf dem $L_p(\Omega)$ definiert.

Wir möchten an dieser Stelle darauf hinweisen, dass wir Funktionen des $L_p(\Omega)$ miteinander identifizieren, wenn sie fast überall im Sinne des Lebesgue-Maßes in Ω übereinstimmen. Genauer gesagt, ist der $L_p(\Omega)$ also kein Raum von Funktionen, sondern von Äquivalenzklassen von Funktionen. Wir haben dieses Wissen vor Augen und schreiben für eine bessere Lesbarkeit kurz $x \in L_p(\Omega)$ statt $[x] \in L_p(\Omega)$ und $x = 0$ in $L_p(\Omega)$, falls $x(t) = 0$ für fast alle $t \in \Omega$ erfüllt ist.

Definition 2.2 ([AF03], S. 27, 2.10)

Wir nennen eine Funktion $x \in \mathcal{M}(\Omega)$ *wesentlich beschränkt* auf Ω , falls eine Konstante K existiert, sodass $|x(t)| \leq K$ f.ü auf Ω gilt. Ferner wird die kleinste obere Schranke dieser Konstanten K als das wesentliche Supremum von $|x|$ auf Ω bezeichnet. Damit definieren wir den Raum der auf Ω wesentlich beschränkten Funktionen

$$L_{\infty}(\Omega) := \left\{ x \in \mathcal{M}(\Omega) \mid \text{ess sup}_{t \in \Omega} |x(t)| < \infty \right\}$$

und führen eine Norm gemäß der Vorschrift

$$\|x\|_\infty := \operatorname{ess\,sup}_{t \in \Omega} |x(t)|$$

ein.

Wir zitieren [AF03], Kapitel 2, Seite 23-49 und fassen die wichtigsten Eigenschaften der vorgestellten Lebesgue-Räume zusammen:

- Für $1 \leq p \leq \infty$ und mit der Norm $\|\cdot\|_p$ ist der Lebesgue-Raum $L_p(\Omega)$ vollständig, also ein Banachraum.²
- Für $1 \leq p < \infty$ ist der Lebesgue-Raum $L_p(\Omega)$ ein separabler Banachraum.³
- Der Raum $L_p(\Omega)$ ist für $1 < p < \infty$ reflexiv.⁴
- Mit dem Skalarprodukt

$$\langle x, y \rangle_2 := \int_{\Omega} x(t)y(t)dt \tag{2.1}$$

wird der $L_2(\Omega)$ zu einem separablen Hilbertraum.^{5,6}

Die Erweiterung des „klassischen“ Ableitungsbegriffs auf die schwache Ableitung führt auf die Definition des Sobolev-Raumes $W_p^k(\Omega)$, welcher all die Funktionen $x \in L_p(\Omega)$ beinhaltet, die für jeden Multiindex α , $|\alpha| \leq k$ eine schwache partielle Ableitung $\mathcal{D}^\alpha x$ bis zur Ordnung k besitzen und diese in dem Lebesgue-Raum $L_p(\Omega)$ liegen. An dieser Stelle möchten wir auf die für jedes Gebiet Ω und mit $1 \leq p \leq \infty$ geltende Inklusion

$$L_p(\Omega) \subset L_{1,\text{loc}}(\Omega) \tag{2.2}$$

hinweisen, vergleiche [AF03], S. 29, 2.15 Corollary. Damit sind die schwachen partiellen Ableitungen $\mathcal{D}^\alpha x$ auch für Funktionen $x \in L_p(\Omega)$ wohldefiniert.

² Ein normierter Raum X heißt vollständig bzw. wird als Banachraum bezeichnet, wenn jede Cauchyfolge aus X in X konvergiert.

³ Ein normierter Raum X ist separabel, wenn es eine abzählbare Teilmenge gibt, die dicht in X liegt. Eine Teilmenge M des normierten Raumes X liegt dicht in X , falls jedes Element $x \in X$ der Grenzwert einer Folge von Elementen aus M ist.

⁴ Der Raum X^* bestehend aus den linearen stetigen Funktionalen $x^*: X \rightarrow \mathbb{K}$ heißt Dualraum des normierten Raumes X . Mit $X^{**} := (X^*)^*$ bezeichnen wir den Bidualraum von X (Dualraum zu X^*). Wir definieren für $x \in X$ die Abbildung $i_x: X^* \rightarrow \mathbb{K}$ vermöge $i_x(x^*) = x^*(x)$. Jedes $x \in X$ definiert durch $i_x(x^*) = x^*(x)$ ein Element i_x des Bidualraumes X^{**} . Die Abbildung $i_X: X \rightarrow X^{**}$, $x \mapsto i_x$ ist linear, isometrisch und injektiv. Die Abbildung i_X heißt kanonische Abbildung von X in X^{**} . Ist i_X zusätzlich surjektiv, so bezeichnen wir den Raum X als reflexiv, vgl. [Wer18], S. 105, Satz III.3.1, Definition III.3.3.

⁵ Ein normierter Raum $(X, \|\cdot\|)$ heißt Prähilbertraum, wenn es ein auf $X \times X$ definiertes Skalarprodukt $\langle \cdot, \cdot \rangle$ mit $\langle x, x \rangle^{1/2} = \|x\|$ für alle $x \in X$ gibt. Ist X zusätzlich vollständig bzgl. der induzierten Norm, so bezeichnen wir X als Hilbertraum, vgl. [Wer18], S. 221, Definition V.1.4.

⁶ Jeder Hilbertraum ist reflexiv, vgl. [Alt12] S. 247, 6.11 Beispiele zur Reflexivität, (1).

Definition 2.3 ([AF03], S. 59 f.)

Für $1 \leq p < \infty$ und $k \in \mathbb{N}_0$ definieren wir den Sobolev-Raum

$$W_p^k(\Omega) := \{x \in L_p(\Omega) \mid \mathcal{D}^\alpha x \in L_p(\Omega) \text{ für alle } \alpha \in \mathbb{N}_0^s \text{ mit } |\alpha| \leq k\}$$

und führen auf ihm die Norm

$$\|x\|_{k,p} := \left(\sum_{0 \leq |\alpha| \leq k} \|\mathcal{D}^\alpha x\|_p^p \right)^{1/p}$$

ein.

Offensichtlich ist $W_p^0(\Omega) = L_p(\Omega)$.

In Erweiterung der Definition 2.3 vereinbaren wir

$$W_{1,\text{loc}}^k(\Omega) := \{x \in L_{1,\text{loc}}(\Omega) \mid \mathcal{D}^\alpha x \in L_{1,\text{loc}}(\Omega) \text{ für alle } \alpha \in \mathbb{N}_0^s \text{ mit } |\alpha| \leq k\} ,$$

sodass aus der Inklusion in (2.2) sofort auch $W_p^k(\Omega) \subset W_{1,\text{loc}}^k(\Omega)$ folgt.

Beschränken wir unsere Betrachtungen auf ein möglicherweise unbeschränktes Intervall und setzen $\Omega := I \subseteq \mathbb{R}$, so liefert der nachfolgende Satz einen nützlichen Zusammenhang zwischen der schwachen Ableitung einer Funktion $x \in W_{1,\text{loc}}^1(I)$ und der punktweise fast überall existierenden 'klassischen' Ableitung ihres Repräsentanten $\tilde{x} \in AC_{\text{loc}}(I)$, bezeichnet mit \tilde{x}' .

Satz 2.4 ([KJF77], S. 274, 5.6.3 Theorem)

Sei $I \subseteq \mathbb{R}$ und $x \in W_{1,\text{loc}}^1(I)$. Dann existiert eine Funktion $\tilde{x} \in AC_{\text{loc}}(I)$, sodass $x = \tilde{x}$ und $Dx = \tilde{x}'$ fast überall auf I gelten.⁷

Demnach können wir also zu jedem Element $x \in W_{1,\text{loc}}^1(I)$ einen lokal absolut stetigen Repräsentanten \tilde{x} mit $x = \tilde{x}$ f.ü. auf I finden. Da $\tilde{x} \in AC_{\text{loc}}(I)$ ist, existiert die 'klassische' Ableitung des Repräsentanten \tilde{x}' fast überall auf I und gemäß des vorangegangenen Satzes stimmt diese mit der schwachen Ableitung von x fast überall auf I überein. Im Zusammenhang mit der Analyse von Differentialgleichungen in Sobolev-Räumen werden wir auf dieses Ergebnis an geeigneter Stelle zurückgreifen.

Häufig findet man in der Literatur auch die Bezeichnung $H_p^k(\Omega)$, wobei dieser Zugang zu den Sobolev-Räumen auf einem Vervollständigkeitsprozess basiert, vgl. [MS64].

Definition 2.5

Sei $1 \leq p < \infty$ und $k \in \mathbb{N}_0$. Der Raum $H_p^k(\Omega)$ ist der Abschluss von $C^\infty(\Omega) \cap W_p^k(\Omega)$ in dem Raum $W_p^k(\Omega)$.

Wir sehen, dass für die Einführung des $H_p^k(\Omega)$ -Raumes der Durchschnitt $C^\infty(\Omega) \cap W_p^k(\Omega)$ herangezogen wird. Offensichtlich ergibt dieser Durchschnitt gerade den Raum all derjenigen Funktionen $x \in C^\infty(\Omega)$, für die $\|x\|_{k,p} < \infty$ gilt. Dieser ist ein normierter linearer Raum,

⁷ Für den Fall eines Gebietes $\Omega \subseteq \mathbb{R}^s$ verweisen wir auf ein ähnliches Resultat in [KJF77], S. 273 f.

jedoch ist er nicht vollständig. Vielmehr führt dessen Vervollständigung dann auf die Definition des $H_p^k(\Omega)$. Bis zur Veröffentlichung der Arbeit von Meyers und Serrin mit dem Titel „ $H = W$ “ wurden die Räume $W_p^k(\Omega)$ und $H_p^k(\Omega)$ im Sinne der hier eingeführten Definitionen strikt voneinander getrennt und nur unter zusätzlichen Glattheitsbedingungen an den Rand des Gebietes Ω miteinander identifiziert. Im Jahre 1964 konnten Meyers und Serrin beweisen, dass die Differenzierung zwischen den Räumen unnötig ist - und zwar unabhängig von der Glätte des Randes $\partial\Omega$ - und demnach beide Zugänge der Sobolev-Räume für jedes Gebiet Ω äquivalent sind.

Satz 2.6 (Satz von Meyers und Serrin [MS64])

Für $1 \leq p < \infty$ gilt

$$W_p^k(\Omega) \equiv H_p^k(\Omega).$$

Insbesondere besagt dieser Satz, dass die unendlich oft differenzierbaren Funktionen in Sobolev-Räumen dicht liegen, der Raum $W_p^k(\Omega)$, $1 \leq p < \infty$ also aus denjenigen Funktionen besteht, welche sich in der Sobolev-Norm durch Funktionen in $C^\infty(\Omega)$ approximieren lassen.

Im Hinblick auf die eingangs motivierte Optimalsteuerungsaufgabe, werden wir uns damit beschäftigen, Anfangswertprobleme für gewöhnliche Differentialgleichungen in Sobolev-Räumen zu studieren. Beachten wir jedoch, dass in einem Sobolev-Raum alle Funktionen, die fast überall gleich sind, miteinander identifiziert werden und in einer Äquivalenzklasse liegen, so können diese natürlich nicht mehr punktweise ausgewertet werden. Darüber hinaus stellt sich sofort die Frage nach der Definition von Rand- bzw. Anfangswerten für Sobolev-Funktionen. Für eine stetige Funktion $x \in C(\bar{\Omega})$ sind die Randwerte durch eine stetige Fortsetzung auf $\partial\Omega$ festgelegt. Sobolev-Funktionen $x \in W_p^k(\Omega)$ sind im Allgemeinen nicht stetig und vor allem nur fast überall auf Ω eindeutig definiert. Da der Rand $\partial\Omega$ eine Teilmenge des \mathbb{R}^s vom Lebesgue-Maß Null ist, darf die Sobolev-Funktion x dort beliebig abgeändert werden, ohne dass sie sich im Sinne des Raumes $L_p(\Omega)$ ändert. Es ergibt sich demnach die Frage, welche sinnvolle Bedeutung wir einem Randwert für eine Äquivalenzklasse $[x] \in W_p^k(\Omega)$ zukommen lassen. Wir werden sehen, dass unter zusätzlichen Regularitätsvoraussetzungen an das Gebiet Ω , diese Frage mit Hilfe der Existenz eines stetigen Vertreters der Äquivalenzklasse sinnvoll beantwortet werden kann.

Definition 2.7 ([AF03], S. 83, 4.9)

Ein beschränktes Gebiet Ω erfüllt die stark lokale Lipschitz-Bedingung, wenn für jeden Punkt $x \in \partial\Omega$ eine Umgebung U_x existiert, sodass $U_x \cap \partial\Omega$ der Graph einer Lipschitz-stetigen Funktion ist.

Satz 2.8 ([AF03], S. 168, 6.3 Theorem von Rellich-Kondrachov, Part III)

Sei Ω ein Gebiet, welches die stark lokale Lipschitz-Bedingung erfüllt und $\Omega_0 \subseteq \Omega$ ein beschränktes Teilgebiet. Seien weiter $j \geq 0$ und $k \geq 1$ ganzzahlig und $1 < p < \infty$. Dann ist die Einbettung

$$W_p^{j+k}(\Omega) \hookrightarrow C^j(\bar{\Omega}_0)$$

kompakt, falls $kp > k$ gilt.

Bemerkung 2.2 ([AF03], S. 169, 6.4 Remarks). *Ist das Gebiet Ω in Satz 2.8 beschränkt, so können wir ohne Einschränkung $\Omega_0 = \Omega$ setzen.*

Wir konzentrieren uns jetzt auf Intervalle und wählen $\Omega = (a, b)$ mit $-\infty < a < b < \infty$. In diesem Fall ist die stark lokale Lipschitz-Bedingung erfüllt und mit $\Omega = \Omega_0$ können wir aus dem Satz 2.8 direkt schließen, dass die Einbettung von dem Raum der Sobolev-Funktionen in den Raum der stetigen Funktionen kompakt ist.

Folgerung 2.1. *Es sei $k \geq 1$ ganzzahlig und $1 < p < \infty$. Dann ist die Einbettung*

$$W_p^k((a, b)) \hookrightarrow C^0([a, b])$$

kompakt, falls $kp > k$ erfüllt ist.

Sei $1 < p < \infty$. Betrachten wir ein $x \in W_p^k((a, b))$, $k \geq 1$ als Funktion, so können wir gemäß des vorangegangenen Satzes stets genau eine Funktion $x^* \in C^0([a, b])$ definieren, die in der gleichen Äquivalenzklasse $[x] \in W_p^k((a, b))$ liegt. Mit anderen Worten können wir für jede Äquivalenzklasse in $W_p^k((a, b))$ einen eindeutig bestimmten stetigen Repräsentanten in $C^0([a, b])$ finden. Durch die Identifikation einer Sobolev-Funktion $x \in W_p^k((a, b))$, $k \geq 1$ mit ihren Repräsentanten $x^* \in C^0([a, b])$ wird eine punktweise Auswertung und insbesondere auch eine sinnvolle Definition von Randwerten möglich.

Bemerkung 2.3. *Setzt man $k = 1$ in dem Satz 2.8, so folgt für $p > 1$ und für alle $j \geq 0$ die Kompaktheit der Einbettung*

$$W_p^{j+1}((a, b)) \hookrightarrow C^j([a, b]).$$

Dies impliziert, dass für eine Sobolev-Funktion $x \in W_p^k((a, b))$ (mit $k \geq 1$ und $p > 1$) sogar ein $(k - 1)$ -mal stetig differenzierbarer Repräsentant $x^ \in C^{k-1}([a, b])$ existiert. Dadurch wird zusätzlich die punktweise Auswertung der Ableitungen bis zur Ordnung $k - 1$ legitimiert und insbesondere können die Randwerte sowohl für x selbst, als auch für $x', x'', \dots, x^{(k-1)}$ eindeutig festgelegt werden.*

Wir fassen abschließend weitere wichtige Eigenschaften der Sobolev-Räume zusammen. Für die entsprechenden Beweise verweisen wir den interessierten Leser auf [AF03], S. 60 f. und [Eva98], Theorem 2, S. 286.

- Für $1 \leq p < \infty$ und versehen mit der Norm $\|\cdot\|_{k,p}$ ist der Raum $W_p^k(\Omega)$ ein separabler Banachraum.
- Der Raum $W_p^k(\Omega)$ ist für $1 < p < \infty$ reflexiv.
- Insbesondere ist der Raum $W_2^k(\Omega)$ ein separabler Hilbertraum mit dem Skalarprodukt

$$\langle x, y \rangle_{k,2} := \sum_{0 \leq |\alpha| \leq k} \langle \mathcal{D}^\alpha x, \mathcal{D}^\alpha y \rangle_2,$$

wobei $\langle \cdot, \cdot \rangle_2$ das in (2.1) definierte Skalarprodukt im $L_2(\Omega)$ meint.

- Für $x \in W_p^1((0, T))$, $0 < T < \infty$ und $1 \leq p \leq \infty$ existiert ein Repräsentant aus $C^0([0, T])$ und es gilt die Darstellung

$$x(t) = x(t_0) + \int_{t_0}^t x'(\tau) d\tau \quad \forall 0 \leq t_0 \leq t \leq T.$$

2.2.2 Die Räume $L_p^n(\Omega)$, $W_p^{k,n}(\Omega)$ und der Raum $L_\infty^{n \times m}(\Omega)$

In der vorliegenden Arbeit werden wir vorrangig mit Funktionenräumen arbeiten, bei denen die Elemente durch vektorwertige beziehungsweise matrixwertige Funktionen gegeben sind.

Definition 2.9

Eine Funktion $x: \Omega \rightarrow \mathbb{R}^n$, $x = (x_1, \dots, x_n)$ mit Komponentenfunktionen $x_i: \Omega \rightarrow \mathbb{R}$, $i = 1, \dots, n$ nennen wir Vektorfunktion (vektorwertige Funktion).

Definition 2.10

Eine Funktion $M: \Omega \rightarrow \mathbb{R}^{n \times m}$ mit Komponentenfunktionen $M_{ij}: \Omega \rightarrow \mathbb{R}$, $i = 1, \dots, n$, $j = 1, \dots, m$ nennen wir Matrixfunktion (matrixwertige Funktion).

Mit $\mathcal{M}^{n \times m}(\Omega)$ bezeichnen wir im Folgenden die Menge aller matrixwertigen Abbildungen $M: \Omega \rightarrow \mathbb{R}^{n \times m}$, deren Komponentenfunktionen Lebesgue-messbar sind, also $M_{ij} \in \mathcal{M}(\Omega)$ für alle $i = 1, \dots, n$, $j = 1, \dots, m$ erfüllt ist. Wählen wir $m = 1$, so erhalten wir die Menge aller vektorwertigen Funktionen $x: \Omega \rightarrow \mathbb{R}^n$ mit Komponentenfunktionen $x_i \in \mathcal{M}(\Omega)$, $i = 1, \dots, n$ und bezeichnen diese mit $\mathcal{M}^n(\Omega)$.

Die elementarste Klasse von matrixwertigen Funktionen sind die konstanten Abbildungen $M(t) = M \in \mathbb{R}^{n \times m}$. Diese werden in der vorliegenden Arbeit jedoch nur eine untergeordnete Rolle spielen. Eine weitaus interessantere Klasse von Matrixfunktionen ist jene, bei der die Komponentenfunktionen alle wesentlich beschränkt sind. Gilt nämlich $M_{ij} \in L_\infty(\Omega)$ für alle $i = 1, \dots, n$, $j = 1, \dots, m$, dann existieren per Definition Konstanten K_{ij} , sodass für alle $i = 1, \dots, n$, $j = 1, \dots, m$ die Ungleichung

$$|M_{ij}(t)| \leq K_{ij} \text{ für fast alle } t \in \Omega \tag{2.3}$$

erfüllt ist. Sei $x \in \mathbb{R}^m$. Mit der Notation $(M(t)x)_i := \sum_{j=1}^m M_{ij}(t)x_j$, $i = 1, \dots, n$ und der Ungleichung von Cauchy-Schwarz erhalten wir

$$\begin{aligned} |M(t)x|_2^2 &= \sum_{i=1}^n |(M(t)x)_i|^2 \\ &= \sum_{i=1}^n \left| \sum_{j=1}^m M_{ij}(t)x_j \right|^2 \leq \sum_{i=1}^n \left(\sum_{j=1}^m |M_{ij}(t)|^2 \cdot \sum_{j=1}^m |x_j|^2 \right) = \sum_{i=1}^n \left(\sum_{j=1}^m |M_{ij}(t)|^2 \cdot |x|_2^2 \right). \end{aligned}$$

Insbesondere ergibt sich dann für alle $x \in \mathbb{R}^m$ mit $|x|_2 = 1$

$$|M(t)x|_2 \leq \left(\sum_{i=1}^n \sum_{j=1}^m |M_{ij}(t)|^2 \right)^{1/2}.$$

Aufgrund der Abschätzungen in (2.3) können wir jetzt direkt auf die Existenz einer Konstanten K mit

$$|M(t)|_2 := \max_{|x|_2=1} |M(t)x|_2 \leq K \text{ für fast alle } t \in \Omega \quad (2.4)$$

schließen. Durch die obige Vorschrift wird eine Matrixnorm, genauer die sogenannte *Spektralnorm* definiert, vgl. Abschnitt 2.2.3, „Exkurs: Matrixnormen im $L_\infty^{n \times m}((0, \infty))$ “. Wir bezeichnen wieder die kleinste obere Schranke einer solchen Konstanten in (2.4) als essentielles Supremum von $|M|_2$ auf Ω und setzen

$$\|M\|_{(\infty)} := \operatorname{ess\,sup}_{t \in \Omega} |M(t)|_2. \quad (2.5)$$

Dies motiviert zu der folgenden Definition des Raumes $L_\infty^{n \times m}(\Omega)$.

Definition 2.11 (vgl. [ASS13], S. 27)

Der Raum

$$L_\infty^{n \times m}(\Omega) := \left\{ M \in \mathcal{M}^{n \times m}(\Omega) \mid \operatorname{ess\,sup}_{t \in \Omega} |M(t)|_2 < \infty \right\}$$

heißt Raum der wesentlich beschränkten Matrixfunktionen und wird mit $L_\infty^{n \times m}(\Omega)$ bezeichnet.

Bemerkung 2.4. *Setzen wir in Definition 2.11 die Dimension $m = 1$, so erhalten wir den Spezialfall der wesentlich beschränkten Vektorfunktionen $L_\infty^n(\Omega)$.*

Die Definition 2.11 lässt sich hierbei auch insofern rechtfertigen, als wir mit $\alpha(t) := |M(t)|_2$ eine Funktion $\alpha: \Omega \rightarrow \mathbb{R}$ erhalten, die aufgrund der vorangegangenen Ausführungen dem Raum $L_\infty(\Omega)$ angehört. Wir schreiben im Folgenden entsprechend $M \in L_\infty^{n \times m}(\Omega)$ und meinen $\alpha \in L_\infty(\Omega)$. Ganz natürlich ergibt sich

$$\|M\|_{(\infty)} := \|\alpha\|_\infty, \quad (2.6)$$

vgl. die Vorschrift in (2.5). Insbesondere erfüllt die damit festgelegte Definition der $\|\cdot\|_{(\infty)}$ -Norm auch alle Normeigenschaften. Würden wir diesen Schemata für die Einführung von Lebesgue- und Sobolev-Räumen für vektorwertige Funktionen folgen, so wäre die Definition

$$L_p^n(\Omega) := \left\{ x \in \mathcal{M}^n(\Omega) \mid \int_\Omega |x(t)|_2^p dt < \infty \right\}$$

intuitiv. Mit anderen Worten würden wir eine Funktion $x = (x_1, \dots, x_n)$ als Element des

Raumes $L_p^n(\Omega)$ identifizieren, wenn die Funktion $\alpha: \Omega \rightarrow \mathbb{R}$ mit $\alpha(t) := |x(t)|_2$ Element des Raumes $L_p(\Omega)$ ist. Ganz analog würden wir

$$\|x\|_{(p)} := \|\alpha\|_p = \left(\int_{\Omega} |x(t)|_2^p dt \right)^{1/p}$$

für die Norm schließen. Tatsächlich lässt sich im maßtheoretischen Zugang über das Bochner-Integral⁸ so der Raum $L_p^n(\Omega)$ analog zum klassischen Lebesgue-Raum konstruieren.⁹ In der vorliegenden Arbeit werden wir jedoch einen anderen Weg einschlagen, welcher intuitiv mit der späteren Interpretation der Elemente des Raumes $L_p^n(\Omega)$ (beziehungsweise $W_p^{k,n}(\Omega)$) einhergeht. Hierzu nutzen wir die im vorangegangenen Kapitel bereits zusammengetragene Theorie zu den Lebesgue- und Sobolev-Räumen reellwertiger Funktionen und konstruieren mittels dieser die Räume $L_p^n(\Omega)$, $W_p^{k,n}(\Omega)$ als kartesisches Produkt. Wir leiten die Theorie also aus der Sicht der Komponentenfunktionen her. Dieses Vorgehen finden wir beispielsweise auch in [AF03]. Anschließend gilt es, die Äquivalenz von diesem Zugang zu jenem über das Bochner-Integral zu verifizieren.

Wir beginnen die Konstruktion des Produktraumes mit dem folgenden Satz über die Charakterisierung von kartesischen Produkträumen.

Satz 2.12 ([AF03], S. 8, 1.23 Theorem)

Seien $(X_j, \|\cdot\|_j)$ für $j = 1, \dots, n$ Banachräume. Weiter sei $X := \sum_{j=1}^n X_j = X_1 \times \dots \times X_n$ der kartesische Produktraum, bestehend aus allen Tupeln $x = (x_1, \dots, x_n)$, $x_j \in X_j$. Mit den Verknüpfungsoperationen

$$x + y = (x_1 + y_1, \dots, x_n + y_n), \quad cx = (cx_1, \dots, cx_n) \tag{2.7}$$

wird X zu einem Vektorraum. Mit den Normen

$$\|x\|_{X,p} := \left(\sum_{j=1}^n \|x_j\|_j^p \right)^{1/p}, \quad 1 \leq p < \infty$$

$$\|x\|_{X,\infty} := \max_{1 \leq j \leq n} \|x_j\|_j$$

wird X zu einem Banachraum. Zusätzlich gelten die folgenden Aussagen:

1. Falls X_j , $j = 1, \dots, n$ separabel ist, so ist auch X separabel.
2. Falls X_j , $j = 1, \dots, n$ reflexiv ist, so ist auch X reflexiv.

Mit Satz 2.12 lassen sich jetzt die im vorangegangenen Kapitel eingeführten Räume $L_p(\Omega)$, $W_p^k(\Omega)$ auf den Fall vektorwertiger Funktionen übertragen. Wir erhalten entsprechend das kartesische Produkt auf dem Raum $L_p(\Omega)$

$$[L_p(\Omega)]^n := \underbrace{L_p(\Omega) \times \dots \times L_p(\Omega)}_{n \text{ mal}}$$

⁸ Das Bochner-Integral ist eine Erweiterung des Lebesgue-Integrals auf banachwertige Abbildungen.

⁹ Eine aus unserer Sicht gute Ausarbeitung zu dieser Thematik ist hier die Masterarbeit [Kre15].

mit den punktweisen Verknüpfungen wie in (2.7). Mit der Norm

$$\|x\|_{[L_p(\Omega)]^n} := \left(\sum_{j=1}^n \|x_j\|_p^p \right)^{1/p}, \quad x = (x_1, \dots, x_n) \in [L_p(\Omega)]^n$$

wird der Raum für $1 \leq p < \infty$ gemäß des Satzes 2.12 zu einem separablen Banachraum. Für $1 < p < \infty$ ist dieser sogar reflexiv. Im Fall $p = 2$ erhalten wir mit dem Skalarprodukt

$$\langle x, y \rangle_{[L_2(\Omega)]^n} := \sum_{j=1}^n \langle x_j, y_j \rangle_2 = \int_{\Omega} x^T(t)y(t)dt$$

einen Hilbertraum. Die Norm ergibt sich in diesem Fall zu

$$\|x\|_{[L_2(\Omega)]^n} = \left(\int_{\Omega} |x(t)|_2^2 dt \right)^{1/2},$$

wobei hier unter dem Integral die euklidische Norm auftritt. Für $1 \leq p < \infty$ lassen sich im Sinne des folgenden Lemmas äquivalente Normen formulieren.

Lemma 2.13

Sei $1 \leq p < \infty$. Die Norm

$$\|x\|_{(p)} := \left(\int_{\Omega} |x(t)|_2^p dt \right)^{1/p}$$

ist äquivalent zu der Norm $\|x\|_{[L_p(\Omega)]^n}$ des Raumes $[L_p(\Omega)]^n$.

Beweis. Aufgrund der Äquivalenz der Normen im euklidischen Standardraum \mathbb{R}^n existieren Konstanten $c, C > 0$ mit der Eigenschaft

$$c|\cdot|_p \leq |\cdot|_2 \leq C|\cdot|_p.$$

Mit den Konstanten $\tilde{c} := c^{1/p}$, $\tilde{C} := C^{1/p}$ impliziert dies für die Normen im Funktionenraum

$$\begin{aligned} \tilde{c}\|x\|_{[L_p(\Omega)]^n} &= \tilde{c} \left(\sum_{j=1}^n \|x_j\|_p^p \right)^{1/p} = \left(\int_{\Omega} c|x(t)|_p^p dt \right)^{1/p} \\ &\leq \left(\int_{\Omega} |x(t)|_2^p dt \right)^{1/p} = \|x\|_{(p)} \\ &\leq \left(\int_{\Omega} C|x(t)|_p^p dt \right)^{1/p} = \tilde{C}\|x\|_{[L_p(\Omega)]^n}. \end{aligned}$$

Folglich sind die Normen $\|\cdot\|_{[L_p(\Omega)]^n}$ und $\|\cdot\|_{(p)}$ äquivalent. □

Auf der zuvor eingeführten äquivalenten Norm $\|\cdot\|_{(p)}$ aufbauend, definieren wir jetzt den Raum $L_p^n(\Omega)$ gemäß der folgenden Definition.

Definition 2.14

Sei $1 \leq p < \infty$. Der Raum

$$L_p^n(\Omega) := \left\{ x \in \mathcal{M}^n(\Omega) \mid \int_{\Omega} |x(t)|_2^p dt < \infty \right\}$$

wird als vektorwertiger Lebesgue-Raum der Ordnung p bezeichnet.

Lemma 2.15

Sei $1 \leq p < \infty$. Der Raum $L_p^n(\Omega)$ versehen mit der Norm

$$\|x\|_{(p)} = \left(\int_{\Omega} |x(t)|_2^p dt \right)^{1/p}$$

genügt den folgenden Eigenschaften:

1. $L_p^n(\Omega)$ ist für $1 \leq p < \infty$ ein separabler Banachraum.
2. $L_p^n(\Omega)$ ist für $1 < p < \infty$ ein reflexiver Raum.
3. $L_2^n(\Omega)$ wird mit dem Skalarprodukt

$$\langle x, y \rangle_{(2)} = \int_{\Omega} x^T(t)y(t)dt$$

ein Hilbertraum.

4. Für $1 \leq p < \infty$ liegt der Raum $C_0^\infty(\Omega, \mathbb{R}^n)$ dicht in dem Raum $L_p^n(\Omega)$.
5. Es gilt die Inklusion $L_p^n(\Omega) \subseteq L_{1,loc}^n(\Omega)$.

Beweis. Wir zeigen zunächst $L_p^n(\Omega) = [L_p(\Omega)]^n$. Sei $x: \Omega \rightarrow \mathbb{R}^n$, $x = (x_1, \dots, x_n)$, $x \in L_p^n(\Omega)$ gegeben. Aus der Messbarkeit der Funktion x folgt die Messbarkeit der Komponenten x_i , $i = 1, \dots, n$ (siehe [Els11], S. 87). Mit Lemma 2.13 folgt

$$\|x_i\|_p \leq \left(\sum_{j=1}^n \|x_j\|_p^p \right)^{1/p} = \|x\|_{[L_p(\Omega)]^n} \leq \frac{1}{c} \|x\|_{(p)} < \infty$$

für jede Komponente x_i , $i = 1, \dots, n$ und mit der Konstanten $\tilde{c} > 0$. Folglich gilt $x_i \in L_p(\Omega)$ und per Definition $x = (x_1, \dots, x_n) \in [L_p(\Omega)]^n$.

In Umkehrung erhalten wir für $x: \Omega \rightarrow \mathbb{R}^n$, $x = (x_1, \dots, x_n)$, $x \in [L_p(\Omega)]^n$ die folgende Implikationskette. Zunächst folgt aus der Messbarkeit aller Komponenten x_i , $i = 1, \dots, n$

die Messbarkeit der vektorwertigen Funktion x (siehe [Els11], S. 87). Mit Lemma 2.13 folgt schließlich

$$\|x\|_{(p)} \leq \tilde{C} \|x\|_{[L_p(\Omega)]^n} = \tilde{C} \left(\sum_{j=1}^n \|x_j\|_p^p \right)^{1/p} < \infty,$$

wobei $\tilde{C} > 0$ und die Normen $\|x_i\|_p$ hierbei nach Konstruktion des Produktraumes $[L_p(\Omega)]^n$ wegen $x_i \in L_p(\Omega)$ endlich sind. Folglich gehört x dem Raum $L_p^n(\Omega)$ an und die Äquivalenz der Räume ist bewiesen.

Die Behauptungen des Lemmas folgen jetzt aus der Äquivalenz der Räume zusammen mit der Normäquivalenz und den Eigenschaften der Grundräume $L_p(\Omega)$, vgl. Seite 11. \square

Wir widmen diesen Abschnitt abschließend noch der Einführung der Sobolev-Räume für vektorwertige Funktionen. Zunächst führen wir wieder den Produktraum

$$\left[W_p^k(\Omega) \right]^n := \underbrace{W_p^k(\Omega) \times \dots \times W_p^k(\Omega)}_{n \text{ mal}}$$

mit den punktweisen Verknüpfungen aus (2.7) ein. Gemäß des Satzes 2.12 erhalten wir mit der Norm

$$\|x\|_{[W_p^k(\Omega)]^n} = \left(\sum_{j=1}^n \|x_j\|_{k,p}^p \right)^{1/p}, \quad x = (x_1, \dots, x_n) \in [W_p^k(\Omega)]^n$$

für $1 \leq p < \infty$ einen separablen Banachraum. Im Fall $1 < p < \infty$ wird der Raum sogar reflexiv. Im Fall $p = 2$ erhalten wir mit dem Skalarprodukt

$$\begin{aligned} \langle x, y \rangle_{[W_2^k(\Omega)]^n} &= \sum_{j=1}^n \langle x_j, y_j \rangle_{k,2} = \sum_{j=1}^n \left(\langle x_j, y_j \rangle_2 + \sum_{0 < |\alpha| \leq k} \langle \mathcal{D}^\alpha x_j, \mathcal{D}^\alpha y_j \rangle_2 \right) \\ &= \langle x, y \rangle_{[L_2(\Omega)]^n} + \sum_{0 < |\alpha| \leq k} \langle \mathcal{D}^\alpha x, \mathcal{D}^\alpha y \rangle_{[L_2(\Omega)]^n} \\ &= \langle x, y \rangle_{(2)} + \sum_{0 < |\alpha| \leq k} \langle \mathcal{D}^\alpha x, \mathcal{D}^\alpha y \rangle_{(2)} \end{aligned}$$

einen Hilbertraum. Die Norm reduziert sich in diesem Fall zu

$$\|x\|_{[W_2^k(\Omega)]^n} := \left(\|x\|_{(2)}^2 + \sum_{0 < |\alpha| \leq k} \|\mathcal{D}^\alpha x\|_{(2)}^2 \right)^{1/2}$$

mit den zuvor eingeführten Normen für die Produkt-Lebesgue-Räume. Für ein allgemeines $1 \leq p < \infty$ lässt sich die folgende Normäquivalenz beweisen.

Lemma 2.16

Sei $\|\cdot\|_{(p)}$ die in Lemma 2.13 definierte Norm. Weiter sei $1 \leq p < \infty$. Die Norm

$$\|x\|_{k,(p)} := \left(\|x\|_{(p)}^p + \sum_{0 < |\alpha| \leq k} \|\mathcal{D}^\alpha x\|_{(p)}^p \right)^{1/p}$$

ist äquivalent zur Norm $\|x\|_{[W_p^k(\Omega)]^n}$ des Raumes $[W_p^k(\Omega)]^n$.

Beweis. Aufgrund der Äquivalenz der Normen $\|\cdot\|_{[L_p(\Omega)]^n}$ und $\|\cdot\|_{(p)}$, die wir in Lemma 2.13 gezeigt haben, existieren Konstanten $c, C > 0$ mit der Eigenschaft

$$c\|\cdot\|_{[L_p(\Omega)]^n} \leq \|\cdot\|_{(p)} \leq C\|\cdot\|_{[L_p(\Omega)]^n}.$$

Dies impliziert für die Sobolev-Norm

$$\begin{aligned} \sqrt[p]{c} \left(\sum_{j=1}^n \|x_j\|_{k,p}^p \right)^{1/p} &= \sqrt[p]{c} \left(\sum_{j=1}^n \left[\|x_j\|_p^p + \sum_{0 < |\alpha| \leq k} \|\mathcal{D}^\alpha x_j\|_p^p \right] \right)^{1/p} \\ &= \left(\left[c\|x\|_{[L_p(\Omega)]^n}^p + \sum_{0 < |\alpha| \leq k} c\|\mathcal{D}^\alpha x\|_{[L_p(\Omega)]^n}^p \right] \right)^{1/p} \\ &\leq \left(\left[\|x\|_{(p)}^p + \sum_{0 < |\alpha| \leq k} \|\mathcal{D}^\alpha x\|_{(p)}^p \right] \right)^{1/p} \\ &\leq \left(\left[C\|x\|_{[L_p(\Omega)]^n}^p + \sum_{0 < |\alpha| \leq k} C\|\mathcal{D}^\alpha x\|_{[L_p(\Omega)]^n}^p \right] \right)^{1/p} \\ &= \sqrt[p]{C} \left(\sum_{j=1}^n \left[\|x_j\|_p^p + \sum_{0 < |\alpha| \leq k} \|\mathcal{D}^\alpha x_j\|_p^p \right] \right)^{1/p} \\ &= \sqrt[p]{C} \left(\sum_{j=1}^n \|x_j\|_{k,p}^p \right)^{1/p} \end{aligned}$$

Folglich sind die Normen $\|\cdot\|_{[W_p^k(\Omega)]^n}$ und $\|\cdot\|_{k,(p)}$ äquivalent. □

Aufbauend auf der Normäquivalenz führen wir jetzt den Sobolevraum $W_p^{k,n}(\Omega)$ ein.

Definition 2.17

Sei $1 \leq p < \infty$ und $k \in \mathbb{N}_0$. Der Raum

$$W_p^{k,n}(\Omega) := \left\{ x \in L_p^n(\Omega) \mid \mathcal{D}^\alpha x \in L_p^n(\Omega) \text{ für alle } \alpha \in \mathbb{N}_0^s \text{ mit } |\alpha| \leq k \right\}$$

heißt Sobolev-Raum der vektorwertigen Funktionen.

Wir zeigen abschließend noch, dass der so definierte Raum ein separabler Banachraum ist.

Lemma 2.18

Sei $1 \leq p < \infty$ und $k \in \mathbb{N}_0$. Der Raum $W_p^{k,n}(\Omega)$ versehen mit der Norm

$$\|x\|_{k,(p)} := \left(\|x\|_{(p)}^p + \sum_{0 < |\alpha| \leq k} \|\mathcal{D}^\alpha x\|_{(p)}^p \right)^{1/p}$$

ist

1. für $1 \leq p < \infty$ ein separabler Banachraum.
2. für $1 < p < \infty$ ein reflexiver Raum.
3. für $p = 2$ mit dem Skalarprodukt

$$\langle x, y \rangle_{k,(2)} := \langle x, y \rangle_{(2)} + \sum_{0 < |\alpha| \leq k} \langle \mathcal{D}^\alpha x, \mathcal{D}^\alpha y \rangle_{(2)}$$

ein Hilbertraum.

4. Falls Ω ein Gebiet ist, welches die stark lokale Lipschitz-Bedingung erfüllt und $\Omega_0 \subseteq \Omega$ ein beschränktes Teilgebiet ist, so ist die Einbettung

$$W_p^{j+r,n}(\Omega) \hookrightarrow C^j(\bar{\Omega}_0, \mathbb{R}^n)$$

für $j \geq 0$, $r \geq 1$, $1 < p < \infty$ und $rp > p$ kompakt.

Bemerkung 2.5 (vgl. [AF03], S. 169, 6.4 Remarks). *Wie bereits in Bemerkung 2.2 erwähnt, können wir, falls Ω beschränkt ist, ohne Einschränkung $\Omega_0 = \Omega$ in Lemma 2.18 Punkt 4 setzen.*

Beweis zu Lemma 2.18. Wir zeigen zunächst wieder $W_p^{k,n}(\Omega) = [W_p^k(\Omega)]^n$. Sei hierzu $x: \Omega \rightarrow \mathbb{R}^n$, $x = (x_1, \dots, x_n)$, $x \in W_p^{k,n}(\Omega)$ gegeben. Gemäß Definition folgt unmittelbar $x \in L_p^n(\Omega)$ und $\mathcal{D}^\alpha x \in L_p^n(\Omega)$ für $|\alpha| \leq k$. Aus dem Beweis von Lemma 2.15 folgt jetzt $x_i \in L_p(\Omega)$ und $\mathcal{D}^\alpha x_i \in L_p(\Omega)$, $|\alpha| \leq k$ und für jede Komponente x_i , $i = 1, \dots, n$. Insgesamt erhalten wir $x_i \in W_p^k(\Omega)$, $i = 1, \dots, n$. Nach Konstruktion gehört jetzt $x = (x_1, \dots, x_n)$ den kartesischen Produktraum $[W_p^k(\Omega)]^n$ an.

In Umkehrung erhalten wir für $x: \Omega \rightarrow \mathbb{R}^n$, $x = (x_1, \dots, x_n)$, $x \in [W_p^k(\Omega)]^n$ die folgende Schlussfolgerung. Zunächst ist per Definition jede Komponente $x_i \in W_p^k(\Omega)$, sodass $x_i \in L_p(\Omega)$ und $\mathcal{D}^\alpha x_i \in L_p(\Omega)$, $|\alpha| \leq k$ gilt. Gemäß des Beweises von Lemma 2.15 folgt $x \in L_p^n(\Omega)$ und $\mathcal{D}^\alpha x = (\mathcal{D}^\alpha x_1, \dots, \mathcal{D}^\alpha x_n) \in L_p^n(\Omega)$. Folglich gilt $x \in W_p^{k,n}(\Omega)$ und die Räume $W_p^{k,n}(\Omega)$, $[W_p^k(\Omega)]^n$ sind äquivalent.

Die Behauptungen des Lemmas folgen jetzt unter Beachtung der Raum- beziehungsweise Normäquivalenz aus den Eigenschaften des Grundraumes $W_p^k(\Omega)$, vgl. Seite 14. \square

2.2.3 Exkurs: Matrixnormen im $L_\infty^{n \times m}((0, \infty))$

Bei der Formulierung unserer Aufgabenstellung treten Abbildungen auf, deren Werte Elemente des $\mathbb{R}^{n \times m}$ sind. Aufgrund dessen haben wir bereits den Begriff einer Matrixfunktion eingeführt, vgl. Definition 2.10. Ziel dieses Abschnittes ist es, die Güte einer wesentlich beschränkten Matrixfunktion $M \in L_\infty^{n \times m}((0, \infty))$ und deren Norm genauer zu studieren. Gemäß der Definition 2.11 ist der Raum $L_\infty^{n \times m}((0, \infty))$ durch die Vorschrift

$$L_\infty^{n \times m}((0, \infty)) := \left\{ M \in \mathcal{M}^{n \times m}((0, \infty)) \mid \operatorname{ess\,sup}_{t \in (0, \infty)} |M(t)|_2 < \infty \right\}$$

festgelegt. Um die entsprechende Norm $\|\cdot\|_{(\infty)}$ (vgl. (2.6)) nun genauer studieren zu können, benötigen wir einige Hilfsmittel aus der linearen Algebra und der numerischen Mathematik, die wir hier zusammenstellen möchten. Für eine umfassende Einführung verweisen wir zum Beispiel auf die Literatur [SB93], [Kan05] und [BR14], die die Grundlage dieses Abschnittes bilden.

Wir beginnen damit, zunächst an den Begriff einer Matrixnorm zu erinnern und führen ihn in Analogie zu einer Norm eines Vektorraumes V über dem Körper \mathbb{K} als eine Norm definiert auf dem Raum der (reellen oder komplexen) Matrizen $\mathbb{K}^{n \times m}$ ein. Dabei beschränken wir uns auf den für uns relevanten Fall, bei dem wir ausschließlich den Körper der reellen Zahlen betrachten, also im Folgenden stets von $\mathbb{K} = \mathbb{R}$ ausgehen.

Definition 2.19 ([BR14], S. 493, Definition 15.5)

Eine Matrixnorm ist eine Funktion $\|\cdot\| : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$, $A \mapsto \|A\|$, die für alle Matrizen $A, B \in \mathbb{R}^{n \times m}$ und Skalare $\alpha \in \mathbb{R}$ die folgenden Eigenschaften erfüllt:

- (i) Positive Definitheit: $\|A\| \geq 0$ und $\|A\| = 0 \Leftrightarrow A = 0$,
- (ii) Homogenität: $\|\alpha A\| = |\alpha| \|A\|$,
- (iii) Dreiecksungleichung: $\|A + B\| \leq \|A\| + \|B\|$.

Weiterhin sprechen wir von einer *submultiplikativen* Matrixnorm, wenn für zwei Matrizen $A \in \mathbb{R}^{n \times m}$ und $B \in \mathbb{R}^{m \times k}$ die Ungleichung

$$\|AB\| \leq \|A\| \|B\|$$

gilt. Interessant für unsere späteren Anwendungen sind insbesondere solche Matrixnormen, welche auf der Grundlage von Vektornormen festgelegt sind und von ihnen als Operatornorm abgeleitet werden können. Das folgende Lemma zeigt, wie aus gegebenen Vektornormen in natürlicher Weise eine Matrixnorm induziert wird.

Lemma 2.20 ([Kan05], S. 17, Lemma 1.15.)

Seien $\|\cdot\|_{\mathbb{R}^m} : \mathbb{R}^m \rightarrow \mathbb{R}$ und $\|\cdot\|_{\mathbb{R}^n} : \mathbb{R}^n \rightarrow \mathbb{R}$ gegebene Vektornormen. Dann definiert die Vorschrift

$$\|A\| := \max_{\|x\|_{\mathbb{R}^m} = 1} \|Ax\|_{\mathbb{R}^n} \tag{2.8}$$

für alle Matrizen $A \in \mathbb{R}^{n \times m}$ eine Matrixnorm auf dem $\mathbb{R}^{n \times m}$.

Die in (2.8) definierte Norm nennen wir *natürliche* oder die durch die zugrunde liegenden Vektornormen *induzierte Matrixnorm*. Wie bemerken zunächst, dass aufgrund der Stetigkeit der Norm auf der kompakten Einheitssphäre

$$S := \{x \in \mathbb{R}^m \mid \|x\|_{\mathbb{R}^m} = 1\}$$

das Maximum in (2.8) nach dem Satz von Weierstraß tatsächlich angenommen wird. Auf den Nachweis der Normaxiome für $\|A\|$ verzichten wir an dieser Stelle und verweisen diesbezüglich auf die angegebene Quelle. Wir möchten jedoch darauf hinweisen, dass für die Norm in (2.8) stets

$$\|A\| = \max_{\|x\|_{\mathbb{R}^m} \leq 1} \|Ax\|_{\mathbb{R}^n} = \max_{x \neq 0} \frac{\|Ax\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^m}} \quad (2.9)$$

gilt. Die Gültigkeit dieser Gleichungen können wir leicht verifizieren. Offensichtlich ist

$$\max_{\|x\|_{\mathbb{R}^m} = 1} \|Ax\|_{\mathbb{R}^n} \leq \max_{\|x\|_{\mathbb{R}^m} \leq 1} \|Ax\|_{\mathbb{R}^n} \quad (2.10)$$

stets korrekt. Ist $0 < \|x\|_{\mathbb{R}^m} \leq 1$, so ergibt sich die Abschätzung

$$\|Ax\|_{\mathbb{R}^n} \leq \frac{\|Ax\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^m}} \leq \max_{x \neq 0} \frac{\|Ax\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^m}}.$$

Gehen wir in der obigen Ungleichung zum Maximum über, erhalten wir

$$\max_{\|x\|_{\mathbb{R}^m} \leq 1} \|Ax\|_{\mathbb{R}^n} \leq \max_{x \neq 0} \frac{\|Ax\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^m}}. \quad (2.11)$$

Wählen wir nun $x \in \mathbb{R}^m \setminus \{0\}$, dann gilt

$$\frac{\|Ax\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^m}} = \left\| A \frac{x}{\|x\|_{\mathbb{R}^m}} \right\|_{\mathbb{R}^n} \leq \max_{\|y\|_{\mathbb{R}^m} = 1} \|Ay\|_{\mathbb{R}^n}$$

und somit auch

$$\max_{x \neq 0} \frac{\|Ax\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^m}} \leq \max_{\|y\|_{\mathbb{R}^m} = 1} \|Ay\|_{\mathbb{R}^n}. \quad (2.12)$$

Tragen wir die Ergebnisse aus (2.10), (2.11) und (2.12) zusammen, lässt sich direkt auf

$$\max_{\|x\|_{\mathbb{R}^m} = 1} \|Ax\|_{\mathbb{R}^n} \leq \max_{\|x\|_{\mathbb{R}^m} \leq 1} \|Ax\|_{\mathbb{R}^n} \leq \max_{x \neq 0} \frac{\|Ax\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^m}} \leq \max_{\|x\|_{\mathbb{R}^m} = 1} \|Ax\|_{\mathbb{R}^n}$$

und somit auf die Gültigkeit der Gleichung

$$\|A\| := \max_{\|x\|_{\mathbb{R}^m} = 1} \|Ax\|_{\mathbb{R}^n} = \max_{\|x\|_{\mathbb{R}^m} \leq 1} \|Ax\|_{\mathbb{R}^n}$$

schließen. Da sich jedes $x \in \mathbb{R}^m$ mit $\|x\|_{\mathbb{R}^m} = 1$ als $x = \frac{y}{\|y\|_{\mathbb{R}^m}}$ mit $y \neq 0$ darstellen lässt,

gilt außerdem

$$\max_{\|x\|_{\mathbb{R}^m}=1} \|Ax\|_{\mathbb{R}^n} = \max_{y \neq 0} \left\| A \frac{y}{\|y\|_{\mathbb{R}^m}} \right\|_{\mathbb{R}^n},$$

womit auch die zweite Gleichheit in (2.9) verifiziert ist.

Betrachten wir die induzierte Matrixnorm und deren Darstellungen in (2.9) genauer, so erkennen wir sofort die Parallelen zu der in der Funktionalanalysis verwendeten Definition einer Operatornorm: Für einen linearen und stetigen Operator $T : X \rightarrow Y$, wobei $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ normierte Räume über \mathbb{R} sind, wird die Operatornorm durch die Vorschrift

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\|x\|_X=1} \|Tx\|_Y = \sup_{\|x\|_X \leq 1} \|Tx\|_Y$$

charakterisiert, vgl. [Wer18], S. 50 f., Definition II.1.3 und Satz II.1.4 (a). Deswegen wird die induzierte Matrixnorm oft auch als Operatornorm bezeichnet.

Bei den induzierten Matrixnormen bekommen wir weitere nützliche Eigenschaften automatisch mitgeliefert: Sei $x \in \mathbb{R}^m$ ein beliebiger, vom Nullvektor verschiedener Vektor, den wir normieren und dafür die Bezeichnung $y := x / \|x\|_{\mathbb{R}^m}$ verwenden. Dann ist $y \in S$ und es gilt $\|Ay\|_{\mathbb{R}^n} \leq \|A\|$. Damit erhalten wir die Ungleichung

$$\|Ax\|_{\mathbb{R}^n} = \|x\|_{\mathbb{R}^m} \|Ay\|_{\mathbb{R}^n} \leq \|x\|_{\mathbb{R}^m} \|A\|,$$

die trivialerweise auch für $x = 0$ erfüllt ist. Sei $\|\cdot\|_{\mathbb{R}^k}$ eine Vektornorm auf dem \mathbb{R}^k . Wir nutzen die vorangegangene Abschätzung und können für die induzierte Norm aus (2.8) folgern, dass

$$\|AB\| := \max_{\|z\|_{\mathbb{R}^k}=1} \|ABz\|_{\mathbb{R}^n} \leq \max_{\|z\|_{\mathbb{R}^k}=1} \|A\| \|Bz\|_{\mathbb{R}^m} = \|A\| \max_{\|z\|_{\mathbb{R}^k}=1} \|Bz\|_{\mathbb{R}^m} = \|A\| \|B\|$$

für alle Matrizen $A \in \mathbb{R}^{n \times m}$ und $B \in \mathbb{R}^{m \times k}$ richtig ist, also jede induzierte Matrixnorm auch submultiplikativ ist. Wir geben die ermittelten Eigenschaften induzierter Matrixnormen zusammenfassend in dem nachfolgenden Lemma an.

Lemma 2.21 (vgl. [Kan05], S. 20, Lemma 1.18.)

Seien $\|\cdot\|_{\mathbb{R}^n}$, $\|\cdot\|_{\mathbb{R}^m}$, $\|\cdot\|_{\mathbb{R}^k}$ gegebene Vektornormen auf dem \mathbb{R}^n , \mathbb{R}^m und dem \mathbb{R}^k . Für die dadurch in Lemma 2.20 eingeführten induzierten Matrixnormen gilt

$$\|Ax\|_{\mathbb{R}^n} \leq \|A\| \|x\|_{\mathbb{R}^m} \quad \text{für alle } A \in \mathbb{R}^{n \times m} \text{ und alle } x \in \mathbb{R}^m, \quad (2.13)$$

sowie

$$\|AB\| \leq \|A\| \|B\| \quad \text{für alle } A \in \mathbb{R}^{n \times m}, B \in \mathbb{R}^{m \times k}.$$

Bemerkung 2.6. Die in (2.13) ermittelte Eigenschaft besagt, dass jede induzierte Matrixnorm mit der Vektornorm, aus der sie abgeleitet wurde, verträglich, ist. Bei der Betrachtung

tion von Abschätzungen, bei denen Vektoren und Matrizen gemeinsam auftreten, erweist sich diese Verträglichkeitsbedingung als sehr nützlich und ist deshalb von großer Bedeutung.

Wir werden uns nun im weiteren Verlauf dieser Arbeit auf die durch die euklidische Vektornorm $|\cdot|_2$ natürlich induzierte Matrixnorm konzentrieren.

Lemma 2.22 (vgl. [Kan05], S. 18, Lemma 1.16., [SK11], S.51 f.)

Sei $A \in \mathbb{R}^{n \times m}$. Für die von der euklidischen Vektornorm $|\cdot|_2$ abgeleitete natürliche bzw. induzierte Matrixnorm $|A|_2$ gilt

$$|A|_2 = \sqrt{\lambda_{\max}(A^T A)}, \quad (2.14)$$

wobei $\lambda_{\max}(A^T A)$ den größten Eigenwert der Matrix $A^T A$ meint. Die Matrixnorm $|A|_2$ bezeichnet man auch als Spektralnorm.

Beweis. Für den Beweis der Gleichung in (2.14) benötigen wir zunächst einige Grundlagen aus der linearen Algebra. Wir erinnern daran, dass für eine allgemeine Matrix $A \in \mathbb{R}^{n \times m}$ die Matrix $A^T A$ eine symmetrische, quadratische $m \times m$ und positiv semidefinite Matrix ist. Die Eigenwerte λ_i ($i = 1, \dots, m$) von $A^T A$ sind demnach alle reell und nicht negativ. Überdies bilden die Eigenvektoren x_1, \dots, x_m eine vollständige Orthonormalbasis im \mathbb{R}^m . Es lässt sich also jeder Vektor $x \in \mathbb{R}^m$ unter Verwendung der Eigenvektoren in der Form

$$x = \sum_{i=1}^m \alpha_i x_i, \quad \alpha_i \in \mathbb{R} \quad (2.15)$$

darstellen, wobei

$$(A^T A)x_i = \lambda_i x_i, \quad x_i \in \mathbb{R}^m, \quad \lambda_i \geq 0 \quad (2.16)$$

und

$$x_i^T x_j = \begin{cases} 1 & i \neq j \\ 0 & i = j \end{cases} \quad (2.17)$$

gelten. Unter Berücksichtigung von (2.15) - (2.17) erhalten wir zunächst für alle $x \in \mathbb{R}^m$

$$\begin{aligned} |Ax|_2^2 &= (Ax)^T (Ax) = x^T A^T A x \\ &= \left(\sum_{i=1}^m \alpha_i x_i \right)^T A^T A \left(\sum_{i=1}^m \alpha_i x_i \right) = \left(\sum_{i=1}^m \alpha_i x_i \right)^T \left(\sum_{i=1}^m \alpha_i \lambda_i x_i \right) = \sum_{i=1}^m \lambda_i \alpha_i^2. \end{aligned} \quad (2.18)$$

Sei nun $\lambda_{\max}(A^T A) := \lambda_1$ der größte Eigenwert der Matrix $A^T A$. Für Vektoren $x \in \mathbb{R}^m$ mit $|x|_2 = 1$ folgt aus der Darstellung in (2.15) sofort, dass

$$\sum_{i=1}^m \alpha_i^2 = 1 \quad (2.19)$$

gilt. Die aus der euklidischen Vektornorm induzierte Matrixnorm ist durch die Vorschrift

$$|A|_2 := \max_{|x|_2=1} |Ax|_2$$

definiert, vgl. Lemma 2.20. Unter Berücksichtigung von (2.18) ergibt sich

$$|A|_2 := \max_{|x|_2=1} |Ax|_2 = \max_{|x|_2=1} \left(\sum_{i=1}^m \lambda_i \alpha_i^2 \right)^{1/2} \leq \max_{|x|_2=1} \left(\lambda_{\max}(A^T A) \sum_{i=1}^m \alpha_i^2 \right)^{1/2}$$

und wegen (2.19)

$$|A|_2 := \max_{|x|_2=1} |Ax|_2 \leq \sqrt{\lambda_{\max}(A^T A)}. \quad (2.20)$$

Wir betrachten nun speziell $x = x_1$, wobei x_1 gerade der (normierte) Eigenvektor der Matrix $A^T A$ zu dem Eigenwert $\lambda_1 = \lambda_{\max}(A^T A)$ ist. Unter Benutzung von (2.16) folgt dann

$$|Ax_1|_2^2 = x_1^T A^T A x_1 = \lambda_{\max}(A^T A) x_1^T x_1 = \lambda_{\max}(A^T A) |x_1|_2^2 = \lambda_{\max}(A^T A),$$

sodass der maximal mögliche Wert $\sqrt{\lambda_{\max}(A^T A)}$ in (2.20) für $x = x_1$ angenommen wird und demnach sogar Gleichheit in (2.20) gilt. \square

Wir erinnern uns daran, dass die Menge aller Eigenwerte einer beliebigen quadratischen Matrix $M \in \mathbb{R}^{n \times n}$

$$\sigma(M) := \{\lambda \in \mathbb{C} \mid \lambda \text{ ist Eigenwert von } M\}$$

als das *Spektrum* von M bezeichnet wird und der betragsgrößte Eigenwert

$$\rho(M) := \max \{|\lambda| \mid \lambda \in \sigma(A)\}$$

der *Spektralradius* von M heißt. Entsprechend können wir die Spektralnorm in (2.14) für eine allgemeine Matrix $A \in \mathbb{R}^{n \times m}$ auch mit Hilfe der Beziehung

$$|A|_2 = \sqrt{\rho(A^T A)} \quad (2.21)$$

beschreiben. Die Bezeichnung als Spektralnorm ist speziell in dem Fall einer symmetrischen Matrix $A \in \mathbb{R}^{n \times n}$ besser nachvollziehbar. Zunächst können wir aus dem Spektralsatz (vgl. z.B. [Kan05], S. 13, Satz 1.8.) die Existenz einer orthogonalen Matrix $Q \in \mathbb{R}^{n \times n}$ mit

$$Q^T A Q = \text{diag} \{\lambda_1, \dots, \lambda_n\} \in \mathbb{R}^{n \times n}$$

folgern. Hierbei sind die Diagonalelemente $\lambda_i, i = 1, \dots, n$ gerade die Eigenwerte von A . Gemeinsam mit der Symmetrie $A^T = A$ und der Orthogonalität¹⁰ $Q^T Q = Q Q^T = I$ können

¹⁰ Mit $I = \text{diag} \{1, \dots, 1\}$ bezeichnen wir die Einheitsmatrix, bei der die Diagonalelemente alle Eins und alle anderen Elemente Null sind.

wir dann auf die Spektralzerlegung der symmetrischen Matrix $A^T A$ schließen

$$Q^T A^T A Q = Q^T A^T (Q Q^T) A Q = (Q^T A^T Q) (Q^T A Q) = \text{diag} \{ \lambda_1^2, \dots, \lambda_n^2 \},$$

wonach die Eigenwerte von $A^T A$ gerade mit den Quadraten der Eigenwerte $\lambda_1^2, \dots, \lambda_n^2$ von A übereinstimmen. Als direkte Folgerung daraus erhalten wir $\rho(A^T A) = \rho(A)^2$ und aus (2.21) ergibt sich dann sofort

$$|A|_2 = \sqrt{\rho(A^T A)} = \rho(A).$$

Für eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ stimmt die Spektralnorm $|A|_2$ also mit dem Spektralradius der Matrix selbst überein. Ist die symmetrische Matrix A zusätzlich positiv semidefinit, sind all ihre Eigenwerte größer oder gleich Null und die Spektralnorm entspricht gerade ihrem größten Eigenwert. Wir halten diese Resultate in dem nachfolgenden Lemma fest.

Lemma 2.23

Für jede Matrix $A \in \mathbb{R}^{n \times n}$ gilt

- (i) $|A|_2 = \rho(A)$, falls A symmetrisch ist,
- (ii) $|A|_2 = \lambda_{\max}(A)$, falls A symmetrisch und positiv semidefinit ist. Dabei ist $\lambda_{\max}(A)$ der größte Eigenwert von A .

Wir möchten die gewonnenen Kenntnisse über die Spektralnorm nun in dem Kontext der auf $(0, \infty)$ wesentlich beschränkten Matrixfunktionen $M : \mathbb{R}_+ \rightarrow \mathbb{R}^{n \times m}$ anwenden. Für eine Funktion $M \in L_\infty^{n \times m}((0, \infty))$ haben wir

$$\|M\|_{(\infty)} := \text{ess sup}_{t \in (0, \infty)} |M(t)|_2$$

festgelegt, vgl. Definition 2.11 und (2.5). Aus der Verträglichkeit zwischen der euklidischen Vektornorm und der Spektralnorm, vgl. Bemerkung 2.6, gewinnen wir die folgende nützliche Abschätzung für fast alle $t \in (0, \infty)$:

$$|M(t)x|_2 \leq |M(t)|_2 |x|_2 \leq \|M\|_{(\infty)} |x|_2 \quad \forall x \in \mathbb{R}^n. \tag{2.22}$$

Als direkte Folgerung aus den Resultaten in dem Lemma 2.22 und dem Lemma 2.23 können wir in Abhängigkeit von den Eigenschaften der Matrix $M = M(t)$ die Norm $\|M\|_{(\infty)}$ genauer charakterisieren.

Lemma 2.24

1. Für jede Matrixfunktion $M \in L_\infty^{n \times m}((0, \infty))$ gilt

$$\|M\|_{(\infty)} = \text{ess sup}_{t \in \mathbb{R}_+} \sqrt{\lambda_{\max}(M^T M)(t)}.$$

Dabei ist $M = M(t)$ eine von t abhängige Matrix. Dementsprechend bezeichnen wir mit $\lambda_{\max}(M^T M)(t)$ den vom Parameter t abhängigen größten Eigenwert der Matrix $M^T M$.

2. Sei $N \in L_{\infty}^{n \times n}((0, \infty))$ mit symmetrischer Matrix $N = N(t)$. Dann ist

$$\|N\|_{(\infty)} = \operatorname{ess\,sup}_{t \in (0, \infty)} |\lambda_{\max}(N)(t)|,$$

wobei mit $\lambda_{\max}(N) = \lambda_{\max}(N)(t)$ der vom Parameter t abhängige betragsgrößte Eigenwert der Matrix N gemeint ist

$$|\lambda_{\max}(N)| = \max \{ |\lambda| \mid \lambda \text{ ist Eigenwert von } N \}.$$

3. Für $N \in L_{\infty}^{n \times n}((0, \infty))$ mit symmetrischer und positiv semidefiniter Matrix $N = N(t)$ gilt

$$\|N\|_{(\infty)} = \operatorname{ess\,sup}_{t \in (0, \infty)} \lambda_{\max}(N)(t).$$

Dabei ist $\lambda_{\max}(N) = \lambda_{\max}(N)(t)$ der vom Parameter t abhängige größte Eigenwert der Matrix N .

In jedem Fall erhalten wir die Gültigkeit der Abschätzung in (2.22).

Die in dem vorangegangenen Lemma gewonnenen Gleichheiten werden wir an späterer Stelle nutzen, um die Voraussetzungen bzw. Anforderungen an die in unsere Aufgabe optimaler Steuerung eingehenden Matrixfunktionen verifizieren zu können.

2.3 Gewichtete Lebesgue- und Sobolev-Räume

2.3.1 Die Räume $L_p(\Omega, \omega)$, $W_p^k(\Omega, S_k)$ und $W_p^k(\Omega, \omega)$

In diesem Abschnitt werden die gewichteten Funktionenräume vorgestellt. Dafür benötigen wir zunächst den Begriff eines Gewichtes.

Definition 2.25 ([KO84], S. 538, 1.1 Definition)

Unter einer Gewichtsfunktion $\omega : \Omega \rightarrow \mathbb{R}$ verstehen wir eine endliche Funktion, die auf Ω fast überall positiv und messbar ist. Die Menge aller Gewichtsfunktionen auf Ω bezeichnen wir mit $\mathcal{W}(\Omega)$.

Damit können wir zunächst die gewichteten Lebesgue-Räume entsprechend den Ausführungen in [KO84], vgl. S. 538, 1.2 Definition, einführen.

Definition 2.26

Sei $1 \leq p < \infty$ und $\omega \in \mathcal{W}(\Omega)$. Dann lässt sich der gewichtete Lebesgue-Raum $L_p(\Omega, \omega)$ durch die Vorschrift

$$L_p(\Omega, \omega) := \left\{ x \in \mathcal{M}(\Omega) : \int_{\Omega} |x(t)|^p \omega(t) dt < \infty \right\}$$

definieren und mit der Norm

$$\|x\|_{p,\omega} := \left(\int_{\Omega} |x(t)|^p \omega(t) dt \right)^{1/p}$$

versehen.

Wählen wir $\omega(t) \equiv 1$, so erhalten wir offensichtlich die „klassischen“ Lebesgue-Räume $L_p(\Omega)$, also $L_p(\Omega) := L_p(\Omega, 1)$. Für die gewichteten Lebesgue-Räume können wir die nachfolgenden Charakteristika festhalten:

- Für $1 \leq p < \infty$ und mit der Norm $\|\cdot\|_{p,\omega}$ sind die Räume $L_p(\Omega, \omega)$ Banachräume, vgl. [DS88], S. 146, 6 Theorem.
- Für $1 < p < \infty$ sind die Räume $L_p(\Omega, \omega)$ reflexiv, vgl. [DS88], S. 288, 2 Corollary.
- Mit dem Skalarprodukt

$$\langle x, y \rangle_{2,\omega} := \int_{\Omega} x(t)y(t)\omega(t)dt \quad (2.23)$$

wird der $L_2(\Omega, \omega)$ zu einem Hilbertraum, vgl. [DS88], S. 243.

Wie bei den Räumen $W_p^k(\Omega)$ möchten wir die gewichteten Sobolev-Räume mit Hilfe der schwachen partiellen Ableitungen definieren. Vorweg benötigen wir dafür noch die nachfolgende Bezeichnung: Sei $k \in \mathbb{N}_0$ und $\alpha \in \mathbb{N}_0^s$ ein Multiindex der Ordnung $|\alpha| \leq k$. Dann bezeichnen wir mit S_k eine Menge von Gewichtsfunktionen definiert auf Ω , welche die Gestalt

$$S_k := \{\omega_\alpha \in \mathcal{W}(\Omega), \forall \alpha : 0 \leq |\alpha| \leq k\} \quad (2.24)$$

besitzt.

Definition 2.27 ([KO84], S. 540)

Für $1 \leq p < \infty$ definieren wir den gewichteten Sobolev-Raum $W_p^k(\Omega, S_k)$ als die Menge aller Funktionen

$$x \in L_p(\Omega, \omega_0) \cap L_{1,\text{loc}}(\Omega), \quad (2.25)$$

die für jeden Multiindex $\alpha \in \mathbb{N}_0^s$ mit $|\alpha| \leq k$ eine schwache partielle Ableitung $\mathcal{D}^\alpha x$ bis zur Ordnung k besitzen, sodass

$$\mathcal{D}^\alpha x \in L_p(\Omega, \omega_\alpha) \cap L_{1,\text{loc}}(\Omega) \quad (2.26)$$

gilt. Durch

$$\|x\|_{k,p,S_k} := \left(\sum_{0 \leq |\alpha| \leq k} \|\mathcal{D}^\alpha x\|_{p,\omega_\alpha}^p \right)^{1/p} \quad (2.27)$$

wird eine Norm auf dem $W_p^k(\Omega, S_k)$ eingeführt.

Sind sämtliche in der Menge S_k enthaltenen Gewichte identisch, so werden sowohl die Funktion selbst als auch deren schwache partielle Ableitungen mit ein und derselben Gewichtsfunktion ω gewichtet. In diesem Fall sprechen wir dann auch von einem uniform gewichteten Sobolev-Raum. Obwohl dieser Sobolev-Raum mit der obigen Definition bereits erfasst ist, formulieren wir für eine bessere Verständlichkeit und Unterscheidung eine eigene Definition.

Definition 2.28

Es sei $\omega \in \mathcal{W}(\Omega)$. Für $1 \leq p < \infty$ definieren wir den uniform gewichteten Sobolev-Raum $W_p^k(\Omega, \omega)$ als die Menge aller Funktionen

$$x \in L_p(\Omega, \omega) \cap L_{1,\text{loc}}(\Omega),$$

die für jeden Multiindex $\alpha \in \mathbb{N}_0^s$ mit $|\alpha| \leq k$ eine schwache partielle Ableitung $\mathcal{D}^\alpha x$ bis zur Ordnung k besitzen, sodass

$$\mathcal{D}^\alpha x \in L_p(\Omega, \omega) \cap L_{1,\text{loc}}(\Omega)$$

gilt. Durch

$$\|x\|_{k,p,\omega} := \left(\sum_{0 \leq |\alpha| \leq k} \|\mathcal{D}^\alpha x\|_{p,\omega}^p \right)^{1/p}$$

wird eine Norm auf dem $W_p^k(\Omega, \omega)$ eingeführt.

Die nachfolgenden Ausführungen beziehen sich auf im Allgemeinen auf den gewichteten Sobolev-Raum $W_p^k(\Omega, S_k)$, gelten aber natürlich auch entsprechend für $W_p^k(\Omega, \omega)$.

Im Hinblick auf eine sinnvolle Nutzung der definierten gewichteten Sobolev-Räume stellt sich sofort die Frage, ob diese Banachräume sind. Im Allgemeinen ist dies nicht für beliebige Gewichtsfunktionen der Fall, jedoch gibt das nachfolgende Lemma einen ersten Hinweis darüber, für welche Klassen von Gewichtsfunktionen wir die Vollständigkeit sichern können.

Lemma 2.29 ([KO84], S. 539, 1.6. Corollary)

Sei $p > 1$ und $\omega \in \mathcal{W}(\Omega)$ eine Gewichtsfunktion mit $\omega^{-1/(p-1)} \in L_{1,\text{loc}}(\Omega)$, dann gilt

$$L_p(\Omega, \omega) \subset L_{1,\text{loc}}(\Omega). \tag{2.28}$$

Die Gültigkeit der Inklusion in (2.28) ist essentiell an die Bedingung $\omega^{-1/(p-1)} \in L_{1,\text{loc}}(\Omega)$ geknüpft. Ist $\omega^{-1/(p-1)} \notin L_{1,\text{loc}}(\Omega)$, so muss die Inklusion nicht erfüllt sein. Setzen wir jedoch das Erfülltsein der angegebenen Bedingung voraus, so macht die Definition schwacher partieller Ableitungen auch für Funktionen $x \in L_p(\Omega, \omega)$ Sinn und wir können festhalten:

Folgerung 2.2. Sei $p > 1$. Erfüllt eine Gewichtsfunktion $\omega_\alpha \in S_k$ die Bedingung $\omega_\alpha^{-1/(p-1)} \in L_{1,\text{loc}}(\Omega)$, so können wir in der Definition 2.27 die entsprechende Forderung in (2.25) bzw. in (2.26) durch $x \in L_p(\Omega, \omega_0)$ bzw. durch $\mathcal{D}^\alpha x \in L_p(\Omega, \omega_\alpha)$ ersetzen.

Analog zu den klassischen Räumen erhalten wir als direkte Folgerung des Lemmas 2.29 die nachfolgende Inklusion.

Folgerung 2.3. *Sei $p > 1$ und $\omega \in \mathcal{W}(\Omega)$ eine Gewichtsfunktion mit $\omega^{-1/(p-1)} \in L_{1,loc}(\Omega)$, dann gilt*

$$W_p^k(\Omega, \omega) \subset W_{1,loc}^k(\Omega).$$

Damit sind wir jetzt in der Lage die Aussagen des Satzes 2.4 auf gewichtete Sobolev-Funktionen zu erweitern.

Folgerung 2.4. *Sei $I \subseteq \mathbb{R}, p > 1$, $\omega \in \mathcal{W}(I)$ eine Gewichtsfunktion mit $\omega^{-1/(p-1)} \in L_{1,loc}(I)$ und $x \in W_p^1(I, \omega)$. Dann existiert ein lokal absolut stetiger Repräsentant $\tilde{x} \in AC_{loc}(I)$, sodass $x = \tilde{x}$ und $\mathcal{D}x = \tilde{x}'$ fast überall auf I gelten.*

Beweis: Nach Voraussetzung können wir von $x \in W_p^1(I, \omega)$ direkt auf $x \in W_{1,loc}^1(I)$ schließen, vgl. Folgerung 2.3. Anwendung des Satzes 2.4 liefert dann die gewünschten Eigenschaften.

Weiter sind wir nun insbesondere in der Lage, die Gewichte des Sobolev-Raumes dahingehend zu charakterisieren, sodass wir einen vollständigen Raum erhalten.

Satz 2.30 ([KO84], S. 540, 1.11 Theorem)

Ist $1 < p < \infty$ und $\omega_\alpha^{-1/(p-1)} \in L_{1,loc}(\Omega)$ für alle $\omega_\alpha \in S_k$ erfüllt, dann ist der $W_p^k(\Omega, S_k)$ ein Banachraum versehen mit der $\|x\|_{k,p,S_k}$ -Norm aus (2.27).

Insbesondere wird der $W_2^k(\Omega, S_k)$ unter den Voraussetzungen des obigen Satzes zu einem Hilbertraum versehen mit dem Skalarprodukt

$$\langle x, y \rangle_{k,2,S_k} := \sum_{0 \leq |\alpha| \leq k} \langle \mathcal{D}^\alpha x, \mathcal{D}^\alpha y \rangle_{2,\omega_\alpha},$$

wobei $\langle \cdot, \cdot \rangle_{2,\omega_\alpha}$ das in (2.23) definierte Skalarprodukt im $L_2(\Omega, \omega_\alpha)$ meint.

Bemerkung 2.7. *In dem Fall des uniform-gewichteten Sobolev-Raumes $W_2^k(\Omega, \omega)$ nutzen wir die Bezeichnung $\langle x, y \rangle_{k,2,\omega}$ und meinen damit entsprechend das uniform-gewichtete Skalarprodukt*

$$\langle x, y \rangle_{k,2,\omega} := \sum_{0 \leq |\alpha| \leq k} \langle \mathcal{D}^\alpha x, \mathcal{D}^\alpha y \rangle_{2,\omega}.$$

In dem Abschnitt zu den „klassischen“ Sobolev-Räumen haben wir den Raum $H^k(\Omega)$ über einen Vervollständigkeitsprozess erhalten und haben gesehen, dass dieser mit dem Raum $W_p^k(\Omega)$ zusammenfällt, vgl. Satz 2.6. Man könnte sich nun die Frage stellen, ob ein ähnliches Vorgehen bei den gewichteten Sobolev-Räumen sinnvoll erscheint. Jedoch treten bereits bei dem Studieren der Inklusion

$$C^\infty(\bar{\Omega}) \subset W_p^k(\Omega, S_k)$$

Schwierigkeiten auf. Diese Relation schließt eine große Anzahl an Gewichtsfunktionen aus und selbst wenn sie für bestimmte Gewichtsfunktionen gültig ist, kann die Abschlussbildung zu Räumen führen, die nichtreguläre Distributionen oder Funktionen mit nichtregulären schwachen Ableitungen enthalten. Beispiele dafür sind in [KO84] angegeben, vgl. Beispiel 1.12, S. 541 und Beispiel 2.5, S. 547 f.

Den gewichteten Raum $\mathring{W}_p^k(\Omega, S_k)$ hingegen können wir unter geeigneten Voraussetzungen an die Gewichtsfunktionen durch eine Abschlussbildung sinnvoll definieren.

Lemma 2.31 ([KO84], S. 551, 4.4 Lemma)

Die Inklusion

$$C_0^\infty(\Omega) \subset W_p^k(\Omega, S_k)$$

ist genau dann erfüllt, wenn $\omega_\alpha \in L_{1,loc}(\Omega)$ für alle $\omega_\alpha \in S_k$ gilt.

Mit Hilfe dieses Lemmas haben wir nun die Möglichkeit, den gewichteten Sobolev-Raum $\mathring{W}_p^k(\Omega, S)$ wie folgt einzuführen:

Definition 2.32 ([KO84], S. 551)

Sei $p > 1$ und die Menge S_k wie in (2.24) festgelegt. Weiter erfüllen alle Gewichtsfunktion $\omega_\alpha \in S_k$ die folgenden Bedingungen:

- (i) $\omega_\alpha \in L_{1,loc}(\Omega)$,
- (ii) $\omega_\alpha^{-1/(p-1)} \in L_{1,loc}(\Omega)$.

Dann definieren wir den gewichteten Sobolev-Raum $\mathring{W}_p^k(\Omega, S_k)$ durch den Abschluss von $C_0^\infty(\Omega)$ bezüglich der Norm $\|\cdot\|_{k,(p),S_k}$.

Setzen wir $\omega_\alpha \equiv 1$ uniform für alle α , so erhalten wir die „klassischen“ Sobolev-Räume $\mathring{W}_p^k(\Omega) := \mathring{W}_p^k(\Omega, 1)$ und $W_p^k(\Omega) := W_p^k(\Omega, 1)$. In diesem Sinne können wir die gewichteten Sobolev-Räume als eine Verallgemeinerung der „klassischen“ Sobolev-Räume auffassen.

Wie bei den „klassischen“ Räumen sind wir daran interessiert, in welcher Art und Weise wir Randwerte einer gewichteten Sobolev-Funktion interpretieren können. Dafür zeigen wir zunächst die Beziehungen zwischen den ungewichteten und gewichteten Sobolev-Räumen auf.

Lemma 2.33 ([Fau09], S. 13, Satz 3.21.)

Sei S_k wie in (2.24) festgelegt.

1. *Gilt $\omega_\alpha \in L_\infty(\Omega)$ für alle $\omega_\alpha \in S_k$, dann erhalten wir die Inklusion*

$$W_p^k(\Omega) \subseteq W_p^k(\Omega, S_k).$$

2. *Ist $\omega_\alpha(t) \geq \kappa > 0$ f.ü. in Ω für alle $\omega_\alpha \in S_k$ erfüllt, so gilt*

$$W_p^k(\Omega, S_k) \subseteq W_p^k(\Omega).$$

Wir können also Klassen von Gewichtsfunktionen angeben, für die die gewichteten Sobolev-Räume mit den ungewichteten Räumen zusammenfallen. Konzentrieren wir uns wieder auf den Fall eines Intervalls $\Omega := I = (a, b)$, $-\infty < a < b < \infty$, so können wir aus dem Satz 2.8 bzw. der Bemerkung 2.1 auf die Existenz eines stetigen Repräsentanten schließen, welcher bis auf $[a, b]$ stetig fortgesetzt werden kann.

Folgerung 2.5. *Sei $k \geq 1$, $1 < p < \infty$ und S_k eine Menge von Gewichtsfunktion, sodass die Äquivalenz $W_p^k((a, b)) = W_p^k((a, b), S_k)$ gilt. Dann existiert für jede Äquivalenzklasse $[x]$ aus $W_p^k((a, b), S_k)$ ein $(k - 1)$ -mal stetig differenzierbarer Repräsentant $x^* \in C^{k-1}([a, b])$.*

Wie bei den „klassischen“ Räumen sind wir damit nun auch in der Lage, Sobolev-Funktionen aus dem gewichteten Raum $W_p^k((a, b), S_k)$ punktweise mit Hilfe eines geeigneten Vertreters der entsprechenden Äquivalenzklasse auf dem abgeschlossenen Intervall $[a, b]$ auszuwerten und insbesondere Randwerte $x(a)$ und $x(b)$ sinnvoll zu interpretieren.

2.3.2 Die Räume $L_p^n(\Omega, \omega)$ und $W_p^{k,n}(\Omega, \omega)$

In Analogie zu Kapitel 2.2.2 definieren wir aufbauend auf Satz 2.12 die zu den gewichteten Lebesgue- und Sobolev-Räumen korrespondierenden Produkträume.

Hierbei können die Überlegungen aus Kapitel 2.2.2 eins-zu-eins übernommen werden, sodass wir im Folgenden die resultierenden Ergebnisse nur in einem zusammenfassenden Charakter präsentieren werden.

Für die gewichteten Lebesgue-Räume notieren wir die folgenden Aussagen.

Definition 2.34

Es sei $\omega \in \mathcal{W}(\Omega)$, sowie $1 \leq p < \infty$. Der Raum

$$L_p^n(\Omega, \omega) := \left\{ x \in \mathcal{M}^n(\Omega) \mid \int_{\Omega} |x(t)|_2^p \omega(t) dt < \infty \right\}$$

wird als gewichteter vektorwertiger Lebesgue-Raum der Ordnung p bezeichnet.

Lemma 2.35

Es sei $\omega \in \mathcal{W}(\Omega)$ und $1 \leq p < \infty$. Der Raum $L_p^n(\Omega, \omega)$ versehen mit der Norm

$$\|x\|_{(p), \omega} = \left(\int_{\Omega} |x(t)|_2^p \omega(t) dt \right)^{1/p}$$

genügt den folgenden Eigenschaften:

1. $L_p^n(\Omega, \omega)$ ist für $1 \leq p < \infty$ ein Banachraum.
2. $L_p^n(\Omega, \omega)$ ist für $1 < p < \infty$ ein reflexiver Raum.

3. $L_2^n(\Omega, \omega)$ wird mit dem Skalarprodukt

$$\langle x, y \rangle_{(2), \omega} = \int_{\Omega} x^T(t)y(t)\omega(t)dt$$

ein Hilbertraum.

4. Gilt $\omega^{-1/(p-1)} \in L_{1,loc}(\Omega)$ mit $p > 1$, so folgt für jede kompakte Teilmenge $\Omega_0 \subseteq \Omega$ die Inklusion $L_p^n(\Omega_0, \omega) \subseteq L_1^n(\Omega_0)$.¹¹

Für die gewichteten Sobolev-Räume erhalten wir die nachfolgende Definition:

Definition 2.36

Sei $1 < p < \infty$ und $k \in \mathbb{N}_0$. Weiter sei $\omega \in \mathcal{W}(\Omega)$ ein Gewicht mit der zusätzlichen Eigenschaft $\omega^{-1/(p-1)} \in L_{1,loc}(\Omega)$. Der Raum

$$W_p^{k,n}(\Omega, \omega) := \left\{ x \in L_p^n(\Omega, \omega) \mid \mathcal{D}^\alpha x \in L_p^n(\Omega, \omega) \text{ für alle } \alpha \in \mathbb{N}_0^s \text{ mit } |\alpha| \leq k \right\}$$

heißt gewichteter Sobolev-Raum der vektorwertigen Funktionen.

Lemma 2.37

Sei $1 < p < \infty$ und $k \in \mathbb{N}_0$. Weiter sei $\omega \in \mathcal{W}(\Omega)$ ein Gewicht mit der zusätzlichen Eigenschaft $\omega^{-1/(p-1)} \in L_{1,loc}(\Omega)$. Der Raum $W_p^{k,n}(\Omega)$ versehen mit der Norm

$$\|x\|_{k,(p),\omega} := \left(\|x\|_{(p),\omega}^p + \sum_{0 < |\alpha| \leq k} \|\mathcal{D}^\alpha x\|_{(p),\omega}^p \right)^{1/p}$$

ist

1. ein Banachraum.
2. für $p = 2$ mit dem Skalarprodukt

$$\langle x, y \rangle_{k,(2),\omega} := \langle x, y \rangle_{(2),\omega} + \sum_{0 < |\alpha| \leq k} \langle \mathcal{D}^\alpha x, \mathcal{D}^\alpha y \rangle_{(2),\omega}$$

ein Hilbertraum.

3. Ist $\Omega := I \subseteq \mathbb{R}$, so existiert zu $x \in W_p^{k,n}(I, \omega)$, $k \geq 1$ ein lokal absolut stetiger Repräsentant $\tilde{x} \in AC_{loc}(I, \mathbb{R}^n)$.¹²

Ist $\omega \in \mathcal{W}(\Omega)$ und gilt $\omega^{-1/(p-1)} \in L_{1,loc}(\Omega)$ mit $p > 1$, so gilt die Inklusion $L_p^n(\Omega, \omega) \subseteq L_{1,loc}^n(\Omega)$. Daraus folgt sofort auch $W_p^{k,n}(\Omega, \omega) \subseteq W_{1,loc}^{k,n}(\Omega)$, womit sich die letzte Eigenschaft des vorangegangenen Lemmas direkt durch Anwendung des Satzes 2.4 ergibt.

¹¹ Mit anderen Worten, es gilt die Inklusion $L_p^n(\Omega, \omega) \subseteq L_{1,loc}^n(\Omega)$.

¹² Für $\Omega \subseteq \mathbb{R}^s$ mit $s > 1$ kann unter dem Begriff „absolute continuous on lines“ eine ähnlich Aussage gefunden werden, vgl.[KJF77], S. 274, Theorem 5.6.3.

2.4 Die Räume $L_2((0, \infty), \omega_{\alpha, \beta})$, $W_2^k((0, \infty), \omega_\beta)$ und $W_2^k((0, \infty), \mathbb{W}_{k, \beta})$

Wir formulieren an späterer Stelle Steuerungsprobleme auf einem unendlichen Zeithorizont in geeigneten gewichteten Lebesgue- und Sobolev-Räumen. Daher sind der Fall $\Omega = (0, \infty)$ und eine spezielle Klasse von Gewichtsfunktionen für uns von besonderem Interesse. In diesem Abschnitt untersuchen wir diese Gewichte genauer und stellen die Definitionen der entsprechend gewichteten Funktionenräume vor.

2.4.1 Die Gewichte $\omega_{\alpha, \beta}(t) := t^\alpha e^{-\beta t}$

Wir studieren eine besondere Klasse von Gewichtsfunktionen und konzentrieren uns dabei auf Gewichte der Gestalt

$$\omega_{\alpha, \beta}(t) := t^\alpha e^{-\beta t}, \quad \alpha \in \mathbb{N}_0, \beta \in \mathbb{R}, \beta > 0. \quad (2.29)$$

Offenbar gilt $\omega_{\alpha, \beta}(t) > 0$ für alle $t \in (0, \infty)$ und insbesondere sind diese Funktionen sogar beliebig oft auf $(0, \infty)$ stetig differenzierbar. Ist $\alpha > 0$, so integrieren wir partiell und erhalten

$$\int_0^\infty \omega_{\alpha, \beta}(t) dt = \int_0^\infty t^\alpha e^{-\beta t} dt = \lim_{T \rightarrow +\infty} \left[-\frac{1}{\beta} t^\alpha e^{-\beta t} \right]_0^T + \frac{\alpha}{\beta} \int_0^\infty t^{\alpha-1} e^{-\beta t} dt.$$

Der Grenzwert ergibt sich zu

$$\lim_{T \rightarrow +\infty} \left[-\frac{1}{\beta} t^\alpha e^{-\beta t} \right]_0^T = \lim_{T \rightarrow +\infty} \left(-\frac{1}{\beta} T^\alpha e^{-\beta T} \right) = \lim_{T \rightarrow +\infty} \left(-\frac{1}{\beta} \cdot \frac{T^\alpha}{e^{\beta T}} \right)$$

und mehrfaches Anwenden (α -mal) der Regel von L'Hospital liefert

$$\lim_{T \rightarrow +\infty} \left(-\frac{1}{\beta} \cdot \frac{T^\alpha}{e^{\beta T}} \right) \stackrel{[\frac{\infty}{\infty}]}{=} \lim_{T \rightarrow +\infty} \left(-\frac{1}{\beta} \cdot \frac{\alpha}{\beta} \cdot \frac{T^{\alpha-1}}{e^{\beta T}} \right) \stackrel{[\frac{\infty}{\infty}]}{=} \dots \stackrel{[\frac{\infty}{\infty}]}{=} \lim_{T \rightarrow +\infty} \left(-\frac{\alpha!}{\beta^{\alpha+1}} \cdot \frac{1}{e^{\beta T}} \right) = 0.$$

Damit erhalten wir

$$\int_0^\infty \omega_{\alpha, \beta}(t) dt = \frac{\alpha}{\beta} \int_0^\infty t^{\alpha-1} e^{-\beta t} dt.$$

Wenden wir die partielle Integration insgesamt α -mal an, so ergibt sich

$$\int_0^\infty \omega_{\alpha, \beta}(t) dt = \frac{\alpha!}{\beta^\alpha} \int_0^\infty e^{-\beta t} dt = \frac{\alpha!}{\beta^\alpha} \cdot \lim_{T \rightarrow +\infty} \left[-\frac{1}{\beta} e^{-\beta t} \right]_0^T = \frac{\alpha!}{\beta^{\alpha+1}} < \infty. \quad (2.30)$$

Ist $\alpha = 0$, so gilt

$$\int_0^\infty \omega_{0, \beta}(t) dt = \int_0^\infty e^{-\beta t} dt = \lim_{T \rightarrow +\infty} \left[-\frac{1}{\beta} e^{-\beta t} \right]_0^T = \frac{1}{\beta} < \infty.$$

Die Gewichte $\omega_{\alpha, \beta}$ sind somit Lebesgue-integrierbar über $(0, \infty)$, also erst recht Funktionen aus dem Raum $L_{1, \text{loc}}((0, \infty))$. Darüber hinaus ist für $p > 1$ auch die Bedingung (ii) in der Definition 2.32 erfüllt, denn für jedes abgeschlossene Intervall $[a, b] \subset (0, \infty)$ gilt

$$\int_a^b \omega_{\alpha, \beta}^{-1/(p-1)}(t) dt = \int_a^b t^{-\alpha/(p-1)} \cdot e^{\beta t/(p-1)} dt < \infty$$

und somit auch $\omega_{\alpha, \beta}^{-1/(p-1)} \in L_{1, \text{loc}}((0, \infty))$.

Schlussendlich möchten wir anmerken, dass die hier betrachteten Gewichte aufgrund der diskutierten Eigenschaften zu einer speziellen Klasse von Gewichtsfunktionen, nämlich den *Dichtefunktionen*, zählen. In Anlehnung an [Mag82], S. 402, 3.1 nennen wir eine Gewichtsfunktion $\omega \in \mathcal{W}((0, \infty))$, für die zusätzlich

$$\int_0^\infty \omega(t) dt < \infty$$

gilt, eine Dichtefunktion, kurz Dichte. Da die hier betrachteten Gewichte $\omega_{\alpha, \beta} \in C(\mathbb{R}_+)$ Lebesgue-integrierbar über \mathbb{R}_+ sind, demnach also die geforderte Eigenschaft erfüllen, können sie auch als Dichte bezeichnet werden, vgl. [Pic14].

2.4.2 Definition und Eigenschaften

Im Mittelpunkt der nachfolgenden Betrachtungen stehen die mit Gewichten der Form (2.29) gewichteten Lebesgue- und Sobolev-Räume $L_2((0, \infty), \omega_{\alpha, \beta})$ und $W_2^k((0, \infty), \omega_\beta)$.

Resultierend aus den Ergebnissen der vorangegangenen Abschnitte möchten wir die Definitionen dieser Räume und deren Eigenschaften in diesem Abschnitt zusammenfassend angeben.

Wir stellen zunächst die gewichteten Lebesgue-Räume $L_2((0, \infty), \omega_{\alpha, \beta})$ vor.

Definition 2.38

Sei $\omega_{\alpha, \beta}(t) := t^\alpha e^{-\beta t}$, $\alpha \in \mathbb{N}_0$, $\beta \in \mathbb{R}$, $\beta > 0$ gegeben. Dann definieren wir den gewichteten Lebesgue-Raum $L_2((0, \infty), \omega_{\alpha, \beta})$ gemäß

$$L_2((0, \infty), \omega_{\alpha, \beta}) := \left\{ x \in \mathcal{M}((0, \infty)) \mid \int_0^\infty |x(t)|^2 \omega_{\alpha, \beta}(t) dt < \infty \right\}$$

und versehen ihn mit der Norm

$$\|x\|_{2,\omega_{\alpha,\beta}} := \left(\int_0^\infty |x(t)|^2 \omega_{\alpha,\beta}(t) dt \right)^{1/2}.$$

Lemma 2.39

Der Raum $L_2((0, \infty), \omega_{\alpha,\beta})$ ist, versehen mit dem Skalarprodukt

$$\langle x, y \rangle_{2,\omega_{\alpha,\beta}} := \int_0^\infty x(t)y(t)\omega_{\alpha,\beta}(t)dt,$$

ein Hilbertraum.

In Anlehnung an Definition 2.28 führen wir den Sobolev-Raum $W_2^k((0, \infty), \omega_\beta)$ ein. Dabei ist $\omega_\beta(t) := \omega_{0,\beta}(t) = e^{-\beta t}$ mit $\beta \in \mathbb{R}, \beta > 0$, vgl. (2.29) mit $\alpha = 0$.

Definition 2.40

Sei $k \in \mathbb{N}_0, \alpha \in \mathbb{N}_0$ und $\omega_\beta(t) := e^{-\beta t}$ mit $\beta \in \mathbb{R}, \beta > 0$. Wir bezeichnen den Raum

$$W_2^k((0, \infty), \omega_\beta) := \{x \in L_2((0, \infty), \omega_\beta) \mid \mathcal{D}^\alpha x \in L_2((0, \infty), \omega_\beta) \forall \alpha : \alpha \leq k\}$$

als den bezüglich ω_β (uniform) gewichteten Sobolev-Raum und versehen ihn mit der Norm

$$\|x\|_{k,2,\omega_\beta} := \left(\|x\|_{2,\omega_\beta}^2 + \sum_{0 < \alpha \leq k} \|\mathcal{D}^\alpha x\|_{2,\omega_\beta}^2 \right)^{1/2}.$$

Die in dem Raum $W_2^k((0, \infty), \omega_\beta)$ enthaltenen Funktionen und deren schwache partielle Ableitungen sind uniform, im Sinne von mit identischen Gewichtsfunktionen ω_β , gewichtet.

Lemma 2.41

Der Raum $W_2^k((0, \infty), \omega_\beta)$ wird mit dem Skalarprodukt

$$\langle x, y \rangle_{k,2,\omega_\beta} := \langle x, y \rangle_{2,\omega_\beta} + \sum_{0 < \alpha \leq k} \langle \mathcal{D}^\alpha x, \mathcal{D}^\alpha y \rangle_{2,\omega_\beta}$$

zu einen Hilbertraum. Darüber hinaus existiert zu jedem $x \in W_2^k((0, \infty), \omega_\beta), k \geq 1$ ein Repräsentant $\tilde{x} \in AC_{loc}((0, \infty))$.

Wir stellen noch den bezüglich der Gewichte aus (2.29) nicht-uniform gewichteten Sobolev-Raum $W_2^k((0, \infty), \mathfrak{w}_{k,\beta})$ vor. Im Unterschied zu den (gleichmäßig) gewichteten Sobolev-Raum $W_2^k((0, \infty), \omega_\beta)$ werden hier die schwachen Ableitungen mit verschiedenen Gewichten des Typs $\omega_{\alpha,\beta}$ versehen. Wie in der Definition 2.27 legen wir für die Definition dieses nicht-uniform gewichteten Raumes eine Menge von Gewichtsfunktionen zugrunde, vgl. (2.24). Ganz konkret betrachten wir dabei eine Menge, die ausschließlich aus Gewichtsfunktionen der Form (2.29) besteht. Um dies zu verdeutlichen, nutzen wir deshalb im Folgenden die Bezeichnung

$\mathfrak{w}_{k, \beta}$ und definieren mit $k \in \mathbb{N}_0$

$$\mathfrak{w}_{k, \beta} := \{\omega_{\alpha, \beta}(t) := t^\alpha e^{-\beta t}, \beta \in \mathbb{R}, \beta > 0 \mid \forall \alpha \in \mathbb{N}_0 : 0 \leq \alpha \leq k\}.$$

Damit können wir den Sobolev-Raum $W_2^k((0, \infty), \mathfrak{w}_{k, \beta})$ wie folgt erklären:

Definition 2.42

Sei $k \in \mathbb{N}_0$ und $\alpha \in \mathbb{N}_0$. Wir bezeichnen den Raum

$$W_2^k((0, \infty), \mathfrak{w}_{k, \beta}) := \{x \in L_2((0, \infty), \omega_\beta) \mid \mathcal{D}^\alpha x \in L_2((0, \infty), \omega_{\alpha, \beta}) \forall \alpha : \alpha \leq k\}.$$

als den bezüglich der Menge $\mathfrak{w}_{k, \beta}$ (nicht-uniform) gewichteten Sobolev-Raum. Durch

$$\|x\|_{k, 2, \mathfrak{w}_{k, \beta}} := \left(\|x\|_{2, \omega_\beta}^2 + \sum_{0 < \alpha \leq k} \|\mathcal{D}^\alpha x\|_{2, \omega_{\alpha, \beta}}^2 \right)^{1/2}$$

wird eine Norm auf dem Raum $W_2^k((0, \infty), \mathfrak{w}_{k, \beta})$ eingeführt.

Lemma 2.43

Mit dem Skalarprodukt

$$\langle x, y \rangle_{k, 2, \mathfrak{w}_{k, \beta}} := \langle x, y \rangle_{2, \omega_\beta} + \sum_{0 < \alpha \leq k} \langle \mathcal{D}^\alpha x, \mathcal{D}^\alpha y \rangle_{2, \omega_{\alpha, \beta}}$$

wird der $W_2^k((0, \infty), \mathfrak{w}_{k, \beta})$ zu einem Hilbertraum. Zu jedem $x \in W_2^k((0, \infty), \mathfrak{w}_{k, \beta})$, $k \geq 1$ existiert ein lokal absolut stetiger Repräsentant $\tilde{x} \in AC_{loc}((0, \infty))$.

Um die Unterschiede zwischen den Räumen deutlich zu machen, geben wir abschließend die angesprochenen Räume beispielhaft für den Fall $k = 2$ an.

Beispiel Der gewichtete Sobolev-Raum $W_2^2((0, \infty), \omega_\beta)$ ist der Raum aller Funktionen $x \in L_2((0, \infty), \omega_\beta)$, für welche die schwachen Ableitungen bis zur Ordnung 2 existieren und $\mathcal{D}^1 x, \mathcal{D}^2 x \in L_2((0, \infty), \omega_\beta)$ gilt. Versehen mit der Norm

$$\|x\|_{2, 2, \omega_\beta} = \left(\|x\|_{2, \omega_\beta}^2 + \|\mathcal{D}^1 x\|_{2, \omega_\beta}^2 + \|\mathcal{D}^2 x\|_{2, \omega_\beta}^2 \right)^2 \quad (2.31)$$

wird der $W_2^2((0, \infty), \omega_\beta)$ zu einem Hilbertraum. Der Raum $W_2^2((0, \infty), \mathfrak{w}_{2, \beta})$ hingegen ist der Raum aller Funktionen $x \in L_2((0, \infty), \omega_\beta)$, für welche die schwachen Ableitungen bis zur Ordnung 2 existieren, die schwache Ableitung $\mathcal{D}^1 x$ jedoch zu dem Raum $L_2((0, \infty), \omega_{1, \beta})$ gehört und $\mathcal{D}^2 x$ in dem Raum $L_2((0, \infty), \omega_{2, \beta})$ liegt. Der nicht-uniform gewichtete Raum $W_2^2((0, \infty), \mathfrak{w}_{2, \beta})$ wird mit der Norm

$$\|x\|_{2, 2, \mathfrak{w}_{2, \beta}} = \left(\|x\|_{2, \omega_\beta}^2 + \|\mathcal{D}^1 x\|_{2, \omega_{1, \beta}}^2 + \|\mathcal{D}^2 x\|_{2, \omega_{2, \beta}}^2 \right)^2$$

zu einen Hilbertraum. Es ist zu beachten, dass hier im Unterschied zu (2.31) verschiedene Gewichte auftreten.

2.5 Asymptotische Steuerbarkeit dynamischer Systeme

Die Frage, ob ein gegebenes dynamisches System in vorgegebener Weise von außen aktiv durch eine Steuergröße, welche sich in Abhängigkeit von der Zeit oder von dem Systemzustand verändern kann, beeinflusst werden kann, ist Gegenstand der Steuerungs- oder auch Kontrolltheorie (aus dem Englischen „control theory“). Hierbei werden sowohl die Eigenschaften dynamischer Systeme analysiert als auch Möglichkeiten der gezielten Einflussnahme auf diese untersucht. In Abhängigkeit der Zielstellung ergeben sich verschiedene Problem- und Aufgabenstellungen. Eine Aufgabe ist es, einen Gleichgewichtszustand des dynamischen Systems mit Hilfe einer geeigneten Steuerungsfunktion zu stabilisieren. Wie dies im Zusammenhang zur Lösung von Optimalsteuerungsproblemen steht, zeigen wir in diesem Abschnitt auf. Dafür geben wir einen knappen Einblick in die Stabilisierung dynamischer Systeme und führen entsprechende Definitionen ein, wobei wir darauf hinweisen möchten, dass die Inhalte dieses Abschnittes einen Auszug aus unserer Publikation [PB18] darstellen.

Wir beginnen damit, die Stabilitätstheorie nach Lyapunov vorzustellen, vgl. [Lya92]. Dazu betrachten wir ein allgemeines nichtlineares autonomes Differentialgleichungssystem

$$\dot{x}(t) = \varphi(x(t)) \quad (2.32)$$

mit vektorwertigem Zustand $x(t) \in \mathbb{R}^n$, einer stetigen Abbildung $\varphi : D \rightarrow \mathbb{R}^n$ und der offenen Menge $D \subset \mathbb{R}^n$, $0 \in D$. Weiterhin sei x_G ein Gleichgewicht (oder auch Ruhelage, Ruhelösung) von (2.32), das heißt, es ist

$$\varphi(x_G) = 0.$$

Definition 2.44 (vgl. [GJ16], S. 100, Definition 7.3)

1. Das Gleichgewicht x_G von (2.32) heißt stabil im Sinne von Lyapunov, falls es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, sodass für alle Lösungen x von (2.32) mit $\|x(0) - x_G\| < \delta$ die Ungleichung $\|x(t) - x_G\| < \varepsilon$ für alle $t \geq 0$ gilt.
2. Das Gleichgewicht x_G von (2.32) heißt asymptotisch stabil, wenn es stabil im Sinne von Lyapunov ist und außerdem ein $\delta > 0$ mit der folgenden Eigenschaft existiert: Für jede Lösung x von (2.32) mit $\|x(0) - x_G\| \leq \delta$ gilt die Grenzwertbeziehung

$$\lim_{t \rightarrow \infty} \|x(t) - x_G\| = 0.$$

3. Das Gleichgewicht x_G von (2.32) heißt exponentiell stabil, falls Konstanten $\alpha, \beta > 0$ existieren, sodass für die Lösungen x von (2.32) mit $\|x(0) - x_G\| \leq \delta$ die Ungleichung

$$\|x(t) - x_G\| \leq \alpha \|x(0) - x_G\| e^{-\beta t} \quad \forall t \geq 0$$

folgt.

Die vorgestellten Stabilitätsbegriffe lassen sich in natürlicher Weise auf Kontrollsysteme übertragen. Im Unterschied zu dem System (2.32) wird nun versucht, das Systemverhalten mit

Hilfe einer Steuerung zu verändern, um zum Beispiel Gleichgewichte stabilisieren zu können oder den Zustand an eine vorgegebene Referenztrajektorie anzunähern. Dazu betrachten wir ein allgemeines nichtlineares autonomes Kontrollsystem der Form

$$\dot{x}(t) = \tilde{\varphi}(x(t), u(t)) \quad (2.33)$$

mit vektorwertigem Zustand x und vektorwertiger Steuerung u . Dabei ist $\tilde{\varphi} : \tilde{D} \rightarrow \mathbb{R}^n$ mit $\tilde{D} := D \times U$ eine auf \tilde{D} stetige Abbildung, $D \subset \mathbb{R}^n$ ist offen und $U \subset \mathbb{R}^m$ ist abgeschlossen. Weiterhin nennen wir x_G ein Gleichgewicht des Kontrollsystems (2.33), falls $\tilde{\varphi}(x_G, 0) = 0$ gilt, wobei wir o.B.d.A. $0 \in U$ voraussetzen.

Definition 2.45

1. Das Gleichgewicht x_G von (2.33) heißt steuerbar im Sinne von Lyapunov, falls eine Steuerung u existiert, sodass alle Lösungen x von (2.33) stabil im Sinne von Lyapunov sind.
2. Das Gleichgewicht x_G von (2.33) heißt asymptotisch steuerbar, falls eine Steuerung u existiert, sodass alle Lösungen x von (2.33) asymptotisch stabil sind.
3. Das Gleichgewicht x_G von (2.33) heißt exponentiell steuerbar, falls eine Steuerung u existiert, sodass alle Lösungen x von (2.33) exponentiell stabil sind.

Für die Berechnung einer stabilisierenden Steuerung entwickelte Lyapunov die sogenannte *Direkte Methode*, vgl. [Lya92], [Hah59]. Diese basiert auf der Berechnung von sogenannten Lyapunov-Funktionen für nichtlineare Systeme, mit deren Hilfe sich Aussagen über die Stabilität eines Gleichgewichtes gewinnen lassen, ohne die exakte Lösung des Systems genauer zu kennen. Die Schwierigkeit dieser Methode besteht darin, dass es kein allgemeines Konstruktionsverfahren für Lyapunov-Funktionen gibt. In [Kal60] untersuchte Kalman 1960 die Steuerbarkeit dynamischer Systeme im Zusammenhang zu Aufgaben der optimalen Steuerung. Er verfolgte den Zugang, eine asymptotisch stabilisierende Steuerung durch Lösung des folgenden Optimalsteuerungsproblems zu finden:

Minimiere

$$J_{t_1}((x, u); (t_0, \xi)) = \int_{t_0}^{t_1} L(x(t), u(t)) dt + \nu(x(t_1)) \quad (2.34)$$

bezüglich

$$\begin{aligned} \dot{x}(t) &= \tilde{\varphi}(x(t), u(t)) \quad \text{f.ü. auf } (t_0, t_1), \\ x(0) &= \xi \in \mathbb{R}^n, \\ u(t) &\in U \quad \text{f.ü. auf } (t_0, t_1). \end{aligned}$$

Bemerkung 2.8.

1. Es wird vorausgesetzt, dass die Funktion $L : D \times U \rightarrow \mathbb{R}_+$ in (2.34) in all ihren Argumenten zweimal stetig differenzierbar ist. Zusätzlich wird gefordert, dass eine strikte Legendre-Bedingung erfüllt sein muss. Die Hesse Matrix $L_{vv}(\xi, v)$ von L muss bezüglich des zweiten Argumentes v und für alle (ξ, v) strikt positiv sein.
2. Es wird angemerkt, dass $t_1 = \infty$ gewählt werden kann. Welche Interpretation des Integrals im Zielfunktional (2.34) für diesen Fall gewählt wird, bleibt offen.
3. Es wird angenommen, dass die obige Aufgabe optimaler Steuerung eine Lösung besitzt. In welchen Funktionenräumen die zugrunde gelegte Aufgabenstellung diskutiert wird, wird nicht genauer erörtert. Für den Fall eines endlichen Zeithorizontes können die Räume $W_1^{1,n}([t_0, t_1]) \times L_\infty((t_0, t_1))$ herangezogen werden.

Eine Antwort auf die Frage nach einen geeigneten Integralbegriff wird in den Arbeiten von Sonntag gegeben, vgl. [Son83], [Son98]. Hier wird eine Interpretation im Sinne des uneigentlichen Lebesgue-Integrals

$$L_{\lim} - \int_{t_0}^{\infty} f(t)dt = \lim_{t_1 \rightarrow \infty} \int_{t_0}^{t_1} f(t)dt$$

vorgenommen. Der für den endlichen Zeithorizont gewählte Zustandsraum

$$W_1^{1,n}([t_0, t_1]) = AC^n([t_0, t_1])$$

wird zu dem Raum der lokal absolut stetigen Funktionen $AC_{\text{loc}}^n(\mathbb{R}_+)$ erweitert, wodurch wir die folgende Aufgabenstellung erhalten:

Minimiere

$$J_\infty((x, u); (t_0, t_1, \xi)) = L_{\lim} - \int_{t_0}^{\infty} L(x(t), u(t))dt \tag{2.35}$$

bezüglich

$$\begin{aligned} (x, u) &\in AC_{\text{loc}}^n(\mathbb{R}_+) \times L_{1,\text{loc}}(\mathbb{R}_+), \\ \dot{x}(t) &= \tilde{\varphi}(x(t), u(t)) \quad \text{f.ü. auf } \mathbb{R}_+, \\ x(0) &= \xi \in \mathbb{R}^n, \\ u(t) &\in U \quad \text{f.ü. auf } \mathbb{R}_+. \end{aligned}$$

Bemerkung 2.9.

1. Im Vergleich der Aufgaben aus (2.34) und (2.35) für $t_1 \rightarrow \infty$ lassen sich verschiedene Schwierigkeiten feststellen. Die Funktion v in (2.34) muss derart gewählt werden, dass die zugehörige Folge optimaler Prozesse $(x(t, t_1), u(t, t_1))$ im Sinne eines geeigneten

Konvergenzbegriffes gegen eine Lösung (\hat{x}, \hat{u}) konvergiert und \hat{x} asymptotisch stabil ist. In [IK02] wird gezeigt, dass die Stabilitätseigenschaft gesichert werden kann, wenn ν eine Lyapunov-Funktion ist.

2. Die Wahl eines Integralbegriffes muss mit Sorgfalt erfolgen und ist sowohl für eine wohldefinierte Formulierung der Aufgabenstellung als auch für die Diskussion der Existenz von Lösungen essentiell, vgl. [LPW08], [PL17].
3. Für ein festes $t_1 < \infty$ ist der betrachtete Funktionenraum $W_1^{1,n}([t_0, t_1])$ ein Banachraum. Der Übergang zu dem unendlichen Zeithorizont liefert als Zustandsraum den Raum der lokal absolut stetigen Funktionen $AC_{loc}^n(\mathbb{R}_+)$, welcher ein lokal konvexer Hausdorff-Raum ist.

In vielen Arbeiten werden Receding Horizon - Strategien oder auch im Fall des linearquadratischen Regulators das Lösen der korrespondierenden Matrix-Riccati-Gleichung zur Bestimmung der stabilisierenden Steuerung vorgeschlagen, vgl. z.B. [Kal60, IK02, Son98, Son83]. Wir hingegen verfolgen die Philosophie, ein Steuerungsproblem mit a priori gegebenem unendlichen Zeithorizont derart zu formulieren, dass die optimale Lösung dieser Aufgabe die exponentielle Steuerbarkeit des zugrunde liegenden dynamischen Systems liefert und wir dieses ohne sukzessive Zerlegung des Zeithorizontes und mit Hilfe eines Pontryaginsches Maximumprinzips studieren können. Dazu führen wir sowohl für den Zustand als auch für die Steuerung gewichtete Funktionenräume ein und interpretieren das Integral im Zielfunktional als eigentliches Lebesgue-Integral. Insbesondere bauen wir bei der Wahl der gewichteten Funktionenräume auf den folgenden Resultaten auf:

Lemma 2.46 ([PB18], Lemma 2, Lemma 3)

1. Ist $x \in W_2^1(\mathbb{R}_+, \omega)$, $\omega(t) := e^{\beta t}$, $\beta > 0$, so folgt $x \in W_1^1(\mathbb{R}_+)$.
2. Ist $x \in W_2^1(\mathbb{R}_+, \omega)$, $\omega(t) := e^{\beta t}$, $\beta > 0$ und $\|x\|_{1,2,\omega} \leq 1$, dann gilt für alle $t \in \mathbb{R}_+$ die Ungleichung

$$|x(t)| \leq c\sqrt{t}e^{-\frac{\beta}{2}t}$$

mit einer Konstanten $c > 0$.

Die Beweise sind in der angegebenen Quelle zu finden. Als Folgerung des obigen Lemmas erhalten wir unter den angegebenen Voraussetzungen, dass $\lim_{t \rightarrow \infty} x(t) = 0$ gilt und x exponentiell stabil im Sinne von Lyapunov ist. Entsprechend legen wir im Kontext der Stabilitätstheorie Gewichte der Form $\omega(t) := e^{\beta t}$, $\beta > 0$ zugrunde. Basierend auf Resultaten aus der Existenztheorie und einem Pontryaginschen Maximumprinzip können wir die Lösung der Steuerungsaufgabe studieren und erhalten eine Steuerung, welche die betrachtete Dynamik exponentiell stabilisiert. Wir werden dieses Vorgehen an dem Beispiel des mathematischen Pendels in dem Kapitel 6 demonstrieren und damit eine weitere Anwendung der in dieser Arbeit vorgeschlagenen Pseudospektralmethode aufzeigen.

3 Ergebnisse aus der Approximationstheorie in Hilberträumen

Im Rahmen dieser Arbeit soll eine indirekte Spektralmethode zur Lösung linear-quadratischer Steuerungsaufgaben über dem unendlichen Zeithorizont $[0, \infty)$ entwickelt werden. Dabei wird die Wahl eines geeigneten Zustands- und Steuerungsraumes auf gewichtete Funktionenräume fallen, welche insbesondere auch Hilberträume sind. Auf eine theoretische Auseinandersetzung mit der Approximation in Hilberträumen kann daher nicht verzichtet werden. Wir konzentrieren uns dabei auf die Vorstellung der wichtigsten Resultate aus der Approximationstheorie in Hilberträumen und beginnen unsere Ausführungen mit der Formulierung eines allgemeinen Approximationsproblems, wobei wir unweigerlich auf den Begriff einer Reihenentwicklung und auf die Definition von Projektions-Operatoren treffen werden. Aufbauend darauf zeigen wir den Zusammenhang zu orthogonalen Funktionensystemen auf und studieren anschließend deren essentielle Eigenschaften, wobei wir insbesondere auf deren Vollständigkeit und Abgeschlossenheit eingehen möchten. Da die polynomiale Approximation in dem gewichteten Lebesgue-Raum $L_2((a, b), \omega)$ für unsere Weiterarbeit von besonderem Interesse ist, studieren wir die beste polynomiale Approximation in diesen Räumen genauer und schließen dieses Kapitel mit der Vorstellung grundlegender Eigenschaften orthogonaler Polynome ab.

3.1 Beste Approximation und Orthogonalprojektion

Bereits in den frühen Ansätzen der numerischen Analysis stellte sich die Aufgabe, komplizierte Funktionen näherungsweise durch einfachere Objekte darzustellen. Formal lässt sich diese Problemstellung auf einige zentrale Grundbegriffe zurückführen, vgl. [Göp73], S. 5, Definition 2.1.1.

Das allgemeine Problem der besten Approximation Sei (X, d) ein metrischer Raum, $M \subseteq X$ eine nichtleere Teilmenge und x ein vorgegebenes Element aus X . Das allgemeine Approximationsproblem besteht darin, eine Lösung der Aufgabe

$$\inf_{y \in M} d(x, y) \rightarrow \text{Inf!} \quad (3.1)$$

zu finden. Ein Element $x_0 \in M$ mit der Eigenschaft $d(x, x_0) = \inf_{y \in M} d(x, y)$ heißt *beste Approximation* oder auch *Bestapproximation* von x bezüglich M . Da ein Abstand zur Menge M gemessen wird, findet man in der Literatur für x_0 auch die Bezeichnung *metrische Projektion* P von x auf M .

Wir suchen also ein Element $x_0 \in M$, welches den kürzesten Abstand von einem Element x des Raumes X zu der Menge M realisiert. Inhalt der Approximationstheorie ist es, Fragen nach der Existenz und Eindeutigkeit einer solchen besten Approximation zu studieren. Im Allgemeinen kann die Aufgabe in (3.1) eine, mehrere oder auch keine Lösung besitzen, weshalb durch die Zuordnung $P : x \mapsto P(x)$ zunächst eine mengenwertige Abbildung auf X definiert wird. Später werden wir sehen, unter welchen Voraussetzungen sich diese Abbildung genauer charakterisieren lässt.

Weiterhin von Interesse sind Methoden zur Konstruktion und Berechnung einer Bestapproximation, à priori Abschätzungen des Fehlers und die sich in diesem Kontext ergebene Konvergenzfrage. Antworten auf diese Fragen können in Abhängigkeit der Struktureigenschaften des Raumes X und der Menge M gegeben werden. Einen tieferen Einblick in die Approximation in speziellen Räumen oder in die klassische Fourier-Approximation liefert Schönhage in [Sch71]. Ein historisch bedeutsames Beispiel für die Approximation von reellwertigen Funktionen auf kompakten Intervallen ist das Folgende:

Beispiel Sei $C([a, b])$ der Vektorraum aller über dem kompakten Intervall $[a, b] \subset \mathbb{R}$ stetigen reellwertigen Funktionen $f : [a, b] \rightarrow \mathbb{R}$ versehen mit der Norm

$$\|f\|_\infty := \max_{t \in [a, b]} |f(t)|.$$

Dann liefert der fundamentale Approximationssatz von Weierstraß, dass jede Funktion $f \in C([a, b])$ beliebig gut durch Polynome approximiert werden kann. Mit anderen Worten können wir zu jeder beliebigen Zahl $\varepsilon > 0$ ein Polynom p finden, welches für alle $t \in [a, b]$ die Abschätzung

$$|f(t) - p(t)| < \varepsilon$$

erfüllt, vgl. [Wer18], S. 31, Satz I.2.11. Damit kann jede auf dem Kompaktum $[a, b]$ stetige Funktion f durch eine Folge von Polynomen im Sinne der gleichmäßigen Konvergenz auf $[a, b]$ approximiert werden.

Wir konzentrieren uns im Folgenden auf die Approximation in normierten Räumen, in denen in natürlicher Weise für $x, y \in X$ durch $d(x, y) := \|x - y\|$ eine Metrik erklärt ist. Vor allem die Approximation in reellen Hilberträumen ist für unsere Ausführungen von besonderem Interesse. Hier ergeben sich unter zusätzlichen Eigenschaften der Menge M wertvolle Charakterisierungen für die beste Approximation. Darüber hinaus liefert der Projektionssatz Aufschluss über deren Existenz und Eindeutigkeit.

Wir beginnen mit der Wiederholung grundlegender Definitionen, wodurch sich das geometrische Konzept der Orthogonalität in Hilberträumen mit Hilfe des Skalarproduktes abstrahieren lässt.

Definition 3.1 ([Wer18], S. 242, Definition V.3.1)

Sei X ein Prähilbertraum. Zwei Elemente $x, y \in X$ heißen orthogonal, in Zeichen $x \perp y$, falls $\langle x, y \rangle = 0$ gilt. Wir nennen zwei Teilmengen $M, N \subseteq X$ orthogonal, falls $\langle x, y \rangle = 0 \forall x \in M, y \in N$ gilt. In diesem Fall schreiben wir $M \perp N$. Das orthogonale Komplement von M

wird durch die Menge

$$M^\perp := \{y \in X : x \perp y \ \forall x \in M\}$$

beschrieben.

Bemerkung 3.1 (vgl. [Wer18], S.243). *Es ergeben sich die folgenden Eigenschaften orthogonaler Elemente und orthogonaler Komplemente:*

- (i) *Sind zwei Elemente orthogonal zueinander, also ist $x \perp y$, dann gilt: $\|x\|^2 + \|y\|^2 = \|x + y\|^2$ (Satz des Pythagoras).*
- (ii) *Das orthogonale Komplement M^\perp ist stets ein abgeschlossener Unterraum von X .*

Mit Kenntnis der vorangegangenen Definitionen können wir direkt in die Approximationstheorie in reellen Hilberträumen einsteigen und beginnen mit dem bedeutsamen *Projektionssatz*, welcher uns sofort eine entsprechende Existenzaussage liefert.

Satz 3.2 ([Wer18], S. 243, Satz V.3.2)

Sei H ein reeller Hilbertraum und $M \subset H$ eine nichtleere, abgeschlossene und konvexe Teilmenge. Dann existiert zu jedem $x \in H$ ein eindeutig bestimmtes Element $x_0 \in M$ mit

$$\inf_{y \in M} \|x - y\| = \|x - x_0\|.$$

Unter den Voraussetzungen des Projektionssatzes können wir jedem Element $x \in H$ eine eindeutig bestimmte Bestapproximation x_0 bezüglich der Menge M zuordnen und erhalten damit also nicht nur eine Existenzaussage, sondern insbesondere auch die Eindeutigkeit der Bestapproximation. Darüber hinaus wird durch die eindeutige Zuordnung $H \ni x \mapsto x_0 \in M$ eine (im Allgemeinen nicht lineare) Abbildung $P : H \rightarrow M, P(x) = x_0$ definiert, die sich genauer charakterisieren lässt:

Lemma 3.3 ([Wer18], S. 244, Lemma V.3.3)

Sei M eine abgeschlossene und konvexe Teilmenge des reellen Hilbertraumes H . Dann sind für $x \in H$ die nachfolgenden Aussagen äquivalent:

- (i) $\inf_{y \in M} \|x - y\| = \|x - x_0\|$
- (ii) $\langle x - x_0, y - x_0 \rangle \leq 0$ für alle $y \in M$.

Durch die Variationsungleichung in (ii) wird eine notwendige und hinreichende Optimalitätsbedingung für das Approximationsproblem zur Verfügung gestellt.

Der Fall eines abgeschlossenen Unterraumes stellt eine Besonderheit bezüglich der besten Approximation in Hilberträumen dar.

Satz 3.4 (vgl. [Wer18], S. 245, Theorem V.3.4)

Sei H ein reeller Hilbertraum und $U \neq \{0\}$ ein abgeschlossener Unterraum von H . Dann gilt:

- (i) *Zu jedem $x \in H$ existiert eine eindeutige Zerlegung $x = x_0 + x_1$ mit $x_0 \in U$ und $x_1 \in U^\perp$, es ist also $H = U \oplus U^\perp$. (Erster Satz von Riesz)*

(ii) Das Element x_0 heißt orthogonale Projektion von x auf U und ist die eindeutig bestimmte beste Approximation von x bezüglich U , das heißt es gilt

$$\|x - x_0\| \leq \|x - y\| \quad \text{für alle } y \in U.$$

(iii) Durch $P_U : H \rightarrow U$ mit $P_U(x) = x_0$ wird eine lineare beschränkte Projektion¹ von H auf U definiert. Für jedes $x \in H$ ist die Gleichung

$$\langle x - P_U(x), y \rangle = 0 \quad \forall y \in U$$

erfüllt, also stets $x - P_U(x) \in U^\perp$. Daher bezeichnet man P_U als Orthogonalprojektion von H auf U .

Für abgeschlossene Unterräume U von H ist die metrische Projektion $P_U : H \rightarrow U$ also insbesondere ein linearer Operator und wird dann als orthogonale Projektion von H auf U bezeichnet.

Wir formulieren das eingangs allgemein gestellte Problem der besten Approximation nun für den Fall eines reellen Hilbertraumes und in Bezug auf einen abgeschlossenen Unterraum. Anschließend fassen wir die Ergebnisse aus dem vorangegangenen Satz zusammen und diskutieren diese bezüglich der gestellten Approximationsaufgabe.

Beste Approximation bezüglich eines abgeschlossenen Unterraumes Es sei H ein reeller Hilbertraum und $U \neq \{0\}$ ein abgeschlossener Unterraum von H . Für ein beliebiges $x \in H$ möchten wir ein Element $x_0 \in U$ finden, sodass dieses die Aufgabe

$$\inf_{y \in U} \|x - y\| \rightarrow \text{Inf!} \tag{3.2}$$

löst.

Die Aussage (ii) des Satzes 3.4 garantiert uns die Existenz einer eindeutig bestimmten Lösung $x_0 \in U$ der gestellten Aufgabe. Darüber hinaus kann diese mit der in (iii) definierten Orthogonalprojektion $P_U(x)$ identifiziert werden, das heißt, für $x \in H$ ist $P_U(x)$ der eindeutig bestimmte Punkt $x_0 \in U$, welcher die Aufgabe in (3.2) löst. Da U ein Unterraum von H ist, erhalten wir aufgrund der Linearität von U und als direkte Folgerung aus dem Lemma 3.3 noch eine genauere Charakterisierung der Bestapproximation $P_U(x) = x_0$. Wir halten diese in einem Lemma fest.

Lemma 3.5 ([GRT09], S. 13, Satz 2.4)

Sei U ein linearer Teilraum des reellen Hilbertraumes H und $x \in H$. Dann ist $x_0 \in U$ die eindeutig bestimmte beste Approximation von x bezüglich U , das heißt es gilt

$$\inf_{y \in U} \|x - y\| = \|x - x_0\|$$

genau dann, wenn

$$\langle x - x_0, y \rangle = 0 \tag{3.3}$$

¹ Eine Projektion auf einem Vektorraum X ist eine Abbildung $P : X \rightarrow X$ mit $P^2 = P$.

für alle $y \in U$ erfüllt ist.

In Bezug auf den abgeschlossenen Unterraum U und aufgrund des Zusammenhangs $P_U(x) = x_0$ ergibt sich damit automatisch auch eine Bestimmungsgleichung für die orthogonale Projektion:

$$\langle x - P_U(x), y \rangle = 0 \quad y \in U. \quad (3.4)$$

Häufig wird in der Literatur, insbesondere in Büchern über Approximationstheorie, dieser Zusammenhang genutzt, um die orthogonale Projektion mit Hilfe von (3.4) zu definieren. Da diese Gleichung insbesondere eine hinreichende Bedingung zum Auffinden des Elementes bester Approximation bzw. zum Bestimmen der orthogonalen Projektion von $x \in H$ bezüglich U darstellt, werden wir auf dieses Resultat an späterer Stelle noch einmal zurückkommen.

3.2 Orthonormalsysteme und Fourier-Entwicklung

Hauptgegenstand dieses Abschnittes sind Fourier-Entwicklungen nach Orthonormalsystemen. Da diesen bei der Approximation in Hilberträumen eine besondere Bedeutung zukommt, werden wir den Zusammenhang zu Approximationsaufgaben hervorheben und die für uns wichtigen Resultate zusammenstellen. Mit H meinen wir im gesamten Abschnitt stets einen reellen Hilbertraum.

In dem vorangegangenen Abschnitt haben wir das Konzept der Orthogonalität in Hilberträumen vorgestellt. Wir bauen darauf auf und führen den Begriff eines Orthogonal- bzw. Orthonormalsystems ein.

Definition 3.6 ([Dav75], S. 163, Definition 8.3.1)

Eine Teilmenge $M \subset H$ heißt Orthogonalsystem, wenn für alle Elemente aus M gilt, dass sie paarweise orthogonal zueinander sind, also $\langle x, y \rangle = 0$ für alle $x, y \in M$, $x \neq y$ richtig ist. Ein Orthogonalsystem, für welches zusätzlich $\langle x, x \rangle = 1$ für alle $x \in M$ erfüllt ist, heißt Orthonormalsystem (kurz: ONS).

Betrachten wir eine Folge von Elementen x_1, x_2, x_3, \dots aus H , so bildet diese also ein Orthogonalsystem, falls die nachfolgenden Gleichungen gelten:

$$\langle x_i, x_j \rangle = 0, \quad i = 1, 2, \dots, \quad j = 1, 2, \dots, \quad i \neq j.$$

Die Folge bildet ein ONS, falls die Gleichungen

$$\langle x_i, x_j \rangle = \begin{cases} 0, & \text{für } i \neq j \\ 1, & \text{für } i = j \end{cases}$$

erfüllt sind. Betrachten wir eine endliche Folge orthogonaler Elemente x_1, x_2, \dots, x_N aus H mit $x_k \neq 0$, $k = 1, 2, \dots, N$, so lässt sich leicht zeigen, dass diese auch linear unabhängig

ist, vgl. [Dav75], S. 164, Theorem 8.3.2. Insbesondere ist jedes ONS linear unabhängig, denn sind x_1, \dots, x_N paarweise verschiedene Elemente eines ONS aus H , so gilt

$$\left\| \sum_{k=1}^N a_k x_k \right\|^2 = \left\langle \sum_{k=1}^N a_k x_k, \sum_{k=1}^N a_k x_k \right\rangle = \sum_{k=1}^N \sum_{l=1}^N a_k a_l \langle x_k, x_l \rangle = \sum_{k=1}^N |a_k|^2$$

mit Skalaren a_1, \dots, a_N . Ist nun $\sum_{k=1}^N a_k x_k = 0$ erfüllt, dann folgt sofort aus der obigen Gleichung $\sum_{k=1}^N |a_k|^2 = 0$, also stets $a_1 = \dots = a_N = 0$. Andersherum muss eine linear unabhängige Menge aus H jedoch nicht notwendig orthogonal sein. Der nachfolgende Satz jedoch wird zeigen, dass wir ausgehend von einer solchen Menge stets Linearkombinationen finden können, welche orthogonal sind, wir können die Menge also *orthogonalisieren*.

Satz 3.7 (Gram-Schmidt-Verfahren, [Wer18], S. 251, Satz V.4.2)

Sei mit $\{x_k : k \in \mathbb{N}\}$ eine linear unabhängige Teilmenge von H gegeben. Dann existiert ein Orthonormalsystem E mit²

$$\overline{\text{lin } E} = \overline{\text{lin}\{x_k : k \in \mathbb{N}\}}.$$

Der Beweis basiert auf der Konstruktion einer Folge, welche gemäß der folgenden Vorschrift gebildet wird:

$$\begin{aligned} \tilde{e}_1 &= x_1 \quad \text{und} \quad e_1 = \frac{\tilde{e}_1}{\|\tilde{e}_1\|}, \\ \tilde{e}_2 &= x_2 - \langle x_2, e_1 \rangle e_1 \quad \text{und} \quad e_2 = \frac{\tilde{e}_2}{\|\tilde{e}_2\|}, \\ &\vdots \\ \tilde{e}_{n+1} &= x_{n+1} - \sum_{j=1}^n \langle x_{n+1}, e_j \rangle e_j \quad \text{und} \quad e_{n+1} = \frac{\tilde{e}_{n+1}}{\|\tilde{e}_{n+1}\|}. \end{aligned}$$

Die Menge $E := \{e_1, e_2, \dots\}$ verkörpert dann ein Orthonormalsystem mit $x_k \in \text{lin } E$ und $e_k \in \text{lin}\{x_r : r \in \mathbb{N}\}$ für alle $k \in \mathbb{N}$.

Wir möchten die Anwendung des Gram-Schmidt-Verfahrens an einem Beispiel vorführen.

Beispiel Wir wenden das Gram-Schmidt-Verfahren auf die Monome $\{1, t, t^2, \dots\}$ in dem gewichteten Lebesgue-Raum $L_2((0, \infty), \omega_{\alpha, \beta})$ mit dem Skalarprodukt

$$\langle x, y \rangle_{2, \omega_{\alpha, \beta}} := \int_0^{\infty} x(t)y(t)\omega_{\alpha, \beta}(t)dt$$

an und führen die Berechnung der ersten zwei Elemente e_1, e_2 beispielhaft vor. Dabei nutzen wir die bereits in dem Abschnitt 2.4.1 vorgestellte Gewichtsfunktion $\omega_{\alpha, \beta}(t) := t^\alpha e^{-\beta t}$ mit

² Mit $\text{lin } E$ wird die lineare Hülle von E bezeichnet.

$\alpha \in \mathbb{N}_0$ und $\beta > 0$. Wir beginnen mit $\tilde{e}_1 = 1$ und berechnen die entsprechende Norm, vgl. (2.30):

$$\|\tilde{e}_1\|_{2,\omega_{\alpha,\beta}} = \left(\int_0^\infty 1 \cdot t^\alpha e^{-\beta t} dt \right)^{1/2} = \left(\frac{\alpha!}{\beta^{\alpha+1}} \right)^{1/2}.$$

Somit erhalten wir das erste normierte Element

$$e_1 = \left(\frac{\alpha!}{\beta^{\alpha+1}} \right)^{-1/2} = \left(\frac{\beta^{\alpha+1}}{\alpha!} \right)^{1/2}.$$

Die Berechnung von e_2 erfolgt in mehreren Teilschritten. Zunächst bestimmen wir \tilde{e}_2 gemäß der folgenden Vorschrift:

$$\tilde{e}_2 = t - \langle t, e_1 \rangle_{2,\omega_{\alpha,\beta}} \cdot e_1. \quad (3.5)$$

Im nächsten Schritt betrachten wir den Ausdruck

$$\langle t, e_1 \rangle_{2,\omega_{\alpha,\beta}} \cdot e_1 = \frac{\beta^{\alpha+1}}{\alpha!} \int_0^\infty t^{\alpha+1} e^{-\beta t} dt \quad (3.6)$$

separat und werten das auftretende Integral mit Hilfe von partieller Integration aus:

$$\int_0^\infty t^{\alpha+1} e^{-\beta t} dt = \lim_{T \rightarrow +\infty} \left[-\frac{1}{\beta} t^{\alpha+1} e^{-\beta t} \right]_0^T + \frac{\alpha+1}{\beta} \int_0^\infty t^\alpha e^{-\beta t} dt.$$

Die Berechnung des Grenzwertes und des Integralwertes erfolgt dabei wie in dem Abschnitt 2.4.1 und führt zu dem nachfolgenden Ergebnis:

$$\int_0^\infty t^{\alpha+1} e^{-\beta t} dt = \frac{\alpha+1}{\beta} \cdot \frac{\alpha!}{\beta^{\alpha+1}}. \quad (3.7)$$

Wir setzen dieses Resultat in (3.6) ein

$$\langle t, e_1 \rangle_{2,\omega_{\alpha,\beta}} \cdot e_1 = \frac{\beta^{\alpha+1}}{\alpha!} \cdot \frac{\alpha+1}{\beta} \cdot \frac{\alpha!}{\beta^{\alpha+1}} = \frac{\alpha+1}{\beta}$$

und erhalten dann gemäß der Vorschrift in (3.5) das Element \tilde{e}_2 :

$$\tilde{e}_2 = t - \frac{\alpha+1}{\beta} = \frac{1}{\beta} (\beta t - (1 + \alpha)).$$

Für die Normierung berechnen wir den Ausdruck

$$\begin{aligned}\|\tilde{e}_2\|_{2,\omega_{\alpha,\beta}}^2 &= \int_0^\infty \left(t - \frac{\alpha+1}{\beta}\right)^2 t^\alpha e^{-\beta t} dt \\ &= \int_0^\infty t^{\alpha+2} e^{-\beta t} dt - \frac{2(\alpha+1)}{\beta} \int_0^\infty t^{\alpha+1} e^{-\beta t} dt + \left(\frac{\alpha+1}{\beta}\right)^2 \int_0^\infty t^\alpha e^{-\beta t} dt, \quad (3.8)\end{aligned}$$

indem wir die Integralausdrücke getrennt auswerten. Dabei benutzen wir wieder partielle Integration und die Ergebnisse aus (2.30) und (3.7):

$$\begin{aligned}1) \quad & \int_0^\infty t^{\alpha+2} e^{-\beta t} dt = \lim_{T \rightarrow +\infty} \left[-\frac{1}{\beta} t^{\alpha+2} e^{-\beta t} \right]_0^T + \frac{\alpha+2}{\beta} \int_0^\infty t^{\alpha+1} e^{-\beta t} dt = \frac{\alpha+2}{\beta} \cdot \frac{\alpha+1}{\beta} \cdot \frac{\alpha!}{\beta^{\alpha+1}} \\ 2) \quad & \frac{2(\alpha+1)}{\beta} \int_0^\infty t^{\alpha+1} e^{-\beta t} dt = \frac{2(\alpha+1)}{\beta} \cdot \frac{\alpha+1}{\beta} \cdot \frac{\alpha!}{\beta^{\alpha+1}} \\ 3) \quad & \left(\frac{\alpha+1}{\beta}\right)^2 \int_0^\infty t^\alpha e^{-\beta t} dt = \left(\frac{\alpha+1}{\beta}\right)^2 \frac{\alpha!}{\beta^{\alpha+1}}.\end{aligned}$$

Eintragen der Ergebnisse aus 1)-3) in (3.8) führt zu

$$\begin{aligned}\|\tilde{e}_2\|_{2,\omega_{\alpha,\beta}}^2 &= \frac{\alpha+2}{\beta} \cdot \frac{\alpha+1}{\beta} \cdot \frac{\alpha!}{\beta^{\alpha+1}} - \frac{2(\alpha+1)}{\beta} \cdot \frac{\alpha+1}{\beta} \cdot \frac{\alpha!}{\beta^{\alpha+1}} + \left(\frac{\alpha+1}{\beta}\right)^2 \frac{\alpha!}{\beta^{\alpha+1}} \\ &= \frac{(\alpha+1)!}{\beta^{\alpha+1}} \left\{ \frac{\alpha+2}{\beta^2} - \frac{2(\alpha+1)}{\beta^2} + \frac{\alpha+1}{\beta^2} \right\} \\ &= \frac{(\alpha+1)!}{\beta^{\alpha+1}} \cdot \frac{1}{\beta^2}.\end{aligned}$$

Schlussendlich können wir das zweite normierte Element e_2 berechnen und erhalten

$$\begin{aligned}e_2 &= \left(\frac{(\alpha+1)!}{\beta^{\alpha+1}} \cdot \frac{1}{\beta^2}\right)^{-1/2} \cdot \frac{1}{\beta} \cdot (\beta t - (1+\alpha)) = \left(\frac{\beta^{\alpha+1}}{(\alpha+1)!}\right)^{1/2} \cdot \beta \cdot \frac{1}{\beta} \cdot (\beta t - (1+\alpha)) \\ &= \left(\frac{\beta^{\alpha+1}}{(\alpha+1)!}\right)^{1/2} \cdot (\beta t - (1+\alpha)).\end{aligned}$$

Da die Berechnungen mit dem Verfahren nach Gram-Schmidt sehr aufwendig sind, verzichten wir auf die Konstruktion weiterer Elemente. Wir möchten jedoch darauf hinweisen, dass wir durch die Anwendung des Verfahrens auf $\{1, t, t^2, \dots\}$ in $L_2((0, \infty), \omega_{\alpha,\beta})$ die sogenannten normierten *verallgemeinerten Laguerre-Polynome* erhalten haben. Aufgrund des zusätzlichen Parameters $\beta > 0$ in der Gewichtsfunktion stellen sie eine Verallgemeinerung der klassischen orthonormalen Laguerre-Polynome dar, welche ein ONS in dem Raum $L_2((0, \infty), t^\alpha e^{-t})$ bilden. Wir werden die verallgemeinerten Laguerre-Polynome an späterer Stelle genauer studieren und dort insbesondere auch die Polynome höheren Grades mit Hilfe einer Rekursionsformel angeben.

Orthonormalsysteme spielen eine entscheidende Rolle in der Approximationstheorie und sind für die Fourier-Reihen-Entwicklung von Funktionen von besonderer Bedeutung. Bevor wir diesen Zusammenhang genauer studieren, führen wir die nachfolgende Definition ein.

Definition 3.8 ([Dav75], S.169, Definition 8.4.1; [GRT09], S.17, Definition 2.3.)

Sei $\{x_k : k \in \mathbb{N}\} \subset H$ ein ONS und x ein beliebiges Element aus H . Die Reihe $\sum_{k=1}^{\infty} \langle x, x_k \rangle x_k$ heißt Fourier-Reihe von x bezüglich des gegebenen ONS, in Zeichen

$$x \sim \sum_{k=1}^{\infty} \langle x, x_k \rangle x_k \quad (3.9)$$

und die Zahl $\langle x, x_k \rangle$ wird als zugehöriger k -ter Fourierkoeffizient von x bezeichnet.

Bemerkung 3.2. Ist $\{x_k : k \in \mathbb{N}\} \subset H$, $x_k \neq 0$, $k = 1, 2, \dots$ ein Orthogonalsystem, so können wir dessen Elemente gemäß der Vorschrift

$$\widetilde{x}_k := \frac{x_k}{\|x_k\|}, \quad k = 1, 2, \dots$$

normieren. Für $x \in H$ ist die entsprechende Fourier-Reihe dann durch

$$x \sim \sum_{k=1}^{\infty} \left\langle x, \frac{x_k}{\|x_k\|} \right\rangle \frac{x_k}{\|x_k\|} = \sum_{k=1}^{\infty} \frac{\langle x, x_k \rangle}{\langle x_k, x_k \rangle} x_k$$

gegeben.

Um die Konvergenzeigenschaften der Fourier-Reihen untersuchen zu können, notieren wir zunächst den Satz des Pythagoras für Orthogonalsysteme, den man durch einfaches nachrechnen bestätigen kann:

Satz 3.9 (Satz des Pythagoras, [GRT09], S. 16, Satz 2.6)

Ist $\{x_1, \dots, x_N\} \subset H$ ein Orthogonalsystem, dann gilt

$$\|x_1 + \dots + x_N\|^2 = \|x_1\|^2 + \dots + \|x_N\|^2.$$

Betrachten wir die Fourier-Reihe aus (3.9) für ein $x \in H$ so gelten für die N -te Partialsumme der Reihe die folgenden Überlegungen: Für beliebiges $N \in \mathbb{N}$ und mit $a_k := \langle x, x_k \rangle$, $k = 1, \dots, N$ ist die Ungleichung

$$\left\| x - \sum_{k=1}^N a_k x_k \right\|^2 \geq 0$$

offensichtlich immer erfüllt. Mit Hilfe der binomischen Formel, des Satzes von Pythagoras und durch Ausnutzen der Orthogonalitäts- und Orthonormalitätseigenschaften des Systems

$\{x_1, \dots, x_N\}$ gelangen wir zu den folgenden Rechnungen:

$$\begin{aligned} 0 \leq \left\| x - \sum_{k=1}^N a_k x_k \right\|^2 &= \|x\|^2 - 2 \left\langle x, \sum_{k=1}^N a_k x_k \right\rangle + \left\| \sum_{k=1}^N a_k x_k \right\|^2 \\ &= \|x\|^2 - 2 \sum_{k=1}^N a_k \langle x, x_k \rangle + \sum_{k=1}^N \|a_k x_k\|^2 \\ &= \|x\|^2 - 2 \sum_{k=1}^N a_k \langle x, x_k \rangle + \sum_{k=1}^N |a_k|^2. \end{aligned}$$

Durch eine quadratische Ergänzung erhalten wir

$$\begin{aligned} 0 \leq \|x\|^2 - 2 \sum_{k=1}^N a_k \langle x, x_k \rangle + \sum_{k=1}^N |a_k|^2 - \sum_{k=1}^N |\langle x, x_k \rangle|^2 + \sum_{k=1}^N |\langle x, x_k \rangle|^2 \\ = \|x\|^2 + \sum_{k=1}^N |a_k - \langle x, x_k \rangle|^2 - \sum_{k=1}^N |\langle x, x_k \rangle|^2 \end{aligned}$$

und da a_k gerade der k -te Fourierkoeffizienten von x ist, folgt

$$\|x\|^2 - \sum_{k=1}^N |\langle x, x_k \rangle|^2 \geq 0.$$

Da N beliebig gewählt wurde, folgt daraus sofort die Besselsche Ungleichung und damit dann auch die Konvergenz der Fourier-Reihe.

Satz 3.10 (vgl. [Wer18], S. 252, Satz V.4.3)

Ist $\{x_k : k \in \mathbb{N}\} \subset H$ ein Orthonormalsystem und $x \in H$, dann gilt die Besselsche Ungleichung

$$\sum_{k=1}^{\infty} |\langle x, x_k \rangle|^2 \leq \|x\|^2$$

und die Fourier-Reihe aus (3.9) konvergiert gegen ein Element aus H .

Da die Partialsummen der Fourier-Reihe in (3.9) eine Cauchy-Folge in H bilden, erhält man die Konvergenz der Fourier-Reihe als direkte Folgerung aus der Besselschen Ungleichung. Dieses Ergebnis ist von entscheidender Bedeutung für die Approximation in Hilberträumen, weshalb wir eine kurze Beweisskizze dieser Folgerung angeben möchten. Sei s_N die N -te Partialsumme der Reihe aus (3.9). Dann gilt für $M < N$ die Rechnung

$$\|s_N - s_M\|^2 = \left\| \sum_{k=1}^N \langle x, x_k \rangle x_k - \sum_{k=1}^M \langle x, x_k \rangle x_k \right\|^2 = \left\| \sum_{k=M+1}^N \langle x, x_k \rangle x_k \right\|^2 = \sum_{k=M+1}^N |\langle x, x_k \rangle|^2.$$

Da aus der Besselschen Ungleichung sofort die Konvergenz der rechtsstehenden Reihe folgt, ist die Folge der Partialsummen also eine Cauchy-Folge und aufgrund der Vollständigkeit von

H können wir damit direkt auf deren Konvergenz schließen. Dies liefert uns dann schließlich die Konvergenz der Fourier-Reihe aus (3.9) gegen ein Element aus H , da diese per Definition konvergiert, falls die aus ihren Partialsummen s_N gebildete Folge in H konvergiert.

Wir können also mit Hilfe der Besselschen Ungleichung garantieren, dass für jedes $x \in H$ und mit einem ONS $\{x_k : k \in \mathbb{N}\} \subset H$ die Fourier-Reihe von x immer einen wohldefinierten Wert, entweder in dem Sinne einer endlichen Summe oder als konvergente Reihe, liefert.

Stellen wir uns die Frage, unter welchen Bedingungen die Gleichheit in der Besselschen Ungleichung gilt, so wird sich zeigen, dass hier ein Zusammenhang zu der bezüglich eines gegebenen ONS beliebig genauen Approximation eines Hilbertraumelementes besteht. Dafür benötigen wir eine zusätzliche Eigenschaft, die sicherstellt, dass das vorliegende ONS auch umfangreich genug ist. Das nachfolgende Beispiel aus [GRT09], S. 17 verdeutlicht dessen Notwendigkeit.

Beispiel Betrachten wir den Hilbertraum $H = \mathbb{R}^3$ und das ONS $\{x_1, x_2\}$ mit $x_1 = (1, 0, 0)^T$ und $x_2 = (0, 1, 0)^T$. Möchten wir den Vektor $x = (1, 1, 1)^T$ durch eine Linearkombination aus Elementen des ONS

$$\sum_{k=1}^2 \alpha_k x_k = (\alpha_1, \alpha_2, 0)^T$$

annähern, so stellen wir fest, dass stets

$$\left\| x - \sum_{k=1}^2 \alpha_k x_k \right\| = \sqrt{(1 - \alpha_1)^2 + (1 - \alpha_2)^2 + 1^2} \geq 1$$

gilt. Das heißt, es ist uns nicht möglich x durch das vorliegende ONS beliebig genau zu approximieren. Der Grund liegt darin, dass das gegebene System nicht maximal ist.

Definition 3.11

Ein ONS $\{x_k : k \in \mathbb{N}\} \subset H$ heißt maximal, wenn aus $x \in H$ und $\langle x, x_k \rangle = 0 \forall k \in \mathbb{N}$ stets folgt, dass $x = 0$ gilt.

Insbesondere besagt die vorangegangene Definition, dass es im Falle eines maximalen ONS kein Element $x \neq 0$ aus H gibt, welches zu allen Elementen des ONS orthogonal ist. Das heißt, dass ein maximales ONS nicht zu einem echt größeren ONS erweitert werden kann. In der Literatur finden man dafür auch die Bezeichnung Orthonormalbasis.

Definition 3.12 ([Wer18], S. 250, Definition V.4.1)

Ein ONS $E \subset H$ heißt Orthonormalbasis, wenn die nachfolgende Implikation gilt:

$$E \subset T, T \text{ ist ONS in } H \Rightarrow T = E.$$

Diese Charakterisierung von Orthonormalsystemen führt uns nun zu dem fundamentalen Satz der Orthogonal- bzw. Fourier-Entwicklung in Hilberträumen.

Satz 3.13 (vgl. [Dav75], S. 191, Theorem 8.9.1.)

Es sei mit $\{x_k : k \in \mathbb{N}\} \subset H$ ein Orthonormalsystem gegeben. Dann sind die folgenden Aussagen äquivalent:

(i) Für alle $x \in H$ gilt die Fourier-Reihendarstellung

$$x = \sum_{k=1}^{\infty} \langle x, x_k \rangle x_k, \quad (3.10)$$

wobei die Konvergenz der Fourier-Reihe im Sinne der Konvergenz ihrer Partialsummen in der Norm des Raumes H zu verstehen ist:

$$\lim_{N \rightarrow \infty} \left\| x - \sum_{k=1}^N \langle x, x_k \rangle x_k \right\| = 0.$$

(ii) Für jedes $x \in H$ gilt Gleichheit in der Besselschen Ungleichung, die sogenannte Parseval'sche Gleichung

$$\|x\|^2 = \langle x, x \rangle = \sum_{k=1}^{\infty} |\langle x, x_k \rangle|^2$$

und für alle $x, y \in H$ gilt die erweiterte Parseval'sche Gleichung

$$\langle x, y \rangle = \sum_{k=1}^{\infty} \langle x, x_k \rangle \langle x_k, y \rangle.$$

(iii) Das ONS ist maximal.

(iv) Ein Element aus H ist eindeutig durch seine Fourierkoeffizienten bestimmt. Das heißt, aus $\langle y, x_k \rangle = \langle w, x_k \rangle$, $k = 1, 2, \dots$ folgt stets, dass $y = w$ gilt.

(v) Es ist $H = \overline{\text{lin}\{x_k : k \in \mathbb{N}\}}$.

Bemerkung 3.3. Ein ONS, welches eine der Eigenschaften des vorangegangenen Satzes erfüllt, ist eine Orthonormalbasis und wird auch als vollständiges ONS bezeichnet.

An dieser Stelle möchten wir darauf hinweisen, dass eine Orthonormalbasis keine Vektorraumbasis sein muss, es sei denn, der betrachtete Hilbertraum besitzt eine endliche Dimension. Die Eigenschaft (i) des Satzes 3.13 zeigt jedoch, wieso dennoch von einer „Basis“ gesprochen wird: jedes Element $x \in H$ kann mittels einer Fourier-Reihe und den Basiselementen dargestellt werden. Aufgrund dessen spricht man hier auch von einer sogenannten 'Basisdarstellung'. Die Eindeutigkeit der Fourier-Reihendarstellung in (3.10) ergibt sich direkt aus der Stetigkeit des Skalarproduktes. Angenommen

$$x = \sum_{k=1}^{\infty} b_k x_k$$

sei eine weitere Darstellung von x . Dann gilt

$$\langle x, x_j \rangle = \left\langle \sum_{k=1}^{\infty} b_k x_k, x_j \right\rangle = \sum_{k=1}^{\infty} \langle b_k x_k, x_j \rangle = \sum_{k=1}^{\infty} b_k \langle x_k, x_j \rangle = b_j \quad \forall j \in \mathbb{N},$$

das heißt, die Fourierkoeffizienten von x bezüglich $\{x_k : k \in \mathbb{N}\}$ sind eindeutig bestimmt.

Wir stellen nun den Zusammenhang zwischen einer Fourier-Reihe und der Orthogonalprojektion im Kontext der Approximationstheorie her.

Beste Approximation bezüglich der abgeschlossenen linearen Hülle eines ONS

Sei $\{x_k : k \in \mathbb{N}\}$ ein ONS des reellen Hilbertraumes H . Dann wählen wir als Unterraum von H gerade die abgeschlossene lineare Hülle des ONS und setzen $U := \overline{\text{lin}\{x_k : k \in \mathbb{N}\}}$. Da U ein abgeschlossener (linearer) Unterraum von H ist, können wir auf die Aufgabenstellung „Beste Approximation bezüglich eines abgeschlossenen Unterraumes“ verweisen und die Resultate der Diskussion aus dem Abschnitt 3.1, S. 48 heranziehen.

Demnach wissen wir, dass eine Lösung von (3.2) mit $U = \overline{\text{lin}\{x_k : k \in \mathbb{N}\}}$ existiert, diese sogar eindeutig bestimmt ist und mit der orthogonalen Projektion $P_U : H \rightarrow U$ zusammenfällt. Der nächste Satz wird jetzt zeigen, dass die Orthogonalprojektion eines Elementes $x \in H$ auf $\overline{\text{lin}\{x_k : k \in \mathbb{N}\}}$ gerade dessen Fourier-Entwicklung bezüglich des vorliegenden ONS entspricht.

Satz 3.14 ([Wer18], S. 254, Satz V.4.8)

Sei $\{x_k : k \in \mathbb{N}\} \subset H$ ein Orthonormalsystem. Die Abbildung $P_U : x \mapsto \sum_{k=1}^{\infty} \langle x, x_k \rangle x_k$ ist die Orthogonalprojektion von H auf die abgeschlossene lineare Hülle $U := \overline{\text{lin}\{x_k : k \in \mathbb{N}\}}$.

Die eindeutige Lösung der Approximationsaufgabe in (3.2) mit $U = \overline{\text{lin}\{x_k : k \in \mathbb{N}\}}$ ist also durch die entsprechende Fourier-Entwicklung von x gegeben, das heißt $x_0 = \sum_{k=1}^{\infty} \langle x, x_k \rangle x_k$ ist die Bestapproximation bzw. die orthogonale Projektion von x auf $\overline{\text{lin}\{x_k : k \in \mathbb{N}\}}$.

Ist das ONS zusätzlich vollständig, so folgt aufgrund der vorangegangenen Diskussion bezüglich des Satzes 3.13, dass wir dann jedes $x \in H$ sogar beliebig genau durch dessen Fourier-Reihe approximieren können.

Wir gehen jetzt auf eine Approximationsaufgabe bezüglich eines speziell konstruierten und endlich-dimensionalen Teilraums $U_N \subset H$ ein. Dafür zitieren wir zunächst einen Satz, der uns Auskunft über eine Minimaleigenschaft der abgebrochenen Fourier-Reihen gibt.

Satz 3.15 ([Dav75], S. 171, Theorem 8.5.1)

Es seien $\{x_k : k \in \mathbb{N}\} \subset H$ ein ONS und $x \in H$ beliebig. Dann gilt

$$\left\| x - \sum_{k=1}^N \langle x, x_k \rangle x_k \right\| \leq \left\| x - \sum_{k=1}^N a_k x_k \right\|$$

für alle möglichen Kombinationen von Konstanten a_1, \dots, a_N .

Aufbauend auf den vorangegangenen Satz formulieren wir die entsprechende Aufgabe der besten Approximation.

Beste Approximation bezüglich der linearen Hülle einer endlichen Orthonormalbasis Sei $\{x_k : k \in \mathbb{N}\} \subset H$ ein ONS. Dann ist jede endliche Menge $\{x_1, x_2, \dots, x_N\}$ linear unabhängig und bildet eine Orthonormalbasis des N -dimensionalen Unterraumes $U_N := \text{lin}\{x_1, x_2, \dots, x_N\} \subset H$. Da U_N offensichtlich abgeschlossen³ ist, ist die Existenz und Eindeutigkeit einer Lösung der Approximationsaufgabe in (3.2) mit $U = U_N$ gesichert. Jedes Element $y \in U_N$ lässt sich als Linearkombination der Basiselemente x_k , $k = 1, \dots, N$ ausdrücken, also gemäß der Vorschrift

$$y = \sum_{k=1}^N a_k x_k$$

mit $a_k \in \mathbb{R}$ festlegen. Wir suchen gerade die Linearkombination von x_1, \dots, x_N , für welche der Ausdruck $\left\| x - \sum_{k=1}^N a_k x_k \right\|$ minimal wird. Der vorangegangene Satz liefert, dass die gesuchte Kombination durch die Summe $\sum_{k=1}^N \langle x, x_k \rangle x_k$ gegeben ist. Somit ist die Lösung der Approximationsaufgabe bezüglich U_N und damit auch die orthogonale Projektion von $x \in H$ auf U_N durch die bei N abgebrochene Fourier-Reihe von x bezüglich des ONS $\{x_k : k \in \mathbb{N}\}$ gegeben.

Dasselbe Ergebnis hätten wir auch mit Hilfe der Resultate des Lemmas 3.5 herleiten können: Da U_N ein linearer Unterraum von H ist, ist $x_0 \in U_N$ genau dann eine Bestapproximation von $x \in H$ bezüglich U_N , wenn

$$\langle x - x_0, y \rangle = 0 \quad \forall y \in U_N$$

gilt, vgl. (3.3). Diese Forderung ist für alle Basiselemente x_j , $j = 1, \dots, N$ des Unterraumes U_N zu erfüllen und zusammen mit der Darstellung für $x_0 \in U_N$, $x_0 = \sum_{k=1}^N a_k x_k$ erhalten wir dann aus der obigen Orthogonalitätsforderung für alle $j = 1, \dots, N$ die folgenden Gleichungen:

$$\begin{aligned} \langle x - x_0, x_j \rangle &= \left\langle x - \sum_{k=1}^N a_k x_k, x_j \right\rangle = \langle x, x_j \rangle - \left\langle \sum_{k=1}^N a_k x_k, x_j \right\rangle = \langle x, x_j \rangle - \sum_{k=1}^N a_k \langle x_k, x_j \rangle = 0 \\ \Leftrightarrow \langle x, x_j \rangle &= \sum_{k=1}^N a_k \langle x_k, x_j \rangle. \end{aligned}$$

Da

$$\langle x_k, x_j \rangle = \begin{cases} 0, & \text{für } k \neq j \\ 1, & \text{für } k = j \end{cases}$$

gilt, folgt aus der letzten Zeile eine Berechnungsvorschrift für die Koeffizienten a_k :

$$a_k = \langle x, x_k \rangle, \quad k = 1, \dots, N.$$

Die Koeffizienten stimmen mit den Fourierkoeffizienten von x überein und das Element

³ Endlich-dimensionale Unterräume von normierten Räumen sind abgeschlossen, vgl. [Wer18], S.28, Korollar I.2.6.

bester Approximation bezüglich U_N ist demnach durch die N -te Fourier-Partialsomme von x bezüglich der vorliegenden Orthonormalbasis gegeben:

$$x_0 = \sum_{k=1}^N \langle x, x_k \rangle x_k.$$

Da sich die Bestapproximation von $x \in H$ bezüglich U_N mit der entsprechenden Orthogonalprojektion identifizieren lässt, vgl. Satz 3.4, können wir festhalten, dass die orthogonale Projektion von H auf U_N durch die Abbildung $P_N : x \mapsto \sum_{k=1}^N \langle x, x_k \rangle x_k$ gegeben ist, also

$$P_N(x) = \sum_{k=1}^N \langle x, x_k \rangle x_k$$

für alle $x \in H$ gilt.

Damit kennen wir jetzt die Bedeutung von Orthonormalsystemen in Bezug auf die beste Approximation in Hilberträumen. Insbesondere wissen wir, dass wir mit Hilfe eines vollständigen ONS, also einer Orthonormalbasis, sogar jedes Element $x \in H$ durch seine Fourier-Reihe darstellen und beliebig genau approximieren können. Offen bisher blieb die Frage nach der Existenz einer Orthonormalbasis. Diese Frage lässt sich mit Hilfe des nächsten Satzes schnell beantworten.

Satz 3.16 ([Wer18], S. 254, Satz V.4.9 (a))

Jedes Orthonormalsystem in einem Hilbertraum H kann zur einer Orthonormalbasis erweitert werden. Insbesondere besitzt jeder nicht-triviale Hilbertraum eine Orthonormalbasis.

Der Beweis ist eine direkte Konsequenz aus dem Auswahlaxiom bzw. dem Zornschen Lemma und kann zum Beispiel in [Wei00], S. 56, Satz 1.57 nachgelesen werden. Im Fall eines separablen Hilbertraumes H ergibt sich die Existenz einer Orthonormalbasis auch ohne Verwendung des Zornschen Lemmas, da der Beweis mit Hilfe des Verfahrens von Gram-Schmidt konstruiert werden kann: Da H separabel ist, können wir eine abzählbare dichte Menge $M := \{x_n\}_{n \in \mathbb{N}}$ in H auswählen, die wir o.B.d.A. als linear unabhängig annehmen können (angenommen die Menge wäre nicht linear unabhängig, so können wir „überflüssige“ Elemente sukzessive streichen und erhalten dann eine abzählbare Menge, deren lineare Hülle dicht ist). Wir wenden das Gram-Schmidtsche Orthonormierungsverfahren (vgl. Satz 3.7) an und erhalten ein Orthonormalsystem E mit $\overline{\text{lin } E} = \overline{\text{lin } M}$. Da aufgrund der Dichtheit $\overline{\text{lin } M} = H$ gilt, ist E eine Orthonormalbasis.

Für abzählbare Orthonormalbasen gilt der folgende Satz:

Satz 3.17 ([Wer18], S. 257, Korollar V.4.10)

In einem unendlichdimensionalen Hilbertraum H sind die folgenden Aussagen äquivalent:

- (i) H ist separabel.
- (ii) H besitzt eine abzählbare Orthonormalbasis.

Wir haben gesehen, dass die Fourier-Entwicklung in Hilberträumen ein mathematisch schönes Werkzeug zur Approximation von Elementen aus H darstellt und konnten insbesondere die Zusammenhänge zur orthogonalen Projektion diskutieren. Abschließen möchten wir diesen Abschnitt mit einem Hinweis auf eine weitere interessante Eigenschaft, die uns die Fourier-Entwicklung bezüglich einer Orthonormalbasis liefert: Ist E eine Orthonormalbasis von H , so lässt sich durch den Zusammenhang zwischen einem Element $x \in H$ und seinen Fourier-Koeffizienten, das heißt durch die Vorschrift $Tx := \{\langle x, x_k \rangle\}_{k \in \mathbb{N}}$, eine Abbildung in einen Folgenraum definieren, die ein isometrischer Isomorphismus ist.

Bemerkung 3.4 ([Wer18], S. 257, Satz V.4.12 und Korollar V.4.13). *Ist E eine Orthonormalbasis von H , so ist H isometrisch isomorph zu dem Folgenraum⁴ $l^2(E)$. Ist eine der Bedingung in dem Satz 3.17 erfüllt, so ist H isometrisch isomorph zu dem Folgenraum l^2 .*

Die Fourier-Entwicklung bezüglich einer Orthonormalbasis E gibt uns damit auch die Möglichkeit, den zugrunde liegenden Hilbertraum H mit dem Folgenraum $l^2(E)$ bzw. l^2 zu identifizieren. Dabei bleibt nicht nur die algebraische Struktur der Räume, sondern auch das Skalarprodukt und die Norm erhalten.

3.3 Polynomiale Approximation in $L_2((a, b), \omega)$

In den vorangegangenen Abschnitten haben wir uns mit den theoretischen Grundlagen der Approximation in Hilberträumen auseinandergesetzt. Machen wir uns bewusst, dass wir das allgemeine Approximationsproblem auch als die Suche nach einfachen Funktionen interpretieren können, welche eine vorgegebene und im Allgemeinen komplizierte Funktion approximieren sollen, so wird sofort klar, weshalb die Approximation durch Polynome von besonderer Bedeutung ist. Durch ihre analytisch einfache Darstellung erhalten wir dadurch handliche Ausdrücke, welche im Hinblick auf numerische Verfahren von besonderer Bedeutung sind und beispielsweise für sehr rechenaufwendige Funktionsauswertungen genutzt werden. Darüber hinaus liefert die Integration und Differentiation auch wieder Polynome, wodurch die Klasse der Polynome einen besonders attraktiven Kandidaten in der Approximationstheorie darstellt.

Sei $I \subseteq \mathbb{R}$ gegeben. Für die weitere Diskussion führen wir die Bezeichnung $\mathcal{P}_N(I)$ ein und meinen damit die Menge aller Polynomfunktionen $p : I \rightarrow \mathbb{R}$ mit reellen Koeffizienten und vom Grad $\leq N$ mit $N \in \mathbb{N}_0$. Mit der üblichen Addition und skalaren Multiplikation ist $\mathcal{P}_N(I)$ ein Vektorraum der Dimension $N + 1$ und die wohl bekannteste Basis von $\mathcal{P}_N(I)$ ist die Monombasis $\{1, t, t^2, \dots, t^N\}$, in der sich jedes Polynom $p \in \mathcal{P}_N(I)$ als eindeutige Linearkombination der Form

$$p(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_N t^N, \quad t \in I$$

⁴ Für $1 \leq p < \infty$ und mit einer reellen oder komplexen Zahlenfolge $\{a_n\}_{n \in \mathbb{N}}$ sind die l^p -Folgenräume durch $l^p := \{\{a_n\}_{n \in \mathbb{N}} : a_n \in \mathbb{K}, \sum_{n=1}^{\infty} |a_n|^p < \infty\}$ definiert und werden für $x = \{a_n\}_{n \in \mathbb{N}} \in l^p$ mit der Norm $\|x\|_p = \left(\sum_{n=1}^{\infty} |a_n|^p\right)^{1/p}$ versehen. Normiert man den endlichdimensionalen Raum \mathbb{K}^n mit einer l^p -Norm, so wird dieser mit $l^p(n)$ bezeichnet, vgl. [Wer18], S. 10 ff.

mit reellen Koeffizienten a_0, \dots, a_N entwickeln lässt. Es gilt

$$\mathcal{P}_N(I) = \text{lin} \{1, t, t^2, \dots, t^N\} \quad \text{für } N \in \mathbb{N}_0.$$

3.3.1 Aufgabenstellung

In dem Abschnitt 3.1, vgl. S. 46 haben wir den fundamentalen Approximationssatz von Weierstrass angesprochen. Er besagt, dass jede stetige Funktion $f \in C([a, b])$ auf einem Kompaktum $[a, b] \subset \mathbb{R}$ bezüglich der Supremumsnorm⁵ $\|\cdot\|_\infty$ beliebig genau durch Polynome approximiert werden kann. Insbesondere liegt der Unterraum der Polynomfunktionen auf $[a, b]$ dicht in dem Raum $C([a, b])$, woraus sofort auch auf dessen Separabilität geschlossen werden kann, vgl. [Wer18], S. 33, Korollar I.2.11. Dieses Resultat legt den Grundstein für die klassische Theorie der Polynomapproximation und korrespondiert zu dem folgenden Problem der besten Approximation mit endlichem Intervall (a, b) :

$$\left\{ \begin{array}{l} \text{Finde, für ein gegebenes } N \in \mathbb{N}_0 \text{ und eine Funktion } f \in C([a, b]), \\ \text{ein Polynom } p^* \in \mathcal{P}_N([a, b]), \text{ sodass} \\ \qquad \|f - p^*\|_\infty \leq \|f - q\|_\infty \quad \forall q \in \mathcal{P}_N([a, b]) \\ \text{gilt.} \end{array} \right. \quad (3.11)$$

Diese Aufgabe wird auch *Tschebyscheffsche* Approximationsaufgabe genannt und besitzt stets eine eindeutig bestimmte Lösung. Der Beweis dieser Aussage basiert auf der Theorie Haarscher Räume und kann zum Beispiel in [Sch71], Kapitel 6, S. 148 ff. oder [Isk18], Kapitel 5, S. 139 ff. nachgeschlagen werden. Damit ist garantiert, dass für jede Funktion $f \in C([a, b])$ ein eindeutig bestimmtes Polynom $p^* \in \mathcal{P}_N([a, b])$ existiert, welches die Aufgabe (3.11) löst und darüber hinaus ergibt sich die gleichmäßige Konvergenz von $p^* \in \mathcal{P}_N([a, b])$ gegen f (für $N \rightarrow \infty$) als eine direkte Folgerung aus dem Weierstraßschen Approximationssatz. Zusätzlich zeigt der Beweis des Weierstraßschen Satzes zwar einen Weg auf, mit dem das Näherungspolynom explizit angegeben werden kann, erfordert jedoch einen sehr großen rechnerischen Aufwand und erscheint daher nicht praktikabel. Auch ist das Lösen der in (3.11) formulierten Aufgabe der besten Approximation und damit die Berechnung von p^* nicht trivial, da die zugrunde liegende Norm, welche gleichmäßige Konvergenz impliziert, eine sehr starke Forderung mit sich zieht. Daher erscheint es sinnvoll, das zugrunde liegende Problem der besten polynomialen Approximation nun allgemeiner im gewichteten L_2 -Sinn zu studieren.

Bemerkung 3.5. *Auch wenn das klassische Resultat von Weierstraß nur für endliche Grundintervalle (a, b) gilt, so lassen sich dennoch ähnliche Ergebnisse für unbeschränkte Intervalle unter zusätzlichen Wachstumsbedingungen herleiten. Ist $f \in C([0, \infty))$ und gilt*

$$\lim_{t \rightarrow +\infty} f(t)e^{-\rho t} = 0$$

⁵ Für $f \in C([a, b])$ ist die Supremumsnorm durch $\|f\|_\infty := \max_{t \in [a, b]} |f(t)|$ definiert.

für ein gegebenes $\rho > 0$, dann existiert für jedes $\varepsilon > 0$ ein Polynom p , sodass die Ungleichung

$$|f(t) - p(t)| e^{-\rho t} < \varepsilon$$

für alle $t \in [0, \infty)$ erfüllt ist, vgl. [Fun92], Theorem 6.1.4, S.95. Ist $f \in C((-\infty, \infty))$ und erfüllt für ein gegebenes $\rho > 0$ die Eigenschaft

$$\lim_{t \rightarrow \pm\infty} f(t) e^{-\rho t^2} = 0,$$

so erhalten wir die gleiche Existenzaussage wie oben, jedoch mit

$$|f(t) - p(t)| e^{-\rho t^2} < \varepsilon$$

für alle $t \in (-\infty, \infty)$, vgl. [Fun92], S.95, Theorem 6.1.5.

Wir konzentrieren uns im Folgenden auf die gewichteten Hilberträume $L_2((a, b), \omega)$. Dabei betrachten wir offene Intervalle $(a, b) \subseteq \mathbb{R}$, wobei wir auch $a = -\infty$ oder $b = \infty$ zulassen werden, und gehen stets von einer auf (a, b) definierten Gewichtsfunktion ω im Sinne der Definition 2.25, S. 29 aus. Sprechen wir von Orthogonalität in dem Raum $L_2((a, b), \omega)$, so beziehen wir uns dabei auf das Skalarprodukt $\langle \cdot, \cdot \rangle_{2, \omega}$ in (2.23), S. 30.

Die Aufgabe der **besten polynomialen Approximation im $L_2((a, b), \omega)$** lautet:

$$\left\{ \begin{array}{l} \text{Es seien ein } N \in \mathbb{N}_0 \text{ und eine Funktion } f \in L_2((a, b), \omega) \text{ gegeben. Gesucht ist} \\ \text{ein Polynom } p^* \in \mathcal{P}_N((a, b)), \text{ sodass} \\ \|f - p^*\|_{2, \omega} \leq \|f - q\|_{2, \omega} \quad \forall q \in \mathcal{P}_N((a, b)) \\ \text{gilt.} \end{array} \right. \quad (3.12)$$

Da der Raum $L_2((a, b), \omega)$ ein Hilbertraum ist, können wir die Aufgabe (3.12) mit den allgemeinen Methoden der Approximationstheorie in Hilberträumen behandeln. Diese haben wir in den vorangegangenen Abschnitten ausführlich erörtert und können diese nun auf die obige Problemstellung anwenden, vgl. speziell mit den Resultaten des Abschnittes 3.2, S. 58 f.

Unter der natürlichen Voraussetzung, dass in dem Raum $L_2((a, b), \omega)$ alle Polynome enthalten sind, können wir mit Hilfe des Verfahrens von Gram-Schmidt ausgehend von den Monomen $\{1, t, t^2, \dots\}$ ein Orthogonalsystem $\{p_0, p_1, \dots\} \subset L_2((a, b), \omega)$ konstruieren. Offensichtlich bildet $\{p_0, p_1, \dots, p_N\}$ eine bzgl. des $L_2((a, b), \omega)$ -Skalarproduktes $\langle \cdot, \cdot \rangle_{2, \omega}$ orthogonale Basis von $\mathcal{P}_N((a, b))$ und insbesondere ist $\mathcal{P}_N((a, b))$ als endlich-dimensionaler Unterraum von $L_2((a, b), \omega)$ auch abgeschlossen. Mit diesen Vorüberlegungen können wir sofort die Fragen nach der Existenz, Eindeutigkeit und sogar nach der Darstellbarkeit einer Lösung von (3.12) beantworten:

- Die Aufgabe (3.12) ist eindeutig lösbar, das heißt für jede Funktion $f \in L_2((a, b), \omega)$ existiert genau ein Polynom $p^* \in \mathcal{P}_N((a, b))$, welches die Gleichung in (3.12) löst.

- Das Polynom $p^* \in \mathcal{P}_N((a, b))$ ist genau dann die beste Approximation von $f \in L_2((a, b), \omega)$ bezüglich $\mathcal{P}_N((a, b))$, wenn

$$\langle f - p^*, q \rangle_{2, \omega} = 0 \quad \forall q \in \mathcal{P}_N((a, b))$$

gilt.

- Das Polynom bester Approximation $p^* \in \mathcal{P}_N((a, b))$ ist durch die N -te Fourier-Partialsumme von f bezüglich des Orthogonalsystems $\{p_0, p_1, \dots\}$ gegeben und lässt sich demnach gemäß der Vorschrift

$$p^* := \sum_{k=0}^N \hat{f}_k p_k \tag{3.13}$$

mit den Fourier-Koeffizienten $\hat{f}_k := \frac{\langle f, p_k \rangle_{2, \omega}}{\|p_k\|_{2, \omega}^2}$ angeben.

Darüber hinaus wissen wir, dass die beste polynomiale Approximation p^* einer Funktion $f \in L_2((a, b), \omega)$ mit ihrer eindeutigen orthogonalen Projektion auf $\mathcal{P}_N((a, b))$ zusammenfällt. Sei

$$P_N : L_2((a, b), \omega) \rightarrow \mathcal{P}_N((a, b))$$

die Orthogonalprojektion von $L_2((a, b), \omega)$ auf $\mathcal{P}_N((a, b))$. Dann erfüllt die Abbildung P_N die Bedingung

$$\langle f - P_N(f), q \rangle_{2, \omega} = 0 \quad \forall q \in \mathcal{P}_N((a, b))$$

und da $P_N(f) = p^*$ gilt, erhalten wir für die Orthogonalprojektion von f auf $\mathcal{P}_N((a, b))$ die einfache Darstellung aus (3.13), also

$$P_N(f) = \sum_{k=0}^N \hat{f}_k p_k.$$

Wir haben die Aufgabe in (3.12), das heißt insbesondere die Existenz und Eindeutigkeit einer Lösung dieser Aufgabe, unter der Voraussetzung, dass alle Polynome in dem zugrunde liegenden Hilbertraum $L_2((a, b), \omega)$ enthalten sind, diskutiert. In dem nächsten Abschnitt möchten wir analysieren, wann diese Voraussetzung erfüllt ist und welche Resultate wir bezüglich der Dichtheit der Polynome in dem Raum $L_2((a, b), \omega)$ daraus ableiten können.

3.3.2 Voraussetzungen und Vorbetrachtungen

Soll eine Funktion $f \in L_2((a, b), \omega)$ durch Polynome approximiert werden, so ist natürlich vorab zu fordern, dass alle Polynome in dem Raum $L_2((a, b), \omega)$ enthalten sind. Ist dies gewährleistet, so können wir uns vertiefend der folgenden Frage widmen:

Lässt sich jede Funktion $f \in L_2((a, b), \omega)$ beliebig genau durch Polynome approximieren?

Um diesen Sachverhalt genauer analysieren zu können, wiederholen wir als erstes den Begriff einer dichten Teilmenge eines normierten Raumes:

Definition 3.18 (vgl. [Zei95], S.84, Definition 1)

Sei $(X, \|\cdot\|)$ ein normierter Raum. Eine Teilmenge $M \subset X$ liegt dicht in X (bezüglich $\|\cdot\|$), falls $\overline{M} = X$ gilt, das heißt, falls zu jedem $x \in X$ und jedem $\varepsilon > 0$ ein $y = y(x, \varepsilon) \in M$ existiert, sodass $\|y - x\| < \varepsilon$ gilt.

Bemerkung 3.6. Eine äquivalente Formulierung der obigen Definition lautet: Eine Teilmenge $M \subset X$ heißt dicht in dem normierten Raum $(X, \|\cdot\|)$, falls zu jedem $x \in X$ eine konvergente Folge $\{y_n\}_{n \in \mathbb{N}}$ aus M mit $\|y_n - x\| \rightarrow 0$ für $n \rightarrow \infty$ existiert.

Mit Hilfe dieser Definition können wir die aufgeworfene Fragestellung präzisieren und formulieren sie wie folgt:

Liegen die algebraischen Polynome (bzgl. $\|\cdot\|_{2,\omega}$) dicht in dem Raum $L_2((a,b),\omega)$?

Wir beantworten die Frage schrittweise und konzentrieren uns erst einmal auf ein endliches Grundintervall (a, b) , $-\infty < a < b < \infty$. Fordern wir, dass

$$\mu_0 := \int_a^b \omega(t) dt < \infty \tag{3.14}$$

gilt, so ist die natürliche Voraussetzung, dass alle Polynome auf (a, b) Elemente des Raumes $L_2((a, b), \omega)$ sind, erfüllt. Darüber hinaus kann gezeigt werden, dass die Polynomfunktionen dicht in $L_2((a, b), \omega)$ liegen und demnach also eine beliebig genaue polynomiale Approximierbarkeit für endliche Intervalle gegeben ist. Wir geben an dieser Stelle die grundlegende Beweisidee an und beziehen uns dabei auf die Untersuchungen in [Sch71], S. 45 f.

Unter der Voraussetzung, dass die Bedingung (3.14) erfüllt ist, liegt der Unterraum der stetigen Funktionen mit kompaktem Träger dicht in $L_2(\mathbb{R}, \omega)$, vgl. [Sch71], S. 45 f., Satz 2.10. Um nun die Dichtheit der Polynome zu zeigen, müssen wir zu jeder Funktion $f \in L_2((a, b), \omega)$ und jedem $\varepsilon > 0$ ein Polynom $p = p(f, \varepsilon)$ finden, sodass

$$\|f - p\|_{2,\omega} < \varepsilon$$

gilt. Sei also $f \in L_2((a, b), \omega)$ gegeben und $\varepsilon > 0$ beliebig gewählt. Dann folgt sofort aus der Dichtheit der stetigen Funktionen mit kompakten Träger, dass eine Funktion $g = g(\varepsilon) \in C([a, b])$ mit

$$\|f - g\|_{2,\omega} < \frac{\varepsilon}{2} \tag{3.15}$$

existiert. Weiterhin erhalten wir auf endlichen Intervallen für jede stetige Funktion u und alle Polynome v die Abschätzung

$$\begin{aligned} \|u - v\|_{2,\omega}^2 &= \int_a^b |u(t) - v(t)|^2 \omega(t) dt \\ &\leq \max_{t \in [a,b]} |u(t) - v(t)|^2 \int_a^b \omega(t) dt = \|u - v\|_\infty^2 \int_a^b \omega(t) dt. \end{aligned}$$

Der Satz von Weierstraß erlaubt, für die Funktion $g \in C([a, b])$ ein Polynom $p = p(\varepsilon)$ derart zu wählen, dass unter Beachtung der obigen Rechnung mit $u := g$ und $v := p$ die Ungleichung

$$\|g - p\|_{2, \omega}^2 \leq \|g - p\|_{\infty}^2 \int_a^b \omega(t) dt < \frac{\varepsilon^2}{4} \quad (3.16)$$

erfüllt ist. Durch Anwendung der Minkowski-Ungleichung und mit Hilfe der Abschätzungen aus (3.15) und (3.16) können wir sofort auf

$$\|f - p\|_{2, \omega} \leq \|f - g\|_{2, \omega} + \|g - p\|_{2, \omega} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

schließen. Damit haben wir bewiesen, dass der Unterraum der Polynomfunktionen auf (a, b) dicht in $L_2((a, b), \omega)$ liegt, sofern (a, b) endlich ist.

Betrachten wir Intervalle (a, b) , $a < b$ mit $a = -\infty$ oder $b = \infty$, so lässt sich eine derartige Aussage über die Polynome nicht einfach übertragen, selbst dann nicht, wenn alle Polynome in dem Raum $L_2((a, b), \omega)$ enthalten sind. Um Letzteres garantieren zu können, haben wir für endliche Intervalle lediglich die Voraussetzung aus (3.14) benötigt. Für $a = -\infty$ oder $b = +\infty$ genügt das nicht, weshalb wir hier stets die Endlichkeit der sogenannten *Momente*

$$\mu_k := \int_a^b t^k \omega(t) dt < \infty, \quad k = 0, 1, 2, \dots \quad (3.17)$$

einfordern müssen. Dass dadurch die Dichtheit der Polynome in $L_2((a, b), \omega)$ über unendliche Intervalle jedoch im Allgemeinen nicht gesichert werden kann, zeigt ein Gegenbeispiel in [Sch71], S. 85.

In dem Fall eines endlichen Grundintervalls konnten wir die eingangs aufgeworfene Fragestellung leicht für alle Gewichte, die die Bedingung in (3.14) erfüllen, positiv beantworten. Eine derart allgemeine Antwort können wir für beliebige Gewichtsfunktionen auf unendlichen Bereichen nicht treffen. Dennoch existieren speziell gewählte Gewichte, für die die Dichtheit auch auf unendlichen Intervallen nachgewiesen werden kann. Insbesondere für das unbeschränkte Intervall $(0, \infty)$ und mit Gewichten der Gestalt $\omega_{\alpha, \beta}(t) := t^{\alpha} e^{-\beta t}$, $\alpha > -1, \beta > 0$ ist dies möglich und wir werden an späterer Stelle genauer darauf eingehen. Ein allgemeineres Resultat findet man in [Sch71], S.187, Satz 8.5., wobei hier zusätzlich vorausgesetzt werden muss, dass die Gewichtsfunktion für $|t| \rightarrow \infty$ hinreichend schnell abnimmt.

Abschließen möchten wir unsere Vorbetrachtungen mit dem Hinweis, dass wir die zu dem Raum $L_2((a, b), \omega)$ führende Klassenbildung im Rahmen der polynomialen Approximation vernachlässigen können, vgl. [Sch71], S. 53. Dabei nutzen wir aus, dass jede Äquivalenzklasse aus $L_2((a, b), \omega)$ höchstens einen stetigen Vertreter hat: Sei $(a, b) \subseteq \mathbb{R}$ ein endliches oder unendliches Grundintervall und angenommen, es existieren zwei voneinander verschiedene Funktionen $f, g \in C((a, b))$, die in derselben Äquivalenzklasse des $L_2((a, b), \omega)$ liegen. Nach Voraussetzung gibt es ein $\tau \in (a, b)$ mit $f(\tau) \neq g(\tau)$ und aufgrund der Stetigkeit der Funktionen f und g demnach auch eine Umgebung $U_{\varepsilon}(\tau) := \{t \in \mathbb{R} : |t - \tau| < \varepsilon\}$ ($\varepsilon > 0$) von τ mit $f(t) \neq g(t)$ für alle $t \in U_{\varepsilon}(\tau)$. Damit haben wir jedoch eine Umgebung vom Maß ungleich Null gefunden, auf der f und g verschieden sind. Dies ist ein Widerspruch dazu,

dass die Funktionen aus derselben Äquivalenzklasse gewählt wurden, wonach $f(t) = g(t)$ für fast alle $t \in (a, b)$ erfüllt ist und sich daher die Funktionen nur auf Nullmengen voneinander unterscheiden. Die Annahme ist demnach falsch und es gilt $f(t) = g(t)$ auf ganz (a, b) . Da jedes Polynom auf ganz \mathbb{R} stetig ist, können wir damit sofort schließen, dass jede Äquivalenzklasse höchstens ein Polynom enthält und in diesem Sinne von der Klassenbildung abgesehen werden kann.

3.3.3 Orthogonale Polynome und einige grundlegende Eigenschaften

In der Aufgabe (3.12) wird für festes $N \in \mathbb{N}_0$ der lineare Raum $\mathcal{P}_N((a, b))$ der algebraischen Polynome mit maximalen Grad N als Approximationsraum gewählt. Unter der Voraussetzung (3.14) bzw. (3.17) ist $\mathcal{P}_N((a, b))$ ein endlich-dimensionaler und abgeschlossener Teilraum des $L_2((a, b), \omega)$. Nach den vorangegangenen Ausführungen wissen wir, dass die eindeutig bestimmte Bestapproximation $p^* \in \mathcal{P}_N((a, b))$ einer Funktion $f \in L_2((a, b), \omega)$ explizit mit Hilfe einer Orthogonalbasis von $\mathcal{P}_N((a, b))$ dargestellt und somit die Lösung der Aufgabe (3.12) vollständig bestimmt werden kann. Diese Überlegung führt uns dazu, Möglichkeiten zur Berechnung einer solchen orthogonalen Polynombasis zu untersuchen. Es existiert eine große Auswahl an Literatur, in der die Theorie der orthogonalen Polynome detailliert studiert und deren Eigenschaften dargelegt werden. Wir konzentrieren uns in diesem Abschnitt auf wenige grundlegende Eigenschaften und verweisen für ein umfassendes Studium der orthogonalen Polynome auf die klassischen Werke [Fre69, Fun92, Sze39] und [Tri70] sowie auf Bücher zur Approximationstheorie und numerischen Mathematik, w.z.B. [Dav75, Isk18, Sch71] und [Sto05].

Für die nachfolgenden Ausführungen setzen wir stets voraus, dass die Bedingungen in (3.14) bzw. (3.17) erfüllt sind.

Existenz und Rekursionsformel In dem Abschnitt 3.2, vgl. Satz 3.7, S. 50, haben wir das konstruktive Verfahren von Gram-Schmidt kennengelernt, mit welchem man in jedem Hilbertraum ausgehend von einer linear unabhängigen Teilmenge ein ONS konstruieren kann. Wenden wir den Algorithmus auf die Monombasis $\{1, t, t^2, \dots, t^N\} \subset L_2((a, b), \omega)$ an, so erhalten wir bzgl. des Skalarproduktes $\langle \cdot, \cdot \rangle_{2, \omega}$ orthogonale Polynome p_0, p_1, \dots, p_N , welche eine Basis des Raumes $\mathcal{P}_N((a, b))$ bilden. Insbesondere ist damit die Existenz einer Orthogonalbasis von $\mathcal{P}_N((a, b))$ gesichert und zugleich ein konstruktives Verfahren zur Berechnung der orthogonalen Polynome gegeben. Bei dem Gram-Schmidt Verfahren wird die Normlänge aller Basiselemente auf den Wert Eins skaliert. An dieser Stelle gibt es auch verschiedene andere Möglichkeiten, die orthogonalen Polynome zu normieren. Die am Häufigsten verwendeten Normierungen sind:

- Alle Polynome besitzen einen *führenden Koeffizienten mit dem Wert Eins*, d.h. $p_0 = 1$ und für $k = 1, 2, \dots, N$ ist

$$p_k(t) = t^k + q_{k-1}(t)$$

mit $q_{k-1} \in \mathcal{P}_{k-1}((a, b))$.

- Die Polynomwerte werden bei $t = a$ oder $t = b$ fixiert⁶. So kann beispielsweise gefordert werden, dass alle Polynome bei $t = a$ auf den Wert Eins normiert sind, d.h. $p_k(a) = 1$, $k = 0, 1, \dots, N$.
- Die Länge der Polynome hat den Wert Eins. Entsprechend sind die Polynome in diesem Fall orthonormal, d.h. $\|p_k\| = 1$ für alle $k = 0, 1, \dots, N$. Die derart normierten Polynome werden oft als *normierte Orthogonalpolynome* bezeichnet.

Da es ohne Probleme möglich ist, durch geeignetes Umrechnen andere Darstellungen zu erhalten, und aufgrund dessen, dass die Skalierung mit einem konstanten Faktor (ungleich Null) für die hier betrachtete Orthogonalität ohne Bedeutung ist, sollte sich der Leser nicht durch die in der Literatur verwendeten unterschiedlichen Darstellungen von Orthogonalpolynomen verunsichern lassen. Oftmals ist es für das Aufzeigen bestimmter Resultate oder Zusammenhänge bequemer, eine bestimmte Normierung zu wählen und entsprechend je nach Zusammenhang eine geeignete Normierung der Orthogonalpolynome zu benutzen.

Wir können also festhalten, dass das Verfahren von Gram-Schmidt ein konstruktives Berechnungsverfahren liefert, mit welchem ein bis auf einen Skalierungs- bzw. Normierungsfaktor eindeutig bestimmtes orthogonales Polynomsystem bestimmt werden kann. Dennoch haben wir bereits in dem Beispiel 3.2 festgestellt, dass die dadurch gegebene Konstruktionsvorschrift rechnerisch sehr aufwendig ist. Von großem praktischen Interesse ist deshalb die nachfolgende Rekursionsformel, welche häufig für die numerische Berechnung von Orthogonalpolynomen zu einer gegebenen Gewichtsfunktion ω genutzt wird.

Satz 3.19 (vgl. [Isk18], S. 121, Theorem 4.16)

Zu dem Skalarprodukt $\langle \cdot, \cdot \rangle_{2, \omega}$ gibt es eindeutig bestimmte Polynome $p_k \in \mathcal{P}_k((a, b))$, $k = 0, 1, \dots$ mit

$$\langle p_i, p_j \rangle_{2, \omega} = 0 \text{ für } i \neq j$$

und mit führendem Koeffizienten Eins. Diese Polynome erfüllen die Drei-Term-Rekursionsformel

$$p_k(t) = (t - a_k)p_{k-1}(t) - b_k p_{k-2}(t), \quad k = 1, 2, \dots \quad (3.18)$$

mit den Setzungen $p_{-1} := 0$, $p_0 := 1$ und den Koeffizienten

$$a_k := \frac{\langle t p_{k-1}, p_{k-1} \rangle_{2, \omega}}{\|p_{k-1}\|_{2, \omega}^2} \text{ für } k \geq 1, \quad b_1 := 1, \quad b_k := \frac{\|p_{k-1}\|_{2, \omega}^2}{\|p_{k-2}\|_{2, \omega}^2} \text{ für } k \geq 2. \quad (3.19)$$

Es ist klar, dass die durch (3.18) definierten Polynome p_k exakt den Grad k besitzen. Weiterhin sind die Koeffizienten in (3.19) eindeutig bestimmt. Die Drei-Term-Rekursionsformel stellt daher sogar eine hinreichende Bedingung dar und kann genutzt werden, um die zu dem gewichteten Skalarprodukt $\langle \cdot, \cdot \rangle_{2, \omega}$ eindeutig bestimmten orthogonalen Polynome $p_k \in \mathcal{P}_k((a, b))$, $k = 0, 1, \dots$ mit führendem Koeffizienten Eins zu berechnen, vgl. [Isk18], S. 121,

⁶ Es gilt $p_k(a) \neq 0$ und $p_k(b) \neq 0$, vgl. Satz 3.20, S. 68

Bemerkung 4.17. Damit haben wir eine nützliche und effiziente Möglichkeit gefunden, mit der wir die Orthogonalpolynome leicht konstruieren können.

Eine weitere Beobachtung ist die Folgende: Für $N \in \mathbb{N}$ bilden die orthogonalen Polynome p_0, \dots, p_{N-1} aus dem vorangegangenen Satz offensichtlich eine Basis in $\mathcal{P}_{N-1}((a, b))$. Demnach lässt sich jedes Polynom $q \in \mathcal{P}_{N-1}((a, b))$ als Linearkombination der p_0, \dots, p_{N-1} schreiben und wir erhalten die Darstellung

$$q(t) = \sum_{j=0}^{N-1} c_j p_j(t), \quad c_j \in \mathbb{R}.$$

Für das orthogonale Polynom p_N ergibt sich dann sofort die einfache Rechnung

$$\langle q, p_N \rangle_{2,\omega} = c_0 \langle p_0, p_N \rangle_{2,\omega} + \dots + c_{N-1} \langle p_{N-1}, p_N \rangle_{2,\omega} = 0$$

und wir können festhalten, dass das Polynom p_N zu jedem Polynom kleineren Grades orthogonal ist, also für $q \in \mathcal{P}_{N-1}((a, b))$ stets

$$\langle q, p_N \rangle_{2,\omega} = 0$$

gilt, vgl. [Sto05], S.175, (3.6.9) Korollar.

Nullstellen und Gauß-Quadratur Abschließen möchten wir diesen Abschnitt mit einem kurzen (aber auf keinen Fall vollständigen) Überblick zu den Nullstellen orthogonaler Polynome und deren Zusammenhang mit der Gauß-Quadratur. Sie spielen im Rahmen von Kollokationsmethoden eine zentrale Rolle und werden zum Beispiel als Stützstellen Gaußscher Quadraturformeln gewählt. Aufgrund dessen ist das Studium wertvoller Eigenschaften der Nullstellen orthogonaler Polynome von besonderem Interesse und kann vielfältig in der auf Seite 66 angegebenen Literatur nachgelesen werden.

Wir möchten zunächst auf die folgenden nützlichen Eigenschaften der Nullstellen aufmerksam machen:

Satz 3.20 (vgl. [Sto05], S.175, (3.6.10) Satz)

Die Nullstellen t_1, \dots, t_N des orthogonalen Polynoms $p_N \in \mathcal{P}_N((a, b))$ sind alle reell, einfach und im Inneren des betrachteten Intervalls (a, b) gelegen.

Sollen nun die Nullstellen der orthogonalen Polynome numerisch berechnet werden, so kann dies mit Hilfe der sogenannten Eigenwert-Methode getan werden. In der vorangegangenen Diskussion haben wir gesehen, dass die orthogonalen Polynome mit Hilfe der Drei-Term-Rekursionsformel aus Satz 3.19 berechnet werden können. Studiert man nun deren Zusammenhang zu speziellen reellen Tridiagonalmatrizen genauer, so ergibt sich das nachfolgende Resultat:

Satz 3.21 (vgl. [Sto05], S. 179, (3.6.20) Satz)

Die Eigenwerte der Tridiagonalmatrix

$$T_N = \begin{bmatrix} a_1 & \sqrt{b_2} & & & & \\ \sqrt{b_2} & a_2 & \sqrt{b_3} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \sqrt{b_{N-1}} & a_{N-1} & \sqrt{b_N} & \\ & & & \sqrt{b_N} & a_N & \end{bmatrix}$$

entsprechen den Nullstellen $t_k, k = 1, \dots, N$ des Orthogonalpolynoms $p_N \in \mathcal{P}_N((a, b))$. Dabei sind a_k, b_k die Koeffizienten der Drei-Term-Rekursionsformel aus (3.18).

Für die Eigenwertberechnung gibt es verschiedene numerische Verfahren, wie zum Beispiel das Bisektions- oder das QR-Verfahren, vgl. [Sto05]. Die eigentliche Schwierigkeit bei der Eigenwert-Methode besteht darin, dass die Koeffizienten a_k, b_k à priori bestimmt werden müssen und deren Berechnung im Allgemeinen eine numerisch schwer lösbare Aufgabe darstellt. Eine weitere Möglichkeit zur Nullstellenberechnung orthogonaler Polynome ist durch eine iterative Methode gegeben, vgl. [STW11], S. 56. Für unsere Zwecke sind die hier beschriebenen Methoden ausreichend, weshalb wir auf weitere Verfahren nicht weiter eingehen werden.

Der nachfolgende Satz zeigt nun den Zusammenhang zwischen orthogonalen Polynomen und Gaußschen Quadraturformeln auf.

Satz 3.22 (vgl. [Sto05], S. 184, (3.6.12) Satz)

Sei $q \in \mathcal{P}_{2N-1}((a, b))$ ein beliebig gewähltes Polynom vom Grad $\leq 2N - 1$. Darüber hinaus seien t_1, \dots, t_N und w_1, \dots, w_N aus dem Raum der reellen Zahlen gewählt. Dann gilt

$$\int_a^b q(t)\omega(t)dt = \sum_{k=1}^N q(t_k)w_k$$

genau dann, wenn die nachfolgenden Eigenschaften erfüllt sind:

- (i) Die Stützstellen t_1, \dots, t_N sind gerade die Nullstellen des N -ten orthogonalen Polynoms $p_N \in \mathcal{P}_N((a, b))$ von Satz 3.19.
- (ii) Die Gewichte w_1, \dots, w_N sind alle strikt positiv und lösen das lineare Gleichungssystem

$$\sum_{k=1}^N p_i(t_k)w_k = \begin{cases} \langle p_0, p_0 \rangle_{2, \omega} & \text{falls } i = 0, \\ 0 & \text{falls } i = 1, 2, \dots, N - 1 \end{cases} .$$

Sind $\{t_k\}_{k=1}^N$ also die Nullstellen des N -ten orthogonalen Polynoms $p_N \in \mathcal{P}_N((a, b))$, dann existiert eine Menge von Quadraturgewichten $\{w_k\}_{k=1}^N$, sodass

$$\int_a^b q(t)\omega(t)dt = \sum_{k=1}^N q(t_k)w_k \quad \forall q \in \mathcal{P}_{2N-1} \tag{3.20}$$

gilt. Die obige Quadraturformel nennt man *Gaußsche Quadraturformel*. Die Gaußsche Quadraturformel hat den maximal möglichen Genauigkeitsgrad $2N - 1$ und ist daher optimal, vgl. z.B. [SS76], S. 211, (8.4.2).

Wir haben bereits aufgezeigt, dass die Nullstellen des N -ten Orthogonalpolynoms mit Hilfe der Tridiagonalmatrizen T_N aus Satz 3.21 berechnet werden können. Damit haben wir also zugleich eine Variante, um die Stützstellen der Gaußschen Quadraturformel (3.20) zu bestimmen. Offen bleibt, wie die zugehörigen Quadraturgewichte zu ermitteln sind. Auch hier lässt sich wieder ein Zusammenhang zu den Tridiagonalmatrizen T_N herstellen.

Satz 3.23 (vgl. [Sto05], S. 187, (3.6.21) Satz)

Sei t_k ein Eigenwert der Tridiagonalmatrix T_N und wird mit $v^k = (v_1^k, \dots, v_N^k)^T$ der gemäß

$$T_N v^k = t_k v^k$$

berechnete zugehörige Eigenvektor mit der Normierung

$$(v^k)^T \cdot v^k = \langle p_0, p_0 \rangle_{2,\omega} = \int_a^b \omega(t) dt$$

bezeichnet. Dann ist für die Quadraturgewichte die Identität

$$w_k = (v_1^k)^2, \quad k = 1, 2, \dots, N.$$

erfüllt.

Offensichtlich ist allein die erste Komponente v_1^k des Eigenvektors v^k zur Bestimmung des Gewichtes w_k von Interesse. An späterer Stelle wird die Ermittlung der Stützstellen und Gewichte mit Hilfe der Eigenwerte und -vektoren für die Implementierung eines Algorithmus benötigt. Dafür genügen die bisher aufgezeigten Ergebnisse und wir können unseren kurzen Einblick in die Theorie der orthogonalen Polynome beenden. Eine ausführlichere Diskussion zu der effektiven numerischen Berechnung der Eigenwerte und -vektoren mit Hilfe eines QR-Verfahrens geben GOLUB und WELSCH in [GW69].

In Abhängigkeit des gewählten Integrationsintervalls (a, b) und der jeweiligen Gewichtsfunktion ω ergeben sich verschiedene orthogonale Polynomsysteme. Die klassischen orthogonalen Polynomsysteme sind die Jacobi-, Laguerre- und Hermite-Polynome. Wir interessieren uns speziell für die bereits erwähnten *verallgemeinerten Laguerre-Polynome*, deren Eigenschaften und mit ihnen zusammenhängende Approximationsergebnisse wir in einem eigenen Abschnitt untersuchen werden.

4 Linear-quadratische Steuerungsaufgaben mit unendlichem Zeithorizont

4.1 Formulierung der Aufgabe $(LQ)_\infty$

Ziel dieser Arbeit ist es, eine indirekte numerische Lösungsmethode für Steuerungsaufgaben mit unendlichem Zeithorizont vorzustellen. Dabei konzentrieren wir uns auf Aufgaben mit einer linear-quadratischen Zielfunktion und einem System linearer Zustandsgleichungen. Das grundlegende Optimierungsproblem $(LQ)_\infty$, auf welches sich alle weiteren Ergebnisse der vorliegenden Arbeit beziehen, möchten wir in diesem Kapitel vorstellen und formulieren es wie folgt:

Minimiere

$$J(x, u) = \int_0^\infty f(t, x(t), u(t)) \omega_\beta(t) dt \quad (4.1)$$

mit dem Integranden

$$f(t, \xi, v) := \frac{1}{2} \{ \xi^T W(t) \xi + v^T R(t) v \} + w^T(t) \xi + r^T(t) v \quad (4.2)$$

bezüglich

$$(x, u) \in W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta), \omega_\beta(t) := e^{-\beta t}, \beta > 0,$$

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + C(t) \quad \text{f.ü. auf } (0, \infty), \quad (4.3)$$

$$x(0) = x_0 \in \mathbb{R}^n, \quad (4.4)$$

$$u(t) \in U \subseteq \mathbb{R}^m \quad \text{f.ü. auf } (0, \infty). \quad (4.5)$$

Dabei sind $W : [0, \infty) \rightarrow \mathbb{R}^{n \times n}$, $R : [0, \infty) \rightarrow \mathbb{R}^{m \times m}$, $w : [0, \infty) \rightarrow \mathbb{R}^n$, $r : [0, \infty) \rightarrow \mathbb{R}^m$ und $A : [0, \infty) \rightarrow \mathbb{R}^{n \times n}$, $B : [0, \infty) \rightarrow \mathbb{R}^{n \times m}$, $C : [0, \infty) \rightarrow \mathbb{R}^n$ gegebene Funktionen, die wir noch genauer spezifizieren werden.

Das zu minimierende Zielfunktional liegt in Form eines Integralfunktionals vor. Wir verstehen das Integral in (4.1) als Lebesgue-Integral und werden im Folgenden bei der Verwendung eines Integrals auch immer von einem Integral im Lebesgueschen Sinne ausgehen. Im Hinblick auf eine wie hier betrachtete Steuerungsaufgabe mit unendlichem Zeithorizont ist es wichtig, genau zwischen den Integraltypen zu differenzieren, da aufgrund der verschiedenen Summenbildungen die einzelnen Integraltypen zu unterschiedlichen Integralwerten führen können. Ein

Beispiel, bei dem die zulässigen Mengen und dadurch die Optimalwerte in Abhängigkeit des gewählten Integraltyps unterschiedlich ausfallen, wird in [PLW08] vorgestellt.

Als Funktionenraum wählen wir für den Zustand x den (uniform) gewichteten Sobolev-Raum $W_2^{1,n}((0, \infty), \omega_\beta)$, während wir für die Steuerung u den gewichteten Lebesgue-Raum $L_2^m((0, \infty), \omega_\beta)$ zugrunde legen. Gemäß der Diskussion in dem Abschnitt 2.4.1 stellen wir fest, dass es sich bei der hier verwendeten Gewichtsfunktion $\omega_\beta(t) := e^{-\beta t}$, $\beta \in \mathbb{R}$, $\beta > 0$ sogar um eine Dichtefunktion handelt, das heißt es ist $\omega_\beta \in C([0, \infty)) \cap L_1((0, \infty))$.

Weiterhin stimmt das Gewicht der Funktionenräume mit der im Zielfunktional (4.1) auftauchenden Gewichtsfunktion $\omega_\beta(t) := e^{-\beta t}$, $\beta > 0$ überein. Tritt im Zielfunktional ein Gewicht $\tilde{\omega}$ verschieden von ω_β auf, so kann gemäß der Rechnung

$$\begin{aligned} J(x, u) &= \int_0^\infty f(t, x(t), u(t)) \tilde{\omega}(t) dt = \int_0^\infty f(t, x(t), u(t)) \tilde{\omega}(t) \cdot \frac{\omega_\beta(t)}{\omega_\beta(t)} dt \\ &= \int_0^\infty \tilde{f}(t, x(t), u(t)) \omega_\beta(t) dt \end{aligned}$$

das Gewicht ω_β erzeugt werden. Der Integrand ist dabei durch

$$\tilde{f}(t, \xi, v) := f(t, x(t), u(t)) \frac{\tilde{\omega}(t)}{\omega_\beta(t)}$$

gegeben und die Voraussetzungen müssen nun für diesen geprüft werden. Im Rahmen dieser Arbeit verzichten wir auf solche Fälle und beziehen uns im Folgenden auf Aufgaben der Gestalt $(LQ)_\infty$, bei denen das Gewicht der Funktionenräume stets mit der im Zielfunktional (4.1) auftauchenden Gewichtsfunktion übereinstimmt.

Die Tatsache, dass die gewählten Funktionenräume Hilberträume sind, liefert den Vorteil, dass die Aufgabe $(LQ)_\infty$ sowohl im Kontext der Hilbertraumtheorie analysiert als auch klassische Methoden der konvexen Funktionalanalysis genutzt werden können. Darüber hinaus ist durch die Raumwahl eine sinnvolle Interpretation der Anfangswertaufgabe in (4.3)-(4.4) gegeben. Diesen Aspekt werden wir in dem Abschnitt 4.3 genauer studieren. Für die nachfolgenden Untersuchungen führen wir noch weitere Begriffe und Definitionen ein.

Ein Funktionenpaar $(x, u) \in W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta)$ oder auch *Steuerprozess* nennen wir *zulässig* für die Aufgabe $(LQ)_\infty$, falls u der in (4.5) vorgegebenen Beschränkung genügt, x eine Lösung der Dynamik in (4.3) zur gegebenen Anfangsbedingung (4.4) ist und (x, u) einen endlichen Wert des Lebesgue-Integrals in (4.1) liefert. Die Funktion x wird dann als die zu der Steuerung u gehörige *Zustandstrajektorie* bezeichnet.

Definition 4.1

Ein zulässiger Steuerprozess (x^*, u^*) wird als **global optimal** in der Aufgabe $(LQ)_\infty$ bezeichnet, falls alle anderen zulässigen Paare (x, u) die Bedingung

$$J(x, u) - J(x^*, u^*) \geq 0$$

erfüllen.

In der Literatur existiert eine Vielzahl unterschiedlicher Definitionen von Optimalität für Steuerungsaufgaben mit einem unendlichen Zeithorizont. Diese sind notwendig, da es a priori häufig nicht möglich ist, die Existenz und Endlichkeit des Lebesgue-Integrals als gegeben anzunehmen. Eine Auswahl unterschiedlicher globaler und lokaler Optimalitätsbegriffe wird in [Lyk10], S. 22 ff. angeführt und deren Verwendung diskutiert. Da wir bei einem zulässigen Steuerprozess die Existenz und die Endlichkeit des Zielfunktional einfordern, ist es sinnvoll, im Rahmen dieser Arbeit das Optimalitätskriterium gemäß der Definition 4.1 zugrunde zu legen. Insbesondere werden wir sehen, dass unsere Grundvoraussetzungen an das linear-quadratische Steuerungsproblem $(LQ)_\infty$ und insbesondere diejenigen, welche wir an die im Zielfunktional auftretenden Funktionen stellen, die Endlichkeit des Lebesgue-Integrals in (4.1) für Prozesse $(x, u) \in W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta)$ garantieren. Darüber hinaus benötigen wir bestimmte Voraussetzungen an $(LQ)_\infty$, um einen Existenz- und Eindeutigkeitssatz zu beweisen und notwendige Optimalitätsbedingungen zu untersuchen. Dabei können wir auf die umfassende Forschungsarbeit unserer Arbeitsgruppe zurückgreifen. Wir stellen deren Inhalte an dieser Stelle zusammenfassend dar, um insbesondere aufzuzeigen, welches Grundgerüst bezüglich der Beweisführung bereits vorhanden ist und von uns in den nächsten Abschnitten verwendet werden kann.

In der Arbeit *Hilbert Space Treatment of Optimal Control Problems with Infinite Horizon* stellt Pickenhain eine Klasse von Steuerungsproblemen über den unendlichen Zeithorizont $(0, \infty)$ vor, vgl. [Pic14]. Die Besonderheit liegt in der Wahl der Funktionenräume. Für den skalaren Zustand wird ein gewichteter Sobolev- und für die skalare Steuerung ein gewichteter Lebesgue-Raum verwendet. Bedingt durch diese Raumwahl, lässt sich die Aufgabenklasse in den Kontext der Hilbertraumtheorie einbetten und entsprechend derer Beweistechniken studieren. Pickenhain formuliert notwendige Optimalitätsbedingungen für den Fall eines konvexen, kompakten Steuerbereiches in Form eines Pontryaginschen Maximumprinzips, vgl. [Pic14], S. 173, Theorem 1. Die dort genutzte Beweistechnik legt den Grundstein für die weiteren Arbeiten in unserer Forschungsgruppe bezüglich notwendiger Optimalitätsbedingungen. Das Wesentliche des Beweises ist die geschickte Verwendung eines Trennungssatzes, welche durch die Formulierung in Hilberträumen legitimiert wurde. Zusätzlich konnte in dieser Arbeit gezeigt werden, dass sich aufgrund der Wahl der gewichteten Funktionenräume auch in dem Fall des unendlichen Zeithorizontes natürliche Transversalitätsbedingungen für die Adjungierte ergeben.

Die spezielle Klasse linear-quadratischer Steuerungsaufgaben über $(0, \infty)$ studiert Pickenhain in *Infinite Horizon Optimal Control Problems in the Light of Convex Analysis*, vgl. [Pic15]. Dabei untersucht sie Aufgaben vom Typ $(LQ)_\infty$ mit skalarwertigen Zustand und skalarwertiger Steuerung ($n = m = 1$). Wieder liefern ein gewichteter Zustands- und Steuerungsraum die Möglichkeit, die Aufgabe im Rahmen der Hilbertraumtheorie zu untersuchen. Hinzu kommt die linear-quadratische Struktur der Aufgabe. Diese wird gekonnt genutzt, um sowohl die Existenz einer Lösung als auch ein Pontryaginschen Maximumprinzips zu beweisen. Im Vergleich zu [Pic14] kann zusätzlich auf die Kompaktheit des Steuerbereiches verzichtet werden, das heißt, die notwendigen Bedingungen gelten auch für einen konvexen und (nur) abgeschlossenen Steuerbereich, vgl. [Pic15], S. 183 ff., Theorem 4. Eine entscheidende Rolle spielen dabei die Wahl und die speziellen Charakteristika der Gewichtsfunktion.

Aufgaben der Klasse $(LQ)_\infty$ mit einer vektorwertigen Zustands- und Steuergröße wurden in

der Arbeit von Ziemann und Pickenhain studiert, vgl. [ZP18]. Hierin werden ausgehend von einem konvexen, kompakten Steuerbereich und aufbauend auf den Techniken und Methoden von [Pic15] notwendige Optimalitätsbedingungen bewiesen. Eine detaillierte Beweisdarstellung des Pontryaginschen Maximumprinzips für diese Aufgaben findet man in [Zie15].

In *Regulator Problems on Unbounded Domains* haben wir das linear-quadratische Steuerungsproblem $(LQ)_\infty$ für einen unbeschränkten Steuerbereich $U = \mathbb{R}^m$ und zunächst für $w(t) \equiv 0$, $r(t) \equiv 0$, $C(t) \equiv 0$ und $B(t) \equiv 1$ untersucht und notwendige Optimalitätsbedingungen nachgewiesen, vgl. [PB18], S. 849, Theorem 2. Indem wir diese Setzungen fallen lassen, möchten wir das Studium der Aufgabenklasse $(LQ)_\infty$ in dieser Arbeit nun fortführen. Dabei ergibt sich der Vorteil, dass wir bei der Beweisstruktur an die Methoden und Argumente der aufgeführten Arbeiten anknüpfen können. Insbesondere bei der Existenztheorie greifen wir dabei auf Resultate und Ideen von Lykina zurück. Sie studierte im Rahmen ihrer Dissertationsschrift [Lyk10] und in der Arbeit *An Existence Theorem for a Class of Infinite Horizon Optimal Control Problems*, vgl. [Lyk16] ausführlich die Fragestellung der Existenz einer optimalen Steuerung für Aufgaben mit unendlichem Zeithorizont. Auch hier liegt die Schlüsselidee darin, gewichtete Sobolev- bzw. gewichtete Lebesgue-Räume als Zustand- bzw. Steuerungsräume einzuführen. Mit Hilfe des verallgemeinerten Satzes von Weierstraß wird ein Existenzsatz präsentiert, dessen Beweis Argumente enthält, auf welche wir in den nachfolgenden Abschnitten aufbauen können.

Mit der Vorstellung der für unsere weiteren Ausführungen relevanten Arbeiten möchten wir diesen Abschnitt schließen und stellen die für die Aufgabe $(LQ)_\infty$ benötigten Grundvoraussetzungen vor.

4.2 Voraussetzungen

Die bereits erwähnten Grundvoraussetzungen an die Aufgabe $(LQ)_\infty$ lauten wie folgt:

(V1) Die Funktionen aus dem Zielfunktional wählen wir gemäß

$$\begin{aligned} W &\in L_\infty^{n \times n}((0, \infty)) \cap C([0, \infty), \mathbb{R}^{n \times n}), \quad R \in L_\infty^{m \times m}((0, \infty)) \cap C([0, \infty), \mathbb{R}^{m \times m}), \\ w &\in L_2^n((0, \infty), \omega_\beta) \cap C([0, \infty), \mathbb{R}^n), \quad r \in L_2^m((0, \infty), \omega_\beta) \cap C([0, \infty), \mathbb{R}^m). \end{aligned}$$

und die Funktionen aus der Differentialgleichung seien durch

$$\begin{aligned} A &\in L_\infty^{n \times n}((0, \infty)) \cap C([0, \infty), \mathbb{R}^{n \times n}) \\ B &\in L_\infty^{n \times m}((0, \infty)) \cap C([0, \infty), \mathbb{R}^{n \times m}) \\ C &\in L_\infty^n((0, \infty)) \cap C([0, \infty), \mathbb{R}^n) \end{aligned}$$

festgelegt.

(V2) An die Struktur der Matrizen $W(t)$ und $R(t)$ fordern wir das Erfülltsein der folgenden Bedingungen:

(V2a) Die Matrizen $W(t)$ und $R(t)$ sind für alle $t \in \mathbb{R}_+$ symmetrisch, das heißt es gilt $W^T(t) = W(t)$ und $R^T(t) = R(t)$ für alle $t \in \mathbb{R}_+$. (Hieraus folgt sofort, dass $W(t)$ und $R(t)$ nur reelle Eigenwerte besitzen.)

(V2b) Die Matrizen $W(t)$ und $R(t)$ sind positiv semidefinit, das heißt es gilt für alle $t \in \mathbb{R}_+$

$$\xi^T W(t) \xi \geq 0 \quad \forall \xi \in \mathbb{R}^n \quad \text{und} \quad v^T R(t) v \geq 0 \quad \forall v \in \mathbb{R}^m.$$

(V3) In Abhängigkeit der Matrix $A(t)$ und mit Hilfe der Notationen

$$\underline{W} := \inf_{t \in [0, \infty)} \lambda_{\min}(W)(t) \quad \text{und} \quad \underline{R} := \inf_{t \in [0, \infty)} \lambda_{\min}(R)(t),$$

wobei $\lambda_{\min}(W)(t)$ den vom Parameter t abhängigen kleinsten Eigenwert der symmetrischen Matrix $W(t)$ bezeichnet und $\lambda_{\min}(R)(t)$ der vom Parameter t abhängige kleinste Eigenwert der symmetrischen Matrix $R(t)$ ist, soll eine der nachfolgenden Bedingungen gelten:

(V3a) In dem Fall $A \equiv 0$ fordern wir, dass $\underline{W} > 0$ und $\underline{R} > 0$ gilt.

(V3b) Ist $A \not\equiv 0$, so setzen wir voraus, dass $\underline{W} > 0$ ist und darüber hinaus die Ungleichung

$$\left(\underline{R} - \underline{W} \cdot \frac{\|B\|_{(\infty)}^2}{\|A\|_{(\infty)}^2} \right) > 0$$

erfüllt ist.

(V4) Der Steuerbereich U sei eine abgeschlossene und konvexe Teilmenge des \mathbb{R}^m .

(V5) Für die Menge der in $(LQ)_\infty$ zulässigen Prozesse \mathcal{Z} setzen wir voraus, dass diese nicht leer ist und mindestens zwei voneinander verschiedene Elemente enthält. Es existiert also ein Paar $(\bar{x}, \bar{u}) \in \mathcal{Z}$, d.h. $\mathcal{Z} \neq \emptyset$ und es ist $\mathcal{Z} \neq \{(\bar{x}, \bar{u})\}$.

In der Voraussetzung (V1) legen wir die Funktionenräume für die in die Aufgabe eingehenden Funktionen fest. Dadurch wird sichergestellt, dass das Zielfunktional und die rechte Seite der Dynamik wohldefiniert sind. Aus der linear-quadratischen Struktur des Integranden

$$f(t, \xi, v) := \frac{1}{2} \{ \xi^T W(t) \xi + v^T R(t) v \} + w^T(t) \xi + r^T(t) v$$

folgt sofort, dass die Funktion $f(t, \xi, v)$ stetig differenzierbar in den Variablen ξ und v ist. Da die Stetigkeit der Funktionen W, R, w, r in (V1) gefordert wird, ergibt sich die Stetigkeit des Integranden bezüglich all seiner Argumente t, ξ, v . Durch die Forderungen in (V1) und (V2a) ist gewährleistet, dass das Zielfunktional für jedes Paar $(x, u) \in W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta)$, also insbesondere auch für diejenigen Paare, welche den Bedingungen (4.3)-(4.5) genügen, einen endlichen Wert liefert (vgl. Lemma 4.4 und Lemma 4.5). Somit kann aufgrund der Behandlung der Steuerungsaufgabe in den gewichteten Funktionenräumen die

Existenz des Lebesgue-Integrals in (4.1) gesichert werden und wir können im Folgenden von einem endlichen Zielfunktionswert im Sinne des Lebesgue-Integrals ausgehen.

Die $\|\cdot\|_{(\infty)}$ -Norm der Matrixfunktionen $W \in L_{\infty}^{n \times n}((0, \infty))$, $R \in L_{\infty}^{m \times m}((0, \infty))$, $A \in L_{\infty}^{n \times n}((0, \infty))$ und $B \in L_{\infty}^{n \times m}((0, \infty))$ kann mit Hilfe der entsprechenden Eigenwerte beschrieben werden, vgl. Abschnitt 2.2.3, speziell Lemma 2.24. Darüber hinaus ergibt sich aufgrund der Stetigkeit dieser Funktionen (vgl. Voraussetzung (V1)) die Identität zum klassischen Supremum, es gilt also

$$\|\cdot\|_{(\infty)} := \operatorname{ess\,sup}_{t \in (0, \infty)} |\cdot|_2 = \sup_{t \in (0, \infty)} |\cdot|_2.$$

Basierend auf der Voraussetzung (V2), nach der die Matrizen $W(t)$ und $R(t)$ symmetrisch und positiv semidefinit sind, erhalten wir

$$\|W\|_{(\infty)} = \sup_{t \in (0, \infty)} \lambda_{\max}(W)(t), \quad \|R\|_{(\infty)} = \sup_{t \in (0, \infty)} \lambda_{\max}(R)(t).$$

Dabei ist $\lambda_{\max}(W)(t)$ der vom Parameter t abhängige größte Eigenwert der Matrix $W(t)$ und entsprechend $\lambda_{\max}(R)(t)$ der vom Parameter t abhängige größte Eigenwert der Matrix $R(t)$. Für die Matrizen $A(t)$ und $B(t)$ ergibt sich

$$\|A\|_{(\infty)} = \sup_{t \in (0, \infty)} \sqrt{\lambda_{\max}(A^T A)(t)}, \quad \|B\|_{(\infty)} = \sup_{t \in (0, \infty)} \sqrt{\lambda_{\max}(B^T B)(t)},$$

wobei $\lambda_{\max}(A^T A)(t)$ den vom Parameter t abhängigen größten Eigenwert der Matrix $(A^T A)(t)$ und $\lambda_{\max}(B^T B)(t)$ den vom Parameter t abhängigen größten Eigenwert der Matrix $(B^T B)(t)$ bezeichnet.

Bemerkung 4.1. Die Voraussetzung (V3) zieht eine weitere Eigenschaft der Matrixfunktionen $W \in L_{\infty}^{n \times n}((0, \infty))$ und $R \in L_{\infty}^{m \times m}((0, \infty))$ nach sich. Ist nämlich $\underline{W} > 0$, so folgt auch $\lambda_{\min}(W)(t) > 0$ für alle $t \in [0, \infty)$. Da $W(t)$ für alle $t \in [0, \infty)$ symmetrisch ist, ergibt sich daraus sofort deren positive Definitheit und es gilt

$$\xi^T W(t) \xi \geq \lambda_{\min}(W)(t) |\xi|_2^2 \geq \underline{W} |\xi|_2^2 \quad \forall \xi \in \mathbb{R}^n \quad (4.6)$$

für alle $t \in [0, \infty)$, vgl. z.B. [GK99], S. 330, Lemma B.4. und Satz B.5. Daher wird die Existenz einer von t unabhängigen Konstanten $k > 0$ garantiert, sodass für alle $t \in [0, \infty)$

$$\xi^T W(t) \xi \geq k |\xi|_2^2 \quad \forall \xi \in \mathbb{R}^n$$

gilt. Dies entspricht gerade der Charakterisierung einer positiv definiten Matrixfunktion. Analog kann die Diskussion bezüglich $\underline{R} > 0$ geführt werden. Entsprechend können wir bei dem Erfülltsein der Voraussetzung (V3) von positiv definiten Matrixfunktionen $W \in L_{\infty}^{n \times n}((0, \infty))$ und $R \in L_{\infty}^{m \times m}((0, \infty))$ sprechen und meinen damit die Existenz zweier von t unabhängigen Konstanten $k_1, k_2 > 0$, sodass die obige Ungleichung für $W(t)$ bzw. analog für $R(t)$ gilt.

Mit Hilfe der eingeführten Voraussetzungen (V1) – (V5), von deren Erfülltsein wir im

Folgenden stets ausgehen werden, können wir uns nun einer genauen Analyse des linear-quadratischen Steuerungsproblem widmen und beginnen damit, die zugrunde gelegte Anfangswertaufgabe in (4.3)-(4.4) zu analysieren.

4.3 Differentialgleichungen in Sobolev-Räumen

Um die Wohldefiniertheit der Steuerungsaufgabe $(LQ)_\infty$ sicherzustellen, ist es unabdingbar, das Anfangswertproblem in (4.3)-(4.4) sinnvoll zu interpretieren. Für Sobolev-Funktionen $x \in W_2^{1,n}((0, \infty), \omega_\beta)$ können wir zunächst nur von einer schwach formulierten Differentialgleichung

$$\mathcal{D}x(\cdot) = \varphi(\cdot, x(\cdot)) \text{ auf } (0, \infty)$$

ausgehen. Wieso wir dennoch berechtigt sind, zu einer punktweisen Formulierung überzugehen, soll in diesem Abschnitt erläutert werden.

Wir motivieren die Diskussion einer Differentialgleichung bzw. eines Anfangswertproblems in geeigneten Sobolev-Räumen, indem wir zunächst einen Blick auf ein klassisches Anfangswertproblem der Form

$$\dot{x} = \varphi(t, x), \quad x(t_0) = x_0 \tag{4.7}$$

werfen. Dabei sei die Funktion $\varphi : D \rightarrow \mathbb{R}^n$ auf einer offenen Menge $D \subseteq \mathbb{R} \times \mathbb{R}^n$ definiert und $(t_0, x_0) \in D$. Im Folgenden verstehen wir die auftretenden Funktionen stets als n -dimensionale Spaltenvektoren,

$$x(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix}, \quad \varphi(t) = \begin{pmatrix} \varphi_1(t, x) \\ \vdots \\ \varphi_n(t, x) \end{pmatrix}$$

und vereinbaren, dass die Ableitungen und Integrale der Vektorfunktionen komponentenweise zu interpretieren sind. Entsprechend sind auch alle Aussagen bezogen auf eine Vektorfunktion, komponentenweise zu verstehen (z.B. bedeutet $x = x(t)$ sei stetig differenzierbar, dass $x_i = x_i(t)$ für alle $i = 1, \dots, n$ stetig differenzierbar ist).

In der klassischen Theorie ist eine auf einem Intervall $I \subseteq \mathbb{R}$ mit $t_0 \in I$ definierte Funktion $x = x(t) : I \rightarrow \mathbb{R}^n$ eine Lösung der obigen Anfangswertaufgabe (in I), wenn x in I differenzierbar ist und die folgenden Eigenschaften besitzt:

1. Der Graph von x ist in D enthalten, d.h. $(t, x(t)) \in D$ für alle $t \in I$.
2. Es gilt $\dot{x}(t) = \varphi(t, x(t))$ für alle $t \in I$ und $x(t_0) = x_0$.

Nach dem Satz von Peano¹ genügt es, die bloße Stetigkeit der rechten Seite φ zu fordern, um die Existenz einer Lösung zu garantieren. Darüber hinaus lässt sich unter dieser Vorausset-

¹ Der Beweis des Satzes geht auf den italienischen Mathematiker Giuseppe Peano (1858-1932) zurück und ist zum Beispiel in [Heu04], S. 136 zu finden.

zung die Anfangswertaufgabe äquivalent in eine Integralgleichung umschreiben, d.h. eine auf I stetige Funktion x löst die Anfangswertaufgabe (4.7) auf I genau dann, wenn die folgende Integralgleichung gilt:

$$x(t) = x_0 + \int_{t_0}^t \varphi(\tau, x(\tau)) d\tau \quad \text{für alle } t \in I. \quad (4.8)$$

Zusätzlich liefert die Stetigkeit der rechten Seite φ automatisch eine höhere Regularität der klassischen Lösung. Ist nämlich x eine Lösung von (4.7) auf I , so ist x per Definition differenzierbar, also insbesondere auch stetig auf I . Aufgrund der Voraussetzung $\varphi \in C(D, \mathbb{R}^n)$ ist auch die Komposition $t \mapsto \varphi(t, x(t))$ auf I stetig und da $\dot{x}(t) = \varphi(t, x(t))$ für alle $t \in I$ gilt, ist x sogar stetig differenzierbar auf I , d.h. $x \in C^1(I, \mathbb{R}^n)$. Ist φ also stetig, so folgt sofort, dass eine klassische Lösung von (4.7) eine stetig differenzierbare Funktion sein muss.

Im Kontext der optimalen Steuerung ergibt sich in natürlicher Weise ein Anfangswertproblem, bei dem die rechte Seite der Differentialgleichung in der Form $\tilde{\varphi}(t, x(\cdot), u(\cdot))$ auftritt. Da die auftretende Steuerung u nicht notwendiger Weise stetig sein muss, sondern lediglich als Lebesgue-messbar vorausgesetzt wird, entstehen demnach Funktionen

$$\varphi(t, x(\cdot)) := \tilde{\varphi}(t, x(\cdot), u(\cdot))$$

mit möglicher Unstetigkeit bezüglich t . Da der klassische Existenzsatz auf der Stetigkeit der rechten Seite beruht, stößt man in dem Fall einer Differentialgleichung mit unstetiger rechter Seite an die Grenzen des klassischen Lösungsbegriffs. Daher ist es erforderlich, den klassischen Lösungsbegriff zu erweitern. Ist die Funktion $\varphi = \varphi(t, x)$ stetig in x , jedoch unstetig in t , studiert man die sogenannten verallgemeinerten Lösungen im Sinne von Carathéodory.

Wie bereits erwähnt, ist in der klassischen Theorie (mit $\varphi \in C(D, \mathbb{R}^n)$ und $x \in C^1(I, \mathbb{R}^n)$) die Integralgleichung (4.8) äquivalent zu dem Anfangswertproblem (4.7). Dies folgt direkt durch Anwenden des Hauptsatzes der Differential- und Integralgleichung und liefert uns damit insbesondere die Möglichkeit, die Funktion x selbst aus ihrer Ableitung wiederherstellen zu können. Lässt man nun die Stetigkeit der rechten Seite fallen, so müssen wir uns fragen, unter welchen möglichst allgemeinen Voraussetzungen es dennoch realisierbar ist, von einer Differentialgleichung zu einer Integralgleichung überzugehen. Die Antwort liefert der Hauptsatz der Differential- und Integralrechnung für das Lebesgue-Integral (H. LEBESGUE (1904), G. VITALI (1905)).

Satz 4.2 ([Els11], S. 304, 4.1.4)

- (1) Ist die Funktion $f : [a, b] \rightarrow \mathbb{R}$ mit $-\infty < a < b < \infty$ absolut stetig und setzt man $f'(t) := 0$ für alle $t \in [a, b]$, in denen f nicht differenzierbar ist, so ist f' Lebesgue-integrierbar auf $[a, b]$ und es gilt

$$f(t) - f(a) = \int_a^t f'(\tau) d\tau, \quad a \leq t \leq b.$$

- (2) Ist umgekehrt die Funktion $f : [a, b] \rightarrow \mathbb{R}$ Lebesgue-integrierbar auf $[a, b]$, so ist die

Integralfunktion

$$F(t) := \int_a^t f(\tau) d\tau, \quad a \leq t \leq b$$

absolut stetig auf $[a, b]$ und es gilt $F' = f$ fast überall in $[a, b]$.

Der vorangegangene Satz besagt, dass die absolut stetigen Funktionen gerade die aus den (Lebesgue-)integrierbaren Funktionen f entstehenden Integralfunktionen F sind und darüber hinaus absolut stetige Funktion aus ihrer Ableitung rekonstruiert werden können. Darauf aufbauend ergibt sich das Lösungskonzept nach Carathéodory. Kehren wir zu dem Anfangswertproblem in (4.7) zurück und betrachten dieses nun mit einer in t unstetigen rechten Seite φ . Wir fordern, dass die Funktion φ die sogenannten *Carathéodory-Bedingungen* erfüllt. Diese lauten, vgl. [Fil88], S. 3 ff.:

Sei $I \subset \mathbb{R}$ ein Intervall, $D \subset \mathbb{R}^n$. Die Funktion $\varphi : I \times D \rightarrow \mathbb{R}^n$ erfüllt die Carathéodory-Bedingungen (auf $I \times D$), falls sie die folgenden Eigenschaften besitzt:

1. Die Funktion $\varphi = \varphi(t, \xi)$ ist für fast alle $t \in I$ stetig in ξ .
2. Die Funktion $\varphi = \varphi(t, \xi)$ ist für jedes $\xi \in D$ messbar in t .
3. Es existiert eine nichtnegative, messbare Funktion $m : I \rightarrow \mathbb{R}$, sodass $|\varphi(t, \xi)| \leq m(t)$ für alle $(t, \xi) \in I \times D$ erfüllt ist. In dem Fall, dass t in I unbeschränkt ist, soll diese Bedingung auf jedem endlichen Teilintervall von I erfüllt sein.

Die Differentialgleichung $\dot{x} = \varphi(t, x)$ heißt *Carathéodory-Gleichung*, falls φ die Carathéodory-Bedingungen erfüllt. Eine Funktion $x : I \rightarrow \mathbb{R}^n$ heißt Lösung der Carathéodory-Gleichung, falls sie auf jedem Teilintervall $[a, b] \subset I$ absolut stetig ist und die Gleichung

$$\dot{x}(t) = \varphi(t, x(t)) \text{ für fast alle } t \in I$$

erfüllt. Gilt zusätzlich $x(t_0) = x_0$, so ist x eine Carathéodory-Lösung des Anfangswertproblems in (4.7).

Analog zum klassischen Existenzsatz kann nun ein verallgemeinerter Satz für Carathéodory-Lösungen formuliert werden:

Satz 4.3 ([Wal00], S. 128, XVIII. Existenzsatz)

Es sei $I = [t_0, t_0 + a]$, $a > 0$. Erfüllt die Funktion $\varphi : I \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ die Carathéodory-Bedingungen, so besitzt die Aufgabe (4.7) auf dem Intervall I mindestens eine Lösung im Sinne von Carathéodory.

Unter den Voraussetzungen des obigen Satzes möchten wir die folgenden Bemerkungen notieren:

Bemerkung 4.2.

1. *Wie im klassischen Fall kann die Eindeutigkeit der Lösung durch eine zusätzliche Lipschitzbedingung gesichert werden, vgl. [Wal00], S. 128, Aussage (b).*

2. Aus der Gültigkeit der Carathéodory-Bedingungen und durch Anwenden des Hauptsatzes (vgl. Satz 4.2) ergibt sich (wie im klassischen Fall) die Äquivalenz zu einer Integralgleichung. Finden wir also eine absolut stetige Funktion $x : I \rightarrow \mathbb{R}^n$, die der Integralgleichung

$$x(t) = x(t_0) + \int_{t_0}^t \varphi(\tau, x(\tau)) d\tau \quad \text{für fast alle } t \in I \quad (4.9)$$

genügt, so ist diese auch eine Lösung des Anfangswertproblems (4.7) im Sinne von Carathéodory.

3. Es genügt eine in I stetige Lösung von (4.9) zu finden: Aus der Stetigkeit von x folgt, dass $t \mapsto \varphi(t, x(t))$ eine auf I Lebesgue-integrierbare Funktion ist. Der Hauptsatz der Differential- und Integralrechnung liefert dann automatisch, dass x auch absolut stetig ist.
4. Da jede stetig differenzierbare Funktion auch absolut stetig ist, folgt dass der klassische Lösungsbegriff mit dem einer verallgemeinerten Lösung im Sinne von Carathéodory zusammenfällt.

Das Lösungskonzept von Carathéodory stellt eine Erweiterung der klassischen Theorie dar, indem die Klasse der absolut stetigen Funktionen herangezogen wird. Für den Fall eines endlichen Intervalls $I = [t_0, t_0 + a]$, $a > 0$ ist die Existenz einer Lösung des betrachteten Anfangswertproblems durch den vorangegangenen Satz garantiert und da die klassische Ableitung absolut stetiger Funktionen nur fast überall existiert, muss entsprechend die Differentialgleichung nur fast überall erfüllt sein.

Aufbauend darauf, möchten wir nun die Verbindung zu Differentialgleichungen in gewichteten Sobolev-Räumen, wie wir sie in der Aufgabe $(LQ)_\infty$ vorfinden, herstellen. Dazu halten wir zunächst fest, dass mit einem $T \in (0, \infty)$ die zwei Formen von Differentialgleichungen

$$\mathcal{D}x(\cdot) = \varphi(\cdot, x(\cdot)) \quad \text{auf } (0, T) \quad (4.10)$$

$$\dot{x}(t) = \varphi(t, x(t)) \quad \text{f.ü. auf } (0, T) \quad (4.11)$$

betrachtet in dem Raum der absolut stetigen Funktionen $AC^n([0, T])$ äquivalent sind. Dies ergibt sich als direkte Folgerung aus der Tatsache, dass die fast überall existierende klassische Ableitung einer absolut stetigen Funktionen mit ihrer schwachen Ableitung zusammenfällt. Letzteres wiederum folgt aus der Gültigkeit der Formel für partielle Integration für absolut stetige Funktionen, wie wir sofort einsehen: Für eine absolut stetige Funktion $f : [a, b] \rightarrow \mathbb{R}$ mit $-\infty < a < b < \infty$ gilt

$$\int_a^b f'(t)\phi(t)dt = - \int_a^b f(t)\phi'(t)dt \quad \forall \phi \in C_0^\infty((a, b)).$$

Demnach ist die für fast alle $t \in [a, b]$ existierende klassische Ableitung $f'(t)$ Repräsentant der schwachen Ableitung von f in (a, b) . Um nun den Übergang zu Differentialgleichungen in einem Sobolev-Raum zu vollziehen, benötigen wir den folgenden Zusammenhang: Die Klassen

$W_1^{1,n}((0, T))$ und $AC^n([0, T])$ sind identisch in dem Sinne, dass für jedes $f \in W_1^{1,n}((0, T))$ eine Funktion $g \in AC^n([0, T])$ (und umgekehrt) mit $f(t) = g(t)$ fast überall auf $(0, T)$ existiert². Demnach können wir also jede Sobolev-Funktion $x \in W_1^{1,n}((0, T))$ mit ihrem absolut stetigen Repräsentanten identifizieren und diesen sogar stetig auf $[0, T]$ fortsetzen, wodurch die Formulierung eines Anfangswertes $x(0) = x_0$ gerechtfertigt ist. Schließlich folgt, dass die Formulierungen der Differentialgleichung (4.10) und (4.11) auch in dem Raum $W_1^{1,n}((0, T))$ äquivalent zu verstehen sind.

Fordern wir zusätzlich ein, dass die Funktion φ die Carathéodory-Bedingungen erfüllt, so liefert der Satz 4.3 die Existenz einer Lösung des zugehörigen Anfangswertproblems mit $x(0) = x_0$ und wir können wieder zu der zu (4.10) und (4.11) äquivalenten Integralgleichung

$$x(t) = x(0) + \int_0^t \varphi(\tau, x(\tau)) d\tau \quad \text{für fast alle } t \in (0, T)$$

übergehen. Diese Überlegungen lassen sich nun wie folgt für gewichtete Sobolev-Funktionen auf den unendlichen Zeithorizont $(0, \infty)$ übertragen: Bezeichnen wir mit x_T die Einschränkung einer Funktion $x \in W_2^{1,n}((0, \infty), \omega)$ mit $\omega_\beta(t) := e^{-\beta t}$, $\beta > 0$ auf das beschränkte Intervall $(0, T)$. Offenbar gilt dann auch $x_T \in W_2^{1,n}((0, T), \omega_\beta)$ und aufgrund der Gültigkeit der Inklusion

$$W_2^{1,n}((0, T), \omega_\beta) \subseteq W_2^{1,n}((0, T)) \subseteq W_1^{1,n}((0, T))$$

folgt sofort $x_T \in W_1^{1,n}((0, T))$. Aufbauend auf der vorangegangenen Diskussion der Differentialgleichungen auf dem endlichen Intervall $(0, T)$, ergibt sich daraus nicht nur die Äquivalenz der Differentialgleichungen

$$\begin{aligned} \mathcal{D}x(\cdot) &= \varphi(\cdot, x(\cdot)) \quad \text{auf } (0, \infty) \\ \dot{x}(t) &= \varphi(t, x(t)) \quad \text{f.ü. auf } (0, \infty) \end{aligned}$$

für Funktionen $x \in W_2^{1,n}((0, \infty), \omega_\beta)$, sondern auch wieder die Möglichkeit eines Überganges zu der äquivalenten Integralgleichung

$$x(t) = x(0) + \int_0^t \varphi(\tau, x(\tau)) d\tau \quad \text{für fast alle } t \in (0, \infty),$$

vorausgesetzt φ erfüllt die Carathéodory-Bedingungen. Für den unendlichen Zeithorizont $(0, \infty)$ erhalten wir damit durch eine geeignete Lokalisierung des Lösungsbegriffs wieder ein Anfangswertproblem im Sinne von Carathéodory. Zusammenfassend können wir festhalten, dass das in (4.3) formulierte Anfangswertproblem in dem Raum $W_2^{1,n}((0, \infty), \omega_\beta)$ wohldefiniert ist. Die Existenz der eindeutig bestimmten lokal absolut stetigen Repräsentanten erlaubt es, zu einer punktwisen Formulierung der Differentialgleichung überzugehen. Ganz natürlich stellt sich somit der Raum $AC_{loc}^n((0, \infty))$ als geeigneter Lösungsraum heraus. Dies

² Der Zusammenhang zwischen den absolut stetigen Funktionen und den Sobolev-Funktionen ist wesentlicher Bestandteil der Arbeit [Dik16]. Entsprechend verweisen wir den Leser für eine ausführlichere Darstellung auf die angegebene Literatur.

liefert uns zusätzlich die wertvolle Möglichkeit, die zugrunde gelegte Differentialgleichung basierend auf der verallgemeinerten Theorie nach Carathéodory zu studieren, gleichzeitig aber auch die Steuerungsaufgabe in dem Kontext von Hilberträumen zu formulieren.

4.4 Formulierung als Optimierungsproblem in einem Hilbertraum

In diesem Abschnitt untersuchen wir die Steuerungsaufgabe $(LQ)_\infty$ mit Hilfe von Operatoren in Hilberträumen. Dieses Vorgehen wird es uns im Anschluss erlauben, die Existenz und Eindeutigkeit einer Lösung von $(LQ)_\infty$ mit Hilfe des Satzes von Weierstraß zu beweisen. Wir beginnen damit, den quadratischen Term der Zielfunktion durch eine symmetrische Bilinearform und den linearen Teil der Zielfunktion durch ein stetiges lineares Funktional zu beschreiben. Dazu definieren wir zunächst den Raum

$$\mathcal{H} := W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta),$$

welcher durch die Norm

$$\|(x, u)\|_{\mathcal{H}} := \left(\|x\|_{1,(2),\omega_\beta}^2 + \|u\|_{(2),\omega_\beta}^2 \right)^{1/2}$$

zu einem Hilbertraum wird. Auf diesem führen wir die Abbildung $\mathcal{Q} : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ durch

$$\mathcal{Q}(z_1; z_2) = \mathcal{Q}((x_1, u_1); (x_2, u_2)) := \int_0^\infty \{x_1^T(t)W(t)x_2(t) + u_1^T(t)R(t)u_2(t)\}\omega_\beta(t)dt \quad (4.12)$$

ein und definieren $q : \mathcal{H} \rightarrow \mathbb{R}$ durch

$$q(z) = q((x, u)) := \int_0^\infty \{w^T(t)x(t) + r^T(t)u(t)\}\omega_\beta(t)dt. \quad (4.13)$$

Mit diesen Abbildungen können wir das Zielfunktional von (4.1) gemäß

$$J(x, u) = \frac{1}{2}\mathcal{Q}((x, u); (x, u)) + q((x, u))$$

schreiben. Wir halten die Eigenschaften der eingeführten Abbildungen \mathcal{Q} und q in den nachfolgenden zwei Lemmata fest.

Lemma 4.4

Durch die Abbildung $\mathcal{Q} : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$, definiert in (4.12), wird eine symmetrische Bilinearform auf \mathcal{H} definiert. Diese ist beschränkt und damit auch stetig auf \mathcal{H} .

Beweis. Es lässt sich leicht nachrechnen, dass \mathcal{Q} für alle $z_1, \hat{z}_1, z_2, \hat{z}_2 \in \mathcal{H}$ und $\lambda \in \mathbb{R}$ den

nachfolgenden Bedingungen genügt:

$$\begin{aligned}\mathcal{Q}(z_1 + \widehat{z}_1; z_2) &= \mathcal{Q}(z_1; z_2) + \mathcal{Q}(\widehat{z}_1; z_2), \\ \mathcal{Q}(z_1; z_2 + \widehat{z}_2) &= \mathcal{Q}(z_1; z_2) + \mathcal{Q}(z_1; \widehat{z}_2), \\ \mathcal{Q}(\lambda z_1; z_2) &= \mathcal{Q}(z_1; \lambda z_2) = \lambda \mathcal{Q}(z_1; z_2).\end{aligned}$$

Demnach wird durch \mathcal{Q} eine Bilinearform auf \mathcal{H} festgelegt. Aufgrund der Symmetrie der Matrizen $W(t)$ und $R(t)$ gilt offensichtlich für alle $t \in \mathbb{R}_+$ und alle $\xi_1, \xi_2 \in \mathbb{R}^n$

$$\xi_1^T W(t) \xi_2 = \xi_1^T W^T(t) \xi_2 = (W(t) \xi_1)^T \xi_2 = \xi_2^T W(t) \xi_1$$

und analog auch $v_1^T R(t) v_2 = v_2^T R(t) v_1$ für alle $v_1, v_2 \in \mathbb{R}^m$. Daraus ergibt sich direkt die Symmetrie von \mathcal{Q} , es ist also $\mathcal{Q}(z_1; z_2) = \mathcal{Q}(z_2; z_1)$ für alle $z_1, z_2 \in \mathcal{H}$ richtig.

Um die Beschränktheit von \mathcal{Q} auf \mathcal{H} zu zeigen, müssen wir die Existenz einer Konstanten $c > 0$ mit

$$|\mathcal{Q}(z_1; z_2)| \leq c \cdot \|z_1\|_{\mathcal{H}} \cdot \|z_2\|_{\mathcal{H}} \quad \forall z_1, z_2 \in \mathcal{H}$$

nachweisen. Da $W \in L_{\infty}^{n \times n}((0, \infty))$ gilt, erhalten wir für jedes $x_1 \in L_2^n((0, \infty), \omega_{\beta})$ aufgrund der Verträglichkeit der Spektral- mit der euklidischen Vektornorm, vgl. (2.22), die Abschätzung

$$\int_0^{\infty} |W(t)x_1(t)|_2^2 \omega_{\beta}(t) dt \leq \|W\|_{(\infty)}^2 \cdot \int_0^{\infty} |x_1(t)|_2^2 \omega_{\beta}(t) dt = \|W\|_{(\infty)}^2 \cdot \|x_1\|_{(2), \omega_{\beta}}^2 < \infty,$$

also auch $Wx_1 \in L_2^n((0, \infty), \omega_{\beta})$. Für Funktionen $x_1, x_2 \in L_2^n((0, \infty), \omega_{\beta})$ können wir also die Ungleichung von Cauchy-Schwarz anwenden und erhalten

$$\left| \langle Wx_1, x_2 \rangle_{(2), \omega_{\beta}} \right| \leq \langle Wx_1, Wx_1 \rangle_{(2), \omega_{\beta}}^{1/2} \cdot \langle x_2, x_2 \rangle_{(2), \omega_{\beta}}^{1/2}$$

beziehungsweise in äquivalenter Form

$$\begin{aligned}& \left| \int_0^{\infty} (W(t)x_1(t))^T x_2(t) \omega_{\beta}(t) dt \right| \\ & \leq \left(\int_0^{\infty} (W(t)x_1(t))^T (W(t)x_1(t)) \omega_{\beta}(t) dt \right)^{1/2} \cdot \left(\int_0^{\infty} x_2^T(t)x_2(t) \omega_{\beta}(t) dt \right)^{1/2} \\ & = \left(\int_0^{\infty} |W(t)x_1(t)|_2^2 \omega_{\beta}(t) dt \right)^{1/2} \cdot \left(\int_0^{\infty} |x_2(t)|_2^2 \omega_{\beta}(t) dt \right)^{1/2}.\end{aligned}$$

Beachten wir wieder die Verträglichkeit der auftauchenden Normen, so können wir weiter

auf

$$\begin{aligned} \left| \int_0^\infty (W(t)x_1(t))^T x_2(t)\omega_\beta(t) dt \right| &\leq \|W\|_{(\infty)} \cdot \left(\int_0^\infty |x_1(t)|_2^2 \omega_\beta(t) dt \right)^{1/2} \cdot \left(\int_0^\infty |x_2(t)|_2^2 \omega_\beta(t) dt \right)^{1/2} \\ &= \|W\|_{(\infty)} \cdot \|x_1\|_{(2),\omega_\beta} \cdot \|x_2\|_{(2),\omega_\beta} \end{aligned}$$

schließen und erhalten aufgrund der Symmetrie von $W(t)$ ($t \in \mathbb{R}_+$) die Ungleichung

$$\left| \int_0^\infty x_1^T(t)W(t)x_2(t)\omega_\beta(t) dt \right| \leq \|W\|_{(\infty)} \cdot \|x_1\|_{(2),\omega_\beta} \cdot \|x_2\|_{(2),\omega_\beta}. \quad (4.14)$$

In gleicher Weise ergibt sich

$$\left| \int_0^\infty u_1^T(t)R(t)u_2(t)\omega_\beta(t) dt \right| \leq \|R\|_{(\infty)} \cdot \|u_1\|_{(2),\omega_\beta} \|u_2\|_{(2),\omega_\beta} \quad (4.15)$$

für $R \in L_\infty^{m \times m}((0, \infty))$ und $u_1, u_2 \in L_2^m((0, \infty), \omega_\beta)$. Zusammengefasst können wir aus (4.14) und (4.15) für alle $(z_1; z_2) := ((x_1, u_1); (x_2, u_2)) \in \mathcal{H}$ folgern, dass

$$\begin{aligned} |\mathcal{Q}(z_1; z_2)| &= |\mathcal{Q}((x_1, u_1); (x_2, u_2))| \\ &= \left| \int_0^\infty \{x_1^T(t)W(t)x_2(t) + u_1^T(t)R(t)u_2(t)\} \omega_\beta(t) dt \right| \\ &\leq \left| \int_0^\infty x_1^T(t)W(t)x_2(t)\omega_\beta(t) dt \right| + \left| \int_0^\infty u_1^T(t)R(t)u_2(t)\omega_\beta(t) dt \right| \\ &\leq \|W\|_{(\infty)} \cdot \|x_1\|_{(2),\omega_\beta} \|x_2\|_{(2),\omega_\beta} + \|R\|_{(\infty)} \cdot \|u_1\|_{(2),\omega_\beta} \|u_2\|_{(2),\omega_\beta} \\ &\leq \|W\|_{(\infty)} \cdot \|x_1\|_{1,(2),\omega_\beta} \|x_2\|_{1,(2),\omega_\beta} + \|R\|_{(\infty)} \cdot \|u_1\|_{(2),\omega_\beta} \|u_2\|_{(2),\omega_\beta} \end{aligned}$$

gilt. Wir wählen $c := \sqrt{2} \max\{\|W\|_{(\infty)}, \|R\|_{(\infty)}\}$ und unter Verwendung der Ungleichung

$$a + b \leq \sqrt{2} (a^2 + b^2)^{1/2} \quad \forall a, b \geq 0 \quad (a, b \in \mathbb{R}), \quad (4.16)$$

können wir weiter auf

$$\begin{aligned}
 |\mathcal{Q}(z_1; z_2)| &\leq c \cdot \left(\|x_1\|_{1,(2),\omega_\beta}^2 \|x_2\|_{1,(2),\omega_\beta}^2 + \|u_1\|_{(2),\omega_\beta}^2 \|u_2\|_{(2),\omega_\beta}^2 \right)^{1/2} \\
 &\leq c \cdot \left(\|x_1\|_{1,(2),\omega_\beta}^2 \|x_2\|_{1,(2),\omega_\beta}^2 + \|x_1\|_{1,(2),\omega_\beta}^2 \|u_2\|_{(2),\omega_\beta}^2 \right. \\
 &\quad \left. + \|u_1\|_{(2),\omega_\beta}^2 \|x_2\|_{1,(2),\omega_\beta}^2 + \|u_1\|_{(2),\omega_\beta}^2 \|u_2\|_{(2),\omega_\beta}^2 \right)^{1/2} \\
 &= c \cdot \left(\|x_1\|_{1,(2),\omega_\beta}^2 + \|u_1\|_{(2),\omega_\beta}^2 \right)^{1/2} \cdot \left(\|x_2\|_{1,(2),\omega_\beta}^2 + \|u_2\|_{(2),\omega_\beta}^2 \right)^{1/2} \\
 &= c \cdot \|(x_1, u_1)\|_{\mathcal{H}} \cdot \|(x_2, u_2)\|_{\mathcal{H}} = c \cdot \|z_1\|_{\mathcal{H}} \cdot \|z_2\|_{\mathcal{H}}
 \end{aligned}$$

schließen, womit die Beschränktheit von \mathcal{Q} auf \mathcal{H} bewiesen ist. Da eine Bilinearform genau dann stetig ist, wenn sie beschränkt ist³, sind alle Behauptungen des Lemmas bewiesen. \square

Lemma 4.5

Durch die Abbildung $q : \mathcal{H} \rightarrow \mathbb{R}$, definiert in (4.13), wird ein lineares und beschränktes Funktional auf \mathcal{H} definiert. Insbesondere ist q demnach auch stetig.

Beweis. Einfaches Nachrechnen liefert sofort, dass q linear ist. Wir verwenden die Dreiecksungleichung und die Ungleichung von Cauchy-Schwarz und sehen ein, dass für alle $z := (x, u) \in \mathcal{H}$ die Abschätzung

$$\begin{aligned}
 |q(z)| &= |q((x, u))| = \left| \int_0^\infty \{w^T(t)x(t) + r^T(t)u(t)\} \omega_\beta(t) dt \right| \\
 &\leq \left| \int_0^\infty w^T(t)x(t) \omega_\beta(t) dt \right| + \left| \int_0^\infty r^T(t)u(t) \omega_\beta(t) dt \right| \\
 &\leq \left(\int_0^\infty |w(t)|_2^2 \omega_\beta(t) dt \right)^{1/2} \left(\int_0^\infty |x(t)|_2^2 \omega_\beta(t) dt \right)^{1/2} \\
 &\quad + \left(\int_0^\infty |r(t)|_2^2 \omega_\beta(t) dt \right)^{1/2} \left(\int_0^\infty |u(t)|_2^2 \omega_\beta(t) dt \right)^{1/2} \\
 &= \|w\|_{(2),\omega_\beta} \|x\|_{(2),\omega_\beta} + \|r\|_{(2),\omega_\beta} \|u\|_{(2),\omega_\beta} \\
 &\leq c \cdot \left(\|x\|_{(2),\omega_\beta}^2 + \|u\|_{(2),\omega_\beta}^2 \right)^{1/2} = c \cdot \|z\|_{\mathcal{H}}
 \end{aligned}$$

³ vgl. z.B. [AE06], S. 180, 4.1 Satz

richtig ist. Dabei haben wir $c := \sqrt{2} \max\{\|w\|_{(2),\omega_\beta}, \|r\|_{(2),\omega_\beta}\}$ gesetzt und die Ungleichung in (4.16) genutzt. Somit erhalten wir die Beschränktheit von q auf \mathcal{H} und insbesondere ist q demnach ein lineares und stetiges Funktional auf \mathcal{H} . \square

Wir wollen nun auch die lineare Systemdynamik (4.3) mit Hilfe einer Operatorgleichung beschreiben. Dazu führen wir den Hilbertraum

$$\mathcal{Y} := L_2^n((0, \infty), \omega_\beta) \times \mathbb{R}^n$$

ein, auf welchem durch

$$\|(w_1, w_2)\|_{\mathcal{Y}} := \left(\|w_1\|_{(2),\omega_\beta}^2 + |w_2|_2^2 \right)^{1/2}$$

eine Norm definiert wird. Der Operator $\mathcal{A} : \mathcal{H} \rightarrow \mathcal{Y}$ und $\mathbf{b} \in \mathcal{Y}$ werden gemäß der Vorschriften

$$\mathcal{A}(x, u) := \begin{pmatrix} A(\cdot)x(\cdot) + B(\cdot)u(\cdot) - \dot{x}(\cdot) \\ x(0) \end{pmatrix}, \quad \mathbf{b} := \begin{pmatrix} -C(\cdot) \\ x_0 \end{pmatrix}. \quad (4.17)$$

festgelegt.

Lemma 4.6

Der Operator $\mathcal{A} : \mathcal{H} \rightarrow \mathcal{Y}$, definiert in (4.17), ist ein linearer und beschränkter Operator auf \mathcal{H} . Insbesondere ist \mathcal{A} demnach auch stetig.

Beweis. Die Linearität von \mathcal{A} lässt sich leicht nachrechnen. Für die Beschränktheit haben wir die Existenz einer Konstanten $c \in \mathbb{R}$ nachzuweisen, sodass

$$\|\mathcal{A}(x, u)\|_{\mathcal{Y}} \leq c \cdot \|(x, u)\|_{\mathcal{H}}$$

für alle $(x, u) \in \mathcal{H}$ erfüllt ist. Wir nutzen wieder die Ungleichung in (4.16), nun jedoch in der Form

$$(a + b)^2 \leq 2(a^2 + b^2) \quad \forall a, b \geq 0 \quad (a, b \in \mathbb{R})$$

und erhalten gemeinsam mit der Dreiecksungleichung für alle $(x, u) \in \mathcal{H}$ die Abschätzung

$$\begin{aligned}
 \|\mathcal{A}(x, u)\|_{\mathcal{Y}}^2 &= |x(0)|_2^2 + \|A(t)x(t) + B(t)u(t) - \dot{x}(t)\|_{(2),\omega_\beta}^2 \\
 &= |x(0)|_2^2 + \int_0^\infty |A(t)x(t) + B(t)u(t) - \dot{x}(t)|_2^2 \omega_\beta(t) dt \\
 &\leq |x(0)|_2^2 + 2 \int_0^\infty \left\{ |A(t)x(t) + B(t)u(t)|_2^2 + |\dot{x}(t)|_2^2 \right\} \omega_\beta(t) dt \\
 &\leq |x(0)|_2^2 + 4 \left(\int_0^\infty |A(t)x(t)|_2^2 \omega_\beta(t) dt + \int_0^\infty |B(t)u(t)|_2^2 \omega_\beta(t) dt \right) \\
 &\quad + 2 \int_0^\infty |\dot{x}(t)|_2^2 \omega_\beta(t) dt.
 \end{aligned}$$

Die Verträglichkeit der Normen gemäß (2.22) liefert

$$\begin{aligned}
 \|\mathcal{A}(x, u)\|_{\mathcal{Y}}^2 &\leq |x(0)|_2^2 + 4 \cdot \left(\|A\|_{(\infty)}^2 \|x\|_{(2),\omega_\beta}^2 + \|B\|_{(\infty)}^2 \|u\|_{(2),\omega_\beta}^2 + \|\dot{x}\|_{(2),\omega_\beta}^2 \right) \\
 &\leq |x(0)|_2^2 + \tilde{c} \cdot \left(\|x\|_{(2),\omega_\beta}^2 + \|u\|_{(2),\omega_\beta}^2 + \|\dot{x}\|_{(2),\omega_\beta}^2 \right) \\
 &\leq c \cdot \left(\|x\|_{1,(2),\omega_\beta}^2 + \|u\|_{(2),\omega_\beta}^2 \right) = c \cdot \|(x, u)\|_{\mathcal{H}}^2
 \end{aligned}$$

mit $\tilde{c} := \max \{4 \|A\|_{(\infty)}^2, 4 \|B\|_{(\infty)}^2, 4\}$. Da $x \in W_2^{1,n}((0, \infty), \omega_\beta)$ gilt, können wir diese Funktion mit einem lokal absolut stetigen Repräsentanten identifizieren, welcher insbesondere in $t = 0$ stetig fortgesetzt werden kann. Daher gilt $|x(0)|_2^2 < \infty$ und die obige Abschätzung kann mit einem hinreichend groß gewähltem $c \in \mathbb{R}$ vollzogen werden, womit wir die Beschränktheit und damit auch die Stetigkeit von \mathcal{A} bewiesen haben. \square

Mit Hilfe der eingeführten Operatoren und der Menge

$$\mathcal{U} := \{u \in L_2^m((0, \infty), \omega_\beta) : u(t) \in U \text{ f.ü. auf } (0, \infty)\}$$

können wir die zulässige Menge der linear-quadratischen Steuerungsaufgabe $(LQ)_\infty$ durch

$$\mathcal{Z} := \{(x, u) \in \mathcal{H} : u \in \mathcal{U}, \mathcal{A}(x, u) = \mathbf{b}\}$$

beschreiben. Insgesamt lässt sich damit das Problem $(LQ)_\infty$ äquivalent als quadratisches Optimierungsproblem in dem Hilbertraum \mathcal{H} gemäß der Form

$$(LQ)_\infty \quad J(x, u) = \frac{1}{2} \mathcal{Q}((x, u); (x, u)) + q(x, u) \rightarrow \inf! \quad (4.18)$$

bzgl. $(x, u) \in \mathcal{Z}$.

notieren.

4.5 Existenz und Eindeutigkeit einer Lösung

Um die Existenz und auch die Eindeutigkeit einer Lösung von $(LQ)_\infty$ sicherstellen zu können, benutzen wir eine verallgemeinerte Version des Satzes von Weierstraß. In Anlehnung an [Göp73], S. 93, Satz 26 formulieren wir den Weierstraß'schen Satz in einem Banachraum B wie folgt:

*Ein schwach unterhalbstetiges Funktional F nimmt auf einer nichtleeren schwach kompakten Teilmenge M von B seinen Minimalwert an.*⁴

Um dieses Resultat auf die Aufgabe $(LQ)_\infty$ anwenden zu können, müssten wir sowohl die schwache Unterhalbstetigkeit des Zielfunktional J , als auch die schwache Kompaktheit der zulässigen Menge \mathcal{Z} nachweisen. Da uns an dieser Stelle jedoch die Kompaktheitseigenschaft von \mathcal{Z} fehlt, werden wir einen kleinen Umweg gehen, indem wir auf die entsprechenden Niveau-Mengen $\mathcal{N}(J, \alpha)$ von J zurückgreifen. Dieses Vorgehen erlaubt es uns, das obige Resultat auf die Durchschnittsmengen $\mathcal{N}(J, \alpha) \cap \mathcal{Z}$ anzuwenden und daraus die gewünschte Existenz einer Lösung von $(LQ)_\infty$ zu folgern.

Für den Beweis des Existenzsatzes gehen wir in mehreren Schritten vor.

1. Zunächst zeigen wir, dass das Zielfunktional J schwach unterhalbstetig ist.
2. Im zweiten Schritt betrachten wir die erwähnten Durchschnittsmengen $\mathcal{N}(J, \alpha) \cap \mathcal{Z}$ und werden nachweisen, dass diese schwach kompakt sind.
3. Schlussendlich wenden wir den Satz von Weierstraß auf die Mengen $\mathcal{N}(J, \alpha) \cap \mathcal{Z}$ an und können damit schließen, dass das Problem $(LQ)_\infty$ (mindestens) eine Lösung besitzt.

Um zusätzlich auch die Eindeutigkeit einer Lösung von $(LQ)_\infty$ garantieren zu können, benötigen wir die strikte Konvexität des Zielfunktional. Hinreichend dafür ist, dass die Bilinearform Q auf dem Kern der Abbildung \mathcal{A} koerzitiv ist. Wir werden den entsprechenden Nachweis gemeinsam mit der Unterhalbstetigkeit von J in dem ersten Beweisschritt erbringen. Damit erhalten wir am Ende unserer Ausführungen nicht nur einen Existenz-, sondern auch einen Eindeutigkeitssatz, wodurch die Existenz einer eindeutig bestimmten Lösung für das Steuerungsproblem $(LQ)_\infty$ gesichert ist, vgl. [PB18].

Bevor wir uns dem Studium des Zielfunktional widmen, halten wir noch das folgende Resultat bezüglich der zulässigen Menge \mathcal{Z} fest:

Lemma 4.7

Die zulässige Menge \mathcal{Z} ist nichtleer, konvex und bezüglich der Normtopologie von \mathcal{H} abgeschlossen.

⁴ Sei X ein normierter Raum. Eine Folge $\{x_n\}_{n \in \mathbb{N}}$ in X heißt schwach konvergent gegen $\bar{x} \in X$, falls für jedes lineare und stetige Funktional x^* auf X gilt, dass $\lim_{n \rightarrow \infty} x^*(x_n) = x^*(\bar{x})$. Wir schreiben dafür $x_n \rightharpoonup \bar{x}$, vgl. [Wer18], S. 120, Definition III.3.6. Ein auf X definiertes Funktional F heißt schwach unterhalbstetig an der Stelle $\bar{x} \in X$, falls für jede schwach konvergente Folge $\{x_n\}_{n \in \mathbb{N}}$ aus X gilt, dass $\liminf_{n \rightarrow \infty} F(x_n) \geq F(\bar{x})$ erfüllt ist, vgl. [Göp73], S. 90. Wir nennen eine nichtleere Menge $M \subseteq X$ schwach kompakt, wenn jede Folge in M eine schwach konvergente Teilfolge mit Grenzwert in M besitzt, vgl. [Wer18], S. 467, Theorem VIII.6.1.

Beweis. Die Voraussetzung (V5) garantiert $\mathcal{Z} \neq \emptyset$. Aus der Konvexität des Steuerbereichs U , vgl. Voraussetzung (V4), und der Linearität der Dynamik können wir sofort auf die Konvexität der Menge \mathcal{Z} schließen. Sei nun $\{(x_N, u_N)\}_{N \in \mathbb{N}}$ eine konvergente Folge aus \mathcal{Z} mit $\{(x_N, u_N)\} \xrightarrow{\mathcal{H}} (x_0, u_0)$ und einem Grenzelement $(x_0, u_0) \in W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta)$. Um die Abgeschlossenheit bezüglich der Menge \mathcal{U} zu beweisen, betrachten wir zunächst die Komponentenfolge $\{u_N\}_{N \in \mathbb{N}} \subset \mathcal{U}$, welche im Sinne der $L_2^m((0, \infty), \omega_\beta)$ -Norm gegen die Funktion $u_0 \in L_2^m((0, \infty), \omega_\beta)$ konvergiert. Wir folgen den Argumenten in [Lyk10], S. 85 und stellen fest: Da das durch die Gewichtsfunktion $\omega_\beta(t) = e^{-\beta t}$, $\beta > 0$ induzierte Maß

$$\mu_{\omega_\beta}(0, \infty) := \int_0^\infty \omega_\beta(t) dt < \infty$$

endlich ist, existiert eine Teilfolge $\{u_{N_k}\}_{k \in \mathbb{N}} \subseteq \{u_N\}_{N \in \mathbb{N}}$, die fast überall auf $(0, \infty)$ punktweise gegen die Funktion u_0 konvergiert, vgl. [KF61], S. 86, Theorem 3. Demnach gibt es eine Lebesguesche Nullmenge \mathcal{N} , sodass für alle $t \in (0, \infty) \setminus \mathcal{N}$ die punktweise Konvergenz

$$u_{N_k}(t) \rightarrow u_0(t) \tag{4.19}$$

gilt. Aus $\{u_N\}_{N \in \mathbb{N}} \subset \mathcal{U}$ folgt außerdem, dass für fast alle $t \in (0, \infty)$ und jede Zahl $N \in \mathbb{N}$ die Inklusion $u_N(t) \in U$ erfüllt ist. Wir definieren die Menge

$$\mathcal{B} := \bigcup_{k=1}^\infty \{t \in (0, \infty) : u_{N_k}(t) \notin \mathcal{U}\}.$$

Aufgrund der Sigma-Additivität des Lebesgue Maßes ist diese ebenfalls vom Lebesgue-Maß Null und für alle $t \in (0, \infty) \setminus \mathcal{B}$ gilt $u_{N_k}(t) \in U$. Weiterhin ist auch $\mathcal{N} \cup \mathcal{B}$ eine Lebesguesche Nullmenge und wir erhalten, dass die punktweise Konvergenz in (4.19) sowie die Inklusion $u_{N_k}(t) \in U$ für alle $t \in (0, \infty) \setminus (\mathcal{N} \cup \mathcal{B})$ bestehen bleibt. Aus der Abgeschlossenheit der Menge U , vgl. Voraussetzung (V4), folgt, dass $u_0(t) \in U$ für alle $t \in (0, \infty) \setminus (\mathcal{N} \cup \mathcal{B})$ gilt und das heißt, dass $u_0(t) \in U$ für fast alle $t \in (0, \infty)$ richtig ist. Somit folgt $u_0 \in \mathcal{U}$, woraus sich die Abgeschlossenheit von \mathcal{Z} bezüglich \mathcal{U} ergibt. Letztlich liefert die Stetigkeit des Operators \mathcal{A} , vgl. Lemma 4.6,

$$\lim_{N \rightarrow \infty} \mathcal{A}(x_N, u_N) = \mathcal{A}\left(\lim_{N \rightarrow \infty} x_N, \lim_{N \rightarrow \infty} u_N\right) = \mathcal{A}(x_0, u_0)$$

und damit $\mathcal{A}(x_0, u_0) = \mathbf{b}$. Zusammengefasst erhalten wir, dass $(x_0, u_0) \in \mathcal{Z}$ gilt, womit die Abgeschlossenheit der Menge \mathcal{Z} bewiesen ist. \square

Das Zielfunktional J

Betrachten wir noch einmal die linear-quadratische Struktur des Integranden

$$f(t, \xi, v) := \frac{1}{2} \{\xi^T W(t) \xi + v^T R(t) v\} + w^T(t) \xi + r^T(t) v,$$

so sehen wir sofort ein, dass uns die positive Semidefinitheit der (symmetrischen) Matrizen

$W(t)$ und $R(t)$ (mit $t \in \mathbb{R}_+$) die Konvexität von f in ξ und v garantiert. Diese Konvexität des Integranden im Zielfunktional erlaubt es uns, direkt auf dessen schwache Unterhalbstetigkeit zu schließen.

Lemma 4.8

Das Zielfunktional J aus (4.1) bzw. (4.18) ist schwach unterhalbstetig auf \mathcal{H} .

Beweis. Für den Nachweis der schwachen Unterhalbstetigkeit des Funktionals J auf $\mathcal{H} = W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta)$, wählen wir die schwach konvergenten Folgen $\{x_N\}_{N \in \mathbb{N}}$ und $\{u_N\}_{k \in \mathbb{N}}$ mit $x_N \in W_2^{1,n}((0, \infty), \omega_\beta)$ und $u_N \in L_2^m((0, \infty), \omega_\beta)$ und $x_N \rightharpoonup x_0$ in $W_2^{1,n}((0, \infty), \omega_\beta)$ und $u_N \rightharpoonup u_0$ in $L_2^m((0, \infty), \omega_\beta)$. Da der Integrand (stetig) differenzierbar in ξ und v ist, folgt aus der Konvexität von f in ξ und v und der Symmetrie der Matrizen $W(t)$ und $R(t)$ (mit $t \in \mathbb{R}_+$) die folgende Ungleichung:

$$\begin{aligned} f(t, x_N(t), u_N(t)) &\geq \frac{1}{2} \{x_0^T(t)W(t)x_0(t) + u_0^T(t)R(t)u_0(t)\} + w^T(t)x_0(t) + r^T(t)u_0(t) \\ &\quad + (W(t)x_0(t) + w(t))^T(x_N(t) - x_0(t)) + (R(t)u_0(t) + r(t))^T(u_N(t) - u_0(t)) \\ &= \frac{1}{2} \{x_0^T(t)W(t)x_0(t) + u_0^T(t)R(t)u_0(t)\} + w(t)^T x_N(t) + r(t)^T u_N(t) \\ &\quad + x_0^T(t)W(t)(x_N(t) - x_0(t)) + u_0^T(t)R(t)(u_N(t) - u_0(t)). \end{aligned}$$

Integration der erhaltenen Ungleichung liefert die folgende Abschätzung für das Zielfunktional:

$$\begin{aligned} \infty &> \int_0^\infty f(t, x_N(t), u_N(t)) \omega_\beta(t) dt \\ &\geq \int_0^\infty \left\{ \frac{1}{2} x_0^T(t)W(t)x_0(t) + \frac{1}{2} u_0^T(t)R(t)u_0(t) + w(t)^T x_N(t) + r(t)^T u_N(t) \right\} \omega_\beta(t) dt \\ &\quad + \int_0^\infty \left\{ x_0^T(t)W(t)(x_N(t) - x_0(t)) + u_0^T(t)R(t)(u_N(t) - u_0(t)) \right\} \omega_\beta(t) dt. \end{aligned}$$

Die Existenz und die Endlichkeit der uneigentlichen Integrale auf der rechten Seite in der vorangegangenen Ungleichung ist aufgrund der Grundvoraussetzungen, insbesondere der Wahl der Funktionenräume in (V1) gesichert (vgl. Lemma 4.4 und Lemma 4.5). Der Übergang zum Limes inferior liefert

$$\liminf_{N \rightarrow \infty} J(x_N, u_N) \geq l_1 + l_2$$

mit

$$l_1 := \liminf_{N \rightarrow \infty} \int_0^{\infty} \left\{ \frac{1}{2} x_0^T(t) W(t) x_0(t) + \frac{1}{2} u_0^T(t) R(t) u_0(t) + w(t)^T x_N(t) + r(t)^T u_N(t) \right\} \omega_{\beta}(t) dt$$

$$l_2 := \liminf_{N \rightarrow \infty} \int_0^{\infty} \left\{ x_0^T(t) W(t) (x_N(t) - x_0(t)) + u_0^T(t) R(t) (u_N(t) - u_0(t)) \right\} \omega_{\beta}(t) dt.$$

Die Ausdrücke l_1 und l_2 können nun separat abgeschätzt werden. Auf deren Herleitung möchten wir an dieser Stelle verzichten und verweisen für eine detaillierte Beweisführung auf [Lyk10], S. 73 ff. Die Idee dabei ist es, zunächst diejenigen Teilfolgen $\{x_{N^k}\}_{N^k \in \mathbb{N}} \subseteq \{x_N\}_{N \in \mathbb{N}}$ und $\{u_{N^k}\}_{N^k \in \mathbb{N}} \subseteq \{u_N\}_{N \in \mathbb{N}}$ zu betrachten, die den entsprechenden Limes inferior in l_1 bzw. l_2 realisieren. Für den Ausdruck l_1 wird anschließend ein geschicktes Diagonalverfahren zur Konstruktion einer weiteren Teilfolge genutzt, um die Gültigkeit der Ungleichung

$$l_1 \geq \int_0^{\infty} \left\{ \frac{1}{2} x_0^T(t) W(t) x_0(t) + \frac{1}{2} u_0^T(t) R(t) u_0(t) + w(t)^T x_0(t) + r(t)^T u_0(t) \right\} \omega_{\beta}(t) dt = J(x_0, u_0)$$

nachzuweisen. Für den zweiten Ausdruck ergibt sich $l_2 = 0$. Zusammengefasst erhält man damit, dass die Ungleichung

$$\liminf_{N \rightarrow \infty} J(x_N, u_N) \geq l_1 + \underbrace{l_2}_{=0} \geq J(x_0, u_0)$$

für jede schwach konvergente Folge $\{(x_N, u_N)\}_{N \in \mathbb{N}} \in \mathcal{H}$ mit $\{(x_N, u_N)\} \xrightarrow{\mathcal{H}} (x_0, u_0) \in \mathcal{H}$ richtig ist - das Funktional J demnach also schwach unterhalbstetig auf \mathcal{H} ist. \square

Die Eindeutigkeit einer Lösung von $(LQ)_{\infty}$ kann durch die strikte Konvexität der Zielfunktion sichergestellt werden. Bevor wir diesen Nachweis erbringen, formulieren wir noch ein dafür hinreichendes Hilfsresultat.

Lemma 4.9

Die Bilinearform \mathcal{Q} ist auf dem Kern der Abbildung \mathcal{A} koerzitiv (stark positiv).

Beweis. Die Bilinearform \mathcal{Q} heißt koerzitiv, wenn $\mathcal{Q}(z; z) \geq \alpha \cdot \|z\|_{\mathcal{H}}^2$ für eine Konstante $\alpha > 0$ und alle $z \in \ker \mathcal{A}$ richtig ist. Der Kern der Abbildung \mathcal{A} ist durch

$$\ker \mathcal{A} := \{(x, u) \in \mathcal{H} : \mathcal{A}(x, u) = 0\}$$

definiert. Wir betrachten ein beliebiges Element $z := (\zeta, \vartheta) \in \ker \mathcal{A}$ und unterscheiden zwei Fälle.

Fall 1: Sei $A \equiv 0$. Dann ist $\dot{\zeta}(t) = B(t)\vartheta(t)$, $\zeta(0) = 0$ und zusammen mit der Normverträglich-

lichkeit, vgl. (2.22), erhalten wir

$$\|\dot{\zeta}\|_{(2),\omega_\beta}^2 = \int_0^\infty |B(t)\vartheta(t)|_2^2 \omega_\beta(t) dt \leq \|B\|_{(\infty)}^2 \int_0^\infty |\vartheta(t)|_2^2 \omega_\beta(t) dt = \|B\|_{(\infty)}^2 \|\vartheta\|_{(2),\omega_\beta}^2$$

und damit

$$\|\vartheta\|_{(2),\omega_\beta}^2 \geq \frac{1}{\|B\|_{(\infty)}^2} \|\dot{\zeta}\|_{(2),\omega_\beta}^2.$$

Zusammen mit der Voraussetzung (V3a), also $\underline{W} > 0$ und $\underline{R} > 0$, und der daraus resultierenden Ungleichung (4.6) (analog für $R(t)$), können wir auf die folgende Abschätzung schließen:

$$\begin{aligned} \mathcal{Q}((\zeta, \vartheta); (\zeta, \vartheta)) &= \int_0^\infty \{\zeta^T(t)W(t)\zeta(t) + \vartheta^T(t)R(t)\vartheta(t)\} \omega_\beta(t) dt \\ &\geq \underline{W} \int_0^\infty |\zeta(t)|_2^2 \omega_\beta(t) dt + \underline{R} \int_0^\infty |\vartheta(t)|_2^2 \omega_\beta(t) dt \\ &\geq \frac{1}{2} \underline{W} \|\zeta\|_{(2),\omega_\beta}^2 + \frac{1}{2} \underline{R} \|\vartheta\|_{(2),\omega_\beta}^2 + \frac{1}{2} \underline{R} \|\vartheta\|_{(2),\omega_\beta}^2 \\ &\geq \frac{1}{2} \underline{W} \|\zeta\|_{(2),\omega_\beta}^2 + \frac{1}{2 \|B\|_{(\infty)}^2} \underline{R} \|\dot{\zeta}\|_{(2),\omega_\beta}^2 + \frac{1}{2} \underline{R} \|\vartheta\|_{(2),\omega_\beta}^2 \\ &\geq \frac{1}{2} \min \left\{ \underline{W}, \frac{1}{\|B\|_{(\infty)}^2} \underline{R}, \underline{R} \right\} \left(\|\zeta\|_{(2),\omega_\beta}^2 + \|\dot{\zeta}\|_{(2),\omega_\beta}^2 + \|\vartheta\|_{(2),\omega_\beta}^2 \right) \end{aligned}$$

Mit $\|(\zeta, \vartheta)\|_{\mathcal{H}} := \|\zeta\|_{1,(2),\omega_\beta}^2 + \|\vartheta\|_{(2),\omega_\beta}^2 = \|\zeta\|_{(2),\omega_\beta}^2 + \|\dot{\zeta}\|_{(2),\omega_\beta}^2 + \|\vartheta\|_{(2),\omega_\beta}^2$ erhalten wir die gewünschte Eigenschaft.

Fall 2: Sei $A \neq 0$. Dann gilt $\dot{\zeta}(t) = A(t)\zeta(t) + B(t)\vartheta(t)$, $\zeta(0) = 0$. Anwenden der Dreiecksungleichung und Ausnutzen der Normverträglichkeit, vgl. (2.22), liefert

$$\begin{aligned} \|\dot{\zeta}\|_{(2),\omega_\beta}^2 &= \int_0^\infty |A(t)\zeta(t) + B(t)\vartheta(t)|_2^2 \omega_\beta(t) dt \leq \int_0^\infty (|A(t)\zeta(t)|_2 + |B(t)\vartheta(t)|_2)^2 \omega_\beta(t) dt \\ &\leq 2 \int_0^\infty |A(t)\zeta(t)|_2^2 \omega_\beta(t) dt + 2 \int_0^\infty |B(t)\vartheta(t)|_2^2 \omega_\beta(t) dt \\ &\leq 2 \max \{ \|A\|_{(\infty)}^2, \|B\|_{(\infty)}^2 \} \left(\|\zeta\|_{(2),\omega_\beta}^2 + \|\vartheta\|_{(2),\omega_\beta}^2 \right). \end{aligned}$$

Setzen wir $c := 2 \max \{ \|A\|_{(\infty)}^2, \|B\|_{(\infty)}^2 \}$, so ergibt sich aus der obigen Abschätzung

$$\|\zeta\|_{(2),\omega_\beta}^2 \geq \frac{1}{c} \|\dot{\zeta}\|_{(2),\omega_\beta}^2 - \|\vartheta\|_{(2),\omega_\beta}^2.$$

und damit folgt

$$\begin{aligned}
 \mathcal{Q}((\zeta, \vartheta); (\zeta, \vartheta)) &\geq \frac{1}{2}\underline{W} \|\zeta\|_{(2),\omega_\beta}^2 + \frac{1}{2}\underline{W} \|\zeta\|_{(2),\omega_\beta}^2 + \underline{R} \|\vartheta\|_{(2),\omega_\beta}^2 \\
 &\geq \frac{1}{2}\underline{W} \|\zeta\|_{(2),\omega_\beta}^2 + \frac{1}{2c}\underline{W} \|\dot{\zeta}\|_{(2),\omega_\beta}^2 - \frac{1}{2}\underline{W} \|\vartheta\|_{(2),\omega_\beta}^2 + \underline{R} \|\vartheta\|_{(2),\omega_\beta}^2 \\
 &\geq \frac{1}{2}\underline{W} \min\left\{1, \frac{1}{c}\right\} \left(\|\zeta\|_{(2),\omega_\beta}^2 + \|\dot{\zeta}\|_{(2),\omega_\beta}^2 \right) + \left(\underline{R} - \frac{1}{2}\underline{W} \right) \|\vartheta\|_{(2),\omega_\beta}^2.
 \end{aligned}$$

Da in (V3b) die Bedingung $\left(\underline{R} - \frac{1}{2}\underline{W}\right) > 0$ gefordert wurde, erhalten wir somit auch in diesem Fall das gewünschte Resultat. \square

Aufbauend auf dem vorangegangenen Lemma können wir zeigen, dass die eben ermittelte Koerzitivität von \mathcal{Q} die strikte Konvexität des Zielfunktionals J nach sich zieht, vgl. [ASS13], S. 22.

Lemma 4.10

Ist die Abbildung \mathcal{Q} auf dem Kern der Abbildung \mathcal{A} koerzitiv, so ist das Funktional J aus (4.1) bzw. (4.18) strikt konvex auf der zulässigen Menge \mathcal{Z} .

Beweis. Für $z_1, z_2 \in \mathcal{Z}$, $z_1 \neq z_2$ ist $z_1 - z_2 \in \ker \mathcal{A}$. Aus der Koerzitivität folgt dann die Existenz einer Konstanten $\alpha > 0$, sodass

$$\alpha \cdot \|z_1 - z_2\|_{\mathcal{H}}^2 \leq \mathcal{Q}(z_1 - z_2; z_1 - z_2) = \mathcal{Q}(z_1; z_1) - 2\mathcal{Q}(z_1; z_2) + \mathcal{Q}(z_2; z_2)$$

gilt und damit

$$2\mathcal{Q}(z_1; z_2) \leq \mathcal{Q}(z_1; z_1) + \mathcal{Q}(z_2; z_2) - \alpha \cdot \|z_1 - z_2\|_{\mathcal{H}}^2$$

richtig ist. Wir multiplizieren die letzte Ungleichung mit $\lambda(\lambda - 1)$, wobei $\lambda \in (0, 1)$ gilt, und erhalten

$$2\lambda(\lambda - 1)\mathcal{Q}(z_1; z_2) \leq \lambda(\lambda - 1) [\mathcal{Q}(z_1; z_1) + \mathcal{Q}(z_2; z_2)] - \alpha \lambda(\lambda - 1) \cdot \|z_1 - z_2\|_{\mathcal{H}}^2.$$

Damit können wir für alle $z_1, z_2 \in \mathcal{Z}$ mit $z_1 \neq z_2$ auf die Abschätzung

$$\begin{aligned}
 \mathcal{Q}((1 - \lambda)z_1 + \lambda z_2; (1 - \lambda)z_1 + \lambda z_2) &= (1 - \lambda)^2 \mathcal{Q}(z_1; z_1) + 2\lambda(1 - \lambda)\mathcal{Q}(z_1; z_2) + \lambda^2 \mathcal{Q}(z_2; z_2) \\
 &\leq \left[(1 - \lambda)^2 + \lambda(1 - \lambda) \right] \mathcal{Q}(z_1; z_1) + \left[\lambda(1 - \lambda) + \lambda^2 \right] \mathcal{Q}(z_2; z_2) - \alpha \lambda(\lambda - 1) \cdot \|z_1 - z_2\|_{\mathcal{H}}^2 \\
 &= (1 - \lambda)\mathcal{Q}(z_1; z_1) + \lambda\mathcal{Q}(z_2; z_2) - \alpha \lambda(\lambda - 1) \cdot \|z_1 - z_2\|_{\mathcal{H}}^2 \\
 &< (1 - \lambda)\mathcal{Q}(z_1; z_1) + \lambda\mathcal{Q}(z_2; z_2)
 \end{aligned}$$

schließen, womit bewiesen ist, dass \mathcal{Q} strikt konvex auf \mathcal{Z} ist. Da q linear ist, erhalten wir

dann sofort für alle $z_1, z_2 \in \mathcal{Z}$, $z_1 \neq z_2$ und $\lambda \in (0, 1)$ die Abschätzung

$$\begin{aligned} J((1-\lambda)z_1 + \lambda z_2) &= \frac{1}{2} \mathcal{Q}((1-\lambda)z_1 + \lambda z_2; (1-\lambda)z_1 + \lambda z_2) + q((1-\lambda)z_1 + \lambda z_2) \\ &< (1-\lambda) \cdot \frac{1}{2} \mathcal{Q}(z_1; z_1) + (1-\lambda) \cdot q(z_1) + \lambda \cdot \frac{1}{2} \mathcal{Q}(z_2; z_2) + \lambda \cdot q(z_2) \\ &= (1-\lambda)J(z_1) + \lambda J(z_2) \end{aligned}$$

und somit auch die strikte Konvexität von J auf \mathcal{Z} . □

Da die zulässige Menge \mathcal{Z} konvex ist und auch das Zielfunktional J (strikt) konvex auf \mathcal{Z} ist, handelt es sich bei der Aufgabe in (4.18) um ein konvexes Optimierungsproblem. Demnach erlaubt uns die Formulierung der Steuerungsaufgabe $(\text{LQ})_\infty$ in den gewichteten Funktionenräumen die vorteilhafte Behandlung der Steuerungsaufgabe über den unendlichen Zeithorizont $[0, \infty)$ mit Hilfe klassischer Methoden aus der konvexen Analysis.

Die Niveaumengen $\mathcal{N}_{\mathcal{Z}}(J, \alpha)$

Für $\alpha \in \mathbb{R}$ sind die *Niveaumengen von J* durch

$$\mathcal{N}(J, \alpha) := \{(x, u) \in \mathcal{H} : J(x, u) \leq \alpha\}$$

definiert. Da in unserer Aufgabenstellung Nebenbedingungen auftreten, betrachten wir anstelle der Niveaumengen von J die Durchschnittsmengen

$$\mathcal{N}_{\mathcal{Z}}(J, \alpha) := \mathcal{N}(J, \alpha) \cap \mathcal{Z} = \{(x, u) \in \mathcal{Z} : J(x, u) \leq \alpha\},$$

also gerade die *Niveaumengen der restringierten Optimierungsaufgabe* $(\text{LQ})_\infty$, vgl. (4.18). Wir werden nun zeigen, dass diese Niveaumengen schwach kompakt sind. Dafür formulieren wir zunächst das folgende Resultat:

Lemma 4.11

Für jedes zulässige Paar $(x, u) \in \mathcal{Z}$ ist die Niveaumenge $\mathcal{N}_{\mathcal{Z}}(J, J(x, u))$ konvex, abgeschlossen und beschränkt (bezüglich der Normtopologie von \mathcal{H}).

Beweis: Für den Nachweis der Konvexität der Niveaumenge wählen wir $z_1 := (x_1, u_1)$, $z_2 := (x_2, u_2) \in \mathcal{N}_{\mathcal{Z}}(J, J(x, u))$ und $\lambda \in [0, 1]$ beliebig. Das Zielfunktional J ist strikt konvex (auf \mathcal{Z}), also insbesondere auch konvex, vgl. Lemma 4.10. Somit ergibt sich

$$\begin{aligned} J((1-\lambda)z_1 + \lambda z_2) &\leq (1-\lambda)J(z_1) + \lambda J(z_2) \\ &= (1-\lambda)J(x_1, u_1) + \lambda J(x_2, u_2) \\ &\leq (1-\lambda)J(x, u) + \lambda J(x, u) = J(x, u) \end{aligned}$$

und daher $(1-\lambda)z_1 + \lambda z_2 \in \mathcal{N}_{\mathcal{Z}}(J, J(x, u))$.

Aus der Stetigkeit der Abbildungen \mathcal{Q} und q (vgl. Lemma 4.4 und Lemma 4.5) folgt sofort die Stetigkeit des Zielfunctionals J auf \mathcal{H} . Aufgrund dessen können wir direkt auf die Abgeschlossenheit von $\mathcal{N}_{\mathcal{Z}}(J, J(x, u))$ schließen. Sei nämlich $\{(x_N, u_N)\}_{N \in \mathbb{N}}$ eine konvergente

Funktionenfolge aus $\mathcal{N}_{\mathcal{Z}}(J, J(x, u))$, die in der Norm des Raumes \mathcal{H} gegen (x_0, u_0) konvergiert, so folgt zunächst aus der Abgeschlossenheit von \mathcal{Z} , dass $(x_0, u_0) \in \mathcal{Z}$ gilt, vgl. Lemma 4.7. Die Stetigkeit von J führt zu der Abschätzung

$$J(x_0, u_0) = J\left(\lim_{N \rightarrow \infty} x_N, \lim_{N \rightarrow \infty} u_N\right) = \lim_{N \rightarrow \infty} J(x_N, u_N) \leq J(x, u).$$

Demnach ist $(x_0, u_0) \in \mathcal{N}_{\mathcal{Z}}(J, J(x, u))$ und somit die Niveaumenge $\mathcal{N}_{\mathcal{Z}}(J, J(x, u))$ abgeschlossen.

Zum Beweis der Beschränktheit sei $(\bar{x}, \bar{u}) \in \mathcal{Z}$ beliebig, aber fest gewählt. Für $(x, u) \in \mathcal{Z}$ definieren wir $\zeta := x - \bar{x}$, $\vartheta := u - \bar{u}$ und betrachten die Differenz

$$\begin{aligned} J(x, u) - J(\bar{x}, \bar{u}) &= \frac{1}{2} \mathcal{Q}((x, u); (x, u)) - \frac{1}{2} \mathcal{Q}((\bar{x}, \bar{u}); (\bar{x}, \bar{u})) + q(x, u) - q(\bar{x}, \bar{u}) \\ &= \frac{1}{2} \mathcal{Q}((x, u); (x, u)) - \frac{1}{2} \mathcal{Q}((\bar{x}, \bar{u}); (\bar{x}, \bar{u})) + q(\zeta, \vartheta). \end{aligned} \quad (4.20)$$

Dabei haben wir in der zweiten Gleichung die Linearität der Abbildung q ausgenutzt. Wir möchten weiter umformen und wählen dafür zunächst beliebige Elemente $z_1, z_2 \in \mathcal{H}$. Aufgrund der Linearität und Symmetrie von \mathcal{Q} gelten die folgenden Gleichungen:

$$\begin{aligned} &\frac{1}{2} \mathcal{Q}(z_1 - z_2; z_1 - z_2) + \mathcal{Q}(z_2; z_1 - z_2) \\ &= \frac{1}{2} [\mathcal{Q}(z_1; z_1) - \mathcal{Q}(z_1; z_2) - \mathcal{Q}(z_2; z_1) + \mathcal{Q}(z_2; z_2)] + \mathcal{Q}(z_2; z_1) - \mathcal{Q}(z_2; z_2) \\ &= \frac{1}{2} [\mathcal{Q}(z_1; z_1) - \mathcal{Q}(z_1; z_2) - \mathcal{Q}(z_1; z_2) + \mathcal{Q}(z_2; z_2)] + \mathcal{Q}(z_1; z_2) - \mathcal{Q}(z_2; z_2) \\ &= \frac{1}{2} \mathcal{Q}(z_1; z_1) - \frac{1}{2} \mathcal{Q}(z_2; z_2). \end{aligned}$$

Setzen wir in der obigen Rechnung $z_1 := (x, u)$, $z_2 := (\bar{x}, \bar{u})$, so können wir damit die Gleichung (4.20) weiter umschreiben und erhalten

$$J(x, u) - J(\bar{x}, \bar{u}) = \frac{1}{2} \mathcal{Q}((\zeta, \vartheta); (\zeta, \vartheta)) + \mathcal{Q}((\bar{x}, \bar{u}); (\zeta, \vartheta)) + q(\zeta, \vartheta). \quad (4.21)$$

Sowohl die Bilinearform \mathcal{Q} als auch die Abbildung q ist beschränkt, vgl. Lemma 4.4 und Lemma 4.5. Demnach existieren Konstanten $c_1, c_2 > 0$, sodass

$$\mathcal{Q}((\bar{x}, \bar{u}); (\zeta, \vartheta)) + q(\zeta, \vartheta) \geq -c_1 \cdot \|(\bar{x}, \bar{u})\|_{\mathcal{H}} \cdot \|(\zeta, \vartheta)\|_{\mathcal{H}} - c_2 \cdot \|(\zeta, \vartheta)\|_{\mathcal{H}}$$

gilt. Da $(\zeta, \vartheta) \in \ker \mathcal{A}$ gilt, folgt aus der Koerzivitätseigenschaft von \mathcal{Q} die Existenz einer Konstanten $\alpha > 0$, sodass

$$\mathcal{Q}((\zeta, \vartheta); (\zeta, \vartheta)) \geq \alpha \cdot \|(\zeta, \vartheta)\|_{\mathcal{H}}^2$$

gilt, vgl. Lemma 4.9. Gemeinsam mit (4.21) können wir daher für alle $(x, u) \in \mathcal{Z}$ auf

$$\begin{aligned} J(x, u) &= \frac{1}{2} \mathcal{Q}((\zeta, \vartheta); (\zeta, \vartheta)) + \mathcal{Q}((\bar{x}, \bar{u}); (\zeta, \vartheta)) + q(\zeta, \vartheta) + J(\bar{x}, \bar{u}) \\ &\geq \frac{\alpha}{2} \|(x - \bar{x}, u - \bar{u})\|_{\mathcal{H}}^2 - c_1 \cdot \|(\bar{x}, \bar{u})\|_{\mathcal{H}} \cdot \|(x - \bar{x}, u - \bar{u})\|_{\mathcal{H}} \\ &\quad - c_2 \cdot \|(x - \bar{x}, u - \bar{u})\|_{\mathcal{H}} + J(\bar{x}, \bar{u}) \end{aligned}$$

schließen. Angenommen, es existiert eine Folge $\{(x_N, u_N)\}_{N \in \mathbb{N}} \subset \mathcal{N}_{\mathcal{Z}}(J, J(x, u))$, sodass $\|(x_N, u_N)\|_{\mathcal{H}} \rightarrow +\infty$ für $N \rightarrow \infty$ gilt. Die Elemente der Folge sind zulässig, also ist $(x_N, u_N) \in \mathcal{Z}$ für alle $N \in \mathbb{N}$ richtig. Nach der vorangegangenen Diskussion folgt dann aber auch

$$\lim_{N \rightarrow \infty} J(x_N, u_N) = +\infty.$$

Dies steht jedoch im Widerspruch zu der Voraussetzung, dass die Folge $\{(x_N, u_N)\}$ aus der Niveaumenge $\mathcal{N}_{\mathcal{Z}}(J, J(x, u))$ gewählt war, also $J(x_N, u_N) \leq J(x, u)$ gilt. Die Annahme war demnach falsch, das heißt, eine solche unbeschränkte Folge kann es nicht geben. Somit ist die Behauptung wahr und die Menge $\mathcal{N}_{\mathcal{Z}}(J, J(x, u))$ für jedes $(x, u) \in \mathcal{Z}$ beschränkt. \square

Aufbauend auf dem vorangegangenen Lemma können wir auf die schwache Kompaktheit der Mengen $\mathcal{N}_{\mathcal{Z}}(J, J(x, u))$ schließen. Dafür verwenden wir einfach, dass jede konvexe, abgeschlossene und (norm-)beschränkte Teilmenge $M \subseteq B$ eines reflexiven Banachraumes B schwach kompakt ist (vgl. [Göp73], S.93, Satz 26), und erhalten somit das gewünschte Resultat.

Lemma 4.12

Für jedes zulässige Paar $(x, u) \in \mathcal{Z}$ ist die Niveaumenge $\mathcal{N}_{\mathcal{Z}}(J, J(x, u))$ schwach kompakt.

An dieser Stelle können wir nun alle Ergebnisse dieses Abschnittes zusammenfassen und die Existenz und die Eindeutigkeit einer Lösung der Aufgabe $(LQ)_{\infty}$ folgern.

Satz 4.13

Die Voraussetzungen (V1)-(V5) seien erfüllt. Dann hat die Aufgabe $(LQ)_{\infty}$ eine eindeutig bestimmte Lösung.

Beweis: Wie bereits angekündigt, erhalten wir aus dem verallgemeinerten Satz von Weierstraß ein Existenzresultat für eine Lösung der Aufgabe $(LQ)_{\infty}$, indem wir die entsprechenden Niveaumengen $\mathcal{N}_{\mathcal{Z}}$ des Zielfunktional J betrachten. Dazu wählen wir ein Paar $(x_1, u_1) \in \mathcal{Z}$, deren Existenz aufgrund der Voraussetzung (V5) gesichert ist. Da (x_1, u_1) selbst ein Element der Niveaumenge $\mathcal{N}_{\mathcal{Z}}(J, J(x_1, u_1))$ ist, gilt $\mathcal{N}_{\mathcal{Z}}(J, J(x_1, u_1)) \neq \emptyset$. Nun nimmt das schwach unterhalbstetige Funktional J auf der schwach kompakten Menge $\mathcal{N}_{\mathcal{Z}}(J, J(x_1, u_1))$ seinen Minimalwert an, das heißt, es existiert ein Funktionenpaar $(x^*, u^*) \in \mathcal{N}_{\mathcal{Z}}(J, J(x_1, u_1))$ mit

$$J(x^*, u^*) \leq J(x, u) \quad \forall (x, u) \in \mathcal{N}_{\mathcal{Z}}(J, J(x_1, u_1)).$$

Für alle anderen in der Aufgabe $(LQ)_\infty$ zulässigen Paare $(x, u) \in \mathcal{Z} \setminus \mathcal{N}_Z(J, J(x_1, u_2))$ ist

$$J(x, u) > J(x_1, u_2) \geq J(x^*, u^*).$$

Daher gilt

$$J(x^*, u^*) = \inf_{(x, u) \in \mathcal{Z}} J(x, u) \quad (4.22)$$

und (x^*, u^*) ist demnach eine Lösung der Aufgabe $(LQ)_\infty$. Insbesondere besitzen $(LQ)_\infty$ und die Hilfsaufgabe

$$J(x, u) \rightarrow \inf!$$

$$\text{bzgl. } (x, u) \in \mathcal{N}_Z(J, J(x_1, u_1))$$

mit $(x_1, u_1) \in \mathcal{Z}$ dasselbe optimale Paar (x^*, u^*) und denselben Optimalwert $J(x^*, u^*)$. Die strikte Konvexität von J garantiert die Eindeutigkeit der Lösung. Um dies einzusehen, sei (\tilde{x}, \tilde{u}) eine weitere Lösung von $(LQ)_\infty$, es gilt also $J(x^*, u^*) = J(\tilde{x}, \tilde{u})$. Da die Menge \mathcal{Z} konvex ist, ist auch $(x_2, u_2) \in \mathcal{Z}$ mit $x_2 := \frac{1}{2}(x^* + \tilde{x})$, $u_2 := \frac{1}{2}(u^* + \tilde{u})$. Angenommen es wäre $(x^*, u^*) \neq (\tilde{x}, \tilde{u})$, dann liefert die strikte Konvexität von J die Ungleichung

$$J(x_2, u_2) < \frac{1}{2}J(x^*, u^*) + \frac{1}{2}J(\tilde{x}, \tilde{u}) = J(x^*, u^*),$$

welche im Widerspruch zu (4.22) und damit zu der Optimalität von (x^*, u^*) steht. Daher war die Annahme falsch und es muss $(x^*, u^*) = (\tilde{x}, \tilde{u})$ gelten, was die Eindeutigkeit der Lösung von $(LQ)_\infty$ beweist. \square

4.6 Optimalitätsbedingungen

In dem ersten Teil dieses Abschnittes formulieren und beweisen wir zunächst notwendige Optimalitätsbedingungen für die linear-quadratische Steuerungsaufgabe $(LQ)_\infty$. Am Ende des Abschnittes werden wir dann sehen, dass wir sogar hinreichende Bedingungen erhalten.

Bevor wir ein Maximumprinzip für die Aufgabe $(LQ)_\infty$ formulieren, benötigen wir noch eine zusätzliche Voraussetzung, die sich auf das dynamische System

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + C(t)$$

bezieht, vgl. (4.3). Konkret stellen wir dabei eine Bedingung an die Eingangsmatrix $B(t)$:

(V6) Wir fordern, dass

$$\underline{B} := \inf_{t \in [0, \infty)} \lambda_{\min}(B^T B)(t) > 0 \quad (4.23)$$

gilt, wobei $\lambda_{\min}(B^T B)(t)$ den vom Parameter t abhängigen kleinsten Eigenwert der Matrix $(B^T B)(t)$ bezeichnet.

Satz 4.14 (Maximumprinzip für $(LQ)_\infty$)

Es sei $U = \mathbb{R}^m$. Weiter seien die Voraussetzungen (V1)-(V3) und (V5)-(V6) erfüllt und $(x^*, u^*) \in \mathcal{Z}$ optimal in der Aufgabe $(LQ)_\infty$ im Sinne der Definition 4.1. Dann existieren Multiplikatoren (λ_0, y_0) mit

$$\lambda_0 = 1, \tag{N}$$

$$y_0 \in W_2^{1,n}((0, \infty), \omega_\beta^{-1}), \omega_\beta^{-1}(t) = e^{\beta t}, \beta > 0 \tag{T}$$

sodass die Maximumbedingung

$$H(t, x^*(t), u^*(t), y_0(t), \lambda_0) = \max_{v \in \mathbb{R}^m} H(t, x^*(t), v, y_0(t), \lambda_0) \quad \text{f.ü. auf } (0, \infty) \tag{M}$$

und die kanonische Gleichung

$$\dot{y}_0(t) = -\nabla_\xi H(t, x^*(t), u^*(t), y_0(t), \lambda_0) \quad \text{f.ü. auf } (0, \infty) \tag{K}$$

erfüllt sind. Dabei heißt die Funktion $H : \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$, definiert durch

$$\begin{aligned} H(t, \xi, v, \eta, \lambda_0) &:= -\lambda_0 \left(\frac{1}{2} \{ \xi^T W(t) \xi + v^T R(t) v \} + w^T(t) \xi + r^T(t) v \right) \omega_\beta(t) \\ &\quad + \eta^T (A(t) \xi + B(t) v + C(t)) \end{aligned}$$

Pontryagin Funktion.

Beweis. Wie bereits in Abschnitt 4.1, S. 73f. erwähnt, können wir bei der Beweisführung auf die Ergebnisse unserer Forschungsgruppe zurückgreifen. Insbesondere folgen wir den Beweisschritten in [Pic15, ZP18, PB18].

Schritt 1: In dem ersten Schritt beweisen wir ein Variationslemma und betrachten dafür ein beliebiges zulässiges Funktionenpaar $(x, u) \in \mathcal{Z}$ und das in der Aufgabe $(LQ)_\infty$ im Sinne der Definition 4.1 optimale Paar $z^* := (x^*, u^*) \in \mathcal{Z}$. Wir zeigen nun, dass für alle $\zeta := x - x^* \in W_2^{1,n}((0, \infty), \omega_\beta)$ und alle $\vartheta := u - u^* \in L_2^m((0, \infty), \omega_\beta)$ die einseitige Variation des Zielfunktional J in (x^*, u^*) existiert und nicht negativ ist, das heißt, es gilt

$$0 \leq \delta^+ J((x^*, u^*); (\zeta, \vartheta)) := \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \{ J(x^* + \varepsilon \zeta, u^* + \varepsilon \vartheta) - J(x^*, u^*) \}.$$

Mithilfe der symmetrischen Bilinearform \mathcal{Q} und des linearen Funktionals q können wir leicht die einseitige Richtungsableitung von J an der Stelle (x^*, u^*) in Richtung $h := (\zeta, \vartheta)$ berech-

nen und erhalten für den obigen Grenzwert die Rechnung

$$\begin{aligned}
 & \lim_{\varepsilon \rightarrow 0+0} \frac{1}{\varepsilon} \{J(z^* + \varepsilon h) - J(z^*)\} \\
 &= \lim_{\varepsilon \rightarrow 0+0} \frac{1}{\varepsilon} \left\{ \frac{1}{2} \mathcal{Q}(z^* + \varepsilon h; z^* + \varepsilon h) + q(z^* + \varepsilon h) - \frac{1}{2} \mathcal{Q}(z^*; z^*) - q(z^*) \right\} \\
 &= \lim_{\varepsilon \rightarrow 0+0} \frac{1}{\varepsilon} \left\{ \frac{1}{2} \mathcal{Q}(z^*; z^*) + \varepsilon \mathcal{Q}(z^*; h) + \frac{\varepsilon^2}{2} \mathcal{Q}(h; h) + q(z^*) + \varepsilon q(h) - \frac{1}{2} \mathcal{Q}(z^*; z^*) - q(z^*) \right\} \\
 &= \lim_{\varepsilon \rightarrow 0+0} \frac{1}{\varepsilon} \left\{ \varepsilon \mathcal{Q}(z^*; h) + \frac{\varepsilon^2}{2} \mathcal{Q}(h; h) + \varepsilon q(h) \right\} = \lim_{\varepsilon \rightarrow 0+0} \left\{ \mathcal{Q}(z^*; h) + \frac{\varepsilon}{2} \mathcal{Q}(h; h) + q(h) \right\} \\
 &= \mathcal{Q}(z^*; h) + q(h) < \infty.
 \end{aligned}$$

Die Endlichkeit des Grenzwertes ergibt sich sofort aus der Beschränktheit von \mathcal{Q} und q auf $W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta)$, vgl. Lemma 4.4 und Lemma 4.5. Weiter folgt aus der Linearität und der Stetigkeit von \mathcal{Q} und q , dass auch die Abbildung $h \mapsto \mathcal{Q}(z^*; h) + q(h)$ linear und stetig in $h := (\zeta, \vartheta) \in W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta)$ ist. Daher existiert die einseitige Gâteaux-Ableitung $\delta^+ J((x^*, u^*); (\zeta, \vartheta))$ und ist durch

$$\begin{aligned}
 & \delta^+ J((x^*, u^*); (\zeta, \vartheta)) \\
 &= \mathcal{Q}((x^*, u^*); (\zeta, \vartheta)) + q(\zeta, \vartheta) \\
 &= \int_0^\infty \{ (x^*(t))^T W(t) \zeta(t) + (u^*(t))^T R(t) \vartheta(t) + w^T(t) \zeta(t) + r^T(t) \vartheta(t) \} \omega_\beta(t) dt \quad (4.24)
 \end{aligned}$$

gegeben. Aus der Optimalität von (x^*, u^*) ergibt sich, dass $\delta^+ J((x^*, u^*); (\zeta, \vartheta))$ nicht negativ ist.

Schritt 2: Wir führen die Bezeichnungen

$$\begin{aligned}
 a_1(\vartheta) &:= \int_0^\infty \{ (u^*(t))^T R(t) \vartheta(t) + r^T(t) \vartheta(t) \} \omega_\beta(t) dt \\
 b_1(\zeta) &:= \int_0^\infty \{ (x^*(t))^T W(t) \zeta(t) + w^T(t) \zeta(t) \} \omega_\beta(t) dt \\
 a_2(\vartheta) &:= -B(\cdot) \vartheta \\
 b_2(\zeta) &:= \dot{\zeta} - A(\cdot) \zeta
 \end{aligned}$$

ein. Die dadurch definierten Operatoren besitzen die folgenden Eigenschaften:

- Der Operator $a_1 : L_2^m((0, \infty), \omega_\beta) \rightarrow \mathbb{R}$ ist linear, beschränkt und stetig.
- Der Operator $b_1 : W_2^{1,n}((0, \infty), \omega_\beta) \rightarrow \mathbb{R}$ ist linear, beschränkt und stetig.
- Der Operator $a_2 : L_2^m((0, \infty), \omega_\beta) \rightarrow L_2^n((0, \infty), \omega_\beta)$ ist linear, beschränkt und stetig.

- Der Operator $b_2 : W_2^{1,n}((0, \infty), \omega_\beta) \rightarrow L_2^n((0, \infty), \omega_\beta)$ ist linear, beschränkt und stetig.

Die Linearität der Operatoren lässt sich leicht nachrechnen. Die Beschränktheit von a_1 ergibt sich direkt aus der Abschätzung

$$\begin{aligned} |a_1(\vartheta)| &\leq \left| \int_0^\infty (u^*(t))^T R(t) \vartheta(t) \omega_\beta(t) dt \right| + \left| \int_0^\infty r^T(t) \vartheta(t) \omega_\beta(t) dt \right| \\ &\leq \|R\|_{(\infty)} \cdot \|u^*\|_{(2), \omega_\beta} \cdot \|\vartheta\|_{(2), \omega_\beta} + \|r\|_{(2), \omega_\beta} \cdot \|\vartheta\|_{(2), \omega_\beta}. \end{aligned}$$

Dabei sind wir wie in dem Beweis der Beschränktheit von Q und q vorgegangen, vgl. Lemma 4.4 und Lemma 4.5. Wir haben die Dreiecksungleichung und die Ungleichung von Cauchy-Schwarz verwendet und darüber hinaus die Verträglichkeit der Spektral- mit der euklidischen Vektornorm, vgl. (2.22), ausgenutzt. Mit denselben Argumenten erhalten wir auch

$$|b_1(\zeta)| \leq \|W\|_{(\infty)} \cdot \|x^*\|_{(2), \omega_\beta} \cdot \|\zeta\|_{(2), \omega_\beta} + \|w\|_{(2), \omega_\beta} \cdot \|\zeta\|_{(2), \omega_\beta}.$$

Für den Operator a_2 gilt

$$\|a_2(\vartheta)\|_{(2), \omega_\beta}^2 = \int_0^\infty |B(t) \vartheta(t)|_2^2 \omega_\beta(t) dt \leq \|B\|_{(\infty)}^2 \cdot \|\vartheta\|_{(2), \omega_\beta}^2$$

und in ähnlicher Weise ergibt sich

$$\|b_2(\zeta)\|_{(2), \omega_\beta}^2 = \|\dot{\zeta} - A\zeta\|_{(2), \omega_\beta}^2 \leq \|\dot{\zeta}\|_{(2), \omega_\beta}^2 + \|A\|_{(\infty)}^2 \cdot \|\zeta\|_{(2), \omega_\beta}^2,$$

womit auch die Beschränktheit von a_2 und b_2 bewiesen ist. Aus der Linearität und der Beschränktheit der Operatoren a_1, b_1, a_2, b_2 folgt sofort auch deren Stetigkeit.

Mit den eingeführten Operatoren können wir die Menge der Variationen M_0 durch

$$\begin{aligned} M_0 = \{ &(z_0, z) \in \mathbb{R} \times L_2^n((0, \infty), \omega_\beta) \mid z_0 = a_1(\vartheta) + b_1(\zeta), \\ &z = a_2(\vartheta) + b_2(\zeta), \\ &\zeta := x - x^* \in W_2^{1,n}((0, \infty), \omega_\beta), \|\zeta\|_{1,(2), \omega_\beta} \leq 1, \\ &\zeta(0) = 0, \\ &\vartheta := u - u^* \in L_2^m((0, \infty), \omega_\beta), \|\vartheta\|_{(2), \omega_\beta} \leq 1 \} \end{aligned}$$

definieren.

Schritt 3: In dem dritten Beweisschritt studieren wir die Menge M_0 genauer. Wir weisen zunächst nach, dass die Menge M_0 konvex und abgeschlossen ist.

Konvexität von M_0 : Wir betrachten zwei Elemente $(z_0, z), (\hat{z}_0, \hat{z}) \in M_0$ mit zugehörigen Funktionen $(\zeta, \vartheta) \in W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta)$ und $(\hat{\zeta}, \hat{\vartheta}) \in W_2^{1,n}((0, \infty), \omega_\beta) \times$

$L_2^m((0, \infty), \omega_\beta)$. Dann gilt aufgrund der Linearität der Operatoren a_1, b_1, a_2, b_2 mit $\lambda \in [0, 1]$

$$\begin{aligned} (1 - \lambda) z_0 + \lambda \hat{z}_0 &= (1 - \lambda) a_1(\vartheta) + (1 - \lambda) b_1(\zeta) + \lambda a_1(\hat{\vartheta}) + \lambda b_1(\hat{\zeta}) \\ &= a_1\left((1 - \lambda)\vartheta + \lambda\hat{\vartheta}\right) + b_1\left((1 - \lambda)\zeta + \lambda\hat{\zeta}\right) \end{aligned}$$

und

$$\begin{aligned} (1 - \lambda) z + \lambda \hat{z} &= (1 - \lambda) a_2(\vartheta) + (1 - \lambda) b_2(\zeta) + \lambda a_2(\hat{\vartheta}) + \lambda b_2(\hat{\zeta}) \\ &= a_2\left((1 - \lambda)\vartheta + \lambda\hat{\vartheta}\right) + b_2\left((1 - \lambda)\zeta + \lambda\hat{\zeta}\right). \end{aligned}$$

Weiterhin ist $(1 - \lambda)\zeta + \lambda\hat{\zeta} \in W_2^{1,n}((0, \infty), \omega_\beta)$ und $(1 - \lambda)\vartheta + \lambda\hat{\vartheta} \in L_2^m((0, \infty), \omega_\beta)$. Darüber hinaus gilt $(1 - \lambda)\zeta(0) + \lambda\hat{\zeta}(0) = 0$ und Anwenden der Dreiecksungleichung liefert die Normbeschränkung

$$\left\| (1 - \lambda)\zeta + \lambda\hat{\zeta} \right\|_{1,(2),\omega_\beta} \leq (1 - \lambda) \|\zeta\|_{1,(2),\omega_\beta} + \lambda \|\hat{\zeta}\|_{1,(2),\omega_\beta} \leq (1 - \lambda) + \lambda = 1.$$

Analog erhalten wir auch

$$\left\| (1 - \lambda)\vartheta + \lambda\hat{\vartheta} \right\|_{(2),\omega_\beta} \leq 1.$$

Daher ist die Konvexkombination von (z_0, z) und (\hat{z}_0, \hat{z}) in der Menge M_0 enthalten und diese demnach konvex.

Abgeschlossenheit von M_0 : Wir bezeichnen mit \overline{M}_0 den Abschluss der Menge M_0 . Da \overline{M}_0 die Vereinigung von M_0 selbst und dem Rand von M_0 ist, folgt unmittelbar das $M_0 \subseteq \overline{M}_0$ gilt. Wir zeigen nun, dass auch die Inklusion $\overline{M}_0 \subseteq M_0$ richtig ist, woraus sich die Identität $M_0 = \overline{M}_0$ ergibt und M_0 demnach abgeschlossen ist.

Wir betrachten ein beliebiges Paar $(\bar{z}_0, \bar{z}) \in \overline{M}_0$. Dann existiert eine Folge $\{(z_0^N, z^N)\}_{N \in \mathbb{N}}$ mit $(z_0^N, z^N) \in M_0$ die gegen (\bar{z}_0, \bar{z}) konvergiert, das heißt die Komponentenfolgen konvergieren in \mathbb{R} bzw. in dem Raum $L_2^n((0, \infty), \omega_\beta)$ gegen das jeweilige Grenzelement:

$$\{z_0^N\}_{N \in \mathbb{N}} \rightarrow \bar{z}_0 \quad \text{in } \mathbb{R}, \quad \{z^N\}_{N \in \mathbb{N}} \rightarrow \bar{z} \quad \text{in } L_2^n((0, \infty), \omega_\beta).$$

Da die Folge $\{(z_0^N, z^N)\}_{N \in \mathbb{N}}$ in M_0 enthalten ist, gilt für die Elemente dieser Folge

$$z_0^N = a_1(\vartheta^N) + b_1(\zeta^N), \quad z^N = a_2(\vartheta^N) + b_2(\zeta^N) \quad (4.25)$$

mit

$$\begin{aligned} \zeta^N &\in W_2^{1,n}((0, \infty), \omega_\beta), \quad \|\zeta^N\|_{1,(2),\omega_\beta} \leq 1, \quad \zeta^N(0) = 0, \\ \vartheta^N &\in L_2^m((0, \infty), \omega_\beta), \quad \|\vartheta^N\|_{(2),\omega_\beta} \leq 1. \end{aligned}$$

Die Räume $W_2^{1,n}((0, \infty), \omega_\beta)$ und $L_2^m((0, \infty), \omega_\beta)$ sind reflexiv. Der Satz von Eberlein-Schmuljan⁵ besagt, dass die abgeschlossene Einheitskugel in diesen Räumen schwach kom-

⁵ vgl. [Göp73], S. 94; [DS88], S. 430; [Yos95], S. 141

pakt ist. Demnach existieren schwach konvergente Teilfolgen $\{\zeta^{N_1}\}_{N_1 \in \mathbb{N}} \subseteq \{\zeta^N\}_{N \in \mathbb{N}}$ und $\{\vartheta^{N_1}\}_{N_1 \in \mathbb{N}} \subseteq \{\vartheta^N\}_{N \in \mathbb{N}}$ mit:

$$\begin{aligned} \{\zeta^{N_1}\} &\rightharpoonup \tilde{\zeta}, \quad \tilde{\zeta} \in W_2^{1,n}((0, \infty), \omega_\beta), \quad \|\tilde{\zeta}\|_{1,(2),\omega_\beta} \leq 1, \\ \{\vartheta^{N_1}\} &\rightharpoonup \tilde{\vartheta}, \quad \tilde{\vartheta} \in L_2^m((0, \infty), \omega_\beta), \quad \|\tilde{\vartheta}\|_{(2),\omega_\beta} \leq 1. \end{aligned}$$

Zusätzlich gilt $\tilde{\zeta}(0) = 0$, vgl. [Lyk10], S. 86, Lemma 7. Die Stetigkeit der Operatoren a_1, b_1, a_2, b_2 erlaubt es uns, in (4.25) für $N_1 \rightarrow \infty$ zur Grenze überzugehen:

$$\bar{z}_0 = a_1(\tilde{\vartheta}) + b_1(\tilde{\zeta}), \quad \bar{z} = a_2(\tilde{\vartheta}) + b_2(\tilde{\zeta}).$$

Zusammengefasst haben wir damit gezeigt, dass (\bar{z}_0, \bar{z}) in der Menge M_0 enthalten ist und M_0 demnach abgeschlossen ist.

Wir schließen das Studium der Menge der Variationen M_0 mit einer letzten Charakterisierung ab: Für $\alpha > 0$ ist das Paar $(-\alpha, 0) \in \mathbb{R} \times L_2^n((0, \infty), \omega_\beta)$ kein Element von M_0 .

Wir führen einen Widerspruchsbeweis. Angenommen, es gilt $(-\alpha, 0) \in M_0$ für $\alpha > 0$. Aus der Definition von M_0 folgt dann die Existenz eines Funktionenpaares $(\zeta, \vartheta) \in W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta)$ mit

$$-\alpha = a_1(\vartheta) + b_1(\zeta), \tag{4.26}$$

$$0 = a_2(\vartheta) + b_2(\zeta), \tag{4.27}$$

$$\zeta(0) = 0, \quad \|\zeta\|_{1,(2),\omega_\beta} \leq 1, \quad \|\vartheta\|_{(2),\omega_\beta} \leq 1.$$

Aus (4.27) erhalten wir sofort $A(\cdot)\zeta(\cdot) + B(\cdot)\vartheta(\cdot) - \dot{\zeta}(\cdot) = 0$. Mit $\zeta := x - x^*$ und $\vartheta := u - u^*$, wobei $(x^*, u^*) \in \mathcal{Z}$ das in der Aufgabe $(\text{LQ})_\infty$ optimale Paar ist, folgt

$$\begin{aligned} \mathcal{A}(x, u) &= \mathcal{A}(\zeta + x^*, \vartheta + u^*) = \mathcal{A}(\zeta, \vartheta) + \mathcal{A}(x^*, u^*) \\ &= \left(\begin{array}{c} A(\cdot)\zeta(\cdot) + B(\cdot)\vartheta(\cdot) - \dot{\zeta}(\cdot) \\ \zeta(0) \end{array} \right) + \mathbf{b} = \mathbf{b}. \end{aligned}$$

Daher ist $(x, u) \in \mathcal{Z}$ zulässig in der Aufgabe $(\text{LQ})_\infty$. In dem ersten Beweisschritt haben wir gezeigt, dass dann die Gâteaux-Ableitung $\delta^+ J((x^*, u^*); (\zeta, \vartheta))$ durch die Darstellung in (4.24) gegeben ist und nicht negativ ist. Bringen wir die Gleichung aus (4.26) in diese Darstellung ein, so erhalten wir

$$0 \leq \delta^+ J((x^*, u^*); (\zeta, \vartheta)) = a_1(\vartheta) + b_1(\zeta) = -\alpha,$$

was im Widerspruch zu der Annahme $\alpha > 0$ steht.

Schritt 4: Der vierte Beweisschritt besteht darin, die Mengen $\{(-\alpha, 0)\} \in \mathbb{R} \times L_2^n((0, \infty), \omega_\beta)$, $\alpha > 0$ und M_0 voneinander zu trennen. Daraus ergibt sich eine Variationsungleichung, aus der sich alle Bedingungen des Satzes ableiten lassen.

Wir haben am Ende des dritten Beweisschrittes gezeigt, dass $(-\alpha, 0) \in \mathbb{R} \times L_2^n((0, \infty), \omega_\beta)$ für $\alpha > 0$ kein Element von M_0 ist. Wir trennen dieses Element von der konvexen und

abgeschlossenen Menge M_0 . Aus dem Trennungssatz von Hahn-Banach⁶ folgt die Existenz eines linearen und stetigen Funktionals $q_\alpha^* \in \mathbb{R} \times L_2^n((0, \infty), \omega_\beta)$, sodass die Ungleichung

$$\langle q_\alpha^*, (-\alpha, 0) \rangle \leq \gamma_{1,\alpha} < \gamma_{2,\alpha} \leq \langle q_\alpha^*, (z_0, z) \rangle \quad \forall (z_0, z) \in M_0$$

mit positiven Konstanten $\gamma_{1,\alpha}, \gamma_{2,\alpha}$ richtig ist. Mit $q_\alpha^* = (\lambda_\alpha, y_\alpha)$ liest sich die obige Ungleichung wie folgt:

$$-\lambda_\alpha \alpha < \lambda_\alpha z_0 + \langle y_\alpha, z \rangle_{(2), \omega_\beta} \quad \forall (z_0, z) \in M_0. \quad (4.28)$$

Da das Paar $(0, 0)$ offensichtlich ein Element der Menge M_0 ist, muss die Ungleichung in (4.28) auch für dieses Element gelten. Daraus folgt $\lambda_\alpha > 0$ für jedes $\alpha > 0$. Nun dividieren wir die Ungleichung (4.28) durch $\lambda_\alpha > 0$ und erhalten

$$-\alpha < z_0 + \langle \tilde{y}_\alpha, z \rangle_{(2), \omega_\beta} \quad \forall (z_0, z) \in M_0 \quad (4.29)$$

mit $\tilde{y}_\alpha = y_\alpha / \lambda_\alpha$. Diese Variationsungleichung bildet das wesentliche Kernelement für die weitere Beweisführung. Verwenden wir die Definitionen der Operatoren a_1, b_1, a_2, b_2 , so lässt sich die Ungleichung aus (4.29) in der folgenden Form schreiben: Für alle $\alpha > 0$ und alle (ζ, ϑ) mit

$$\begin{aligned} \zeta &\in W_2^{1,n}((0, \infty), \omega_\beta), \|\zeta\|_{1,(2), \omega_\beta} \leq 1, \zeta(0) = 0, \\ \vartheta &\in L_2^m((0, \infty), \omega_\beta), \|\vartheta\|_{(2), \omega_\beta} \leq 1, \end{aligned}$$

gilt

$$-\alpha < a_1(\vartheta) + b_1(\zeta) + \langle \tilde{y}_\alpha, a_2(\vartheta) + b_2(\zeta) \rangle_{(2), \omega_\beta}. \quad (4.30)$$

Darüber hinaus ist diese Ungleichung für alle $\alpha > 0$ richtig, also speziell auch für $\alpha = 1/N$, $N \in \mathbb{N}$ und mit der Bezeichnung $\tilde{y}_N := \tilde{y}_\alpha$ erfüllt.

Schritt 5: Der Inhalt dieses Beweisschrittes ist es, die Beschränktheit der Folge $\{\tilde{y}_N\}_{N \in \mathbb{N}}$ in dem gesamten $L_2^n((0, \infty), \omega_\beta)$ nachzuweisen. Dies ermöglicht es uns, auf die Existenz einer schwach konvergenten Teilfolge zu schließen, um mit Hilfe dieser und der Variationsungleichung aus (4.30) auf die Maximumbedingung (M) und auf die kanonische Gleichung (K) schließen zu können.

Wir nutzen die Wahl des unbeschränkten Steuerbereiches $U = \mathbb{R}^m$ aus und es wird deutlich, wieso die zusätzliche Voraussetzung (V6) gestellt werden muss. Dafür halten wir $x = x^*$ fest (dann ist $\zeta \equiv 0$) und variieren ausschließlich bezüglich der Steuerung. Aus (4.30) erhalten wir dann für alle $\alpha > 0$ die Ungleichung

$$-\alpha < a_1(\vartheta) + \langle \tilde{y}_\alpha, a_2(\vartheta) \rangle_{(2), \omega_\beta} \quad \forall \vartheta \in L_2^m((0, \infty), \omega_\beta), \|\vartheta\|_{(2), \omega_\beta} \leq 1.$$

⁶ vgl. [Wer18], S. 116, Theorem III.2.4

Daraus können wir unter Berücksichtigung von $\|\vartheta\|_{(2),\omega_\beta} \leq 1$ weiter auf

$$-\alpha\|\vartheta\|_{(2),\omega_\beta} < a_1(\vartheta) + \langle \tilde{y}_\alpha, a_2(\vartheta) \rangle_{(2),\omega_\beta} \quad \forall \vartheta \in L_2^m((0, \infty), \omega_\beta)$$

schließen. Aus der Linearität von a_1 und a_2 folgt dann die Abschätzung

$$\left| a_1(\vartheta) + \langle \tilde{y}_\alpha, a_2(\vartheta) \rangle_{(2),\omega_\beta} \right| < \alpha\|\vartheta\|_{(2),\omega_\beta} \quad \forall \vartheta \in L_2^m((0, \infty), \omega_\beta), \quad (4.31)$$

woraus wir schließlich für alle $\vartheta \in L_2^m((0, \infty), \omega_\beta)$ und alle $\alpha > 0$ die Ungleichung

$$\left| \langle \tilde{y}_\alpha, a_2(\vartheta) \rangle_{(2),\omega_\beta} \right| - |a_1(\vartheta)| \leq \left| a_1(\vartheta) + \langle \tilde{y}_\alpha, a_2(\vartheta) \rangle_{(2),\omega_\beta} \right| < \alpha\|\vartheta\|_{(2),\omega_\beta}$$

erhalten. Speziell für $\alpha = 1/N$, $N \in \mathbb{N}$ können wir zusammen mit der Beschränktheit von a_1 aus der letzten Ungleichung auf

$$\left| \langle \tilde{y}_N, a_2(\vartheta) \rangle_{(2),\omega_\beta} \right| \leq c_1\|\vartheta\|_{(2),\omega_\beta} \quad \forall \vartheta \in L_2^m((0, \infty), \omega_\beta), \forall N \in \mathbb{N} \quad (4.32)$$

schließen, wobei wir $\tilde{y}_N := \tilde{y}_\alpha$ gesetzt haben. Für den nächsten Schritt definieren wir die Menge

$$\mathbf{Z}_1 := \{z \in L_2^n((0, \infty), \omega_\beta) \mid z = a_2(\vartheta), \vartheta \in L_2^m((0, \infty), \omega_\beta)\}$$

und werden zeigen, dass $\mathbf{Z}_1 = L_2^n((0, \infty), \omega_\beta)$ richtig ist. Dafür wählen wir eine beliebiges $z \in L_2^n((0, \infty), \omega_\beta)$ und betrachten die Gleichung

$$z = a_2(\vartheta) = -B(\cdot)\vartheta.$$

Entsprechend der Definition der Lebesgue-Räume sind die obigen Funktionen genau dann äquivalent, wenn sie außerhalb einer Lebesgueschen Nullmenge \mathcal{N} übereinstimmen,

$$z(t) = -B(t)\vartheta(t) \quad \forall t \in (0, \infty) \setminus \mathcal{N}.$$

Multiplikation der obigen Gleichung mit $B^T(t)$ ergibt

$$B^T(t)z(t) = -(B^T B)(t)\vartheta(t) \quad \forall t \in (0, \infty) \setminus \mathcal{N}. \quad (4.33)$$

Dabei ist $(B^T B)(t)$ (für alle $t \in \mathbb{R}_+$) eine symmetrische, quadratische $m \times m$ und positiv semidefinite Matrix, deren Eigenwerte alle reell sind. Aufgrund der Voraussetzung (V6) sind die Eigenwerte insbesondere auch echt größer Null, weshalb $(B^T B)(t)$ sogar eine positiv definite, also reguläre Matrix ist. Wir multiplizieren (4.33) mit der inversen Matrix $(B^T B)^{-1}(t)$ und erhalten

$$(B^T B)^{-1}(t)B^T(t)z(t) = -(B^T B)^{-1}(t)(B^T B)(t)\vartheta(t) = -\vartheta(t) \quad \forall t \in (0, \infty) \setminus \mathcal{N}. \quad (4.34)$$

Aus der Stetigkeit von $B(\cdot)$ auf \mathbb{R}_+ folgt sofort auch die Stetigkeit der Funktionen $B^T(\cdot)$ und $(B^T B)(\cdot)$. Mit Hilfe der Cramer'schen Regel⁷ können wir auf die Stetigkeit der Inversen $(B^T B)^{-1}(\cdot)$ auf \mathbb{R}_+ schließen. Dies führt zu den folgenden Rückschlüssen: Die Vorschrift in

⁷ vgl. z.B. [Göll17], S. 146, Satz 2.68

(4.34) definiert fast überall auf $(0, \infty)$ eine Funktion ϑ . Aus der Stetigkeit der Funktionen $(B^T B)^{-1}(\cdot), B^T(\cdot)$ folgt auch deren Messbarkeit.⁸ Die Funktion $z \in L_2^n((0, \infty), \omega_\beta)$ ist per Definition messbar. Als Produkt messbarer Funktion⁹ erhalten wir auch die Messbarkeit der Funktion ϑ und es bleibt zu zeigen, dass $\vartheta \in L_2^m((0, \infty), \omega_\beta)$ gilt. Aus (4.34) folgt

$$\|\vartheta\|_{(2), \omega_\beta}^2 = \int_0^\infty \left| (B^T B)^{-1}(t) B^T(t) z(t) \right|_2^2 \omega_\beta(t) dt \leq \int_0^\infty \left| (B^T B)^{-1}(t) \right|_2^2 \left| B^T(t) z(t) \right|_2^2 \omega_\beta(t) dt.$$

Für beliebiges, aber festes $t \in \mathbb{R}_+$ erhalten wir aus der Theorie der linearen Algebra die folgenden Aussagen: Da $(B^T B)(t)$ symmetrisch ist, ist auch $(B^T B)^{-1}(t)$ eine symmetrische Matrix. Darüber hinaus gilt: Ist $\lambda(t)$ ein Eigenwert der regulären Matrix $(B^T B)(t)$, so ist $1/\lambda(t)$ ein Eigenwert von $(B^T B)^{-1}(t)$. Da $(B^T B)(t)$ eine positiv definite Matrix ist, demnach also all deren Eigenwerte positiv sind, ergibt sich aufgrund der vorangegangenen Aussage auch die positive Definitheit von $(B^T B)^{-1}(t)$. Für die Spektralnorm folgt dann $\left| (B^T B)^{-1}(t) \right|_2 = \lambda_{\max}(B^T B)^{-1}(t)$, wobei $\lambda_{\max}(B^T B)^{-1}(t)$ der von $t \in \mathbb{R}_+$ abhängige größte Eigenwert der Matrix $(B^T B)^{-1}(t)$ ist, vgl. Lemma 2.23. Darüber hinaus ist $1/\lambda_{\max}(B^T B)^{-1}(t)$ ein Eigenwert von $(B^T B)(t)$ mit

$$\frac{1}{\lambda_{\max}(B^T B)^{-1}(t)} \geq \lambda_{\min}(B^T B)(t) > 0.$$

Mit Hilfe dieser Überlegungen können wir weiter auf

$$\begin{aligned} \|\vartheta\|_{(2), \omega_\beta}^2 &\leq \int_0^\infty \left| (B^T B)^{-1}(t) \right|_2^2 \left| B^T(t) z(t) \right|_2^2 \omega_\beta(t) dt \\ &\leq \int_0^\infty \left[\lambda_{\max}(B^T B)^{-1}(t) \right]^2 \left| B^T(t) z(t) \right|_2^2 \omega_\beta(t) dt \\ &\leq \int_0^\infty \frac{1}{\left[\lambda_{\min}(B^T B)(t) \right]^2} \left| B^T(t) z(t) \right|_2^2 \omega_\beta(t) dt \end{aligned}$$

schließen. Da aus $B \in L_\infty^{n \times m}((0, \infty))$ sofort auch $B^T \in L_\infty^{m \times n}((0, \infty))$ folgt, erhalten wir zusammen mit der Voraussetzung in (4.23)

$$\|\vartheta\|_{(2), \omega_\beta}^2 \leq \frac{1}{\underline{B}^2} \int_0^\infty \left| B^T(t) \right|_2^2 |z(t)|_2^2 \omega_\beta(t) dt \leq \frac{1}{\underline{B}^2} \cdot \left\| B^T \right\|_{(\infty)}^2 \cdot \|z\|_{(2), \omega_\beta}^2 < \infty, \quad (4.35)$$

also $\vartheta \in L_2^m((0, \infty), \omega_\beta)$. Zusammengefasst haben wir damit gezeigt, dass für jedes $z \in L_2^n((0, \infty), \omega_\beta)$ die Existenz einer Funktion $\vartheta \in L_2^m((0, \infty), \omega_\beta)$ mit $z = a_2(\vartheta) = -B(\cdot)\vartheta$ garantiert ist und somit $z \in \mathbf{Z}_1$ folgt. Daher gilt offenbar $\mathbf{Z}_1 = L_2^n((0, \infty), \omega_\beta)$ und der Operator a_2 ist demnach surjektiv.

Nun kehren wir zu der Ungleichung aus (4.32) zurück. Für jedes $\vartheta \in L_2^m((0, \infty), \omega_\beta)$ resultiert aus der Vorschrift $z = a_2(\vartheta) = -B(\cdot)\vartheta$ ein $z \in L_2^n((0, \infty), \omega_\beta)$, für welches aufgrund der vorangegangenen Betrachtungen die Ungleichung in (4.35) gilt. Daher folgt ausgehend von

⁸ vgl. z.B. [For12], §4 Integration messbarer Funktionen

⁹ vgl. z.B. [For12], §4 Integration messbarer Funktionen

(4.32) die Abschätzung

$$\left| \langle \tilde{y}_N, a_2(\vartheta) \rangle_{(2), \omega_\beta} \right| \leq \tilde{c}_1 \cdot \|a_2(\vartheta)\|_{(2), \omega_\beta} \quad \forall \vartheta \in L_2^m((0, \infty), \omega_\beta), \forall N \in \mathbb{N}, \quad (4.36)$$

was bedeutet, dass die Folge $\{\tilde{y}_N\}_{N \in \mathbb{N}}$ auf der Menge Z_1 beschränkt ist. Die Bedingung in (4.36) zieht aufgrund von $Z_1 = L_2^n((0, \infty), \omega_\beta)$ die Beschränktheit der Folge $\{\tilde{y}_N\}_{N \in \mathbb{N}}$ in dem gesamten Raum $L_2^n((0, \infty), \omega_\beta)$ nach sich.

Da dieser reflexiv ist, erhalten wir als direkte Folgerung aus dem Satz von Eberlein-Schmuljan die Existenz einer schwach konvergenten Teilfolge von $\{\tilde{y}_N\}_{N \in \mathbb{N}}$, welche gegen ein Grenzelement $\tilde{y}_0 \in L_2^n((0, \infty), \omega_\beta)$ konvergiert. Wir bezeichnen diese im Folgenden wieder mit $\{\tilde{y}_N\}_{N \in \mathbb{N}}$. Damit haben wir alle Vorkehrungen getroffen, um nun die Bedingungen des Satzes zu beweisen. Wir beginnen mit der Maximumbedingung (M).

Schritt 6: (Maximumbedingung) Für den Nachweis der Maximumbedingung (M) halten wir wieder $x = x^*$ fest und variieren nur bezüglich der Steuerung. Die Variationsungleichung (4.30) geht entsprechend der Diskussion in dem vorangegangenen Beweisschritt in

$$\left| a_1(\vartheta) + \langle \tilde{y}_N, a_2(\vartheta) \rangle_{(2), \omega_\beta} \right| < \frac{1}{N} \|\vartheta\|_{(2), \omega_\beta} \quad \forall \vartheta \in L_2^m((0, \infty), \omega_\beta), \forall N \in \mathbb{N}$$

über, vgl. Ungleichung (4.31) mit $\alpha = 1/N$. Gehen wir nun zur Grenze über, so erhalten wir für $N \rightarrow \infty$

$$a_1(\vartheta) + \langle \tilde{y}_0, a_2(\vartheta) \rangle_{(2), \omega_\beta} = 0 \quad \forall \vartheta \in L_2^m((0, \infty), \omega_\beta). \quad (4.37)$$

Der linear-quadratische Integrand $f(t, \xi, v)$ des Zielfunktional in (4.1) ist (stetig) differenzierbar und insbesondere auch konvex in v . Die Konvexität erlaubt uns die nachfolgende Rechnung mit $\vartheta := u - u^*$ und unter Beachtung der Symmetrie der Matrix $R(t)$ für alle $t \in \mathbb{R}_+$:

$$\begin{aligned} & \int_0^\infty \nabla_v^T f(t, x^*(t), u^*(t)) \cdot \vartheta(t) \omega_\beta(t) dt \leq \int_0^\infty \{f(t, x^*(t), u(t)) - f(t, x^*(t), u^*(t))\} \omega_\beta(t) dt \\ \Leftrightarrow a_1(\vartheta) &= \int_0^\infty \{(u^*(t))^T R(t) \vartheta(t) + r^T(t) \vartheta(t)\} \omega_\beta(t) dt \\ &\leq \int_0^\infty \left\{ \frac{1}{2} \left[u(t)^T R(t) u(t) - (u^*(t))^T R(t) u^*(t) \right] + r^T(t) (u(t) - u^*(t)) \right\} \omega_\beta(t) dt. \end{aligned} \quad (4.38)$$

Aus (4.37) erhalten wir durch Einsetzen der Definition von a_2

$$a_1(\vartheta) = - \langle \tilde{y}_0, a_2(\vartheta) \rangle_{(2), \omega_\beta} = \int_0^\infty \tilde{y}_0^T(t) B(t) (u(t) - u^*(t)) \omega_\beta(t) dt$$

und gemeinsam mit der Abschätzung aus (4.38) ergibt sich

$$0 \leq \int_0^{\infty} \left\{ \frac{1}{2} \left[u(t)^T R(t) u(t) - (u^*(t))^T R(t) u^*(t) \right] + r^T(t) (u(t) - u^*(t)) \right\} \omega_{\beta}(t) \\ + [\tilde{y}_0 \omega_{\beta}(t)]^T B(t) (u(t) - u^*(t)) dt.$$

Wir können leicht überprüfen, dass die obige Ungleichung mit $y_0 := \tilde{y}_0 \omega_{\beta}$ gerade der integrierten Maximumbedingung

$$\int_0^{\infty} H(t, x^*(t), u^*(t), y_0(t), \lambda_0) dt \geq \int_0^{\infty} H(t, x^*(t), u(t), y_0(t), \lambda_0) dt \quad \forall u \in L_2^m((0, \infty), \omega_{\beta})$$

entspricht. Folgen wir abschließend den Techniken in [Wag09] (vgl. S. 553-554, Beweis zu Theorem 2.3), so ergibt sich daraus die fast überall auf $(0, \infty)$ geltende Maximumbedingung (M).

Schritt 7: (Kanonische Gleichung) Die kanonische Gleichung (K) erhalten wir durch Variation des Zustandes bei gleichzeitigem Festhalten der Steuerung $u = u^*$ (dann ist $\vartheta \equiv 0$). Aus (4.30) erhalten wir für alle $\alpha > 0$ die Ungleichung

$$-\alpha < b_1(\zeta) + \langle \tilde{y}_{\alpha}, b_2(\zeta) \rangle_{(2), \omega_{\beta}} \quad \forall \zeta \in W_2^{1,n}((0, \infty), \omega_{\beta}), \|\zeta\|_{1,(2), \omega_{\beta}} \leq 1, \zeta(0) = 0.$$

Daraus können wir wieder unter Berücksichtigung von $\|\zeta\|_{1,(2), \omega_{\beta}} \leq 1$ weiter

$$-\alpha \|\zeta\|_{1,(2), \omega_{\beta}} < b_1(\zeta) + \langle \tilde{y}_{\alpha}, b_2(\zeta) \rangle_{(2), \omega_{\beta}} \quad \forall \zeta \in W_2^{1,n}((0, \infty), \omega_{\beta}), \zeta(0) = 0$$

schließen. Mit denselben Überlegungen wie in dem vorherigen Beweisschritt ergibt sich aus der Linearität von b_1 und b_2 die Ungleichung

$$\left| b_1(\zeta) + \langle \tilde{y}_{\alpha}, b_2(\zeta) \rangle_{(2), \omega_{\beta}} \right| < \alpha \|\zeta\|_{1,(2), \omega_{\beta}} \quad \forall \zeta \in W_2^{1,n}((0, \infty), \omega_{\beta}), \zeta(0) = 0$$

und mit $\alpha = 1/N$ geht diese für alle $N \in \mathbb{N}$ in

$$\left| b_1(\zeta) + \langle \tilde{y}_N, b_2(\zeta) \rangle_{(2), \omega_{\beta}} \right| < \frac{1}{N} \|\zeta\|_{1,(2), \omega_{\beta}} \quad \forall \zeta \in W_2^{1,n}((0, \infty), \omega_{\beta}), \zeta(0) = 0$$

über. Der Grenzübergang $N \rightarrow \infty$ liefert

$$b_1(\zeta) + \langle \tilde{y}_0, b_2(\zeta) \rangle_{(2), \omega_{\beta}} = 0 \quad \forall \zeta \in W_2^{1,n}((0, \infty), \omega_{\beta}), \zeta(0) = 0.$$

Gemäß der Definition von b_1 und b_2 hat die obige Gleichung für alle $\zeta \in W_2^{1,n}((0, \infty), \omega_{\beta})$

mit $\zeta(0) = 0$ die Form

$$\begin{aligned} 0 &= \int_0^{\infty} \left\{ (x^*(t))^T W(t) \zeta(t) + w^T(t) \zeta(t) + \tilde{y}_0^T(t) [\dot{\zeta}(t) - A(t) \zeta(t)] \right\} \omega_{\beta}(t) dt \\ &= \int_0^{\infty} \left\{ [\tilde{y}_0^T(t) \omega_{\beta}(t)] \dot{\zeta}(t) + [(x^*(t))^T W(t) + w^T(t) - \tilde{y}_0^T(t) A(t)] \omega_{\beta}(t) \zeta(t) \right\} dt. \end{aligned}$$

Aufgrund der Inklusion $C_0^{\infty}([0, \infty), \mathbb{R}^n) \subseteq W_2^{1,n}((0, \infty), \omega_{\beta})$ gilt die erhaltende Variationsgleichung insbesondere auch für alle Testfunktionen. Ein Vergleich mit der Definition der schwachen Ableitung liefert, dass $\tilde{y}_0 \omega_{\beta}$ eine schwache Ableitung gemäß der Vorschrift

$$[\tilde{y}_0^T(t) \omega_{\beta}(t)]' = [(x^*(t))^T W(t) + w^T(t) - \tilde{y}_0^T(t) A(t)] \omega_{\beta}(t)$$

besitzt. Definieren wir wieder $y_0 := \tilde{y}_0 \omega_{\beta}$, so erhalten wir die fast überall auf $(0, \infty)$ geltende kanonische Gleichung (K):

$$\dot{y}_0(t) = [(x^*(t))^T W(t) + w^T(t)] \omega_{\beta}(t) - y_0^T(t) A(t) = -\nabla_{\xi} H(t, x^*(t), u^*(t), y_0(t), 1).$$

Schritt 8: (Transversalitätsbedingung) Der letzte Beweisschritt beinhaltet den Nachweis der Transversalitätsbedingungen. Für die Adjungierte $y_0 := \tilde{y}_0 \omega_{\beta}$ gilt:

$$\begin{aligned} \int_0^{\infty} |y_0(t)|_2^2 e^{\beta t} dt &= \int_0^{\infty} |\tilde{y}_0|_2^2 e^{-2\beta t} e^{\beta t} dt = \int_0^{\infty} |\tilde{y}_0|_2^2 \omega_{\beta}(t) dt < \infty, \\ \int_0^{\infty} |\dot{y}_0(t)|_2^2 e^{\beta t} dt &= \int_0^{\infty} \left| [(x^*(t))^T W(t) + w(t)] \omega_{\beta}(t) - y_0^T(t) A(t) \right|_2^2 e^{\beta t} dt \\ &= \int_0^{\infty} \left| (x^*(t))^T W(t) + w(t) - \tilde{y}_0^T(t) A(t) \right|_2^2 e^{-\beta t} dt \\ &\leq \int_0^{\infty} \left(\left| (x^*(t))^T W(t) + w(t) \right|_2 + \left| \tilde{y}_0^T(t) A(t) \right|_2 \right)^2 \omega_{\beta}(t) dt \\ &\leq 2 \int_0^{\infty} \left| (x^*(t))^T W(t) + w(t) \right|_2^2 \omega_{\beta}(t) dt + 2 \int_0^{\infty} \left| \tilde{y}_0^T(t) A(t) \right|_2^2 \omega_{\beta}(t) dt \\ &\leq 4 \int_0^{\infty} \left| (x^*(t))^T W(t) \right|_2^2 \omega_{\beta}(t) dt + 4 \int_0^{\infty} |w(t)|_2^2 \omega_{\beta}(t) dt + 2 \|A\|_{(\infty)}^2 \|\tilde{y}_0\|_{(2), \omega_{\beta}}^2 \\ &\leq 4 \left(\|W\|_{(\infty)}^2 \|x^*\|_{(2), \omega_{\beta}}^2 + \|w\|_{(2), \omega_{\beta}}^2 + \|A\|_{(\infty)}^2 \|\tilde{y}_0\|_{(2), \omega_{\beta}}^2 \right) < \infty. \end{aligned}$$

Daher ist $y_0 \in W_2^{1,n}((0, \infty), \omega_{\beta}^{-1})$, $\omega_{\beta}^{-1}(t) = e^{\beta t}$, $\beta > 0$ richtig und der Beweis des Satzes abgeschlossen. \square

Bemerkung 4.3. Die Transversalitätsbedingung (T) zieht unmittelbar den Grenzwert

$$\lim_{T \rightarrow \infty} |y_0(T)|_2 = 0$$

bzw.

$$\lim_{T \rightarrow \infty} y_0^{(i)}(T) = 0, \quad i = 1, \dots, n$$

mit $y_0 := (y_0^{(1)}, y_0^{(2)}, \dots, y_0^{(n)})^T$ nach sich. Dies ist eine direkte Folgerung aus der Raumzugehörigkeit $y_0 \in W_2^{1,n}((0, \infty), \omega_\beta^{-1})$ und der Lebesgue-Integrierbarkeit von ω_β über $(0, \infty)$, vgl. (2.30). Der Beweis ergibt sich aus den Arbeiten [Pic14], S. 180, Lemma 6; [Mag82], S. 403, 3.1 Lemma; [Bur98], S. 138 ff., Theorem 4.

Bemerkung 4.4. Der Parameter $\beta > 0$ in der Gewichtsfunktion ω_β wird weder durch die benötigten Voraussetzungen für die Existenz einer eindeutigen Lösung von $(LQ)_{(\infty)}$ noch für die hergeleiteten Optimalitätsbedingungen eingeschränkt, weshalb die gewonnenen Resultate für alle $\beta > 0$ verwendet werden können.

Für Optimalsteuerungsaufgaben mit einem unendlichen Betrachtungshorizont der Klasse $(LQ)_\infty$ liefert der Satz 4.14 notwendige Bedingungen zur Ermittlung möglicher Lösungskandidaten. Aufgrund der linear-quadratischen Struktur der Aufgabe $(LQ)_\infty$ können wir sofort auch darauf schließen, dass das Maximumprinzip sogar hinreichende Optimalitätsbedingungen zur Verfügung stellt. In [PBK16] haben wir dies für den Fall eines skalaren Zustandes und einer skalaren Steuerung mit Hilfe des bekannten Königsweges von Carathéodory bewiesen. Die dortigen Argumente bilden nun auch die Grundlage für den Beweis des nachfolgenden Satzes.

Satz 4.15 (Hinreichende Optimalitätsbedingungen für $(LQ)_\infty$)

Es sei $U = \mathbb{R}^m$, die Voraussetzungen (V1)-(V3), (V5) erfüllt und $(x^*, u^*) \in \mathcal{Z}$ ein für die Aufgabe $(LQ)_\infty$ zulässiger Steuerprozess. Ferner gelte, dass das Paar (x^*, u^*) gemeinsam mit der Adjungierten $y_0 \in W_2^{1,n}((0, \infty), \omega_\beta^{-1})$ die Bedingungen (M) und (K) in dem Satz 4.14 mit $\lambda_0 = 1$ erfüllt. Dann ist (x^*, u^*) die eindeutig bestimmte optimale Lösung der Aufgabe $(LQ)_\infty$ im Sinne der Definition 4.1.

Beweis. Kann nachgewiesen werden, dass für jeden in der Aufgabe $(LQ)_\infty$ beliebig gewählten zulässigen Steuerprozess $(x, u) \in \mathcal{Z}$ die Ungleichung

$$J(x^*, u^*) - J(x, u) \leq 0$$

gilt, so ist die Optimalität von $(x^*, u^*) \in \mathcal{Z}$ bewiesen. Dazu betrachten wir zunächst die punktweise maximierte Pontryaginsche Funktion, welche auch Hamiltonfunktion genannt wird. Entsprechend der in dem Satz 4.14 definierten Pontryaginschen Funktion ist diese durch

$$\mathcal{H}(t, \xi, \eta, 1) := \max_{v \in \mathbb{R}^m} H(t, \xi, v, \eta, 1)$$

gegeben. Die Pontryaginsche Funktion (mit $\lambda_0 = 1$)

$$H(t, \xi, v, \eta, 1) = - \left(\frac{1}{2} \{ \xi^T W(t) \xi + v^T R(t) v \} + w^T(t) \xi + r^T(t) v \right) \omega_\beta(t) \\ + \eta^T (A(t) \xi + B(t) v + C(t))$$

ist als Summe des negativen in (ξ, v) linear-quadratischen Integranden aus (4.2) multipliziert mit dem Gewicht ω_β und der in (ξ, v) linearen rechten Seite der Dynamik aus (4.3) offensichtlich konkav in $(\xi, v) \in \mathbb{R}^n \times \mathbb{R}^m$. Daraus können wir leicht auf die Konkavität der Hamiltonfunktion $\mathcal{H}(t, \xi, \eta, 1)$ in $\xi \in \mathbb{R}^n$ schließen: Dafür sei

$$v_1 = \arg \max_{v \in \mathbb{R}^m} H(t, \xi_1, v, \eta, 1), \quad v_2 = \arg \max_{v \in \mathbb{R}^m} H(t, \xi_2, v, \eta, 1)$$

mit $\xi_1, \xi_2 \in \mathbb{R}^n$ und $\alpha \in [0, 1]$. Dann gilt

$$\begin{aligned} & \alpha \mathcal{H}(t, \xi_1, \eta, 1) + (1 - \alpha) \mathcal{H}(t, \xi_2, \eta, 1) \\ &= \alpha \max_{v \in \mathbb{R}^m} H(t, \xi_1, v, \eta, 1) + (1 - \alpha) \max_{v \in \mathbb{R}^m} H(t, \xi_2, v, \eta, 1) \\ &= \alpha H(t, \xi_1, v_1, \eta, 1) + (1 - \alpha) H(t, \xi_2, v_2, \eta, 1) \\ &\leq H(t, \alpha \xi_1 + (1 - \alpha) \xi_2, \alpha v_1 + (1 - \alpha) v_2, \eta, 1) \\ &\leq \max_{v \in \mathbb{R}^m} H(t, \alpha \xi_1 + (1 - \alpha) \xi_2, v, \eta, 1) = \mathcal{H}(t, \alpha \xi_1 + (1 - \alpha) \xi_2, \eta, 1), \end{aligned}$$

womit die Konkavität von $\mathcal{H}(t, \xi, \eta, 1)$ in $\xi \in \mathbb{R}^n$ für alle $(t, \eta) \in \mathbb{R}_+ \times \mathbb{R}^n$ gezeigt ist. Wir können nun zu der Differenz $J(x^*, u^*) - J(x, u)$ zurückkehren und schätzen ab:

$$\begin{aligned} & J(x^*, u^*) - J(x, u) \\ &= \int_0^\infty f(t, x^*(t), u^*(t)) \omega_\beta(t) - f(t, x(t), u(t)) \omega_\beta(t) dt \\ &= \int_0^\infty y_0^T(t) (A(t)x^*(t) + B(t)u^*(t) + C(t)) - H(t, x^*(t), u^*(t), y_0(t), 1) \\ &\quad - y_0^T(t) (A(t)x(t) + B(t)u(t) + C(t)) + H(t, x(t), u(t), y_0(t), 1) dt \\ &= \int_0^\infty H(t, x(t), u(t), y_0(t), 1) - H(t, x^*(t), u^*(t), y_0(t), 1) - y_0^T(t) (\dot{x}(t) - \dot{x}^*(t)) dt \\ &\leq \int_0^\infty \mathcal{H}(t, x(t), y_0(t), 1) - \mathcal{H}(t, x^*(t), y_0(t), 1) - y_0^T(t) (\dot{x}(t) - \dot{x}^*(t)) dt \end{aligned}$$

Dabei haben wir sowohl die Definition der Pontryaginschen Funktion und der Hamiltonfunktion genutzt, als auch verwendet, dass $(x^*, u^*) \in \mathcal{Z}$ der Dynamik in (4.3) genügt. Die

Konkavität der Hamiltonfunktion in dem Argument ξ führt weiter auf

$$\begin{aligned} J(x^*, u^*) - J(x, u) &\leq \int_0^\infty \nabla_\xi^T \mathcal{H}(t, x^*(t), y_0(t), 1) (x(t) - x^*(t)) - y_0^T(t) (\dot{x}(t) - \dot{x}^*(t)) dt \\ &= \int_0^\infty \nabla_\xi^T H(t, x^*(t), u^*(t), y_0(t), 1) (x(t) - x^*(t)) - y_0^T(t) (\dot{x}(t) - \dot{x}^*(t)) dt \end{aligned}$$

und zusammen mit der Transversalitätsbedingung (T), der adjungierten Gleichung (K) und der Anfangsbedingung $x^*(0) = x(0)$ ergibt sich schließlich

$$\begin{aligned} J(x, u) - J(x^*, u^*) &\leq \int_0^\infty -\dot{y}_0^T(t)(x(t) - x^*(t)) - y_0^T(t)(\dot{x}(t) - \dot{x}^*(t)) dt \\ &= \int_0^\infty -\frac{d}{dt}[y_0^T(t)(x(t) - x^*(t))] dt \\ &= -\lim_{T \rightarrow \infty} y_0^T(T)[x(T) - x^*(T)] + y_0^T(0) \underbrace{[x(0) - x^*(0)]}_{=0} \\ &= -\lim_{T \rightarrow \infty} y_0^T(T)x(T) + \lim_{T \rightarrow \infty} y_0^T(T)x^*(T). \end{aligned} \quad (4.39)$$

Da $x, x^* \in W_2^{1,n}((0, \infty), \omega_\beta)$ und $y_0 \in W_2^{1,n}((0, \infty), \omega_\beta^{-1})$ gilt, folgt $y_0^T x \in W_1^{1,1}((0, \infty))$ und $y_0^T x^* \in W_1^{1,1}((0, \infty))$. Daraus wiederum ergibt sich, dass die Grenzwerte in der Zeile (4.39) identisch Null¹⁰ sind, womit das gewünschte Resultat gezeigt ist. Die Eindeutigkeit der Lösung folgt aus dem Satz 4.13. \square

Aufgrund des vorangegangenen Satzes erhalten wir hinreichende Optimalitätsbedingungen für die Aufgabe $(LQ)_\infty$. Wir werden diese nutzen, um die eindeutig bestimmte, im Sinne der Definition 4.1 und für die Aufgabe $(LQ)_\infty$ optimale Lösung $(x^*, u^*) \in \mathcal{Z}$ zu ermitteln. Dabei gehen wir wie folgt vor:

Unter Gültigkeit der Voraussetzungen (V1)-(V3) und (V5)-(V6) studieren wir die Aufgabe $(LQ)_\infty$ mit $U = \mathbb{R}^m$. Dann liefert uns der Satz 4.13 die Existenz der eindeutig bestimmten optimalen Lösung $(x^*, u^*) \in \mathcal{Z}$ und das Maximumprinzip garantiert uns zusätzlich die Existenz von Multiplikatoren (λ_0, y_0) mit $y_0 \in W_2^{1,n}((0, \infty), \omega_\beta^{-1})$ und $\lambda_0 = 1$. Zusammen mit der adjungierten Funktion y_0 genügt das optimale Paar (x^*, u^*) der Maximumbedingung (M) und der kanonischen Gleichung (K). Aufgrund der Konkavität der Pontryaginschen Funktion in $v \in \mathbb{R}^m$ ist die Bedingung

$$\nabla_v H(t, x^*(t), u^*(t), y_0(t), 1) = 0 \quad \text{f.ü. auf } (0, \infty) \quad (4.40)$$

mit

$$\nabla_v H(t, \xi, v, \eta, 1) = (-R^T(t)v - r(t))e^{-\beta t} + B^T(t)\eta$$

¹⁰ vgl. Bemerkung 5.1

notwendig und hinreichend für die optimale Steuerung $u^* \in L_2((0, \infty), \omega_\beta)$. Direktes Ausrechnen liefert

$$u^*(t) = R^{-1}(t) \left[B^T(t)y_0(t)e^{\beta t} - r(t) \right] \quad \text{f.ü. auf } (0, \infty), \quad (4.41)$$

wobei die Existenz der inversen Matrix $R^{-1}(t), t \in \mathbb{R}_+$ aufgrund der Voraussetzung (V3) gesichert ist. Setzen wir die Beziehung aus (4.41) in die kanonische Gleichung (K) und in die dynamische Zustandsgleichung (4.3) ein, so ergibt sich zusammen mit

$$\nabla_\xi H(t, \xi, v, \eta, 1) = -(W(t)\xi + w(t))e^{-\beta t} + A^T(t)\eta$$

das Differentialgleichungssystem

$$(S1) \begin{cases} \dot{x}^*(t) &= A(t)x^*(t) + B(t)R^{-1}(t) \left[B^T(t)y_0(t)e^{\beta t} - r(t) \right] + C(t) & \text{f.ü. auf } (0, \infty), \\ \dot{y}_0(t) &= (W(t)x^*(t) + w(t))e^{-\beta t} - A^T(t)y_0(t) & \text{f.ü. auf } (0, \infty). \end{cases}$$

Mit der Substitution $y := y_0\omega_\beta^{-1}$ können wir dieses in das folgende System überführen:

$$(S2) \begin{cases} \dot{x}^*(t) &= A(t)x^*(t) + B(t)R^{-1}(t) \left[B^T(t)y(t) - r(t) \right] + C(t) & \text{f.ü. auf } (0, \infty), \\ \dot{y}(t) &= W(t)x^*(t) + w(t) + (\beta \cdot I_{n \times n} - A^T(t))y(t) & \text{f.ü. auf } (0, \infty). \end{cases}$$

Aufgrund von

$$\int_0^\infty |y(t)|_2^2 \omega_\beta(t) dt = \int_0^\infty |y_0(t) e^{\beta t}|_2^2 e^{-\beta t} dt = \int_0^\infty |y_0(t)|_2^2 e^{\beta t} dt < \infty, \quad (4.42)$$

$$\begin{aligned} \int_0^\infty |\dot{y}(t)|_2^2 \omega_\beta(t) dt &= \int_0^\infty |\dot{y}_0(t) e^{\beta t} + \beta y_0(t) e^{\beta t}|_2^2 e^{-\beta t} dt = \int_0^\infty |\dot{y}_0(t) + \beta y_0(t)|_2^2 e^{\beta t} dt \\ &\leq 2 \int_0^\infty \left[|\dot{y}_0(t)|_2^2 + \beta^2 |y_0(t)|_2^2 \right] e^{\beta t} dt < \infty \end{aligned} \quad (4.43)$$

gilt offenbar $y \in W_2^{1,n}((0, \infty), \omega_\beta)$. Im Umkehrschluss bedeutet das: Das System (S2) mit $x^*(0) = x_0$ ist in $W_2^{1,n}((0, \infty), \omega_\beta) \times W_2^{1,n}((0, \infty), \omega_\beta)$ lösbar. Finden wir eine solche Lösung $(x^*, y) \in W_2^{1,n}((0, \infty), \omega_\beta) \times W_2^{1,n}((0, \infty), \omega_\beta)$ von (S2) mit $x^*(0) = x_0$, so ist $y_0 = y\omega_\beta$ eine Funktion des Raumes $W_2^{1,n}((0, \infty), \omega_\beta^{-1})$. Dies kann analog zu dem Vorgehen in (4.42)-(4.43) gezeigt werden. Das Paar $(x^*, y_0) \in W_2^{1,n}((0, \infty), \omega_\beta) \times W_2^{1,n}((0, \infty), \omega_\beta^{-1})$ löst dann die Gleichungen in (S1). Mit Hilfe von y_0 berechnen wir die Steuerung $u^* \in L_2((0, \infty), \omega_\beta)$ aus der Vorschrift (4.41). Offensichtlich haben wir damit ein zulässiges Paar $(x^*, u^*) \in \mathcal{Z}$ und eine adjungierte Funktion $y_0 \in W_2^{1,n}((0, \infty), \omega_\beta^{-1})$ gefunden, welche die Bedingungen (M) und (K) aus dem Satz 4.14 mit $\lambda_0 = 1$ erfüllen. Aus dem Satz 4.15 folgt, dass dieses Paar $(x^*, u^*) \in \mathcal{Z}$ auch die eindeutig bestimmte Lösung der Steuerungsaufgabe $(LQ)_\infty$ ist.

Bemerkung 4.5. Aus den Bedingungen des Maximumprinzips lassen sich weitere Informa-

tionen über Eigenschaften der Lösung $(x^*, u^*) \in \mathcal{Z}$ und der Adjungierten $y \in W_2^{1,n}((0, \infty), \omega_\beta)$ ableiten.

1. Wegen $U = \mathbb{R}^m$ ist die Maximumbedingung (M) zu der Bedingung in (4.40) äquivalent und direktes Auflösen nach u^* führt auf die Vorschrift in (4.41). Aus $y_0 \in W_2^{1,n}((0, \infty), \omega_\beta^{-1})$ folgt $y := y_0 \cdot \omega_\beta^{-1} \in W_2^{1,n}((0, \infty), \omega_\beta)$ und der Satz von Rellich-Kondrachov liefert die Existenz eines stetigen Repräsentanten $\tilde{y} \in C([0, \infty))$, wodurch zusätzlich die Formulierung des Anfangswertes $y(0)$ legitimiert ist, vgl. Bemerkung 2.3. Gemeinsam mit der Voraussetzung (V1) erhalten wir dann aufgrund von (4.41), dass die optimale Steuerung u^* der Aufgabe $(LQ)_\infty$ eine auf $[0, \infty)$ stetige Funktion ist.
2. Es lässt sich zeigen, dass x^* und y_0 durch eine Matrixfunktion P gekoppelt sind, d.h.

$$y_0(t) = P(t)x^*(t) \quad \text{f.ü. auf } (0, \infty),$$

und P durch Lösen einer Matrix-Riccati-Differentialgleichung bestimmt werden kann, vgl. [ASS15]. Setzt man diese Beziehung in (4.41) ein, so erhält man ein Gesetz, mit welchem die optimale Steuerung u^* zur Zeit t direkt aus dem optimalen Zustand $x^*(t)$ berechnet werden kann. Die optimale Steuerung u^* der Aufgabe $(LQ)_\infty$ ist demnach eine sogenannte Rückkopplungssteuerung (feed back control).

3. Unter weiteren Voraussetzungen an die Eingangsdaten lassen sich bessere Regularitätseigenschaften für die Lösung $(x^*, y) \in W_2^{1,n}((0, \infty), \omega_\beta) \times W_2^{1,n}((0, \infty), \omega_\beta)$ des kanonischen Differentialgleichungssystems (S2) gewinnen. So können wir aus den Gleichungen in (S2) auf $(x^*, y) \in W_2^{2,n}((0, \infty), \omega_\beta) \times W_2^{2,n}((0, \infty), \omega_\beta)$ schließen, falls die Funktionen W, R, w, r, A, B, C zusätzlich stetig differenzierbar auf $[0, \infty)$ sind, vgl. [PB19], Lemma 5.

Zusammengefasst möchten wir festhalten, dass sich das Lösen der Steuerungsaufgabe $(LQ)_\infty$ darauf reduzieren lässt, ein Paar $(x, y) \in W_2^{1,n}((0, \infty), \omega_\beta) \times W_2^{1,n}((0, \infty), \omega_\beta)$ mit $x(0) = x_0$ zu finden, welches das System

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} A(t) & B(t)R^{-1}(t)B^T(t) \\ W(t) & \beta \cdot I_{n \times n} - A^T(t) \end{pmatrix} \cdot \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} + \begin{pmatrix} C(t) - B(t)R^{-1}(t)r(t) \\ w(t) \end{pmatrix}$$

fast überall auf $(0, \infty)$ löst. Offen bisher bleibt, wie der Anfangswert der Adjungierten zu bestimmen ist. Wir werden in dem nächsten Kapitel sehen, dass die Transversalitätsbedingung (T) aus dem Satz 4.14, welche uns in Form einer Raumzugehörigkeit zur Verfügung steht, eine entscheidende Rolle dabei spielt und wir daher ein wohlformuliertes Anfangswertproblem erhalten. Dies motiviert uns im Folgenden den Einsatz von Spektralmethoden zu studieren, um das obige Differentialgleichungssystem numerisch zu lösen.

5 Indirekte Spektralmethode zur Lösung linear-quadratischer Steuerungsaufgaben mit unendlichem Zeithorizont

Aufgaben optimaler Steuerung stellen eine häufig sehr komplexe Klasse von Optimierungsaufgaben dar. Im Vergleich zu statischen Optimierungsaufgaben, deren Optimierungsvariablen Elemente des euklidischen Raumes sind, werden Optimalsteuerungsaufgaben in unendlich-dimensionalen Funktionenräumen diskutiert und lassen sich daher als unendlich-dimensionale Optimierungsaufgaben auffassen. Darüber hinaus beinhaltet die Problembeschreibung ein System von Differentialgleichungen, welches die Dynamik des zugrunde liegenden Systems beschreibt. Diese Komplexität optimaler Steuerungsaufgaben führt gewöhnlich dazu, dass diese Aufgaben nur numerisch zu lösen sind. Es existiert eine Vielzahl von Ansätzen, Methoden und Software zur numerischen Lösung von Optimalsteuerungsproblemen. Einen Überblick findet man beispielsweise in [Bet98, Rao09, Sac92, SB92, Str94] und den darin angegebenen Quellen. Allgemein lassen sich die verwendeten numerischen Methoden in die Gruppe der *direkten* und der *indirekten* Verfahren einteilen. In Anlehnung an die genannte Literatur lassen sich diese wie folgt klassifizieren:

Die indirekten Verfahren bauen auf der Verwendung notwendiger Optimalitätsbedingungen auf. Unter der Annahme bestimmter Voraussetzungen liefert das Pontryaginsche Maximumprinzip ein Differentialgleichungssystem, welches aus den Zustands- und Adjungiertendifferentialgleichungen gebildet wird. In einigen Spezialfällen, wie zum Beispiel in dem Fall der linear-quadratischen Steuerungsaufgabe $(LQ)_\infty$, ist das Lösen dieses Differentialgleichungssystems unter Beachtung der Anfangsbedingung an den Zustand und bei gleichzeitiger Ausnutzung der Transversalitätsbedingung äquivalent zur Lösung der Steuerungsaufgabe selbst. Entsprechend kann von der Lösung der ursprünglichen Steuerungsaufgabe zur Lösung einer aus den notwendigen Bedingungen resultierenden Randwertaufgabe übergegangen werden. In Abhängigkeit der zugrunde liegenden Struktur der Differentialgleichungen können verschiedene numerische Verfahren für Randwertaufgaben, welche sowohl Verfahren zur Lösung von Anfangswertproblemen als auch Nullstellenverfahren beinhalten, herangezogen werden. Entsprechend werden in der Literatur zahlreiche numerische Lösungsmethoden für Randwertprobleme angeboten. Zu nennen sind hier neben den Integrationsmethoden, die Finite-Differenzen-Methode (FDM), die Finite-Elemente-Methode (FEM), aber auch die Kollokationsmethoden, vgl. z.B. [AMR95]. Ausgehend von dem einfachen Schießverfahren (single shooting) entwickelte sich die sehr robuste und zuverlässige Mehrzielmethode, auch Mehrfachschießverfahren (multiple shooting) genannt, vgl. [Bul71, Kel68, Osb69]. Die erfolgreiche Anwendung der Mehrzielmethode im Kontext der optimalen Steuerung spiegelt sich sowohl in der hohen Anzahl von Publikationen zu diesem Thema wider, vgl. z.B. [BP84, Mau76, MG75, MP95, Obe86, PR78], wird aber auch durch eine Reihe von entwickelter Software, w.z.B. die leistungsfähige Fortran-Routine BNDSO [OG89], belegt. Um einen

ersten Einblick in die Grundidee der indirekten Verfahren zu bekommen, verweisen wir den Leser auf [FH86], S. 487 ff. Dort wird anhand eines einfachen Steuerungsproblems über einen fest vorgegebenen endlichen Zeithorizont und mit einer gewöhnlichen Zustandsdifferentialgleichung beschrieben, wie man die Theorie des Pontryaginschen Maximumprinzips nutzen und diese numerisch mit dem Einsatz von Schießverfahren, der Methode der konjugierten Gradienten und eines Kollokationsverfahren zur Lösung der Steuerungsaufgabe verwenden kann.

Die direkten Verfahren verfolgen einen anderen Grundgedanken: Hier wird das ursprüngliche Steuerungsproblem durch Diskretisierung der Steuer- und/oder Zustandsvariablen in ein endlichdimensionales Optimierungsproblem überführt, welches sich dann mit Verfahren der nichtlinearen Optimierung (z.B. SQP-Verfahren) lösen lässt. In Anlehnung an [Str94] lassen sich die direkten Verfahren in die direkten Schießverfahren und die direkten Kollokationsverfahren unterteilen. Die direkten Schießverfahren sind dadurch gekennzeichnet, dass sie lediglich auf einer Diskretisierung der Steuerung basieren und der Zustand durch numerische Integration der Differentialgleichung berechnet wird. Bei den direkten Kollokationsverfahren wird zusätzlich zu der Steuerung auch die Zustandsvariable diskretisiert. Die Erfüllung der Differentialgleichung wird an vorgegebenen Kollokationspunkten gefordert. Auch bei den direkten Verfahren wurden zahlreiche Softwarepakete zur numerischen Lösung von Optimalsteuerungsproblemen entwickelt. So wird die Idee des direkten Schießverfahrens zum Beispiel in dem Fortran-Programmpaket OCPID-DAE, vgl. [Ger13], umgesetzt und durch DIRCOL, vgl. [Str99], ein Programmpaket basierend auf der Methode der direkten Kollokation zur Verfügung gestellt.

In Abhängigkeit davon, ob nur die Steuerung oder aber auch der Zustand diskretisiert wird, spricht man von einem reduzierten bzw. vollständigen Diskretisierungsansatz, vgl. z.B. [Ger11]. Wählt man einen reduzierten Ansatz, wird der Zustand als Optimierungsvariable eliminiert. Bei der vollständigen Diskretisierung werden dem Optimierer sowohl die Steuerung als auch der Zustand als Optimierungsvariable übergeben. Der Vorteil des vollständigen Ansatzes besteht darin, dass auch reine Zustandsbeschränkungen und gemischte Steuerungs-Zustandsbeschränkungen berücksichtigt werden können. Im Vergleich zu einem reduzierten Ansatz entstehen jedoch mehr Optimierungsvariablen, weshalb die Dimension des resultierenden Optimierungsproblem bei einer vollständigen Diskretisierung sehr groß werden kann.

Sowohl die indirekten als auch die direkten Verfahren beinhalten eine Diskretisierung und eine Auswertung von Optimalitätsbedingungen. Lediglich die Reihenfolge ist verschieden. Bei den indirekten Methoden findet zunächst die Auswertung der notwendigen Optimalitätsbedingungen statt, anschließend werden diese diskretisiert. Dies vertritt den Standpunkt „First optimize, then discretize“. Im Gegensatz dazu verfolgen die direkten Verfahren die Strategie „First discretize, then optimize“. Dabei ergibt sich aus der Sicht des Anwenders der Vorteil, dass dieser keine Vorkenntnisse aus der Theorie der Optimalsteuerung benötigt, d.h. insbesondere auch kein Wissen zur Formulierung der adjungierten Differentialgleichung für die Anwendung der Methode notwendig ist. Im Vergleich zu den indirekten Verfahren erfordern die direkten Verfahren daher weniger Vorarbeit und werden bevorzugt verwendet. Dennoch können aus der Parametrisierung der Steuerung bzw. der Steuerung und des Zustandes erhebliche Genauigkeitsverluste resultieren. Hier weisen die indirekten Verfahren

einen Vorteil auf. Durch die Berechnung der adjungierten Variablen ist eine Überprüfung der notwendigen Optimalitätsbedingungen möglich, was dazu führt, dass nichtoptimale Lösungskandidaten aussortiert werden können und sich daher bei der Verwendung indirekter Verfahren Lösungen mit hoher Genauigkeit erzielen lassen. Voraussetzung für die Konvergenz eines entsprechenden Verfahrens ist hierbei ein guter Startwert für die Adjungierte.

In der vorliegenden Arbeit beschäftigen wir uns mit einem indirekten Lösungsverfahren. Grundlage dafür bilden die vorangegangenen Kapitel, in denen wir sowohl die Existenz als auch die Eindeutigkeit einer Lösung der Aufgabe $(LQ)_\infty$ untersucht und entsprechend das Pontryaginsche Maximumprinzip studiert haben. Wir werden sehen, dass wir damit das theoretische Werkzeug parat haben, welches benötigt wird, um eine indirekte Methode zur numerischen Lösung von $(LQ)_\infty$ heranzuziehen. Konkret präsentieren wir eine indirekte Kollokationsmethode, wobei wir die gesuchten Funktionen durch Ansätze basierend auf orthogonalen Polynomsystemen approximieren. Die Wahl globaler Ansatzfunktionen führt zu der Klasse der Spektralmethoden, welche sich allgemein in die Familie der gewichteten Residuen einordnen lassen. In Anlehnung an [STW11] möchten wir zunächst deren formale Grundidee erläutern.

Ausgangspunkt ist eine in dem Gebiet Ω gestellte allgemeine Differentialgleichung

$$\mathcal{L}\{x(t)\} = f(t), \quad t \in \Omega, \quad (5.1)$$

wobei \mathcal{L} einen beliebigen Differentialoperator darstellt. Die gesuchte Lösung x wird nun mit Hilfe eines Ansatzes der Form

$$x(t) \approx x_N(t) = \sum_{k=1}^N a_k \phi_k(t) \quad (5.2)$$

approximiert. Dabei sind ϕ_k linear unabhängige Ansatzfunktionen (auch Basisfunktionen genannt) und a_k die noch unbekannt, zu berechnenden Koeffizienten. Letztere sollen derart bestimmt werden, sodass die exakte Lösung x möglichst gut von der Näherungslösung x_N approximiert wird. Einsetzen von x_N in (5.1) führt zu einem lokalen Fehler, dem sogenannten Residuum:

$$\mathbf{R}_N(t) := \mathcal{L}\{x_N(t)\} - f(t) \neq 0, \quad t \in \Omega. \quad (5.3)$$

Bei dem Verfahren der gewichteten Residuen muss dieser Fehler für die Näherungslösung x_N nicht für alle $t \in \Omega$ verschwinden. Vielmehr wird dieser mit sogenannten Testfunktionen ψ_j multipliziert, anschließend über Ω integriert und dadurch gefordert, dass die Differentialgleichung in dem Gebiet Ω nur noch im Mittel erfüllt wird:

$$\langle \mathbf{R}_N, \psi_j \rangle_\Omega := \int_{\Omega} \mathbf{R}_N \psi_j \, d\Omega \stackrel{!}{=} 0, \quad 0 \leq j \leq N, \quad (5.4)$$

Das auftauchende Integral definiert ein (gewichtetes) Funktionenskalarprodukt $\langle \cdot, \cdot \rangle_\Omega$ für auf dem Gebiet Ω definierte Lebesgue-integrierbare Funktionen. Damit fordert die Gleichung (5.4), dass der Fehler der approximierten Differentialgleichung orthogonal zu allen Testfunktionen ist. Da sich die Testfunktionen auch als Gewichtung für das Residuum interpretieren

lassen, wird diese Methode auch als Verfahren der gewichteten Residuen bezeichnet. Durch die Gewichtung, das heißt durch den Einsatz der Testfunktionen, entstehen dabei mehr Möglichkeiten, die obige Gleichung zu erfüllen. Darüber hinaus stellt die Bedingung (5.4) weniger starke Anforderungen an die Näherungslösung x_N als die punktweise Forderung in (5.3), vgl. [Wag17], S. 52 f.

Die Wahl der Ansatzfunktionen in (5.2) und der Testfunktionen in (5.4) wird dabei in Abhängigkeit der gegebenen Differentialgleichung und möglicherweise vorhandenen Randbedingungen getroffen, um sicherzustellen, dass diese bestmöglichst von u_N erfüllt werden. Im Allgemeinen lassen sich dabei die folgenden Methoden unterscheiden, vgl. [CHQZ06, STW11]:

- Galerkin-Verfahren: Die Ansatz- und Testfunktionen sind identisch, das heißt, es gilt $\phi_k = \psi_j$.
- Kollokationsverfahren: Hier wird gefordert, dass die Näherungslösung x_N die Differentialgleichung (5.1) in vorgegebenen Kollokationspunkten $\{t_j\}$ exakt erfüllt. Demnach wird gefordert, dass das Residuum in diesen Punkten verschwindet, $\mathbf{R}_N(t_j) = 0$. Durch die Wahl von Diracschen δ -Funktionen¹ (Distributionen) für die Testfunktionen, $\psi_j(t) = \delta(t - t_j)$, lässt sich das Kollokationsverfahren aus der Gleichung (5.4) herleiten und wird deshalb zu den Verfahren der gewichteten Residuen gezählt.²
- Petrov-Galerkin-Verfahren: Die Ansatz- und Testfunktionen sind verschieden, das heißt, es gilt $\phi_k \neq \psi_j$.

Die Wahl der Testfunktionen ist eines der Hauptmerkmale, wodurch sich Spektralmethoden von FEM und FDM unterscheiden lassen. Bei den Spektralmethoden kommen Testfunktionen zum Einsatz, die globale Glattheitseigenschaften aufweisen, das heißt global unendlich oft differenzierbar sind. Häufig werden dafür trigonometrische Funktionen oder orthogonale Polynome, welche oftmals Eigenfunktionen einer Sturm-Liouville-Eigenwertaufgabe sind, gewählt. Bei FEM und FDM kommen hingegen Testfunktionen mit lokalem Charakter und endlichen Regularitäten zum Einsatz, vgl. [CHQZ06], S. 3. f. und [STW11], S.3 f.

In der Literatur findet man oftmals auch den Begriff einer Pseudospektralmethode, wobei hier jedoch unterschiedliche Interpretationen verwendet werden. Wir orientieren uns an [Fun92], S. 184 und fassen eine Pseudospektralmethode als eine Spektralmethode auf, bei denen globale Kollokationsansätze genutzt werden.

Auch wenn Pseudospektralmethoden ursprünglich vor allem in der Strömungsmechanik zum Einsatz kamen, vgl. [CHQZ06], existiert heute eine große Anzahl von Anwendungen zur numerischen Lösung von Optimalsteuerungsaufgaben. In [Vla88, VD88] wurden direkte Lösungsmethoden basierend auf einer Entwicklung nach Tschebyscheff-Polynomen vorgestellt. Es folgten zahlreiche Arbeiten, in denen Lagrange-Polynome für die Erzeugung polynomialer Ansatzfunktionen verwendet wurden, vgl. beispielsweise [CB89, EKR95, FR00, Wil04, BHTR06]. Ein Vorteil bei der Wahl einer Lagrange-Basis resultiert aus ihrer Interpolationseigenschaft. Diese führt dazu, dass die gesuchten Entwicklungskoeffizienten mit den Werten der Näherungslösung an den Stützstellen übereinstimmen. In den Publikationen

¹ Für die Definition der Dirac- δ -Funktion und deren Eigenschaften siehe z.B. [MV01].

² vgl. z.B. [Bet04], S. 177 f.

[FR08, GPF⁺11, GHR11] werden Pseudospektralmethoden zur Lösung von Steuerungsaufgaben über dem unendlichen Zeithorizont $[0, \infty)$ präsentiert. Die numerischen Methoden basieren dabei auf einer Transformation des unendlichen Zeithorizontes auf den endlichen Horizont $[-1, 1]$. Mit unserer Methode möchten wir eine Möglichkeit präsentieren, bei der auf eine solche Transformation verzichtet werden kann.

Mit diesem Einblick möchten wir unsere Einführung in die verschiedenen numerischen Verfahren abschließen und verweisen für einen detaillierteren Einblick in die Theorie der Spektralmethoden an [CHQZ06], [Fun92] und [STW11]. Alle Quellen beinhalten neben der Darstellung grundlegender Eigenschaften orthogonaler Polynomsysteme entsprechende Resultate aus der Approximationstheorie. Speziell wird die Konstruktion spektraler Methoden und deren Anwendungen anhand verschiedener Problemklassen diskutiert und herausgearbeitet.

Für die Konstruktion der, auf den Pontryaginschen Maximumprinzip basierenden, indirekten Spektralmethode für Steuerungsaufgaben der Form $(LQ)_\infty$ verwenden wir verallgemeinerte Laguerre-Polynome zur Konstruktion von Näherungslösungen. In dem nächsten Abschnitt stellen wir daher zunächst deren grundlegenden Eigenschaften vor und zeigen relevante Resultate aus der Approximationstheorie auf.

5.1 Verallgemeinerte Laguerre-Polynome

Eine bedeutende Rolle bei der Anwendung und Umsetzung einer Spektralmethode kommt den orthogonalen Polynomsystemen zu. In Abhängigkeit des Grundintervalls und der gewählten Gewichtsfunktion ergeben sich verschiedene Orthogonalpolynome. Die Laguerre-Polynome $\mathcal{L}_l^{(\alpha)}$, $l \in \mathbb{N}_0$ bilden ein orthogonales Polynomsystem in dem Lebesgue-Raum $L_2((0, \infty), \omega_\alpha)$ mit

$$\omega_\alpha(t) = t^\alpha e^{-t}, \quad \alpha > -1, \quad t \in (0, \infty) \quad (5.5)$$

und gehören neben den Jacobi- und den Hermite-Polynomen zu den klassischen Orthogonalpolynomen, vgl. [Sze39], S. 29. In dem Abschnitt 2.4.1 haben wir Gewichtsfunktionen der obigen Gestalt untersucht, dabei jedoch einen natürlichen Exponenten $\alpha \in \mathbb{N}_0$ vorausgesetzt. Die Laguerre-Gewichtsfunktion ω_α in (5.5) beinhaltet reelle Exponenten $\alpha > -1$, wobei die Momente

$$\mu_k = \int_0^\infty t^k \omega_\alpha(t) dt = \int_0^\infty t^{\alpha+k} e^{-t} dt \quad (k = 0, 1, 2, \dots)$$

als Lebesgue-Integrale existieren (vgl. [Els11], S. 154) und insbesondere $\mu_k = \Gamma(\alpha + k + 1)$ gilt, wobei $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ mit $z \in \mathbb{R}$, $z > 0$ die Eulersche Gammafunktion meint.³ Es ist $\omega_\alpha \in L_1((0, \infty)) \cap C((0, \infty))$ und $\omega_\alpha(t) > 0$, $t \in (0, \infty)$.

Es existiert eine Vielzahl an Literatur, welche sich mit dem Studium der genannten klassischen Polynomsysteme befasst. Dem interessierten Leser möchte wir an dieser Stelle die

³ Einen Überblick über die grundlegenden Eigenschaften der Gammafunktion findet man in [Fun92], S. 2 ff., [Sze39], S. 14 ff., [Tri70], S. 143 ff.

Arbeiten [Fun92, Sze39, Tri70, Fre69] empfehlen. In der Formulierung unserer Steuerungsaufgabe $(LQ)_\infty$ tritt die Gewichtsfunktion $\omega_\beta(t) = e^{-\beta t}$, $\beta > 0$ auf. Wir müssen den zusätzlich auftauchenden Parameter β bei der Wahl des Polynomsystems berücksichtigen. In natürlicher Weise ist es demnach sinnvoll, einer Spektralmethode zur Lösung von $(LQ)_\infty$ ein auf $(0, \infty)$ definiertes Polynomsystem zugrunde zu legen, welches orthogonal bezüglich der allgemeineren Gewichtsfunktion

$$\omega_{\alpha,\beta}(t) = t^\alpha e^{-\beta t}, \quad \alpha > -1, \beta > 0, t \in (0, \infty) \quad (5.6)$$

ist. Dies führt uns automatisch zu der Klasse der *verallgemeinerten Laguerre-Polynome* $\mathcal{L}_l^{(\alpha,\beta)}$, welche ein auf $(0, \infty)$ definiertes und bezüglich der Gewichtsfunktion $\omega_{\alpha,\beta}$ in (5.6) orthogonales Polynomsystem bilden. In der Arbeit *A new generalized Laguerre spectral approximation and its applications* präsentierten Ben-yu und Xiao-yong das System der verallgemeinerten Laguerre-Polynome, welches durch die Einführung des zusätzlichen Parameters β in der Gewichtsfunktion als eine Verallgemeinerung der klassischen Laguerre-Polynome mit $\omega_\alpha(t) = \omega_{\alpha,1}(t) = t^\alpha e^{-t}$ aufgefasst werden kann, vgl. [ByXy05]. Neben grundlegenden Eigenschaften dieser Polynome stellen sie auch entsprechende Ergebnisse bezüglich der Orthogonalprojektion in dem Raum $L_2((0, \infty), \omega_{\alpha,\beta})$ vor. Wir greifen die für unsere Ausführungen wichtigen Resultate auf und geben diese in den folgenden Abschnitten zusammenfassend an.

5.1.1 Grundeigenschaften

Die verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(\alpha,\beta)}$ lassen sich mit Hilfe der expliziten Vorschrift

$$\mathcal{L}_l^{(\alpha,\beta)}(t) = \frac{1}{l!} t^{-\alpha} e^{\beta t} \frac{d^l}{dt^l} (t^{l+\alpha} e^{-\beta t}), \quad l = 0, 1, \dots \quad (t > 0) \quad (5.7)$$

berechnen, [ByXy05] und die ersten verallgemeinerten Laguerre-Polynome lauten demnach

$$\begin{aligned} \mathcal{L}_0^{(\alpha,\beta)}(t) &= 1, & \mathcal{L}_1^{(\alpha,\beta)}(t) &= -\beta t + \alpha + 1, \\ \mathcal{L}_2^{(\alpha,\beta)}(t) &= \frac{1}{2} \beta^2 t^2 - \beta(\alpha + 2)t + \frac{1}{2}(\alpha + 1)(\alpha + 2), \dots \end{aligned}$$

Wir werden zunächst zeigen, dass wir die Berechnungsformel in (5.7) mit Hilfe einer geeigneten Substitution der unabhängigen Variablen aus den klassischen Laguerre-Polynomen herleiten können. Das klassische Laguerre-Polynom $\mathcal{L}_l^{(\alpha)}$ vom Grad l erhält man aus der Rodrigues' Formel

$$\mathcal{L}_l^{(\alpha)}(\tau) = \frac{1}{l!} \tau^{-\alpha} e^\tau \frac{d^l}{d\tau^l} (\tau^{l+\alpha} e^{-\tau}), \quad \tau \in (0, \infty), \quad (5.8)$$

vgl. [Sze39], S. 101. Mit $\beta > 0$ definieren wir die folgenden stetig differenzierbaren Funktionen

$$\tau := g(t) = \beta \cdot t, \quad t := h(\tau) = \frac{\tau}{\beta} \quad (5.9)$$

und halten fest, dass offensichtlich $g(h(\tau)) = \tau$ und $g(h(t)) = t$ gelten. Mit der Hilfsfunktion

$$F(\tau) := \tau^{l+\alpha} \cdot e^{-\tau}$$

und durch Anwenden der Kettenregel

$$\frac{d}{dt}F(g(t)) = F'(g(t)) \cdot g'(t) = F'(g(t)) \cdot \beta \Rightarrow \frac{d^l}{dt^l}F(g(t)) = F^{(l)}(g(t)) \cdot \beta^l$$

erhalten wir sofort die Beziehung

$$F^{(l)}(g(t)) = \frac{1}{\beta^l} \cdot \frac{d^l}{dt^l}F(g(t)) \Leftrightarrow \frac{d^l}{d\tau^l}F(\tau) = \frac{1}{\beta^l} \cdot \frac{d^l}{dt^l}F(\beta t).$$

Anwenden der Variablentransformation aus (5.9) auf die Formel in (5.8) liefert

$$\begin{aligned} \mathcal{L}_l^{(\alpha)}(\tau) &= \mathcal{L}_l^{(\alpha)}(g(t)) = \frac{1}{l!} (\beta t)^{-\alpha} e^{\beta t} \frac{1}{\beta^l} \frac{d^l}{dt^l}F(\beta t) \\ &= \frac{1}{l!} t^{-\alpha} e^{\beta t} \frac{1}{\beta^{l+\alpha}} \frac{d^l}{dt^l} \left((\beta t)^{l+\alpha} \cdot e^{-\beta t} \right) \\ &= \mathcal{L}_l^{(\alpha, \beta)}(t). \end{aligned} \quad (5.10)$$

Es lassen sich sämtliche Eigenschaften der klassischen Laguerre-Polynome auf die verallgemeinerten Laguerre-Polynome übertragen. Wir konzentrieren uns auf die Folgenden:

Lemma 5.1

Die verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(\alpha, \beta)}$ bilden ein vollständiges Funktionensystem im $L_2((0, \infty), \omega_{\alpha, \beta})$ bezüglich der Gewichtsfunktion aus (5.6). Es gilt die Orthogonalitätsrelation

$$\langle \mathcal{L}_l^{(\alpha, \beta)}, \mathcal{L}_m^{(\alpha, \beta)} \rangle_{\omega_{\alpha, \beta}} := \int_0^\infty \mathcal{L}_l^{(\alpha, \beta)}(t) \mathcal{L}_m^{(\alpha, \beta)}(t) \omega_{\alpha, \beta}(t) dt = \begin{cases} \gamma_l^{(\alpha, \beta)} & l = m \\ 0 & l \neq m \end{cases} \quad (5.11)$$

mit

$$\gamma_l^{(\alpha, \beta)} = \frac{\Gamma(l + \alpha + 1)}{\beta^{\alpha+1} l!}. \quad (5.12)$$

Beweis. Wir beginnen mit dem Beweis der angegebenen Orthogonalitätsrelation. Für die

klassischen Laguerre-Polynome $\mathcal{L}_l^{(\alpha)}$ gilt

$$\int_0^{\infty} \mathcal{L}_l^{(\alpha)}(\tau) \mathcal{L}_m^{(\alpha)}(\tau) \omega_{\alpha}(\tau) d\tau = \begin{cases} \frac{\Gamma(l+\alpha+1)}{l!} & l = m \\ 0 & l \neq m \end{cases} \quad (5.13)$$

mit $\omega_{\alpha}(\tau) = \tau^{\alpha} e^{-\tau}$, vgl. [Tri70], S. 214. Mit der Substitution aus (5.9) erhalten wir $h'(\tau) d\tau = dt \Leftrightarrow d\tau = \beta dt$ und unter Beachtung von (5.10) ergibt sich

$$\begin{aligned} \int_0^{\infty} \mathcal{L}_l^{(\alpha)}(\tau) \mathcal{L}_m^{(\alpha)}(\tau) \tau^{\alpha} e^{-\tau} d\tau &= \beta^{\alpha+1} \int_0^{\infty} \mathcal{L}_l^{(\alpha)}(\beta t) \mathcal{L}_m^{(\alpha)}(\beta t) t^{\alpha} e^{-\beta t} dt \\ &= \beta^{\alpha+1} \int_0^{\infty} \mathcal{L}_l^{(\alpha,\beta)}(t) \mathcal{L}_m^{(\alpha,\beta)}(t) \omega_{\alpha,\beta}(t) dt. \end{aligned} \quad (5.14)$$

Setzen wir nun die Beziehung aus (5.13) in die obige Gleichung ein, so erhalten wir die gewünschte Aussage in (5.11). Auch für den Nachweis der Vollständigkeit nutzen wir die Transformation der unabhängigen Variablen. Ausgangspunkt ist die Vollständigkeit der klassischen Laguerre-Polynome $\mathcal{L}_l^{(\alpha)}$ bezüglich der Gewichtsfunktion in (5.5), vgl. [Tri70], S. 235. Aus der Rechnung in (5.14) erhalten wir

$$\|\mathcal{L}_l^{(\alpha)}\|_{2,\omega_{\alpha}}^2 = \langle \mathcal{L}_l^{(\alpha)}, \mathcal{L}_l^{(\alpha)} \rangle_{2,\omega_{\alpha}} = \beta^{\alpha+1} \langle \mathcal{L}_l^{(\alpha,\beta)}, \mathcal{L}_l^{(\alpha,\beta)} \rangle_{2,\omega_{\alpha,\beta}} = \beta^{\alpha+1} \|\mathcal{L}_l^{(\alpha,\beta)}\|_{2,\omega_{\alpha,\beta}}^2.$$

Mit denselben Überlegungen wie in (5.14) und der Substitution aus (5.9) erhalten wir für eine Funktion $f \in L_2((0, \infty), \omega_{\alpha,\beta})$ die nachfolgenden Rechnungen

$$\|f\|_{2,\omega_{\alpha,\beta}}^2 = \int_0^{\infty} |f(t)|^2 \omega_{\alpha,\beta}(t) dt = \frac{1}{\beta^{\alpha+1}} \int_0^{\infty} |\tilde{f}(\tau)|^2 \omega_{\alpha}(\tau) d\tau = \frac{1}{\beta^{\alpha+1}} \|\tilde{f}\|_{2,\omega_{\alpha}}^2,$$

wobei wir $\tilde{f}(\tau) := f(\frac{\tau}{\beta})$ gesetzt haben. Weiterhin gilt

$$\begin{aligned} \langle \tilde{f}, \frac{\mathcal{L}_l^{(\alpha)}}{\|\mathcal{L}_l^{(\alpha)}\|_{2,\omega_{\alpha}}} \rangle_{2,\omega_{\alpha}} &= \int_0^{\infty} \tilde{f}(\tau) \frac{1}{\|\mathcal{L}_l^{(\alpha)}\|_{2,\omega_{\alpha}}} \mathcal{L}_l^{(\alpha)}(\tau) \omega_{\alpha}(\tau) d\tau \\ &= \int_0^{\infty} f(t) \frac{1}{(\beta^{\alpha+1})^{1/2}} \frac{1}{\|\mathcal{L}_l^{(\alpha,\beta)}\|_{2,\omega_{\alpha,\beta}}} \mathcal{L}_l^{(\alpha,\beta)}(t) \beta^{\alpha} \omega_{\alpha,\beta}(t) \beta dt \\ &= (\beta^{\alpha+1})^{1/2} \langle f, \frac{\mathcal{L}_l^{(\alpha,\beta)}}{\|\mathcal{L}_l^{(\alpha,\beta)}\|_{2,\omega_{\alpha,\beta}}} \rangle_{2,\omega_{\alpha,\beta}}. \end{aligned}$$

Aufgrund der Vollständigkeit der klassischen Laguerre-Polynome genügen sie der Parseval'schen Gleichung für jede Funktion des $L_2((0, \infty), \omega_{\alpha})$, also insbesondere auch für $\tilde{f} \in L_2((0, \infty), \omega_{\alpha})$, vgl. Satz 3.13, (ii). Unter Verwendung der vorangegangenen Rechnungen

können wir daher weiter auf

$$\|f\|_{2,\omega_{\alpha,\beta}}^2 = \frac{1}{\beta^{\alpha+1}} \|\tilde{f}\|_{2,\omega_\alpha}^2 = \frac{1}{\beta^{\alpha+1}} \sum_{l=0}^{\infty} \left| \left\langle \tilde{f}, \frac{\mathcal{L}_l^{(\alpha)}}{\|\mathcal{L}_l^{(\alpha)}\|_{2,\omega_\alpha}} \right\rangle_{2,\omega_\alpha} \right|^2 = \sum_{l=0}^{\infty} \left| \left\langle f, \frac{\mathcal{L}_l^{(\alpha,\beta)}}{\|\mathcal{L}_l^{(\alpha,\beta)}\|_{2,\omega_{\alpha,\beta}}} \right\rangle_{2,\omega_{\alpha,\beta}} \right|^2$$

schließen, womit gezeigt ist, dass die Parseval'sche Gleichung für jedes $f \in L_2((0, \infty), \omega_{\alpha,\beta})$ gilt. Also ist das System der verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(\alpha,\beta)}$ vollständig im $L_2((0, \infty), \omega_{\alpha,\beta})$. \square

Wie bei den klassischen Laguerre-Polynomen lassen sich zahlreiche Eigenschaften und Beziehungen zu den Ableitungen der Polynome herleiten.

Lemma 5.2

Für die Ableitung der verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(\alpha,\beta)}$ gelten die Gleichungen

$$\frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t) = -\beta \mathcal{L}_{l-1}^{(\alpha+1,\beta)}(t) = -\beta \sum_{k=0}^{l-1} \mathcal{L}_k^{(\alpha,\beta)}(t), \quad l \geq 1 \quad (5.15)$$

$$\mathcal{L}_l^{(\alpha,\beta)}(t) = \frac{1}{\beta} \left[\frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t) - \frac{d}{dt} \mathcal{L}_{l+1}^{(\alpha,\beta)}(t) \right], \quad l \geq 1 \quad (5.16)$$

und es ist

$$\mathcal{L}_l^{(\alpha,\beta)}(0) = \frac{\Gamma(l + \alpha + 1)}{\Gamma(\alpha + 1)\Gamma(l + 1)}.$$

Beweis. Die Gleichungen lassen sich wieder direkt mit Hilfe der Transformation aus (5.9) und unter Verwendung der Beziehungen (vgl. [Sze39], S. 101 f., [Tri70], S. 215)

$$\begin{aligned} \frac{d}{d\tau} \mathcal{L}_l^{(\alpha)}(\tau) &= -\mathcal{L}_{l-1}^{(\alpha+1)}(\tau), \quad \mathcal{L}_l^{(\alpha+1)}(\tau) = \sum_{k=0}^l \mathcal{L}_k^{(\alpha)}(\tau), \quad \mathcal{L}_l^{(\alpha)}(0) = \binom{l + \alpha}{l} \\ \mathcal{L}_l^{(\alpha)}(\tau) &= \frac{d}{d\tau} \mathcal{L}_l^{(\alpha)}(\tau) - \frac{d}{d\tau} \mathcal{L}_{l+1}^{(\alpha)}(\tau) \end{aligned} \quad (5.17)$$

nachrechnen. Wir erhalten

$$\begin{aligned} \frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t) &= \frac{d}{dt} \mathcal{L}_l^{(\alpha)}(g(t)) = \frac{d}{d\tau} \mathcal{L}_l^{(\alpha)}(\tau) \cdot g'(t) = \beta \frac{d}{d\tau} \mathcal{L}_l^{(\alpha)}(\tau) \\ &= -\beta \mathcal{L}_{l-1}^{(\alpha+1)}(\tau) = -\beta \sum_{k=0}^{l-1} \mathcal{L}_k^{(\alpha)}(\tau) \\ &= -\beta \mathcal{L}_{l-1}^{(\alpha+1,\beta)}(t) = -\beta \sum_{k=0}^{l-1} \mathcal{L}_k^{(\alpha,\beta)}(t) \end{aligned} \quad (5.18)$$

und durch Einsetzen von (5.18) in (5.17) sofort auch (5.16). Es gilt $\mathcal{L}_l^{(\alpha)}(0) = \mathcal{L}_l^{(\alpha,\beta)}(0)$. Die verschiedenen Darstellungen der Funktionswerte in $t = 0$ ergeben sich aus der Funktionalgleichung der Gammafunktion

$$\Gamma(z + 1) = z\Gamma(z), \quad z > 0.$$

Aus dieser erhalten wir durch iteratives Anwenden

$$\Gamma(z + l) = z(z + 1) \cdot \dots \cdot (z + l - 1)\Gamma(z)$$

und mit $\Gamma(l + 1) = l!$, $l \in \mathbb{N}_0$ gilt

$$\begin{aligned} \frac{\Gamma(l + \alpha + 1)}{\Gamma(\alpha + 1)\Gamma(l + 1)} &= \frac{(\alpha + 1) \cdot (\alpha + 2) \cdot \dots \cdot (\alpha + 1 + l - 1) \cdot \Gamma(\alpha + 1)}{\alpha \cdot \Gamma(\alpha) \cdot l!} \\ &= \frac{(\alpha + 1) \cdot (\alpha + 2) \cdot \dots \cdot (\alpha + l) \cdot \alpha \cdot \Gamma(\alpha)}{\alpha \cdot \Gamma(\alpha) \cdot l!} \\ &= \frac{(\alpha + 1) \cdot (\alpha + 2) \cdot \dots \cdot (\alpha + l)}{l!} = \binom{l + \alpha}{l}. \end{aligned}$$

□

Mit $k \in \mathbb{N}$ liefert uns die Beziehung aus (5.15) die Gleichung

$$\frac{d^k}{dt^k} \mathcal{L}_l^{(\alpha,\beta)}(t) = (-\beta)^k \mathcal{L}_{l-k}^{(\alpha+k,\beta)}(t), \quad l \geq k \quad (5.19)$$

und wir erhalten

$$\begin{aligned} \left\langle \frac{d^k}{dt^k} \mathcal{L}_l^{(\alpha,\beta)}, \frac{d^k}{dt^k} \mathcal{L}_m^{(\alpha,\beta)} \right\rangle_{\omega_{\alpha+k,\beta}} &= (-\beta)^{2k} \langle \mathcal{L}_{l-k}^{(\alpha+k,\beta)}, \mathcal{L}_{m-k}^{(\alpha+k,\beta)} \rangle_{\omega_{\alpha+k,\beta}} \\ &= \begin{cases} (\beta)^{2k} \gamma_{l-k}^{(\alpha+k,\beta)} & l = m \\ 0 & l \neq m \end{cases}. \end{aligned} \quad (5.20)$$

Gemäß Lemma 5.1 bilden die Polynome $\mathcal{L}_{l-k}^{(\alpha+k,\beta)}$, $l \geq k$ ein vollständiges Funktionensystem im $L_2((0, \infty), \omega_{\alpha+k,\beta})$ bezüglich der Gewichtsfunktion $\omega_{\alpha+k,\beta}(t) = t^{\alpha+k} e^{-\beta t}$, $\alpha > -1$, $\beta > 0$. Damit können wir sofort auch auf die Vollständigkeit der Polynome $\frac{d^k}{dt^k} \mathcal{L}_l^{(\alpha,\beta)}$ schließen, denn für eine Funktion $f \in L_2((0, \infty), \omega_{\alpha+k,\beta})$ gilt unter Verwendung von (5.19) und (5.20) die Parseval'sche Gleichung

$$\|f\|_{2,\omega_{\alpha+k,\beta}}^2 = \sum_{l=k}^{\infty} \left| \left\langle f, \frac{\mathcal{L}_{l-k}^{(\alpha+k,\beta)}}{\|\mathcal{L}_{l-k}^{(\alpha+k,\beta)}\|_{2,\omega_{\alpha+k,\beta}}} \right\rangle_{2,\omega_{\alpha+k,\beta}} \right|^2 = \sum_{l=k}^{\infty} \left| \left\langle f, \frac{\frac{d^k}{dt^k} \mathcal{L}_l^{(\alpha,\beta)}}{\|\frac{d^k}{dt^k} \mathcal{L}_l^{(\alpha,\beta)}\|_{2,\omega_{\alpha+k,\beta}}} \right\rangle_{2,\omega_{\alpha+k,\beta}} \right|^2,$$

womit die Vollständigkeit bewiesen ist. Wir halten dieses Ergebnis in dem nachfolgenden Lemma fest.

Lemma 5.3

Es sei $k \in \mathbb{N}$. Die k -ten Ableitungen der verallgemeinerten Laguerre-Polynome $\frac{d^k}{dt^k} \mathcal{L}_l^{(\alpha, \beta)}$, $l \geq k$ bilden ein vollständiges Funktionensystem im $L_2((0, \infty), \omega_{\alpha+k, \beta})$ bezüglich der Gewichtsfunktion $\omega_{\alpha+k, \beta}(t) = t^{\alpha+k} e^{-\beta t}$, $\alpha > -1$, $\beta > 0$.

Abschließen möchten wir diesen Abschnitt mit der Angabe der Laguerre-Differentialgleichung

$$\tau y'' + (\alpha + 1 - \tau)y' + ly = 0, \quad y = \mathcal{L}_l^{(\alpha)}(\tau), \quad l = 0, 1, \dots,$$

deren Lösungen die klassischen Laguerre-Polynome sind, vgl. [Sze39], S. 100. Mit Hilfe der Transformation aus (5.9) und der Rechnungen in (5.18) sehen wir sofort ein, dass die Differentialgleichung der verallgemeinerten Laguerre-Polynome in die Gestalt

$$ty'' + (\alpha + 1 - \beta t)y' + \beta ly = 0, \quad y = \mathcal{L}_l^{(\alpha, \beta)}(t), \quad l = 0, 1, \dots, \tag{5.21}$$

übergeht.

5.1.2 Fourier-Laguerre-Approximation

Die verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(\alpha, \beta)}$ bilden ein vollständiges Funktionensystem in dem Hilbertraum $L_2((0, \infty), \omega_{\alpha, \beta})$. Demnach lässt sich jede Funktion $f \in L_2((0, \infty), \omega_{\alpha, \beta})$ durch die Fourier-Laguerresche Reihe

$$f = \sum_{l=0}^{\infty} \hat{f}_l \mathcal{L}_l^{(\alpha, \beta)} \tag{5.22}$$

mit den Fourierkoeffizienten

$$\hat{f}_l := \frac{\langle f, \mathcal{L}_l^{(\alpha, \beta)} \rangle_{2, \omega_{\alpha, \beta}}}{\|\mathcal{L}_l^{(\alpha, \beta)}\|_{2, \omega_{\alpha, \beta}}^2}, \quad l \geq 0. \tag{5.23}$$

darstellen, vgl. Satz 3.13. Dabei ist die Konvergenz der obigen Reihe natürlich bezüglich der in $L_2((0, \infty), \omega_{\alpha, \beta})$ definierten Norm gemeint. Die Fourier-Laguerresche Reihe von f konvergiert demnach in dem Intervall $(0, \infty)$ im gewichteten Quadratmittel, also bezüglich der gewichteten $L_2((0, \infty), \omega_{\alpha, \beta})$ -Norm, gegen die Funktion f selbst; das bedeutet

$$\lim_{N \rightarrow \infty} \int_0^{\infty} |f(t) - \sum_{l=0}^N \hat{f}_l \mathcal{L}_l^{(\alpha, \beta)}(t)|^2 \omega_{\alpha, \beta}(t) dt = 0. \tag{5.24}$$

Aufgrund der Stetigkeit und der Definitheit des Skalarproduktes lässt sich die obige Forderung äquivalent als

$$\begin{aligned} \lim_{N \rightarrow \infty} \langle f - \sum_{l=0}^N \hat{f}_l \mathcal{L}_l^{(\alpha, \beta)}, f - \sum_{l=0}^N \hat{f}_l \mathcal{L}_l^{(\alpha, \beta)} \rangle_{2, \omega_{\alpha, \beta}} &= \langle f - \sum_{l=0}^{\infty} \hat{f}_l \mathcal{L}_l^{(\alpha, \beta)}, f - \sum_{l=0}^{\infty} \hat{f}_l \mathcal{L}_l^{(\alpha, \beta)} \rangle_{2, \omega_{\alpha, \beta}} = 0 \\ &\Leftrightarrow f = \sum_{l=0}^{\infty} \hat{f}_l \mathcal{L}_l^{(\alpha, \beta)} \quad \text{in } L_2((0, \infty), \omega_{\alpha, \beta}) \end{aligned}$$

schreiben. Zusammenfassend möchten wir für eine Funktion $f \in L_2((0, \infty), \omega_{\alpha, \beta})$ festhalten:

- Die Fourier-Laguerre-Reihenentwicklung in (5.22) ist stets im Sinne der Normkonvergenz des $L_2((0, \infty), \omega_{\alpha, \beta})$ zu interpretieren, vgl. (5.24).
- Die Entwicklung in (5.22) bedeutet Gleichheit zwischen dem Element f und seiner Fourier-Reihe im Sinne des Hilbertraumes $L_2((0, \infty), \omega_{\alpha, \beta})$, das heißt

$$f = \sum_{l=0}^{\infty} \hat{f}_l \mathcal{L}_l^{(\alpha, \beta)} \quad \text{f.ü. auf } (0, \infty).$$

Wir möchten die Fourier-Reihen-Entwicklung mit den verallgemeinerten Laguerre-Polynomen

$$\mathcal{L}_l^{(0, \beta)}(t) = \mathcal{L}_l^{(\beta)}(t), \quad l \geq 0$$

betrachten. Entsprechend der allgemeinen Entwicklung in (5.22) und (5.23) gilt für eine Funktion $f \in L_2((0, \infty), \omega_{\beta})$ mit $\omega_{\beta}(t) = e^{-\beta t}$, $\beta > 0$ die Reihendarstellung

$$f = \sum_{l=0}^{\infty} \hat{f}_l \mathcal{L}_l^{(\beta)}$$

mit den Fourierkoeffizienten

$$\hat{f}_l = \beta \langle f, \mathcal{L}_l^{(\beta)} \rangle_{2, \omega_{\beta}}, \quad l \geq 0. \quad (5.25)$$

Für eine Funktion $f \in W_2^r((0, \infty), \mathfrak{w}_{r, \beta})$, $r \in \mathbb{N}$ mit

$$\mathfrak{w}_{r, \beta} := \left\{ \omega_{k, \beta}(t) = t^k e^{-\beta t}, \beta \in \mathbb{R}, \beta > 0 \mid \forall k : 0 \leq k \leq r \right\} = \{ \omega_{\beta}, \omega_{1, \beta}, \dots, \omega_{r, \beta} \}$$

ist per Definition $\mathcal{D}^k f \in L_2((0, \infty), \omega_{k, \beta})$ für alle $k \in \mathbb{N}_0$ mit $k \leq r$ richtig. Da $\omega_{k, \beta}^{-1} \in L_{1, \text{loc}}((0, \infty))$, $k \in \mathbb{N}_0$ gilt, folgt nach Lemma 2.29 sofort die Inklusion $L_2((0, \infty), \omega_{k, \beta}) \subset L_{1, \text{loc}}((0, \infty))$. Daher ist $\mathcal{D}^k f \in W_{1, \text{loc}}^1((0, \infty))$ für alle k mit $0 \leq k \leq r - 1$ und entsprechend der Aussage des Satzes 2.4 können wir auf die Existenz eines Repräsentanten \tilde{f} mit $\tilde{f}^{(k)} \in AC_{\text{loc}}((0, \infty)) \forall k : 0 \leq k \leq r - 1$ schließen, sodass $f = \tilde{f}$ und $\mathcal{D}^k f = \frac{d^k}{dt^k} \tilde{f} =$

$\tilde{f}^{(k)} \forall k \leq r$ fast überall auf $(0, \infty)$ gelten. Demnach können wir die schwache Ableitung $\mathcal{D}^k f$ der Sobolev-Funktion $f \in W_2^r((0, \infty), \mathfrak{w}_{r,\beta})$ mit der klassischen Ableitung $\tilde{f}^{(k)}$ ihres Repräsentanten identifizieren. In diesem Sinne nutzen wir im Folgenden für eine Funktion $f \in W_2^r((0, \infty), \mathfrak{w}_{r,\beta})$, $r \in \mathbb{N}$ die Notation $\mathcal{D}^k f = \frac{d^k}{dt^k} f = f^{(k)}$. Der Repräsentant \tilde{f} und dessen Ableitungen bis zur Ordnung $r - 1$ sind nur lokal auf $(0, \infty)$ absolut stetig, weshalb eine sinnvolle Interpretation von Funktionswerten in $t = 0$ nicht möglich ist.

Für eine Funktion $f \in W_2^r((0, \infty), \omega_\beta)$, $r \in \mathbb{N}$ mit $\omega_\beta(t) = e^{-\beta t}$, $\beta \in \mathbb{R}$, $\beta > 0$ lassen sich die obigen Überlegungen übertragen und zusätzlich liefert der Satz von Rellich-Kondrachov die Möglichkeit der sinnvollen Definition von Funktionswerten in $t = 0$. Die Einschränkung von f auf $(0, T)$, $T > 0$ ist Element des Raumes $W_2^r((0, T), \omega_\beta)$. Die Gewichtsfunktion ω_β nimmt auf dem Kompaktum $[0, T]$ ein Minimum $m > 0$ und Maximum $M > 0$ an. Daraus ergibt sich die Gleichheit

$$W_2^r((0, T), \omega_\beta) = W_2^r((0, T)).$$

Dies liefert die Existenz eines stetig differenzierbaren Repräsentanten $\tilde{f} \in C^{r-1}([0, T])$, wodurch für eine Funktion aus $W_2^r((0, \infty), \omega_\beta)$ zusätzlich die Formulierung der Anfangswerte $f^{(k)}(0)$, $k = 0, 1, \dots, r - 1$ legitimiert ist, vgl. Bemerkung 2.3.

Bemerkung 5.1. Eine Funktion $f \in W_1^1((0, \infty))$ besitzt einen lokal auf $(0, \infty)$ absolut stetigen Repräsentanten. Darüber hinaus existiert der Grenzwert

$$\lim_{T \rightarrow 0^+} f(T)$$

und es gilt

$$\lim_{T \rightarrow \infty} f(T) = 0,$$

vgl. [Bur98], S. 138, Theorem 4. Der Repräsentant lässt sich demnach stetig in $t = 0$ fortsetzen und ist somit auf jedem Intervall $[0, T]$, $T > 0$ absolut stetig. Daher besitzt die Sobolev-Funktion eine sinnvolle Definition des Wertes $f(0) = f_0$ und der Hauptsatz der Differential- und Integralrechnung liefert uns den für die nachfolgenden Beweise nützlichen Zusammenhang

$$\int_0^\infty f'(t) dt = -f(0).$$

Lemma 5.4

Sei $r \in \mathbb{N}$ und $f \in W_2^r((0, \infty), \omega_\beta) \cap W_2^r((0, \infty), \mathfrak{w}_{r,\beta})$. Dann gilt für $f^{(k)}$ und alle $k \in \mathbb{N}_0$ mit $k \leq r$ die Fourierentwicklung

$$f^{(k)} = \sum_{l=k}^\infty \hat{f}_l \frac{d^k}{dt^k} \mathcal{L}_l^{(\beta)} = \sum_{l=k}^\infty \hat{f}_l (-\beta)^k \mathcal{L}_{l-k}^{(k,\beta)},$$

wobei \hat{f}_l die Fourierkoeffizienten von f aus (5.25) sind.

Beweis. Wir führen den Beweis in zwei Schritten.

Schritt 1: Wir betrachten zunächst eine Hilfsfunktion $g \in W_2^1((0, \infty), \omega_\beta)$, für die zusätzlich $g \in L_2((0, \infty), \omega_{\alpha, \beta})$ und $\mathcal{D}^1 g = g' \in L_2((0, \infty), \omega_{\alpha+1, \beta})$ mit $\omega_{\alpha, \beta}(t) = t^\alpha e^{-\beta t}$, $\alpha > -1$, $\beta > 0$ gilt. Die Vollständigkeit der Polynome $\frac{d}{dt} \mathcal{L}_l^{(\alpha, \beta)}$, $l \geq 1$ in $L_2((0, \infty), \omega_{\alpha+1, \beta})$ liefert die Fourierentwicklung von g' gemäß

$$g' = \sum_{l=1}^{\infty} b_l \frac{d}{dt} \mathcal{L}_l^{(\alpha, \beta)}$$

mit

$$b_l := \frac{\langle g', \frac{d}{dt} \mathcal{L}_l^{(\alpha, \beta)} \rangle_{2, \omega_{\alpha+1, \beta}}}{\left\| \frac{d}{dt} \mathcal{L}_l^{(\alpha, \beta)} \right\|_{2, \omega_{\alpha+1, \beta}}^2}. \quad (5.26)$$

Wir werden nun zeigen, dass die Koeffizienten b_l aus der Reihenentwicklung der Ableitung g' mit den Fourierkoeffizienten a_l der Funktion g selbst übereinstimmen. Diese lassen sich durch

$$a_l := \frac{\langle g, \mathcal{L}_l^{(\alpha, \beta)} \rangle_{2, \omega_{\alpha, \beta}}}{\left\| \mathcal{L}_l^{(\alpha, \beta)} \right\|_{2, \omega_{\alpha, \beta}}^2}, \quad l \geq 0.$$

berechnen, vgl. (5.23). Wir folgen den Ideen in [Fun92], S. 100. Dafür definieren wir $x(t) := g(t) \left[\frac{d}{dt} \mathcal{L}_l^{(\alpha, \beta)}(t) \right] \omega_{\alpha+1, \beta}(t)$. Es ist $x(0) = 0$ und die verallgemeinerte Ableitung von x existiert. Wir zeigen zunächst, dass $x \in W_1^1((0, \infty))$ gilt. Aus der Hölderschen Ungleichung und mit $g \in L_2((0, \infty), \omega_{\alpha, \beta})$ erhalten wir

$$\begin{aligned} \int_0^{\infty} |x(t)| dt &= \int_0^{\infty} |g(t) \frac{d}{dt} \mathcal{L}_l^{(\alpha, \beta)}(t) \omega_{\alpha+1, \beta}(t)| dt = \int_0^{\infty} |g(t) \frac{d}{dt} \mathcal{L}_l^{(\alpha, \beta)}(t) t^{\alpha+1} e^{-\beta t}| dt \\ &= \int_0^{\infty} |g(t) t^{\frac{\alpha}{2}} e^{-\frac{\beta}{2} t} \frac{d}{dt} \mathcal{L}_l^{(\alpha, \beta)}(t) t^{\frac{\alpha}{2}+1} e^{-\frac{\beta}{2} t}| dt \\ &\leq \left(\int_0^{\infty} |g(t)|^2 t^\alpha e^{-\beta t} dt \right)^{1/2} \cdot \left(\int_0^{\infty} |t \cdot \frac{d}{dt} \mathcal{L}_l^{(\alpha, \beta)}(t)|^2 t^\alpha e^{-\beta t} dt \right)^{1/2} < \infty. \quad (5.27) \end{aligned}$$

Da $p(t) := t \cdot \frac{d}{dt} \mathcal{L}_l^{(\alpha, \beta)}(t)$ offensichtlich ein Polynom vom Grad $\leq l$ mit $l \geq 1$ ist, besitzt es die formelle Gestalt

$$p(t) = a_0 + a_1 t + \dots + a_l t^l, \quad a_0, \dots, a_l \in \mathbb{R}$$

und es gilt

$$|p(t)|^2 \leq (|a_0| + |a_1|t + \dots + |a_l|t^l)^2, \quad t \in (0, \infty).$$

Rechnet man das Quadrat auf der rechten Seite der obigen Zeile aus, so erhält man wieder

ein Polynom (mit maximalen Grad $2l$). Gemeinsam mit der Existenz der Momente bezüglich $\omega_{\alpha,\beta}(t) = t^\alpha e^{-\beta t}$ liefert dies schließlich die Endlichkeit des rechtsstehenden Integrals in (5.27). Durch Anwenden der Dreiecksungleichung gelangen wir zu der Ungleichung

$$\int_0^\infty \left| \frac{d}{dt} x(t) \right| dt \leq I_1 + I_2 + I_3$$

mit

$$I_1 := \int_0^\infty |g'(t) \frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t) t^{\alpha+1} e^{-\beta t}| dt, \quad I_2 := \int_0^\infty |g(t) \frac{d^2}{dt^2} \mathcal{L}_l^{(\alpha,\beta)}(t) t^{\alpha+1} e^{-\beta t}| dt,$$

$$I_3 := \int_0^\infty |g(t) \frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t) \{(\alpha+1)t^\alpha e^{-\beta t} - \beta t^{\alpha+1} e^{-\beta t}\}| dt.$$

Wir nutzen wieder die Höldersche Ungleichung und erhalten mit $g' \in L_2((0, \infty), \omega_{\alpha+1, \beta})$ die Abschätzung

$$I_1 = \int_0^\infty |g'(t) t^{\frac{\alpha+1}{2}} e^{-\frac{\beta}{2}t} \frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t) t^{\frac{\alpha+1}{2}} e^{-\frac{\beta}{2}t}| dt$$

$$\leq \left(\int_0^\infty |g'(t)|^2 t^{\alpha+1} e^{-\beta t} dt \right)^{1/2} \cdot \left(\int_0^\infty |t \cdot \frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t)| t^\alpha e^{-\beta t} dt \right)^{1/2} < \infty.$$

Analog zu den vorangegangenen Ausführungen ergeben sich auch $I_2 < \infty$ und $I_3 < \infty$, womit wir auf $x \in W_1^1((0, \infty))$ schließen können. Entsprechend der Bemerkung 5.1 gilt

$$\langle g', \frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)} \rangle_{2, \omega_{\alpha+1, \beta}} = \int_0^\infty g'(t) \frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t) \omega_{\alpha+1, \beta}(t) dt$$

$$= - \int_0^\infty g(t) \left[\frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t) \omega_{\alpha+1, \beta}(t) \right]' dt. \quad (5.28)$$

Einfaches Anwenden der Produktregel führt auf

$$\left[\frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t) \omega_{\alpha+1, \beta}(t) \right]' = \frac{d^2}{dt^2} \mathcal{L}_l^{(\alpha,\beta)}(t) t^{\alpha+1} e^{-\beta t} + \{(\alpha+1)t^\alpha e^{-\beta t} - \beta t^{\alpha+1} e^{-\beta t}\} \frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t)$$

$$= \left\{ t \frac{d^2}{dt^2} \mathcal{L}_l^{(\alpha,\beta)}(t) + (\alpha+1 - \beta t) \frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t) \right\} t^\alpha e^{-\beta t}, \quad t \in (0, \infty).$$

Wir nutzen die Differentialgleichung aus (5.21) und können aus der letzten Gleichung weiter

auf

$$\left[\frac{d}{dt}\mathcal{L}_l^{(\alpha,\beta)}(t)\omega_{\alpha+1,\beta}(t)\right]' = -\beta l \mathcal{L}_l^{(\alpha,\beta)}(t) \cdot t^\alpha e^{-\beta t} = -\beta l \mathcal{L}_l^{(\alpha,\beta)}(t) \cdot \omega_{\alpha,\beta}(t), \quad t \in (0, \infty) \quad (5.29)$$

schließen. Einsetzen dieser Gleichheit in die rechte Seite der Gleichung (5.28) ergibt

$$\langle g', \frac{d}{dt}\mathcal{L}_l^{(\alpha,\beta)} \rangle_{2,\omega_{\alpha+1,\beta}} = \beta l \cdot \int_0^\infty g(t)\mathcal{L}_l^{(\alpha,\beta)}(t) \cdot \omega_{\alpha,\beta}(t) dt = \beta l \cdot \langle g, \frac{d}{dt}\mathcal{L}_l^{(\alpha,\beta)} \rangle_{2,\omega_{\alpha,\beta}}. \quad (5.30)$$

Multiplizieren wir die Gleichung in (5.29) mit $\mathcal{L}_l^{(\alpha,\beta)}(t)$ und integrieren anschließend, so ergibt sich

$$\begin{aligned} \int_0^\infty \left[\frac{d}{dt}\mathcal{L}_l^{(\alpha,\beta)}(t)\omega_{\alpha+1,\beta}(t)\right]' \mathcal{L}_l^{(\alpha,\beta)}(t) dt &= -\beta l \cdot \int_0^\infty \mathcal{L}_l^{(\alpha,\beta)}(t)\mathcal{L}_l^{(\alpha,\beta)}(t)\omega_{\alpha,\beta}(t) dt \\ &= -\beta l \|\mathcal{L}_l^{(\alpha,\beta)}\|_{2,\omega_{\alpha,\beta}}^2. \end{aligned} \quad (5.31)$$

Definieren wir $\hat{x}(t) := \frac{d}{dt}\mathcal{L}_l^{(\alpha,\beta)}(t)\mathcal{L}_l^{(\alpha,\beta)}(t)\omega_{\alpha+1,\beta}(t)$, so lässt sich analog zu den vorangegangenen Ausführungen zeigen, dass auch \hat{x} ein Element des Raumes $W_1^1((0, \infty))$ mit $\hat{x}(0) = 0$ ist. Daher erhalten wir wieder für den Ausdruck auf der linken Seite in (5.31)

$$\int_0^\infty \left[\frac{d}{dt}\mathcal{L}_l^{(\alpha,\beta)}(t)\omega_{\alpha+1,\beta}(t)\right]' \mathcal{L}_l^{(\alpha,\beta)}(t) dt = -\left\|\frac{d}{dt}\mathcal{L}_l^{(\alpha,\beta)}\right\|_{2,\omega_{\alpha+1,\beta}}^2. \quad (5.32)$$

Einsetzen von (5.32) in (5.31) ergibt

$$\left\|\frac{d}{dt}\mathcal{L}_l^{(\alpha,\beta)}\right\|_{2,\omega_{\alpha+1,\beta}}^2 = \beta l \|\mathcal{L}_l^{(\alpha,\beta)}\|_{2,\omega_{\alpha,\beta}}^2. \quad (5.33)$$

Abschließend tragen wir die Ergebnisse aus (5.30) und (5.33) in (5.26) ein, woraus das gewünschte Resultat folgt:

$$b_l = \frac{\langle g', \frac{d}{dt}\mathcal{L}_l^{(\alpha,\beta)} \rangle_{2,\omega_{\alpha+1,\beta}}}{\left\|\frac{d}{dt}\mathcal{L}_l^{(\alpha,\beta)}\right\|_{2,\omega_{\alpha+1,\beta}}^2} = \frac{\beta l \cdot \langle g, \mathcal{L}_l^{(\alpha,\beta)} \rangle_{2,\omega_{\alpha,\beta}}}{\beta l \cdot \|\mathcal{L}_l^{(\alpha,\beta)}\|_{2,\omega_{\alpha,\beta}}^2} = a_l, \quad l \geq 1.$$

Schritt 2: Da $f \in W_2^r((0, \infty), w_{r,\beta})$ vorausgesetzt wurde, ist $f^{(k)} \in L_2((0, \infty), \omega_{k,\beta})$ für alle $0 \leq k \leq r$. Wir wenden nun die Ergebnisse aus Schritt 1 rekursiv an und nutzen dabei insbesondere die Beziehung aus (5.19).

$\alpha = 0$: Wir setzen $g := f \in L_2((0, \infty), \omega_\beta)$, $g' = f' \in L_2((0, \infty), \omega_{1,\beta})$. Dann gilt

$$\begin{aligned} f &= \sum_{l=0}^\infty \hat{f}_l^{(0)} \mathcal{L}_l^{(\beta)}, \\ f' &= \sum_{l=1}^\infty \hat{f}_l^{(0)} \frac{d}{dt}\mathcal{L}_l^{(\beta)} = \sum_{l=1}^\infty \hat{f}_l^{(0)} (-\beta) \mathcal{L}_{l-1}^{(1,\beta)} \end{aligned}$$

mit den Koeffizienten

$$\hat{f}_l^{(0)} := \frac{\langle f, \mathcal{L}_l^{(\beta)} \rangle_{2, \omega_\beta}}{\|\mathcal{L}_l^{(\beta)}\|_{2, \omega_\beta}^2}.$$

$\alpha = 1$: Wir setzen $g := f' \in L_2((0, \infty), \omega_{1, \beta})$, $g' = f^{(2)} \in L_2((0, \infty), \omega_{2, \beta})$. Dann gilt

$$\begin{aligned} f' &= \sum_{l=0}^{\infty} \hat{f}_l^{(1)} \mathcal{L}_l^{(1, \beta)} = \sum_{l=1}^{\infty} \hat{f}_{l-1}^{(1)} \mathcal{L}_{l-1}^{(1, \beta)}, \\ f^{(2)} &= \sum_{l=1}^{\infty} \hat{f}_l^{(1)} \frac{d}{dt} \mathcal{L}_l^{(1, \beta)} = \sum_{l=2}^{\infty} \hat{f}_{l-1}^{(1)} \frac{d}{dt} \mathcal{L}_{l-1}^{(1, \beta)} \end{aligned}$$

mit den Koeffizienten

$$\hat{f}_l^{(1)} := \frac{\langle f', \mathcal{L}_l^{(1, \beta)} \rangle_{2, \omega_{1, \beta}}}{\|\mathcal{L}_l^{(1, \beta)}\|_{2, \omega_{1, \beta}}^2}.$$

Für die Funktion f' haben wir demnach die Darstellungen

$$f' = \sum_{l=1}^{\infty} \hat{f}_l^{(0)} (-\beta) \mathcal{L}_{l-1}^{(1, \beta)} = \sum_{l=1}^{\infty} \hat{f}_{l-1}^{(1)} \mathcal{L}_{l-1}^{(1, \beta)},$$

wobei die Eindeutigkeit der Fourierkoeffizienten den Zusammenhang

$$\hat{f}_{l-1}^{(1)} = \hat{f}_l^{(0)} (-\beta), \quad l \geq 1 \quad (5.34)$$

liefert. Daraus ergibt sich durch rekursives Anwenden des obigen Schemas die Behauptung, dass für $k \in \mathbb{N}$, $k \leq r$ und mit den Koeffizienten

$$\hat{f}_l^{(k)} := \frac{\langle f^{(k)}, \mathcal{L}_l^{(k, \beta)} \rangle_{2, \omega_{k, \beta}}}{\|\mathcal{L}_l^{(k, \beta)}\|_{2, \omega_{k, \beta}}^2} \quad (5.35)$$

die Gleichung

$$\hat{f}_{l-k}^{(k)} = \hat{f}_l^{(0)} (-\beta)^k, \quad l \geq k \quad (5.36)$$

gilt. Wir beweisen diese Vermutung durch Induktion nach $k \in \mathbb{N}$. Der Induktionsanfang ergibt sich aus der vorangegangenen Diskussion, vgl. (5.34). Für den Induktionsschritt nutzen wir wieder das Resultat aus Schritt 1.

$\alpha = k$: Wir setzen $g := f^{(k)} \in L_2((0, \infty), \omega_{k, \beta})$, $g' = f^{(k+1)} \in L_2((0, \infty), \omega_{k+1, \beta})$. Dann

gilt

$$\begin{aligned} f^{(k)} &= \sum_{l=0}^{\infty} \hat{f}_l^{(k)} \mathcal{L}_l^{(k,\beta)}, \\ f^{(k+1)} &= \sum_{l=1}^{\infty} \hat{f}_l^{(k)} \frac{d}{dt} \mathcal{L}_l^{(k,\beta)} = \sum_{l=k+1}^{\infty} \hat{f}_{l-k}^{(k)} \frac{d}{dt} \mathcal{L}_{l-k}^{(k,\beta)}. \end{aligned} \quad (5.37)$$

Die Induktionsvoraussetzung liefert die Gültigkeit der Formel in (5.36) für ein beliebiges $k \in \mathbb{N}$. Damit ergibt sich aus (5.37) die folgende Gleichungskette:

$$f^{(k+1)} = \sum_{l=k+1}^{\infty} \hat{f}_l^{(0)} (-\beta)^k \frac{d}{dt} \mathcal{L}_{l-k}^{(k,\beta)} = \sum_{l=k+1}^{\infty} \hat{f}_l^{(0)} (-\beta)^{k+1} \mathcal{L}_{l-(k+1)}^{(k+1,\beta)}. \quad (5.38)$$

Außerdem gilt für $f^{(k+1)} \in L_2((0, \infty), \omega_{k+1,\beta})$ die Fourier-Darstellung

$$f^{(k+1)} = \sum_{l=0}^{\infty} \hat{f}_l^{(k+1)} \mathcal{L}_l^{(k+1,\beta)} = \sum_{l=k+1}^{\infty} \hat{f}_{l-(k+1)}^{(k+1)} \mathcal{L}_{l-(k+1)}^{(k+1,\beta)}. \quad (5.39)$$

Ein Vergleich der Darstellung aus (5.38) und (5.39) liefert wieder aufgrund der Eindeutigkeit der Koeffizienten die Gültigkeit der Induktionsbehauptung

$$\hat{f}_{l-(k+1)}^{(k+1)} = \hat{f}_l^{(0)} (-\beta)^{k+1}, \quad l \geq k+1.$$

Durch Anwenden von (5.19) erhalten wir für alle $0 \leq k \leq r$ und $f^{(k)} \in L_2((0, \infty), \omega_{k,\beta})$ die Darstellung

$$\begin{aligned} f^{(k)} &= \sum_{l=0}^{\infty} \hat{f}_l^{(k)} \mathcal{L}_l^{(k,\beta)} = \sum_{l=k}^{\infty} \hat{f}_{l-k}^{(k)} \mathcal{L}_{l-k}^{(k,\beta)} \\ &= \sum_{l=k}^{\infty} \hat{f}_l^{(0)} (-\beta)^k \mathcal{L}_{l-k}^{(k,\beta)} = \sum_{l=k}^{\infty} \hat{f}_l^{(0)} \frac{d^k}{dt^k} \mathcal{L}_l^{(\beta)} \end{aligned}$$

mit $\hat{f}_l^{(0)} = \hat{f}_l$. □

Das vorangegangene Lemma liefert uns eine Darstellung für die Fourier-Laguerre-Reihen der k -ten Ableitung einer Funktion $f \in W_2^r((0, \infty), \omega_\beta) \cap W_2^r((0, \infty), \omega_{r,\beta})$, $r \in \mathbb{N}$, $k \leq r$, bei denen lediglich die Fourierkoeffizienten \hat{f}_l der Funktion f selbst auftauchen. Wir möchten dabei noch einmal betonen, dass dies die Konvergenz bezüglich der entsprechenden L_2 -Norm impliziert. Im Allgemeinen ist eine punktweise oder sogar gleichmäßige Konvergenz der Fourier-Laguerre-Reihen nicht gesichert. Bevor wir diese Aspekte genauer studieren, möchten wir uns zunächst der Diskussion des Fehlers widmen, welcher gemessen in der L_2 -Norm bei der Approximation durch verallgemeinerte Laguerre-Polynome auftritt, und stellen dafür die entsprechenden Resultate aus der Literatur zusammen. Dafür benötigen wir noch

weitere Notationen. Wir führen zunächst die Definition

$$\widetilde{W}_2^1((0, \infty), \omega_\beta) := \left\{ f \in W_2^1((0, \infty), \omega_\beta) : f(0) = 0 \right\}$$

ein und stellen fest, dass $\widetilde{W}_2^1((0, \infty), \omega_\beta)$ ein Unterraum des Sobolev-Raumes $W_2^1((0, \infty), \omega_\beta)$ ist. Betrachten wir eine Folge $\{f_N\}_{N \in \mathbb{N}}$ aus $\widetilde{W}_2^1((0, \infty), \omega_\beta)$, welche im Sinne der Norm des Raumes $W_2^1((0, \infty), \omega_\beta)$ gegen das Grenzelement $f_0 \in W_2^1((0, \infty), \omega_\beta)$ konvergiert, so impliziert dies sofort auch die schwache Konvergenz von $\{f_N\}_{N \in \mathbb{N}}$ gegen f_0 . Da $f_N(0) = 0$ für alle $N \in \mathbb{N}$ vorausgesetzt ist, können wir durch Anwenden des Lemmas 7, S. 86 aus [Lyk16] folgern, dass auch $f_0(0) = 0$ gilt. Dies beweist die Abgeschlossenheit von $\widetilde{W}_2^1((0, \infty), \omega_\beta)$. Als abgeschlossener Unterraum des Hilbertraumes $W_2^1((0, \infty), \omega_\beta)$ ist $\widetilde{W}_2^1((0, \infty), \omega_\beta)$ daher selbst ein Hilbertraum, versehen mit dem Skalarprodukt $\langle \cdot, \cdot \rangle_{1,2,\omega_\beta}$ und der dadurch induzierten Norm des Raumes $W_2^1((0, \infty), \omega_\beta)$.⁴ Oftmals ist es vorteilhaft, den Raum $\widetilde{W}_2^1((0, \infty), \omega_\beta)$ mit einer anderen, äquivalenten Norm zu versehen. Grundlage dafür bildet die Poincaré-Ungleichung für Funktionen $f \in \widetilde{W}_2^1((0, \infty), \omega_\beta)$:

$$\|f\|_{2,\omega_\beta} \leq c \|f'\|_{2,\omega_\beta}, \quad c > 0, \quad (5.40)$$

vgl. [Pic15], S. 171, Lemma 2. Daraus können wir direkt für alle $f \in \widetilde{W}_2^1((0, \infty), \omega_\beta)$ die Abschätzung

$$\|f\|_{1,2,\omega_\beta}^2 = \|f\|_{2,\omega_\beta}^2 + \|f'\|_{2,\omega_\beta}^2 \leq C \|f'\|_{2,\omega_\beta}^2$$

mit positiver Konstante C folgern und dies erlaubt die Einführung der zu $\|\cdot\|_{1,2,\omega_\beta}^2$ auf $\widetilde{W}_2^1((0, \infty), \omega_\beta)$ äquivalenten Norm

$$\|f\|_{\widetilde{1},2,\omega_\beta} := \|f'\|_{2,\omega_\beta},$$

welche durch das Skalarprodukt

$$\langle f, g \rangle_{\widetilde{1},2,\omega_\beta} := \langle f', g' \rangle_{2,\omega_\beta} = \int_0^\infty f'(t)g'(t)\omega_\beta(t)dt, \quad f, g \in \widetilde{W}_2^1((0, \infty), \omega_\beta)$$

induziert wird.

Des Weiteren definieren wir für $N \in \mathbb{N}_0$ die Menge

$$\mathring{\mathcal{P}}_N((0, \infty)) := \{p \in \mathcal{P}_N((0, \infty)) : p(0) = 0\}.$$

Als endlich-dimensionaler Teilraum der Hilberträume $L_2((0, \infty), \omega_\beta)$ und $\widetilde{W}_2^1((0, \infty), \omega_\beta)$ ist $\mathring{\mathcal{P}}_N((0, \infty))$ abgeschlossen. Gemäß den Ausführungen in dem Kapitel 3 ist die Orthogonalprojektion

$$P_{N,\alpha,\beta} : L_2((0, \infty), \omega_{\alpha,\beta}) \rightarrow \mathcal{P}_N((0, \infty)) \quad (5.41)$$

⁴ Ist X ein Banachraum und U ein abgeschlossener Unterraum von X , so ist U vollständig, vgl. [Wer18], S. 5, Lemma I.1.3.

durch die Bedingung

$$\langle P_{N,\alpha,\beta}f - f, p \rangle_{2,\omega_{\alpha,\beta}} = 0 \quad \forall p \in \mathcal{P}_N((0, \infty)) \quad (5.42)$$

festgelegt. Wir wissen, dass sie die eindeutig bestimmte beste polynomiale Approximation einer Funktion $f \in L_2((0, \infty), \omega_{\alpha,\beta})$ bezüglich $\mathcal{P}_N((0, \infty))$ liefert. Da die verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(\alpha,\beta)}$ ein orthogonales System des $L_2((0, \infty), \omega_{\alpha,\beta})$ bilden und

$$\mathcal{P}_N((0, \infty)) = \text{lin} \left\{ \mathcal{L}_0^{(\alpha,\beta)}, \mathcal{L}_1^{(\alpha,\beta)}, \dots, \mathcal{L}_N^{(\alpha,\beta)} \right\}$$

gilt, lässt sich die Bestapproximation von $f \in L_2((0, \infty), \omega_{\alpha,\beta})$ bezüglich $\mathcal{P}_N((0, \infty))$ und damit die orthogonale Projektion $P_{N,\alpha,\beta}f$ durch die bei N abgebrochene Fourier-Laguerre-Reihe

$$(P_{N,\alpha,\beta}f)(t) = \sum_{l=0}^N \hat{f}_l \mathcal{L}_l^{(\alpha,\beta)}(t)$$

mit den in (5.23) angegebenen Fourierkoeffizienten angeben. Ist $\alpha = 0$, so schreiben wir $P_{N,\beta} := P_{N,0,\beta}$.

Die orthogonalen Projektionen

$$\mathring{P}_{N,\beta} : L_2((0, \infty), \omega_\beta) \rightarrow \mathring{\mathcal{P}}_N((0, \infty))$$

und

$$\mathring{P}_{N,\beta}^1 : \widetilde{W}_2^1((0, \infty), \omega_\beta) \rightarrow \mathring{\mathcal{P}}_N((0, \infty))$$

lassen sich durch die Bestimmungsgleichungen

$$\langle \mathring{P}_{N,\beta}f - f, p \rangle_{2,\omega_\beta} = 0 \quad \forall p \in \mathring{\mathcal{P}}_N((0, \infty)) \quad (5.43)$$

bzw.

$$\left\langle \frac{d}{dt}(\mathring{P}_{N,\beta}^1 f - f), p' \right\rangle_{2,\omega_\beta} = 0 \quad \forall p \in \mathring{\mathcal{P}}_N((0, \infty)) \quad (5.44)$$

charakterisieren.

Aufbauend auf der Beziehung in (5.36) können wir auf eine nützliche Eigenschaft bezüglich der orthogonalen Projektion in (5.41) schließen.

Folgerung 5.1. *Sei $r \in \mathbb{N}$ und $f \in W_2^r((0, \infty), \omega_\beta) \cap W_2^r((0, \infty), \mathfrak{w}_{r,\beta})$. Für alle $k \in \mathbb{N}_0$ mit $k \leq r$ gilt*

$$P_{N-k,k,\beta}f^{(k)} = \frac{d^k}{dt^k} P_{N,\beta}f.$$

Beweis. Nach Voraussetzung ist $f^{(k)} \in L_2((0, \infty), \omega_{k,\beta})$ und es gilt die Entwicklung

$$f^{(k)} = \sum_{l=0}^{\infty} \hat{f}_l^{(k)} \mathcal{L}_l^{(k,\beta)}$$

mit den Koeffizienten aus (5.35), woraus sofort

$$(P_{N-k,k,\beta} f^{(k)})(t) = \sum_{l=0}^{N-k} \hat{f}_l^{(k)} \mathcal{L}_l^{(k,\beta)}(t) = \sum_{l=k}^N \hat{f}_{l-k}^{(k)} \mathcal{L}_{l-k}^{(k,\beta)}(t), \quad t \in (0, \infty)$$

folgt. Die orthogonale Projektion für $f \in L_2((0, \infty), \omega_\beta)$ ist durch

$$P_{N,\beta} f(t) = \sum_{l=0}^N \hat{f}_l^{(0)} \mathcal{L}_l^{(\beta)}(t), \quad t \in (0, \infty)$$

gegeben. Wir nutzen die Eigenschaft (5.19) und die Beziehung in (5.36) und erhalten mit

$$\frac{d^k}{dt^k} (P_{N,\beta} f)(t) = \sum_{l=k}^N \hat{f}_l^{(0)} \frac{d^k}{dt^k} \mathcal{L}_l^{(\beta)}(t) = \sum_{l=k}^N \hat{f}_l^{(0)} (-\beta)^k \mathcal{L}_{l-k}^{(k,\beta)}(t) = \sum_{l=k}^N \hat{f}_{l-k}^{(k)} \mathcal{L}_{l-k}^{(k,\beta)}(t), \quad t \in (0, \infty)$$

das gewünschte Resultat. \square

Die Grundeigenschaften und die Vollständigkeit der verallgemeinerten Laguerre-Polynome stellen die wesentlichen Hilfsmittel zur Verfügung, um Aussagen über den Fehler der polynomialen Bestapproximation $(P_{N,\beta} f)$ einer Funktion $f \in L_2((0, \infty), \omega_\beta)$ in Abhängigkeit von N und f zu erhalten. Dass die Bestapproximation im $L_2((0, \infty), \omega_\beta)$ für $N \rightarrow \infty$ gegen f selbst konvergiert, ist klar. Das nachfolgende Lemma wird zeigen, dass die Größenordnung des Approximationsfehlers bei der Laguerre-Approximation lediglich durch die Güte der zugrunde gelegten Funktion f bestimmt wird.

Lemma 5.5 (vgl. [ByXy05], S.385, Theorem 2.1)

Seien $r \in \mathbb{N}, k \in \mathbb{N}_0$ mit $0 \leq k \leq r \leq N + 1$, $N \in \mathbb{N}$ fest gewählt. Für jede Funktion $f \in W_2^r((0, \infty), \omega_\beta) \cap W_2^r((0, \infty), \mathfrak{w}_{r,\beta})$ gilt

$$\left\| \frac{d^k}{dt^k} (P_{N,\beta} f - f) \right\|_{2,\omega_{k,\beta}} \leq c(\beta N)^{(k-r)/2} \left\| \frac{d^r}{dt^r} f \right\|_{2,\omega_{r,\beta}}$$

mit einer Konstanten $c > 0$.

Beweis. Nach Voraussetzung ist $f \in W_2^r((0, \infty), \mathfrak{w}_{r,\beta})$ und daher $\frac{d^k}{dt^k} f = f^{(k)} \in L_2((0, \infty), \omega_{k,\beta})$ für alle $0 \leq k \leq r$. Die Polynome $\mathcal{L}_{l-k}^{(k,\beta)}$, $l \geq k$ bilden ein vollständiges System im $L_2((0, \infty), \omega_{k,\beta})$

und genügen daher der Parseval'schen Gleichung. Aufgrund der Gleichung

$$\begin{aligned} |\langle f^{(k)}, \mathcal{L}_{l-k}^{(k,\beta)} \rangle_{2,\omega_{k,\beta}}|^2 &= |\hat{f}_{l-k}|^2 \|\mathcal{L}_{l-k}^{(k,\beta)}\|_{2,\omega_{k,\beta}}^4 = |\hat{f}_l (-\beta)^k|^2 \|\mathcal{L}_{l-k}^{(k,\beta)}\|_{2,\omega_{k,\beta}}^4 \\ &= |\hat{f}_l|^2 (\beta)^{2k} \|\mathcal{L}_{l-k}^{(k,\beta)}\|_{2,\omega_{k,\beta}}^4, \end{aligned}$$

vgl. (5.35) und (5.36), nimmt sie die folgende Gestalt an:

$$\begin{aligned} \|f^{(k)}\|_{2,\omega_{k,\beta}}^2 &= \sum_{l=k}^{\infty} |\langle f^{(k)}, \frac{\mathcal{L}_{l-k}^{(k,\beta)}}{\|\mathcal{L}_{l-k}^{(k,\beta)}\|_{2,\omega_{k,\beta}}} \rangle_{2,\omega_{k,\beta}}|^2 = \sum_{l=k}^{\infty} \frac{1}{\|\mathcal{L}_{l-k}^{(k,\beta)}\|_{2,\omega_{k,\beta}}^2} |\langle f^{(k)}, \mathcal{L}_{l-k}^{(k,\beta)} \rangle_{2,\omega_{k,\beta}}|^2 \\ &= \sum_{l=k}^{\infty} \frac{1}{\|\mathcal{L}_{l-k}^{(k,\beta)}\|_{2,\omega_{k,\beta}}^2} |\hat{f}_l|^2 (\beta)^{2k} \|\mathcal{L}_{l-k}^{(k,\beta)}\|_{2,\omega_{k,\beta}}^4 = \sum_{l=k}^{\infty} |\hat{f}_l|^2 (\beta)^{2k} \|\mathcal{L}_{l-k}^{(k,\beta)}\|_{2,\omega_{k,\beta}}^2 \\ &= \sum_{l=k}^{\infty} |\hat{f}_l|^2 (\beta)^{2k} \gamma_{l-k}^{(k,\beta)}. \end{aligned} \quad (5.45)$$

Aus der Bedingung (5.42) folgt

$$\begin{aligned} \|f^{(k)}\|_{2,\omega_{k,\beta}}^2 &= \|f^{(k)} - P_{N-k,k,\beta} f^{(k)} + P_{N-k,k,\beta} f^{(k)}\|_{2,\omega_{k,\beta}}^2 \\ &= \|f^{(k)} - P_{N-k,k,\beta} f^{(k)}\|_{2,\omega_{k,\beta}}^2 + 2 \langle f^{(k)} - P_{N-k,k,\beta} f^{(k)}, P_{N-k,k,\beta} f^{(k)} \rangle_{2,\omega_{k,\beta}} \\ &\quad + \|P_{N-k,k,\beta} f^{(k)}\|_{2,\omega_{k,\beta}}^2 \\ &= \|f^{(k)} - P_{N-k,k,\beta} f^{(k)}\|_{2,\omega_{k,\beta}}^2 + \|P_{N-k,k,\beta} f^{(k)}\|_{2,\omega_{k,\beta}}^2. \end{aligned} \quad (5.46)$$

Weiterhin gilt

$$\begin{aligned} \|P_{N-k,k,\beta} f^{(k)}\|_{2,\omega_{k,\beta}}^2 &= \langle P_{N-k,k,\beta} f^{(k)}, P_{N-k,k,\beta} f^{(k)} \rangle_{2,\omega_{k,\beta}} \\ &= \left\langle \sum_{l=k}^N \hat{f}_l (-\beta)^k \mathcal{L}_{l-k}^{(k,\beta)}, \sum_{m=k}^N \hat{f}_m (-\beta)^k \mathcal{L}_{m-k}^{(k,\beta)} \right\rangle_{2,\omega_{k,\beta}} \\ &= \sum_{l,m=k}^N \hat{f}_l \hat{f}_m \beta^{2k} \langle \mathcal{L}_{l-k}^{(k,\beta)}, \mathcal{L}_{m-k}^{(k,\beta)} \rangle_{2,\omega_{k,\beta}} = \sum_{l=k}^N |\hat{f}_l|^2 \beta^{2k} \gamma_{l-k}^{(k,\beta)}. \end{aligned} \quad (5.47)$$

Wir tragen die Ergebnisse aus (5.45), (5.46), (5.47) und aus der Folgerung 5.1 zusammen und erhalten die nachfolgende Rechnung:

$$\begin{aligned} \left\| \frac{d^k}{dt^k} (P_{N,\beta} f - f) \right\|_{2,\omega_{k,\beta}}^2 &= \left\| \frac{d^k}{dt^k} P_{N,\beta} f - f^{(k)} \right\|_{2,\omega_{k,\beta}}^2 = \|P_{N-k,k,\beta} f^{(k)} - f^{(k)}\|_{2,\omega_{k,\beta}}^2 \\ &= \|f^{(k)}\|_{2,\omega_{k,\beta}}^2 - \|P_{N-k,k,\beta} f^{(k)}\|_{2,\omega_{k,\beta}}^2 \\ &= \sum_{l=k}^{\infty} |\hat{f}_l|^2 \beta^{2k} \gamma_{l-k}^{(k,\beta)} - \sum_{l=k}^N |\hat{f}_l|^2 \beta^{2k} \gamma_{l-k}^{(k,\beta)} \\ &= \sum_{l=N+1}^{\infty} |\hat{f}_l|^2 \beta^{2k} \gamma_{l-k}^{(k,\beta)}. \end{aligned} \quad (5.48)$$

Für die Partialsummen der (bzgl. der L_2 -Norm) konvergenten Reihen in (5.45) mit $k = r$ und in (5.48) führen wir die Bezeichnungen

$$a_n := \sum_{l=N+1}^n |\hat{f}_l|^2 \beta^{2k} \gamma_{l-k}^{(k,\beta)}, \quad b_n := \sum_{l=r}^n |\hat{f}_l|^2 \beta^{2r} \gamma_{l-r}^{(r,\beta)}$$

ein. Es gilt

$$\begin{aligned} a_n &= \sum_{l=N+1}^n |\hat{f}_l|^2 \beta^{2k} \gamma_{l-k}^{(k,\beta)} \frac{\beta^{2r} \gamma_{l-r}^{(r,\beta)}}{\beta^{2r} \gamma_{l-r}^{(r,\beta)}} \leq \max_{N+1 \leq l \leq n} \left\{ \frac{\gamma_{l-k}^{(k,\beta)}}{\gamma_{l-r}^{(r,\beta)}} \right\} \beta^{2k-2r} \sum_{l=N+1}^n |\hat{f}_l|^2 \beta^{2r} \gamma_{l-r}^{(r,\beta)} \\ &\leq \max_{N+1 \leq l \leq n} \left\{ \frac{\gamma_{l-k}^{(k,\beta)}}{\gamma_{l-r}^{(r,\beta)}} \right\} \beta^{2k-2r} \sum_{l=r}^n |\hat{f}_l|^2 \beta^{2r} \gamma_{l-r}^{(r,\beta)} \\ &\leq \max_{N+1 \leq l \leq n} \left\{ \frac{\gamma_{l-k}^{(k,\beta)}}{\gamma_{l-r}^{(r,\beta)}} \right\} \beta^{2k-2r} b_n \end{aligned} \quad (5.49)$$

und

$$\frac{\gamma_{l-k}^{(k,\beta)}}{\gamma_{l-r}^{(r,\beta)}} = \frac{\Gamma((l-k) + k + 1)}{\beta^{k+1} (l-k)!} \cdot \frac{\beta^{r+1} (l-r)!}{\Gamma((l-r) + r + 1)} = \beta^{r-k} \frac{(l-r)!}{(l-k)!}, \quad l \geq N+1 \geq r \geq k.$$

Außerdem ist für jedes $m \in \mathbb{N}$ mit $m+1-r > 0$

$$\frac{(m-r)!}{(m-k)!} \geq \frac{(m+1-r) \cdot (m-r)!}{(m+1-k) \cdot (m-k)!} = \frac{(m+1-r)!}{(m+1-k)!}, \quad (k \leq r)$$

richtig, woraus folgt, dass das Maximum in (5.49) für $l = N+1$ angenommen wird. Damit können wir weiter auf

$$a_n \leq \frac{\gamma_{N+1-k}^{(k,\beta)}}{\gamma_{N+1-r}^{(r,\beta)}} \beta^{2k-2r} b_n = \beta^{r-k} \frac{(N+1-r)!}{(N+1-k)!} \beta^{2k-2r} b_n = \beta^{k-r} \frac{(N+1-r)!}{(N+1-k)!} b_n$$

schließen, woraus nach dem Vergleichskriterium für konvergente Folgen

$$\lim_{n \rightarrow \infty} a_n \leq \beta^{k-r} \frac{(N+1-r)!}{(N+1-k)!} \lim_{n \rightarrow \infty} b_n$$

folgt. Unter Beachtung von (5.45) und (5.48) liefert die obige Ungleichung die Abschätzung

$$\left\| \frac{d^k}{dt^k} (P_{N,\beta} f - f) \right\|_{2,\omega_{k,\beta}}^2 \leq \beta^{k-r} \frac{(N+1-r)!}{(N+1-k)!} \left\| \frac{d^r}{dt^r} f \right\|_{2,\omega_{r,\beta}}^2. \quad (5.50)$$

Wird r fest in dem Intervall $0 \leq r \leq N+1$ gewählt, so gilt

$$\sqrt{\frac{(N+1-r)!}{N!}} \cong N^{(1-r)/2},$$

das bedeutet

$$\lim_{N \rightarrow \infty} \sqrt{\frac{(N+1-r)!}{N!}} \cdot \frac{1}{N^{(1-r)/2}} = 1$$

vgl. [STW11], S. 119f., Remark 3.7. Aus der obigen Konvergenz folgt die Existenz einer Konstanten $c_1 > 0$ mit

$$\sqrt{\frac{(N+1-r)!}{N!}} \cdot \frac{1}{N^{(1-r)/2}} \leq c_1 \Leftrightarrow \sqrt{\frac{(N+1-r)!}{N!}} \leq c_1 N^{(1-r)/2} \quad \forall N.$$

Analog erhalten wir die Existenz einer Konstanten $c_2 > 0$, sodass für festes k mit $0 \leq k \leq N+1$

$$\sqrt{\frac{(N+1-k)!}{N!}} \leq c_2 N^{(1-k)/2} \quad \forall N$$

gilt. Zusammengefasst können wir daher auf

$$\sqrt{\frac{(N+1-r)!}{(N+1-k)!}} = \frac{\sqrt{\frac{(N+1-r)!}{N!}}}{\sqrt{\frac{(N+1-k)!}{N!}}} \leq \frac{c_1 N^{(1-r)/2}}{c_2 N^{(1-k)/2}} = \frac{c_1}{c_2} N^{(k-r)/2}$$

schließen und das Einsetzen dieser Abschätzung in (5.50) liefert mit $c > 0$ das gewünschte Ergebnis

$$\| \frac{d^k}{dt^k} (P_{N,\beta} f - f) \|_{2,\omega_{k,\beta}}^2 \leq c^2 \beta^{k-r} N^{(k-r)} \| \frac{d^r}{dt^r} f \|_{2,\omega_{r,\beta}}^2.$$

□

Lemma 5.6 (vgl. [ByXy05], S. 391, Theorem 2.3)

Für $r \in \mathbb{N}$ sei $f \in W_2^r((0, \infty), \omega_\beta)$, $f' \in W_2^{r-1}((0, \infty), \mathfrak{w}_{r-1,\beta})$ und $f(0) = 0$. Dann gilt für $N \geq r$ die Ungleichung

$$\| \frac{d}{dt} (\mathring{P}_{N,\beta}^1 f - f) \|_{2,\omega_\beta} \leq c(\beta N)^{(1-r)/2} \| \frac{d^r}{dt^r} f \|_{2,\omega_{r-1,\beta}}$$

mit einer Konstanten $c > 0$.

Beweis. Nach Voraussetzung ist $f \in \widetilde{W}_2^1((0, \infty), \omega_\beta)$. Aufgrund von (5.44) gilt die Gleichung

$$\begin{aligned} \langle f' - \frac{d}{dt} \mathring{P}_{N,\beta}^1 f, \frac{d}{dt} \mathring{P}_{N,\beta}^1 f - p' \rangle_{2,\omega_\beta} &= \langle f' - \frac{d}{dt} \mathring{P}_{N,\beta}^1 f, \frac{d}{dt} \mathring{P}_{N,\beta}^1 f \rangle_{2,\omega_\beta} - \langle f' - \frac{d}{dt} \mathring{P}_{N,\beta}^1 f, p' \rangle_{2,\omega_\beta} \\ &= 0 \quad \forall p \in \mathring{\mathcal{P}}_N((0, \infty)). \end{aligned}$$

Damit erhalten wir für alle $p \in \mathring{\mathcal{P}}_N((0, \infty))$ die folgende Rechnung:

$$\begin{aligned}
 \|f' - p'\|_{2, \omega_\beta}^2 &= \left\| \left(f' - \frac{d}{dt} \mathring{P}_{N, \beta}^1 f \right) + \left(\frac{d}{dt} \mathring{P}_{N, \beta}^1 f - p' \right) \right\|_{2, \omega_\beta}^2 \\
 &= \|f' - \frac{d}{dt} \mathring{P}_{N, \beta}^1 f\|_{2, \omega_\beta}^2 + 2 \langle f' - \frac{d}{dt} \mathring{P}_{N, \beta}^1 f, \frac{d}{dt} \mathring{P}_{N, \beta}^1 f - p' \rangle_{2, \omega_\beta} + \|\frac{d}{dt} \mathring{P}_{N, \beta}^1 f - p'\|_{2, \omega_\beta}^2 \\
 &= \|f' - \frac{d}{dt} \mathring{P}_{N, \beta}^1 f\|_{2, \omega_\beta}^2 + \|\frac{d}{dt} \mathring{P}_{N, \beta}^1 f - p'\|_{2, \omega_\beta}^2 \geq \|f' - \frac{d}{dt} \mathring{P}_{N, \beta}^1 f\|_{2, \omega_\beta}^2. \quad (5.51)
 \end{aligned}$$

Wir setzen

$$q(t) := \int_0^t P_{N-1, \beta} f'(\tau) d\tau.$$

Da offensichtlich $q \in \mathring{\mathcal{P}}_N((0, \infty))$ gilt, gilt die Abschätzung in (5.51) auch für $p = q$. Nach Voraussetzung ist $f' \in W_2^{r-1}((0, \infty), \omega_\beta) \cap W_2^{r-1}((0, \infty), \mathfrak{w}_{r-1, \beta})$, $N \geq r \geq 1$. Wir können demnach die Abschätzung aus (5.50) mit $k = 0$ anwenden und erhalten das zu beweisende Resultat

$$\begin{aligned}
 \|f' - \frac{d}{dt} \mathring{P}_{N, \beta}^1 f\|_{2, \omega_\beta}^2 &\leq \|f' - P_{N-1, \beta} f'\|_{2, \omega_\beta}^2 \\
 &\leq \beta^{-(r-1)} \frac{(N-r+1)!}{N!} \|\frac{d^{r-1}}{dt^{r-1}} f'\|_{2, \omega_{r-1, \beta}}^2 \\
 &\leq c^2 (\beta N)^{1-r} \|\frac{d^r}{dt^r} f\|_{2, \omega_{r-1, \beta}}^2
 \end{aligned}$$

mit $c > 0$. □

Folgerung 5.2 (vgl. [XyH13], S. 1119, Theorem 2.1). *Für $r \in \mathbb{N}$ sei $f \in W_2^r((0, \infty), \omega_\beta)$, $f' \in W_2^{r-1}((0, \infty), \mathfrak{w}_{r-1, \beta})$, $f(0) = 0$ und $N \geq r$. Dann gilt*

$$\|\mathring{P}_{N, \beta}^1 f - f\|_{2, \omega_\beta} \leq c \beta^{(-1-r)/2} N^{(1-r)/2} \|\frac{d^r}{dt^r} f\|_{2, \omega_{r-1, \beta}}$$

mit einer Konstanten $c > 0$.

Beweis. Nach Voraussetzung ist $g := f - \mathring{P}_{N, \beta}^1 f \in W_2^1((0, \infty), \omega_\beta)$ mit $g(0) = 0$. Wir definieren $x(t) := g^2(t) \omega_\beta(t)$ und erhalten

$$\int_0^\infty |x(t)| dt = \int_0^\infty g^2(t) \omega_\beta(t) dt < \infty, \quad \int_0^\infty |x'(t)| dt \leq \kappa \left(\|g\|_{2, \omega_\beta} \cdot \|g'\|_{2, \omega_\beta} + \|g\|_{2, \omega_\beta}^2 \right) < \infty$$

mit einer Konstanten $\kappa > 0$, also ist $g \in W_1^1((0, \infty))$. Aufgrund von Bemerkung 5.1 ergibt sich daher

$$\beta \int_0^\infty [f(t) - \mathring{P}_{N, \beta}^1 f(t)]^2 \omega_\beta(t) dt = 2 \int_0^\infty [f(t) - \mathring{P}_{N, \beta}^1 f(t)] \frac{d}{dt} (f(t) - \mathring{P}_{N, \beta}^1 f(t)) \omega_\beta(t) dt.$$

Für $a, b \in \mathbb{R}$ und mit beliebigen $\varepsilon > 0$ gilt die Ungleichung

$$2ab = (\sqrt{\varepsilon}a + \frac{1}{\sqrt{\varepsilon}}b)^2 - \varepsilon a^2 - \frac{1}{\varepsilon}b^2 \leq 2(\varepsilon a^2 + \frac{1}{\varepsilon}b^2) - \varepsilon a^2 - \frac{1}{\varepsilon}b^2 = \varepsilon a^2 + \frac{1}{\varepsilon}b^2,$$

womit wir (mit $\varepsilon := \frac{\beta}{2}$) weiter auf

$$\begin{aligned} \beta \int_0^\infty [f(t) - \mathring{P}_{N,\beta}^1 f(t)]^2 \omega_\beta(t) dt &\leq \frac{\beta}{2} \int_0^\infty [f(t) - \mathring{P}_{N,\beta}^1 f(t)]^2 \omega_\beta(t) dt \\ &\quad + \frac{2}{\beta} \int_0^\infty \left[\frac{d}{dt} (f(t) - \mathring{P}_{N,\beta}^1 f(t)) \right]^2 \omega_\beta(t) dt \end{aligned}$$

schließen können. Umstellen und Anwenden von Lemma 5.6 ergibt schließlich die zu beweisende Abschätzung:

$$\begin{aligned} \int_0^\infty [f(t) - \mathring{P}_{N,\beta}^1 f(t)]^2 \omega_\beta(t) dt &\leq \frac{4}{\beta^2} \int_0^\infty \left[\frac{d}{dt} (f(t) - \mathring{P}_{N,\beta}^1 f(t)) \right]^2 \omega_\beta(t) dt \\ &\leq c\beta^{-1-r} N^{1-r} \left\| \frac{d^r}{dt^r} f \right\|_{2, \omega_{r-1, \beta}}^2. \end{aligned}$$

□

Lemma 5.7 (vgl. [XyH13], S. 1119, Lemma 2.4)

Für $r \in \mathbb{N}$ sei $f \in W_2^r((0, \infty), \omega_\beta)$, $f' \in W_2^{r-1}((0, \infty), \omega_{r-1, \beta})$, $f(0) = 0$ und $N \geq r$. Dann gilt

$$\|\mathring{P}_{N,\beta} f - f\|_{2, \omega_\beta} \leq c(\beta N)^{(1-r)/2} \left\| \frac{d^r}{dt^r} f \right\|_{2, \omega_{r-1, \beta}}$$

mit einer Konstanten $c > 0$.

Beweis. Der Beweis erfolgt mit den gleichen Argumenten und Ideen wie in dem Lemma 5.6. Analog wie in (5.51) gilt aufgrund von (5.43) die Abschätzung

$$\|\mathring{P}_{N,\beta} f - f\|_{2, \omega_\beta} \leq \|p - f\|_{2, \omega_\beta} \quad \forall p \in \mathring{P}_N((0, \infty)).$$

Da $p - f \in \widetilde{W}_2^1((0, \infty), \omega_\beta)$ gilt, folgt aus der Poincaré-Ungleichung (5.40) die Abschätzung

$$\|\mathring{P}_{N,\beta} f - f\|_{2, \omega_\beta} \leq \widehat{c} \|p' - f'\|_{2, \omega_\beta} \quad \forall p \in \mathring{P}_N((0, \infty)), \widehat{c} > 0. \quad (5.52)$$

Das Polynom

$$q(t) := \int_0^t P_{N-1, \beta} f'(\tau) d\tau$$

aus dem Raum $\mathring{\mathcal{P}}_N((0, \infty))$ setzen wir in (5.52) ein und erhalten wieder durch Anwendung von (5.50) mit $k = 0$ die zu beweisende Ungleichung:

$$\|\mathring{P}_{N,\beta}f - f\|_{2,\omega_\beta} \leq \hat{c} \|P_{N-1,\beta}f' - f'\|_{2,\omega_\beta} \leq c(\beta N)^{(1-r)/2} \left\| \frac{d^r}{dt^r} f \right\|_{2,\omega_{r-1,\beta}}, \quad c > 0.$$

□

Lemma 5.8

Ist $\alpha \geq 0$, so gilt für jedes Polynom $p \in \mathcal{P}_N((0, \infty))$ die Abschätzung

$$\|p'\|_{2,\omega_{\alpha,\beta}} \leq \beta N \|p\|_{2,\omega_{\alpha,\beta}}.$$

Beweis. Jedes Polynom $p \in \mathcal{P}_N((0, \infty))$ lässt sich als Linearkombination der verallgemeinerten Laguerre-Polynome

$$p(t) = \sum_{l=0}^N \hat{p}_l \mathcal{L}_l^{(\alpha,\beta)}(t), \quad \hat{p}_l \in \mathbb{R} \quad (5.53)$$

darstellen, wobei aufgrund von

$$\langle p, \mathcal{L}_m^{(\alpha,\beta)} \rangle_{2,\omega_{\alpha,\beta}} = \left\langle \sum_{l=0}^N \hat{p}_l \mathcal{L}_l^{(\alpha,\beta)}, \mathcal{L}_m^{(\alpha,\beta)} \right\rangle_{2,\omega_{\alpha,\beta}} = \sum_{l=0}^N \hat{p}_l \langle \mathcal{L}_l^{(\alpha,\beta)}, \mathcal{L}_m^{(\alpha,\beta)} \rangle_{2,\omega_{\alpha,\beta}} = \hat{p}_m \gamma_m^{(\alpha,\beta)}$$

offenbar

$$\hat{p}_m = \frac{1}{\gamma_m^{(\alpha,\beta)}} \langle p, \mathcal{L}_m^{(\alpha,\beta)} \rangle_{2,\omega_{\alpha,\beta}}, \quad m = 0, 1, \dots, N$$

gilt. Für die Ableitung

$$p'(t) = \sum_{l=1}^N \hat{p}_l \frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t)$$

ergibt sich unter Beachtung von (5.15) die Rechnung

$$p'(t) = \sum_{l=1}^N \hat{p}_l \frac{d}{dt} \mathcal{L}_l^{(\alpha,\beta)}(t) = \sum_{l=1}^N \hat{p}_l \left(-\beta \sum_{k=0}^{l-1} \mathcal{L}_k^{(\alpha,\beta)}(t) \right) = \sum_{l=0}^{N-1} \left(-\beta \sum_{k=l+1}^N \hat{p}_k \right) \mathcal{L}_l^{(\alpha,\beta)}(t). \quad (5.54)$$

Mit Hilfe des Satzes von Pythagoras (vgl. Satz 3.9) und den Darstellungen in (5.53) und

(5.54) erhalten wir

$$\begin{aligned} \|p\|_{2,\omega_{\alpha,\beta}}^2 &= \left\| \sum_{l=0}^N \hat{p}_l \mathcal{L}_l^{(\alpha,\beta)} \right\|_{2,\omega_{\alpha,\beta}}^2 = \sum_{l=0}^N \|\hat{p}_l \mathcal{L}_l^{(\alpha,\beta)}\|_{2,\omega_{\alpha,\beta}}^2 = \sum_{l=0}^N |\hat{p}_l|^2 \|\mathcal{L}_l^{(\alpha,\beta)}\|_{2,\omega_{\alpha,\beta}}^2 \\ &= \sum_{l=0}^N |\hat{p}_l|^2 \gamma_l^{(\alpha,\beta)} \end{aligned}$$

und analog

$$\|p'\|_{2,\omega_{\alpha,\beta}}^2 = \sum_{l=0}^{N-1} \beta^2 \left(\sum_{k=l+1}^N \hat{p}_k \right)^2 \gamma_l^{(\alpha,\beta)}.$$

Aus der Ungleichung von Cauchy-Schwarz folgt

$$\left(\sum_{k=l+1}^N \hat{p}_k \right)^2 = \left(\sum_{k=l+1}^N \hat{p}_k (\gamma_N^{(\alpha,\beta)})^{1/2} (\gamma_N^{(\alpha,\beta)})^{-1/2} \right)^2 \leq \left(\sum_{k=l+1}^N |\hat{p}_k|^2 \gamma_N^{(\alpha,\beta)} \right) \left(\sum_{k=l+1}^N (\gamma_N^{(\alpha,\beta)})^{-1} \right)$$

und daher

$$\begin{aligned} \|p'\|_{2,\omega_{\alpha,\beta}}^2 &\leq \sum_{l=0}^{N-1} \beta^2 \left(\sum_{k=l+1}^N |\hat{p}_k|^2 \gamma_N^{(\alpha,\beta)} \right) \left(\sum_{k=l+1}^N (\gamma_N^{(\alpha,\beta)})^{-1} \right) \gamma_l^{(\alpha,\beta)} \\ &\leq \sum_{l=0}^{N-1} \beta^2 \|p\|_{2,\omega_{\alpha,\beta}}^2 \left(\sum_{k=l+1}^N (\gamma_N^{(\alpha,\beta)})^{-1} \right) \gamma_l^{(\alpha,\beta)}. \end{aligned}$$

Für $\alpha \geq 0$ und mit Hilfe von (5.12) ist

$$\frac{\gamma_{k+1}^{(\alpha,\beta)}}{\gamma_k^{(\alpha,\beta)}} = \frac{\Gamma(k + \alpha + 1 + 1)}{\beta^{\alpha+1} (k + 1)!} \cdot \frac{\beta^{\alpha+1} k!}{\Gamma(k + \alpha + 1)} = \frac{k + \alpha + 1}{k + 1} \geq 1, \quad (k \in \mathbb{N})$$

das bedeutet $\gamma_{k+1}^{(\alpha,\beta)} \geq \gamma_k^{(\alpha,\beta)}$. Damit können wir weiter abschätzen und gelangen schlussendlich zu der Abschätzung

$$\begin{aligned} \|p'\|_{2,\omega_{\alpha,\beta}}^2 &\leq \beta^2 \|p\|_{2,\omega_{\alpha,\beta}}^2 \sum_{l=0}^{N-1} \gamma_l^{(\alpha,\beta)} \left(\sum_{k=l+1}^N (\gamma_N^{(\alpha,\beta)})^{-1} \right) \leq \beta^2 \|p\|_{2,\omega_{\alpha,\beta}}^2 N \underbrace{\sum_{l=0}^{N-1} \frac{\gamma_l^{(\alpha,\beta)}}{\gamma_{l+1}^{(\alpha,\beta)}}}_{\leq 1} \\ &\leq (\beta N)^2 \|p\|_{2,\omega_{\alpha,\beta}}^2. \end{aligned}$$

□

Die vorangegangenen Resultate beschäftigen sich mit der Konvergenz im quadratischen Mittel von Fourier-Laguerre-Reihe gegen die zugrunde gelegte Funktion $f \in L_2((0, \infty), \omega_{\alpha,\beta})$.

Wir wissen, dass die Reihenentwicklung Gleichheit zwischen dem Element f und seiner Fourier-Reihe im Sinne des Hilbertraumes $L_2((0, \infty), \omega_{\alpha, \beta})$ bedeutet, das heißt

$$f = \sum_{l=0}^{\infty} \hat{f}_l \mathcal{L}_l^{(\alpha, \beta)} \quad \text{f.ü. auf } (0, \infty)$$

mit den Fourier-Koeffizienten

$$\hat{f}_l := \frac{\langle f, \mathcal{L}_l^{(\alpha, \beta)} \rangle_{2, \omega_{\alpha, \beta}}}{\|\mathcal{L}_l^{(\alpha, \beta)}\|_{2, \omega_{\alpha, \beta}}^2}, \quad l \geq 0 \quad (5.55)$$

gilt. Nun möchten wir unsere Betrachtungen auf weitere Konvergenzkonzepte ausweiten und stellen uns die Frage, ob sich die Funktion f in $(0, \infty)$ auch punktweise durch ihre formale Fourier-Laguerre-Reihe approximieren lässt. Wir studieren demnach Voraussetzungen, unter welchen wir

$$f(t) = \sum_{l=0}^{\infty} \hat{f}_l \mathcal{L}_l^{(\alpha, \beta)}(t) \quad \text{f.ü. auf } (0, \infty)$$

schreiben dürfen. In unserer Publikation [PB19] haben wir uns mit genau dieser Thematik beschäftigt und konnten damit einen Beitrag zu existierenden Ergebnissen aus der Theorie der Fourier-Laguerre-Reihen leisten. Wir zitieren im Folgenden die wichtigsten Resultate aus dieser Arbeit und verweisen den interessierten Leser für weitere Ausführungen auf die genannte Quelle.

In Anlehnung an [Tri70], S. 243 f beginnen wir unsere Betrachtungen mit einem Resultat, welches ein hinreichendes Kriterium für die gleichmäßige Konvergenz (welche die punktweise Konvergenz nach sich zieht) einer Fourier-Laguerre-Reihe liefert. In [Tri70] wird die Gewichtsfunktion $\omega_{\alpha}(t) = t^{\alpha} e^{-t}$, $\alpha > -1$ zugrunde gelegt und das Kriterium entsprechend für $\beta = 0$ bewiesen. Wir nutzen die Beweisideen aus [Tri70] und verallgemeinern die dortigen Ausführungen für Gewichte $\omega_{\alpha, \beta}(t) = t^{\alpha} e^{-\beta t}$ mit $\alpha > -1$ und $\beta > 0$.

Satz 5.9 ([PB19], Theorem 4)

Jede Funktion Φ , welche sich mit einer Funktion $f \in L_2((0, \infty), \omega_{\alpha, \beta})$ für alle $t > 0$ in der Form

$$\Phi(t) = \beta e^{\beta t} \int_t^{\infty} f(\tau) e^{-\beta \tau} d\tau$$

darstellen lässt, kann in jedem endlichen Intervall $(0, b)$, $b > 0$ in die absolut und gleichmäßig konvergente Fourier-Reihe

$$\Phi(t) = \sum_{l=0}^{\infty} \hat{\Phi}_l \mathcal{L}_l^{(\alpha, \beta)}(t) \quad (5.56)$$

mit den Koeffizienten

$$\hat{\Phi}_l := \hat{f}_l - \hat{f}_{l+1}, \quad l = 0, 1, \dots$$

entwickelt werden. Dabei sind \hat{f}_l die Fourier-Koeffizienten von f bezüglich der verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(\alpha,\beta)}$, $l = 1, \dots$, vgl. (5.55).

Beweis. Für alle Funktionen $f, g \in L_2((0, \infty), \omega_{\alpha,\beta})$ und die verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(\alpha,\beta)}$, $l = 0, 1, \dots$ gilt die erweiterte Parseval'sche Gleichung (vgl. Satz 3.13), woraus

$$\begin{aligned} \langle f, g \rangle_{2, \omega_{\alpha,\beta}} &= \sum_{l=0}^{\infty} \frac{1}{\|\mathcal{L}_l^{(\alpha,\beta)}\|_{2, \omega_{\alpha,\beta}}} \langle f, \mathcal{L}_l^{(\alpha,\beta)} \rangle_{2, \omega_{\alpha,\beta}} \cdot \frac{1}{\|\mathcal{L}_l^{(\alpha,\beta)}\|_{2, \omega_{\alpha,\beta}}} \langle g, \mathcal{L}_l^{(\alpha,\beta)} \rangle_{2, \omega_{\alpha,\beta}} \quad (5.57) \\ &= \sum_{l=0}^{\infty} \hat{f}_l \int_0^{\infty} g(t) \mathcal{L}_l^{(\alpha,\beta)}(t) \omega_{\alpha,\beta}(t) dt \end{aligned}$$

folgt. Dabei sind \hat{f}_l, \hat{g}_l , $l = 0, 1, \dots$ die Fourier-Laguerre-Koeffizienten der Funktionen f, g , vgl. (5.55). Wir definieren die Funktion

$$g(t) := \tau_{\xi}(t) t^{-\alpha}$$

mit Hilfe der Stufenfunktion

$$\tau_{\xi}(t) := \begin{cases} 0, & 0 \leq t < \xi \\ 1, & t \geq \xi \end{cases},$$

wobei $g \in L_2((0, \infty), \omega_{\alpha,\beta})$ für alle $\xi > 0$ gilt, setzen diese in (5.57) ein und erhalten:

$$\int_{\xi}^{\infty} f(t) e^{-\beta t} dt = \sum_{l=0}^{\infty} \hat{f}_l \int_{\xi}^{\infty} \mathcal{L}_l^{(\alpha,\beta)}(t) e^{-\beta t} dt. \quad (5.58)$$

Das auf der rechten Seite stehende Integral lässt sich mit Hilfe von (5.16) durch partielle Integration noch weiter umformen:

$$\begin{aligned} \int_{\xi}^{\infty} \mathcal{L}_{l-1}^{(\alpha,\beta)}(t) e^{-\beta t} dt &= \frac{1}{\beta} e^{-\beta \xi} [\mathcal{L}_l^{(\alpha,\beta)}(\xi) - \mathcal{L}_{l-1}^{(\alpha,\beta)}(\xi)] + \int_{\xi}^{\infty} [\mathcal{L}_{l-1}^{(\alpha,\beta)}(t) - \mathcal{L}_l^{(\alpha,\beta)}(t)] e^{-\beta t} dt \\ &\Rightarrow \int_{\xi}^{\infty} \mathcal{L}_l^{(\alpha,\beta)}(t) e^{-\beta t} dt = \frac{1}{\beta} e^{-\beta \xi} [\mathcal{L}_l^{(\alpha,\beta)}(\xi) - \mathcal{L}_{l-1}^{(\alpha,\beta)}(\xi)], \quad l \geq 1 \end{aligned}$$

Setzen wir die obige Beziehung in (5.58) ein, so erhalten wir zunächst

$$\int_{\xi}^{\infty} f(t) e^{-\beta t} dt = \sum_{l=0}^{\infty} \hat{f}_l \frac{1}{\beta} e^{-\beta \xi} [\mathcal{L}_l^{(\alpha,\beta)}(\xi) - \mathcal{L}_{l-1}^{(\alpha,\beta)}(\xi)]$$

und anschließend durch Multiplikation mit $\beta e^{\beta\xi}$ für alle $\xi \in (0, b)$ die Gleichung

$$\Phi(\xi) = \sum_{l=0}^{\infty} \hat{f}_l \left[\mathcal{L}_l^{(\alpha, \beta)}(\xi) - \mathcal{L}_{l-1}^{(\alpha, \beta)}(\xi) \right], \quad (5.59)$$

wobei $\mathcal{L}_0^{(\alpha, \beta)}(\xi) := 0$ gesetzt ist. Die absolute und gleichmäßige Konvergenz der obigen Reihe für alle $\xi \in (0, b)$ folgt aus der Gültigkeit der erweiterten Parseval'sche Gleichung und kann analog mit den Methoden aus [Tri70], S. 40f. mit Hilfe des Konvergenzkriteriums nach Cauchy nachgewiesen werden. Daher ist eine Umordnung der Reihenglieder erlaubt und führt auf die Reihe in (5.56). \square

Bemerkung 5.2. *Multiplizieren wir die Reihe in (5.56) bzw. in (5.59) mit dem Faktor $e^{-\beta t}$, so folgt die gleichmäßige Konvergenz der Reihe sogar auf dem ganzen Intervall $(0, \infty)$.*

Wir können den vorangegangenen Satz für verschiedene Setzungen der Funktionen Φ und f anwenden. Der nachfolgende Satz zeigt, wie wir die gewonnenen Resultate im Kontext einer Fourier-Laguerre-Reihe in gewichteten Sobolev-Räumen nutzen können. Dafür führen wir einen weiteren nicht-uniform gewichteten Sobolev-Raum ein: Mit der Menge von Gewichtsfunktionen

$$\overline{\omega}_{2, \beta} := \{\omega_\beta, \omega_{1, \beta}, \omega_{1, \beta}\} = \{e^{-\beta t}, te^{-\beta t}, te^{-\beta t}\}.$$

definieren wir den Sobolev-Raum

$$W_2^2((0, \infty), \overline{\omega}_{2, \beta}) := \left\{ f \in L_2((0, \infty), \omega_\beta) \mid \mathcal{D}^1 f, \mathcal{D}^2 f \in L_2((0, \infty), \omega_{1, \beta}) \right\}$$

Im Unterschied zu den bisher verwendeten Sobolev-Räumen $W_2^r((0, \infty), \omega_{r, \beta})$, $r \in \mathbb{N}$ werden nun die distributionellen Ableitungen $\mathcal{D}^1 f, \mathcal{D}^2 f$ mit der gleichen Funktion $\omega_{1, \beta}$ gewichtet. Für eine Funktion $y \in W_2^2((0, \infty), \overline{\omega}_{2, \beta})$ gilt $y \in L_2((0, \infty), \omega_\beta)$ und $\mathcal{D}^1 y, \mathcal{D}^2 y \in L_2((0, \infty), \omega_{1, \beta})$. Aufgrund von $\omega_\beta^{-1}, \omega_{1, \beta}^{-1} \in L_{1, \text{loc}}((0, \infty))$ folgt $y, \mathcal{D}^1 y \in W_{1, \text{loc}}^1((0, \infty))$ und daher die Existenz eines Repräsentanten \tilde{y} mit $\tilde{y}, \tilde{y}' \in AC_{\text{loc}}((0, \infty))$, sodass $y = \tilde{y}$ und $\mathcal{D}^1 y = \tilde{y}'$ fast überall auf $(0, \infty)$ gelten, vgl. Lemma 2.29 und Satz 2.4. In diesem Sinne identifizieren wir im Folgenden die schwache Ableitung $\mathcal{D}^1 y$ der Sobolev-Funktion $y \in W_2^2((0, \infty), \overline{\omega}_{2, \beta})$ mit der klassischen Ableitung \tilde{y}' ihres Repräsentanten. Ist $y \in W_2^2((0, \infty), \omega_\beta) \cap W_2^2((0, \infty), \overline{\omega}_{2, \beta})$, so liefert der Satz von Rellich-Kondrachov zusätzlich eine sinnvolle Definition der Anfangswerte $y(0), y'(0)$, vgl. den Ausführungen auf Seite 127.

Satz 5.10 ([PB19], Theorem 5)

Es sei $y \in W_2^2((0, \infty), \omega_\beta) \cap W_2^2((0, \infty), \overline{\omega}_{2, \beta})$ mit $y(0) = y'(0) = 0$. Dann lässt sich y' in jedem Intervall $(0, b)$, $b > 0$ in die absolut und gleichmäßig konvergente Fourier-Laguerre-Reihe

$$y'(t) = \sum_{l=0}^{\infty} (\hat{f}_l - \hat{f}_{l+1}) \mathcal{L}_l^{(1, \beta)}(t) \quad (5.60)$$

entwickeln, wobei \hat{f}_l die Fourier-Koeffizienten von $f := y' - \frac{1}{\beta}\mathcal{D}^1(y') \in L_2((0, \infty), \omega_{1,\beta})$ bezüglich der verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(1,\beta)}$, $l = 1, \dots$ sind.

Beweis. Es lässt sich leicht nachrechnen, dass $y' \cdot \omega_\beta \in W_1^1([0, \infty))$ aus der Raumzugehörigkeit $y \in W_2^1((0, \infty), \omega_\beta)$ folgt. Aufgrund der Beziehungen

$$y'(t)e^{-\beta t} = \int_0^t \mathcal{D}^1(y'(\tau)e^{-\beta\tau})d\tau, \quad \int_0^\infty \mathcal{D}^1(y'(t)e^{-\beta t})dt = -y'(0) = 0$$

gilt

$$y'(t)e^{-\beta t} = -\int_t^\infty \mathcal{D}^1(y'(\tau)e^{-\beta\tau})d\tau = \beta \int_t^\infty \left[y'(\tau) - \frac{1}{\beta}\mathcal{D}^1(y'(\tau)) \right] e^{-\beta\tau} d\tau.$$

Mit der Setzung

$$f := y' - \frac{1}{\beta}\mathcal{D}^1(y') \tag{5.61}$$

folgt

$$y'(t) = \beta e^{\beta t} \int_t^\infty f(\tau)e^{-\beta\tau}d\tau$$

und wir können nun den Satz 5.9 für $\Phi = y'$ und mit $f \in L_2((0, \infty), \omega_{1,\beta})$ aus (5.61) anwenden. Daraus ergibt sich sofort die punktweise und gleichmäßige Konvergenz der Reihe in (5.60) in jedem Intervall $(0, b)$, $b > 0$. Setzen wir wieder $\mathcal{L}_0^{(1,\beta)}(t) := 0$, so lässt sich die Reihe auch in der Form

$$y'(t) = \sum_{l=0}^\infty \hat{f}_l \left[\mathcal{L}_l^{(1,\beta)}(t) - \mathcal{L}_{l-1}^{(1,\beta)}(t) \right], \tag{5.62}$$

schreiben, vgl. (5.59). □

Mit dem vorangegangenen Satz lässt sich die Frage der gleichmäßigen Konvergenz auf Intervallen $(0, b)$, $b > 0$ der Fourier-Laguerre-Reihe aus (5.60) bzw. (5.62) für Sobolev-Funktionen $y \in W_2^2((0, \infty), \omega_\beta) \cap W_2^2((0, \infty), \overline{\omega}_{2,\beta})$ mit $y(0) = y'(0) = 0$ beantworten. Über das Konvergenzverhalten dieser Reihe in $t = 0$ haben wir bisher jedoch noch nichts erfahren. Erörtert haben wir diese Fragestellung in der Publikation [PB19], vgl. Lemma 6. Auf die Angabe des Beweises verzichten wir an dieser Stelle und verweisen auf die angegebene Quelle. Das Resultat jedoch formulieren wir in dem nachfolgendem Lemma.

Lemma 5.11 ([PB19], Lemma 6)

Unter den Voraussetzungen des Satzes 5.10 konvergiert die Fourier-Laguerre-Reihe in (5.62)

gleichmäßig in jedem Intervall $[0, b)$, $b > 0$ mit

$$\hat{f}_0 = - \sum_{l=1}^{\infty} \hat{f}_l$$

und es gilt

$$y(t) = t \left[\hat{f}_0 + \sum_{l=1}^{\infty} \frac{\hat{f}_l}{l+1} \mathcal{L}_l^{(1,\beta)}(t) \right]. \quad (5.63)$$

Die gleichmäßige Konvergenz der Reihe in (5.62) auf jedem Intervall $[0, t)$, $t > 0$ erlaubt die gliedweise Integration, woraus sich für $y \in W_1^1([0, t))$ die nachfolgenden Rechnungen ergeben:

$$\begin{aligned} y(t) &= \int_0^t y'(\tau) d\tau = \hat{f}_0 t + \sum_{l=1}^{\infty} \hat{f}_l \int_0^t [\mathcal{L}_l^{(1,\beta)}(\tau) - \mathcal{L}_{l-1}^{(1,\beta)}(\tau)] d\tau \\ &= \hat{f}_0 t + \frac{1}{\beta} \sum_{l=1}^{\infty} \hat{f}_l [\mathcal{L}_l^{(\beta)}(t) - \mathcal{L}_{l+1}^{(\beta)}(t)] - \frac{1}{\beta} \sum_{l=1}^{\infty} \hat{f}_l [\mathcal{L}_l^{(\beta)}(0) - \mathcal{L}_{l+1}^{(\beta)}(0)] \\ &= \hat{f}_0 t + \frac{1}{\beta} \sum_{l=1}^{\infty} \hat{f}_l [\mathcal{L}_l^{(\beta)}(t) - \mathcal{L}_{l+1}^{(\beta)}(t)]. \end{aligned}$$

Nutzen wir den folgenden Zusammenhang zwischen den verallgemeinerten Laguerre-Polynomen (vgl. [ByXy05], S. 284)

$$-\beta t \mathcal{L}_l^{(1,\beta)}(t) = (l+1) [\mathcal{L}_l^{(\beta)}(t) - \mathcal{L}_{l+1}^{(\beta)}(t)],$$

so ergibt sich die Darstellung in (5.63).

Bemerkung 5.3. Neben der in dieser Arbeit vorgestellten Spektralmethode zur Lösung von Steuerungsproblemen über dem unendlichen Zeithorizont $[0, \infty)$ wurde in unserer Publikation [PB19] eine weitere Spektralmethode vorgestellt. Dabei verfolgten wir den Ansatz, die zu einer Aufgabe optimaler Steuerung vom Typ $(LQ)_{(\infty)}$ duale Aufgabe zu konstruieren (siehe auch [PBK16]) und diese dann durch einen speziell gewählten Fourier-Laguerre-Reihenansatz zu approximieren. Die resultierende duale Aufgabe ist eine Variationsaufgabe, formuliert in dem Durchschnitt der gewichteten Sobolev-Räume $W_2^2((0, \infty), \omega_\beta) \cap W_2^2((0, \infty), \overline{\omega}_{2,\beta})$, welche wir direkt in eine endlich dimensionale Optimierungsaufgabe überführen können. Bei beiden Methoden führt die Wahl der gewichteten Sobolev-Räume zu wohldefinierten Aufgaben, für welche, basierend auf den theoretischen Vorarbeiten, Fourier-Laguerre-Ansätze für numerische Verfahren über dem unendlichen Horizont $[0, \infty)$ natürlich erscheinen. Dabei wird bei keinem der Verfahren der unendliche Zeithorizont abgeschnitten oder dieser auf einen endlichen transformiert. Für detaillierte Informationen über das genaue Konstruktions- und Approximationsschema basierend auf der Dualitätstheorie verweisen wir auf die genannte Publikation.

5.2 Formulierung der Methode für $(LQ)_\infty$

Wir betrachten die linear-quadratische Steuerungsaufgabe $(LQ)_\infty$, vgl. S. 71 mit $U = \mathbb{R}^m$ und unter Gültigkeit der Voraussetzungen (V1)-(V3), (V5)-(V6). Dann besitzt die Aufgabe $(LQ)_\infty$ eine eindeutig bestimmte Lösung $(x^*, u^*) \in W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta)$, vgl. Satz 4.13. Um diese zu ermitteln, können wir entsprechend den Ausführungen in Abschnitt 4.6, S. 111-113 in zwei Schritten vorgehen:

Schritt 1: Finde ein Funktionenpaar $(x, y) \in W_2^{1,n}((0, \infty), \omega_\beta) \times W_2^{1,n}((0, \infty), \omega_\beta)$ mit $x(0) = x_0$, welches das System

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} A(t) & B(t)R^{-1}(t)B^T(t) \\ W(t) & \beta \cdot I_{n \times n} - A^T(t) \end{pmatrix} \cdot \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} + \begin{pmatrix} C(t) - B(t)R^{-1}(t)r(t) \\ w(t) \end{pmatrix}$$

fast überall auf $(0, \infty)$ löst.

Schritt 2: Bestimme die zugehörige Steuerung $u \in L_2((0, \infty), \omega_\beta)$ aus der Gleichung

$$u(t) = R^{-1}(t) \left[B^T(t)y(t) - r(t) \right] \quad \text{f.ü. auf } (0, \infty). \quad (5.64)$$

Bemerkung 5.4.

1. Die Existenz einer Lösung des Differentialgleichungssystems in Schritt 1 ist durch die Existenz der Lösung $(x^*, u^*) \in W_2^{1,n}((0, \infty), \omega_\beta) \times L_2^m((0, \infty), \omega_\beta)$ der zugrunde liegenden Steuerungsaufgabe und durch die Gültigkeit des Maximumprinzips gesichert.
2. Das in Schritt 1 und 2 gefundene Paar (x, u) ist die gesuchte und eindeutig bestimmte Lösung der Steuerungsaufgabe, $(x, u) = (x^*, u^*)$.

Wir möchten das in Schritt 1 zu lösende System mit Hilfe von Spektralmethoden numerisch lösen. Konkret schlagen wir eine Kollokationsmethode vor, welche einfach vom Nutzer anzuwenden und zu implementieren ist. Der Anfangswert $y(0) = d \in \mathbb{R}^n$ ist dabei ein noch zu bestimmender Parameter. Wir werden zunächst zeigen, wie wir die Transversalitätsbedingung geschickt mit den kanonischen Gleichungen verknüpfen können, um diesen festzulegen. Dabei verwenden wir im Folgenden die Notation $M_{i,j}(t)$ für den Eintrag einer Matrix $M(t)$ in der Zeile i und in der Spalte j . Mit $v_i(t)$ meinen wir die i -te Komponente eines Vektors $v(t)$.

5.2.1 Anfangswert der Adjungierten

Für die Bestimmung des Anfangswertes $y(0) = d \in \mathbb{R}^n$ nutzen wir die durch das Maximumprinzip gewonnene Transversalitätsbedingung, aus welcher sich die Raumzugehörigkeit $y \in W_2^{1,n}((0, \infty), \omega_\beta)$ ergibt. Wir demonstrieren die Herleitung für eine Komponentenfunktion $y^{(\nu)} \in W_2^1((0, \infty), \omega_\beta)$ mit $\nu \in \{1, \dots, n\}$. Aus der Raumzugehörigkeit von $y^{(\nu)}$ folgt,

dass $y^{(\nu)} \cdot \omega_\beta \in W_1^1((0, \infty))$ gilt und wir können den Zusammenhang

$$\int_0^\infty \left[y^{(\nu)}(t) \omega_\beta(t) \right]' dt = -y^{(\nu)}(0) \cdot \omega_\beta(0)$$

aus Bemerkung 5.1 verwenden. Da weiterhin $\dot{\omega}_\beta(t) = -\beta \cdot \omega_\beta(t)$ und $\omega_\beta(0) = 1$ gelten, ergibt sich daraus die folgende Gleichung:

$$\int_0^\infty \dot{y}^{(\nu)}(t) \omega_\beta(t) dt - \beta \int_0^\infty y^{(\nu)}(t) \omega_\beta(t) dt = -y^{(\nu)}(0). \quad (5.65)$$

Die kanonische Gleichung für $y^{(\nu)}$ lautet

$$\dot{y}^{(\nu)}(t) = \sum_{j=1}^n W_{\nu,j}(t) x^{(j)}(t) - \sum_{j=1}^n A_{\nu,j}(t) y^{(j)}(t) + \beta y^{(\nu)}(t) + w_\nu(t) \quad \text{f.ü. auf } (0, \infty).$$

Einsetzen dieser Gleichung in (5.65) führt auf die Beziehung

$$\begin{aligned} & \int_0^\infty \left\{ \sum_{j=1}^n W_{\nu,j}(t) x^{(j)}(t) - \sum_{j=1}^n A_{\nu,j}(t) y^{(j)}(t) + \beta y^{(\nu)}(t) + w_\nu(t) \right\} \omega_\beta(t) dt - \beta \int_0^\infty y^{(\nu)}(t) \omega_\beta(t) dt \\ &= \int_0^\infty \left\{ \sum_{j=1}^n W_{\nu,j}(t) x^{(j)}(t) - \sum_{j=1}^n A_{\nu,j}(t) y^{(j)}(t) + w_\nu(t) \right\} \omega_\beta(t) dt = -y^{(\nu)}(0) \end{aligned}$$

und somit zu

$$d_\nu = \int_0^\infty \left\{ \sum_{j=1}^n A_{\nu,j}(t) y^{(j)}(t) - \sum_{j=1}^n W_{\nu,j}(t) x^{(j)}(t) - w_\nu(t) \right\} \omega_\beta(t) dt. \quad (5.66)$$

Die zunächst noch als unbestimmt angenommenen Anfangswerte $d = (d_1, d_2, \dots, d_N)^T$, $d_\nu \in \mathbb{R}$, $\nu = 1, \dots, n$ lassen sich nun durch die obige Gleichung festlegen. Wir nutzen dieses Wissen an späterer Stelle für die Formulierung von Anfangswerten für die Näherungslösung der Adjungierten.

5.2.2 Kollokationsgleichungen

Für die numerische Lösung des Differentialgleichungssystems approximieren wir die gesuchten Funktionen

$$\begin{aligned} x &= (x^{(1)}, \dots, x^{(n)})^T \in W_2^{1,n}((0, \infty), \omega_\beta) \\ y &= (y^{(1)}, \dots, y^{(n)})^T \in W_2^{1,n}((0, \infty), \omega_\beta), \nu = 1, \dots, n \end{aligned}$$

durch Polynomfunktionen

$$x_N^{(1)}, \dots, x_N^{(n)}; y_N^{(1)}, \dots, y_N^{(n)} \in \mathcal{P}_N((0, \infty)).$$

Da $\mathcal{P}_N((0, \infty)) = \text{lin} \{ \mathcal{L}_0^{(\beta)}, \mathcal{L}_1^{(\beta)}, \dots, \mathcal{L}_N^{(\beta)} \}$ gilt, wählen wir die verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(\beta)}$, $l = 0, \dots, N$ als Basisfunktionen, woraus sich Ansätze der Form

$$\begin{aligned} x^{(\nu)}(t) &\approx x_N^{(\nu)}(t) = \sum_{l=0}^N a_l^{(\nu)} \mathcal{L}_l^{(\beta)}(t) \\ y^{(\nu)}(t) &\approx y_N^{(\nu)}(t) = \sum_{l=0}^N b_l^{(\nu)} \mathcal{L}_l^{(\beta)}(t), \quad \nu = 1, \dots, n \end{aligned} \quad (5.67)$$

mit den noch unbekanntenen Koeffizienten $a_l^{(\nu)}, b_l^{(\nu)} \in \mathbb{R}$, $l = 0, \dots, N$ ergeben. Für die Näherungslösungen fordern wir, dass sie die kanonischen Gleichungen an fest vorgegebenen Kollokationspunkten exakt erfüllen. Mit den Bezeichnungen

$$\widehat{B}(t) := (B(t)R^{-1}(t)B^T(t)) \in \mathbb{R}^{n \times n}, \quad \widehat{c}(t) := (C(t) - B(t)R^{-1}(t)r(t)) \in \mathbb{R}^n$$

ergeben sich für $i = 1, \dots, N$ die folgenden Kollokationsgleichungen:

$$\begin{aligned} \dot{x}_N^{(\nu)}(\tau_i^{(N)}) &= \sum_{j=1}^N A_{\nu,j}(\tau_i^{(N)}) \cdot x_N^{(j)}(\tau_i^{(N)}) + \sum_{j=1}^N \widehat{B}_{\nu,j}(\tau_i^{(N)}) \cdot y_N^{(j)}(\tau_i^{(N)}) + \widehat{c}_\nu(\tau_i^{(N)}), \\ \dot{y}_N^{(\nu)}(\tau_i^{(N)}) &= \sum_{j=1}^N W_{\nu,j}(\tau_i^{(N)}) \cdot x_N^{(j)}(\tau_i^{(N)}) - \sum_{j=1}^N A_{\nu,j}^T(\tau_i^{(N)}) \cdot y_N^{(j)}(\tau_i^{(N)}) + \beta y_N^{(\nu)}(\tau_i^{(N)}) \\ &\quad + w_\nu(\tau_i^{(N)}). \end{aligned}$$

Als Kollokationspunkte $\tau_i^{(N)}$, $i = 1, \dots, N$ wählen wir die Nullstellen des N -ten Laguerre-Polynoms $\mathcal{L}_N^{(\beta)}$. Aus (5.67) folgt dann

$$x_N^{(\nu)}(\tau_i^{(N)}) = \sum_{l=0}^{N-1} a_l^{(\nu)} \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}), \quad y_N^{(\nu)}(\tau_i^{(N)}) = \sum_{l=0}^{N-1} b_l^{(\nu)} \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}), \quad \nu = 1, \dots, n \quad (5.68)$$

und differenzieren liefert

$$\dot{x}_N^{(\nu)}(\tau_i^{(N)}) = \sum_{l=1}^N a_l^{(\nu)} \frac{d}{dt} \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}), \quad \dot{y}_N^{(\nu)}(\tau_i^{(N)}) = \sum_{l=1}^N b_l^{(\nu)} \frac{d}{dt} \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}), \quad \nu = 1, \dots, n, \quad (5.69)$$

wobei wir verwendet haben, dass $\frac{d}{dt} \mathcal{L}_0^{(\beta)}(t) = 0$ gilt.

Darüber hinaus fügen wir den zusätzlichen Kollokationspunkt $\tau_0 = 0$ ein und fordern für die Näherungslösungen der Zustände, dass die Bedingung

$$x_N(\tau_0) := (x_N^{(1)}(\tau_0), x_N^{(2)}(\tau_0), \dots, x_N^{(n)}(\tau_0))^T = x_0$$

erfüllt ist. Für die Bestimmung der Anfangswerte $y^{(\nu)}(\tau_0)$ lassen wir die Gleichungen aus (5.66) in unser Kollokationsverfahren einfließen. Dazu diskretisieren wir die auftauchende Integralgleichung mit Hilfe der Gaußschen-Quadraturformel. Für die Kollokationspunkte $\tau_i^{(N)}$, $i = 1, \dots, N$ existiert eine Menge von Quadraturgewichten $w_i^{(N)}$, $i = 1, \dots, N$, sodass die Gaußsche Quadraturformel für alle Polynome vom Grad $\leq 2N - 1$ exakt ist, vgl. Satz 3.22, S. 69. Entsprechend fordern wir, dass die Polynome $y_N^{(\nu)} \in \mathcal{P}_N((0, \infty))$, $\nu = 1, \dots, n$ den diskretisierten Bedingungen

$$y_N^{(\nu)}(0) = \sum_{i=1}^N \left\{ \sum_{j=1}^N A_{\nu,j}(\tau_i^{(N)}) \cdot y_N^{(j)}(\tau_i^{(N)}) - \sum_{j=1}^N W_{\nu,j}(\tau_i^{(N)}) \cdot x_N^{(j)}(\tau_i^{(N)}) - w_\nu(\tau_i^{(N)}) \right\} w_i^{(N)}$$

genügen. Zusammengefasst lässt sich die Kollokationsmethode für das kanonische Differentialgleichungssystem aus Schritt 1 in der folgenden Form schreiben: Finde Polynome $x_N^{(\nu)}, y_N^{(\nu)} \in \mathcal{P}_N((0, \infty))$, $\nu = 1, \dots, n$, welche den Gleichungen

$$\left\{ \begin{array}{l} \dot{x}_N^{(\nu)}(\tau_i^{(N)}) = \sum_{j=1}^n A_{\nu,j}(\tau_i^{(N)}) \cdot x_N^{(j)}(\tau_i^{(N)}) + \sum_{j=1}^n \widehat{B}_{\nu,j}(\tau_i^{(N)}) \cdot y_N^{(j)}(\tau_i^{(N)}) + \widehat{c}_\nu(\tau_i^{(N)}), \quad 1 \leq i \leq N \\ \dot{y}_N^{(\nu)}(\tau_i^{(N)}) = \sum_{j=1}^n W_{\nu,j}(\tau_i^{(N)}) \cdot x_N^{(j)}(\tau_i^{(N)}) - \sum_{j=1}^n A_{\nu,j}^T(\tau_i^{(N)}) \cdot y_N^{(j)}(\tau_i^{(N)}) + \beta y_N^{(\nu)}(\tau_i^{(N)}) \\ \quad + w_\nu(\tau_i^{(N)}), \quad 1 \leq i \leq N \\ x_N(\tau_0) = x_0 \\ y_N^{(\nu)}(\tau_0) = \sum_{i=1}^N \left\{ \sum_{j=1}^N A_{\nu,j}(\tau_i^{(N)}) \cdot y_N^{(j)}(\tau_i^{(N)}) - \sum_{j=1}^N W_{\nu,j}(\tau_i^{(N)}) \cdot x_N^{(j)}(\tau_i^{(N)}) - w_\nu(\tau_i^{(N)}) \right\} w_i^{(N)} \end{array} \right.$$

für alle $\nu = 1, \dots, n$ genügen. Durch Einsetzen der Ansätze aus (5.68) und (5.69) in die obigen Kollokationsgleichungen ergeben sich die Bestimmungsgleichungen für die unbekanntenen Koeffizienten in (5.67). Das gegebene kanonische Differentialgleichungssystem besteht aus insgesamt $2n$ Differentialgleichungen. Aus einer Differentialgleichung resultieren N Kollokationsgleichungen, das bedeutet, die $2n$ Differentialgleichungen werden durch die Diskretisierung in insgesamt $2n \cdot N$ algebraische Gleichungen überführt. Die Anfangswertbedingungen liefern zusätzlich $2n$ Gleichungen. Insgesamt resultiert daher aus dem Kollokationsansatz ein lineares algebraisches Gleichungssystem mit $2n \cdot (N + 1)$ Gleichungen zur Bestimmung der $2n \cdot (N + 1)$ Unbekannten $a_0^{(\nu)}, a_1^{(\nu)}, \dots, a_N^{(\nu)}; b_0^{(\nu)}, b_1^{(\nu)}, \dots, b_N^{(\nu)}$, $\nu = 1, \dots, n$. Die Approximation der Steuerung $u_N := (u_N^{(1)}, u_N^{(2)}, \dots, u_N^{(n)})^T$ lässt sich nach dem Lösen der Kollokationsgleichungen direkt mit Hilfe der aus der Maximumbedingung gewonnenen Beziehung (5.64) ermitteln:

$$u_N(t) = R^{-1}(t) \left[B^T(t) y_N(t) - r(t) \right].$$

5.2.3 Aufstellen der Koeffizientenmatrix

Wir haben bereits erwähnt, dass sich die Kollokationsgleichungen zu einem in $a_l, b_l \in \mathbb{R}$, $l = 0, \dots, N$ linearen Gleichungssystem umformen lassen. Die korrespondierende Koeffizientenmatrix lässt sich dabei durch Einsetzen der Ansätze aus (5.68) und (5.69) ermitteln. Wir demonstrieren das Aufstellen dieser Matrix für eine bessere Übersicht für den Fall eines skalaren Zustandes $x \in W_2^1((0, \infty), \omega_\beta)$ und einer skalaren adjungierten Funktion $y \in W_2^1((0, \infty), \omega_\beta)$. Mit den Ansätzen

$$x(t) \approx x_N(t) = \sum_{l=0}^N a_l \mathcal{L}_l^{(\beta)}(t), \quad y(t) \approx y_N(t) = \sum_{l=0}^N b_l \mathcal{L}_l^{(\beta)}(t)$$

reduziert sich das diskretisierte Gleichungssystem auf $2(N+1)$ Gleichungen. Aus der diskretisierten Zustandsgleichung

$$\dot{x}_N(\tau_i^{(N)}) = A(\tau_i^{(N)}) \cdot x_N(\tau_i^{(N)}) + \widehat{B}(\tau_i^{(N)}) \cdot y_N(\tau_i^{(N)}) + \widehat{c}(\tau_i^{(N)}), \quad 1 \leq i \leq N$$

erhalten wir durch Einsetzen der Polynomansätze und unter Beachtung von (5.68) und (5.69) für $1 \leq i \leq N$ die Gleichungen

$$\sum_{l=1}^N a_l \frac{d}{dt} \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}) = A(\tau_i^{(N)}) \cdot \sum_{l=0}^{N-1} a_l \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}) + \widehat{B}(\tau_i^{(N)}) \cdot \sum_{l=0}^{N-1} b_l \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}) + \widehat{c}(\tau_i^{(N)})$$

und Umstellen liefert

$$\begin{aligned} & -A(\tau_i^{(N)}) \cdot \mathcal{L}_0^{(\beta)}(\tau_i^{(N)}) \cdot a_0 + \sum_{l=1}^{N-1} \left(\frac{d}{dt} \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}) - A(\tau_i^{(N)}) \cdot \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}) \right) \cdot a_l \\ & + \frac{d}{dt} \mathcal{L}_N^{(\beta)}(\tau_i^{(N)}) \cdot a_N - \sum_{l=0}^{N-1} \widehat{B}(\tau_i^{(N)}) \cdot \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}) \cdot b_l = \widehat{c}(\tau_i^{(N)}), \quad 1 \leq i \leq N. \end{aligned}$$

Für die diskretisierte Differentialgleichung der Adjungierten

$$\dot{y}_N(\tau_i^{(N)}) = W(\tau_i^{(N)}) \cdot x_N(\tau_i^{(N)}) + (\beta - A(\tau_i^{(N)})) \cdot y_N(\tau_i^{(N)}) + w(\tau_i^{(N)}), \quad 1 \leq i \leq N$$

gehen wir analog vor und erhalten

$$\begin{aligned} & - \sum_{l=0}^{N-1} W(\tau_i^{(N)}) \cdot \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}) \cdot a_l + (A(\tau_i^{(N)}) - \beta) \cdot \mathcal{L}_0^{(\beta)}(\tau_i^{(N)}) \cdot b_0 \\ & + \sum_{l=1}^{N-1} \left(\frac{d}{dt} \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}) + [A(\tau_i^{(N)}) - \beta] \cdot \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}) \right) \cdot b_l \\ & + \frac{d}{dt} \mathcal{L}_N^{(\beta)}(\tau_i^{(N)}) \cdot b_N = w(\tau_i^{(N)}), \quad 1 \leq i \leq N. \end{aligned}$$

Für den Anfangswert $x_N(\tau_0)$ der Näherungslösung x_N ergibt sich aufgrund von $\mathcal{L}_l^{(\beta)}(0) = 1$, $l = 0, 1, \dots$, vgl. Lemma 5.2, die Bedingung

$$x_N(\tau_0) = x_N(0) = \sum_{l=0}^N a_l \mathcal{L}_l^{(\beta)}(0) \Rightarrow \sum_{l=0}^N a_l = x_0.$$

Für den Anfangswert $y_N(\tau_0)$ erhalten wir aus

$$y_N(\tau_0) = y_N(0) = \sum_{i=1}^N \left\{ A(\tau_i^{(N)}) \cdot y_N(\tau_i^{(N)}) - W(\tau_i^{(N)}) \cdot x_N(\tau_i^{(N)}) - w(\tau_i^{(N)}) \right\} \mathbf{w}_i^{(N)}$$

die Gleichung

$$\begin{aligned} \sum_{l=0}^{N-1} \left(\sum_{i=1}^N W(\tau_i^{(N)}) \cdot \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}) \cdot \mathbf{w}_i^{(N)} \right) \cdot a_l + \sum_{l=0}^{N-1} \left(1 - \sum_{i=1}^N A(\tau_i^{(N)}) \cdot \mathcal{L}_l^{(\beta)}(\tau_i^{(N)}) \cdot \mathbf{w}_i^{(N)} \right) \cdot b_l \\ + b_N = - \sum_{i=1}^N w(\tau_i^{(N)}) \cdot \mathbf{w}_i^{(N)}. \end{aligned}$$

Es resultiert ein Gleichungssystem der Form

$$M^N \cdot (a_0, a_1, \dots, a_N; b_0, b_1, \dots, b_N)^T = z^N$$

mit dem Vektor $z^N \in \mathbb{R}^{2N+2}$, dessen Einträge durch

$$z_i^N := \begin{cases} \hat{c}(\tau_i^{(N)}) & 1 \leq i \leq N, \\ w(\tau_{i-N}^{(N)}) & N+1 \leq i \leq 2N, \\ x_0 & i = 2N+1, \\ - \sum_{k=1}^N w(\tau_k^{(N)}) \cdot \mathbf{w}_k^{(N)} & i = 2N+2 \end{cases}$$

gegeben sind. Die einzelnen Einträge der $(2N+2) \times (2N+2)$ Matrix M^N sowie eine gesamte Darstellung der Koeffizientenmatrix M^N sind auf den nachfolgenden Seiten abgebildet. Dabei haben wir zur Abkürzung die Notationen $\tau_i := \tau_i^{(N)}$ und $\mathbf{w}_i := \mathbf{w}_i^{(N)}$, $i = 1, \dots, N$ verwendet.

$$M_{i,j}^N := \begin{cases} -A(\tau_i^{(N)}) \cdot \mathcal{L}_0^{(\beta)}(\tau_i^{(N)}) & 1 \leq i \leq N, j = 1, \\ \frac{d}{dt} \mathcal{L}_{j-1}^{(\beta)}(\tau_i^{(N)}) - A(\tau_i^{(N)}) \cdot \mathcal{L}_{j-1}^{(\beta)}(\tau_i^{(N)}) & 1 \leq i \leq N, 2 \leq j \leq N, \\ \frac{d}{dt} \mathcal{L}_N^{(\beta)}(\tau_i^{(N)}) & 1 \leq i \leq N, j = N + 1, \\ -\widehat{B}(\tau_i^{(N)}) \cdot \mathcal{L}_{j-N-2}^{(\beta)}(\tau_i^{(N)}) & 1 \leq i \leq N, \\ & N + 2 \leq j \leq 2N + 1, \\ 0 & 1 \leq i \leq N, j = 2N + 2, \\ -W(\tau_{i-N}^{(N)}) \cdot \mathcal{L}_{j-1}^{(\beta)}(\tau_{i-N}^{(N)}) & N + 1 \leq i \leq 2N, 1 \leq j \leq N, \\ 0 & N + 1 \leq i \leq 2N, j = N + 1, \\ [A(\tau_{i-N}^{(N)}) - \beta] \cdot \mathcal{L}_0^{(\beta)}(\tau_{i-N}^{(N)}) & N + 1 \leq i \leq 2N, j = N + 2, \\ \frac{d}{dt} \mathcal{L}_{j-N-2}^{(\beta)}(\tau_{i-N}^{(N)}) + [A(\tau_{i-N}^{(N)}) - \beta] \cdot \mathcal{L}_{j-N-2}^{(\beta)}(\tau_{i-N}^{(N)}) & N + 1 \leq i \leq 2N, \\ & N + 3 \leq j \leq 2N + 1, \\ \frac{d}{dt} \mathcal{L}_N^{(\beta)}(\tau_{i-N}^{(N)}) & N + 1 \leq i \leq 2N, j = 2N + 2, \\ 1 & i = 2N + 1, 1 \leq j \leq N + 1, \\ 0 & i = 2N + 1, \\ & N + 2 \leq j \leq 2N + 2, \\ \sum_{k=1}^N W(\tau_k^{(N)}) \cdot \mathcal{L}_{j-1}^{(\beta)}(\tau_k^{(N)}) \cdot \mathbf{w}_k^{(N)} & i = 2N + 2, 1 \leq j \leq N, \\ 1 - \sum_{k=1}^N A(\tau_k^{(N)}) \cdot \mathcal{L}_{j-N-2}^{(\beta)}(\tau_k^{(N)}) \cdot \mathbf{w}_k^{(N)} & i = 2N + 2, \\ & N + 2 \leq j \leq 2N + 1, \\ 1 & i = j = 2N + 2. \end{cases}$$

$$\begin{pmatrix}
\mathbf{1} & j=2, \dots, N & N+1 & N+2 & j=N+3, \dots, 2N+1 & 2N+2 \\
-A(\tau_1) \cdot \mathcal{L}_0^{(\beta)}(\tau_1) & \frac{d}{dt} \mathcal{L}_{j-1}^{(\beta)}(\tau_1) - A(\tau_1) \cdot \mathcal{L}_{j-1}^{(\beta)}(\tau_1) & \frac{d}{dt} \mathcal{L}_N^{(\beta)}(\tau_1) & -\widehat{B}(\tau_1) \cdot \mathcal{L}_0^{(\beta)}(\tau_1) & -\widehat{B}(\tau_1) \cdot \mathcal{L}_{j-N-2}^{(\beta)}(\tau_1) & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
-A(\tau_N) \cdot \mathcal{L}_0^{(\beta)}(\tau_N) & \frac{d}{dt} \mathcal{L}_{j-1}^{(\beta)}(\tau_N) - A(\tau_N) \cdot \mathcal{L}_{j-1}^{(\beta)}(\tau_N) & \frac{d}{dt} \mathcal{L}_N^{(\beta)}(\tau_N) & -\widehat{B}(\tau_N) \cdot \mathcal{L}_0^{(\beta)}(\tau_N) & -\widehat{B}(\tau_N) \cdot \mathcal{L}_{j-N-2}^{(\beta)}(\tau_N) & 0 \\
N+1 & -W(\tau_1) \cdot \mathcal{L}_0^{(\beta)}(\tau_1) & -W(\tau_1) \cdot \mathcal{L}_{j-1}^{(\beta)}(\tau_1) & [A(\tau_1) - \beta] \cdot \mathcal{L}_0^{(\beta)}(\tau_1) & \frac{d}{dt} \mathcal{L}_{j-N-2}^{(\beta)}(\tau_1) + [A(\tau_1) - \beta] \cdot \mathcal{L}_{j-N-2}^{(\beta)}(\tau_1) & \frac{d}{dt} \mathcal{L}_{j-N-2}^{(\beta)}(\tau_1) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
2N & -W(\tau_N) \cdot \mathcal{L}_0^{(\beta)}(\tau_N) & -W(\tau_N) \cdot \mathcal{L}_{j-1}^{(\beta)}(\tau_N) & [A(\tau_N) - \beta] \cdot \mathcal{L}_0^{(\beta)}(\tau_N) & \frac{d}{dt} \mathcal{L}_{j-N-2}^{(\beta)}(\tau_N) + [A(\tau_N) - \beta] \cdot \mathcal{L}_{j-N-2}^{(\beta)}(\tau_N) & \frac{d}{dt} \mathcal{L}_{j-N-2}^{(\beta)}(\tau_N) \\
2N+1 & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
2N+2 & \sum_{k=1}^N W(\tau_k) \cdot \mathcal{L}_0^{(\beta)}(\tau_k) \cdot \mathbf{w}_k & \sum_{k=1}^N W(\tau_k) \cdot \mathcal{L}_{j-1}^{(\beta)}(\tau_k) \cdot \mathbf{w}_k & \mathbf{1} - \sum_{k=1}^N A(\tau_k) \cdot \mathcal{L}_0^{(\beta)}(\tau_k) \cdot \mathbf{w}_k & \mathbf{1} - \sum_{k=1}^N A(\tau_k) \cdot \mathcal{L}_{j-N-2}^{(\beta)}(\tau_k) \cdot \mathbf{w}_k & \mathbf{1}
\end{pmatrix}$$

6 Beispiele

Alle numerischen Untersuchungen und Berechnungen wurden mit Hilfe der Software MATLAB R2017a und Maple 2017 auf einem Windows System mit einem Intel(R) Core(TM) i7-4600 U CPU @ 2.10 GHz und 8 GB RAM durchgeführt.

6.1 Beispiel von Garg, Hager und Rao

Das folgende Beispiel ist der Arbeit [GHR11] von Garg, Hager und Rao entnommen. Betrachtet wird ein nichtlineares Problem optimaler Steuerung mit einem unendlichen Zeithorizont:

$$J(z, \tilde{u}) = \frac{1}{2} \int_0^{\infty} \{\ln^2 z(t) + \tilde{u}(t)^2\} dt \rightarrow \text{Min} !$$
$$\dot{z}(t) = z(t) \ln z(t) + z(t) \tilde{u}(t), \quad z(0) = e^2, \quad \tilde{u}(t) \in \mathbb{R}.$$

Mit der in [PW12] vorgeschlagenen logarithmischen Zustandstransformation $\tilde{x}(t) := \ln z(t)$ lässt sich die obige Aufgabe in eine linear-quadratische Aufgabe optimaler Steuerung umformen und wir erhalten:

$$J(\tilde{x}, \tilde{u}) = \frac{1}{2} \int_0^{\infty} \{\tilde{x}(t)^2 + \tilde{u}(t)^2\} dt \rightarrow \text{Min} !$$
$$\dot{\tilde{x}}(t) = \tilde{x}(t) + \tilde{u}(t), \quad \tilde{x}(0) = 2, \quad \tilde{u}(t) \in \mathbb{R}.$$

Eine zweite Transformation des Zustandes $x(t) := \tilde{x}(t)e^t$ und der Steuerung $u(t) := \tilde{u}(t)e^t$ führt auf ein linear-quadratisches Steuerungsproblem vom Typ $(LQ)_{\infty}$ mit einem gewichteten Zielfunktional. Die transformierte Aufgabe lautet:

$$J(x, u) = \frac{1}{2} \int_0^{\infty} \{x^2(t) + u^2(t)\} e^{-2t} dt \rightarrow \text{Min} !$$
$$\dot{x}(t) = 2x(t) + u(t), \quad x(0) = 2, \quad u(t) \in \mathbb{R}. \quad (6.1)$$

Die in dem Zielfunktional auftauchende Gewichtsfunktion $\omega_2(t) := e^{-2t}$ motiviert uns, nach Lösungen in gewichteten Funktionenräumen zu suchen. Entsprechend studieren wir die Aufgabe bezüglich

$$(x, u) \in W_2^1((0, \infty), \omega_2) \times L_2((0, \infty), \omega_2), \quad \omega_2(t) := e^{-2t}.$$

Mit $W = R = B = 1$, $w = r = C = 0$ und $A = 2$ entspricht diese Aufgabe der Problemklasse $(LQ)_\infty$ mit skalarlem Zustand und skalarer Steuerung, wobei der Parameter β in der Gewichtsfunktion $\omega_\beta(t) = e^{-\beta t}$ in diesem Beispiel fest gewählt ist, also $\beta = 2$ gilt. Das Beispiel lässt sich mit der in den vorangegangenen Kapiteln vorgestellten Theorie studieren. Die in Abschnitt 4.2 angegebenen Voraussetzungen lassen sich für die Aufgabe in (6.1) leicht verifizieren, insbesondere vereinfacht sich die Forderung (V3b) zu

$$(\underline{R} - \underline{W} \cdot \frac{B^2}{A^2}) > 0$$

und ist mit $\underline{R} = \underline{W} = 1$ erfüllt. Die Existenz einer eindeutig bestimmten Lösung $(x^*, u^*) \in W_2^1((0, \infty), \omega_2) \times L_2((0, \infty), \omega_2)$ ist aufgrund des Satzes 4.13 garantiert. Darüber hinaus gilt (V6) mit $\underline{B} = 1$ und wir können die Bedingungen des Maximumprinzips verwenden, um einen Lösungskandidaten zu ermitteln.

6.1.1 Diskussion der notwendigen Bedingungen

Mit der Pontryaginschen Funktion

$$H(t, \xi, v, \eta, 1) = -\frac{1}{2}(\xi^2 + v^2)e^{-2t} + \eta(2\xi + v)$$

lassen sich die Bedingungen des Satzes 4.14 für unser Beispiel wie folgt formulieren: Ist $(x^*, u^*) \in W_2^1((0, \infty), \omega_2) \times L_2((0, \infty), \omega_2)$ optimal in der Aufgabe aus (6.1) im Sinne der Definition 4.1, dann existieren Multiplikatoren (λ_0, y_0) mit

$$\lambda_0 = 1, \tag{N}$$

$$y_0 \in W_2^1((0, \infty), \omega_2^{-1}), \omega_2^{-1}(t) = e^{2t}, \tag{T}$$

sodass die Maximumbedingung

$$H(t, x^*(t), u^*(t), y_0(t), \lambda_0) = \max_{v \in \mathbb{R}} H(t, x^*(t), v, y_0(t), \lambda_0) \quad \text{f.ü. auf } (0, \infty) \tag{M}$$

und die kanonische Gleichung

$$\dot{y}_0(t) = x^*(t)e^{-2t} - 2y_0(t) \quad \text{f.ü. auf } (0, \infty) \tag{K}$$

erfüllt sind. Wir stellen die Auswertung der notwendigen Bedingungen analog zu den Ausführungen in Abschnitt 4.6, S. 111-113 für dieses Beispiel vor. Zunächst lässt sich die Maximumbedingung direkt auswerten und liefert

$$u^*(t) = y_0(t)e^{2t} \quad \text{f.ü. auf } (0, \infty). \tag{6.2}$$

Das bedeutet, dass sich die optimale Steuerung $u^* \in L_2((0, \infty), \omega_2)$ direkt aus der Adjungierten $y_0 \in W_2^1((0, \infty), \omega_2^{-1})$ berechnen lässt. Das kanonische Gleichungssystem überführen wir durch Einsetzen von (6.2) und mit Hilfe der Substitution $y(t) := y_0(t)e^{2t}$ in ein autonomes

System der folgenden Form:

$$\begin{aligned}\dot{x}^*(t) &= 2x^*(t) + y(t) \\ \dot{y}(t) &= x^*(t).\end{aligned}$$

Dieses besitzt die Struktur eines linearen Differentialgleichungssystems mit konstanten Koeffizienten $\dot{z} = Mz$, wobei

$$z := \begin{pmatrix} x^* \\ y \end{pmatrix}, \quad M = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$$

gewählt sind und $(x^*, y) \in W_2^1((0, \infty), \omega_2)$ gilt. Die Systemlösungen lassen sich durch die Berechnung der Eigenwerte der Koeffizientenmatrix M und die dazugehörigen Eigenvektoren ermitteln. Die Eigenwerte von M sind

$$\lambda_1 = 1 + \sqrt{2}, \quad \lambda_2 = 1 - \sqrt{2}$$

und die zugehörigen Eigenvektoren lauten

$$\mathbf{v}_1 = (1 + \sqrt{2}, 1)^T, \quad \mathbf{v}_2 = (1 - \sqrt{2}, 1)^T.$$

Daher ist die allgemeine Lösung des Differentialgleichungssystems durch

$$z(t) = c_1 e^{\lambda_1 t} \mathbf{v}_1 + c_2 e^{\lambda_2 t} \mathbf{v}_2$$

mit den noch freien Konstanten $c_1, c_2 \in \mathbb{R}$ gegeben. Für den Zustand x^* und die transformierte Adjungierte y erhalten wir damit

$$\begin{aligned}x^*(t) &= c_1(1 + \sqrt{2})e^{(1+\sqrt{2})t} + c_2(1 - \sqrt{2})e^{(1-\sqrt{2})t}, \\ y(t) &= c_1 e^{(1+\sqrt{2})t} + c_2 e^{(1-\sqrt{2})t},\end{aligned}$$

beziehungsweise durch Rücksubstitution für y_0 die Vorschrift

$$y_0(t) = c_1 e^{(\sqrt{2}-1)t} + c_2 e^{(-1-\sqrt{2})t}. \quad (6.3)$$

Die Unbekannten c_1, c_2 ergeben sich aus der Anfangsbedingung für den Zustand und durch Ausnutzen der Transversalitätsbedingung

$$\lim_{T \rightarrow \infty} y_0(T) = 0, \quad (6.4)$$

vgl. Bemerkung 4.3. Zunächst bestimmen wir durch Hinzunahme der Transversalitätsbedingung die Konstante c_1 , indem wir den Grenzübergang $T \rightarrow \infty$ in (6.3) vollziehen:

$$\lim_{T \rightarrow \infty} y_0(T) = \lim_{T \rightarrow \infty} c_1 e^{(\sqrt{2}-1)T} + \underbrace{c_2 e^{(-1-\sqrt{2})T}}_{\rightarrow 0} \stackrel{!}{=} 0$$

Um die Bedingung in (6.4) erfüllen zu können, setzen wir $c_1 = 0$. Die Anfangsbedingung $x^*(0) = 2$ bestimmt die Konstante c_2 :

$$x^*(0) = c_2(1 - \sqrt{2}) = 2 \Leftrightarrow c_2 = -2(1 + \sqrt{2}).$$

Zusammengefasst ergibt sich unter Beachtung von (6.2)

$$x^*(t) = 2e^{(1-\sqrt{2})t}, \quad u^*(t) = -2(1 + \sqrt{2})e^{(1-\sqrt{2})t}.$$

6.1.2 Optimale Lösung

Die Auswertung des Maximumprinzips liefert uns einen zulässigen Lösungskandidaten, welcher durch das Paar

$$x^*(t) = 2e^{(1-\sqrt{2})t}, \quad u^*(t) = -2(1 + \sqrt{2})e^{(1-\sqrt{2})t}$$

bereitgestellt wird. Der Satz 4.15 garantiert, dass das gefundene Paar (x^*, u^*) auch die eindeutig bestimmte Lösung der Steuerungsaufgabe in (6.1) ist. Für den optimalen Prozess (x^*, u^*) lässt sich der Zielfunktionswert wie folgt berechnen:

$$\begin{aligned} J(x^*, u^*) &= \frac{1}{2} \int_0^\infty \{x^*(t)^2 + u^*(t)^2\} e^{-2t} dt = \frac{1}{2} \int_0^\infty \{4e^{2(1-\sqrt{2})t} + 4(1 + \sqrt{2})^2 e^{2(1-\sqrt{2})t}\} e^{-2t} dt \\ &= \int_0^\infty \{2e^{-2\sqrt{2}t} + 2(\sqrt{2} + 1)^2 e^{-2\sqrt{2}t}\} dt = \lim_{b \rightarrow \infty} \left[\left(-\frac{1}{\sqrt{2}} - \frac{(\sqrt{2} + 1)^2}{\sqrt{2}} \right) \cdot \frac{1}{e^{2\sqrt{2}t}} \right]_0^b \\ &= 2(\sqrt{2} + 1). \end{aligned}$$

Die optimale Lösung (x^*, u^*) liefert einen minimalen Zielfunktionswert von $J(x^*, u^*) = 2(1 + \sqrt{2}) \approx 4.828427124746190$.

6.1.3 Anwenden der Spektralmethode

Entsprechend den Ausführungen in dem Abschnitt 5.2.2 lauten die Kollokationsgleichungen für die gesuchten Näherungslösungen $x_N, y_N \in \mathcal{P}_N((0, \infty))$ des gegebenen Beispiels wie folgt:

$$\begin{cases} \dot{x}_N(\tau_i^{(N)}) = 2 \cdot x_N(\tau_i^{(N)}) + y_N(\tau_i^{(N)}), & 1 \leq i \leq N \\ \dot{y}_N(\tau_i^{(N)}) = x_N(\tau_i^{(N)}), & 1 \leq i \leq N \\ x_N(\tau_0) = 2 \\ y_N(\tau_0) = \sum_{i=1}^N \{2 \cdot y_N(\tau_i^{(N)}) - x_N(\tau_i^{(N)})\} \mathbf{w}_i^{(N)} \end{cases}$$

Dabei sind $\tau_i^{(N)}$, $1 \leq i \leq N$ die Nullstellen des verallgemeinerten Laguerre-Polynoms $\mathcal{L}_N^{(2)}$, $w_i^{(N)}$, $i = 1, \dots, N$ die Gewichte der Gauß-Quadratur und $\tau_0 = 0$ der zusätzlich hinzugenommene Kollokationspunkt. Aufgrund der Maximumbedingung gilt für die Steuerung der Zusammenhang in (6.2), womit wir die approximierte Steuerung $u_N \in \mathcal{P}_N((0, \infty))$ aus $u_N(t) = y_N(t)$ berechnen können.

Für dieses Beispiel ist der M-File zum Aufstellen und Lösen des obigen Gleichungssystems in dem Anhang A zu finden.

6.1.4 Auswertung der numerischen Ergebnisse

Wir präsentieren die numerischen Ergebnisse für eine unterschiedliche Anzahl N an Kollokationspunkten bzw. für unterschiedliche Polynomgrade. Bereits für $N = 8$, das heißt für die Verwendung von lediglich 8 Kollokationspunkten, liefert die Spektralmethode eine derart gute Approximation des Zustandes und der Steuerung, dass der exakte Zielfunktionswert bereits bis auf vier Stellen nach dem Komma exakt angenähert werden kann. Konkret werden die Kollokationsgleichungen an den Nullstellen $\tau_i^{(8)}$, $1 \leq i \leq 8$ von $\mathcal{L}_8^{(2)}$ ausgewertet und der Zustand $x^* \in W_2^1((0, \infty), \omega_2)$ und die Adjungierte $y \in W_2^1((0, \infty), \omega_2)$ durch Polynome

$$x_8(t) = \sum_{l=0}^8 a_l \mathcal{L}_l^{(\beta)}(t), \quad y_8(t) = \sum_{l=0}^8 b_l \mathcal{L}_l^{(\beta)}(t) \quad (6.5)$$

mit Maximalgrad 8 approximiert. In diesem Fall liegt der letzte Kollokationspunkt bei $\tau_8^{(8)} = 11.431565868444633$ und es gilt

$$0.085139816152551 \leq \tau_i^{(8)} \leq 11.431565868444633, \quad i = 1, \dots, 8.$$

Die Kollokationsmethode mit $N = 8$ reduziert das Lösen des Optimalsteuerungsproblems über dem unendlichen Zeithorizont $[0, \infty)$ aus (6.1) auf das Lösen eines linearen Systems mit 18 algebraischen Gleichungen. Wir erhalten die Koeffizienten

$$\begin{aligned} a_0 &= 1.656854266539350, & a_1 &= 0.284271332696749, & a_2 &= 0.048773729641144 \\ a_3 &= 0.008371045150114, & a_4 &= 0.001452541259541, & a_5 &= 0.000344202407135 \\ a_6 &= 0.000612673183269, & a_7 &= 0.003331836692477, & a_8 &= -0.004011627569779 \\ b_0 &= -4.000000000000001, & b_1 &= -0.686291466921300, & b_2 &= -0.117748801527802 \\ b_3 &= -0.020201342245515, & b_4 &= -0.003459251945286, & b_5 &= -0.000554169426203 \\ b_6 &= 0.000134235388067, & b_7 &= 0.001359581754604, & b_8 &= -0.001665918346238, \end{aligned}$$

woraus sich durch Einsetzen in die Ansätze aus (6.5) die Näherungslösungen

$$\begin{aligned}
 x_8(t) &= 2 - 0.818737959783879 t + 0.105754466114635 t^2 \\
 &\quad + 0.104213548994575 t^3 - 0.101223996402397 t^4 + 0.040176621829031 t^5 \\
 &\quad - 0.007856892615396 t^6 + 0.000730442764845 t^7 - 0.000025470651237 t^8, \\
 y_8(t) &= -4.828427133269675 + 2.004011627569780 t - 0.441462000450170 t^2 \\
 &\quad + 0.110135203340751 t^3 - 0.048830327387229 t^4 + 0.017197058037457 t^5 \\
 &\quad - 0.003288391646611 t^6 + 0.000303943238484 t^7 - 0.000010577259341 t^8
 \end{aligned}$$

ergeben. Mit $u_8(t) = y_8(t)$ berechnen wir den approximierten Zielfunktionswert unter Verwendung der Gaußschen Quadraturformel und erhalten

$$J(x_8, u_8) = 4.828430474791660.$$

Mit steigender Anzahl der Kollokationspunkte wird deutlich, dass eine Verbesserung der Approximation erreicht wird. Für $N = 10, 12, 14$ sind die Näherungslösungen $x_N(t)$ und $u_N(t)$ in den Abbildungen 6.1 und 6.2 grafisch in einem direkten Vergleich mit den exakten Lösungen $x^*(t)$ und $u^*(t)$ dargestellt.

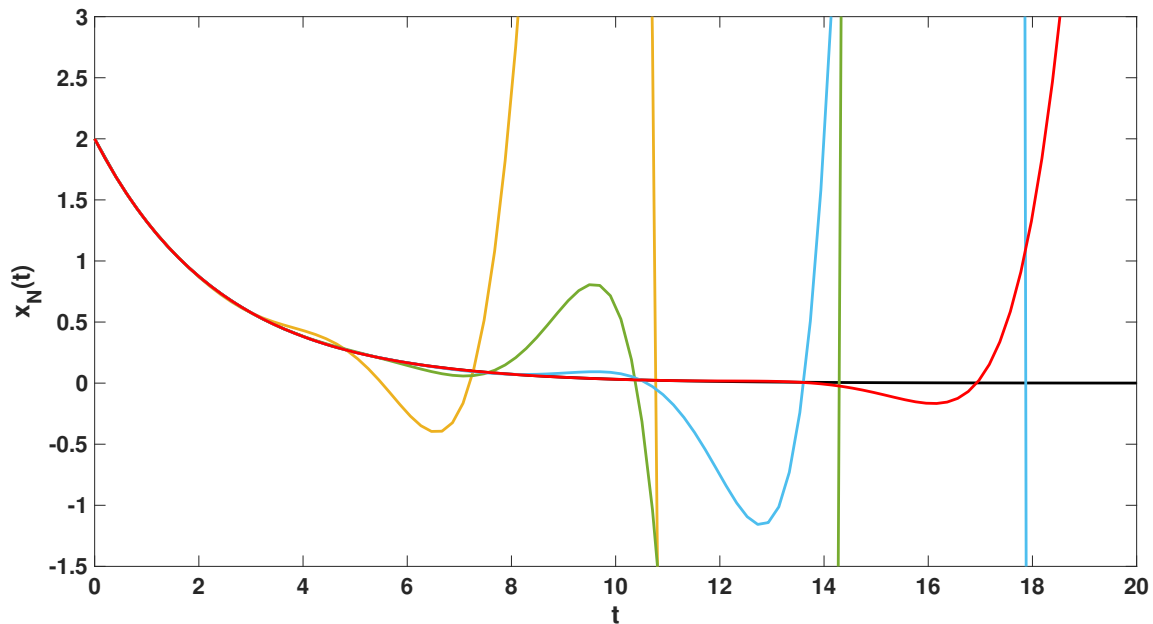


Abbildung 6.1: $\beta = 2$: Näherungslösung $x_N(t)$ des Zustandes für $N = 8$ (—), $N = 10$ (—), $N = 12$ (—), $N = 14$ (—) im Vergleich mit der exakten Lösung $x^*(t)$ (—)

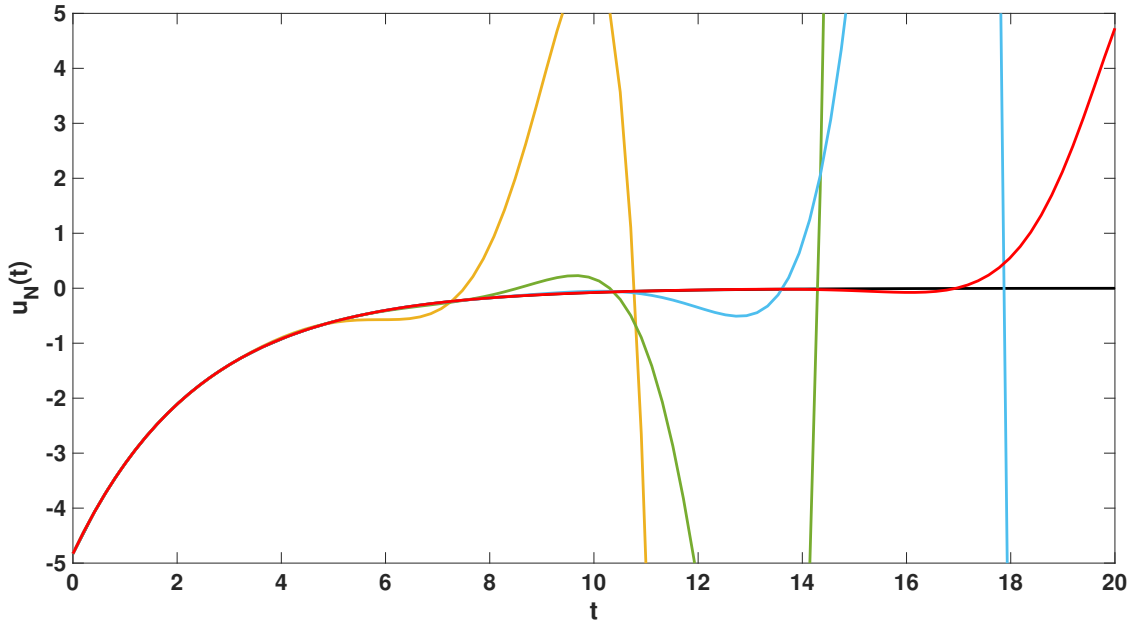


Abbildung 6.2: $\beta = 2$: Näherungslösung $u_N(t)$ der Steuerung für $N = 8$ (—), $N = 10$ (—), $N = 12$ (—), $N = 14$ (—) im Vergleich mit der exakten Lösung $u^*(t)$ (—)

Die jeweils letzten Kollokationspunkte sind durch

$$\tau_{10}^{(10)} = 14.960348506136945, \quad \tau_{12}^{(12)} = 18.549560522233460, \quad \tau_{14}^{(14)} = 22.183040855558701$$

festgelegt. Die Abbildungen lassen erkennen, dass sich die numerischen Lösungen mit wachsendem N auf einem sich vergrößernden Intervall den exakten Lösungen annähern. Das heißt, die Erhöhung von N und damit das Vergrößern des letzten Kollokationspunktes verschiebt den Zeitpunkt, ab dem die numerische Approximation von den exakten Lösungen abweicht nach hinten. Dennoch wird deutlich, dass keine punktweise bzw. lokale Konvergenz zu erwarten ist, was aber auch im Kontext der Theorie der gewichteten Sobolev- und Lebesgue-Räume und im Hinblick auf das quadratische Zielfunktional als nicht sinnvoll erscheint. Entsprechend interessieren wir uns stattdessen für ein globales Fehlermaß gemessen in dem gewichteten L_{2,ω_β} -Sinn. In Tabelle 6.1 sind die L_{2,ω_β} -Fehler der Näherungslösungen für $N = 8, 10, 12, 14$ dargestellt.

Die gewichteten L_{2,ω_β} -Fehler werden für anwachsendes N kleiner. Insbesondere haben wir in diesem Beispiel bereits bei $N = 14$ eine Approximationsgüte erreicht, bei der der L_2 -Fehler des Zustandes in einer Größenordnung von $1.1 \cdot 10^{-7}$ liegt und sich der Fehler der Steuerung in einer Größenordnung von $4.7 \cdot 10^{-8}$ befindet. Diese Genauigkeit spiegelt sich in den numerisch ermittelten Zielfunktionswerten wider, vgl. Tabelle 6.2. Es ist klar zu erkennen, dass der approximierte Zielfunktionswert $J(x_N, u_N)$ bei $N = 14$ bereits so gut

N	$\ x^* - x_N\ _{2,\omega_\beta}$	$\ u^* - u_N\ _{2,\omega_\beta}$
8	$0.37075086564 \cdot 10^{-2}$	$0.15360948664 \cdot 10^{-2}$
10	$0.28266584272 \cdot 10^{-3}$	$0.11707241558 \cdot 10^{-3}$
12	$0.19335025499 \cdot 10^{-4}$	$8.00917154073 \cdot 10^{-6}$
14	$1.13123425284 \cdot 10^{-7}$	$4.68579721511 \cdot 10^{-8}$

Tabelle 6.1: $\beta = 2$: L_{2,ω_β} -Fehler des Zustandes und der Steuerung

ist, dass erst die fünfzehnte Stelle nach dem Komma von dem exakten Wert verschieden ist. Dementsprechend können wir festhalten, dass für dieses Beispiel eine Kollokation mit $N = 14$ Punkten ausreichend ist, um präzise numerische Werte für den optimalen Zielfunktionswert zu ermitteln. Eine bessere Genauigkeit konnten wir in diesem Beispiel auch bei Erhöhung von N nicht erreichen.

N	$J(x_N, u_N)$	$y_N(0)$
8	4.828430474791660	-4.828427133269675
10	4.828427144099940	-4.828427124727055
12	4.828427124836948	-4.828427124746217
14	4.828427124746193	-4.828427124746190

Tabelle 6.2: $\beta = 2$: Approximation des Zielfunktionswertes und des Anfangswertes der Adjungierten

Ein Blick auf die numerisch ermittelten Anfangswerte $y_N(0)$ scheint uns als besonders interessant. Wir erinnern noch einmal daran, dass der Anfangswert der Adjungierten $y(0)$ a priori nicht bekannt ist und als unbestimmter Parameter aufgefasst werden kann. Mit Hilfe der Transversalitätsbedingung war es uns jedoch möglich, eine Bestimmungsgleichung in Integralform herzuleiten und diese durch Diskretisierung mit der Gauß-Quadratur in die Spektralmethode einfließen zu lassen. Da wir die exakte Lösung analytisch mit Hilfe des Maximumprinzips ermitteln konnten, können wir den exakten Anfangswert für einen Vergleich mit dem numerisch ermittelten heranziehen. Der exakte Anfangswert der Adjungierten $y \in W_2^1((0, \infty), \omega_2)$ beträgt

$$y(0) = -2(1 + \sqrt{2}) \approx -4.828427124746190.$$

Die Werte $y_N(0)$ sind in der Tabelle 6.2 erfasst und demonstrieren den erfolgreichen Einsatz der gewählten Bestimmungsgleichung. Bei $N = 14$ stimmt der dadurch numerisch bestimmte Wert bis auf fünfzehn Stellen nach dem Komma mit dem exakten Wert $y(0)$ überein.

6.2 Modifiziertes Beispiel

Wir möchten nun das Beispiel aus (6.1) in dem Sinne abändern, dass wir den Parameter $\beta > 0$, $\beta \neq 2$ in der Gewichtsfunktion variieren und dessen Einfluss auf die Approximationsergebnisse untersuchen. Entsprechend formulieren wir die modifizierte Steuerungsaufgabe

wie folgt:

Minimiere

$$J(x, u) = \frac{1}{2} \int_0^{\infty} \{x^2(t) + u^2(t)\} e^{-\beta t} dt$$

bezüglich

$$(x, u) \in W_2^1((0, \infty), \omega_\beta) \times L_2((0, \infty), \omega_\beta), \omega_\beta(t) := e^{-\beta t}, \beta > 0,$$

$$\begin{aligned} \dot{x}(t) &= 2x(t) + u(t) \quad \text{f.ü. auf } (0, \infty), \\ x(0) &= 2, \quad u(t) \in \mathbb{R} \quad \text{f.ü. auf } (0, \infty). \end{aligned} \quad (6.6)$$

Analog zu dem Beispiel mit $\beta = 2$ stellen wir fest, dass die obige Aufgabe der Problemklasse $(LQ)_\infty$ mit skalarem Zustand und skalarer Steuerung entspricht und den Voraussetzungen aus Abschnitt 4.2 genügt. Daher ist die Existenz der eindeutig bestimmten Lösung $(x^*, u^*) \in W_2^1((0, \infty), \omega_\beta) \times L_2((0, \infty), \omega_\beta)$ aufgrund des Satzes 4.13 garantiert und da auch die Voraussetzung (V6) erfüllt ist, können wir wieder die Bedingungen des Maximumprinzips verwenden, um einen Lösungskandidaten zu ermitteln.

6.2.1 Diskussion der notwendigen Bedingungen

Da die Auswertung der notwendigen Optimalitätsbedingungen mit den gleichen Methoden wie in dem Abschnitt 6.1.1 erfolgt, fassen wir uns hier kurz. Die Pontryaginsche Funktion für das modifizierte Beispiel lautet

$$H(t, \xi, v, \eta, 1) = -\frac{1}{2}(\xi^2 + v^2)e^{-\beta t} + \eta(2\xi + v).$$

Für die Lösung $(x^*, u^*) \in W_2^1((0, \infty), \omega_\beta) \times L_2((0, \infty), \omega_\beta)$ existieren ein Multiplikator $\lambda_0 = 1$ und eine adjungierte Funktion $y_0 \in W_2^1((0, \infty), \omega_\beta^{-1}), \omega_\beta^{-1}(t) = e^{\beta t}$, welche die Maximumbedingung

$$u^*(t) = y_0(t)e^{\beta t} \quad \text{f.ü. auf } (0, \infty) \quad (M)$$

und die kanonische Gleichung

$$\dot{y}_0(t) = x^*(t)e^{-\beta t} - 2y_0(t) \quad \text{f.ü. auf } (0, \infty) \quad (K)$$

erfüllen. Durch direktes Einsetzen der Maximumbedingung und mit Hilfe der Transformation $y(t) := y_0(t)e^{\beta t}$ lässt sich das kanonische Differentialgleichungssystem wieder in ein

autonomes System überführen:

$$\begin{aligned}\dot{x}^*(t) &= 2x^*(t) + y(t) \\ \dot{y}(t) &= x^*(t) + (\beta - 2)y(t).\end{aligned}$$

Im Unterschied zu dem ersten Beispiel betrachten wir nun $\beta \neq 2$, wodurch der Parameter β Einfluss auf die rechte Seite der Adjungiertendifferentialgleichung nimmt. Die Koeffizientenmatrix des obigen Systems lautet

$$M(\beta) = \begin{pmatrix} 2 & 1 \\ 1 & \beta - 2 \end{pmatrix}$$

und es ergeben sich in Abhängigkeit von β die Eigenwerte

$$\lambda_1(\beta) = \frac{\beta}{2} + \frac{1}{2}\sqrt{\beta^2 - 8\beta + 20}, \quad \lambda_2(\beta) = \frac{\beta}{2} - \frac{1}{2}\sqrt{\beta^2 - 8\beta + 20}$$

mit den dazugehörigen Eigenvektoren

$$\mathbf{v}_1(\beta) = \left(\frac{1}{\lambda_1(\beta) - 2}, 1 \right)^T, \quad \mathbf{v}_2(\beta) = \left(\frac{1}{\lambda_2(\beta) - 2}, 1 \right)^T.$$

Entsprechend sind der Zustand x^* und die transformierte Adjungierte y durch die Vorschrift

$$x^*(t) = c_1 \cdot \frac{1}{\lambda_1(\beta) - 2} \cdot e^{\lambda_1(\beta)t} + c_2 \cdot \frac{1}{\lambda_2(\beta) - 2} \cdot e^{\lambda_2(\beta)t}, \quad y(t) = c_1 \cdot e^{\lambda_1(\beta)t} + c_2 \cdot e^{\lambda_2(\beta)t}$$

festgelegt, woraus sich durch Rücksubstitution

$$y_0(t) = c_1 e^{(\lambda_1(\beta) - \beta)t} + c_2 e^{(\lambda_2(\beta) - \beta)t}$$

ergibt. Die Unbekannte c_1 bestimmen wir wieder durch Hinzunahme der Transversalitätsbedingung:

$$\lim_{T \rightarrow \infty} y_0(T) = \lim_{T \rightarrow \infty} c_1 e^{(\lambda_1(\beta) - \beta)T} + c_2 e^{(\lambda_2(\beta) - \beta)T} \stackrel{!}{=} 0.$$

Da $\lambda_2(\beta) - \beta = -\frac{\beta}{2} - \frac{1}{2}\sqrt{\beta^2 - 8\beta + 20} < 0$ für alle $\beta > 0$ gilt, setzen wir wieder $c_1 = 0$ und sichern damit das Erfülltsein der obigen Grenzwertbeziehung für alle $\beta > 0$. Aus dem Anfangswert $x^*(0) = 2$ ergibt sich

$$x^*(0) = c_2 \cdot \frac{1}{\lambda_2(\beta) - 2} = 2 \Leftrightarrow c_2 = 2 \cdot \lambda_2(\beta) - 4 < 0 \quad \forall \beta > 0,$$

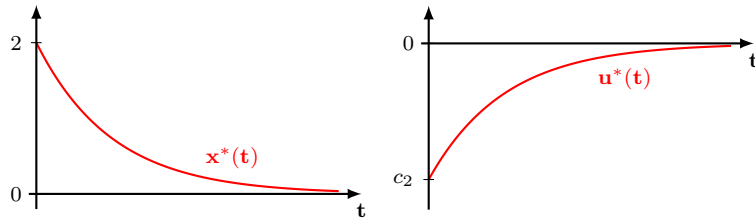
woraus mit Hilfe der Maximumbedingung die Kandidaten

$$x^*(t) = 2 e^{\lambda_2(\beta)t}, \quad u^*(t) = (2 \cdot \lambda_2(\beta) - 4) e^{\lambda_2(\beta)t} \quad (6.7)$$

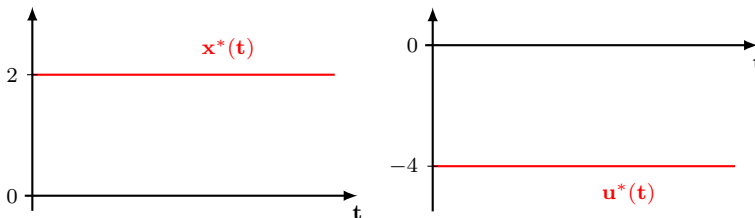
mit der Adjungierten

$$y_0(t) = (2 \cdot \lambda_2(\beta) - 4) e^{(\lambda_2(\beta) - \beta)t}$$

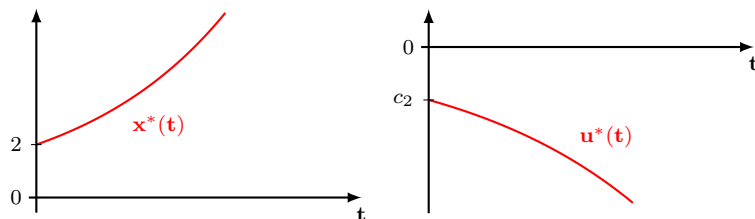
resultieren. Die von dem Eigenwert $\lambda_2(\beta)$ abhängigen Lösungsstrukturen sind in der Abbildung 6.3 dargestellt und zeigen den Einfluss des Parameters $\beta > 0$ auf.



(a) $\lambda_2(\beta) < 0 \Leftrightarrow 0 < \beta < 2.5$



(b) $\lambda_2(\beta) = 0 \Leftrightarrow \beta = 2.5$



(c) $\lambda_2(\beta) > 0 \Leftrightarrow \beta > 2.5$

Abbildung 6.3: Lösungsstruktur in Abhängigkeit von $\lambda_2(\beta)$ bzw. β

6.2.2 Optimale Lösung

Wir haben den zulässigen Lösungskandidaten (6.7) ermittelt, wobei durch den Satz 4.15 garantiert ist, dass das gefundene Paar (x^*, u^*) auch die eindeutig bestimmte Lösung der Steuerungsaufgabe in (6.6) ist. Für alle $\beta > 0$ ist $2\lambda_2(\beta) - \beta < 0$. Daher liefert uns der optimale Prozess (x^*, u^*) aus (6.7) in Abhängigkeit des Parameters $\beta > 0$ den folgenden

Zielfunktionswert:

$$\begin{aligned}
J(x^*, u^*) &= \frac{1}{2} \int_0^\infty \{x^*(t)^2 + u^*(t)^2\} e^{-\beta t} dt = \frac{1}{2} \int_0^\infty \{4e^{2\lambda_2(\beta)t} + (2\lambda_2(\beta) - 4)^2 e^{2\lambda_2(\beta)t}\} e^{-\beta t} dt \\
&= \left[2 + \frac{(2\lambda_2(\beta) - 4)^2}{2} \right] \cdot \int_0^\infty e^{(2\lambda_2(\beta) - \beta)t} dt \\
&= \left[2 + \frac{(2\lambda_2(\beta) - 4)^2}{2} \right] \cdot \lim_{b \rightarrow \infty} \left[\frac{1}{(2\lambda_2(\beta) - \beta)} \cdot e^{(2\lambda_2(\beta) - \beta)t} \right]_0^b \\
&= \frac{2}{\beta - 2\lambda_2(\beta)} + \frac{(2\lambda_2(\beta) - 4)^2}{2\beta - 2\lambda_2(\beta)}.
\end{aligned}$$

6.2.3 Anwenden der Spektralmethode

Der in diesem Beispiel noch zu wählende Parameter $\beta > 0$, $\beta \neq 2$ taucht in der diskretisierten Adjungiertendifferentialgleichung auf. Entsprechend erhalten wir für die zu ermittelnden Näherungslösungen $x_N, y_N \in \mathcal{P}_N((0, \infty))$ und mit denen als Kollokationspunkte gewählten Nullstellen $\tau_i^{(N)}$, $1 \leq i \leq N$ des verallgemeinerten Laguerre-Polynoms $\mathcal{L}_N^{(\beta)}$, den Gewichten der Gauß-Quadratur $\mathbf{w}_i^{(N)}$, $i = 1, \dots, N$ und dem zusätzlichen Kollokationspunkt $\tau_0 = 0$ die nachfolgenden Kollokationsgleichungen:

$$\begin{cases} \dot{x}_N(\tau_i^{(N)}) = 2 \cdot x_N(\tau_i^{(N)}) + y_N(\tau_i^{(N)}), & 1 \leq i \leq N \\ \dot{y}_N(\tau_i^{(N)}) = x_N(\tau_i^{(N)}) + (\beta - 2) \cdot y_N(\tau_i^{(N)}), & 1 \leq i \leq N \\ x_N(\tau_0) = 2 \\ y_N(\tau_0) = \sum_{i=1}^N \{2 \cdot y_N(\tau_i^{(N)}) - x_N(\tau_i^{(N)})\} \mathbf{w}_i^{(N)} \end{cases}$$

Die approximierte Steuerung $u_N \in \mathcal{P}_N((0, \infty))$ berechnen wir aus der Maximumbedingung gemäß der Vorschrift $u_N(t) = y_N(t)$.

6.2.4 Auswertung der numerischen Ergebnisse für $\beta = 1$ und $\beta = 2.4$

Wir untersuchen zunächst die numerischen Ergebnisse für den Fall $\lambda_2(\beta) < 0$, d.h. für Parameter $\beta \in (0, 2.5)$, vgl. Abbildung 6.3 (a). Ist $\beta = 2$, so erhalten wir unser ursprüngliches Beispiel aus (6.1). Wir präsentieren in diesem Abschnitt weitere numerische Resultate für $\beta = 1$ und $\beta = 2.4$. Entsprechend den Ausführungen in den vorangegangenen Abschnitten erhalten wir für $\beta = 1$ den optimalen Zustand und die optimale Steuerung

$$x^*(t) = 2e^{(1/2 - \sqrt{13}/2)t}, \quad u^*(t) = (-3 - \sqrt{13})e^{(1/2 - \sqrt{13}/2)t}$$

mit der Adjungierten $y_0(t) = (-3 - \sqrt{13}) e^{(-1/2 - \frac{\sqrt{13}}{2})t}$. Für den optimalen Zielfunktionswert ergibt sich

$$J(x^*, u^*) = \frac{2\sqrt{13}}{13} + \frac{(-3 - \sqrt{13})^2 \sqrt{13}}{26} \approx 6.605551275463991.$$

Ist $\beta = 2.4$, so ist

$$x^*(t) = 2 e^{(6/5 - \sqrt{41}/5)t}, \quad u^*(t) = \left(-\frac{8}{5} - \frac{2\sqrt{41}}{5}\right) e^{(6/5 - \sqrt{41}/5)t},$$

$$y_0(t) = \left(-\frac{8}{5} - \frac{2\sqrt{41}}{5}\right) e^{(-6/5 - \sqrt{41}/5)t}$$

und es resultiert ein Zielfunktionswert von

$$J(x^*, u^*) = \frac{5}{\sqrt{41}} + \frac{5\sqrt{41}}{164} \cdot \left(-\frac{8}{5} - \frac{2\sqrt{41}}{5}\right)^2 \approx 4.161249694973140.$$

Die Wahl des Parameters β beeinflusst die verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(\beta)}$, $l = 0, 1, \dots$ und entsprechend die Nullstellen von $\mathcal{L}_N^{(\beta)}$. Die daraus resultierenden jeweils letzten Kollokationspunkte sind für unterschiedliche N in der Tabelle 6.3 zusammengestellt.

N	$\tau_N^{(N)}$ für $\beta = 1$	$\tau_N^{(N)}$ für $\beta = 2.4$
8	22.863131736889265	9.526304890370525
10	29.920697012273891	12.466957088447458
12	37.099121044466919	15.457967101861213
14	44.366081711117403	18.485867379632268
16	51.701160339543314	21.542150141476387
18	59.090546435901274	-

Tabelle 6.3: Kollokationspunkte $\tau_N^{(N)}$ für $\beta = 1$ und $\beta = 2.4$

Sowohl für $\beta = 1$ als auch für $\beta = 2.4$ liefert die Spektralmethode bereits für niedrige Polynomgrade gute Approximationsergebnisse. Dennoch wird deutlich, dass für $\beta = 2.4$ eine deutlich höhere Genauigkeit erreicht werden kann als für $\beta = 1$. Dies spiegelt sich sowohl in der Betrachtung der gewichteten L_{2,ω_β} -Fehler als auch bei der Auswertung der Zielfunktionswerte wider, vgl. Tabelle 6.4 und Tabelle 6.5.

Bei $\beta = 1$ haben wir die besten Resultate für $N = 18$ erzielen können. In diesem Fall ist $\tau_{18}^{(18)} \approx 59.01$ und der numerisch ermittelte Zielfunktionswert stimmt bis auf sechs Stellen nach dem Komma mit dem exakten Wert von $J(x^*, u^*)$ überein. Die zugehörigen L_{2,ω_β} -Fehler für x_{18} und u_{18} befinden sich in einer Größenordnung von $0.1 \cdot 10^{-2}$ und $0.2 \cdot 10^{-3}$.

Ist $\beta = 2.4$, so genügt es den Polynomgrad N maximal gleich 16 zu wählen, um den exakten Zielfunktionswert bis auf fünfzehn Stellen nach dem Komma genau zu berechnen, vgl. Tabelle 6.5. Eine derart hohe Genauigkeit konnten wir in dem Fall $\beta = 1$ auch für größere N nicht erreichen. Die deutlich bessere Approximationsgüte spiegelt sich entsprechend auch bei einem Vergleich der L_{2,ω_β} -Fehler wider und wir erkennen, dass die jeweiligen Fehler sowohl für den

N	$\beta = 1$		$\beta = 2.4$	
	$\ x^* - x_N\ _{2,\omega_\beta}$	$\ u^* - u_N\ _{2,\omega_\beta}$	$\ x^* - x_N\ _{2,\omega_\beta}$	$\ u^* - u_N\ _{2,\omega_\beta}$
8	$0.870188467 \cdot 10^{-1}$	$0.697963055 \cdot 10^{-1}$	$0.17096109 \cdot 10^{-1}$	$0.82168150 \cdot 10^{-2}$
10	0.341607909179886	$0.917162796 \cdot 10^{-1}$	$0.20310965 \cdot 10^{-4}$	$9.76195453 \cdot 10^{-6}$
12	0.120235207963304	$0.403821868 \cdot 10^{-1}$	$0.10194189 \cdot 10^{-4}$	$4.89958056 \cdot 10^{-6}$
14	$0.534509442 \cdot 10^{-1}$	$0.149687335 \cdot 10^{-1}$	$1.11463543 \cdot 10^{-9}$	$5.35715309 \cdot 10^{-10}$
16	$0.808145530 \cdot 10^{-2}$	$0.285892230 \cdot 10^{-2}$	$2.91895385 \cdot 10^{-11}$	$1.40364529 \cdot 10^{-11}$
18	$0.103971366 \cdot 10^{-2}$	$0.209089678 \cdot 10^{-3}$	-	-

Tabelle 6.4: $\beta = 1$ und $\beta = 2.4$: L_{2,ω_β} -Fehler des Zustandes und der Steuerung

N	$J(x_N, u_N)$	
	$\beta = 1$	$\beta = 2.4$
8	6.611897853	4.161336719742284
10	6.612396859	4.161249695095972
12	6.608049900	4.161249695004072
14	6.605777942	4.161249694973139
16	6.605564419	4.161249694973140
18	6.605551122	-

Tabelle 6.5: $\beta = 1$ und $\beta = 2.4$: Approximation des Zielfunktionswertes

Zustand als auch für die Steuerung in dem Fall $\beta = 2.4$ signifikant kleiner sind als in dem Fall $\beta = 1$, vgl. Tabelle 6.4. Für $N = 16$ und $\beta = 2.4$ ergeben sich die Fehler in einer Größenordnung von $2.9 \cdot 10^{-11}$ und $1.4 \cdot 10^{-11}$.

Werfen wir nun einen Blick auf die numerisch erzielten Anfangswerte $y_N(0)$. Für $\beta = 1$ beträgt der exakte Anfangswert der Adjungierten $y \in W_2^1((0, \infty), \omega_\beta)$

$$y(0) = -3 - \sqrt{13} \approx -6.605551275463990$$

und für $\beta = 2.4$ gilt

$$y(0) = -\frac{8}{5} - \frac{2\sqrt{41}}{5} \approx -4.161249694973140.$$

Die Auswertung in der Tabelle 6.6 stützt die bisherigen Beobachtungen und zeigt, dass die Genauigkeit für $\beta = 2.4$ größer ist als die für $\beta = 1$. Obwohl bei $\beta = 1$ der Anfangswert der Adjungierten numerisch „nur“ bis auf sechs Nachkommastellen ($N = 18$) genau bestimmt werden kann, demonstrieren unsere Experimente, dass die integrierte Transversalitätsbedingung für die näherungsweise Ermittlung von $y^*(0)$ geeignet ist. Für $\beta = 2.4$ erreichen wir den exakten Wert bis auf dreizehn Nachkommastellen ($N = 16$). Für $N = 10$ erhielten wir diesen sogar bis auf fünfzehn Nachkommastellen genau, jedoch erhöhten wir den Polynomgrad noch bis auf 16, um eine bessere Annäherung an den Zielfunktionswert zu erreichen.

Zusammenfassend können wir festhalten, dass für den Fall $\lambda_2(\beta) < 0$ eine Erhöhung des Parameters β zu deutlich besseren Approximationsergebnissen geführt hat. Besitzt man den

N	$y_N(0)$	
	$\beta = 1$	$\beta = 2.4$
8	-6.60880221	-4.161249694973067
10	-6.60192712	-4.161249694973140
12	-6.60596730	-4.161249694973138
14	-6.60549298	-4.161249694973139
16	-6.60555415	-4.161249694973139
18	-6.60555113	-

Tabelle 6.6: $\beta = 1$ und $\beta = 2.4$: Approximation des Anfangswertes der Adjungierten

Freiheitsgrad, den Parameter β in der Aufgabenstellung zu wählen, so kann dieses Wissen bewusst verwendet werden, um die Genauigkeit des numerisch berechneten Zielfunktionswertes zu verbessern. In unseren Beispielen ging dies zusätzlich damit einher, weniger Kollokationspunkte, also auch einen geringeren Polynomgrad N wählen zu müssen, um dennoch einen hohen Genauigkeitsgrad erzielen zu können. Dies liegt natürlich daran, dass β gerade das exponentielle Abklingverhalten der Gewichtsfunktion $\omega_\beta(t) = e^{-\beta t}$ bestimmt und diese wiederum dem gewichteten Zustands- und Steuerungsraum zugrunde gelegt wurde und auch im Zielfunktional auftaucht. Entsprechend ist durch den Parameter β ein Maß gegeben, mit welchem auf die L_{2,ω_β} -Fehler und aufgrund der quadratischen Struktur des Integranden auch auf die Annäherung an den exakten Wert $J(x^*, u^*)$ positiv Einfluss genommen werden kann.

Auf den nachfolgenden Seiten fügen wir ergänzend die grafischen Darstellungen der erhaltenen Näherungslösungen an. Die Abbildungen 6.4 und 6.5 illustrieren die Annäherung der Polynome x_N und u_N an x^* bzw. u^* mit wachsendem Polynomgrad N für den Fall $\beta = 1$. Wieder ist klar zu erkennen, dass das „Wegspringen“ der Polynome deutlich vor dem letzten Kollokationspunkt $\tau_{18}^{(18)}$ stattfindet.

In den Grafiken 6.6 und 6.7 ist gut erkennbar, dass die Näherungslösungen für $\beta = 2.4$ die exakten Lösungen auf einem ähnlich großem Zeitintervall wie bei $\beta = 1$ überdecken, obwohl die jeweils letzten Kollokationspunkte für $\beta = 2.4$ deutlich kleiner als die für $\beta = 1$ sind. Zusammen mit der vorangegangenen Auswertung lässt dies die Vermutung aufkommen, dass die Größe des letzten Kollokationspunktes im Vergleich zu einer Erhöhung des Parameters β weniger Einfluss auf die Approximationsgüte nimmt.

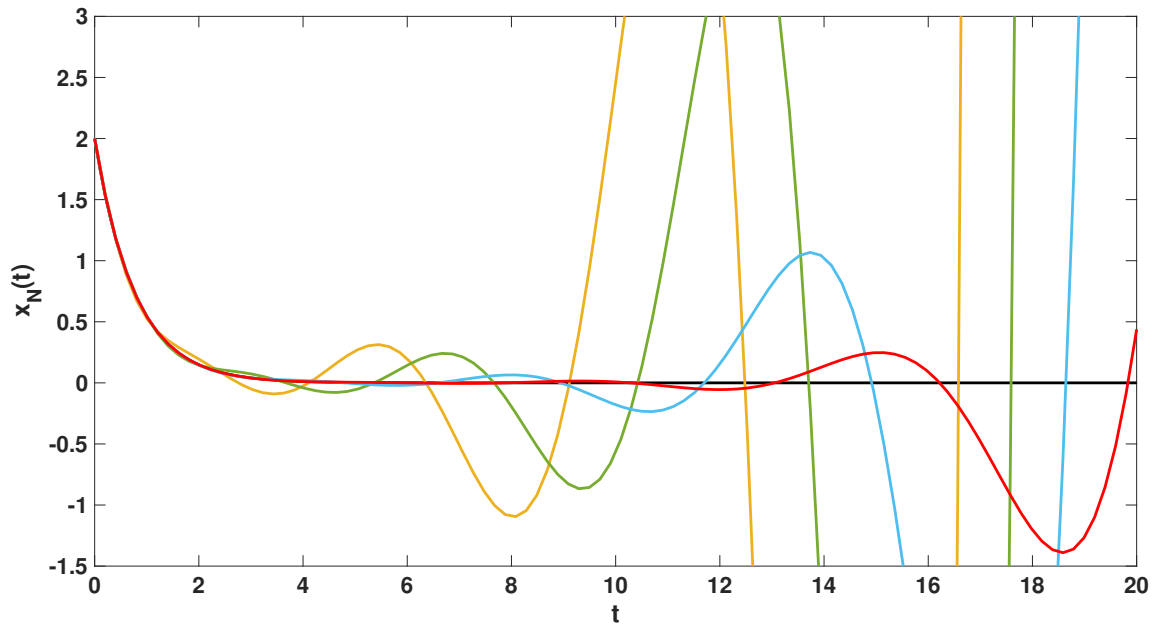


Abbildung 6.4: $\beta = 1$: Näherungslösung $x_N(t)$ des Zustandes für $N = 12$ (—), $N = 14$ (—), $N = 16$ (—), $N = 18$ (—) im Vergleich mit der exakten Lösung $x^*(t)$ (—)

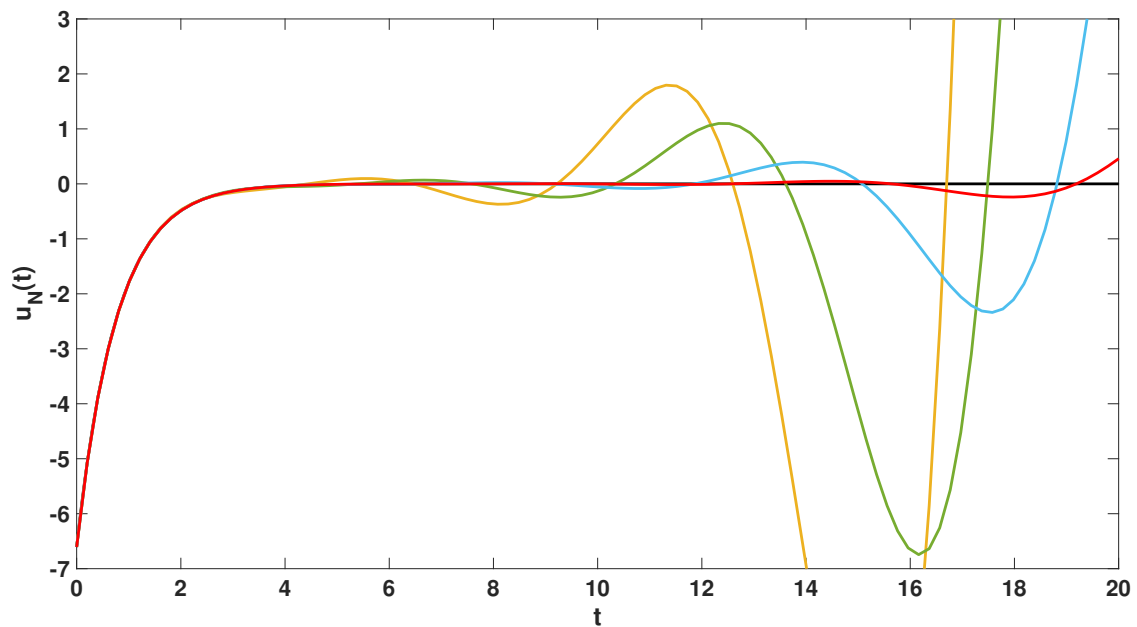


Abbildung 6.5: $\beta = 1$: Näherungslösung $u_N(t)$ der Steuerung für $N = 12$ (—), $N = 14$ (—), $N = 16$ (—), $N = 18$ (—) im Vergleich mit der exakten Lösung $u^*(t)$ (—)

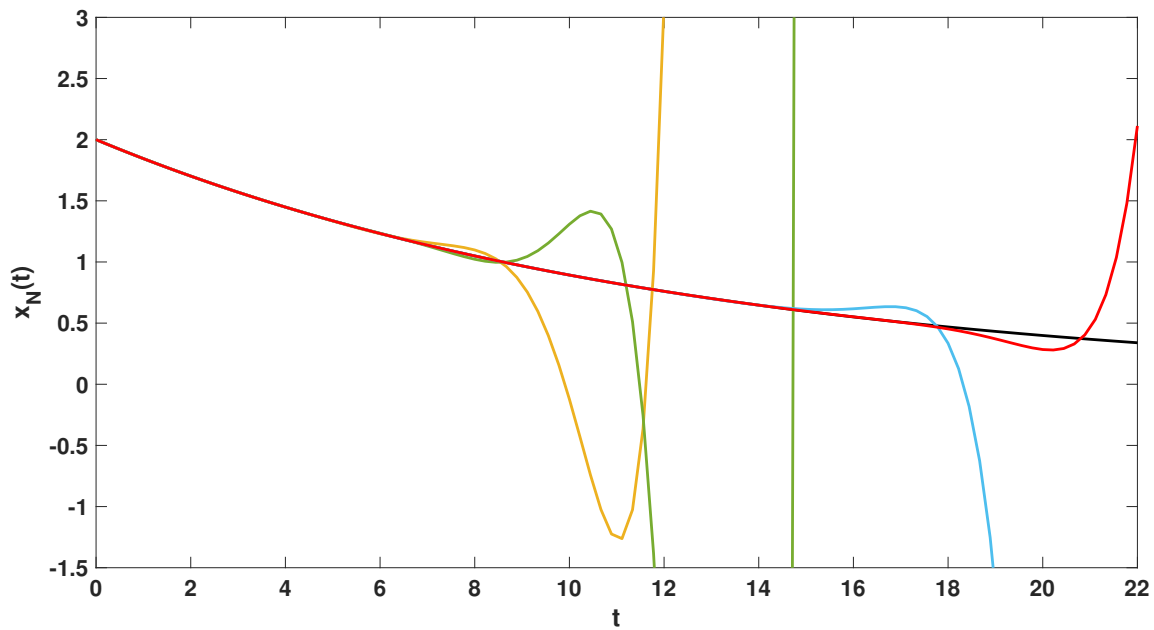


Abbildung 6.6: $\beta = 2.4$: Näherungslösung $x_N(t)$ des Zustandes für $N = 10$ (—), $N = 12$ (—), $N = 14$ (—), $N = 16$ (—) im Vergleich mit der exakten Lösung $x^*(t)$ (—)

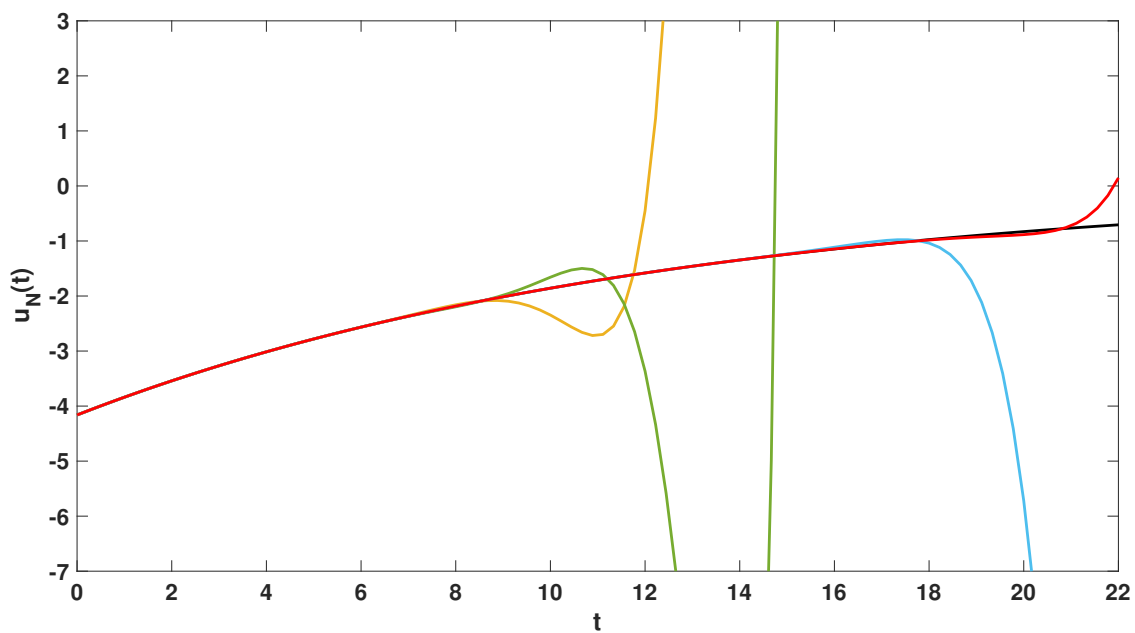


Abbildung 6.7: $\beta = 2.4$: Näherungslösung $u_N(t)$ der Steuerung für $N = 10$ (—), $N = 12$ (—), $N = 14$ (—), $N = 16$ (—) im Vergleich mit der exakten Lösung $u^*(t)$ (—)

6.2.5 Auswertung der numerischen Ergebnisse für $\beta = 2.5$ und $\beta = 25$

In diesem Abschnitt untersuchen wir die numerischen Resultate für den Fall $\lambda_2(\beta) \geq 0$, d.h. für Parameter $\beta \geq 2.5$, vgl. Abbildung 6.3 (b) und (c). Wir haben die Spektralmethode für verschiedene $\beta \geq 2.5$ getestet. Es hat sich bestätigt, dass für kleinere Werte des Parameters β eine höhere Anzahl von Kollokationspunkten benötigt wird, um eine entsprechende Approximationsgenauigkeit erreichen zu können. Zur Demonstration präsentieren wir die Ergebnisse für $\beta = 2.5$ und $\beta = 25$. Ist $\beta = 2.5$, so erhalten wir die konstanten Lösungen

$$x^*(t) = 2, \quad u^*(t) = -4, \quad y_0(t) = -4 \quad (6.8)$$

und einen optimalen Zielfunktionswert von

$$J(x^*, u^*) = 4.$$

Für $\beta = 25$ sind der optimale Zustand und die optimale Steuerung durch

$$x^*(t) = 2e^{(25/2 - \sqrt{445}/2)t}, \quad u^*(t) = (21 - \sqrt{445})e^{(25/2 - \sqrt{445}/2)t}$$

gegeben. Die zugehörige Adjungierte lautet $y_0(t) = (21 - \sqrt{445})e^{(-25/2 - \sqrt{445}/2)t}$ und es resultiert ein Zielfunktionswert von

$$J(x^*, u^*) = \frac{2}{\sqrt{445}} - \frac{\sqrt{445}}{890} \cdot (21 - \sqrt{445})^2 \approx 0.095023109728987.$$

Die jeweils verwendeten letzten Kollokationspunkte sind für unterschiedliche N in der Tabelle 6.7 aufgelistet.

N	$\tau_N^{(N)}$ für $\beta = 2.5$	$\tau_N^{(N)}$ für $\beta = 25$
8	9.145252694755708	0.914525269475571
10	11.968278804909563	1.196827880490955
12	14.839648417786773	1.483964841778677
14	17.746432684446980	1.774643268444697
16	20.680464135817324	2.068046413581733
18	23.636218574360505	-
20	26.609766610246311	-

Tabelle 6.7: Kollokationspunkte $\tau_N^{(N)}$ für $\beta = 2.5$ und $\beta = 25$

Der Fall $\beta = 2.5$ scheint uns besonders interessant, da hier als Spezialfall die in (6.8) angegebenen konstanten Lösungen resultieren. Für die numerische Lösung haben wir 20 Kollokationspunkte benötigt, um die besten numerischen Resultate zu erzielen, vgl. Tabelle 6.8. Dabei konnten wir den Zielfunktionswert bis auf 14 Nachkommastellen exakt berechnen. Bei $\beta = 25$ befinden wir uns in dem Fall $\lambda_2(\beta) > 0$, vgl. Abbildung 6.3, (c). Hier erhielten wir bereits bei 16 Kollokationspunkten eine Zielfunktionswert, welcher bis auf 15 Stellen nach dem Komma mit dem optimalen Wert $J(x^*, u^*)$ übereinstimmt, vgl. Tabelle 6.8.

N	$J(x_N, u_N)$	
	$\beta = 2.5$	$\beta = 25$
8	4.000069059820334	0.095023109767418
10	4.000000544245374	0.095023109729707
12	4.000000000248837	0.095023109728986
14	4.000000000000012	0.095023109728986
16	3.999999999999999	0.095023109728987
18	3.999999999999986	-
20	4.000000000000001	-

Tabelle 6.8: $\beta = 2.5$ und $\beta = 25$: Approximation des Zielfunktionswertes

Ein Blick auf die in der Tabelle 6.9 zusammengefassten L_{2,ω_β} -Fehler ist interessant. Ein Vergleich der Fehler bei festen N zeigt auf, dass die Fehler bei größerem β kleiner ausfallen. Besonders deutlich ist dies bei kleineren N zu erkennen. So sehen wir zum Beispiel bei $N = 8$ und $\beta = 2.5$, dass sich die Fehler für x_8 und u_8 in einer Größenordnung von 0.01 und $7.4 \cdot 10^{-3}$ befinden, währenddessen die Fehler bei $\beta = 25$ für x_8 und u_8 bereits eine Größenordnung von $5.6 \cdot 10^{-7}$ und $1.1 \cdot 10^{-5}$ aufweisen.

N	$\beta = 2.5$		$\beta = 25$	
	$\ x^* - x_N\ _{2,\omega_\beta}$	$\ u^* - u_N\ _{2,\omega_\beta}$	$\ x^* - x_N\ _{2,\omega_\beta}$	$\ u^* - u_N\ _{2,\omega_\beta}$
8	$0.148657803 \cdot 10^{-1}$	$0.743289016 \cdot 10^{-2}$	$5.65027873 \cdot 10^{-7}$	$0.11891147 \cdot 10^{-4}$
10	$0.131969116 \cdot 10^{-2}$	$0.659845582 \cdot 10^{-3}$	$7.73596316 \cdot 10^{-8}$	$1.62821853 \cdot 10^{-6}$
12	$0.282216513 \cdot 10^{-4}$	$0.141108256 \cdot 10^{-4}$	$6.27653585 \cdot 10^{-10}$	$1.32104800 \cdot 10^{-8}$
14	$1.942643523 \cdot 10^{-7}$	$9.713217616 \cdot 10^{-8}$	$8.07223340 \cdot 10^{-11}$	$1.69900473 \cdot 10^{-9}$
16	$8.908625082 \cdot 10^{-9}$	$4.454312545 \cdot 10^{-9}$	$1.46803875 \cdot 10^{-11}$	$3.08985632 \cdot 10^{-10}$
18	$2.894383473 \cdot 10^{-9}$	$1.447191684 \cdot 10^{-9}$	-	-
20	$5.676017169 \cdot 10^{-13}$	$2.838007475 \cdot 10^{-13}$	-	-

Tabelle 6.9: $\beta = 2.5$ und $\beta = 25$: L_{2,ω_β} -Fehler des Zustandes und der Steuerung

Auch diese Beispiele zeigen, dass die Diskretisierung der integrierten Transversalitätsbedingung einen numerisch stabilen Anfangswert $y_N(0)$ liefert, vgl. Tabelle 6.10. Für $\beta = 2.5$ beträgt der exakte Anfangswert der Adjungierten $y \in W_2^1((0, \infty), \omega_\beta)$

$$y(0) = -4$$

und für $\beta = 25$ gilt

$$y(0) = 21 - \sqrt{445} \approx -0.095023109728988.$$

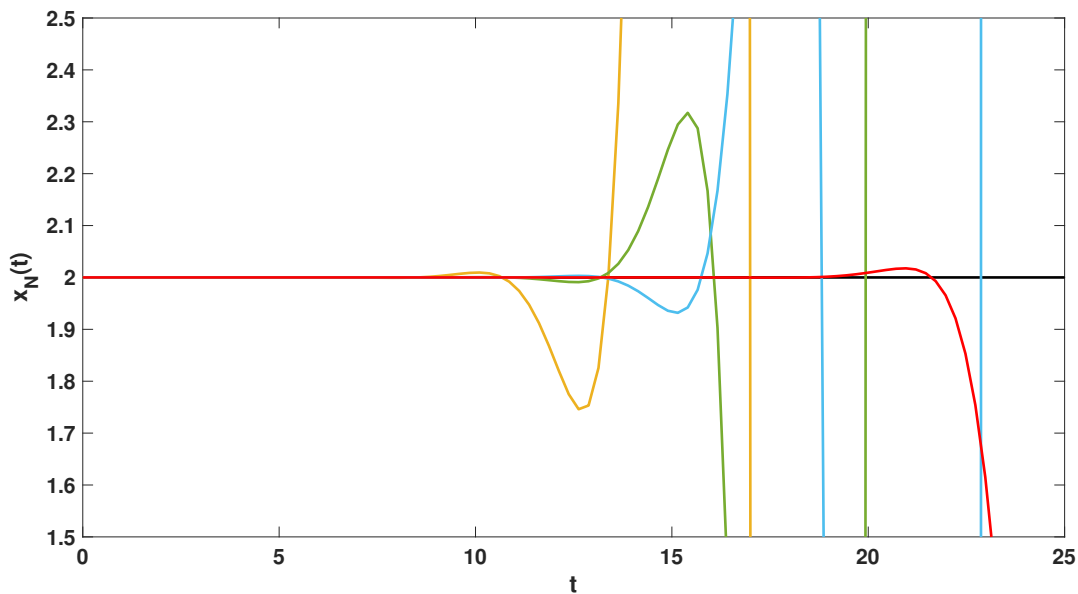
Für $\beta = 25$ weicht der numerische Wert $y_N(0)$ bereits bei einem Polynomgrad von $N = 8$ erst ab der zwölften Nachkommastelle von dem exakten Wert $y(0)$ ab. Ab $N = 10$ wird $y_N(0)$ stabil bis auf 14 Stellen nach dem Komma berechnet. Für $\beta = 2.5$ gelingt bei $N = 8$ sogar eine Genauigkeit bis auf 15 Stellen nach dem Komma.

Die Abbildungen 6.8 und 6.9 illustrieren die Näherungslösungen x_N und u_N für $N = 14, 16, 18, 20$ und für die Wahl $\beta = 2.5$. In den Grafiken 6.10 und 6.11 sind die numerisch

N	$y_N(0)$	
	$\beta = 2.5$	$\beta = 25$
8	-4.0000000000000000	-0.095023109729093
10	-4.0000000000000057	-0.095023109728987
12	-3.999999999999979	-0.095023109728986
14	-4.0000000000000000	-0.095023109728986
16	-4.0000000000000000	-0.095023109728987
18	-4.000000000000027	-
20	-4.0000000000000000	-

Tabelle 6.10: $\beta = 2.5$ und $\beta = 25$: Approximation des Anfangswertes der Adjungierten

ermittelten Approximationen x_N und u_N für $N = 10, 12, 14, 16$ mit $\beta = 25$ dargestellt. Wie auch bei den bisherigen Beispielen verdeutlichen die Grafiken für $\beta = 2.5$, dass die Näherungslösungen die exakten Lösungen zunehmend auf einem Intervall beginnend bei $t = 0$ und endend bei etwa $t = 19$ überdecken, wobei der letzte Kollokationspunkt für $N = 20$ durch $\tau_{20}^{20} \approx 26.60$ gegeben ist. Ist $\beta = 25$, so liegt der letzte Kollokationspunkt für $N = 20$ bei $\tau_{20}^{20} \approx 2.06$ und ist damit deutlich kleiner als in dem Fall $\beta = 2.5$. Entsprechend springen die Näherungslösungen zeitiger von den exakten Lösungen ab, was in den Abbildungen 6.10 und 6.11 sehr gut zu erkennen ist.

Abbildung 6.8: $\beta = 2.5$: Näherungslösung $x_N(t)$ des Zustandes für $N = 14$ (—), $N = 16$ (—), $N = 18$ (—), $N = 20$ (—) im Vergleich mit der exakten Lösung $x^*(t)$ (—)

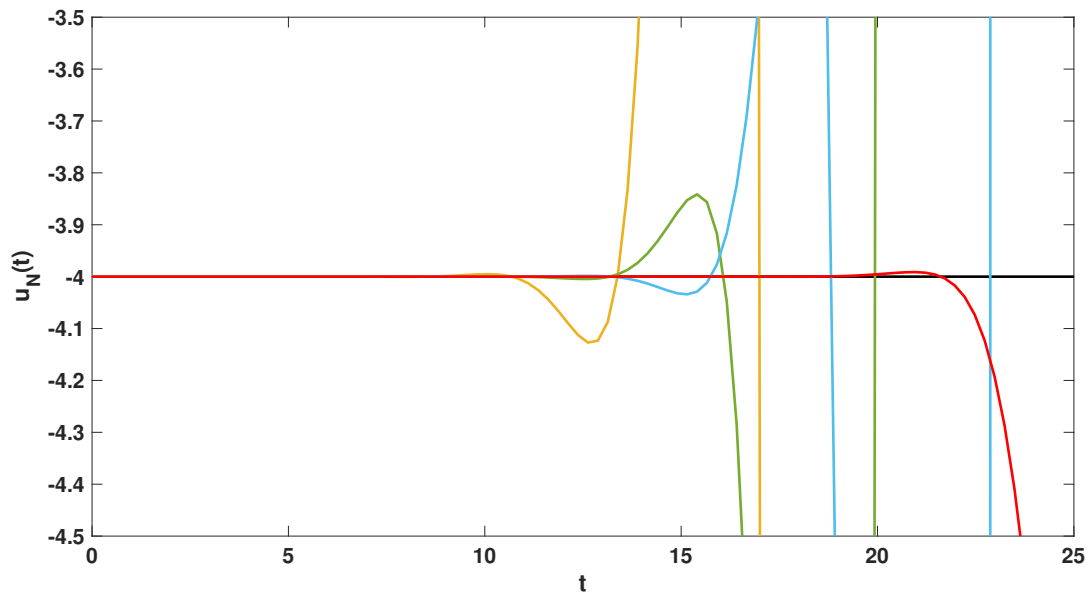


Abbildung 6.9: $\beta = 2.5$: Näherungslösung $u_N(t)$ der Steuerung für $N = 14$ (—), $N = 16$ (—), $N = 18$ (—), $N = 20$ (—) im Vergleich mit der exakten Lösung $u^*(t)$ (—)

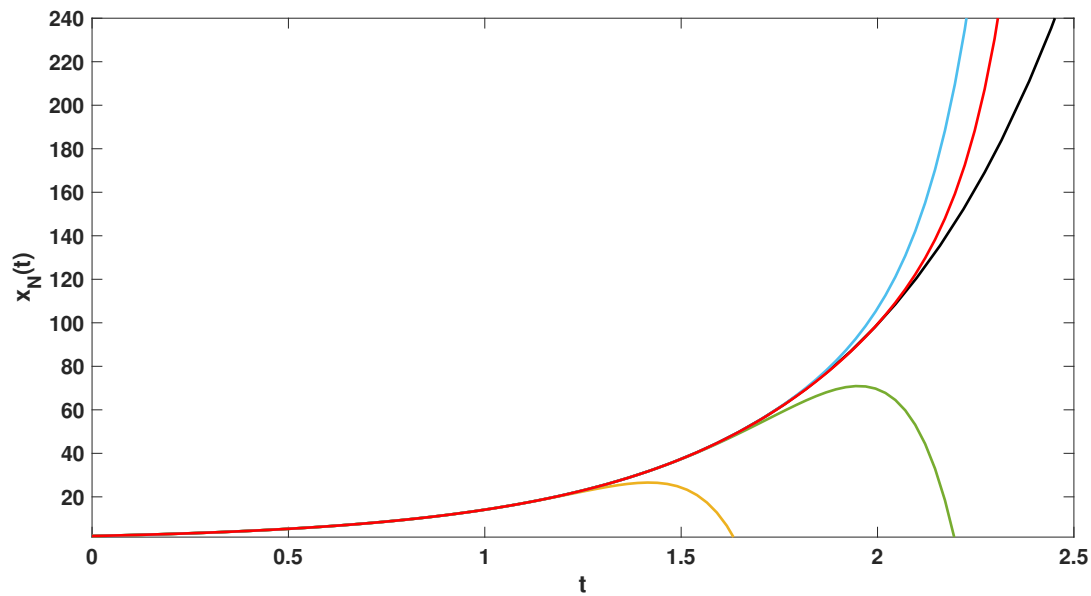


Abbildung 6.10: $\beta = 25$: Näherungslösung $x_N(t)$ des Zustandes für $N = 10$ (—), $N = 12$ (—), $N = 14$ (—), $N = 16$ (—) im Vergleich mit der exakten Lösung $x^*(t)$ (—)

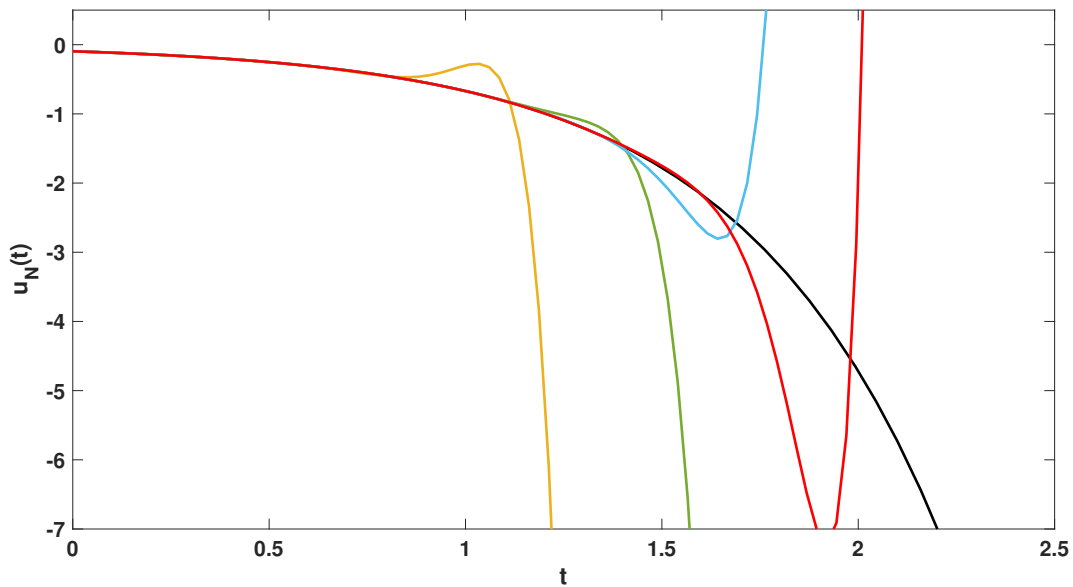


Abbildung 6.11: $\beta = 25$: Näherungslösung $u_N(t)$ der Steuerung für $N = 10$ (—), $N = 12$ (—), $N = 14$ (—), $N = 16$ (—) im Vergleich mit der exakten Lösung $u^*(t)$ (—)

6.3 Mathematisches Pendel

Das mathematische Pendel ist ein klassisches Beispiel aus der Mechanik zur Beschreibung von Pendelschwingungen, vgl. z.B. [Heu04], [GJ16]. Ein Massenpunkt M mit der Masse m wird durch einen starren, masselosen Stab der Länge $l = 1$ an einem festen Punkt aufgehängt, vgl. Abbildung 6.12. Der Massenpunkt kann hierbei frei (in nur einer Ebene) hin und her schwingen; Reibungseffekte im Aufhängepunkt und der Luftwiderstand werden vernachlässigt.

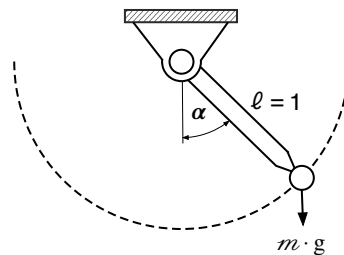


Abbildung 6.12: Mathematisches Pendel

Die Bewegungsgleichung lässt sich wie folgt mathematisch herleiten: Sei α der Winkel zwischen der Pendelstange und der Vertikalen. Nach Auslenkung mit dem Winkel α wirkt auf

M lediglich die Schwerkraft, genauer: ihr Anteil in Richtung der Tangente an die Kreisbahn, die das Pendel macht. Mit der Erdbeschleunigung $g = 9.81 \text{ m/s}^2$ ist diese tangentielle Komponente durch $-g m \sin(\alpha)$ gegeben. Die Bogenlänge s ist gleich $l \cdot \alpha$ und gemeinsam mit dem Newtonschen Kraftgesetz $F = m \cdot a$ (a ist die Beschleunigung) ergibt sich in Abhängigkeit der Zeit t für M die Bewegungsgleichung

$$-g m \sin(\alpha(t)) = m \ddot{\alpha}(t).$$

Die obige nichtlineare Differentialgleichung zweiter Ordnung lässt sich mit Hilfe der Hilfsvariablen $x = (x_1, x_2)^T := (\alpha, \dot{\alpha})^T \in \mathbb{R}^2$ als System von zwei gekoppelten Differentialgleichungen erster Ordnung schreiben:

$$\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} = \underbrace{\begin{pmatrix} x_2(t) \\ -g \sin(x_1(t)) \end{pmatrix}}_{=: \varphi(x(t))}. \quad (6.9)$$

Das System (6.9) besitzt unendlich viele Gleichgewichtspunkte $x_k^* = (k\pi, 0)^T$, $k \in \mathbb{Z}$, vgl. [GJ16], S. 3 ff., d.h. es ist

$$\varphi(x_k^*(t)) = 0 \quad \forall t \in \mathbb{R}.$$

Für gerade k beschreiben alle x_k^* das ruhig nach unten hängende Pendel; ist k ungerade, so wird die Situation des senkrecht nach oben stehenden Pendels dargestellt. Wir erkennen leicht, dass auch schon die kleinste Störung des nach oben stehenden Pendels ein Umfallen des Pendels bewirkt und sich das Pendel unabhängig von der Stärke des Stoßes weit von dem Gleichgewicht entfernt. Entsprechend spricht man in diesem Fall von einem instabilen Gleichgewicht. Hingegen führt eine leichte Störung des senkrecht nach unten hängenden Pendels lediglich dazu, dass das Pendel beginnt, leicht zu pendeln. Daher sind die Gleichgewichte x_k^* für gerade k stabil. Diese physikalischen Interpretationen lassen sich mathematisch genau formulieren und können zum Beispiel in [GJ16], Kapitel 7 und 8 nachgeschlagen werden.

Wir interessieren uns an dieser Stelle für die Linearisierung des Systems (6.9) in dem instabilen Gleichgewicht $x_1^* = (\pi, 0)^T$:

$$\dot{x}(t) = \underbrace{\begin{pmatrix} 0 & 1 \\ g & 0 \end{pmatrix}}_{=: A} x(t). \quad (6.10)$$

Das linearisierte System beschreibt approximativ die Abweichung des Pendels von der Ruhelage und besitzt das instabile Gleichgewicht $x_G = (0, 0)^T$, welches dem Gleichgewicht x_1^* des nichtlinearen Systems entspricht. Daher ergibt sich auch bei dem linearisierten System die instabile Situation, dass sich die Lösungen von (6.10) für Anfangswerte $x_0 \approx 0$, $x_0 \neq 0$ vom Gleichgewicht entfernen, was dem Umfallen des Pendels entspricht. Unser Anliegen ist es, dieses Umfallen zu verhindern und dafür das linearisierte System gezielt durch eine

Steuerung u von außen zu beeinflussen:

$$\dot{x}(t) = \begin{pmatrix} 0 & 1 \\ g & 0 \end{pmatrix} x(t) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} u(t). \quad (6.11)$$

Stellen wir uns vor, dass das Pendel zum Anfangszeitpunkt $t_0 = 0$ durch einen kleinen Stoß aus der Ruhelage x_G gebracht wird, so soll die Steuerung das Pendel derart ausbalancieren, sodass die Lösungen nahe bei dem Gleichgewicht bleiben. Wir modellieren das Anstoßen des Pendels, indem wir einen Anfangswert $x(0) \in \mathbb{R}^2$ wählen, der nicht mit dem Gleichgewicht $x_G = (0, 0)^T$ übereinstimmt, sondern von diesen etwas entfernt ist. Dazu setzen wir $x(0) = (\delta, 0)^T$, $\delta > 0$ und suchen nach einer Steuerung, welche das dynamische System ausgehend von diesem Anfangswert asymptotisch stabilisiert.

Wie in dem Abschnitt 2.5 bereits angekündigt, möchten wir unser Vorgehen zum Auffinden einer asymptotisch stabilisierenden Steuerung an dem Beispiel des mathematischen Pendels demonstrieren. Wir verfolgen wir dabei den Ansatz, ein linear-quadratisches Optimalsteuerungsproblem über dem unendlichen Zeithorizont $[0, \infty)$ zu formulieren und dabei geeignet gewichtete Funktionenräume zu verwenden. Dabei lässt sich das Steuerungsproblem als Hilfsproblem zum Auffinden der gesuchten, das dynamische System stabilisierenden, Steuerung interpretieren.

Für die Formulierung der Steuerungsaufgabe definieren wir zunächst die Diagonalmatrix $D \in \mathbb{R}^{2 \times 2}$, deren Einträge die (reellen) Eigenwerte von A sind und die aus der Diagonalisierung der Systemmatrix $A = T \cdot D \cdot T^{-1}$ resultierende Transformationsmatrix $T \in \mathbb{R}^{2 \times 2}$,

$$D := \begin{pmatrix} -\sqrt{g} & 0 \\ 0 & \sqrt{g} \end{pmatrix}, \quad T := \begin{pmatrix} -\frac{1}{\sqrt{g}} & \frac{1}{\sqrt{g}} \\ 1 & 1 \end{pmatrix}.$$

Mit Hilfe der Zustandstransformation $\tilde{x} = T^{-1} \cdot x$ erhalten wir die transformierte Systemgleichung

$$\dot{\tilde{x}}(t) = D \cdot \tilde{x}(t) + T^{-1} \cdot u(t)$$

und formulieren das folgende Optimalsteuerungsproblem:

Minimiere

$$J(\tilde{x}, u) = \frac{1}{2} \int_0^{\infty} \{ \tilde{x}^T(t) W \tilde{x}(t) + u^T(t) R u(t) \} e^{\beta t} dt$$

bezüglich

$$(\tilde{x}, u) \in W_2^{1,2} \left((0, \infty), e^{\beta t} \right) \times L_2^2 \left((0, \infty), e^{\beta t} \right), \beta > 0,$$

$$\begin{aligned} \dot{\tilde{x}}(t) &= D \tilde{x}(t) + T^{-1} u(t) && \text{f.ü. auf } (0, \infty), \\ \tilde{x}(0) &= \tilde{x}_0(\delta), u(t) \in \mathbb{R}^2 && \text{f.ü. auf } (0, \infty) \end{aligned} \quad (6.12)$$

mit den Ansätzen

$$W := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad R := \frac{1}{\varepsilon} \begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix}, \quad \varepsilon := 0.1$$

und dem von dem Störparameter $\delta > 0$ abhängigen Anfangswert

$$\tilde{x}_0(\delta) := \left(-\frac{\sqrt{g} \cdot \delta}{2}, \frac{\sqrt{g} \cdot \delta}{2} \right)^T.$$

Bemerkung 6.1.

- Im Unterschied zu der in den bisher betrachteten Aufgaben der Klasse $(LQ)_{(\infty)}$ auftauchenden Gewichtsfunktion $\omega_\beta(t) := e^{-\beta t}$, $\beta > 0$ beinhaltet die Formulierung der obigen linear-quadratischen Steuerungsaufgabe das dazu inverse Gewicht $\omega_\beta^{-1}(t) := e^{\beta t}$, $\beta > 0$. Diese Wahl resultiert aus dem Kontext der Stabilitätstheorie (vgl. Abschnitt 2.5), das heißt, aus dem Ziel, Lösungen mit der Eigenschaft

$$\lim_{T \rightarrow \infty} \tilde{x}(T) = 0$$

zu finden. Die Raumwahl $\tilde{x} \in W_2^{1,2}((0, \infty), \omega_\beta^{-1})$ garantiert uns das geforderte asymptotische Verhalten des dynamischen Systems, vgl. Lemma 2.46.

- Die entsprechende Aufgabenklasse, in welche die obige Formulierung fällt, haben wir in [PB18] studiert. Wir greifen auf die dort gewonnenen Resultate zurück und nutzen diese im Folgenden zur Analyse der Optimalsteuerungsaufgabe in (6.12).
- Die Spektralnorm von D ist $|D|_2 = \sqrt{g} > 0$. Die Matrizen W, R sind so gewählt, dass alle in [PB18] formulierten Regularitätsvoraussetzungen erfüllt sind. Es folgt die Existenz der eindeutig bestimmten Lösung der Optimalsteuerungsaufgabe in (6.12), vgl. [PB18], Theorem 1, und die Gültigkeit eines Pontryaginschen Maximumprinzips, vgl. [PB18], Theorem 2. Die resultierenden notwendigen Bedingungen sind gleichzeitig auch hinreichend für den global optimalen Steuerprozess im Sinne der Definition 4.1, vgl. [PB18], Theorem 3.

6.3.1 Diskussion der notwendigen Bedingungen

Wir verwenden das Maximumprinzip aus [PB18] (vgl. Theorem 2), um einen Lösungskandidaten der Optimalsteuerungsaufgabe in (6.12) zu ermitteln. Die Pontryaginsche Funktion für diese Aufgabe lautet

$$H(t, \xi, v, \eta, 1) = -\frac{1}{2}(\xi^T W \xi + v^T R v)e^{\beta t} + \eta^T (D\xi + T^{-1}v).$$

Für die Lösung $(\tilde{x}^*, u^*) \in W_2^{1,2}((0, \infty), \omega_\beta^{-1}) \times L_2^2((0, \infty), \omega_\beta^{-1})$ existieren ein Multiplikator $\lambda_0 = 1$ und eine adjungierte Funktion $y \in W_2^1((0, \infty), \omega_\beta)$, welche die Maximumbedingung

$$u^*(t) = R^{-1}(T^{-1})^T y(t) e^{-\beta t} \quad \text{f.ü. auf } (0, \infty) \quad (6.13)$$

erfüllen und das kanonische Differentialgleichungssystem

$$\begin{aligned} \dot{\tilde{x}}^*(t) &= D\tilde{x}^*(t) + T^{-1}u(t) & \text{f.ü. auf } (0, \infty), \\ \dot{y}(t) &= W\tilde{x}^*(t)e^{\beta t} - Dy(t) & \text{f.ü. auf } (0, \infty) \end{aligned}$$

lösen. Durch Einsetzen der Maximumbedingung und mit Hilfe der Transformation $\hat{x}(t) := \tilde{x}^*(t)e^{\beta t}$ überführen wir die kanonischen Differentialgleichungen in das folgende autonome System:

$$\begin{aligned} \dot{\hat{x}}(t) &= [D + \beta \cdot I_{2 \times 2}] \hat{x}(t) + T^{-1}R^{-1}(T^{-1})^T y(t) & \text{f.ü. auf } (0, \infty), \\ \dot{y}(t) &= W\hat{x}(t) - Dy(t) & \text{f.ü. auf } (0, \infty). \end{aligned} \quad (6.14)$$

Bemerkung 6.2. Bei den vorangegangenen Beispielen haben wir eine Transformation der Adjungierten verwendet. Dies führte dazu, dass wir die kanonischen Differentialgleichungen in dem gewichteten Sobolev-Raum $W_2^1((0, \infty), \omega_\beta)$ studieren konnten. Da die verallgemeinerten Laguerre-Polynome $\mathcal{L}_l^{(\beta)}$ ein vollständiges Orthogonalsystem in diesem Raum bilden, war dies der Ausgangspunkt für den Approximationsansatz in der Spektralmethode. Bei der aus der Stabilisierung resultierenden Optimalsteuerungsaufgabe aus (6.12) liegt der Zustandsraum $W_2^{1,2}((0, \infty), \omega_\beta^{-1})$ zugrunde. Um ein System von Differentialgleichungen zu erhalten, welches in dem Funktionenraum $W_2^{1,2}((0, \infty), \omega_\beta) \times W_2^{1,2}((0, \infty), \omega_\beta)$ formuliert ist, haben wir deshalb an dieser Stelle eine Zustandstransformation genutzt. Dies ermöglicht es uns an späterer Stelle die gesuchte asymptotisch stabilisierende Steuerung mit Hilfe des gewohnten Approximationsansatzes zu bestimmen.

Die Lösungen $(\hat{x}, y) \in W_2^{1,2}((0, \infty), \omega_\beta) \times W_2^{1,2}((0, \infty), \omega_\beta)$ des autonomen Systems erhalten wir wieder durch die Berechnung der Eigenwerte und Eigenvektoren der zugehörigen Koeffizientenmatrix $M \in \mathbb{R}^{4 \times 4}$. Für das gegebene System lautet diese

$$M := \begin{pmatrix} \beta - \sqrt{g} & 0 & \frac{\varepsilon}{2} & 0 \\ 0 & \beta + \sqrt{g} & 0 & \frac{\varepsilon}{2} \\ 1 & 0 & \sqrt{g} & 0 \\ 0 & 1 & 0 & -\sqrt{g} \end{pmatrix}$$

und es ergeben sich die von dem Parameter $\beta > 0$ abhängigen Eigenwerte

$$\begin{aligned} \lambda_1(\beta) &= \frac{\beta}{2} + \frac{1}{2} \sqrt{\beta^2 + 4\sqrt{g}\beta + 2\varepsilon + 4g}, & \lambda_2(\beta) &= \frac{\beta}{2} - \frac{1}{2} \sqrt{\beta^2 + 4\sqrt{g}\beta + 2\varepsilon + 4g} \\ \lambda_3(\beta) &= \frac{\beta}{2} + \frac{1}{2} \sqrt{\beta^2 - 4\sqrt{g}\beta + 2\varepsilon + 4g}, & \lambda_4(\beta) &= \frac{\beta}{2} - \frac{1}{2} \sqrt{\beta^2 - 4\sqrt{g}\beta + 2\varepsilon + 4g} \end{aligned}$$

und Eigenvektoren

$$\begin{aligned} \mathbf{v}_1(\beta) &= \left(0, \frac{\varepsilon}{k_1(\beta) - \beta - 2\sqrt{g}}, 0, 1 \right)^T, & \mathbf{v}_2(\beta) &= \left(0, \frac{\varepsilon}{-k_1(\beta) - \beta - 2\sqrt{g}}, 0, 1 \right)^T, \\ \mathbf{v}_3(\beta) &= \left(\frac{\varepsilon}{k_2(\beta) - \beta + 2\sqrt{g}}, 0, 1, 0 \right)^T, & \mathbf{v}_4(\beta) &= \left(\frac{\varepsilon}{-k_2(\beta) - \beta + 2\sqrt{g}}, 0, 1, 0 \right)^T. \end{aligned}$$

mit

$$k_1(\beta) := \sqrt{\beta^2 + 4\sqrt{g}\beta + 4g + 2\varepsilon}, \quad k_2(\beta) := \sqrt{\beta^2 - 4\sqrt{g}\beta + 4g + 2\varepsilon}.$$

Die Zustände und Adjungierten $\hat{x}^{(1)}, \hat{x}^{(2)}, y^{(1)}, y^{(2)} \in W_2^1((0, \infty), \omega_\beta)$ sind dementsprechend durch die Vorschriften

$$\begin{aligned} \hat{x}^{(1)}(t) &= c_3 \cdot \frac{\varepsilon}{k_2(\beta) - \beta + 2\sqrt{g}} \cdot e^{\lambda_3(\beta)t} + c_4 \cdot \frac{\varepsilon}{-k_2(\beta) - \beta + 2\sqrt{g}} \cdot e^{\lambda_4(\beta)t}, \\ \hat{x}^{(2)}(t) &= c_1 \cdot \frac{\varepsilon}{k_1(\beta) - \beta - 2\sqrt{g}} \cdot e^{\lambda_1(\beta)t} + c_2 \cdot \frac{\varepsilon}{-k_1(\beta) - \beta - 2\sqrt{g}} \cdot e^{\lambda_2(\beta)t}, \\ y^{(1)}(t) &= c_3 \cdot e^{\lambda_3(\beta)t} + c_4 \cdot e^{\lambda_4(\beta)t}, \quad y^{(2)}(t) = c_1 \cdot e^{\lambda_1(\beta)t} + c_2 \cdot e^{\lambda_2(\beta)t} \end{aligned}$$

mit Konstanten $c_1, c_2, c_3, c_4 \in \mathbb{R}$ festgelegt. Um diese unbekanntenen Konstanten zu bestimmen, transformieren wir die Adjungierte $y \in W_2^{1,2}((0, \infty), \omega_\beta)$ gemäß der Vorschrift $y_0(t) := y(t)e^{-\beta t}$, um wie bei den vorangegangenen Beispielen die aus der Raumzugehörigkeit $y_0 \in W_2^{1,2}((0, \infty), \omega_\beta^{-1})$ resultierenden Grenzwertbeziehungen

$$\begin{aligned} \lim_{T \rightarrow \infty} y_0^{(1)}(T) &= \lim_{T \rightarrow \infty} c_3 \cdot e^{(\lambda_3(\beta) - \beta)T} + c_4 \cdot e^{(\lambda_4(\beta) - \beta)T} \stackrel{!}{=} 0, \\ \lim_{T \rightarrow \infty} y_0^{(2)}(T) &= \lim_{T \rightarrow \infty} c_1 \cdot e^{(\lambda_1(\beta) - \beta)T} + c_2 \cdot e^{(\lambda_2(\beta) - \beta)T} \stackrel{!}{=} 0 \end{aligned}$$

zu nutzen. Da für alle $\beta > 0$ die Ungleichungen $\lambda_2(\beta) - \beta < 0$ und $\lambda_4(\beta) - \beta < 0$ gelten, setzen wir $c_1 = c_3 = 0$, um die Gültigkeit der obigen Bedingungen sicherzustellen. Die Konstanten c_2, c_4 lassen sich aus der Anfangsbedingung für den Zustand $\hat{x}(0) = \tilde{x}_0(\delta)$ festlegen: Aus

$$\hat{x}(0) = \begin{pmatrix} c_4 \cdot \frac{\varepsilon}{-k_2(\beta) - \beta + 2\sqrt{g}} \\ c_2 \cdot \frac{\varepsilon}{-k_1(\beta) - \beta - 2\sqrt{g}} \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} -\frac{\sqrt{g}\delta}{2} \\ \frac{\sqrt{g}\delta}{2} \end{pmatrix}$$

erhalten wir

$$c_2 = \frac{\sqrt{g}\delta}{2\varepsilon} \cdot (-k_1(\beta) - \beta - 2\sqrt{g}), \quad c_4 = \frac{-\sqrt{g}\delta}{2\varepsilon} \cdot (-k_2(\beta) - \beta + 2\sqrt{g}).$$

Damit sind die Lösungen $(\hat{x}, y) \in W_2^{1,2}((0, \infty), \omega_\beta) \times W_2^{1,2}((0, \infty), \omega_\beta)$ des Differentialglei-

chungssystem in (6.14) mit $\hat{x}(0) = \tilde{x}_0(\delta)$ durch

$$\hat{x}^{(1)}(t) = \frac{-\sqrt{g}\delta}{2} \cdot e^{\lambda_4(\beta)t}, \quad \hat{x}^{(2)}(t) = \frac{\sqrt{g}\delta}{2} \cdot e^{\lambda_2(\beta)t},$$

$$y^{(1)}(t) = \frac{-\sqrt{g}\delta}{2\varepsilon} \cdot (-k_2(\beta) - \beta + 2\sqrt{g}) \cdot e^{\lambda_4(\beta)t}, \quad y^{(2)}(t) = \frac{\sqrt{g}\delta}{2\varepsilon} \cdot (-k_1(\beta) - \beta - 2\sqrt{g}) \cdot e^{\lambda_2(\beta)t}$$

gegeben. Mit Hilfe der Maximumbedingung (6.13) und durch die Rücktransformation $\tilde{x}^*(t) := \hat{x}(t)e^{-\beta t}$ können wir die Lösungskandidaten $(\tilde{x}^*, u^*) \in W_2^{1,2}((0, \infty), \omega_\beta^{-1}) \times L_2^2((0, \infty), \omega_\beta^{-1})$ für die Aufgabe in (6.12) angeben:

$$\tilde{x}^{*(1)}(t) = \frac{-\sqrt{g}\delta}{2} \cdot e^{(\lambda_4(\beta)-\beta)t}, \quad \tilde{x}^{*(2)}(t) = \frac{\sqrt{g}\delta}{2} \cdot e^{(\lambda_2(\beta)-\beta)t},$$

$$u^{*(1)}(t) = \frac{\delta}{4} \cdot (-k_2(\beta) - \beta + 2\sqrt{g}) \cdot e^{(\lambda_4(\beta)-\beta)t} + \frac{\delta}{4} \cdot (-k_1(\beta) - \beta - 2\sqrt{g}) \cdot e^{(\lambda_2(\beta)-\beta)t}, \quad (6.15)$$

$$u^{*(2)}(t) = \frac{-\sqrt{g}\delta}{4} \cdot (-k_2(\beta) - \beta + 2\sqrt{g}) \cdot e^{(\lambda_4(\beta)-\beta)t} + \frac{\sqrt{g}\delta}{4} \cdot (-k_1(\beta) - \beta - 2\sqrt{g}) \cdot e^{(\lambda_2(\beta)-\beta)t}. \quad (6.16)$$

Wir haben bereits erwähnt, dass das Maximumprinzip für die Aufgabe in (6.12) sogar hinreichende Kriterien bereitstellt. Daher ist durch die ermittelten Kandidaten die eindeutig bestimmte Lösung der Aufgabe aus (6.12) gegeben. Da $y \cdot e^{-\beta t} \in W_2^{1,2}((0, \infty), \omega_\beta^{-1})$ gilt, erhalten wir aus der Berechnungsvorschrift in (6.13) eine stetige und beschränkte Steuerung u^* , welche in Abhängigkeit der Störung $\delta > 0$ und des Parameters $\beta > 0$ gegeben ist. Darüber hinaus stabilisiert die in (6.15)-(6.16) angegebene Steuerung u^* das linearisierte dynamische System des mathematischen Pendels aus (6.11) exponentiell im Sinne von Lyapunov (vgl. Definition 2.45, S. 41), wobei mit $x = T \cdot \tilde{x}$ der zugehörige stabilisierte Zustand $x^* \in W_2^{1,2}((0, \infty), \omega_\beta^{-1})$ durch

$$x^{*(1)}(t) = \frac{\delta}{2} \cdot e^{(\lambda_4(\beta)-\beta)t} + \frac{\delta}{2} \cdot e^{(\lambda_2(\beta)-\beta)t},$$

$$x^{*(2)}(t) = \frac{-\sqrt{g}\delta}{2} \cdot e^{(\lambda_4(\beta)-\beta)t} + \frac{\sqrt{g}\delta}{2} \cdot e^{(\lambda_2(\beta)-\beta)t}$$

mit

$$\lim_{T \rightarrow \infty} x^{*(1)}(T) = 0, \quad \lim_{T \rightarrow \infty} x^{*(2)}(T) = 0$$

gegeben ist.

6.3.2 Anwenden der Spektralmethode

Das Differentialgleichungssystem aus (6.14) ist für $\hat{x} = (\hat{x}^{(1)}, \hat{x}^{(2)})^T \in W_2^{1,2}((0, \infty), \omega_\beta)$ und $y = (y^{(1)}, y^{(2)})^T \in W_2^{1,2}((0, \infty), \omega_\beta)$ formuliert. Analog zu den bisherigen Beispielen su-

chen wir nach Näherungslösungen $\hat{x}_N^{(1)}, \hat{x}_N^{(2)}, y_N^{(1)}, y_N^{(2)} \in \mathcal{P}_N((0, \infty))$, welche die Differentialgleichungen an den Nullstellen $\tau_i^{(N)}$, $1 \leq i \leq N$ des verallgemeinerten Laguerre-Polynoms $\mathcal{L}_N^{(\beta)}$ exakt erfüllen. Zusammen mit der Anfangsbedingung für den Zustand, der integrierten Transversalitätsbedingung, den Gewichten der Gauß-Quadratur $w_i^{(N)}$, $i = 1, \dots, N$ und dem zusätzlichen Kollokationspunkt $\tau_0 = 0$ erhalten wir die nachfolgenden $4N + 4$ Kollokationsgleichungen:

$$\left\{ \begin{array}{l}
 \dot{\hat{x}}_N^{(1)}(\tau_i^{(N)}) = (\beta - \sqrt{g}) \cdot \hat{x}_N^{(1)}(\tau_i^{(N)}) + \frac{\varepsilon}{2} \cdot y_N^{(1)}(\tau_i^{(N)}), \quad 1 \leq i \leq N \\
 \dot{\hat{x}}_N^{(2)}(\tau_i^{(N)}) = (\beta + \sqrt{g}) \cdot \hat{x}_N^{(2)}(\tau_i^{(N)}) + \frac{\varepsilon}{2} \cdot y_N^{(2)}(\tau_i^{(N)}), \quad 1 \leq i \leq N \\
 \dot{y}_N^{(1)}(\tau_i^{(N)}) = \hat{x}_N^{(1)}(\tau_i^{(N)}) + \sqrt{g} \cdot y_N^{(1)}(\tau_i^{(N)}), \quad 1 \leq i \leq N \\
 \dot{y}_N^{(2)}(\tau_i^{(N)}) = \hat{x}_N^{(2)}(\tau_i^{(N)}) - \sqrt{g} \cdot y_N^{(2)}(\tau_i^{(N)}), \quad 1 \leq i \leq N \\
 \hat{x}_N^{(1)}(\tau_0) = -\frac{\sqrt{g} \cdot \delta}{2}, \quad \hat{x}_N^{(2)}(\tau_0) = \frac{\sqrt{g} \cdot \delta}{2} \\
 y_N^{(1)}(\tau_0) = \sum_{i=1}^N \left\{ (\beta - \sqrt{g}) \cdot y_N^{(1)}(\tau_i^{(N)}) - \hat{x}_N^{(1)}(\tau_i^{(N)}) \right\} w_i^{(N)} \\
 y_N^{(2)}(\tau_0) = \sum_{i=1}^N \left\{ (\beta + \sqrt{g}) \cdot y_N^{(2)}(\tau_i^{(N)}) - \hat{x}_N^{(2)}(\tau_i^{(N)}) \right\} w_i^{(N)}
 \end{array} \right.$$

Die Approximation der gesuchten Steuerung, welche das linearisierte System asymptotisch stabilisiert, berechnen wir im Anschluss direkt aus der Maximumbedingung in (6.13).

6.3.3 Auswertung der numerischen Ergebnisse für $\delta = 0.5$ und $\delta = 3$

Da sich die in (6.12) formulierte Optimalsteuerungsaufgabe als Hilfsaufgabe zum Auffinden einer das linearisierte Pendelsystem asymptotisch stabilisierenden Steuerung interpretieren lässt, ziehen wir bei diesem Beispiel nicht den Zielfunktionswert als Gütekriterium heran. Vielmehr sind wir daran interessiert eine möglichst gute Approximation der gesuchten Steuerung zu finden, weshalb wir hier den gewichteten L_{2, ω_β} -Fehler als Maß zur Bewertung der Approximationsgüte verwenden. Da der Parameter $\beta > 0$ wieder frei zu wählen ist, sind wir bei unseren numerischen Experimenten wie folgt vorgegangen: Für eine gegebene Störung $\delta > 0$ des Anfangswertes $x(0) = (\delta, 0)^T$ haben wir zu Beginn $\beta = 4$ gesetzt. Anschließend berechneten wir die Approximationen für eine steigende Anzahl N des Polynomgrades, für welche wir die entsprechenden L_{2, ω_β} -Fehler zwischen $u_N^{(1)}$ und $u^{*(1)}$ bzw. $u_N^{(2)}$ und $u^{*(2)}$ und deren Summe auswerteten. Konnten wir trotz der Erhöhung von N keine Verbesserung der Fehlerordnung (bezogen auf die Summe der L_{2, ω_β} -Fehler) mehr feststellen, so wählten wir einen größeren Wert von β und wiederholten das Vorgehen. Die Tabellen 6.11 und 6.12 dokumentieren die Ergebnisse für $\delta = 0.5$ und $\delta = 3$. Bei beiden Störungen haben wir $\beta = 4, 20, 50$ und $\beta = 100$ betrachtet.

Für $\delta = 0.5$ wurden die besten Approximationen bei $\beta = 20$ und $N = 18$ erzielt. In diesem Fall ist der letzte Kollokationspunkt $\tau_{18}^{(18)} \approx 2.95$ und die Summe der L_{2, ω_β} -Fehler befindet

sich in einer Größenordnung von $5.9 \cdot 10^{-10}$. Wie auch bei den bisherigen Beispielen führt eine Erhöhung des Polynomgrades N bei jedem Wert von β zu besseren Ergebnissen. Für $\beta = 4$ mussten wir N bis auf 20 erhöhen, um einen Fehler von $6.3 \cdot 10^{-9}$ zu erreichen. Für größere Werte von β erreichten wir diese Fehlerordnung bereits mit einem geringeren Polynomgrad, was mit weniger Kollokationspunkten einhergeht: für $\beta = 20$ genügte es, $N = 16$ zu wählen, um einen Fehler der Größe $5.4 \cdot 10^{-9}$ zu erreichen; bei $\beta = 50$ und $\beta = 100$ erhielten wir $9.4 \cdot 10^{-9}$ und $9.0 \cdot 10^{-9}$ bereits bei $N = 14$. Daraus abgeleitet lässt sich vermuten, dass eine Erhöhung des Parameters β mit einer geringeren Anzahl N und damit mit dem Lösen eines kleineren Gleichungssystems einhergeht, um eine vorgegebene Fehlerschranke zu erreichen. Dennoch zeigen die Einträge der Tabelle 6.12 auch, dass die Approximationsergebnisse nicht automatisch verbessert werden können, wenn β beliebig erhöht wird. Für dieses Beispiel konnte der geringe Fehler bei $\beta = 20$ und $N = 18$ trotz ansteigenden β und N nicht weiter verkleinert werden.

β	N	$\ u^{*(1)} - u_N^{(1)}\ _{2,\omega_\beta}$	$\ u^{*(2)} - u_N^{(2)}\ _{2,\omega_\beta}$	Σ
4	8	$0.249281438 \cdot 10^{-1}$	$0.780760271 \cdot 10^{-1}$	0.1030041710
4	10	$0.135020877 \cdot 10^{-3}$	$0.422832197 \cdot 10^{-3}$	$0.557853074 \cdot 10^{-3}$
4	12	$8.637741579 \cdot 10^{-6}$	$0.269388559 \cdot 10^{-4}$	$0.355765975 \cdot 10^{-4}$
4	14	$4.361207018 \cdot 10^{-6}$	$0.136595909 \cdot 10^{-4}$	$0.180207979 \cdot 10^{-4}$
4	16	$2.523107157 \cdot 10^{-8}$	$7.887505594 \cdot 10^{-8}$	$1.041061275 \cdot 10^{-7}$
4	18	$2.227431695 \cdot 10^{-7}$	$6.977986645 \cdot 10^{-7}$	$9.205418340 \cdot 10^{-7}$
4	20	$1.664953838 \cdot 10^{-9}$	$4.706285246 \cdot 10^{-9}$	$6.371239084 \cdot 10^{-9}$
20	8	$0.192143367 \cdot 10^{-2}$	$0.442296884 \cdot 10^{-2}$	$0.634440251 \cdot 10^{-2}$
20	10	$0.790732086 \cdot 10^{-4}$	$0.263059772 \cdot 10^{-3}$	$0.342132980 \cdot 10^{-3}$
20	12	$1.301744535 \cdot 10^{-6}$	$4.945994781 \cdot 10^{-6}$	$6.247739316 \cdot 10^{-6}$
20	14	$1.125769224 \cdot 10^{-9}$	$1.264535800 \cdot 10^{-8}$	$1.377112722 \cdot 10^{-8}$
20	16	$1.097298401 \cdot 10^{-9}$	$4.353710600 \cdot 10^{-9}$	$5.451009001 \cdot 10^{-9}$
20	18	$3.135595403 \cdot 10^{-10}$	$2.859613677 \cdot 10^{-10}$	$5.995209080 \cdot 10^{-10}$
50	8	$0.228027344 \cdot 10^{-4}$	$0.373430581 \cdot 10^{-3}$	$0.396233316 \cdot 10^{-3}$
50	10	$0.455723350 \cdot 10^{-3}$	$0.132450455 \cdot 10^{-2}$	$0.178022791 \cdot 10^{-2}$
50	12	$4.462699888 \cdot 10^{-8}$	$4.910863675 \cdot 10^{-7}$	$5.357133664 \cdot 10^{-7}$
50	14	$3.227924728 \cdot 10^{-9}$	$6.254322895 \cdot 10^{-9}$	$9.482247623 \cdot 10^{-9}$
100	8	$6.925314090 \cdot 10^{-6}$	$0.241733238 \cdot 10^{-3}$	$0.248658552 \cdot 10^{-3}$
100	10	$8.733138572 \cdot 10^{-6}$	$0.845399410 \cdot 10^{-4}$	$0.932730796 \cdot 10^{-4}$
100	12	$1.764436980 \cdot 10^{-6}$	$4.771700226 \cdot 10^{-6}$	$6.536137206 \cdot 10^{-6}$
100	14	$1.287714240 \cdot 10^{-9}$	$7.784176059 \cdot 10^{-9}$	$9.071890299 \cdot 10^{-9}$

Tabelle 6.11: L_{2,ω_β} -Fehler der Steuerung für $\delta = 0.5$ und $\beta = 4, 20, 50, 100$

Die Tabelle 6.12 beinhaltet die numerischen Ergebnisse für $\delta = 3$. In diesem Fall konnten die besten Ergebnisse mit einem Fehler der Größenordnung von $9.9 \cdot 10^{-10}$ für $\beta = 50$ und $N = 24$ ermittelt werden. Der letzte Kollokationspunkt hierbei ist $\tau_{(24)}^{24} \approx 1.62$. Ähnlich wie bei den Ergebnissen für $\delta = 0.5$ ging hier die Erhöhung von β mit einer geringeren Anzahl von N einher. Zum Beispiel konnte bei $\beta = 4$ ein Fehler von $3.4 \cdot 10^{-8}$ mit 20 Kollokationspunkten ermittelt werden, während für $\beta = 100$ nur 14 Punkte genügten, um einen Fehler von $5.4 \cdot 10^{-8}$ zu erzielen.

Positiv hervorzuheben ist, dass sowohl für $\delta = 0.5$ also auch für $\delta = 3$ sehr gute Approximationsergebnisse berechnet werden konnten. Ein Vergleich der jeweils besten Ergebnisse macht deutlich, dass die geringe Störung einen kleineren Wert für β und einen geringeren Polynomgrad N erfordert: um einen Fehler von $5.9 \cdot 10^{-10}$ zu erhalten musste $\beta = 20$ und $N = 18$ gesetzt werden. Für die deutlich größere Störung $\delta = 3$ war es notwendig, $\beta = 50$ und $N = 24$ zu wählen, um ein ähnliches Resultat, nämlich $9.9 \cdot 10^{-10}$, zu erreichen.

β	N	$\ u^{*(1)} - u_N^{(1)}\ _{2,\omega_\beta}$	$\ u^{*(2)} - u_N^{(2)}\ _{2,\omega_\beta}$	Σ
4	8	0.1495688552	0.4684561969	0.6180250521
4	10	$0.810127021 \cdot 10^{-3}$	$0.253699675 \cdot 10^{-2}$	$0.334712377 \cdot 10^{-2}$
4	12	$0.518250175 \cdot 10^{-4}$	$0.161631785 \cdot 10^{-3}$	$0.213456802 \cdot 10^{-3}$
4	14	$0.261670990 \cdot 10^{-4}$	$0.819574080 \cdot 10^{-4}$	$0.108124507 \cdot 10^{-3}$
4	16	$1.513309528 \cdot 10^{-7}$	$4.732023272 \cdot 10^{-7}$	$6.245332800 \cdot 10^{-7}$
4	18	$1.336505923 \cdot 10^{-6}$	$4.186839995 \cdot 10^{-6}$	$5.523345918 \cdot 10^{-6}$
4	20	$8.096121774 \cdot 10^{-9}$	$2.640884092 \cdot 10^{-8}$	$3.450496269 \cdot 10^{-8}$
20	8	$0.115285961 \cdot 10^{-1}$	$0.265377953 \cdot 10^{-1}$	$0.380663915 \cdot 10^{-1}$
20	10	$0.474439665 \cdot 10^{-3}$	$0.157835864 \cdot 10^{-2}$	$0.205279831 \cdot 10^{-2}$
20	12	$7.810540735 \cdot 10^{-6}$	$0.296766383 \cdot 10^{-4}$	$0.374871790 \cdot 10^{-4}$
20	14	$6.485556448 \cdot 10^{-9}$	$7.669599076 \cdot 10^{-8}$	$8.318154721 \cdot 10^{-8}$
20	16	$6.353100094 \cdot 10^{-9}$	$2.770586503 \cdot 10^{-8}$	$3.405896512 \cdot 10^{-8}$
20	18	$5.189030725 \cdot 10^{-10}$	$8.810683942 \cdot 10^{-9}$	$9.329587014 \cdot 10^{-9}$
50	8	$0.136816336 \cdot 10^{-3}$	$0.224058405 \cdot 10^{-2}$	$0.237740039 \cdot 10^{-2}$
50	10	$0.273433782 \cdot 10^{-2}$	$0.794701895 \cdot 10^{-2}$	$0.106813567 \cdot 10^{-1}$
50	12	$2.677889682 \cdot 10^{-7}$	$2.945999604 \cdot 10^{-6}$	$3.213788572 \cdot 10^{-6}$
50	14	$1.936228079 \cdot 10^{-8}$	$3.741886822 \cdot 10^{-8}$	$5.678114901 \cdot 10^{-8}$
50	16	$1.360470814 \cdot 10^{-9}$	$9.322955678 \cdot 10^{-9}$	$1.068342649 \cdot 10^{-8}$
50	18	$4.359079672 \cdot 10^{-11}$	$1.381114315 \cdot 10^{-9}$	$1.424705112 \cdot 10^{-9}$
50	20	$8.330523309 \cdot 10^{-11}$	$1.191685571 \cdot 10^{-9}$	$1.274990804 \cdot 10^{-9}$
50	22	$4.532390656 \cdot 10^{-10}$	$1.006082797 \cdot 10^{-9}$	$1.459321863 \cdot 10^{-9}$
50	24	$1.169041393 \cdot 10^{-10}$	$8.780354255 \cdot 10^{-10}$	$9.949395648 \cdot 10^{-10}$
100	8	$0.415514445 \cdot 10^{-4}$	$0.145039758 \cdot 10^{-2}$	$0.14919490 \cdot 10^{-2}$
100	10	$0.523993989 \cdot 10^{-4}$	$0.507238241 \cdot 10^{-3}$	$0.55963764 \cdot 10^{-3}$
100	12	$0.105869057 \cdot 10^{-4}$	$0.286294809 \cdot 10^{-4}$	$0.39216386 \cdot 10^{-4}$
100	14	$8.356407451 \cdot 10^{-9}$	$4.622111077 \cdot 10^{-8}$	$5.45775182 \cdot 10^{-8}$

Tabelle 6.12: L_{2,ω_β} -Fehler der Steuerung für $\delta = 3$ und $\beta = 4, 20, 50, 100$

Die Näherungslösung für die Steuerung haben wir direkt durch Auswertung der Maximumbedingung (6.13) berechnet. Entsprechend beeinflusst die Güte der numerisch ermittelten Adjungierten die Approximationsgüte der Steuerungen. Da die Approximationen der Adjungierten wiederum durch die integrierte Transversalitätsbedingungen bestimmt werden, möchten wir abschließend noch einen Blick auf die Genauigkeit der numerisch ermittelten Anfangswerte $y_N^{(1)}(0), y_N^{(2)}(0)$ werfen: Für $\delta = 0.5$, $\beta = 20$ und $N = 18$ ist

$$y_N^{(1)}(0) = 2.151661860993174 \cdot 10^2, \quad y_N^{(2)}(0) = -4.113390064444905 \cdot 10^2$$

und

$$y^{(1)}(0) = 2.151661860994834 \cdot 10^2, \quad y^{(2)}(0) = -4.113390064444913 \cdot 10^2.$$

Die Anfangswerte $y^{(1)}(0)$ und $y^{(2)}(0)$ konnten in dem obigen Fall bis auf 9 und 11 Nachkommastellen exakt berechnet werden. Für $\delta = 3$, $\beta = 50$ und $N = 24$ erhielten wir

$$y_N^{(1)}(0) = 4.109645346942826 \cdot 10^3, \quad y_N^{(2)}(0) = -5.286821429128540 \cdot 10^3.$$

Ein Vergleich mit den analytisch ermittelten Werten

$$y^{(1)}(0) = 4.109645347038001 \cdot 10^3, \quad y^{(2)}(0) = -5.286821429087426 \cdot 10^3$$

zeigt, dass hierbei die numerischen Anfangswerte bis auf 5 und 6 Stellen nach dem Komma mit den Werten von $y^{(1)}(0)$ und $y^{(2)}(0)$ übereinstimmen.

Abschließend veranschaulichen die Grafiken auf den nachfolgenden Seiten die Näherungen $u_N^{(1)}, u_N^{(2)}$ der gesuchten asymptotisch stabilisierenden Steuerung $u^{*(1)}, u^{*(2)}$ für $\delta = 0.5$ und $\delta = 3$, vgl. Abbildung 6.13 und Abbildung 6.15. Die zugehörigen Näherungen $x_N^{(1)}, x_N^{(2)}$ des linearisierten Pendelsystems sind in den Abbildungen 6.14 und 6.16 zu finden. Da bei allen Grafiken lediglich zu Beginn des Zeitintervalls $[0, \infty)$ geringe Abweichungen zu den exakten Lösungen zu erkennen sind, haben wir das Intervall $[0, 5]$ zugrunde gelegt. Wir haben die Grafiken auch für $t > 5$ betrachtet, jedoch war dort optisch kein Unterschied mehr zwischen den numerischen und exakten Lösungen zu erkennen. Insbesondere wurde dabei deutlich, dass die Zustände für $T \rightarrow \infty$ gegen Null konvergieren, also jeweils

$$\lim_{T \rightarrow \infty} x_N^{(1)}(T) = 0, \quad \lim_{T \rightarrow \infty} x_N^{(2)}(T) = 0$$

erfüllt ist.

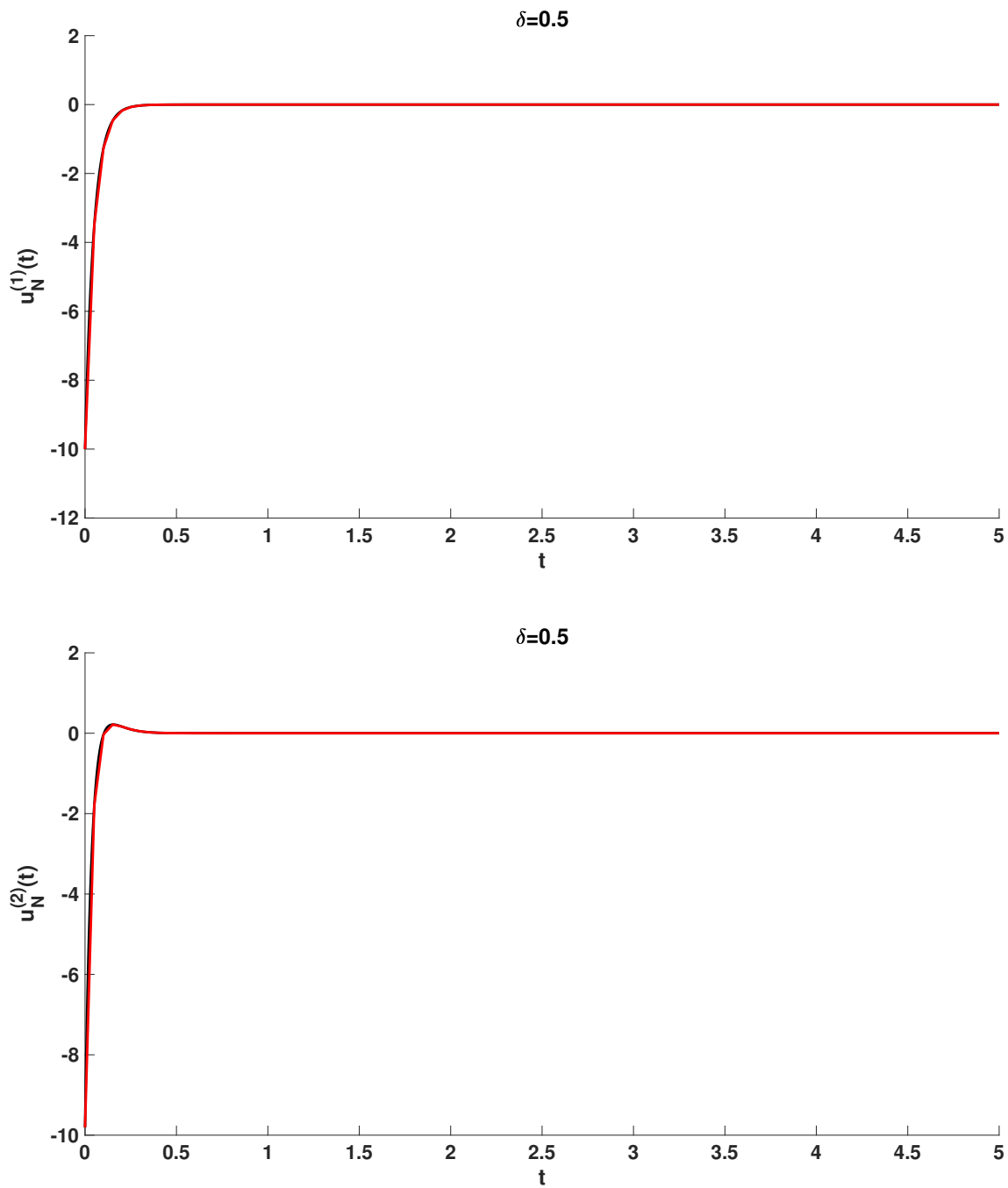


Abbildung 6.13: $\delta = 0.5$ und $\beta = 20$: Näherungslösungen $u_N^{(1)}(t), u_N^{(2)}(t)$ der Steuerungen für $N = 18$ (—) im Vergleich mit den exakten Lösungen $u^{*(1)}(t), u^{*(2)}(t)$ (—)

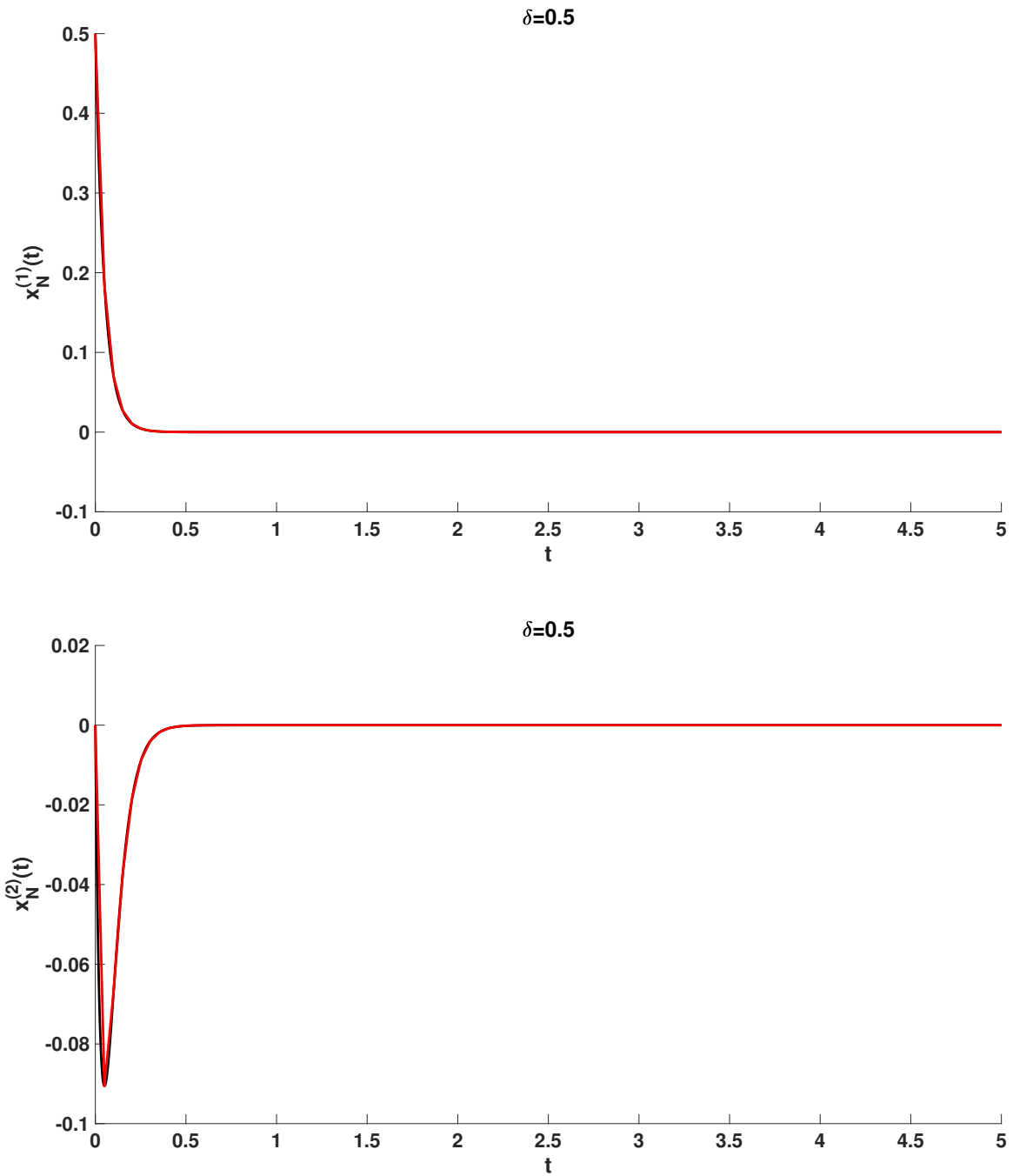


Abbildung 6.14: $\delta = 0.5$ und $\beta = 20$: Näherungslösungen $x_N^{(1)}(t), x_N^{(2)}(t)$ der Zustände für $N = 18$ (—) im Vergleich mit den exakten Lösungen $x^{*(1)}(t), x^{*(2)}(t)$ (—)

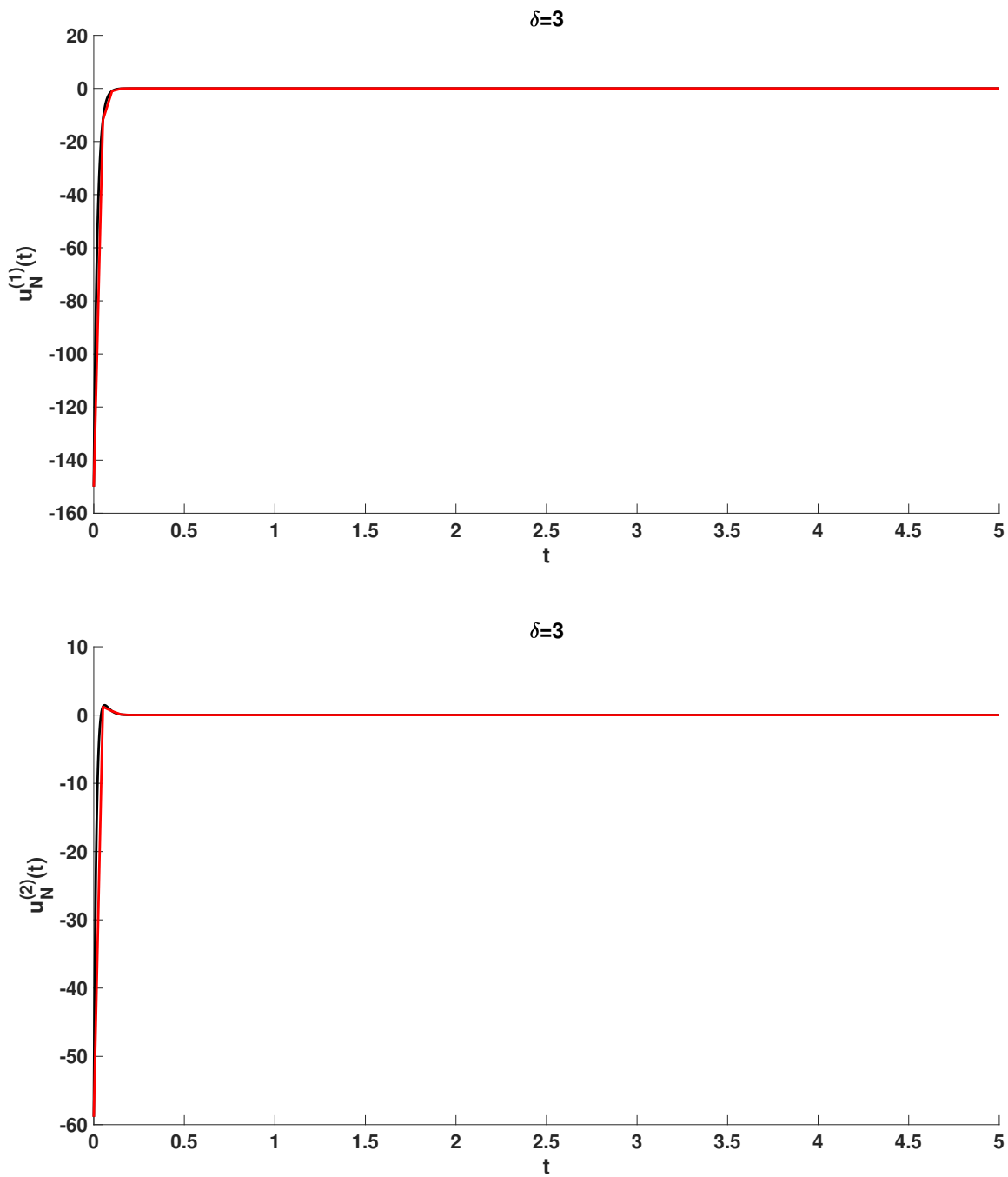


Abbildung 6.15: $\delta = 3$ und $\beta = 50$: Näherungslösungen $u_N^{(1)}(t), u_N^{(2)}(t)$ der Steuerungen für $N = 24$ (—) im Vergleich mit den exakten Lösungen $u^{*(1)}(t), u^{*(2)}(t)$ (—)

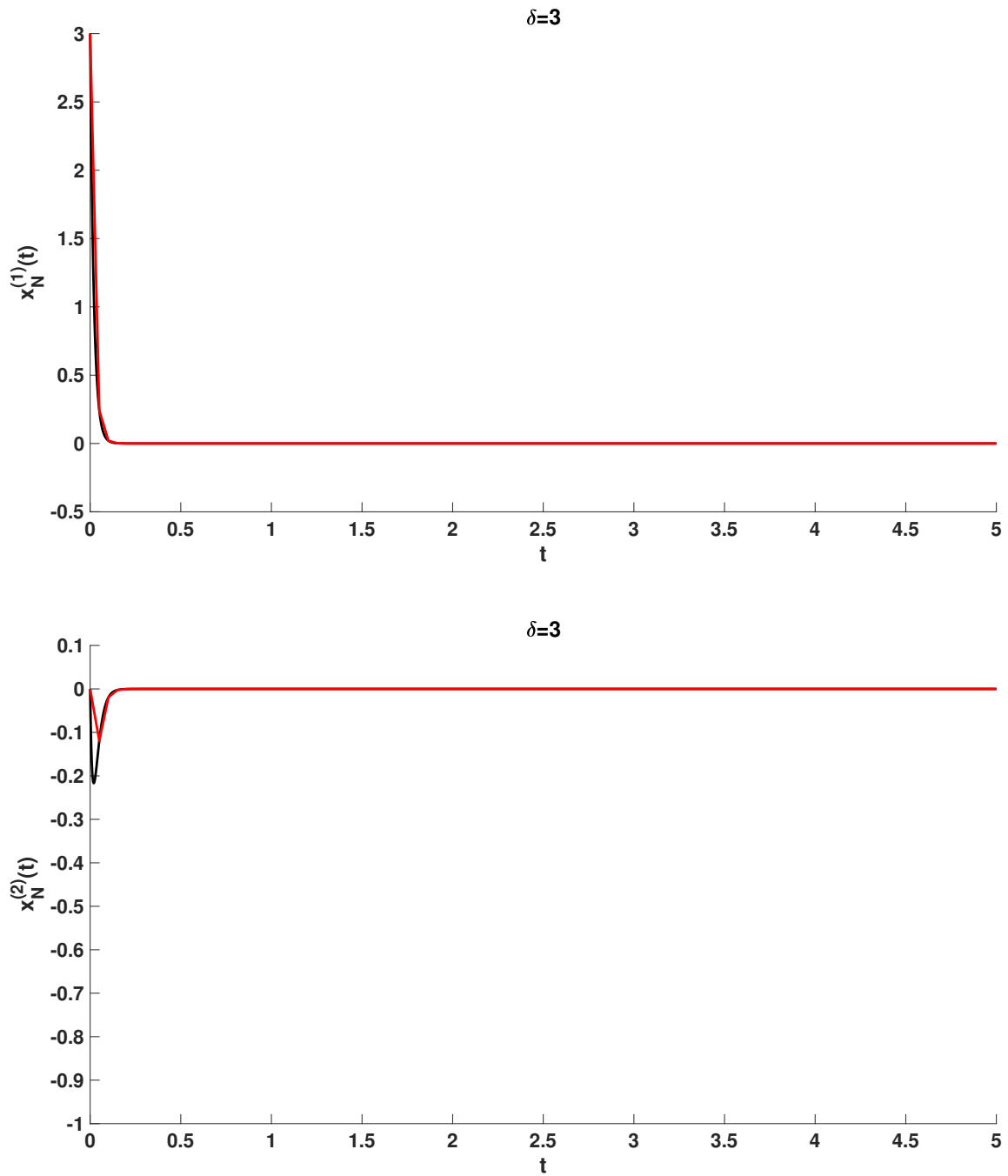


Abbildung 6.16: $\delta = 3$ und $\beta = 50$: Näherungslösungen $x_N^{(1)}(t), x_N^{(2)}(t)$ der Zustände für $N = 24$ (—) im Vergleich mit den exakten Lösungen $x^{*(1)}(t), x^{*(2)}(t)$ (—)

6.4 Beispiel mit Steuerungsbeschränkungen

In diesem Abschnitt diskutieren wir ein Beispiel, bei dem die Steuerung punktweise beschränkt ist. In Anlehnung an den Ausführungen in [Pic15] motivieren wir unser Vorgehen an dem bereits vorgestellten Beispiel in (6.1), S. 157. Auf der Suche nach Lösungen

$$(x, u) \in W_2^1((0, \infty), \omega_2) \times L_2((0, \infty), \omega_2), \omega_2(t) := e^{-2t}$$

erhielten wir durch Auswertung der Optimalitätsbedingungen den optimalen Steuerprozess

$$x^*(t) = 2e^{(1-\sqrt{2})t}, \quad u^*(t) = -2(1 + \sqrt{2})e^{(1-\sqrt{2})t} \quad (6.17)$$

Nun stellen wir fest, dass

$$u^*(t) \in U := [-2(1 + \sqrt{2}), 2(1 + \sqrt{2})] \quad \text{f.ü. auf } (0, \infty)$$

gilt. Die Menge U ist beschränkt, konvex und abgeschlossen. Wie in [Pic15] präsentiert, kann das Beispiel (6.1) auch als Optimalsteuerungsaufgabe mit Steuerungsbeschränkungen formuliert werden:

Minimiere

$$J(x, u) = \frac{1}{2} \int_0^\infty \{x^2(t) + u^2(t)\} e^{(-2+\beta)t} \omega_\beta(t) dt$$

bezüglich

$$\begin{aligned} (x, u) &\in W_2^1((0, \infty), \omega_\beta) \times L_2((0, \infty), \omega_\beta), \\ \omega_\beta(t) &:= e^{-\beta t}, \quad 4 < \beta < 2(1 + \sqrt{2}), \\ \dot{x}(t) &= 2x(t) + u(t) \text{ f.ü. auf } (0, \infty), \quad x(0) = 2, \\ u(t) &\in [-6, 6] \text{ f.ü. auf } (0, \infty). \end{aligned} \quad (6.18)$$

Zur Behandlung der obigen Aufgabe wird in [Pic15] ein Maximumprinzip vom Pontryagin Typ bewiesen (vgl. Theorem 5) und verifiziert, dass die optimale Lösung aus (6.17) zusammen mit der Adjungierten

$$y(t) = -2(1 + \sqrt{2})e^{(-1-\sqrt{2})t}$$

die daraus resultierenden notwendigen Optimalitätsbedingungen erfüllen. Angeregt durch diese Strategie möchten wir uns nun einer Aufgabe widmen, bei welcher die Menge $U \subseteq \mathbb{R}$ wie im obigen Beispiel durch sogenannte Box-Restriktionen mit konstanten Schranken gegeben ist, also

$$U := \{v \in \mathbb{R} : -\bar{u} \leq v \leq \bar{u}\} \quad (6.19)$$

mit einer fest vorgegebenen Zahl $\bar{u} > 0$ aus \mathbb{R} . Ausgangspunkt soll die Aufgabenstellung in (6.18) sein. Wir modifizieren diese, indem wir den Anfangswert des Zustandes $x(0) = x_0 > 0$ als freien reellen Parameter auffassen und Box-Restriktionen der Form (6.19) einfügen. Mit diesen Überlegungen formulieren wir die folgende Steuerungsaufgabe:

Minimiere

$$J(x, u) = \frac{1}{2} \int_0^{\infty} \{x^2(t) + u^2(t)\} e^{(-2+\beta)t} \omega_{\beta}(t) dt$$

bezüglich

$$\begin{aligned} (x, u) &\in W_2^1((0, \infty), \omega_{\beta}) \times L_2((0, \infty), \omega_{\beta}), \\ \omega_{\beta}(t) &:= e^{-\beta t}, \quad 4 < \beta < 2(1 + \sqrt{2}), \\ \dot{x}(t) &= 2x(t) + u(t) \text{ f.ü. auf } (0, \infty), \quad x(0) = x_0 > 0, \\ u(t) &\in [-\bar{u}, \bar{u}] \text{ f.ü. auf } (0, \infty), \quad \bar{u} > 0. \end{aligned} \tag{6.20}$$

Im Folgenden sei $\bar{u}(\sqrt{2} - 1) < x_0 < \frac{\beta}{2}$ vorausgesetzt. Da die Existenz einer Lösung dieser Optimalsteuerungsaufgabe durch Anwendung des Theorems 3 aus [Pic15] gesichert ist, konzentrieren wir uns jetzt darauf, eine Lösung von (6.20) mit Hilfe der notwendigen Optimalitätsbedingungen zu berechnen. Dabei nutzen wir aus, dass uns das Maximumprinzip gleichzeitig auch wieder hinreichende Optimalitätsbedingungen zur Verfügung stellt, vgl. [Pic15], S. 189.

6.4.1 Diskussion der notwendigen Bedingungen

Wie wenden das Maximumprinzip aus [Pic15], vgl. Theorem 5, an. Demnach ist die Pontryaginische Funktion der Aufgabe (6.20) durch

$$H(t, \xi, v, \eta, 1) = -\frac{1}{2}(\xi^2 + v^2)e^{-2t} + \eta(2\xi + v)$$

gegeben und für den optimalen Prozess $(x^*, u^*) \in W_2^1((0, \infty), \omega_{\beta}) \times L_2((0, \infty), \omega_{\beta})$ existieren Multiplikatoren $\lambda_0 = 1$ und $y_0 \in W_2^1((0, \infty), \omega_{\beta}^{-1})$, $\omega_{\beta}^{-1}(t) = e^{\beta t}$, sodass die Maximumbedingung

$$H(t, x^*(t), u^*(t), y_0(t), \lambda_0) = \max_{v \in U} H(t, x^*(t), v, y_0(t), \lambda_0) \quad \text{f.ü. auf } (0, \infty) \quad (M)$$

und die kanonische Gleichung

$$\dot{y}_0(t) = x^*(t)e^{-2t} - 2y_0(t) \quad \text{f.ü. auf } (0, \infty) \quad (K)$$

erfüllt sind. Wir beginnen mit der Diskussion der Maximumbedingung, welche im Gegensatz zu den bisher betrachteten Beispielen und aufgrund der Box-Restriktionen das folgende restringierte Optimierungsproblem in v beinhaltet:

$$\begin{aligned}
 -H(t, x^*(t), v, y_0(t)) &\rightarrow \inf! & (6.21) \\
 \text{u.d.N. } v - \bar{u} &\leq 0 \\
 -\bar{u} - v &\leq 0
 \end{aligned}$$

Aus der Diskussion des entsprechenden Karush-Kuhn-Tucker-Systems resultiert die Hilfsfunktion

$$v^*(t, y_0(t)) = \begin{cases} -\bar{u}, & \text{falls } y_0(t)e^{2t} < -\bar{u} \\ y_0(t)e^{2t}, & \text{falls } y_0(t)e^{2t} \in [-\bar{u}, \bar{u}] \\ \bar{u}, & \text{falls } y_0(t)e^{2t} > \bar{u} \end{cases} \quad (6.22)$$

zur Berechnung der optimalen Steuerung. Es gilt nun deren Lösungsstruktur weiter zu analysieren. Dafür lassen wir zunächst die Box-Restriktionen fallen und ermitteln die sogenannte freie Lösung analog zu dem Vorgehen in dem Abschnitt 6.1.1. Wir erhalten

$$x^*(t) = x_0 e^{(1-\sqrt{2})t}, \quad u^*(t) = \frac{x_0}{1-\sqrt{2}} e^{(1-\sqrt{2})t}$$

mit der Adjungierten

$$y_0(t) = \frac{x_0}{1-\sqrt{2}} e^{(-1-\sqrt{2})t}.$$

In der Abbildung 6.17 ist der Verlauf von u^* (—) für den Fall der freien Lösung skizziert und es wird klar, wie sich die Struktur bei Beachtung der Box-Restriktionen ändert. Entsprechend liegt der folgende Ansatz nahe:

$$u^*(t) = \begin{cases} -\bar{u}, & \text{falls } t \leq \bar{t} \\ y_0(t)e^{2t}, & \text{falls } t > \bar{t} \end{cases} \quad (6.23)$$

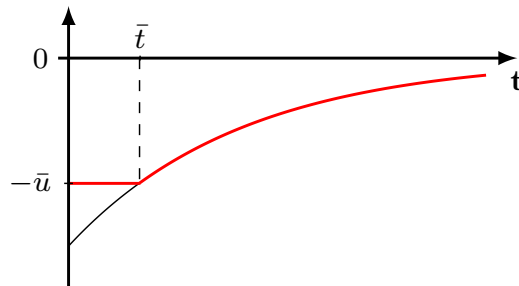


Abbildung 6.17: Ansatz für die Steuerung bei Box-Restriktionen $[-\bar{u}, \bar{u}]$, $\bar{u} > 0$ mit $\bar{u}(\sqrt{2} - 1) < x_0$

Für das weitere Vorgehen gehen wir von der Existenz eines Umschaltzeitpunktes $\bar{t} > 0$ aus und unterstellen eine Lösungsstruktur der Form (6.23). Für $t \leq \bar{t}$ ergibt sich mit $u^*|_{[0, \bar{t}]} = -\bar{u}$ das folgende kanonische Differentialgleichungssystem:

$$\begin{aligned}\dot{x}^*(t) &= 2x^*(t) - \bar{u} \\ \dot{y}_0(t) &= x^*(t)e^{-2t} - 2y_0(t).\end{aligned}$$

Unter Beachtung des Anfangswertes $x^*(0) = x_0$ lassen sich die Lösungen dieses Systems leicht berechnen und sind durch

$$\begin{aligned}x^*(t) &= \left(x_0 - \frac{\bar{u}}{2}\right)e^{2t} + \frac{\bar{u}}{2}, \quad t \leq \bar{t} \\ y_0(t) &= De^{-2t} + \left(\frac{x_0}{2} - \frac{\bar{u}}{4}\right) + \frac{\bar{u}}{2}te^{-2t}, \quad t \leq \bar{t}\end{aligned}$$

mit einer Konstanten $D \in \mathbb{R}$ gegeben. Für $t > \bar{t}$ lassen sich mit $u^*|_{(\bar{t}, \infty)} = y_0(t)e^{2t}$ die Lösungen in allgemeiner Form wie bei der freien Lösung bestimmen, wobei $c \in \mathbb{R}$ ein noch freier Parameter ist:

$$x^*(t) = c(1 - \sqrt{2})e^{(1-\sqrt{2})t}, \quad u^*(t) = ce^{(1-\sqrt{2})t}, \quad y_0(t) = ce^{(-1-\sqrt{2})t}.$$

Um die Unbekannten $c, D \in \mathbb{R}$ zu bestimmen, nutzen wir die analytische Güte des Zustandes und der Adjungierten aus. Da x^* dem gewichteten Sobolev-Raum $W_2^1((0, \infty), \omega_\beta)$ angehört, können wir auf die Existenz eines stetigen Repräsentanten schließen und fordern

$$\lim_{t \rightarrow \bar{t}} x^*(t) \stackrel{!}{=} c(1 - \sqrt{2})e^{(1-\sqrt{2})\bar{t}} \Rightarrow \left(x_0 - \frac{\bar{u}}{2}\right)e^{2\bar{t}} + \frac{\bar{u}}{2} \stackrel{!}{=} c(1 - \sqrt{2})e^{(1-\sqrt{2})\bar{t}}.$$

Daraus ergibt sich

$$c = \frac{(2x_0 - \bar{u})e^{2\bar{t}} + \bar{u}}{2(1 - \sqrt{2})} e^{(\sqrt{2}-1)\bar{t}}. \quad (6.24)$$

Analog lässt sich für die Adjungierte $y_0 \in W_2^1((0, \infty), \omega_\beta^{-1})$ zusammen mit (6.24) aus

$$\lim_{t \rightarrow \bar{t}} y_0(t) \stackrel{!}{=} ce^{(-1-\sqrt{2})\bar{t}} \Rightarrow De^{-2\bar{t}} + \left(\frac{x_0}{2} - \frac{\bar{u}}{4}\right) + \frac{\bar{u}}{2}\bar{t}e^{-2\bar{t}} \stackrel{!}{=} ce^{(-1-\sqrt{2})\bar{t}}$$

der bisher noch freie Parameter $D \in \mathbb{R}$ durch

$$D = \frac{(2x_0 - \bar{u})e^{2\bar{t}} + \bar{u}}{2(1 - \sqrt{2})} + \left(\frac{\bar{u}}{4} - \frac{x_0}{2}\right)e^{2\bar{t}} - \frac{\bar{u}\bar{t}}{2} \quad (6.25)$$

festlegen. Aus der Regularität der Eingangsdaten der Aufgabe und der Maximumbedingung folgt die Stetigkeit der Steuerung u^* , vgl. [ZP18], 7.2. Daher kann der Umschaltzeitpunkt unter Beachtung von (6.24) aus

$$\lim_{t \rightarrow \bar{t}} u^*(t) \stackrel{!}{=} -\bar{u} \Rightarrow c(1 - \sqrt{2})e^{(1-\sqrt{2})\bar{t}} \stackrel{!}{=} -\bar{u}$$

gemäß

$$\bar{t} = \frac{1}{2} \ln \left| \frac{-3\bar{u} + 2\sqrt{2}\bar{u}}{2x_0 - \bar{u}} \right|$$

berechnet werden. Die sinnvolle Forderung $\bar{t} > 0$ ist für $\bar{u}(\sqrt{2} - 1) < x_0 < \frac{\bar{u}}{2}$ gesichert. Insgesamt erhalten wir

$$x^*(t) = \begin{cases} (x_0 - \frac{\bar{u}}{2})e^{2t} + \frac{\bar{u}}{2}, & \text{falls } t \leq \bar{t} \\ c(1 - \sqrt{2})e^{(1-\sqrt{2})t}, & \text{falls } t > \bar{t} \end{cases}, \quad u^*(t) = \begin{cases} -\bar{u}, & \text{falls } t \leq \bar{t} \\ ce^{(1-\sqrt{2})t}, & \text{falls } t > \bar{t} \end{cases}, \quad (6.26)$$

$$y_0(t) = \begin{cases} De^{-2t} + (\frac{x_0}{2} - \frac{\bar{u}}{4}) + \frac{\bar{u}}{2}te^{-2t}, & \text{falls } t \leq \bar{t} \\ ce^{(-1-\sqrt{2})t}, & \text{falls } t > \bar{t} \end{cases}$$

mit denen in (6.24) und (6.25) angegebenen Konstanten.

Wir möchten im Folgenden die Aufgabe (6.20) für den Anfangswert $x_0 = 1$ und mit $\bar{u} = \frac{11}{5}$ numerisch lösen. Für diese Setzungen ergeben sich für $c, D \in \mathbb{R}$ die Werte

$$c \approx -2.509292496, \quad D \approx -2.454966375$$

Der zugehörige Umschaltzeitpunkt und Zielfunktionswert sind durch

$$\bar{t} \approx 0.3175740526, \quad J(x^*, u^*) \approx 1.209503362707199$$

gegeben. Die Grafik 6.18 illustriert das Erfülltsein der Schaltbedingung (6.22): Es ist klar zu erkennen, dass $y_0(t)e^{2t} < -\bar{u}$ für $t \in [0, \bar{t}]$ gilt und $y_0(t)e^{2t} \in (-\bar{u}, \bar{u})$ für $t > \bar{t}$ richtig ist. Demnach ist die Maximumbedingung für die gesetzten Parameter erfüllt und durch (x^*, u^*) eine optimale Lösung für die Aufgabe (6.20) gegeben.

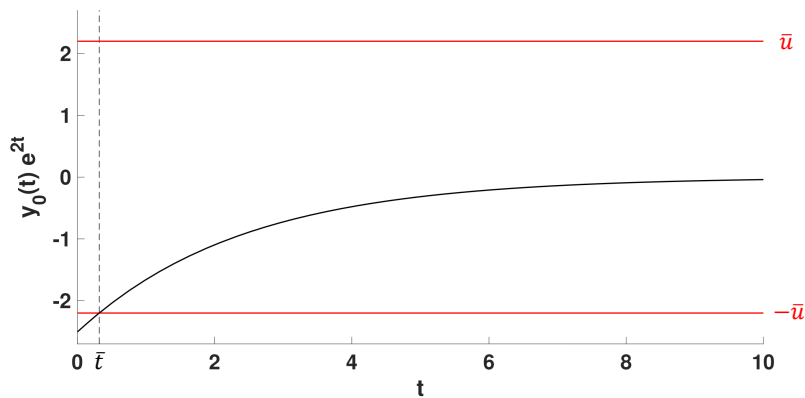


Abbildung 6.18: Darstellung der Schaltbedingung für $x_0 = 1$ und $\bar{u} = \frac{11}{5}$

6.4.2 Anwenden der Spektralmethode

Durch die Transformation $y(t) := y_0(t)e^{\beta t}$ überführen wir das kanonische Differentialgleichungssystem in das System

$$\begin{aligned}\dot{x}^*(t) &= 2x^*(t) + u^*(t) \\ \dot{y}(t) &= x^*(t)e^{(\beta-2)t} + (\beta-2)y(t)\end{aligned}\tag{6.27}$$

mit $(x^*, y) \in W_2^1((0, \infty), \omega_\beta) \times W_2^1((0, \infty), \omega_\beta)$. Die in der Maximumbedingung (M) formulierte quadratische Optimierungsaufgabe in \mathbb{R} , vgl. (6.21), haben wir durch Auswerten der Karush-Kuhn-Tucker-Bedingungen analysiert. Gemäß (6.22) können wir die optimale Steuerung u^* aus der Projektionsformel und der Hilfsfunktion

$$v^*(t, y(t)) = \mathbb{P}_{[-\bar{u}, \bar{u}]} \left\{ y(t)e^{(2-\beta)t} \right\} \quad \text{f.ü. auf } (0, \infty)$$

bestimmen. Dabei meinen wir für $a, b \in \mathbb{R}$, $a \leq b$ mit $\mathbb{P}_{[a,b]}$ die Projektion von \mathbb{R} auf $[a, b]$, also

$$\mathbb{P}_{[a,b]}(v) := \min \{b, \max \{a, v\}\}.$$

Wir setzen die Projektionsformel in das System (6.27) ein und suchen nach Näherungslösungen $x_N, y_N \in \mathcal{P}_N((0, \infty))$, welche das folgende nichtlineare Gleichungssystem lösen:

$$\begin{cases} \dot{x}_N(\tau_i^{(N)}) = 2 \cdot x_N(\tau_i^{(N)}) + \min \left\{ \bar{u}, \max \left\{ -\bar{u}, y_N(\tau_i^{(N)}) \cdot e^{(2-\beta) \cdot \tau_i^{(N)}} \right\} \right\}, & 1 \leq i \leq N \\ \dot{y}_N(\tau_i^{(N)}) = x_N(\tau_i^{(N)}) \cdot e^{(\beta-2) \cdot \tau_i^{(N)}} + (\beta-2) \cdot y_N(\tau_i^{(N)}), & 1 \leq i \leq N \\ x_N(\tau_0) = 1 \\ y_N(\tau_0) = \sum_{i=1}^N \left\{ 2 \cdot y_N(\tau_i^{(N)}) - x_N(\tau_i^{(N)}) \cdot e^{(\beta-2) \cdot \tau_i^{(N)}} \right\} \mathbf{w}_i^{(N)} \end{cases}$$

Wie bisher sind $\tau_i^{(N)}$, $1 \leq i \leq N$ die Nullstellen des verallgemeinerten Laguerre-Polynoms $\mathcal{L}_N^{(\beta)}$, $\tau_0 = 0$ und $\mathbf{w}_i^{(N)}$, $i = 1, \dots, N$ sind die Gewichte der Gauß-Quadratur. Die Approximation der Steuerung berechnen wir gemäß

$$u_N(t) = \mathbb{P}_{[-\bar{u}, \bar{u}]} \left\{ y_N(t) \cdot e^{(2-\beta)t} \right\}.$$

Im Vergleich zu den bisherigen Beispielen liefert das hier betrachtete Beispiel neben der Behandlung der Box-Restriktionen zwei weitere interessante Aspekte:

- Der Parameter β aus der Gewichtsfunktion $\omega_\beta(t) := e^{-\beta t}$ nimmt keinen Einfluss auf die Lösungsstruktur (6.26), taucht jedoch explizit in dem aus den Kollokationsgleichungen resultierenden Gleichungssystem auf. Wir werden untersuchen, inwiefern β die Näherungslösungen beeinflusst.

- Die bisher gelösten Differentialgleichungssysteme waren autonom, wohingegen das System (6.27) explizit von t abhängt. Ein wesentliches Merkmal von Kollokationsverfahren besteht in der einfachen Behandlung zeitabhängiger Systeme und dieses Beispiel verdeutlicht, dass dies keine Schwierigkeiten bei der Diskretisierung nachsichzieht.

6.4.3 Auswertung der numerischen Ergebnisse für $x_0 = 1$ und $\bar{u} = 11/5$

Der Parameter β ist gemäß der Aufgabenstellung (6.20) aus dem Intervall $(4, 2(1 + \sqrt{2}))$ zu wählen. Diese Festlegung resultiert aus der Raumwahl und den Voraussetzungen für die Anwendung des Maximumprinzips, vgl. [Pic15]. Bei unseren numerischen Experimenten haben wir Berechnungen für unterschiedliche $\beta \in (4, 2(1 + \sqrt{2}))$ durchgeführt und die Resultate für $\beta = 4.01, 4.41, 4.81$ in der Tabelle 6.13 zusammengefasst.

β	N	$J(x_N, u_N)$	$ J(x_N, u_N) - J(x^*, u^*) $	\bar{t}_N	$ \bar{t}_N - \bar{t} $
4.01	5	1.226939998	0.017436636	0.5488877164	0.23131366
4.01	6	1.209993981	$4.906183251 \cdot 10^{-4}$	0.3619691941	$0.44395141 \cdot 10^{-1}$
4.01	7	1.209446672	$5.669050434 \cdot 10^{-5}$	0.3433516607	$0.25777608 \cdot 10^{-1}$
4.01	8	1.209434562	$6.880068115 \cdot 10^{-5}$	0.3289864688	$0.11412416 \cdot 10^{-1}$
4.01	9	1.209497236	$6.126123765 \cdot 10^{-6}$	0.3281051120	$0.10531059 \cdot 10^{-1}$
4.41	5	1.208820462	$6.829001291 \cdot 10^{-4}$	0.3192497718	$0.16757192 \cdot 10^{-2}$
4.41	6	1.209699667	$1.963044150 \cdot 10^{-4}$	0.2522106260	$0.65363426 \cdot 10^{-1}$
4.41	7	1.209481143	$2.221924477 \cdot 10^{-5}$	0.3493990931	$0.31825040 \cdot 10^{-1}$
4.41	8	1.209484790	$1.857184973 \cdot 10^{-5}$	0.3230823961	$0.55083435 \cdot 10^{-2}$
4.41	9	1.209561365	$5.800277681 \cdot 10^{-5}$	0.3259535464	$0.83794938 \cdot 10^{-2}$
4.81	5	1.209081745	$4.216168472 \cdot 10^{-4}$	0.4393503546	0.12177630
4.81	6	1.221308140	0.011804777	0.4320477152	0.11447366
4.81	7	1.210608143	0.001104780	0.2131190124	0.1044550402
4.81	8	1.209658068	$1.547055611 \cdot 10^{-4}$	0.3157562692	$0.18177834 \cdot 10^{-2}$
4.81	9	1.209598862	$9.550022047 \cdot 10^{-5}$	0.3198452866	$0.22712340 \cdot 10^{-2}$

Tabelle 6.13: Approximierter Zielfunktionswert $J(x_N, u_N)$ und Umschaltzeitpunkt \bar{t}_N für $\beta = 4.01, 4.41, 4.81$

Zunächst möchten wir erwähnen, dass alle Näherungslösungen u_N die angenommene Schaltstruktur (vgl. (6.26)) mit einem Schaltzeitpunkt widerspiegelten. Da die Frage der Eindeutigkeit der optimalen Lösung für die betrachtete Aufgabe bisher offen bleibt, war dies nicht mit Sicherheit zu erwarten. Darüber hinaus resultiert wie bei den bisherigen Beispielen auch bei dieser Aufgabe die Erhöhung des Polynomgrades N in einer genaueren Berechnung von $J(x_N, u_N)$. Für $\beta = 4.01$ liefern uns die Näherungslösungen bei der Verwendung von $N = 9$ Kollokationspunkten den Zielfunktionswert $J(x_N, u_N) \approx 1.209497236$. Damit weicht dieser um $\approx 6.126 \cdot 10^{-6}$ von $J(x^*, u^*)$ ab. Der numerisch ermittelte Schaltzeitpunkt \bar{t}_N liegt bei 0.3281051120 . Ein Erhöhung von β auf 4.41 bzw. 4.81 bei gleichbleibenden $N = 9$ führt dazu, dass der Fehler $|J(x_N, u_N) - J(x^*, u^*)|$ auf $\approx 5.8 \cdot 10^{-5}$ bzw. $9.55 \cdot 10^{-5}$ ansteigt. Die Berechnung des Umschaltzeitpunktes hingegen verbessert sich. Für $\beta = 4.81$ und $N = 9$ ist $|\bar{t}_N - \bar{t}| \approx 0.227 \cdot 10^{-2}$. Das Ziel der Optimierung ist eine bestmögliche Annäherung an

$J(x^*, u^*)$. Gemessen daran lieferte die Wahl $\beta = 4.01$ und $N = 9$ die besten Approximationsergebnisse für dieses Beispiel.

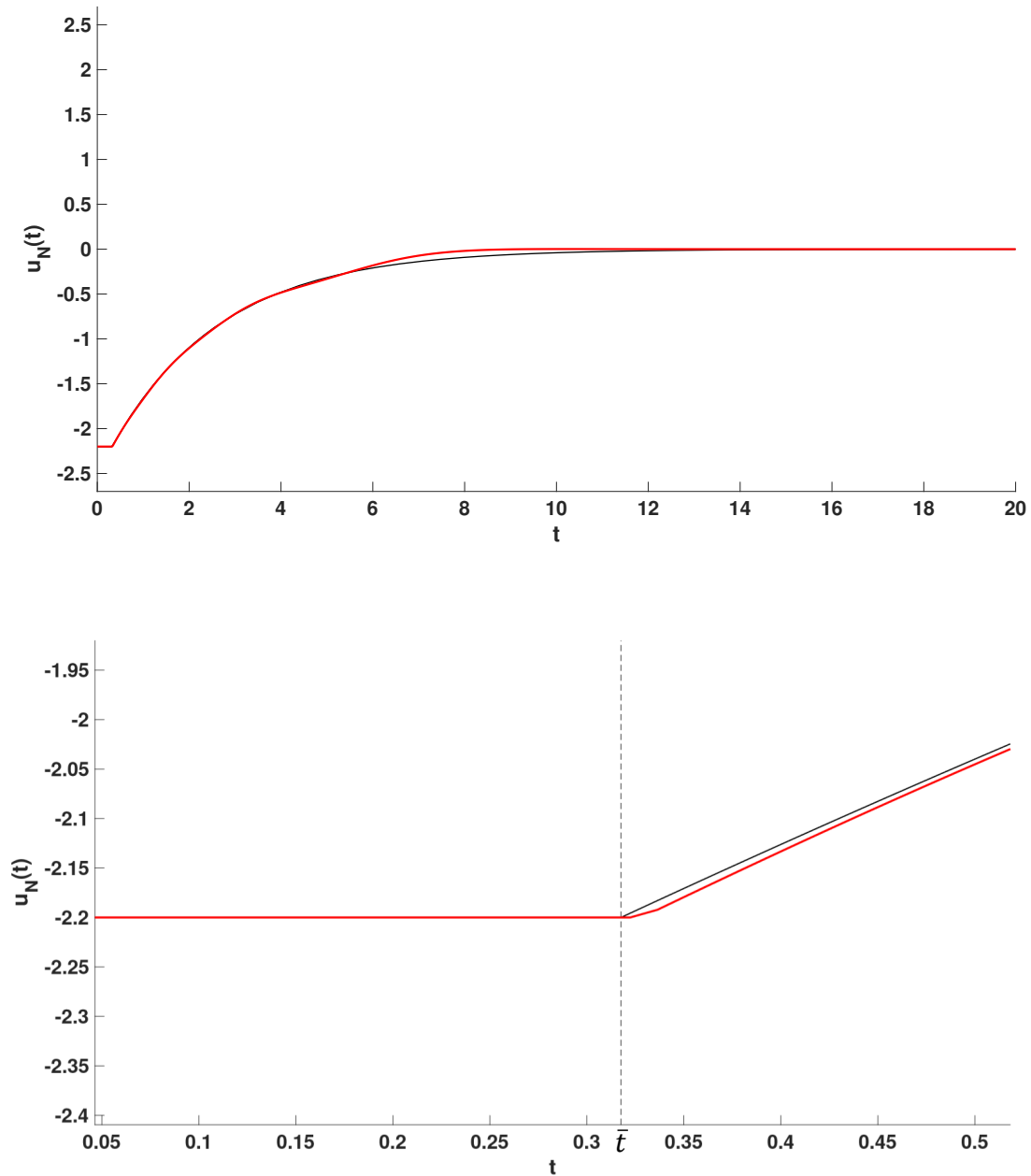


Abbildung 6.19: $\beta = 4.01$: Näherungslösung $u_N(t)$ der Steuerung für $N = 9$ (—) im Vergleich mit der optimalen Steuerung $u^*(t)$ (—) auf $[0, 20]$ und in der Nähe des Umschaltzeitpunktes \bar{t}

Die zugehörige numerisch ermittelte Steuerung u_N ist in der Abbildung 6.19 dargestellt, wobei die zweite Grafik verdeutlicht, wie gut der Umschaltzeitpunkt \bar{t} angenähert wird. Der letzte Kollokationspunkt ist dabei durch $\tau_9^{(9)} \approx 6.57$ gegeben. Den analytisch ermittelten Anfangswert der Adjungierten

$$y(0) \approx -2.504966372928010$$

können wir wieder gut durch die numerische Lösung annähern und erhalten $y_9(0) \approx -2.514$.

Auf Basis der Ergebnisse für $\beta = 4.01, 4.41, 4.81$ lässt sich festhalten, dass umso näher der Parameter β an der linken Intervallgrenze 4 gewählt wurde, desto bessere numerische Ergebnisse konnten erzielt werden. Wir haben die numerische Lösungsmethode auch für die Werte $\beta = 2$ und $\beta = 5$ getestet. Da beide Werte außerhalb des vorgegebenen Intervalls $(4, 2(1 + \sqrt{2}))$ liegen, für welches die Gültigkeit des Maximumprinzips gesichert ist, erscheint uns die Auswertung der entsprechenden Ergebnisse als besonders spannend.

Betrachten wir zunächst $\beta = 2$. Auch hier erhielten wir für $N = 9$ die besten Resultate und die Näherungslösungen liefern den Zielfunktionswert $J(x_N, u_N) \approx 1.209812086$, welcher um $\approx 3.087 \cdot 10^{-4}$ von $J(x^*, u^*)$ abweicht, vgl. Tabelle 6.14.

β	N	$J(x_N, u_N)$	$ J(x_N, u_N) - J(x^*, u^*) $
2	5	1.908641819	0.699138457
2	6	1.259643779	0.050140416
2	7	1.218262324	0.008758961
2	8	1.210148850	$6.454879291 \cdot 10^{-4}$
2	9	1.209812086	$3.087234472 \cdot 10^{-4}$
5	5	1.207149428	0.002353933
5	6	1.209289543	$2.138192155 \cdot 10^{-4}$
5	7	1.209496775	$6.587034886 \cdot 10^{-6}$
5	8	1.209493948	$9.413865232 \cdot 10^{-6}$
5	9	1.209616124	$1.127621937 \cdot 10^{-4}$

Tabelle 6.14: Approximierter Zielfunktionswert $J(x_N, u_N)$ für $\beta = 2, 5$

Interessant ist ein Blick auf die grafische Darstellung der zugehörigen Näherungslösung u_N , vgl. Abbildung 6.20. Im Gegensatz zu den bisherigen Werten für β liefert $\beta = 2$ eine komplett neue Schaltstruktur mit insgesamt sechs Umschaltzeiten:

$$\begin{aligned} \bar{t}_N^1 &\approx 0.3159033702, & \bar{t}_N^2 &\approx 7.748970352, & \bar{t}_N^3 &\approx 8.164094356 \\ \bar{t}_N^4 &\approx 12.34378999, & \bar{t}_N^5 &\approx 8.944676125, & \bar{t}_N^6 &\approx 12.32571063. \end{aligned}$$

Da die Frage nach der Eindeutigkeit der Lösung für das vorliegende Beispiel offen ist, kann dies ein Hinweis auf eine weitere optimale Steuerung sein. Ist $\beta = 5$, so erhielten wir die besten Resultate für $N = 7$ und erhalten wieder die ursprünglich angenommene Struktur mit einem Umschaltzeitpunkt, vgl. Abbildung 6.20. Die Ergebnisse liefern sogar eine vergleichbar gute Annäherung an den Zielfunktionswert wie für $\beta = 4.01$, nämlich

$J(x_N, u_N) \approx 1.209496775$. Dies entspricht einer Abweichung zu dem analytisch bestimmten Wert von $\approx 6.587 \cdot 10^{-6}$.

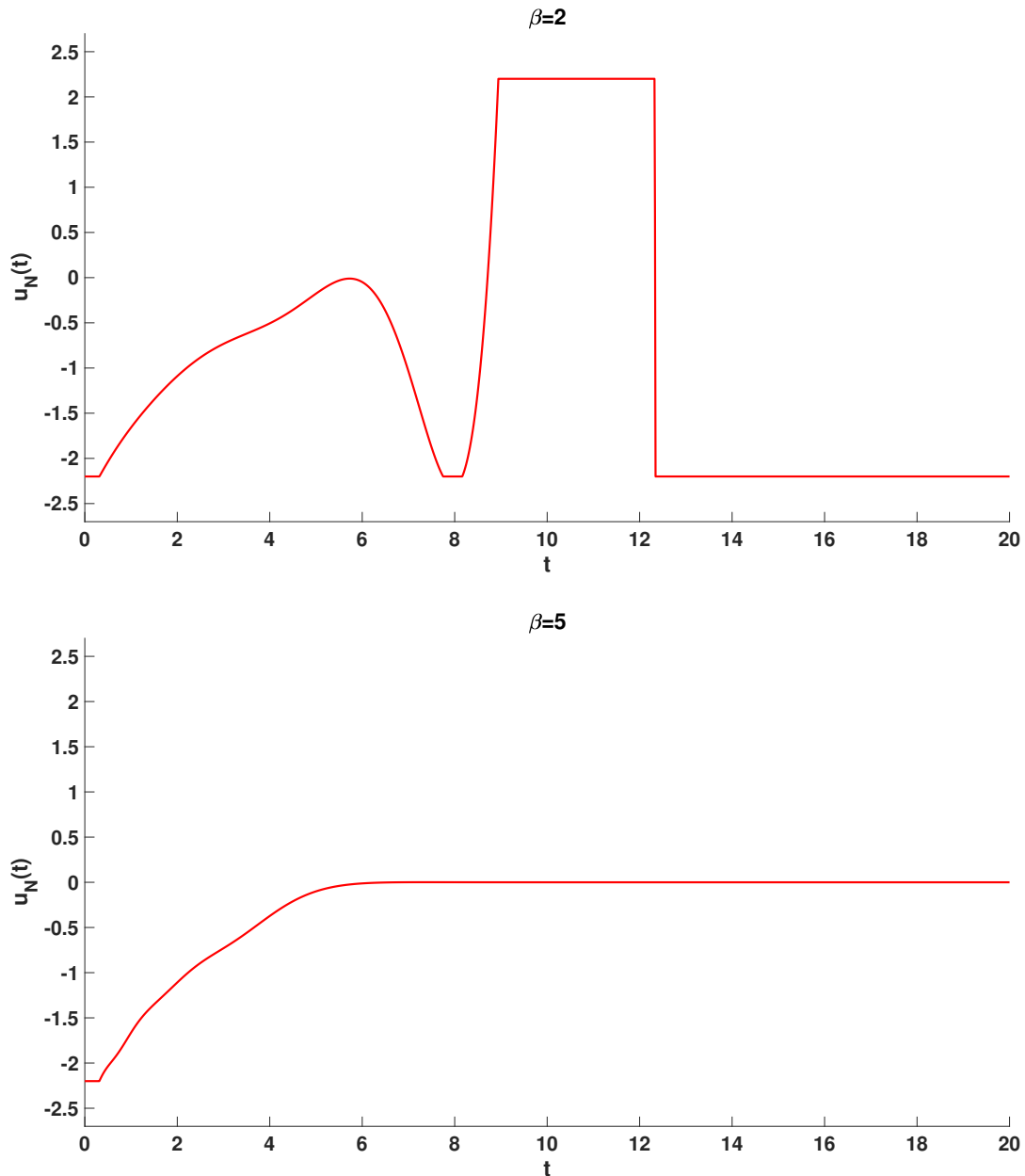


Abbildung 6.20: Näherungslösung $u_N(t)$ der Steuerung für $\beta = 2$ und $N = 9$ und für $\beta = 5$ und $N = 7$

Abschließend können wir festhalten, dass die besten Approximationen für den Zielfunktionswert erzielt wurden, wenn β innerhalb des Intervalls $(4, 2(1 + \sqrt{2}))$ gewählt wurde. Für

diesen Parameterbereich kann sowohl die Existenz einer Lösung als auch die Gültigkeit des Maximumprinzips garantiert werden. Die vorgestellte Spektralmethode liefert gute Näherungslösungen und verdeutlicht an diesem Beispiel neben der einfachen Umsetzung von Steuerungsbeschränkungen in Form von konstanten Box-Restriktionen auch die unkomplizierte Behandlung von explizit von der Zeit t abhängigen Systemen.

7 Abschließende Bemerkungen

In der vorliegenden Arbeit wurden linear-quadratische Optimalsteuerungsaufgaben mit dem a priori gegebenen unendlichen Zeithorizont $[0, \infty)$ untersucht. Diese Aufgabenklasse ist für die Approximation nichtlinearer Aufgaben von großer Bedeutung, tritt bei Verfahren der sequentiellen quadratischen Optimierung auf und entsteht im Kontext der asymptotischen Steuerbarkeit dynamischer Systeme. In der gesamten Arbeit wurde ein Zugang mittels speziell gewichteter Funktionenräume gewählt und der Lebesguesche Integralbegriff zugrunde gelegt. Konkret wurden Zustände und Steuerungen betrachtet, welche Elemente eines gewichteten Sobolev- bzw. eines gewichteten Lebesgue-Raumes sind. Durch diese Raumwahl konnten die linear-quadratischen Optimalsteuerungsaufgaben über $[0, \infty)$ in eine gewichtete Hilbertraumtheorie eingebettet und mit Methoden der konvexen Analysis behandelt werden. Neben der Existenz und der Eindeutigkeit der Lösung wurde ein Maximumprinzip vom Pontryagin-Typ bewiesen. Dieses liefert die kanonischen Gleichungen, die Maximumbedingung aber auch die Existenz einer adjungierten Funktion, welche einer natürlichen Transversalitätsbedingung genügt. Aufbauend auf den durch das Maximumprinzip gegebenen notwendigen Optimalitätsbedingungen wurde ein indirektes Spektralverfahren zur Lösung linear-quadratischer Optimalsteuerungsaufgaben über $[0, \infty)$ vorgestellt. Bei der Konstruktion des Verfahrens spielte die erhaltene Transversalitätsbedingung eine zentrale Rolle: Mit Hilfe dieser ist es gelungen, eine Bestimmungsgleichung für den Anfangswert der adjungierten Funktion in Integralform herzuleiten. Durch eine Kollokation des kanonischen Differentialgleichungssystems und unter Verwendung der integrierten Transversalitätsbedingung ließ sich das Auffinden einer Näherungslösung für die Optimalsteuerungsaufgabe auf das Lösen eines linearen Gleichungssystems reduzieren. Die genutzten Approximationsansätze basieren auf dem über $(0, \infty)$ vollständigen Funktionensystem der verallgemeinerten Laguerre-Polynome.

Die Anwendung des indirekten Spektralverfahrens wurde anhand verschiedener Beispiele im Detail erläutert. Neben der ausführlichen Herleitung der Kollokationsgleichungen und der integrierten Transversalitätsbedingung beinhaltet diese Arbeit im Anhang auch den Quellcode für eine Implementierung in Matlab. Bei den numerischen Experimenten wurden die Approximationsergebnisse in Abhängigkeit des gewählten Polynomgrades N bzw. der Anzahl der verwendeten Kollokationspunkte ausgewertet und unter verschiedenen Gesichtspunkten betrachtet. War der Parameter $\beta > 0$ aus der Gewichtsfunktion $\omega_\beta(t) := e^{-\beta t}$ in der Aufgabenstellung nicht fest vorgegeben, so wurde zusätzlich dessen Einfluss auf die Lösungen untersucht. Bei allen Beispielen lag ein besonderes Augenmerk auf dem numerisch ermittelten Anfangswert der Adjungierten, welcher allein durch die integrierte Transversalitätsbedingung bestimmt wurde. Die numerischen Resultate bestätigten die Wahl dieser Bedingung und zeigten, dass der Anfangswert der adjungierten Funktion damit numerisch sehr genau berechnet werden konnte.

Neben Beispielen der optimalen Steuerung wurde auch die asymptotische Stabilisierbarkeit eines linearisierten dynamischen Systems untersucht. Nachdem der theoretische Zusammenhang zwischen dem Gebiet der Optimalsteuerungsaufgaben über $[0, \infty)$ und der asymptotischen Steuerbarkeit dynamischer Systeme in dem Kontext gewichteter Sobolev-Räume in dieser Arbeit herausgearbeitet wurde, wurde dieser anhand des mathematischen Pendels aufgezeigt. Mit dem Ziel eine das System asymptotisch stabilisierende Steuerung zu finden, wurde das indirekte Spektralverfahren genutzt und damit eine weitere Anwendungsmöglichkeit dieser Methode demonstriert. Auch wurde an diesem Beispiel deutlich, dass das Auftreten von mehreren Zuständen und Steuerungen keine zusätzlichen Schwierigkeiten bereitet und sich diese in dem numerischen Verfahren leicht umsetzen lassen. Zusätzlich wurde eine Aufgabe mit Box-Restriktionen vorgestellt, um veranschaulichen zu können, wie Steuerungsbeschränkungen im Rahmen des vorgestellten Spektralverfahrens behandelt werden können.

Alle in dieser Arbeit präsentierten numerischen Experimente führten zu zufriedenstellenden Approximationsergebnissen. Dennoch ergeben sich natürlich weitere Anknüpfungspunkte für die zukünftige Arbeit.

Aktuell arbeiten wir an der Fertigstellung eines Konvergenzbeweises. Mit den in dieser Arbeit gelieferten Resultaten zur punktweisen Konvergenz der verallgemeinerten Fourier-Laguerre-Reihen haben wir den Ausgangspunkt geschaffen. Die Grundidee des Beweises basiert auf diesen Ergebnissen und beinhaltet zusätzlich Methoden aus der Dualitätstheorie, wodurch der Übergang zu einer Variationsaufgabe möglich ist. Aufbauend darauf lassen sich die globalen Approximationsansätze für eine auf der dualen Seite direkten Diskretisierung verwenden. Da die verallgemeinerten Fourier-Reihen der gesuchten Lösungen zulässig für die duale Aufgabe sind, ist ein direktes Abschätzen zu den numerisch ermittelten Lösungen möglich und ein Konvergenzbeweis auf diesem Weg scheint vielversprechend.

Zusätzlich gilt es, das durch die indirekte Diskretisierung entstandene Gleichungssystem, dessen Lösungseigenschaften und die zugehörige Koeffizientenmatrix genauer zu untersuchen. Insbesondere erzeugten die in dieser Arbeit vorgestellten Beispiele eine singuläre, schlecht konditionierte Koeffizientenmatrix. Dabei konnte numerisch nachgewiesen werden, dass der Rang der Koeffizientenmatrix M^N mit dem Rang der erweiterten Koeffizientenmatrix (M^N, z^N) übereinstimmt. Entsprechend war die Lösbarkeit des Gleichungssystems gesichert. Dennoch konnte festgestellt werden, dass eine beliebige Erhöhung von N nicht unmittelbar zu einer Verbesserung der Approximationsergebnisse führte. Wir führen dies sowohl auf Rundungsfehler aber auch auf die schlechte Konditionszahl zurück. Diese Aspekte erfordern in Zukunft noch genauere Untersuchungen.

Abschließend lässt sich sagen, dass die vorliegende Arbeit die notwendige Theorie und ein darauf aufbauendes indirektes Spektralverfahren zur Lösung linear-quadratischer Optimalsteuerungsaufgaben über dem unendlichen Zeithorizont $[0, \infty)$ liefert. Die vorgestellte Methode stellt aus Sicht der Autorin eine anwendbare Grundlage dar, um die Weiterentwicklung auf allgemeinere Aufgabenklassen voranzutreiben. Dies schließt die Anpassung des Lösungsverfahrens für nichtlineare Optimalsteuerungsaufgaben ein, beinhaltet aber auch die Untersuchung, wie sich zusätzliche Nebenbedingungen, wie zum Beispiel Zustandsbeschränkungen, auf die Approximationsergebnisse auswirken.

A Quellcode für das Beispiel aus Abschnitt 6.1

Bei Aufruf der Funktion $LQR(\text{Polynomgrad}, \text{beta})$ übergibt der Anwender den gewünschten Polynomgrad N bzw. die zu verwendende Anzahl an Kollokationspunkten und den Parameter β aus der Gewichtsfunktion. Nachdem die Momente der Gewichtsfunktion $\omega_\beta(t) = e^{-\beta t}$, $\beta > 0$ berechnet wurden, werden diese gemeinsam mit den Eingabeparametern N und β an *Kollokation.m* übergeben. In *Kollokation.m* werden zunächst die Quadraturgewichte (Zeile 32-35) und Kollokationspunkte (Zeile 30) bestimmt. Für die Berechnung der verallgemeinerten Laguerre-Polynome (Zeile 18-21) wird die Funktion *LaguerreGenBeta(varargin)* verwendet. Diese basiert im Wesentlichen auf der Funktion *LaguerreGen(varargin)* von Matthias Trampisch, welche die verallgemeinerten Laguerre-Polynome $\mathcal{L}_N^{(\alpha,0)}$ für $\alpha > -1$ (und $\beta = 0$) berechnet. Der zugehörige M-File ist unter

<https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/26218/versions/5/previews/LaguerreGen.m/index.html>

zu finden. Die Funktion wurde insofern erweitert, dass der Parameter $\beta > 0$ ergänzt wurde und nun die verallgemeinerten Laguerre-Polynome $\mathcal{L}_N^{(\alpha,\beta)}$ ermittelt werden können. In den Zeilen 72-140 wird die Koeffizientenmatrix des aus den Kollokationsgleichungen resultierenden Gleichungssystem aufgestellt. Das lineare System wird anschließend mit dem Backslash-Operator gelöst (Zeile 169).

A.1 LQR.m

```
1 function [weights, tau, kn, r] = LQR(Polynomgrad, beta)
2 format long;
3 n = 2 * Polynomgrad + 1;
4 global N;
5 N = Polynomgrad;
6 momente = factorial(0:n-1) ./ beta .^(1:n);
7 [weights, tau, kn, r] = Kollokation(momente, beta, N)
8 end
```

A.2 Kollokation.m

```
1 function [weights, tau, kn, r] = Kollokation(momente, beta, N)
2
3 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
4 % Berechnung der Quadraturstellen und -gewichte
5 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```

7  global T;
8  T=zeros(N,N);

10 Hauptdiagonale = 1/beta.*[1:2:2*N-1];
11 Nebendiagonale = 1/beta.*[1:1:N-1];

13 T = diag(Hauptdiagonale);
14 T = T + diag(Nebendiagonale,1);
15 T = T + diag(Nebendiagonale,-1);

18 global LagPoly;
19 for k = 1:N+1
20   LagPoly{k} = LaguerreGenBeta(k-1,0,beta); % Orthogonale L-Polynome
21 end

23 for j= 1:N+1
24   LagPoly{j}=flipud(LagPoly{j});
25   AblLagPoly{j}=polyder(LagPoly{j}); %Ableitung der orthog. L-Polynome
26 end

28 [V,D]= eig(T);      % Eigenwerte bestimmen
29 global tau;          % Nullstellen von  $L_N$ 
30 tau = diag(D);      % Quad.-stellen sind Eigenw. von T

32 weights = zeros(N,1); % Quadraturgewichte sind skalierte
33 for k = 1:N          % Skalarprodukte der Eigenvektoren
34   weights(k) = momente(1) * V(1,k)^2 / (V(:,k)' * V(:,k));
35 end

37 % Koeffizient im Fehlerterm
38 kn = integ(mult(LagPoly{N+1},LagPoly{N+1}),momente);

40 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
41 % einige Unterfunktionen zur Behandlung der auftretenden Polynome %
42 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

44 % Polynome miteinander multiplizieren
45 function p = mult(s,t)
46 p = zeros(length(s)+length(t)-1,1);
47 for j = 1:length(s);
48 p(j:length(t)+j-1) = p(j:length(t)+j-1) + s(j) * t(:);
49 end
50 end

52 % Polynome addieren
53 function r = add(p,q)
54 if length(q) >= length(p)
55   r = q(:);
56   r(1:size(p,1)) = r(1:size(p,1)) + p(:);
57 else
58   r = p(:);
59   r(1:size(q,1)) = r(1:size(q,1)) + q(:);
60 end
61 end

```

```

63 % Integral auswerten
64 function r = integ(p,momente)
65 r = max(abs((p(:))'*momente(1:length(p))'),eps);
66 end

68 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
69 % Koeffizientenmatrix M^N: i Zeilen, j Spalten %
70 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

72 m = 2*N+2; % Anzahl Zeilen /Spalten von M^N
73 global M;
74 M=zeros(m,m); % Nullmatrix wird erzeugt

76 % Diskretisierte DGL: Zustand
77 for i=1:N
78     for j=1:m

80         if j==1
81             M(i,j)=-2*polyval(LagPoly{1},tau(i));
82         end

84         if j>=2 && j<=N
85             M(i,j)=polyval(AblLagPoly{j},tau(i))-2*polyval(LagPoly{j},tau(i));
86         end

88         if j==N+1
89             M(i,j)=polyval(AblLagPoly{N+1},tau(i));
90         end

92         if j>=N+2 && j<=2*N+1
93             M(i,j)=-polyval(LagPoly{j-(N+1)},tau(i));
94         end

96     end
97 end

99 % Diskretisierte DGL: Adjungierte
100 for i=N+1:2*N
101     for j=1:m

103         if j>=1 && j<=N
104             M(i,j)=-polyval(LagPoly{j},tau(i-N));
105         end

107         if j==N+2
108             M(i,j)=(2-beta)*polyval(LagPoly{1},tau(i-N));
109         end

111         if j>=N+3 && j<= 2*N+1
112             M(i,j)=polyval(AblLagPoly{j-(N+1)},tau(i-N))
113             +(2-beta)*polyval(LagPoly{j-(N+1)},tau(i-N));
114         end

116         if j==2*N+2
117             M(i,j)=polyval(AblLagPoly{N+1},tau(i-N));
118         end

```

A Quellcode für das Beispiel aus Abschnitt 6.1

```

120     end
121 end

123 % Anfangswert Zustand
124 i=2*N+1;
125     for j=1:m
126
127         if j>=1 && j<=N+1
128             M(i , j)=1;
129         end
130
131     end

133 % Anfangswert Adjungierte /(T)
134 i=2*N+2;
135     for j=1:m
136
137         if j>=1 && j<=N
138             M(i , j)=sum(weights(1:N).*polyval(LagPoly{j},tau(1:N)));
139         end
140
141         if j>=N+2 && j<=2*N+1
142             M(i , j)=1-2*sum(weights(1:N).*polyval(LagPoly{j-(N+1)},tau(1:N)));
143         end
144
145         if j==2*N+2
146             M(i , j)=1;
147         end
148
149     end

151 %-----
152 %-----

154 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
155 % Vektor z^N %
156 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

158 % Vektor z^N (rechte Seite des GS)
159 global z;
160 z=zeros(m,1); % Nullvektor wird erzeugt
161 z(2*N+1,1)=2; % Anfangswert Zustand x_0

163 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
164 % M^N*Koeff=z^N %
165 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

167 % Loesen des GS M^N*Koeff=z^N
168 global Koeff;
169 Koeff=M\z

171 global r;
172 r=beta;

174 end

```

A.3 LaguerreGenBeta.m

```

1 function y = LaguerreGenBeta(varargin)
2 %LaguerreGen calculates the generalized Laguerre polynomial L{n,alpha,beta}
3 %
4 % The function computes the generalized Laguerre polynomial L{n,alpha,beta}
5 % and is based on LaguerreGen(varargin).m from Matthias Trampisch.
6 % Here we considered the Laguerre-weight function including the additional
7 % parameter beta: t^{alpha}* e^{-beta t}

10 % Input:
11 % - N = nonnegative integer as degree level
12 % - alpha >= -1 real number (input is optional)
13 % - beta >0 real number
14 %
15 % The output is formatted as a polynomial vector of length (n+1).
16 % [a_0 a_1 a_2 ... a_N]
17 % The lowest coefficient is the first element.
18 % Ln = a_0+a_1*t+a_2*t^2+...+a_N*t^N
19 %

21 %%% ===== %
22 %                set default parameters and rename input
23 % ===== %

24 if (nargin == 2)           %only one parameter "n" supplied
25     N = varargin{1};
26     alpha = 0;           %set default value for alpha
27     beta = 0;
28 elseif (nargin == 3)     %at least two parameters supplied
29     N = varargin{1};
30     alpha = varargin{2};
31     beta = varargin{3};
32 end;
33 %%% ===== %
34 %                error checking of input parameters
35 % ===== %
36 %if (nargin == 0) || (nargin > 2) || (n~=abs(round(n))) || (alpha < -1)
37 %    error('n must be integer, and (optional) alpha >= -1');
38 %end;
39 %%% ===== %
40 %                Recursive calculation of generalized Laguerre polynomial
41 % ===== %
42 L=zeros(N+1);           %reserve memory for faster storage
43 switch N
44     case 0
45         L(1,:)=1;
46     otherwise           %n>1 so we need to do recursion
47         L(1,:)=[zeros(1,N), 1];
48         L(2,:)=[zeros(1, N-1), -1, (alpha+1)];
49         for i=3:N+1
50             A1 = 1/(i-1) * (conv([zeros(1, N-1), -1, (2*(i-1)+alpha-1)], L(i-1,:)));
51             A2 = 1/(i-1) * (conv([zeros(1, N), ((i-1)+alpha-1)], L(i-2,:)));
52             B1=A1(length(A1)-N:1:length(A1));
53             B2=A2(length(A2)-N:1:length(A2));
54             L(i,:)=B1-B2;           % i-th row corresponds to L{i-1, alpha}

```

A Quellcode für das Beispiel aus Abschnitt 6.1

```
55     end;
56 end;
57 %% ===== %
58 %                               Define output
59 % ===== %
60 p=flipr(L(N+1,:)); %last row is the gen. Laguerre polynomial L{n, alpha}
61 Lbeta=zeros(1,N+1);
62
63 for i=1:N+1
64     j=i-1;
65     betaPotenz=beta^j;
66     Lbeta(i)=p(i)*betaPotenz;
67 end;
68
69 LbetaN=flipr(Lbeta);
70 y=LbetaN();
```


Literaturverzeichnis

- [AC79] AUBIN, J. P. ; CLARKE, F. H.: Shadow Prices and Duality for a Class of Optimal Control Problems. In: *SIAM Journal on Control and Optimization* 17 (1979), S. 567–586
- [AE06] AMANN, H. ; ESCHER, J.: *Analysis II*. Birkhäuser, Basel, 2006
- [AF03] ADAMS, R. A. ; FOURNIER, J.J.F.: *Sobolev Spaces*. Second. Academic Press, 2003
- [AK70] ARROW, K. ; KRUZ, M.: *Public Investment, the Rate of Return, and Optimal Fiscal Policy*. New York: RFF Press, 1970
- [AK08] ASEEV, S. M. ; KRYAZHIMSKIY, A. V.: Shadow Prices in Infinite-horizon Optimal Control Problems with Dominating Discounts. In: *Applied Mathematics and Computation* 204 (2008), S. 519–531
- [Alt12] ALT, H. W.: *Lineare Funktionalanalysis, Eine anwendungsorientierte Einführung*. 6., überarbeitete Auflage. Springer-Verlag Berlin Heidelberg, 2012
- [AMR95] ASCHER, U. M. ; MATTHEIJ, R. M. M. ; RUSSELL, R. D.: *Numerical solution of boundary value problems for ordinary differential equations*. SIAM. Classics in Applied Mathematics, 1995
- [Ase17] ASEEV, S. M.: Existence of an optimal control in infinite-horizon problems with unbounded set of control constraints. In: *Proceedings of the Steklov Institute of Mathematics* 297 (2017), S. 1–10
- [ASS13] ALT, W. ; SCHNEIDER, C. ; SEYDENSCHWANZ, M.: *EAGLE-STARHILFE: Optimale Steuerung - Theorie und numerische Verfahren*. Edition am Gutenbergplatz Leipzig, 2013
- [ASS15] ALT, W. ; SCHNEIDER, C. ; SEYDENSCHWANZ, M.: *Ergänzungen zu EAGLE-STARHILFE: Optimale Steuerung - Theorie und numerische Verfahren*. Fakultät für Mathematik und Informatik, Friedrich-Schiller-Universität Jena, 2015
- [AV15] ASEEV, S. M. ; VELIOV, V. M.: Maximum principle for infinite-horizon optimal control problems with dominating discount. In: *Proceedings of the Steklov Institute of Mathematics* 291 (2015), S. 22–39
- [Bet98] BETTS, J. T.: A Survey of Numerical Methods for Trajectory Optimization. In: *Journal of Guidance, Control, and Dynamics* 21 (1998), S. 193–207
- [Bet04] BETTEN, J.: *Finite Elemente für Ingenieure 2. Variationsrechnung, Energiemethoden, Näherungsverfahren, Nichtlinearitäten, Numerische Integrationen*. Springer-Verlag Berlin Heidelberg, 2004

- [BHTR06] BENSON, D. A. ; HUNTINGTON, G. T. ; THORVALDSEN, T. ; RAO, A. V.: Direct Trajectory Optimization and Costate Estimation via an Orthogonal Collocation Method. In: *Journal of Guidance, Control and Dynamics* 29 (2006), S. 1435–1440
- [BP84] BOCK, H. G. ; PLITT, K. J.: A Multiple Shooting Algorithm for Direct Solution of Optimal Control Problems. In: *IFAC Proceedings Volumes* 17 (1984), S. 1603–1608
- [BR14] BANERJEE, S. ; ROY, A.: *Linear Algebra and Matrix Analysis for Statistics*. CRC Press, Taylor & Francis Group, 2014 (Chapman & Hall/CRC texts in Statistical Science)
- [Bul71] BULIRSCH, R.: Die Mehrzielmethode zur numerischen Lösung von nichtlinearen Randwertproblemen und Aufgaben der optimalen Steuerung / Report der Carl-Cranz-Gesellschaft. 1971. – Forschungsbericht
- [Bur98] BURENKOV, V. I.: *Sobolev Spaces on Domains*. Vieweg+Teubner Verlag, Wiesbaden, 1998
- [ByXy05] BEN-YU, G. ; XIAO-YONG, Z.: A new generalized Laguerre spectral approximation and its applications. In: *Journal of Computational and Applied Mathematics* 184 (2005), S. 382–403
- [BZ18] BURTCHEM, A. ; ZIEMANN, T.: Analytische Hilfsmittel der Steuerungstheorie. In: *Preprint-Reihe Mathematik M-01/2018, BTU Cottbus-Senftenberg* (2018)
- [Car95] CARLSON, D. A.: An existence theorem for hereditary Lagrange problems on an unbounded interval. In: *Boundary Value Problems for Functional Differential Equations, World Scientific Publishing, Singapore* (1995), S. 73–83
- [CB89] CUTHRELL, J. E. ; BIEGLER, L. T.: Simultaneous optimization and solution methods for batch reactor control profiles. In: *Computers & Chemical Engineering* 13 (1989), S. 49–62
- [CHL91] CARLSON, D. ; HAURIE, A. B. ; LEIZAROWITZ, A.: *Infinite Horizon Optimal Control*. Springer-Verlag Berlin Heidelberg, 1991
- [CHQZ06] CANUTO, C. ; HUSSAINI, M.Y. ; QUARTERONI, A. ; ZANG, Th.A.: *Spectral Methods*. Springer-Verlag Berlin Heidelberg, 2006
- [Dav75] DAVIS, P. J.: *Interpolation and Approximation*. Dover Publications, 1975
- [Dik16] DIKARIEV, I.: *Ein Beitrag zur Existenz optimaler Lösungen bei Steuerungsprobleme mit unendlichem Zeithorizont*, Brandenburgische Technische Universität Cottbus-Senftenberg, Masterarbeit, 2016
- [DS88] DUNFORD, N. ; SCHWARTZ, J. T.: *Linear Operators Part I: General Theory*. Wiley Classics Library Edition, 1988
- [EKR95] ELNAGAR, G. ; KAZEMI, M. A. ; RAZZAGHI, M.: The Pseudospectral Legendre Method for Discretizing Optimal Control Problems. In: *IEEE Transactions on Automatic Control* 40 (1995), S. 1793 – 1796

-
- [Els11] ELSTRODT, J.: *Maß- und Integrationstheorie*. 7. Springer Berlin Heidelberg, 2011 (Springer-Lehrbuch)
- [Eva98] EVANS, L. C.: *Graduate Studies in Mathematics*. Bd. 19: *Partial Differential Equations*. Second. American Mathematical Society, 1998
- [Fau09] FAUSTMANN, M.: *Gewichtete Sobolev-Räume und nicht "klassische" lineare elliptische Randwertprobleme*, Technische Universität Wien, Bachelorarbeit, 2009
- [FH86] FEICHTINGER, G. ; HARTL, R. F.: *Optimale Kontrolle ökonomischer Prozesse. Anwendungen des Maximumprinzips in den Wirtschaftswissenschaften*. Walter De Gruyter & Co., 1986
- [Fil88] FILIPPOV, A. F. ; ARSCOTT, F.M. (Hrsg.): *Differential Equations with Discontinuous Righthand Sides*. Springer Netherlands, 1988
- [For12] FORSTER, O.: *Analysis 3. Maß- und Integrationstheorie, Integralsätze im \mathbb{R}^n und Anwendungen*. Vieweg+Teubner Verlag, 2012
- [FR00] FAHROO, F. ; ROSS, I. M.: Direct Trajectory Optimization by a Chebyshev Pseudospectral Method. In: *Proceedings of the 2000 American Control Conference* (2000)
- [FR08] FAHROO, F. ; ROSS, I. M.: Pseudospectral Methods for Infinite-Horizon Nonlinear Optimal Control Problems. In: *Journal of Guidance, Control, and Dynamics* 31 (2008), S. 927–936
- [Fre69] FREUD, G.: *Orthogonale Polynome*. Birkhäuser, Basel, 1969
- [Fun92] FUNARO, D.: *Lecture Notes in Physics Monographs*. Bd. 8: *Polynomial Approximation of Differential Equations*. Springer-Verlag Berlin Heidelberg, 1992
- [GCF⁺08] GRASS, D. ; CAULKINS, J. P. ; FEICHTINGER, G. ; TRAGLER, G. ; BEHRENS, D. A.: *Optimal Control of Nonlinear Processes with Applications in Drugs, Corruption, and Terror*. Springer-Verlag Berlin Heidelberg, 2008
- [Ger11] GERDTS, M.: *Optimal Control of ODEs and DAEs*. Walter De Gruyter, 2011
- [Ger13] GERDTS, M.: *User's Guide OCPID-DAE1 Optimal Control and Parameter Identification with Differential-Algebraic Equations of Index 1, Version 1.3*, 2013
- [GHR11] GARG, D. ; HAGER, W. W. ; RAO, A. V.: Pseudospectral Methods for Solving Infinite-Horizon Optimal Control Problems. In: *Automatica* 47 (2011), Nr. 4, S. 829–837
- [GJ16] GRÜNE, L. ; JUNGE, O.: *Gewöhnliche Differentialgleichungen*. Springer Spektrum, Wiesbaden, 2016
- [GK99] GEIGER, C. ; KANZOW, C.: *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer, Berlin, Heidelberg, 1999
- [Göl17] GÖLLMANN, L.: *Lineare Algebra*. Springer Spektrum, Berlin, Heidelberg, 2017

- [Göp73] GÖPFERT, A.: *Mathematische Optimierung in allgemeinen Vektorräumen*. BSB B. G. Teubner Verlagsgesellschaft, 1973 (Mathematisch-Naturwissenschaftliche Bibliothek, Band 58)
- [GPF⁺11] GARG, D. ; PATTERSON, M. A. ; FRANCOLIN, C. ; DARBY, C. L. ; HUNTINGTON, G. T. ; HAGER, W. W.: Direct trajectory optimization and costate estimation of finite-horizon and infinite-horizon optimal control problems using a Radau pseudospectral method. In: *Computational Optimization and Applications* 49 (2011)
- [GRT09] GÖPFERT, A. ; RIEDRICH, T. ; TAMMER, C.: *Angewandte Funktionalanalysis*. Vieweg+Teubner Verlag, 2009
- [GW69] GOLUB, G.H. ; WELSCH, J.H.: Calculation of Gauss quadrature rules. In: *Mathematics of Computation* 23 (1969), S. 221–230
- [Hah59] HAHN, W.: *Theorie und Anwendung der direkten Methode von Ljapunov*. Springer-Verlag Berlin Heidelberg, 1959 (Ergebnisse der Mathematik und ihrer Grenzgebiete. 2. Folge 22)
- [Hal74] HALKIN, H.: Necessary Conditions for Optimal Control Problems with Infinite Horizons. In: *Econometrica* 42 (1974), S. 267–272
- [Hau03] HAURIE, A.: Integrated Assessment Modeling for Global Climate Change: An Infinite Horizon Optimization Viewpoint. In: *Environmental Modeling & Assessment* 8 (2003), S. 117–132
- [Heu04] HEUSER, H.: *Gewöhnliche Differentialgleichungen*. Vieweg+Teubner Verlag, 2004
- [IK02] ITO, K. ; KUNISCH, K.: Receding horizon optimal control for infinite dimensional systems. In: *ESAIM: Control, Optimisation and Calculus of Variations* 8 (2002), S. 741–760
- [Isk18] ISKE, A.: *Approximation*. Springer Spektrum, 2018
- [Kal60] KALMAN, R. E.: Contributions to the Theory of Optimal Control. In: *Boletín de la Sociedad Matemática Mexicana* 5 (1960), S. 102–119
- [Kan05] KANZOW, C.: *Numerik linearer Gleichungssysteme. Direkte und iterative Verfahren*. Springer, Berlin, Heidelberg, 2005
- [Kel68] KELLER, H. B.: *Numerical Methods for Two-Point Boundary-Value Problems*. Blaisdell Publishing Company, 1968
- [KF61] KOLMOGOROV, A.N. ; FOMIN, S.V.: *Elements of the Theory of Functions and Functional Analysis*. Bd. 2, Measure. The Lebesgue Integral. Hilbert Space. Graylock Press, 1961
- [KJF77] KUFNER, A. ; JOHN, O. ; FUCIK, S.: *Function Spaces*. Czechoslovak Academy of Science, 1977
- [KO84] KUFNER, A. ; OPIC, B.: How to define reasonably weighted Sobolev spaces. In: *Commentationes Mathematicae Universitatis Carolinae* 25 (1984), S. 537–554

-
- [Kre15] KREUTER, M.: *Sobolev Spaces of Vector-Valued Functions*, Ulm University Faculty of Mathematics and Economics, Master Thesis, 2015. https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.020/abschlussarbeiten/MA_Marcel_Kreuter.pdf
- [LL79] LANDAU, L. D. ; LIFSCHITZ, E. M.: *Lehrbuch der theoretischen Physik III, Quantenmechanik*. Akademie-Verlag Berlin, 1979
- [LPW08] LYKINA, V. ; PICKENHAIN, S. ; WAGNER, M.: Different interpretations of the improper integral objective in an infinite horizon control problem. In: *Journal of Mathematical Analysis and Applications* 340 (2008), S. 498–510
- [Lya92] LYAPUNOV, A. M.: *The general problem of the stability of motion. Translated and edited by A.T. Fuller*. Taylor & Francis, 1992
- [Lyk10] LYKINA, V.: *Beiträge zur Theorie der Optimalsteuerungsprobleme mit unendlichem Zeithorizont*, Fakultät für Mathematik, Naturwissenschaften und Informatik der Brandenburgischen Technischen Universität Cottbus, Diss., 2010
- [Lyk16] LYKINA, V.: An Existence Theorem for a Class of Infinite Horizon Optimal Control Problems. In: *Journal of Optimization Theory and Applications* 169 (2016), April, S. 50–73
- [Mag80] MAGILL, M. J. P.: An existence criterion for models of resource allocation over an infinite horizon. In: *Economics Letters* 5 (1980), S. 209–213
- [Mag82] MAGILL, M. J. P.: Pricing Infinite Horizon Programs. In: *Journal of Mathematical Analysis and Applications* 88 (1982)
- [Mau76] MAURER, H.: Numerical solution of singular control problems using multiple shooting techniques. In: *Journal of Optimization Theory and Applications* 18 (1976), S. 235–257
- [MG75] MAURER, H. ; GILLESSEN, W.: Application of multiple shooting to the numerical solution of optimal control problems with bounded state variables. In: *Computing* 15 (1975), S. 105–126
- [Mic82] MICHEL, P.: On the Transversality Condition in Infinite Horizon Optimal Problems. In: *Econometrica* 50 (1982), S. 975–985
- [MP95] MAURER, H. ; PESCH, H. J.: Solution differentiability for parametric nonlinear control problems with control-state constraints. In: *Journal of Optimization Theory and Applications* 86 (1995), S. 285–309
- [MS64] MEYERS, N. ; SERRIN, J.: $H=W$. In: *Proc. Natl. Acad. Sci. USA* 51 (1964), S. 1055–1056
- [MV01] MEYBERG, K. ; VACHENAUER, P.: *Höhere Mathematik 2. Differentialgleichungen, Funktionentheorie, Fourier-Analyse, Variationsrechnung*. Springer-Verlag Berlin Heidelberg, 2001 (4)

- [Obe86] OBERLE, H. J.: Numerical solution of minimax optimal control problems by multiple shooting technique. In: *Journal of Optimization Theory and Applications* 50 (1986), S. 331–357
- [OG89] OBERLE, H. J. ; GRIMM, W.: BNDSCO - A Program for the Numerical Solution of Optimal Control Problems. In: *Report 515, Institut for Flight Systems Dynamics, German Aerospace Research Establishment DLR* (1989)
- [Os69] OSBORNE, M.R.: On shooting methods for boundary value problems. In: *Journal of Mathematical Analysis and Applications* 27 (1969), S. 417–433
- [PB18] PICKENHAIN, S. ; BURTCHEM, A.: Regulator Problems on Unbounded Domains, Stationarity-Optimal Control-Asymptotic Controllability. In: *Vietnam Journal of Mathematics* (2018)
- [PB19] PICKENHAIN, S. ; BURTCHEM, A.: Problems in the Calculus of Variations on Unbounded Intervals - Fourier-Laguerre Analysis and Approximations. In: *Vietnam Journal of Mathematics* (2019), S. 1–22
- [PBK16] PICKENHAIN, S. ; BURTCHEM, A. ; KOLO, K.: An indirect pseudospectral method for the solution of linear-quadratic optimal control problems with infinite horizon. In: *Optimization* 65 (2016), S. 609–633
- [Pic09] PICKENHAIN, S.: On Adequate Transversality Conditions for Infinite Horizon Optimal Control Problems—A Famous Example of Halkin. In: *Dynamic Systems, Economic Growth, and the Environment* (2009), S. 3–22
- [Pic14] PICKENHAIN, S.: Hilbert Space Treatment of Optimal Control Problems with Infinite Horizon. In: *Bock, H. G.; Phu, H. X.; Rannacher R.; Schloeder, J.P.: Modeling, Simulation and Optimization of Complex Processes - HPSC 2012, Springer* (2014), S. 169–182
- [Pic15] PICKENHAIN, S.: Infinite Horizon Optimal Control Problems in the Light of Convex Analysis in Hilbert Spaces. In: *Set-Valued and Variational Analysis* 23 (2015), S. 169–189
- [Pic19] PICKENHAIN, S.: Infinite Horizon Problems in the Calculus of Variations. The Role of Transformations with an Application to the Brachistochrone Problem. In: *Set-Valued and Variational Analysis* 27 (2019), S. 331–354
- [PL17] PICKENHAIN, S. ; LYKINA, V.: Weighted functional spaces approach in infinite horizon optimal control problems: A systematic analysis of hidden opportunities and advantages. In: *Journal of Mathematical Analysis and Applications* 454 (2017), S. 195–218
- [PLW08] PICKENHAIN, S. ; LYKINA, V. ; WAGNER, M.: On the lower semicontinuity of functionals involving Lebesgue or improper Riemann integrals in infinite horizon optimal control problems. In: *Control and Cybernetics* 37 (2008), S. 451 – 468
- [PR78] PESCH, H.J. ; RENTROP, P.: Numerical Solution of the Flow between Two-Counter-Rotating Infinite Plane Disks by Multiple Shooting. In: *Journal of Applied Mathematics and Mechanics* 58 (1978), S. 23–28

-
- [PW12] PICKENHAIN, S. ; WENZKE, D.: Transformation of Infinite Horizon Optimal Control Problems. In: *Preprint Reihe Mathematik M-02/2012, BTU Cottbus* (2012)
- [Rao09] RAO, A.V.: A Survey of Numerical Methods for Optimal Control. In: *Advances in Astronautical Sciences* 135 (2009), S. 497–528
- [Sac92] SACHS, E.W.: Topics in Modern Computational Methods for Optimal Control Problems. In: KRABS, W. (Hrsg.) ; ZOWE, J. (Hrsg.): *Modern Methods of Optimization. Lecture Notes in Economics and Mathematical Systems* Bd. 378, Springer, Berlin, Heidelberg, 1992, S. 313–348
- [SB92] STRYK, O. v. ; BULIRSCH, R.: Direct and Indirect Methods for Trajectory Optimization. In: *Annals of Operations Research* 37 (1992), S. 357–373
- [SB93] STOER, J. ; BULIRSCH, R.: *Introduction to Numerical Analysis*. Second Edition. Springer-Verlag New York, 1993
- [Sch71] SCHÖNHAGE, A.: *Approximationstheorie*. Walter De Gruyter & Co, 1971
- [Sei99] SEIERSTAD, A.: Necessary Conditions for Nonsmooth, Infinite-Horizon, Optimal Control Problems. In: *Journal of Optimization Theory and Applications* 103 (1999), S. 201–229
- [SK11] SCHWARZ, R. ; KÖCKLER, N.: *Numerische Mathematik*. 8. Auflage. Vieweg+Teubner Verlag, 2011
- [Son83] SONTAG, E. D.: A Lyapunov-Like Characterization of Asymptotic Controllability. In: *SIAM Journal on Control and Optimization* 21 (1983), Nr. 3, S. 462–471
- [Son98] SONTAG, E. D.: *Mathematical Control Theory. Deterministic Finite Dimensional Systems*. Springer-Verlag New York, 1998
- [SS76] SCHMEISSER, G. ; SCHIRMEIER, H.: *Praktische Mathematik*. Walter De Gruyter GmbH & Co. KG, 1976
- [Sto05] STOER, J.: *Numerische Mathematik 1*. Bd. 9. Springer-Verlag Berlin Heidelberg, 2005
- [Str94] STRYK, O. v.: *Numerische Lösung optimaler Steuerungsprobleme: Diskretisierung, Parameteroptimierung und Berechnung der adjungierten Variablen*, Technische Universität München, Diss., 1994
- [Str99] STRYK, O.: *User's Guide for DIRCOL: A Direct Collocation Method for the Numerical Solution of Optimal Control Problems, Version 2.1*, 1999
- [STW11] SHEN, J. ; TANG, T. ; WANG, L.: *Springer Series in Computational Mathematics*. Bd. 41: *Spectral Methods. Algorithms, Analysis and Applications*. Springer-Verlag Berlin Heidelberg, 2011
- [Sze39] SZEGÖ, G.: *Orthogonal Polynomials*. Bd. XXIII. American Mathematical Society, 1939

- [Tau10] TAUCHNITZ, N.: *Das Pontrjaginsche Maximumprinzip für eine Klasse hybrider Steuerungsprobleme mit Zustandsbeschränkungen und seine Anwendung*, Fakultät für Mathematik, Naturwissenschaften und Informatik der Brandenburgischen Technischen Universität Cottbus, Diss., 2010
- [Tri70] TRICOMI, F.G.: *Vorlesungen über Orthogonalreihen*. Springer-Verlag Berlin Heidelberg, 1970
- [VD88] VLASSEN BROECK, J. ; DOOREN, R. van: A Chebyshev Technique for Solving Non-linear Optimal Control Problems. In: *IEEE Transactions on Automatic Control* 33 (1988), Nr. 4, S. 333–340
- [Vla88] VLASSEN BROECK, J.: A chebyshev polynomial method for optimal control with state constraints. In: *Automatica* 24 (1988), S. 499–506
- [Wag09] WAGNER, M.: Pontryagin’s Maximum Principle for Multidimensional Control Problems in Image Processing. In: *Journal of Optimization Theory and Applications* 140 (2009), Mar, Nr. 3, S. 543–576
- [Wag17] WAGNER, M.: *Lineare und nichtlineare FEM*. Springer Vieweg, Wiesbaden, 2017
- [Wal00] WALTER, W.: *Gewöhnliche Differentialgleichungen*. 7. Springer-Verlag Berlin Heidelberg, 2000
- [Wei00] WEIDEMANN, J.: *Lineare Operatoren in Hilberträumen*. Vieweg+Teubner Verlag, 2000
- [Wer18] WERNER, D.: *Funktionalanalysis*. Bd. 8. Springer Spektrum, 2018
- [Wil04] WILLIAMS, P.: Jacobi Pseudospectral Method for Solving Optimal Control Problems. In: *Journal of Guidance, Control and Dynamics* 27 (2004), S. 293–297
- [XyH13] XIAO-YONG, Z. ; HUAFENG, W.: Generalized Laguerre approximation and its application to ordinary differential equation. In: *Applied Mathematics & Information Sciences* 7, No. 3 (2013), S. 1117–1124
- [Yos95] YOSIDA, K.: *Functional Analysis*. 6. Springer-Verlag Berlin Heidelberg, 1995
- [Zei95] ZEIDLER, E.: *Applied Mathematical Sciences*. Bd. 108: *Applied Functional Analysis*. Springer-Verlag New York, 1995
- [Zie15] ZIEMANN, T.: *Ein Maximumprinzip für eine Klasse von Optimalsteuerungsaufgaben über eeine unendlichen Zeithorizont mit vektoriellen Zuständen und Steuerungen*, Brandenburgische Technische Universität Cottbus-Senftenberg, Masterarbeit, 2015
- [ZP18] ZIEMANN, T. ; PICKENHAIN, S.: First-order necessary optimality conditions for infinite horizon optimal control problems with linear dynamics and convex objective. In: *Optimization* 67 (2018), S. 757–782