Diskretisierung Elliptischer Steuerungsprobleme

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalis

vorgelegt dem Rat der Fakultät für Mathematik und Informatik
der Friedrich–Schiller–Universität Jena

eingereicht von Dipl.-Math. Nils Bräutigam

geb. am 23.07.1979 in Pößneck (Thüringen)

Jena, 24.04.2006

Gutachter:

1. **Prof. Dr. Walter Alt**
2. **Prof. Dr. Christoph Büskens**

Tag der letzten Prüfung des Rigorosums: 27. Juni 2006
Tag der öffentlichen Verteidigung: 30. Juni 2006

Für Messane

und ihre Geschwister

Zusammenfassung

Im Mittelpunkt der Untersuchungen steht das folgende Problem der Optimalen Steuerung

$$\min J(z,u) = \frac{1}{2}\int_0^T |z(t) - z_d(t)|^2 + \nu\,|u(t)|^2\,dt$$

unter den Nebenbedingungen

$$\begin{aligned} -\ddot{z}(t) + Az(t) &= Bu(t) + e(t) && \forall' t \in [0,T] \\ z(0) = z(T) &= 0 \\ a \le u(t) &\le b && \forall' t \in [0,T]. \end{aligned}$$

Diese Problemstellung beschreibt den gleichzeitigen Wunsch den gewünschten Zustand z_d unter minimalem Verbrauch der eingesetzten Energie u zu erreichen. Auf Grund der kontinuierlichen Bedingung über das Intervall $[0,T]$ entfällt die Möglichkeit einer direkten Implementierung solcher Probleme. Nach theoretischen Überlegungen über Lösbarkeit und Eindeutigkeit der Lösung ist es das Ziel, über eine geeignete Diskretisierung das Problem numerisch zu lösen. Wir verwenden die Methode der Finiten Elemente und die Methode der Finiten Differenzen, woraus sich diskrete Steuerungsprobleme ergeben, deren numerische Berechnung durchführbar ist.

Für den Abstand der Lösungen der beiden Probleme entwickeln wir in Abhängigkeit von der Gitterweite oberer Schranken. In beiden Fällen gelingt es uns im quadratischen Mittel quadratische Konvergenzordnung zu erreichen. Ferner definieren wir eine neue zulässige Steuerung und zeigen für diese punktweise quadratische Konvergenz gegen die exakte Lösung.

Die hergeleiteten theoretischen Resultate untermalen wir an Hand eines Beispielproblems, bei dem die optimale Steuerung bekannt und somit die Aussagen über den Fehler nachprüfbar sind. Den Rahmen bildet eine konkrete Anwendung der stationären Wärmeverteilung in einem Stab.

Vorwort

Die vorliegende Dissertation entstand während meiner zweijährigen Arbeit am Institut für Angewandte Mathematik der Friedrich-Schiller-Universität Jena. Unterstützt durch ein Stipendium des Freistaates Thüringen erhielt ich so die Gelegenheit meinem Wissensdurst über das Studium der Mathematik hinaus nachzugehen. Viele Hinweise und Anregungen gingen in dieser Zeit bei mir ein, so unter anderem die Beachtung und Unterscheidung der Vielzahl von Abschätzungskonstanten. Diese versah ich mit der Nummer der (Un)Gleichung, in der sie das erste Mal auftraten, so dass der Leser jederzeit ihren konkreten Wert nachschlagen kann. Die dadurch wachsende Komplexität so mancher Abschätzung wird durch die Gewissheit gerechtfertigt, keine Abhängigkeit von den wichtigen Größen zu übersehen. Die definierten diskreten Probleme erhalten die einheitliche Bezeichnung $(\mathrm{StP})_h$. Im jeweiligen Abschnitt ist damit jeweils das zu Beginn definierte Problem gemeint. Darüber hinaus habe ich stets versucht, bei den Erläuterung auf eine unnötige Allgemeinheit zu verzichten, damit dem Leser die für die Problemstellung wichtigen Konzepte und Resultate nicht durch zu viel Peripherie verdeckt werden.

Die Arbeit gliedert sich in sieben Kapitel. An erster Stelle steht die praktische Rechtfertigung der theoretischen Arbeiten durch eine konkrete Anwendung. Nach einer kurzen Einleitung inklusive Vorstellung der verwendeten Notationen sowie der mathematischen Problemstellung folgt die theoretische Betrachtung der Aufgaben. Einer Einführung in die Diskretisierung und der damit verbundenen diskreten Konzepte folgend, stellen wir in Kapitel 4 die Methode der Finiten Elemente vor. Anschließen wird sich die Methode der Finiten Differenzen in Kapitel 5. Die Unterschiede und Gemeinsamkeiten, sowie eine Abstraktion der Vorgehensweisen und die Herleitung von abstrakten Konvergenzaussagen sind der Inhalt von Kapitel 6. Anschließend kehren wir zur Anwendung zurück und berechnen auf den verschiedenen Wegen Lösungen für das Problem. Im Anhang finden wir Erweiterungen zu den Ergebnissen, sowie wichtige Hilfsresultate.

An dieser Stelle möchte ich es nicht versäumen, all jenen Menschen zu danken, mit deren Hilfe die Entstehung dieser Arbeit erst möglich gewesen ist. So danke ich dem Betreuer meiner Arbeit, Herrn Prof. Dr. Walter Alt, für seine fachliche Hilfe sowie für so manch anregendes Gespräch darüber hinaus. Weiterhin möchte ich den Mitarbeitern am Institut für Angewandte Mathematik in ihrer Gesamtheit sowie Dr. Dieter Kai-

ser, Dr. Joachim Jüngel und Dr. Michael Fritzsche im Speziellen meinen Dank für die wertvollen Diskussionen und Anregungen aussprechen. Meinen Freunden Katrin Hilse, Thomas Milde und Tilman Zscheckel danke ich für das aufwendige Lesen und Korrigieren der (fast) fertigen Dissertation und darüber hinaus für das wertvolle Aufmuntern. Frau Heidemarie Langner zeichnete sich für das Fernhalten jeglicher Bürokratie von meiner Person verantwortlich, was nicht genug der Anerkennung verdient. Neben der mathematischen Seite, beinhaltet der gesamte Rahmen der Promotion auch einen menschlichen Aspekt. In dieser Hinsicht lässt sich Hilfe und Unterstützung selten in Worten und Mengen darstellen. Mein Dank richtet sich an meine Eltern und meine Familie, die mich diesen Lebensweg einschlagen ließen und fortwährend unterstützen.

An letzter Stelle möchte ich einer Person meinen Dank aussprechen, deren Rolle einen unendlich großen Rahmen einnimmt. Ohne meine Frau Mandy wäre diese Leistung nicht in Ansätzen durchführbar gewesen. Dein Zuspruch und Vertrauen ist die Quelle meiner Kraft eines jeden Tages.

Inhaltsverzeichnis

Glossar

Grundlagen und Funktionen

(a,b), $[a,b]$	offenes bzw. abgeschlossenes Intervall
$\mathbb{R}$	Menge der reellen Zahlen
$\mathbb{R}^n$	Vektorraum der reellen n-dimensionalen Vektoren
0_n	Nullvektor im $\mathbb{R}^n$
I_n	n-dimensionale Einheitsmatrix
$\forall'$	für fast alle t
t_i	Stützstelle zur Unterteilung des Intervalls $[0,T]$
T_i	Teilintervall $[t_i, t_{i+1})$
S_i	Mittelpunkt des Intervalls T_i
u	Steuerung
$z(u)$, $p(u)$	Zustand bzw. adjungierter Zustand zu u
e	Störung auf der rechten Seite der Systemgleichung
y	Hilfsvariable für die rechte Seite der Systemgleichung
$\bar{u}$	Optimale Steuerung, d.h. Lösung von (StP)
u_h	diskrete optimale Steuerung, d.h. Lösung von $(\mathrm{StP})_h$
$\bar{u}_h$	semi-diskrete optimale Steuerung, d.h. Lösung von $(\mathrm{StP})_{\frac{h}{2}}$
u_h^*	Bestapproximation von $\bar{u}$ im Raum $U_h[0,T,\mathbb{R}^m]$
$\bar{p} = p(\bar{u})$	Adjungierter Zustand zur optimalen Steuerung
$z_h(u)$, $p_h(u)$	diskreter Zustand bzw. diskreter adjungierter Zustand zu u
$p_h^*(u)$	diskreter adjungierter Zustand zu Steuerung u und $z(u)$
$\tilde{u}$	aus den Optimalitätsbedingungen berechnete zulässige Steuerung
α, β, ζ	äquivalente Vektoren zu den Funktionen u_h, z_h, p_h
γ	Fehler bei der Diskretisierung der Systemgleichung
$\widetilde{J}$, $\widetilde{J}_h$	Zielfunktional bezüglich Zustand und Steuerung bzw. dessen Diskretisierung
J, J_h	Zielfunktional bezüglich der Steuerung bzw. dessen Diskretisierun
z_d, ν	gewünschter Zustand bzw. Gewichtungsparameter im Zielfunktional
A, B	Parametermatrizen in der Systemgleichung
U^{ad}, U_h^{ad}	zulässige Menge und deren Diskretisierung
$\mathcal{L}$	Lagrange-Funktion
H	Hamilton-Funktion

Funktionenräume und Operatoren

$\mathscr{D}^j$	j-te schwache Ableitung
$L_p[0,T,\mathbb{R}^n]$	Raum der p-integrierbaren Funktionen mit Wertebereich $\mathbb{R}^n$
$W_p^k[0,T,\mathbb{R}^n]$	Raum der absolut stetigen Funktionen, deren k-te schwache Ableitung im $L_p[0,T,\mathbb{R}^n]$ enthalten ist
$W_{p,0}^k[0,T,\mathbb{R}^n]$	Raum der Funktionen aus $W_p^k[0,T,\mathbb{R}^n]$ mit verschwindenden Randwerten
$C[0,T,\mathbb{R}^n]$	Raum der stetigen Funktion mit Wertebereich $\mathbb{R}^n$
$C^l[0,T,\mathbb{R}^n]$	Raum der l-mal klassisch differenzierbaren Funktionen
$C_0^\infty[0,T,\mathbb{R}^n]$	Raum der beliebig oft differenzierbaren Funktionen und verschwindenden Randwerten
$\bigvee_a^b f$	totale Variation von f auf dem Intervall $[a,b]$
$BV[0,T,\mathbb{R}^m]$	Raum aller Funktionen von bechränkter Variation
$P_m[0,T,\mathbb{R}^n]$	Raum der Polynome vom Grad höchstens m und Wertebereich $\mathbb{R}^n$
X'	Dualraum zu X
$\mathscr{L}(X,Y)$	Raum der linearen Funktionale zwischen X und Y
ℓ, $\ell\|_X$	lineares Funktional bzw. dessen Einschränkung auf die Menge X
$U_h[0,T,\mathbb{R}^n]$	Raum der stückweise kosntante Funktionen
$V_h[0,T,\mathbb{R}^n]$	Raum der stückweise linearen Funktionen
$V_{h,0}[0,T,\mathbb{R}^n]$	Raum der stückweise linearen Funktionen mit verschwindenden Randwerten
$u_h^{(j)}$, $v_h^{(j)}$	Basisfunktionen von $U_h[0,T,\mathbb{R}^m]$ bzw. $V_{h,0}[0,T,\mathbb{R}^n]$
$\|\cdot\|_p$	Norm auf $L_p[0,T,\mathbb{R}^n]$, d.h. $\|f\|_p = \left(\int_0^T \|f(t)\|^p\right)^{\frac{1}{p}}$
$\|\cdot\|_\infty$	Norm auf $L_\infty[0,T,\mathbb{R}^n]$, d.h. $\|f\|_\infty = \operatorname{ess\,sup}\ \|f(t)\|$
$\|\cdot\|_{k,p}$	Norm auf $W_p^k[0,T,\mathbb{R}^n]$, d.h. $\|f\|_{k,p} = \left(\sum_{j=0}^k \|\mathscr{D}^j f(t)\|_p^p\right)^{\frac{1}{p}}$
$\|\cdot\|_{k,p}$	Semi-Norm auf $W_p^k[0,T,\mathbb{R}^n]$, d.h. $\|f\|_{k,p} = \|\mathscr{D}^k f(t)\|_p$
$\|\cdot\|$	Abkürzung für $\|\cdot\|_2$ bzw. Matrixnorm
$\langle\cdot,\cdot\rangle$	Skalarprodukt auf dem $L_2[0,T,\mathbb{R}^n]$
$a(\cdot,\cdot)$	elliptische und beschränkte Bilinearform auf $W_{2,0}^1[0,T,\mathbb{R}^n]$
$a_h(\cdot,\cdot)$	elliptische und beschränkte Bilinearform auf $V_{h,0}[0,T,\mathbb{R}^n]$
$\langle\cdot,\cdot\rangle_h$	diskretes Skalarprodukt auf dem $V_h[0,T,\mathbb{R}^n]$
$\langle\cdot,\cdot\rangle_C = \langle\cdot,C\cdot\rangle$	diskretes Skalarprodukt mit positiv definiter Matrix C
$\|\cdot\|_h$	diskrete Norm auf $V_h[0,T,\mathbb{R}^n]$
$(\cdot)_{\mathsf{h}}$, $(\cdot)_{\bar{\mathsf{h}}}$	vorwärts bzw. rückwärts gerichtete Euler-Approximation
$(\cdot)_{\mathsf{h}\bar{\mathsf{h}}}$	Hintereinanderausführung bzw. zentraler Differenzenstern
Δ_x	Laplace-Operator bezüglich der x-Variable
Δ	Laplace-Operator bezüglich der einzigen Variable
$\mathcal{T}$	Transformation $\mathcal{T}\cdot = B\cdot + e$
$\mathcal{S}$, $\mathcal{S}^*$	Lösungsoperator für die Systemgleichung und adjungierter Operator
$\mathcal{S}_h$, $\mathcal{S}_h^*$	diskrete Approximationen der Operatoren
$\Pi_{[a,b]}$	Projektionsoperator auf die zulässige Menge
P_0, P_1	Diskretisierungsoperatoren auf den Raum $U_h[0,T,\mathbb{R}^m]$ bzw. $V_h[0,T,\mathbb{R}^n]$

Werte der Konstanten

$c^o_{2.4}$	$\max\{1, \|A\|\}$
$c^u_{2.4}$	$\frac{1}{2}\min\{1, \frac{1}{T^2}\} = {c_{2.6}}^{-1}$
$c_{2.4}$	$\sqrt{{c^o_{2.4}}^3 c^u_{2.4}}$
$c_{2.6}$	$\sqrt{2}\max\{1, T\}$
$c_{2.7}$	$c_{2.4}c^2_{2.6} = 2\max\{1, \|A\|\}\max\{1, T^2\}$
$c_{4.9}$	$c^3_{2.6}c^3_{2.7}c_{2.4}$
$c_{4.13}$	$\frac{T}{4}\|B\|\, c_{2.7}$
$c_{4.14}$	$\max\{c_{4.9}, T^2 c_{4.13}\}$
c^Π	$\frac{\|B\|}{\nu}$
$c_{4.15}$	$c^\Pi(c_{4.14} + \frac{1}{8}c_{2.6}c_{2.7})$
$c_{5.8}$	$1 + \frac{T^2}{\sqrt{2}}\|A\|$
$c_{5.9}$	$\frac{T}{2}c_{5.8}c^o_{2.4}c_{2.6}c_{2.7}$

$C^{\bar{u}}_{4.14}$	$\|\mathcal{T}\bar{u}\|_\infty + \|z_d\|_\infty + \mathsf{V}^T_0\dot{u} + \|\dot{\bar{u}}\|_\infty$
$C^{\bar{u}}_{5.13}$	$\|\mathcal{T}\bar{u}\|_{1,\infty} + \mathsf{V}^T_0(\dot{\mathcal{T}\bar{u}}) + \|z_d\|_{2,\infty}$

Kapitel 1

Einführung

1.1 Anwendung

In einer Zeit, in der mehr und mehr über Energiepreise diskutiert wird, gibt es selbstverständlich auch Ideen über den sparsamen Umgang mit Energie, also ihren Verbrauch zu minimieren. Diese Überlegungen sind natürlich nicht neu, rücken aber durch Diskussionen über Rohstoffe und deren Verwertung regelmäßig in den Mittelpunkt. Energie zu sparen, aber dennoch seine Ziele so vollkommen wie möglich zu erreichen, diese Aufgabe stellen sich Privatpersonen im täglich Leben, Führungspersonen im geschäftlichen Bereich und auch Ingenieure im produktiven Bereich. Dabei ist der Begriff „Energie“ allgemein als „Kosten“ zu interpretieren, die einen negativen Nutzen für die handelnde Person nach sich zieht. Eines dieser Probleme aus dem ingenieuer-wissenschaftlichen Bereich ist die Untersuchung der Wärmeverteilung in einem Medium. An Hand empirischer Messungen wurden bereits Zusammenhänge erforscht und Modelle entwickelt. Eines dieser Modelle betrifft die Verbindung zwischen von außen zugeführter Energie in Form von Wärme und der vorherrschenden Temperatur in einem Gegenstand bzw. Raum. Als Beispiel stellen wir uns einen Backofen vor, in dem an einem kalten Sonntagmorgen Backwaren erhitzt werden sollen. Der dreidimensionale Raum des Ofens unterliegt nach dem Anschalten einer stetigen Energiezufuhr und es stellt sich eine typische Wärmeverteilung ein. Es ist allerdings nicht das Ziel an jedem Ort des Ofens die gleiche Temperatur zu erreichen, sondern diese nach den Vorstellungen des Anwenders zu regeln. Dazu betrachten wir die Energiezufuhr für jeden Punkt justierbar und untersuchen die Abhängigkeit von Temperatur und zugeführter Energie. Als Modell für den Zusammenhang zwischen beiden Größen gilt die parabolische Differentialgleichung

$$\frac{\partial z(t,x)}{\partial t} - \Delta_x z(t,x) = u(t,x) + e(t,x).$$

Die Funktion e soll dabei den „Abfluss“ von Energie, z.b. durch das Fenster im Ofen, darstellen. Als anderes Anwendungsgebiet verweisen wir auf die Suche nach einer optimalen Heizstrategie für einen Raum, z.b. dem Wohnzimmer, oder der gesamten Wohnung mit Interaktionen zwischen den einzelnen Zimmern. Lassen wir die zeitliche Ände-

rung der Temperatur außen vor, so sprechen wir von *stationärer* Wärmeverteilung und die Gleichung reduziert sich auf

$$-\Delta z(x) = u(x) + e(x).$$

Die Ortsvariable x stellt im Beispiel des Ofens einen dreidimensionalen Vektor dar. Wir vereinfachen dieses Modell und betrachten einen Stab der Länge $T = 1$, den wir einer gesteuerten Energiezufuhr aussetzen und einen spontanen Energieverlust annehmen. An den beiden Stabenden ist die Temperatur fest auf einem konstanten Niveau vorgeschrieben und die Temperatur soll an jedem inneren Punkt des Stabs einen gewünschten Wert annehmen. Die dabei eingesetzte Energie unterliegt hinsichtlich ihrer Größe Einschränkungen und es ist das Ziel, so wenig wie möglich davon zu verbrauchen. Kombinieren wir beide Forderungen, so suchen wir ein Minimum des Funktionals

$$J(z,u) = \frac{1}{2}\int_0^1 |z(t) - z_d(t)|^2 + \nu|u(t)|^2\,dt,$$

wobei $\nu > 0$ als reeller Parameter die Wichtigkeit des Energieverbrauchs regelt. Der Faktor $\frac{1}{2}$ spielt nur eine formale Rolle und besitzt keinen Einfluss auf die Lösung. Der erste Summand misst die Abweichung von der vorgegebenen Funktion z_d und der zweite den Einsatz von Energie über den gesamten Stab der Länge 1. Wir verwenden von nun an die übliche Schreibweise mit t als unabhängiger Variable. Die Lösung $(\bar{z}, \bar{u})$ gibt dann in Abhängigkeit der Parameter z_d, e und ν den optimalen Wärmeeintrag auf den Stab und die dazu eingesetzte Energie an.

Bei der Festlegung der Parameterfunktionen beachten wir das vorgegebene Ziel. So soll die Temperatur im Zentrum des Stabes wesentlich höher sein, als an den Enden. Daher wählen wir für z_d eine Glockenkurve mit dem Maximum im Mittelpunkt

$$z_d(t) = 5e^{\frac{1}{4}}(e^{(t-t^2)} - 1).$$

Den äußeren Einfluss modellieren wir mit der Funktion $e(t) = -\frac{1}{4}\sin(\pi t)$. Zusammen mit der Beschränkung $|u(t)| \leq 1$ stellt sich uns das Steuerungsproblem in der Form

$$\text{(StP)} \qquad \min_u \frac{1}{2}\int_0^1 |z(t) - z_d(t)|^2 + \nu|u(t)|^2\,dt, \qquad u \in U^{ad},$$

mit den Nebenbedingungen

$$\begin{aligned} -\ddot{z}(t) &= u(t) + e(t) \qquad \forall' t \in [0,1] \\ z(0) &= z(1) = 0 \end{aligned}$$

und

$$U^{ad} = \Big\{u \in L_2[0,1,\mathbb{R}] : -1 \leq u(t) \leq 1,\ \forall' t \in [0,1]\Big\}.$$

Zur Veranschaulichung seien die beiden Funktionen in den Abbildungen 1.1 und 1.2 dargestellt.

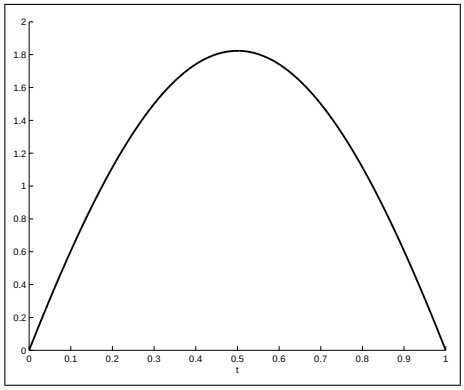

Abbildung 1.1: Gewünschter Zustand $z_d(t) = 5e^{\frac{1}{4}}(e^{(t-t^2)} - 1)$

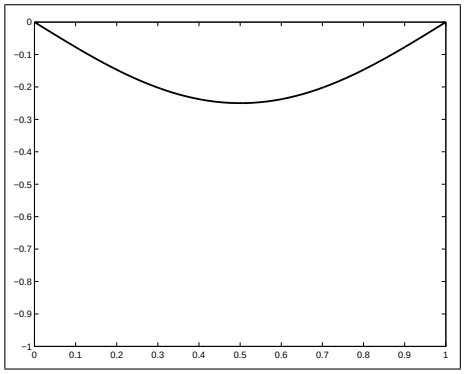

Abbildung 1.2: Wärmeaustrag $e(t) = \frac{1}{4}\sin(\pi t)$

1.2 Funktionenräume

Für die zu lösenden Probleme benötigen wir verschiedenste Aspekte aus der Analysis, speziell der linearen Analysis. So sei $[0, T]$ ein abgeschlossenes und beschränktes Intervall. Die Menge aller auf $[0, T]$ stetigen Funktionen mit reellem Wertebereich bezeichnen wir mit $C[0, T]$. Zusammen mit der Norm $\|x\|_\infty = \sup_{t\in[0,T]} |x(t)|$ ist $C[0, T]$ ein vollständiger, normierter Raum, d.h. ein Banach-Raum. Weiterhin bezeichnen wir mit $C^k[0, T]$, $0 \leq k < \infty$, den Raum aller Funktionen, deren k-te klassische Ableitung stetig auf $[0, T]$ ist. Das bedeutet, die Ableitungen existieren auf dem offenen Intervall $(0, T)$ und sind stetig bis in die Randpunkte, also auf ganz $[0, T]$, fortsetzbar. Diese Bezeichnung gilt für alle $k \in \mathbb{N}$ und wir definieren die Menge aller beliebig oft differenzierbaren Funktionen durch

$$C^\infty[0, T] = \bigcap_{k\in\mathbb{N}} C^k[0, T].$$

Es gilt folgende Inklusionskette für $k \leq l$:

$$C^\infty[0, T] \subset C^l[0, T] \subset C^k[0, T] \subset C[0, T].$$

An dieser Stelle erscheint in Übereinstimmung mit der in der Literatur gebräuchlichen Schreibweise ein weiterer Funktionenraum. So umfasst der Raum $C_0^\infty[0,T]$ alle Funktionen $f : [0,T] \to \mathbb{R}$ mit $f \in C^\infty[0,T]$ und verschwindenden Randwerten. Für mehrdimensionales Grundgebiet gestaltet sich die Definition des Raumes C_0^∞ komplizierter. In unserem Fall mit einem abgeschlossenen und beschränkten Intervall als Definitionsbereich sind beide Definitionsmöglichkeiten äquivalent.

In den L_p-Normen $\|x\|_p = \left(\int_0^T |x(t)|^p \, dt\right)^{\frac{1}{p}}$ ist $C[0,T]$ nicht vollständig. Wir definieren aber mit Hilfe dieser Normen einen größeren Funktionenraum. So bezeichnen wir mit $\mathscr{L}_p[0,T]$ die Menge aller Funktionen deren L_p-Norm endlich ist. Diese Menge zusammen mit der entsprechenden Norm bildet noch keinen normierten Raum, da aus einer verschwindenden Norm einer Funktion noch nicht auf das Nullelement geschlossen werden kann; $\|\cdot\|_p$ ist auf $\mathscr{L}_p[0,T]$ demnach eine Semi-Norm. Erst durch Identifizierung aller Funktionen, die sich nur auf einer Nullmenge bezüglich des Lebesgue-Maßes unterscheiden, entsteht ein normierter Raum von Äquivalenzklassen, bezeichnet mit $L_p[0,T]$. Falls wir für $u, v \in L_p[0,T]$ schreiben $u = v$, so meinen wir $u \sim v$, d.h. u und v sind in der gleichen Äquivalenzklasse, was gleichbedeutend ist zu $u(t) = v(t)$ für fast alle $t \in [0,T]$. Wir sprechen dennoch von Funktionen und meinen damit einen Repräsentanten dieser Äquivalenzklasse. Falls sich in dieser eine stetige Funktion befindet, so wählen wir dieses Element als Repräsentant.

Setzen wir speziell $p = 2$, so definieren wir auf dem Raum $L_2[0,T]$ das Skalarprodukt

$$\langle x, y\rangle_{L_2[0,T]} = \int_0^T x(t)y(t)\, dt, \tag{1.1}$$

welches die Norm $\|\cdot\|_2$ erzeugt. Damit wird der $L_2[0,T]$ zum Hilbert-Raum. Weitere wichtige Funktionenräume erhalten wir durch die Betrachtung von schwachen Ableitungen. Dazu benutzen wir die Schreibweise $\mathscr{D}^j f$ für die j-te klassische Ableitung der Funktion f. Falls f nur von einer Zeitvariablen abhängig ist, verwenden wir auch die Schreibweise $\dot{f}$, $\ddot{f}$ usw. Erfüllt eine Funktion $v \in L_2[0,T]$

$$\int_0^T f(t)\mathscr{D}^1\varphi(t)\, dt = -\int_0^T v(t)\varphi(t)\, dt \qquad \forall \varphi \in C_0^\infty[0,T],$$

so bezeichnen wir sie als schwache Ableitung von f und verwenden dafür die Symbole $\mathscr{D}^1 f$ bzw. $\dot{f}$ synonym. Das stellt keinen Widerspruch zu vorherigen Definition dar. Wenn eine Funktion im klassischen Sinn differenzierbar ist, dann existiert die schwache Ableitung und beide stimmen überein. Für höhere Ableitungen iterieren wir die Definition und schreiben $\mathscr{D}^j$ für $j \in \mathbb{N}$. Darauf aufbauend bezeichnen wir mit $W_p^k[0,T]$ den Raum der absolut stetigen Funktionen $f : [0,T] \to \mathbb{R}$ mit $\mathscr{D}^j f \in L_p[0,T]$ für jede natürliche Zahl $j \le k$. Definieren wir auf $W_p^k[0,T]$ die Norm

$$\|x\|_{W_p^k[0,T]} = \left(\sum_{j=0}^{k} \|\mathscr{D}^j x\|^p_{L_p[0,T]}\right)^{\frac{1}{p}},$$

so sind die Funktionenräume vollständig, also Banach-Räume. Für $p = 2$ führen wir darüber hinaus das Skalarprodukt

$$\langle x, y\rangle_{W_2^k[0,T]} = \sum_{j=0}^{k} \langle \mathscr{D}^j x, \mathscr{D}^j y\rangle_{L_2[0,T]} = \sum_{j=0}^{k} \int_0^T \mathscr{D}^j x(t) \mathscr{D}^j y(t)\, dt$$

ein und erhalten so einen Hilbert-Raum. Die Norm, die durch dieses Skalarprodukt induziert wird, lautet

$$\|x\|_{W_2^k[0,T]} = \left(\sum_{j=0}^{k} \|\mathscr{D}^j x\|^2_{L_2[0,T]} \right)^{\frac{1}{2}} = \left(\sum_{j=0}^{k} \int_0^T |\mathscr{D}^j x(t)|^2\, dt \right)^{\frac{1}{2}}.$$

Nehmen wir nur die höchsten Ableitungen, so erhalten wir eine Semi-Norm, welche definiert ist durch

$$|x|_{W_2^k[0,T]} = \|\mathscr{D}^k x\|_{L_2[0,T]} = \left(\int_0^T |\mathscr{D}^k x(t)|^2\, dt \right)^{\frac{1}{2}}.$$

Die Funktion $|\cdot|_{W_2^k[0,T]}$ ist keine Norm, da man aus $|x|_{W_2^k[0,T]} = 0$ nicht auf $x = 0$ schließen kann.

Der Fall $p = \infty$ spielt bei den eben eingeführten Funktionenräumen eine Sonderrolle. So ist $L_\infty[0,T]$ der Raum aller Funktionen, die bis auf eine Nullmenge beschränkt sind, d.h. für die gilt

$$\|f\|_\infty = \operatorname*{ess\,sup}_{t\in[0,T]} |f(t)| < \infty.$$

Weiterhin bezeichnen wir mit $W_\infty^k[0,T]$ den Raum aller absolut stetigen Funktionen, für die jede schwache Ableitung bis zur Ordnung k ein Element von $L_\infty[0,T]$ ist.

Es ist offensichtlich, dass für $0 \le k \le l$ und $1 \le p \le \infty$ gilt

$$C^l[0,T] \subset W_p^k[0,T].$$

Die Inklusionen $W_p^k[0,T] \subset C^l[0,T]$ gelten unter bestimmten Voraussetzungen an k und l in Abhängigkeit von $1 \le p < \infty$, und zwar müssen sie die Ungleichung

$$k > l + \frac{1}{p}$$

erfüllen (vgl. Adams [1] (1975)). Die einfache Gestalt der Bedingung liegt an der Dimension $n = 1$ der Grundmenge $[0,T]$. Bei höher dimensionalen Gebieten nehmen sie entsprechend komplexere Form an. Sie sind dann so zu verstehen, dass jede Äquivalenzklasse von $W_p^k[0,T]$ eine l-mal stetig differenzierbare Funktion enthält. Es gilt sogar

genauer die Stetigkeit der Abbildung $Id : W_p^k[0,T] \to C^l[0,T]$ mit $Id\, f = f$ und somit folgende Abschätzung mit einer von f unabhängigen Konstante c

$$\|f\|_{C^l[0,T]} \leq c\, \|f\|_{W_p^k[0,T]}.$$

Für eine exakte Formulierung des zu Grunde liegenden Resultats, dem Sobolew'schen Einbettungssatz, verweisen wir auf den Anhang, Satz A.3.2.

Fügen wir zu den Sobolew-Räumen $W_p^k[0,T]$ für $k \geq 1$ noch das Subscript „0" hinzu, so meinen wir damit die Menge der Funktionen aus $W_p^k[0,T]$, die an den Stellen 0 und T, verschwinden, also

$$W_{p,0}^k[0,T] = \{f \in W_p^k[0,T] : f(0) = f(T) = 0\}.$$

Die Forderung $k \geq 1$ berechtigt zu dieser Schreibweise, denn wir haben bereits gesehen, dass jede Funktion aus $W_p^k[0,T]$ für $p > 1$ mindestens stetig auf $[0,T]$ ist. Als interessante Eigenschaft des Raumes $W_{p,0}^2[0,T]$ möchten wir an dieser Stelle die Normierung durch die (sonst nur als Semi-Norm verwendbare) Funktion $|\cdot|_{W_p^2[0,T]}$, denn durch die vorgegebenen Werte auf dem Rand des Definitionsgebietes wird der Schluss $|f|_{W_{p,0}^2[0,T]} = 0 \Longrightarrow f(t) = 0$ für fast alle $t \in [0,T]$ gesichert. Im Fall $k = 1$ würde sogar die Vorgabe nur eines Randwertes ausreichen.

Da der Raum $L_p[0,T]$ der Spezialfall der $W_p^k[0,T]$ mit der geringsten Einschränkung $k = 0$ ist, versteht sich die Inklusion $W_p^k[0,T] \subset L_p[0,T]$ für $k \in \mathbb{N}$ von selbst. Betrachten wir nun den Raum $W_p^k[0,T]$ zusammen mit der Norm $\|\cdot\|_{L_p[0,T]}$, so stellen wir zunächst fest, dass er in dieser Norm nicht vollständig ist. Weiterhin wissen wir um die Dichtheitseigenschaft des Raumes $C^\infty[0,T]$ in $L_p[0,T]$ in dessen Norm, d.h. jede Funktion aus $L_p[0,T]$ kann in der L_p-Norm durch eine Folge von beliebig oft differenzierbaren Funktionen approximiert werden. Da $C^\infty[0,T]$ auch in allen $W_p^k[0,T]$, $k \in \mathbb{N}$, enthalten ist, liegen diese Räume dicht in $W_p^l[0,T]$ mit $l \leq k$ und somit auch in $L_p[0,T]$.

Die vorgestellten Funktionenräume besitzen alle die Eigenschaft der unendlichen Dimension. Allerdings enthalten sie endlich-dimensionale Unterräume. Eine zentrale Rolle spielen dabei die Räume der Polynome bis zu einem bestimmten Grad m. Sie werden mit $\mathcal{P}_m$ bezeichnet, also ist

$$\mathcal{P}_m[0,T] = \Big\{ f \in C[0,T] : f(t) = \sum_{k=0}^{m} a_k t^k,\ a_k \in \mathbb{R} \Big\}.$$

Für spätere Anwendungen ist es notwendig, vektorwertige Funktionen einzuführen. So nennen wir $L_p[0,T,\mathbb{R}^n]$ den Raum aller Äquivalenzklassen von Funktionen $x : [0,T] \to \mathbb{R}^n, t \mapsto (x_1(t), \ldots, x_n(t))$, deren L_p-Norm durch

$$\|x\|_p = \left(\int_0^T |x(t)|_2^p \, dt \right)^{\frac{1}{p}}$$

definiert ist. In dieser Definition bezeichnet $|\cdot|_2$ die Euklidische Norm eines Vektors im $\mathbb{R}^n$. Falls keine Möglichkeit zur Verwechslung besteht, werden wir den Index weglassen und alternativ bei $|\cdot|$ auch vom Betrag eines Vektors sprechen. Die Bezeichnung $W_p^k[0,T,\mathbb{R}^n]$ ist analog zum eindimensionalen Fall aufzufassen. Das Skalarprodukt für den Fall $p=2$ lautet dann

$$\langle x,y\rangle_{W_2^k[0,T,\mathbb{R}^n]} = \sum_{j=0}^{k} \langle \mathscr{D}^j x, \mathscr{D}^j y\rangle_{L_2[0,T,\mathbb{R}^n]} = \sum_{j=0}^{k} \int_0^T (\mathscr{D}^j x)(t)^\mathsf{T} (\mathscr{D}^j y)(t)\,dt.$$

Für $k=0$ erhalten wir als Spezialfall das Skalarprodukt für den $L_2[0,T,\mathbb{R}^n]$.

Mit $C^k[0,T,\mathbb{R}^n]$ bezeichnen wir den Raum aller Funktionen $x:[0,T]\to\mathbb{R}^n$, $t\mapsto (x_1(t),\ldots,x_n(t))$, die in jeder Komponente k-mal stetig differenzierbar sind. Mit der Norm

$$\|x\|_{C^k} = \max_{0\le j\le k} \|x^{(j)}\|_{C^k} = \max_{0\le j\le k} \max_{t\in[0,T]} |x^{(j)}(t)|$$

wird auch dieser Raum zu einem Banach-Raum. Ebenfalls analog dem eindimensionalen Fall führen wir den Raum der Polynome bis zum Grad m ein. Es ist

$$\mathcal{P}_m[0,T,\mathbb{R}^n] = \Big\{ f\in C[0,T,\mathbb{R}^n] : f(t) = \sum_{k=0}^{m} a_k t^k,\ a_k\in\mathbb{R}^n \Big\}. \tag{1.2}$$

Da wir Normen und Semi-Normen sowie die eventuell zugehörigen Skalarprodukte sehr häufig benutzen werden, ist es angebracht eine Kurzschreibweise einführen. Wir setzen analog zu $\|\cdot\|_{L_p} = \|\cdot\|_p$ auch für $\|\cdot\|_{W_p^k}$ kurz $\|\cdot\|_{k,p}$ und für $|\cdot|_{W_p^k}$ kurz $|\cdot|_{k,p}$. Bei den Skalarprodukten sparen wir uns im Standardfall L_2 den Index. Die Dimension der zu Grunde liegenden Räume lassen wir bei den Kurzbezeichnungen solange keine Möglichkeit zur Verwechslung besteht außen vor.

Für zwei gegebene lineare Räume X und Y betrachten wir alle linearen und beschränkten Abbildungen mit dem Definitionsbereich X und dem Wertebereich Y und bezeichnen diese Menge mit $\mathscr{L}(X,Y)$. Jedes Element $\ell\in\mathscr{L}(X,Y)$ erfüllt demnach die folgenden Voraussetzungen

$$\begin{aligned} &\ell(\lambda x_1+\mu x_2) = \lambda\ell(x_1)+\mu\ell(x_2) \qquad \forall x_1,x_2\in X, \forall\lambda,\mu\in\mathbb{R} \\ &\sup_{v\in X\setminus\{0\}} \frac{\|\ell(v)\|_Y}{\|v\|_X} \le c < \infty \end{aligned}$$

Mit den Verknüpfungen $(\ell_1+\ell_2)(x) = \ell_1(x)+\ell_2(x)$ und $\alpha\ell_1(x) = \ell_1(\alpha x)$ für $\ell_1,\ell_2\in\mathscr{L}(X,Y)$ und $\alpha\in\mathbb{R}$ wird $\mathscr{L}(X,Y)$ ein linearer Raum. Man kann zeigen, dass die Vollständigkeit von Y die Vollständigkeit von $\mathscr{L}(X,Y)$ impliziert (vgl. Werner [24] (2004)), daher schließen wir für $\mathscr{L}(X,\mathbb{R})$ auf die Eigenschaft eines Banach-Raumes. Diesen Spezialfall nennen wir den Raum der linearen beschränkten Funktionale oder auch *Dualraum*. Wir führen für ihn die abkürzende Schreibweise X' ein.

Ein wichtiges Resultat für die Relation eines Hilbert-Raumes und seines Dualraumes ist der Darstellungssatz von Fréchet und Riesz (siehe Anhang, Satz A.3.5). Weiterhin folgt aus der Definition des Dualraumes folgende Implikation für die Räume X, Y

$$X \subset Y \Longrightarrow Y' \subset X',$$

d.h. betrachten wir für ein Funktional $\ell \in Y'$ die Einschränkung auf dem Raum Y, so gehört $\ell|_X$ zum Dualraum von X.

1.3 Problemstellung

Für die Einführung des zu betrachtenden Problems definieren wir zunächst die notwendigen Bezeichnungen. So seien $z \in W_2^2[0,T,\mathbb{R}^n]$, $u \in L_2[0,T,\mathbb{R}^m]$ und $z_d, e \in W_\infty^2[0,T,\mathbb{R}^n]$. Die reelle Matrix A mit der Dimension $n \times n$ sei symmetrisch. B sei ebenfalls eine reelle Matrix mit der Dimension $n \times m$. Weiterhin seien $a, b \in \mathbb{R}^m$ mit $a_k < b_k$ für $k = 1, \ldots, n$ und $\nu \in \mathbb{R}$ mit $\nu > 0$.

All jene Parameter erhalten nun einen Platz im Steuerungsproblem $(\widetilde{\mathrm{StP}})$, welches wir als Ausgangspunkt für die weiteren Untersuchungen verwenden. Wir betrachten

$$(\widetilde{\mathrm{StP}}) \qquad \min_{z,u} \widetilde{J}(z,u) = \min_{z,u} \frac{1}{2}\|z - z_d\|_2^2 + \frac{\nu}{2}\|u\|_2^2$$

unter den Nebenbedingungen

$$\begin{aligned} -\ddot{z}(t) + Az(t) &= Bu(t) + e(t) \qquad \forall' t \in [0,T] \\ z(0) &= z(T) = 0_n \end{aligned} \tag{1.3}$$

und den Kontrollrestriktionen

$$a \le u(t) \le b \qquad \forall' t \in [0,T].$$

Bemerkung: Die Voraussetzungen an A, e und z_d sind aus formalen Gründen so strikt. Die Symmetrie von A benötigen wir später für den Nachweis der Existenz und Eindeutigkeit der Lösung von (1.3). Für die theoretischen Betrachtungen zum Problem $(\widetilde{\mathrm{StP}})$ hätten auch die Forderungen $z_d, e \in L_2[0,T,\mathbb{R}^n]$ ausgereicht. Allerdings benötigen wir in den Kapiteln 4 und 5 für die Berechnung einer Näherungslösung durch Diskretisierung des Problems $(\widetilde{\mathrm{StP}})$ die schärferen Voraussetzungen. $\Diamond$

Mit Hilfe der Bezeichnung $y = Bu + e \in L_2[0,T,\mathbb{R}^n]$ ist (1.3) äquivalent zu

$$-\ddot{z}(t) + Az(t) = y(t) \qquad \forall' \in [0,T], \qquad z(0) = z(T) = 0_n. \tag{1.4}$$

Ab sofort sprechen wir synonym von der *Systemgleichung* und meinen die auf den jeweiligen Kontext bezogene Gleichung (1.3) oder (1.4).

Wir definieren weiterhin folgende Menge der zulässigen Funktionen

$$U^{ad} = \{u \in L_2[0,T,\mathbb{R}^m] : a \leq u(t) \leq b \quad \forall' t \in [0,T]\},$$

und den Projektionsoperator $\Pi_{[a,b]} : C[0,T,\mathbb{R}^m] \to C[0,T,\mathbb{R}^m]$ durch

$$\Pi_{[a,b]}(f(x)) = \max\{a, \min\{b, f(x)\}\}, \qquad f \in L_2[0,T,\mathbb{R}^n],$$

wobei sowohl die Relationen als auch die Funktionen „max“ und „min“ komponentenweise anzuwenden sind.

Bemerkung: Wir sind bereits an dieser Stelle in der Lage weitere Aussagen über die Lösungen $(\bar{z}, \bar{u})$ von $(\widetilde{\text{StP}})$ zu treffen. Auf Grund der Kontrollrestriktionen ist $\bar{u}$ wesentlich beschränkt, also $\bar{u} \in L_\infty[0,T,\mathbb{R}^m]$, woraus folgt, dass auch $\bar{z}$ eine beschränkte zweite Ableitung besitzt, also $\bar{z} \in W^2_\infty[0,T,\mathbb{R}^n]$. Später werden wir an Hand der Optimalitätsbedingungen für (StP) weitere Eigenschaften der Lösung herleiten. $\Diamond$

Kapitel 2

Theoretische Grundlagen

2.1 Die Variationsgleichung der Systemgleichung

Für eine effiziente Lösung des Steuerungsproblems ($\widetilde{\text{StP}}$) untersuchen wir, ob eine Darstellung des Zustands z als Funktion der Steuerung u existiert. Das ist aber nur dann sinnvoll, falls die Systemgleichung (1.3) für jedes $u \in L_2[0,T,\mathbb{R}^m]$ bzw. falls die Variationsgleichung (1.4) für jedes $y \in L_2[0,T,\mathbb{R}^n]$ eine eindeutig bestimmte Lösung $z \in W^2_{2,0}[0,T,\mathbb{R}^n]$ besitzt.

Im ersten Schritt leiten wir eine andere Formulierung der Systemgleichung (1.4) her und zeigen die Äquivalenz der Lösungen. Wir sind dabei an einer Lösung im schwachen Sinne interessiert, d.h. es muss gelten

$$\int_0^T (-\ddot{z}(t) + Az(t))^\mathsf{T} v(t)\,dt = \int_0^T y(t)^\mathsf{T} v(t)\,dt \qquad \forall v \in W^1_{2,0}[0,T,\mathbb{R}^n].$$

Diese Identität schreiben wir mit Hilfe partieller Integration um zu

$$\int_0^T \dot{z}(t)^\mathsf{T}\dot{v}(t) + z(t)^\mathsf{T} Av(t)\,dt = \int_0^T y(t)^\mathsf{T} v(t)\,dt \qquad \forall v \in W^1_{2,0}[0,T,\mathbb{R}^n]. \tag{2.1}$$

Die Gleichung (2.1) nennen wir das zu (1.4) gehörige *Variationsproblem* oder die *Variationsgleichung*.

Der Grund für die Einführung der schwachen Formulierung von (1.4) ist der leichtere Zugang zu einer Existenz- und Eindeutigkeitsaussage einer Lösung von (2.1). Der Nutzen für unsere Problemstellung erschließt sich uns durch Lemma 2.1.2, denn die Lösungen beider Gleichungen sind äquivalent. Dazu benötigen wir zunächst noch ein Hilfsresultat, bevor wir den Beweis in Angriff nehmen.

Lemma 2.1.1 (Variationslemma). *Seien $x, y \in L_2[0,T,\mathbb{R}^n]$ und es gelte*

$$\int_0^T x(t)^\mathsf{T} v(t) + y(t)^\mathsf{T}\dot{v}(t)\,dt = 0 \qquad \forall v \in W^1_{2,0}[0,T,\mathbb{R}^n].$$

Dann ist $y \in W_2^1[0,T,\mathbb{R}^n]$ *und* $\dot{y}(t) = x(t)$ *für fast alle* $t \in [0,T]$.

Beweis: siehe Alt [2] (Lemma 5.2.2). □

Lemma 2.1.2. *Sei* $y \in L_2[0,T,\mathbb{R}^n]$. *Dann ist* $z \in W_{2,0}^2[0,T,\mathbb{R}^n]$ *Lösung der Systemgleichung* (1.4) *genau dann, wenn* z *auch Lösung der Variationsgleichung* (2.1) *ist.*

Beweis: Sei $z \in W_{2,0}^2[0,T,\mathbb{R}^n]$ zunächst Lösung der Systemgleichung (1.4). Dann ist z trivialerweise auch schwache Lösung, erfüllt also (2.1).

Für die Rückrichtung schreiben wir zunächst die Variationsgleichung in der Form

$$\int_0^T \dot{z}(t)^\mathsf{T}\dot{v}(t) + z(t)^\mathsf{T}Av(t) - y(t)^\mathsf{T}v(t)\,dt = \int_0^T \big(Az(t) - y(t)\big)^\mathsf{T}v(t) + \dot{z}(t)^\mathsf{T}\dot{v}(t)\,dt = 0 \qquad \forall v \in W_{2,0}^1[0,T,\mathbb{R}^n]$$

und benutzen anschließend das Variationslemma (Lemma 2.1.1). Demnach ist $\dot{z} \in W_2^1[0,T,\mathbb{R}^n]$ und

$$\ddot{z}(t) = Az(t) - y(t) \Longleftrightarrow -\ddot{z}(t) + Az(t) = y(t) \qquad \forall' t \in [0,T].$$

Daher gilt für eine Lösung $z \in W_{2,0}^1[0,T,\mathbb{R}^n]$ von (2.1) einerseits die strengere Glattheitseigenschaft $z \in W_{2,0}^2[0,T,\mathbb{R}^n]$ und darüber hinaus löst es auch die ursprüngliche Systemgleichung. □

Die Nebenbedingung im Steuerungsproblem ($\widetilde{\text{StP}}$) können wir demnach durch ihre Variationsgleichung ersetzen. Die Lösung des Problems wird auf Grund der Aussage von Lemma 2.1.2 nicht beeinflusst und somit verzichten wir im Folgenden auf eine Unterscheidung der Problemformulierungen.

2.2 Eindeutigkeit der Lösung der Systemgleichung

In diesem Abschnitt ist es unser Ziel, die eindeutige Lösbarkeit der Variationsgleichung (2.1) für jedes $y \in L_2[0,T,\mathbb{R}^n]$ nachzuweisen. Dafür definieren wir die Bilinearform $a(\cdot,\cdot) : W_{2,0}^1[0,T,\mathbb{R}^n] \times W_{2,0}^1[0,T,\mathbb{R}^n] \to \mathbb{R}$ durch

$$a(z,v) = \int_0^T \dot{z}(t)^\mathsf{T}\dot{v}(t) + z(t)^\mathsf{T}Av(t)\,dt \tag{2.2}$$

und zeigen zunächst Eigenschaften dieser Funktion. Die Symmetrie ist äquivalent zur Symmetrie der Matrix A, welche wir bei der Problemstellung voraus gesetzt hatten. Für den Nachweis weiterer Eigenschaften benötigen wir noch ein Hilfsresultat.

Lemma 2.2.1 (Poincaré'sche Ungleichung). *Sei $z \in W^1_{2,0}[0,T,\mathbb{R}^n]$. Dann gilt*

$$\|z\|_2 \leq T\|\dot{z}\|_2. \tag{2.3}$$

Beweis: Da z aus dem Raum $W^1_2[0,T,\mathbb{R}^n]$ stammt, ist speziell $z(0) = 0$. Daher gilt unter Verwendung der Hölder'schen Ungleichung aus Lemma A.3.1

$$\|z\|_2^2 = \int_0^T \Big| \int_0^t \dot{z}(s)\, ds\Big|^2 dt \leq \int_0^T \left(\int_0^T |\dot{z}(s)|\, ds\right)^2 dt \overset{(A.2)}{\leq} \int_0^T T\|\dot{z}\|_2^2\, dt = T^2\|\dot{z}\|_2^2.$$

□

Korollar 2.2.2. *Auf dem Raum $W^1_{2,0}[0,T,\mathbb{R}^n]$ sind die Normen $\|z\|_{1,2}$ und $\|\dot{z}\|_2$ äquivalent, d.h.*

$$\|\dot{z}\|_2 \leq \|z\|_{1,2} \leq \sqrt{2}\max\{1,T\}\,\|\dot{z}\|_2 \qquad \forall z \in W^1_{2,0}[0,T,\mathbb{R}^n].$$

Beweis: Die erste Ungleichung ist per Definition erfüllt. Für die Abschätzung nach oben benutzen wir die Poincaré'sche Ungleichung

$$\|z\|_{1,2} \leq \sqrt{T^2\|\dot{z}\|_2^2 + \|\dot{z}\|_2^2} \leq \sqrt{2}\max\{1,T\}\,\|\dot{z}\|_2.$$

□

Definition 2.2.3 (Elliptizität). Sei $[H, \|\cdot\|_H]$ ein normierter Raum. Eine Bilinearform $Q : H \times H \to \mathbb{R}$ heißt *elliptisch auf H* bzw. *H-elliptisch*, falls

$$Q(v,v) \geq c\,\|v\|_H^2, \qquad \forall v \in H$$

mit einer von v unabhängigen Konstante $c > 0$ gilt. ◊

Lemma 2.2.4. *Falls die Matrix A positiv semi-definit ist, so ist die Bilinearform $a(\cdot,\cdot)$ auf $W^1_{2,0}[0,T,\mathbb{R}^n]$ elliptisch und beschränkt, d.h. es gilt*

$$\frac{1}{c^u_{2.4}}\|z\|_{1,2}^2 \leq a(z,z) \leq c^o_{2.4}\|z\|_{1,2}^2 \qquad \forall z \in W^1_{2,0}[0,T,\mathbb{R}^n] \tag{2.4}$$

mit $c^u_{2.4} = 2\max\{1,T^2\}$ und $c^o_{2.4} = \max\{1,\|A\|\}$.

Beweis: Seien $z, v \in W^1_{2,0}[0,T,\mathbb{R}^n]$. Die Beschränktheit leiten wir wie folgt her:

$$\begin{aligned}
a(z,v) &= \langle \dot{z}, \dot{v}\rangle + \langle Az, v\rangle \\
&\leq \|\dot{z}\|_2\|\dot{v}\|_2 + \|A\|\|z\|_2\|v\|_2 \\
&\leq \max\{1,\|A\|\}\,(\|\dot{z}\|_2\|\dot{v}\|_2 + \|z\|_2\|v\|_2) \\
&\overset{(A.1)}{\leq} c^o_{2.4}\Big(\|\dot{z}\|_2^2 + \|z\|_2^2\Big)^{\frac{1}{2}}\Big(\|\dot{v}\|_2^2 + \|v\|_2^2\Big)^{\frac{1}{2}} \\
&= c^o_{2.4}\,\|z\|_{1,2}\|v\|_{1,2}
\end{aligned}$$

Daraus folgt auch sofort die rechte Ungleichung der Aussage. Für die Elliptizitätsaussage schließen wir mit positiv semi-definiten Matrix A

$$\begin{aligned} a(z,z) &= \|\dot{z}\|_2^2 + \underbrace{\langle Az, z\rangle}_{\geq 0} \geq \frac{1}{2}(\|\dot{z}\|_2^2 + \|\dot{z}\|_2^2) \\ &\overset{(2.3)}{\geq} \frac{1}{2}(\|\dot{z}\|_2^2 + \frac{1}{T^2}\|z\|_2^2) \geq \frac{1}{2}\min\{1, \frac{1}{T^2}\}\,\|z\|_{1,2}^2. \end{aligned}$$

Mit $\left(\frac{1}{2}\min\{1, \frac{1}{T^2}\}\right)^{-1} = 2\max\{1, T^2\}$ folgt die Aussage. □

Die Bilinearform $a(\cdot,\cdot)$ definiert demnach wegen der Symmetrie von A ein Skalarprodukt auf dem Raum $W_{2,0}^1[0,T,\mathbb{R}^n]$ und induziert die Norm $\|\cdot\|_a = \sqrt{a(\cdot,\cdot)}$. Der Raum $W_{2,0}^1[0,T,\mathbb{R}^n]$ ist auf Grund der Beschränktheit von $a(\cdot,\cdot)$ auch mit der Norm $\|\cdot\|_a$ vollständig und damit ein Hilbert-Raum.

Da y aus dem Hilbert-Raum $L_2[0,T,\mathbb{R}^n]$ stammt, existiert nach dem Satz von Fréchet-Riesz (siehe Anhang, Satz A.3.5) ein lineares beschränktes Funktional $\ell : L_2[0,T,\mathbb{R}^n] \to \mathbb{R}$ mit $\ell(v) = \int_0^T y(t)^\mathsf{T} v(t)\,dt = \langle y, v\rangle$. Da ℓ natürlich auch auf dem Teilraum $W_{2,0}^1[0,T,\mathbb{R}^n]$ linear und beschränkt ist, wenden wir wieder den Satz von Fréchet-Riesz an. Es gibt demnach genau ein $z \in W_{2,0}^1[0,T,\mathbb{R}^n]$ mit $l(v) = \langle z, v\rangle_{W_{2,0}^1} = a(z,v)$, da $a(\cdot,\cdot)$ auf diesem Raum ein Skalarprodukt darstellt, mit dem $W_{2,0}^1[0,T,\mathbb{R}^n]$ zum Hilbert-Raum wird. Daher folgern wir mit eben diesem Element z

$$a(z,v) = \ell(v) = \int_0^T y(t)^\mathsf{T} v(t)\,dt \qquad \forall v \in W_{2,0}^1[0,T,\mathbb{R}^n].$$

Nach dem Beweis der Existenz einer eindeutig bestimmten Lösung von (2.1) im Raum $W_{2,0}^1[0,T,\mathbb{R}^n]$ folgt im letzten Schritt die Anwendung von Lemma 2.1.2. Danach ist dieses $z \in W_{2,0}^1[0,T,\mathbb{R}^n]$ ebenfalls Lösung der eigentlichen Systemgleichung (1.4).

Weiterhin setzen wir in (2.1) die spezielle Funktion z für v ein, also

$$\int_0^T \dot{z}(t)^\mathsf{T}\dot{z}(t) + z(t)^\mathsf{T} A z(t)\,dt = \int_0^T y(t)^\mathsf{T} z(t)\,dt.$$

Weil A positiv semi-definit ist, folgern wir mit Hilfe der Cauchy-Schwarz-Ungleichung

$$\|z\|_2^2 \overset{(2.3)}{\leq} T^2\|\dot{z}\|_2^2 \leq T^2\|y\|_2\|z\|_2.$$

Ist z nicht äquivalent zur Nullfunktion, so folgt durch beidseitiges Kürzen

$$\|z\|_2 \leq T^2\|y\|_2.$$

Im anderen Fall ist die Abschätzung trivial.

Diese Überlegungen anschließend formulieren wir die Resultate in einem Satz.

Satz 2.2.5. *Ist die Matrix A symmetrisch und positiv semi-definit, so besitzt die Systemgleichung* (1.4) *für jedes $y \in L_2[0,T,\mathbb{R}^n]$ eine eindeutig bestimmte Lösung $z \in W^2_{2,0}[0,T,\mathbb{R}^n]$ und es gilt*

$$\|z\|_2 \leq T^2 \|y\|_2. \tag{2.5}$$

Beweis: In den Ausführungen vor der Formulierung dieses Satzes nahmen wir den Beweis bereits vorweg. Zusammenfassend sei an dieser Stelle die zweimalige Anwendung des Satzes von Fréchet-Riesz (siehe Anhang, Satz A.3.5) betont. Dies ist möglich, da in der Variationsgleichung (2.1) die Gleichheit für alle $v \in W^1_{2,0}[0,T,\mathbb{R}^n]$ gefordert wird. Weiterhin definiert die linke Seite ein Skalarprodukt und die rechte Seite ein lineares beschränktes Funktional, welches sogar auf dem größeren Raum $L_2[0,T,\mathbb{R}^n]$ definiert ist. Somit schließen wir von der Funktion $y \in L_2[0,T,\mathbb{R}^n]$ auf die Existenz und Eindeutigkeit des Funktionals $\ell \in (W^1_{2,0}[0,T,\mathbb{R}^n])'$ und somit auf die Existenz und Eindeutigkeit der Lösung $z \in W^1_{2,0}[0,T,\mathbb{R}^n]$.

Mit Lemma 2.1.2 folgern wir für diese Funktion z die stärkere Differenzierbarkeitseigenschaft $z \in W^2_{2,0}[0,T,\mathbb{R}^n]$ und vollenden damit den Beweis. □

Bemerkung: Der Beweis von Satz 2.2.5 beruht auf der zweimaligen Anwendung des Darstellungssatzes von Fréchet-Riesz (siehe Anhang, Satz A.3.5). Eine andere Möglichkeit besteht durch Betrachtung des Funktionals

$$I(z) = a(z,z) - \ell(z)$$

mit dem linearen beschränkten Funktional $\ell(z) = \int_0^T y(t)^\mathsf{T} z(t)\,dt$. Suchen wir dessen Minimum im Raum $W^1_{2,0}[0,T,\mathbb{R}^n]$ so ist (1.4) die notwendige Bedingung für die Optimalität eines z. Auf Grund der vorausgesetzten Elliptizität von $a(\cdot,\cdot)$ auf $W^1_{2,0}[0,T,\mathbb{R}^n]$ folgt mit Alt [2] (Satz 2.5.2) die strikte Konvexität von I und somit die Existenz und Eindeutigkeit eines Minimums bzw. einer Lösung von (1.4).

Die Aussage lässt sich auch mit deutlich allgemeineren Voraussetzungen beweisen. So stellt $a(z,\cdot) \;:\; W^1_2[0,T,\mathbb{R}^n] \to \mathbb{R}$ ebenfalls ein lineares beschränktes Funktional dar und die Symmetrie der Bilinearform $a(\cdot,\cdot)$ ist daher nicht notwendig. Weiterhin können wir die Räume $L_2[0,T,\mathbb{R}^n]$ und $W^1_2[0,T,\mathbb{R}^n]$ durch beliebige Hilbert-Räume H und V mit der Einschränkung $V \subset H \subset V'$ ersetzen. Wir verweisen dazu auf das *Lax-Milgram-Lemma* (Lemma A.3.6) im Anhang. ◊

Korollar 2.2.6. *Unter den Voraussetzungen und mit den Bezeichnungen aus Satz 2.2.5 gelten ähnliche Abschätzungen auch in anderen Normen.*

$$\|z\|_{1,2} \leq c_{2.6}\,\|y\|_2 \tag{2.6}$$

$$\|z\|_{2,2} \leq \sqrt{2} c_{2.7}\,\|y\|_2 \tag{2.7}$$

$$\|z\|_\infty \leq T^{\frac{3}{2}} \|y\|_2 \leq T^2 \|y\|_\infty. \tag{2.8}$$

mit den von z und y unabhängigen Konstanten $c_{2.6} = \sqrt{2}\max\{1,T\}$ und $c_{2.7} = c^2_{2.6} c^o_{2.4}$.

Beweis: Mit Hilfe von Lemma 2.2.4 schätzen wir ab

$$\frac{1}{\sqrt{2}} \min\{1, \frac{1}{T}\} \|z\|_{1,2} \leq \|z\|_a = \|\ell\|_{(W^1_{2,0})'}$$
$$= \sup_{v \in W^1_{2,0}} \frac{|\ell(v)|}{\|v\|_{W^1_{2,0}}} = \sup_{v \in W^1_{2,0}} \frac{|\langle y, v\rangle|}{\|v\|_{W^1_{2,0}}} \leq \sup_{v \in W^1_{2,0}} \frac{\|y\|_2 \|v\|_2}{\|v\|_{W^1_{2,0}}} \leq \|y\|_2.$$

Die Suprema beziehen sich dabei auf alle Elemente außer der Nullfunktion und es gilt somit

$$\|z\|_{1,2} \leq \frac{\sqrt{2}}{\min\{1, T^{-1}\}} \|y\|_2 = \sqrt{2} \max\{1, T\} \|y\|_2.$$

Damit ergibt auch die zweite Aussage, denn es gilt

$$\|\ddot{z}\|_2 \leq \|A\| \|z\|_2 + \|y\|_2 \leq (T^2 \|A\| + 1) \|y\|_2 \leq c^2_{2.6} c^o_{2.4} \|y\|_2,$$

und somit auch

$$\|z\|_{2,2} = \sqrt{\|z\|^2_{1,2} + \|\ddot{z}\|^2_2} \leq \sqrt{c^2_{2.6} + (c^2_{2.6} c^o_{2.4})^2} \|y\|_2 \leq \sqrt{2} c^2_{2.6} c^o_{2.4} \|y\|_2.$$

Weiterhin verweisen wir auf den Sobolew'schen Einbettungssatz (siehe Anhang, Satz A.3.2) und schließen auf Grund von $z(0) = 0_n$

$$\|z\|^2_\infty \overset{(A.4)}{\leq} T \|\dot{z}\|^2_2 \leq T \|y\|_2 \|z\|_2,$$

und somit auch auf

$$\|z\|_\infty \leq T^2 \|y\|_\infty,$$

falls z nicht äquivalent zur Nullfunktion ist. In diesem Fall wäre die Abschätzung ohnehin trivial. □

Die eben bewiesene Ungleichung (2.8) verwenden wir im nächsten Schritt für eine stärkere Aussage bezüglich der Relation der Normen von Lösung und rechter Seite der Systemgleichung.

Lemma 2.2.7. *Sei $z \in W^2_{2,0}[0, T, \mathbb{R}^n]$ die Lösung von (1.4) für die rechte Seite $y \in L_2[0, T, \mathbb{R}^n]$ und für $t \in [0, T]$ sei $Y(t) = \int_0^t y(s)\, ds$. Dann gilt*

$$\|z\|_\infty \leq c_{2.7} \|Y\|_1.$$

Beweis: Wir betrachten zunächst für $A = 0$ die Variationsgleichung (2.1) mit deren Lösung $z^0 \in W^1_{2,0}[0, T, \mathbb{R}^n]$, d.h.

$$\int_0^T \dot{z}^0(t)^\mathsf{T} \dot{v}(t)\, dt = \int_0^T y(t)^\mathsf{T} v(t)\, dt = -\int_0^T Y(t)^\mathsf{T} \dot{v}(t)\, dt \quad \forall v \in W^1_{2,0}[0, T, \mathbb{R}^n].$$

Also ist

$$\int_0^T (\dot{z}^0(t) + Y(t))^\mathsf{T} \dot{v}(t)\, dt = 0 \qquad \forall v \in W^1_{2,0}[0, T, \mathbb{R}^n].$$

Mit Alt [2] (2005, Lemma 5.2.1) folgt daraus

$$\dot{z}^0(t) + Y(t) = \text{const.} =: q \qquad \forall' t \in (0, T).$$

Bilden wir das Integral über $[0, T]$ und beachten die Randwertvorgaben $z^0(0) = z^0(T) = 0_n$, so erhalten wir

$$Tq = \int_0^T \dot{z}^0(t) + Y(t)\, dt = \underbrace{z^0(T) - z^0(0)}_{0} + \int_0^T Y(t)\, dt \Longrightarrow q = \frac{1}{T} \int_0^T Y(t)\, dt.$$

Also ist

$$\begin{aligned} z^0(t) &= z^0(0) + \int_0^t \dot{z}^0(s)\, ds = tq - \int_0^t Y(s)\, ds \\ &= \frac{t}{T} \int_0^T Y(s)\, ds - \int_0^t Y(s)\, ds \\ &= (\frac{t}{T} - 1) \int_0^t Y(s)\, ds + \frac{t}{T} \int_t^T Y(s)\, ds. \end{aligned}$$

Gehen wir nun zum Betrag über

$$|z^0(t)| \leq (1 - \frac{t}{T}) \int_0^t |Y(s)|\, ds + \frac{t}{T} \int_t^T |Y(s)|\, ds,$$

so folgt

$$\|z^0\|_\infty \leq \|Y\|_1 = \int_0^T |Y(t)|\, ds.$$

Weiterhin ist $z - z^0$ die Lösung der Randwertaufgabe

$$\begin{aligned} a(z, v) &= -\int_0^T z^0(t)^\mathsf{T} A v(t)\, dt \qquad \forall v \in W^1_{2,0}[0, T, \mathbb{R}^n] \\ z(0) &= z(T) = 0_n. \end{aligned}$$

Damit folgt unter Beachtung von Korollar 2.2.6

$$\begin{aligned} \|z\|_\infty &\leq \|z - z^0\|_\infty + \|z^0\|_\infty \overset{(2.8)}{\leq} T^2 \|A\| \, \|z^0\|_\infty + \|z^0\|_\infty \\ &\leq (1 + T^2 \|A\|) \, \|Y\|_1 \leq c_{2.7} \, \|Y\|_1. \end{aligned}$$

□

Die zuletzt durchgeführten Betrachtungen rechtfertigen die Schreibweise $z = z(u)$. Die Funktion $z(u)$ ist somit das Bild von u unter einem Operator, den wir mit $\mathcal{S} \circ \mathcal{T}$ bezeichnen, also $\mathcal{S} \circ \mathcal{T} : L_2[0,T,\mathbb{R}^m] \to L_2[0,T,\mathbb{R}^n] \to W^2_{2,0}[0,T,\mathbb{R}^n]$. Dabei ist $\mathcal{T}$ die affin-lineare Transformation $B \cdot + e$, welche auf die eindeutige Lösbarkeit der Systemgleichung (1.4) keinen Einfluss besitzt und $\mathcal{S}$ derjenige Operator, der jedem $y \in L_2[0,T,\mathbb{R}^n]$ die eindeutige Lösung der Systemgleichung (1.4) zuweist, also

$$y = \mathcal{T}u = Bu + e, \qquad z = \mathcal{S}y. \tag{2.9}$$

Die Abbildung $\mathcal{S}$ ist auf Grund der Form der Differentialgleichung ebenfalls linear und die Beschränktheit zeigten wir bereits in Korollar 2.2.6. Darüber hinaus stellt der Operator $\mathcal{S}$ auch einen Endomorphismus auf dem $L_\infty[0,T,\mathbb{R}^n]$ dar, denn für jede beschränkte rechte Seite y ist das Bild aus dem Raum $W^2_\infty[0,T,\mathbb{R}^n]$ und damit beschränkt. Das Funktional $\widetilde{J}$ in unserem Optimierungsproblem ($\widetilde{\text{StP}}$) ist demnach allein von der Steuerung u abhängig und das Problem ($\widetilde{\text{StP}}$) verändert sich zu

$$\text{(StP)} \qquad \min_u J(\mathcal{S}\mathcal{T}u, u) = \min_u \frac{1}{2}\|\mathcal{S}\mathcal{T}u - z_d\|_2^2 + \frac{\nu}{2}\|u\|_2^2, \quad u \in U^{ad}.$$

Die Transformation $\mathcal{T}$ aus (2.9) verwenden wir weiterhin zur Verkürzung des Notationsaufwands.

Diese Problemstellung ist weiteren Untersuchungen wesentlich zugänglicher, denn nun ist es möglich, bezüglich der Steuerung als alleiniger Variable zu optimieren. Die Differentialgleichung als Nebenbedingung findet sich implizit im Zielfunktional wieder und es bleiben die Kontrollrestriktionen, denen ausschließlich u unterliegt. Wir erhalten also ein Steuerungsproblem mit Restriktionen, dessen zulässige Menge nichtleer, abgeschlossen und konvex ist.

2.3 Die adjungierte Gleichung

Wie im letzten Abschnitt hergeleitet, beschreiben wir die Zuordnung $y \mapsto z$ mit Hilfe des linearen und beschränkten Operators $\mathcal{S}$. Dieser bildet den $L_2[0,T,\mathbb{R}^n]$ zunächst in den Raum $W^2_{2,0}[0,T,\mathbb{R}^n]$ ab. Jener Raum stellt allerdings einen Unterraum des $L_2[0,T,\mathbb{R}^n]$ dar, daher ist $\mathcal{S}$ ein Endomorphismus. Als Operator auf einem Hilbert-Raum betrachten wir seinen adjungierten Operator $\mathcal{S}^*$. Die beiden stehen über die Forderung

$$\langle \mathcal{S}^* u, w\rangle = \langle u, \mathcal{S}w\rangle \qquad \forall u, w \in L_2[0,T,\mathbb{R}^n]$$

in Verbindung. In unserem Fall ist es möglich, den adjungierten Operator zu bestimmen, wie folgendes Lemma zeigt.

Lemma 2.3.1. *Der Operator $\mathcal{S} : L_2[0,T,\mathbb{R}^n] \to L_2[0,T,\mathbb{R}^n]$ ist selbstadjungiert, d.h. es gilt*

$$\langle \mathcal{S}u, w\rangle = \langle u, \mathcal{S}w\rangle \qquad \forall u, w \in L_2[0,T,\mathbb{R}^n].$$

Beweis: Wir benutzen die Bezeichnungen $z(u) = \mathcal{S}u$ und $z(w) = \mathcal{S}w$. Dann schließen wir mit Hilfe partieller Integration, der Symmetrie von A und unter Beachtung der Randwerte von $z(u)$ und $z(w)$

$$\begin{aligned}\langle z(u), w\rangle &= \langle z(u), -\ddot{z}(w) + Az(w)\rangle \\ &= \int_0^T -z(u)(t)^\mathsf{T}\ddot{z}(w)(t) + z(u)(t)^\mathsf{T}Az(w)(t)\,dt \\ &= \int_0^T \dot{z}(u)(t)^\mathsf{T}\dot{z}(w)(t) + z(u)(t)^\mathsf{T}Az(w)(t)\,dt \\ &= \int_0^T -\ddot{z}(u)(t)^\mathsf{T}z(w)(t) + (Az(u)(t))^\mathsf{T}z(w)(t)\,dt \\ &= \langle u, z(w)\rangle.\end{aligned}$$

□

Wir wissen demnach um die Identität von $\mathcal{S}$ und $\mathcal{S}^*$ und sind somit in der Lage eine neue Größe einzuführen, wie es z.B. auch in Tröltzsch [25] (2005) geschieht.

Definition 2.3.2 (Adjungierter Zustand). Sei $z \in L_2[0,T,\mathbb{R}^n]$. Die eindeutig bestimmte Lösung $p \in W^2_{2,0}[0,T,\mathbb{R}^n]$ der *adjungierten Gleichung*

$$\begin{aligned}-\ddot{p}(t) + Ap(t) &= z(t) - z_d(t) \qquad \forall' t \in [0,T] \\ p(0) &= p(T) = 0_n,\end{aligned} \tag{2.10}$$

nennen wir den *adjungierten Zustand*. ◊

Bemerkung: Den adjungierten Zustand bzw. die adjungierte Gleichung (2.10) führten wir soeben per Definition ein. Dies ermöglichte uns der Umstand, dass der adjungierte Operator zu $\mathcal{S}$ bereits auf Grund dessen Selbstadjungiertheit bekannt ist. Eine allgemeinere Herleitung des adjungierte Operators bzw. der adjungierten Gleichung mit Hilfe des formalen *Lagrange-Prinzips* verschieben wir daher in den Anhang (siehe Abschnitt A.1). ◊

Wie im Fall der Systemgleichung transformieren wir (2.10) in die Form einer Variationsgleichung. Diese lautet dann

$$\int_0^T \dot{p}(t)^\mathsf{T}\dot{v}(t) + p(t)^\mathsf{T}Av(t)\,dt = \int_0^T (z(t) - z_d(t))^\mathsf{T}v(t)\,dt \qquad \forall v \in W^1_{2,0}[0,T,\mathbb{R}^n]. \tag{2.11}$$

Mit den gleichen Argumenten wie in Lemma 2.1.2 folgt anschließend die Äquivalenz der beiden Lösungen. Weiterhin ergeben sich aus Satz 2.2.5 die folgenden Abschätzungen.

Lemma 2.3.3. *Seien $y \in L_2[0,T,\mathbb{R}^n]$ und $z \in W^2_{2,0}[0,T,\mathbb{R}^n]$ die eindeutig bestimmte Lösung der Systemgleichung* (1.4)*. Dann existiert eine eindeutig bestimmte Lösung $p \in W^2_{2,0}[0,T,\mathbb{R}^n]$ der adjungierten Gleichung* (2.10) *und für diese gilt:*

$$\|p\|_2 \leq T^2 \|z - z_d\|_2 \leq T^2 c_{2.6} \left(\|y\|_2 + \|z_d\|_2\right) \tag{2.12}$$

$$\|p\|_{1,2} \leq c_{2.6} \|z - z_d\|_2 \leq c^2_{2.6} \left(\|y\|_2 + \|z_d\|_2\right) \tag{2.13}$$

$$\|p\|_{2,2} \leq \sqrt{2} c_{2.7} \|z - z_d\|_2 \leq \sqrt{2} c_{2.6} c_{2.7} \left(\|y\|_2 + \|z_d\|_2\right). \tag{2.14}$$

Beweis: Der Beweis verläuft analog zu jenem von Satz 2.2.5. Die Subtraktion von z_d besitzt keinen Einfluss auf die Existenz und Eindeutigkeit der Lösung. Wir erhalten zunächst

$$\|z - z_d\|_2 \leq \|z\|_2 + \|z_d\|_2 \leq \|z\|_{1,2} + \|z_d\|_2 \leq c_{2.6} \|y\|_2 + \|z_d\|_2$$

und daraus folgernd

$$\|p\|_{1,2} \leq c_{2.6} \|z - z_d\|_2 \leq c^2_{2.6} \left(\|y\|_2 + \|z_d\|_2\right).$$

Die Ungleichung (2.14) folgt mit der adjungierten Gleichung (2.10), denn es ist

$$\|\ddot{p}\|_2 \leq \|A\| \, \|p\|_2 + \|z - z_d\|_2 \leq \left(1 + T^2 \|A\|\right) \|z - z_d\|_2 \leq c_{2.7} \|z - z_d\|_2$$

und somit

$$\begin{aligned} \|p\|_{2,2} = \sqrt{\|p\|^2_{1,2} + \|\ddot{p}\|^2_2} &\leq \sqrt{2} \max\{c_{2.6}, c_{2.7}\} \|z - z_d\|_2 \leq \sqrt{2} c_{2.7} \|z - z_d\|_2 \\ &\leq \sqrt{2} c_{2.6} c_{2.7} \left(\|y\|_2 + \|z_d\|_2\right). \end{aligned}$$

□

Auch an dieser Stelle ist es möglich das eben bewiesene Stabilitätsresultat auf Abschätzungen in der L_∞-Norm zu erweitern.

Korollar 2.3.4. *Sei nun $y \in L_\infty[0,T,\mathbb{R}^n]$. Dann gilt mit den Bezeichnungen aus Lemma 2.3.3*

$$\|p\|_\infty \leq T^{\frac{3}{2}} \|z - z_d\|_2 \leq T^2 c_{2.6} \left(\|y\|_\infty + \|z_d\|_\infty\right). \tag{2.15}$$

Beweis: Wir übernehmen die Abschätzungen aus Korollar 2.2.6 und benutzen für die zweite Aussage zusätzlich die Dreiecks-Ungleichung,

$$\|p\|_\infty \leq T^{\frac{3}{2}} \|z - z_d\|_2 \leq T^{\frac{3}{2}} (T^2 \|y\|_2 + \|z_d\|_2) \leq T^2 c_{2.6} \left(\|y\|_\infty + \|z_d\|_\infty\right).$$

□

In Lemma 2.3.1 benutzten wir bereits die Schreibweise $z(u)$ für die Verdeutlichung der Abhängigkeit des Zustandes von der rechten Seite der Systemgleichung (1.4). Auch bei der Behandlung des adjungierten Zustands verwenden wir später die Steuerung als Argument, d.h.

$$z(u) = \mathcal{S}(Bu + e), \qquad p(u) = \mathcal{S}^*(z(u) - z_d) = \mathcal{S}^*(\mathcal{S}(Bu + e) - z_d).$$

Bemerkung: Auch wenn wir in Lemma 2.3.1 die Gleichheit der Operatoren $\mathcal{S}$ und $\mathcal{S}^*$ herleiteten, bezeichnen wir sie dennoch mit unterschiedlichen Symbolen, da sie eine unterschiedliche Interpretation innerhalb der Problembehandlung erfahren. So ist z.B. der Wertevorrat von $\mathcal{S}^*$ wesentlich geringer, da wir den Operator ausschließlich auf Funktionen mit stärkeren Glattheitseigenschaften anwenden. ◊

2.4 Eigenschaften der optimalen Steuerung

Nach der Herleitung der adjungierten Gleichung untersuchen wir nun, ob das Problem (StP) überhaupt eine Lösung besitzt und falls ja, ob sie eindeutig bestimmt ist. Darüber hinaus ist eine Charakterisierung der Lösung durch die Fréchet-Ableitung an der Optimalstelle von großer Bedeutung, daher berechnen wir diese zuvor.

Da J auf dem Hilbert-Raum $L_2[0,T,\mathbb{R}^n]$ definiert ist, schreiben wir die Auswertung seiner Ableitung auf einem Element aus diesem Raum als Skalarprodukt und schließen so mit Hilfe der Kettenregel auf

$$\begin{aligned} J'(u)(h) &= \left(\frac{\partial \frac{1}{2}\|\mathcal{S}\mathcal{T}u - z_d\|_2^2}{\partial(\mathcal{S}\mathcal{T}u - z_d)} \frac{\partial(\mathcal{S}\mathcal{T}u - z_d)}{\partial(\mathcal{T}u)} \frac{\partial \mathcal{T}u}{\partial u} + \nu \frac{\partial \|u\|_2^2}{\partial u} \right)(h) \\ &= ((\mathcal{S}\mathcal{T}u - z_d) \circ \mathcal{S} \circ B + \nu u)(h) \\ &= \langle B^\mathsf{T} \mathcal{S}^*(\mathcal{S}\mathcal{T}u - z_d) + \nu u, h \rangle \end{aligned}$$

für beliebiges $h \in L_2[0,T,\mathbb{R}^m]$. Dabei beachten wir die Linearität von $\mathcal{S}$ und daraus folgernd $\mathcal{S}'(y)(h) = \mathcal{S}h$ für alle $y \in L_2[0,T,\mathbb{R}^n]$.

Lemma 2.4.1. *Das Problem* (StP) *besitzt eine eindeutig bestimmte Lösung $\bar{u}$. Die Bedingung*

$$\bar{u} = \Pi_{[a,b]}\left(-\frac{1}{\nu} B^\mathsf{T} p(\bar{u}) \right) \tag{2.16}$$

ist notwendig und hinreichend für die Optimalität der Steuerung $\bar{u}$ mit dem adjungierten Zustand $p(\bar{u}) = \mathcal{S}^*(\mathcal{S}(B\bar{u} + e) - z_d)$. Darüber hinaus ist die optimale Steuerung $\bar{u}$ in $[0,T]$ fast überall differenzierbar und ihre Ableitung wesentlich beschränkt, d.h. es gilt $\bar{u} \in W_\infty^1[0,T,\mathbb{R}^m]$.

Beweis: Das Funktional J ist strikt konvex und die Lösungsmenge U^{ad} nichtleer, konvex, beschränkt und abgeschlossen. Demnach schließen wir mit Alt [2] (2005, Satz

2.5.2) auf eine eindeutig bestimmte Lösung $\bar{u} \in U^{ad}$ von (StP).

Der Operator $\Pi_{[a,b]}$ ist eine Projektion auf eine abgeschlossene und konvexe Menge, daher ist nach dem Satz über die Charakterisierung einer Projektion (siehe Anhang, Satz A.3.4) die Bedingung

$$\langle B^{\mathsf{T}} p(\bar{u}) + \nu \bar{u}, u - \bar{u} \rangle \geq 0 \qquad \forall u \in U^{ad}. \tag{2.17}$$

zu (2.16) äquivalent. Wir werden im weiteren Verlauf diese *Variationsungleichung* verwenden.

Da J Fréchet-differenzierbar ist, existiert auch die Richtungsableitung für jede Richtung und sie stimmt mit der Fréchet-Ableitung überein. Die notwendige Optimalitätsbedingung lautet dann mit Alt [2] (2005, Satz 4.2.2)

$$\langle J'(\bar{u}), u - \bar{u} \rangle \geq 0 \qquad \forall u \in U^{ad}. \tag{2.18}$$

Auf Grund der Konvexität der Zielfunktion ist die Optimalitätsbedingung auch hinreichend (vgl. Alt [2] (2005, Lemma 4.3.6)).

Für die Ableitung des Zielfunktionals an der Stelle $\bar{u}$ gilt mit den Gedanken vor dem Lemma

$$J'(\bar{u})(u - \bar{u}) = (B^{\mathsf{T}} \mathcal{S}^*(\mathcal{S}\mathcal{T}\bar{u} - z_d) + \nu\bar{u})(u - \bar{u}) = \langle B^{\mathsf{T}} \mathcal{S}^*(\mathcal{S}\mathcal{T}\bar{u} - z_d) + \nu\bar{u}, u - \bar{u} \rangle.$$

Mit (2.18) und $p(\bar{u}) = \mathcal{S}^*(\mathcal{S}\mathcal{T}\bar{u} - z_d)$ folgt dann die Variationsungleichung

$$\langle J'(\bar{u}), u - \bar{u} \rangle = \langle B^{\mathsf{T}} p(\bar{u}) + \nu\bar{u}, u - \bar{u} \rangle \geq 0 \qquad \forall u \in U^{ad},$$

und somit auf Grund von $\bar{u} \in U^{ad}$ die Aussage des Lemmas.

Weiterhin leiten wir aus (2.16) die Lipschitz-Stetigkeit von $\bar{u}$ her. Der Operator $\Pi_{[a,b]}$ ist eine Projektion auf eine konvexe und abgeschlossene Menge, daher gilt $\|\Pi_{[a,b]}\| \leq 1$. Es folgt dann für beliebige $t_1, t_2 \in [0, T]$

$$|\bar{u}(t_1) - \bar{u}(t_2)| \leq \frac{\|B\|}{\nu} |p(\bar{u})(t_1) - p(\bar{u})(t_2)| \leq \frac{\|B\|}{\nu} \|\dot{p}(\bar{u})\|_\infty |t_1 - t_2|.$$

Damit ist $\bar{u}$ Lipschitz-stetig, daher auch absolut stetig und es existiert eine schwache Ableitung auf $[0, T]$. Deren Beschränktheit leiten wir sofort aus der Lipschitz-Stetigkeit ab und demnach folgt $\dot{\bar{u}} \in W^1_\infty[0, T, \mathbb{R}^m]$. □

Bemerkung:

(1) Die eben bewiesene Lipschitz-Stetigkeit von $\bar{u}$ liegt an der Darstellungsmöglichkeit (2.16) und an der stetigen Differenzierbarkeit des adjungierten Zustandes, welche aus $\bar{p} \in W^2_{2,0}[0, T, \mathbb{R}^n]$ und Satz A.3.2 folgt. Aus diesem Grund sind insbesondere die Werte der optimalen Steuerung an allen Stellen $t \in [0, T]$ eindeutig bestimmt.

(2) Die Aussage von Lemma 2.4.1 lässt sich auch in anderer, sehr anschaulicher Weise formulieren. Die optimale Steuerung $\bar{u}$ ist ein Fixpunkt der Zuordnung

$$u \mapsto \Pi_{[a,b]}\left(-\frac{1}{\nu}B^{\mathsf{T}}p(u)\right) = \Pi_{[a,b]}\left(-\frac{1}{\nu}B^{\mathsf{T}}\mathcal{S}^*(\mathcal{S}\mathcal{T}u - z_d)\right)$$

als Operator auf dem $L_2[0,T,\mathbb{R}^m]$.

(3) Wir behalten die Bezeichnungen des letzten Lemmas weiterhin bei und verwenden für die optimale Steuerung das Symbol $\bar{u}$. Die zugehörigen Zustände heißen dann $\bar{z} = z(\bar{u})$ und $\bar{p} = p(\bar{u})$.

$\Diamond$

Die Erweiterung von Lemma 2.4.1 auf unbeschränktes U^{ad} übernehmen wir aus Tröltzsch [25] (2005).

Korollar 2.4.2 (Tröltzsch [25], Satz 2.16). *Die Menge U^{ad} im Steuerungsproblem* (StP) *sei nicht beschränkt, d.h. $a_k = -\infty$ und/oder $b_k = \infty$ für $k \in \{1,\dots,n\}$, so besitzt* (StP) *dennoch eine eindeutig bestimmte Lösung. Die Formel* (2.16) *bleibt ebenfalls erhalten.*

Beweis: Sei $\underline{u} \geq 0$ das Infimum vom $J(u)$ über der Menge U^{ad}. Dann gilt für $\|u\|_2^2 > \frac{2}{\nu}(\underline{u}+1)$

$$J(u) \geq \frac{\nu}{2}\|u\|_2^2 > \underline{u} + 1.$$

Also wenden wir Lemma 2.4.1 an mit der nichtleeren, konvexen, abgeschlossenen und beschränkten Menge

$$U^{ad} \cap \Big\{u \in L_2[0,T,\mathbb{R}^m] : \|u\|_2^2 \leq \frac{2}{\nu}(\underline{u}+1)\Big\}.$$

Die Gleichung (2.16) ist von der Beschränktheit von U^{ad} unabhängig. $\square$

Bemerkung: Die Aussage des letzten Korollars folgt anschaulich sofort aus der strikten Konvexität des Zielfunktionals J. Sie bedingt nämlich $J \to \infty$ falls $\|u\| \to \infty$, daher ist das Minimum ebenfalls beschränkt. $\Diamond$

Eine wichtige Eigenschaft der Variationsungleichung (2.17) weisen wir im folgenden Lemma nach. Wir benötigen sie später bei der Betrachtung von diskreten Lösungen des Steuerungsproblems (StP).

Lemma 2.4.3. *Die Funktion $\bar{u} \in L_2[0,T,\mathbb{R}^m]$ erfüllt genau dann die Variationsungleichung* (2.17)*, wenn für fast alle $t \in [0,T]$ die punktweise Variationsungleichung*

$$\big(B^{\mathsf{T}}p(\bar{u})(t) + \nu\bar{u}(t)\big)^{\mathsf{T}}\big(u - \bar{u}(t)\big) \geq 0 \qquad \forall u \in U$$

mit $U = \{u \in \mathbb{R}^m : a \leq u \leq b\}$ Gültigkeit besitzt.

Beweis: Nehmen wir an, es existiere eine Funktion $u \in U^{ad}$ und eine Menge $\mathcal{N} \subset [0,T]$ vom Maß größer als Null, so dass gelte

$$\big(B^{\mathsf{T}}\bar{p}(t) + \nu\bar{u}(t)\big)^{\mathsf{T}}\big(u(t) - \bar{u}(t)\big) < 0 \qquad \forall t \in \mathcal{N}.$$

Definieren wir mit Hilfe dieser Menge die Funktion

$$u_{\mathcal{N}}(t) = \begin{cases} u(t), & t \in \mathcal{N} \\ \bar{u}(t), & t \notin \mathcal{N} \end{cases},$$

so folgte auf Grund von $u_{\mathcal{N}} \in U^{ad}$ sofort

$$\int_0^T \big(B^{\mathsf{T}}\bar{p}(t) + \nu\bar{u}(t)\big)^{\mathsf{T}}\big(u_{\mathcal{N}}(t) - \bar{u}(t)\big)\,dt = \int_{\mathcal{N}} \big(B^{\mathsf{T}}\bar{p}(t) + \nu\bar{u}(t)\big)^{\mathsf{T}}\big(u(t) - \bar{u}(t)\big)\,dt < 0$$

im Widerspruch zur Voraussetzung (2.17).

Die Rückrichtung ist offensichtlich. □

2.5 Steuerungsproblem mit beliebigen Randwerten

Bei der Vorstellung des Steuerungsproblems (StP) lautete die Forderung an die Randwerte $z(0) = z(T) = 0_n$. Durch diese Vorgabe war es möglich die Existenz und Eindeutigkeit einer Lösung der Systemgleichung (1.4) und die Stabilitätsungleichungen (2.5) sowie die darauf folgenden Abschätzungen zu zeigen. Die Verallgemeinerung auf beliebige Randwerte $z(0) = z_0$ und $z(T) = z_T$ mit $z_0, z_T \in \mathbb{R}^n$ untersuchen wir an dieser Stelle.

Dazu setzen wir die Funktion r an, als Summe der Funktion z und einer Geraden zwischen den beiden Punkten z_0 und z_T, d.h.

$$r(t) = z(t) + \left(z_0 + t\frac{z_T - z_0}{T}\right) \qquad \forall t \in [0,T].$$

Damit folgt $r(0) = z_0$, $r(T) = z_T$ und

$$-\ddot{z}(t) + Az(t) = -\ddot{r}(t) + A\left(r(t) + z_0 + t\frac{z_T - z_0}{T}\right) = y(t) \qquad \forall' t \in [0,T]$$

was mit $y^*(t) = y(t) - A\left(z_0 + t\frac{z_T - z_0}{T}\right)$ gleichbedeutend ist zu

$$\begin{aligned} -\ddot{r}(t) + Ar(t) &= y^*(t) \qquad \forall' t \in [0,T] \\ r(0) &= z_0 \\ r(T) &= z_T \end{aligned} \tag{2.19}$$

Da wir aus der Existenz einer Lösung von (2.19) auf die gleiche Art und Weise auch auf die Existenz einer Lösung von (1.4) mit Nullrandbedingungen schließen, sind die beiden Formulierungen äquivalent. Für den Beweis von (2.6) war es nötig, die Problemstellung mit verschwindenden Randwerten zu benutzen, aber wie wir eben zeigten, bedeutet das keine Einschränkung an die in der Systemgleichung geforderten Randwerte. Bei genauer Untersuchung, wie z.B. in Alt [2] (2005), stellt ich heraus, dass die Aussage von Satz 2.2.5 nur für Funktionen $r = r^1 - r^2$ mit r^1, r^2 Lösungen von (2.19) gelten muss. Daher zieht eine Einschränkung auf verschwindende Randwerte kein Verlust an der Allgemeinheit der Problemstellung nach sich.

Kapitel 3

Allgemeine Aspekte der Diskretisierung

3.1 Diskretisierung und diskrete Räume

Zur Lösung des Steuerungsproblems (StP) ist es nötig, das unendlich-dimensionale Problem zu diskretisieren und für die entstehenden diskreten, endlich-dimensionalen Probleme die Lösungen zu berechnen. Dafür benötigen wir Fehlerabschätzungen für die exakte optimale Steuerung und die durch Lösen der diskreten Probleme berechneten Funktionen. Diesen Weg gehen wir mittels zweier verschiedener Methoden, für die wir zuvor allgemeine Konzepte der Diskretisierung einführen.

Zunächst unterteilen wir das Intervall $[0,T]$ mit Hilfe der Zwischenstellen $0 = t_0 < t_1 < \ldots < t_{N-1} < t_N = T$ in N Intervalle der konstanten Länge $h = t_{i+1} - t_i$. Das bedeutet also $t_i = ih = iT/N$ für $i = 0, \ldots, N$. Wir bezeichnen mit T_i die links abgeschlossenen und rechts offenen Intervalle $[t_i, t_{i+1}[$ für $i = 0, \ldots, N-2$ und das beidseitig abgeschlossene Intervall $[t_i, t_{i+1}]$ für $i = N-1$. Auf der Basis dieser Zerlegung definieren wir endlich-dimensionale Unterräume von $L_2[0,T,\mathbb{R}^m]$ bzw. von $W_2^1[0,T,\mathbb{R}^n]$ unter Verwendung der aus (1.2) bekannten Vektorräume aller Polynome

$$U_h[0,T,\mathbb{R}^m] = \{u_h \in L_\infty[0,T,\mathbb{R}^m] : u_h|_{T_i} \in \mathcal{P}_0[t_i, t_{i+1}, \mathbb{R}^m], i = 0, \ldots, N-1\},$$
$$V_h[0,T,\mathbb{R}^n] = \{v_h \in C[0,T,\mathbb{R}^n] : v_h|_{T_i} \in \mathcal{P}_1[t_i, t_{i+1}, \mathbb{R}^n], i = 0, \ldots, N-1\}.$$

Es ist auf Grund der endlichen Dimension offensichtlich, dass beide Räume versehen mit dem L_2-Skalarprodukt Hilbert-Räume sind. Wir verwenden daher die Norm $\|\cdot\|_2$.

Oftmals benötigen wir auch Funktionen v_h aus $V_h[0,T,\mathbb{R}^n]$ mit den Randbedingungen $v_h(0) = v_h(T) = 0$, daher definieren wir zusätzlich den Raum

$$V_{h,0}[0,T,\mathbb{R}^n] = V_h[0,T,\mathbb{R}^n] \cap W_{2,0}^1[0,T,\mathbb{R}^n].$$

Diese endlich-dimensionalen Räume statten wir mit Basen aus und schreiben jedes

Element als Linearkombination der Basiselemente. Dazu benutzen wir folgende Darstellung

$$U_h[0,T,\mathbb{R}^m] = \underbrace{U_h[0,T,\mathbb{R}] \times \cdots \times U_h[0,T,\mathbb{R}]}_{m-\text{mal}}$$

des Raumes $U_h[0,T,\mathbb{R}^m]$ und definieren für $U_h[0,T,\mathbb{R}]$ die Basis

$$u_h^{(j)}(t) = \begin{cases} 1, & t \in T_j \\ 0, & \text{sonst} \end{cases}, \qquad j = 0, \ldots, N-1,$$

also die charakteristischen Funktionen der Intervalle T_j.

Für $V_h[0,T,\mathbb{R}^n]$ gehen wir analog von

$$V_h[0,T,\mathbb{R}^n] = \underbrace{V_h[0,T,\mathbb{R}] \times \cdots \times V_h[0,T,\mathbb{R}]}_{n-\text{mal}}$$

aus und definieren die Funktionen

$$v_h^{(j)}(t) = \begin{cases} \dfrac{t - t_{j-1}}{h}, & t \in T_{j-1} \\ \dfrac{t_{j+1} - t}{h}, & t \in T_j \\ 0, & \text{sonst} \end{cases}, \qquad j = 1, \ldots, N-1,$$

sowie

$$v_h^{(0)}(t) = \begin{cases} \dfrac{t_1 - t}{h}, & t \in T_0 \\ 0, & \text{sonst} \end{cases} \quad \text{und} \quad v_h^{(N)}(t) = \begin{cases} \dfrac{t - t_{N-1}}{h}, & t \in T_{N-1} \\ 0, & \text{sonst} \end{cases}.$$

Die $v_h^{(j)}$ sind die sogenannten *Hut*-Funktionen und erstrecken sich jeweils über das Intervall $[t_{j-1}, t_{j+1}]$ bzw. über das jeweilige Randintervall T_0 und T_{N-1}. Die Menge $(v_h^{(0)}, \ldots, v_h^{(N)})$ bildet dann die gesuchte Basis von $V_h[0,T,\mathbb{R}]$. Man weist leicht nach, dass diese Mengen von Funktionen wirklich Basen der jeweiligen Räume sind.

Wir besitzen demnach eine Darstellung beliebiger Funktionen $u_h \in U_h[0,T,\mathbb{R}]$ und $v_h \in V_h[0,T,\mathbb{R}]$ durch

$$u_h = \sum_{j=0}^{N-1} \alpha_j u_h^{(j)} \qquad \text{bzw.} \qquad v_h = \sum_{j=0}^{N} \beta_j v_h^{(j)}.$$

Die Dimensionen von $U_h[0,T,\mathbb{R}]$ und $V_h[0,T,\mathbb{R}]$ bestimmen sich durch die Anzahl der Basiselemente zu N und $N+1$. Daraus leiten wir nun die Darstellung beliebiger Funktionen $u_h \in U_h[0,T,\mathbb{R}^m]$ und $v_h \in V_h[0,T,\mathbb{R}^n]$ ab. Dazu definieren wir die Basiselemente

$u_h^{(j,k)}$ durch

$$u_h^{(j,k)} = (0, \ldots, 0, \underbrace{u_h^{(j)}}_{\text{k-te Stelle}}, 0, \ldots, 0)^{\mathsf{T}}, \qquad j = 0, \ldots, N-1;\ k = 1, \ldots, m.$$

Die Dimension von $U_h[0, T, \mathbb{R}^m]$ bestimmt sich somit zu Nm und jedes Element $u_h \in U_h[0, T, \mathbb{R}^m]$ lässt sich darstellen durch

$$u_h = \sum_{j=0}^{N-1} \sum_{k=1}^{m} \alpha_{j,k} u_h^{(j,k)} = \sum_{j=0}^{N-1} \alpha_j u_h^{(j)}.$$

Im zweiten Teil der Gleichung benutzten wir bereits eine andere Schreibweise. Die Koeffizienten α_j sind Vektoren der Dimension m und zusammen mit den eindimensionalen Basisfunktionen $u_h^{(j)}$ erhalten wir die gleiche Funktion. Bilden wir aus den Koeffizienten $\alpha_{j,k}$ den Vektor

$$\alpha = (\alpha_{0,1}, \ldots, \alpha_{0,m}, \ldots, \alpha_{N-1,1}, \ldots, \alpha_{N-1,m})^{\mathsf{T}} \in \mathbb{R}^{Nm},$$

so gibt es eine eineindeutige Zuordnung $u_h \longleftrightarrow \alpha$, die jeder Funktion aus $U_h[0, T, \mathbb{R}^m]$ den Vektor ihrer Koeffizienten zuordnet. Somit ist $U_h[0, T, \mathbb{R}^m]$ zum $\mathbb{R}^{Nm}$ isomorph und wir identifizieren jedes Element $u_h \in U_h[0, T, \mathbb{R}^m]$ mit dem zugehörigen Vektor α.

Analog gehen wir vor bei der Betrachtung einer Basis von V_h mit Hilfe der oben definierten Funktionen $v_h^{(j)}$. So seien

$$v_h^{(j,k)} = (0, \ldots, 0, \underbrace{v_h^{(j)}}_{\text{k-te Stelle}}, 0, \ldots, 0)^{\mathsf{T}} \qquad j = 0, \ldots, N;\ k = 1, \ldots, n.$$

Damit folgt für die Dimension von $V_h[0, T, \mathbb{R}^n]$ der Wert $(N+1)n$ und die Darstellbarkeit von Funktionen $z_h, p_h \in V_h[0, T, \mathbb{R}^n]$ durch

$$z_h = \sum_{j=0}^{N} \sum_{k=1}^{n} \beta_{j,k} v_h^{(j,k)} = \sum_{j=0}^{N} \beta_j v_h^{(j)}, \qquad p_h = \sum_{j=0}^{N} \sum_{k=1}^{n} \zeta_{j,k} v_h^{(j,k)} = \sum_{j=0}^{N} \zeta_j v_h^{(j)}.$$

Die Bezeichnungen z_h und p_h sind an dieser Stelle mit Bedacht in dieser Form gewählt. Wir werden später damit den diskretne Zustand bzw. den diskreten adjugnierten Zustand bezeichnen. Auch hier weisen wir auf die vereinfachende Schreibweise mit den eindimensionalen Basisfunktionen hin. Ferner existieren wieder eineindeutige Zuordnungen zwischen den Funktionen aus dem Raum $V_h[0, T, \mathbb{R}^n]$ und den zugehörigen Vektoren der Koeffizienten, also

$$z_h \longleftrightarrow \beta = (\beta_{0,1}, \ldots, \beta_{0,n}, \ldots, \beta_{N,1}, \ldots, \beta_{N,n})^{\mathsf{T}} \in \mathbb{R}^{(N+1)n},$$
$$p_h \longleftrightarrow \zeta = (\zeta_{0,1}, \ldots, \zeta_{0,n}, \ldots, \zeta_{N,1}, \ldots, \zeta_{N,n})^{\mathsf{T}} \in \mathbb{R}^{(N+1)n}.$$

Betrachten wir abschließend noch den Raum $V_{h,0}[0,T,\mathbb{R}^n]$, für dessen Elemente an den Stellen $t=0$ und $t=T$ eine Nullstelle vorgegeben ist. Für eine Basis dieses Raumes benötigen wir demnach für die Ränder keine Basisfunktionen, denn für die Koeffizienten β_0 und β_N wäre der Wert 0_n vorgeschrieben. Wir stellen somit die Funktionen aus $V_{h,0}[0,T,\mathbb{R}^n]$ durch eine Linearkombination wie folgt dar

$$z_h = \sum_{j=1}^{N-1} \sum_{k=1}^{n} \beta_{j,k} v_h^{(j,k)} = \sum_{j=1}^{N-1} \beta_j v_h^{(j)}.$$

Der zugeordnete Koeffizientenvektor besitzt dann die Gestalt

$$\beta = (\beta_{1,1}, \ldots, \beta_{1,n}, \ldots, \beta_{N-1,1}, \ldots, \beta_{N-1,n})^{\mathsf{T}} \in \mathbb{R}^{(N-1)n}.$$

Wir finden also auch hier bijektive Zuordnungen $V_h[0,T,\mathbb{R}^n] \longleftrightarrow \mathbb{R}^{(N+1)n}$ bzw. $V_{h,0}[0,T,\mathbb{R}^n] \longleftrightarrow \mathbb{R}^{(N-1)n}$ und identifizieren so jedes Element der jeweiligen Räume mit der zugehörigen Matrix ihrer Koeffizienten.

3.2 Diskrete Normen

Im letzten Abschnitt führten wir ausgehend von einer Unterteilung des Intervalls $[0,T]$ Vektorräume stückweise konstanter und stückweise linearer Funktionen ein. Für beide Räume übernahmen wir das Skalarprodukt $\langle z, v\rangle = \int_0^T z(t)^{\mathsf{T}} v(t)\,dt$ und die zugehörige Norm $\|\cdot\|_2$ vom Ausgangsraum $L_2[0,T,\mathbb{R}^n]$. Ein anderer Weg der Definition einer geeigneten Norm ist die Approximation des Integrals mittels der Quadraturformel

$$\int_0^T z(t)^{\mathsf{T}} v(t)\,dt \approx h \sum_{i=0}^{N} z(t_i)^{\mathsf{T}} v(t_i) \qquad z, v \in C[0,T,\mathbb{R}^n].$$

Setzen wir auf der rechten Seite zwei Funktionen $z_h, v_h \in V_h[0,T,\mathbb{R}^n]$ ein, so stellt der Ausdruck ein alternatives Skalarprodukt

$$\langle z_h, v_h\rangle_h := h \sum_{i=0}^{N} z_h(t_i)^{\mathsf{T}} v_h(t_i)$$

auf $V_h[0,T,\mathbb{R}^n]$ dar. Die von ihm induzierte Norm lautet dann

$$\|z_h\|_h = \sqrt{h \sum_{i=0}^{N} |z_h(t_i)|^2}.$$

Auf dem Raum $U_h[0,T,\mathbb{R}^n]$ lautet die Definition dieser diskreten Norm ähnlich, es ist

$$\|u_h\|_h = \sqrt{h \sum_{i=0}^{N-1} |u_h(S_i)|^2},$$

wobei S_i den Mittelpunkt des Intervalls T_i für $i = 0, \ldots, N-1$ bezeichnet.

Für die weiteren Betrachtungen definieren wir zwecks besserer Übersicht abkürzende Schreibweisen. So mögen die Subscripte h und $\bar{\mathsf{h}}$ für die vorwärts bzw. rückwärts gerichtete Euler-Approximation der ersten Ableitung auf dem Intervall rechts bzw. links der Stützstelle stehen. Das bedeutet $(z_i)_{\mathsf{h}} = \frac{z_{i+1}-z_i}{h}$, $i = 0, \ldots, N-1$, und $(z_i)_{\bar{\mathsf{h}}} = \frac{z_i - z_{i-1}}{h}$, $i = 1, \ldots, N$. Für die Hintereinanderausführung von $(\cdot)_{\bar{\mathsf{h}}}$ und $(\cdot)_{\mathsf{h}}$ schreiben wir $(\cdot)_{\bar{\mathsf{h}}\mathsf{h}} = (\cdot)_{\mathsf{h}\bar{\mathsf{h}}}$. Mit diesen Symbolen bestimmen wir die L_2-Norm der Funktion $z_h \in V_h[0, T, \mathbb{R}^n]$ exakt, denn mit $\beta_i = z_h(t_i)$ für $i = 0, \ldots, N$ ist

$$\begin{aligned}
\int_0^T |z_h(t)|^2 \, dt &= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} z_h(t)^{\mathsf{T}} z_h(t) \, dt \\
&= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} (t-t_i)^2 (\beta_i)_{\mathsf{h}}^{\mathsf{T}} (\beta_i)_{\mathsf{h}} + 2(t-t_j)(\beta_i)_{\mathsf{h}}^{\mathsf{T}} \beta_i + \beta_i^{\mathsf{T}} \beta_i \, dt \\
&= h \sum_{i=0}^{N-1} \frac{(\beta_{i+1}-\beta_i)^{\mathsf{T}}(\beta_{i+1}-\beta_i)}{3} + (\beta_{i+1}-\beta_i)^{\mathsf{T}} \beta_i + \beta_i^{\mathsf{T}} \beta_i \\
&= h \sum_{i=0}^{N-1} \frac{(\beta_{i+1}-\beta_i)^{\mathsf{T}}(\beta_{i+1}-\beta_i)}{3} + \beta_i^{\mathsf{T}} \beta_{i+1} \\
&= \beta^{\mathsf{T}} \underbrace{\frac{h}{6} \begin{pmatrix} 2I & I & & & \\ I & 4I & I & & \\ & & \ddots & & \\ & & I & 4I & I \\ & & & I & 2I \end{pmatrix}}_{=:\mathfrak{L}} \beta
\end{aligned}$$

und somit

$$\|z_h\|_2^2 = \beta^{\mathsf{T}} \mathfrak{L} \beta \qquad \forall z_h \in V_h[0, T, \mathbb{R}^n]. \tag{3.1}$$

Durch Vorgabe verschwindender Randwerte, also $z_h \in V_{h,0}[0, T, \mathbb{R}^n]$ und $\beta_i = z_h(t_i)$ für $i = 1, \ldots, N-1$, vereinfacht sich die Formel zu

$$\|z_h\|_2^2 = \beta^{\mathsf{T}} \underbrace{\frac{h}{6} \begin{pmatrix} 4I & I & & & \\ I & 4I & I & & \\ & & \ddots & & \\ & & I & 4I & I \\ & & & I & 4I \end{pmatrix}}_{\mathfrak{L}_0} \beta, \qquad \forall z_h \in V_{h,0}[0, T, \mathbb{R}^n]. \tag{3.2}$$

Die Matrizen $\mathfrak{L}$ und $\mathfrak{L}_0$ besitzen demnach die Dimensionen $(N+1)n \times (N+1)n$ und

$(N-1)n \times (N-1)n$. Wir kennen demnach zwei verschiedene Normen auf dem Raum $V_{h,0}[0,T,\mathbb{R}^n]$. In welcher Beziehung sie stehen zeigt folgendes Lemma.

Lemma 3.2.1. *Die Normen $\|\cdot\|_h$ bzw. $\|\cdot\|_2$ sind auf den Räumen $V_h[0,T,\mathbb{R}^n]$ bzw. $V_{h,0}[0,T,\mathbb{R}^n]$ äquivalent und die Abschätzungskonstanten sind unabhängig von h, d.h.*

$$\frac{1}{\sqrt{6}}\|u_h\|_h \le \|u_h\|_2 \le \|u_h\|_h \qquad \forall u_h \in V_h[0,T,\mathbb{R}^n] \tag{3.3}$$

$$\frac{1}{\sqrt{3}}\|z_h\|_h \le \|z_h\|_2 \le \|z_h\|_h \qquad \forall z_h \in V_{h,0}[0,T,\mathbb{R}^n]. \tag{3.3a}$$

Beweis: Wieder mit den Bezeichnungen $\beta_i = z_h(t_i)$ für $i = 1,\dots,N-1$ und $\beta = (\beta_1^\mathsf{T},\dots,\beta_{N-1}^\mathsf{T})^\mathsf{T}$ gilt offensichtlich

$$\|z_h\|_h^2 = h|\beta|^2.$$

Weiterhin wissen wir mit Hilfe des Satzes von Gerschgorin (siehe Anhang, Satz A.3.3) für die Eigenwerte λ der Matrix $\mathfrak{L}_0$ folgendes

$$|\lambda - \frac{2}{3}h| \le \frac{h}{3} \Longrightarrow \frac{h}{3} \le \lambda \le h.$$

Dann ist mit $\|z_h\|_2^2 = \beta^\mathsf{T}\mathfrak{L}_0\beta$

$$\frac{1}{3}\|z_h\|_h^2 \le \lambda_{\min}(\mathfrak{L}_0)\sum_{i=1}^{N-1}|\beta_i|^2 \le \langle \mathfrak{L}_0\beta,\beta\rangle \le \lambda_{\max}(\mathfrak{L}_0)\sum_{i=1}^{N-1}|\beta_i|^2 \le \|z_h\|_h^2.$$

Für den Raum $V_h[0,T,\mathbb{R}^n]$ beachten wir folgende Veränderungen. Der Vektor α besteht jetzt aus $N+1$ Komponenten, d.h. für die Eigenwerte der Matrix $\mathfrak{L}$ schließen wir mit dem gleichen Argument wie eben und beachten die zusätzlichen Bedingungen für die erste und letzte Komponente

$$|\lambda - \frac{2}{3}h| \le \frac{h}{3} \Longrightarrow \frac{h}{3} \le \lambda \le h, \qquad |\lambda - \frac{h}{3}| \le \frac{h}{6} \Longrightarrow \frac{h}{6} \le \lambda \le \frac{h}{2}.$$

□

Betrachten wir nun die Ableitung einer stückweise linearen Funktion z_h, so gilt $\dot{z}_h(t) = \frac{\beta_{i+1}-\beta_i}{h} = (\beta_i)_\mathsf{h}$ für $t \in T_i$ und $i = 0,\dots,N-1$, d.h. $\dot{z}_h$ ist stückweise konstant und es folgt

$$\|\dot{z}_h\|_2 = \sqrt{\int_0^T \dot{z}_h(t)^\mathsf{T}\dot{z}_h(t)\,dt} = \sqrt{\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\frac{|\beta_{i+1}-\beta_i|^2}{h^2}\,dt}$$

$$= \sqrt{h\sum_{i=0}^{N-1}\frac{|\beta_{i+1}-\beta_i|^2}{h^2}} = \sqrt{h\sum_{i=0}^{N-1}|(\beta_i)_\mathsf{h}|^2}.$$

Den Vektor $((\beta_0)_{\mathsf{h}}^{\mathsf{T}}, \ldots, (\beta_{N-1})_{\mathsf{h}}^{\mathsf{T}})^{\mathsf{T}}$ interpretieren wir als Ableitung der Funktion z_h und somit lautet eine äquivalente Formulierung der letzten Gleichung

$$\|\dot{z}_h\|_2 = \|\dot{z}_h\|_h,$$

d.h. bei stückweise konstanten Funktionen liefern beide Normen $\|\cdot\|_2$ und $\|\cdot\|_h$, den gleichen Wert. Allgemein gilt

$$\|u_h\|_2 = \|u_h\|_h \qquad \forall u_h \in U_h[0,T,\mathbb{R}^n].$$

Die Berechnung der Norm $\|\cdot\|_h$ geschieht dann unter Verwendung der Werte am jeweiligen Mittelpunkt der Intervalle. Daraus folgern wir mit Lemma 3.2.1

$$\|z_h\|_{1,2} = \sqrt{\|z_h\|_2^2 + \|\dot{z}_h\|_2^2} \leq \sqrt{\|z_h\|_h^2 + \|\dot{z}_h\|_h^2}$$

und analog zu den Betrachtungen bei kontinuierlichen Funktionen zeigen wir auch im diskreten Fall, dass der Ausdruck $\|\dot{z}_h\|_2$ schon eine äquivalente Norm auf $V_{h,0}[0,T,\mathbb{R}^n]$ darstellt. Wir erinnern an dieser Stelle an Lemma 2.2.1 mit der darin bewiesenen Poincaré'schen Ungleichung und leiten zunächst eine diskrete Variante davon her.

Lemma 3.2.2 (Diskrete Poincaré'sche Ungleichung). *Sei $z_h \in V_{h,0}[0,T,\mathbb{R}^n]$ mit $\beta_i = z_h(t_i)$ für $i = 0,\ldots,N$ und $\|z_h\|_h = \sqrt{h\sum_{i=1}^{N-1}|\beta_i|^2}$. Dann gilt*

$$\|z_h\|_h \leq \frac{T}{\sqrt{2}}\sqrt{h\sum_{i=0}^{N-1}|(\beta_i)_{\mathsf{h}}|^2}. \tag{3.4}$$

Beweis: Auf Grund von $\beta_0 = 0_n$ gilt zunächst

$$\beta_k = h\sum_{i=0}^{k-1}\frac{\beta_{i+1}-\beta_i}{h}, \qquad k = 1,\ldots,N$$

und somit schließen wir mit Hilfe der Hölder'schen Ungleichung (siehe Anhang, Lemma A.3.1)

$$\begin{aligned}|\beta_k|^2 &\leq h^2\left(\sum_{i=0}^{k-1}\frac{|\beta_{i+1}-\beta_i|}{h}\right)^2 \overset{(A.1)}{\leq} h^2k\sum_{i=0}^{k-1}\frac{|\beta_{i+1}-\beta_i|^2}{h^2}\\ &\leq h^2k\sum_{i=0}^{N-1}\frac{|\beta_{i+1}-\beta_i|^2}{h^2} = h^2k\sum_{i=0}^{N-1}|(\beta_i)_{\mathsf{h}}|^2 = hk\|\dot{z}_h\|_h^2.\end{aligned}$$

Damit gilt

$$\|z_h\|_h^2 = h\sum_{k=1}^{N-1}|\beta_k|^2 \leq h^2\|\dot{z}_h\|_h^2\sum_{k=1}^{N-1}k \leq h^2\frac{N(N-1)}{2}\|\dot{z}_h\|_h^2 \leq \frac{T^2}{2}\|\dot{z}_h\|_h^2.$$

□

Gehen wir zurück zu den Gedanken direkt vor dem Lemma und schätzen dort mit der eben hergeleiteten Ungleichung weiter ab, so folgt

$$\|z_h\|_{1,2} \leq \sqrt{\|z_h\|_2^2 + \|\dot{z}_h\|_2^2} \overset{(2.3)}{\leq} \sqrt{\frac{T^2}{2}\|\dot{z}_h\|_2^2 + \|\dot{z}_h\|_2^2} \leq c_{2.6}\,\|\dot{z}_h\|_h.$$

Die gleichen Randwertbedingungen bilden im nächsten Lemma die Argumente beim Beweis einer diskreten Variante des Sobolew'schen Einbettungssatzes (siehe Anhang, Satz A.3.2) dar. Wie wir sehen, stimmen sogar die Konstanten im Vergleich zum kontinuierlichen Fall überein.

Lemma 3.2.3. *Sei $z_h \in W_2^1[0,T,\mathbb{R}^n]$ mit $z_h(t_i) = \beta_i$ für $i = 0,\ldots,N$. Dann gilt*

$$\|z_h\|_\infty \leq \sqrt{2}\max\{\frac{1}{\sqrt{T}}, \sqrt{T}\}\,\sqrt{\|z\|_h^2 + \|\dot{z}_h\|_h^2}.$$

Darüber hinaus gilt für $z_h \in W_{2,0}^1[0,T,\mathbb{R}^n]$

$$\|z_h\|_\infty \leq \sqrt{T}\|\dot{z}_h\|_h. \tag{3.5}$$

Beweis: Sei j derjenige Index mit $|\beta_j| = \min\limits_{0\leq i\leq N}|\beta_i|$. Dann gilt wegen $|\beta_j| \leq \frac{1}{N}\sum_{k=0}^N |\beta_k|$ und $Nh = T$

$$|\beta_i| \leq |\beta_j| + h\sum_{k=j}^{i-1}\frac{|\beta_{k+1}-\beta_k|}{h} \leq \frac{h}{T}\sum_{k=0}^{N}|\beta_j| + h\sum_{k=0}^{N-1}\frac{|\beta_{k+1}-\beta_k|}{h}.$$

Auf Grund der Minimaleigenschaft von β_j folgt dann für $i = 0,\ldots,N$

$$\begin{aligned}
|\beta_i| &\leq \sqrt{T}\|\dot{z}_h\|_h + \frac{h}{T}\sum_{k=0}^{N}|\beta_k| \overset{(A.1)}{\leq} \sqrt{T}\|\dot{z}_h\|_h + \frac{1}{\sqrt{T}}\|z_h\|_h \\
&\leq \max\{\frac{1}{\sqrt{T}}, \sqrt{T}\}\,(\|z_h\|_h + \|\dot{z}_h\|_h) \\
&\overset{(A.1)}{\leq} \sqrt{2}\max\{\frac{1}{\sqrt{T}}, \sqrt{T}\}\,\sqrt{\|z_h\|_h^2 + \|\dot{z}_h\|_h^2}.
\end{aligned}$$

Wegen $\beta_0 = 0_n$ setzen wir für die zweite Ungleichung $j = 0$ und es folgt mit den gleichen Überlegungen wie eben

$$|\beta_i| \leq \sqrt{T}\|\dot{z}_h\|_h, \qquad i = 0,\ldots,N.$$

Da die rechten Seiten jeweils unabhängig vom Index i sind, bilden wir links das Maximum über alle i und vollenden somit den Beweis. □

Zum Abschluss dieses Abschnitts stellen wir ein Konzept vor, welches wir zum Beschreiben des Verhaltens bzw. der Form einer Funktion einsetzen. Dazu benötigen wir Zerlegungen des Intervalls $[0,T]$. Darunter ist jeweils eine Menge von Stellen $\tau_k \in [0,T]$, $k = 1,\ldots,m$ mit $m \in \mathbb{N}$ und $0 \leq \tau_1 < \tau_2 < \ldots < \tau_{m-1} < \tau_m \leq T$ zu verstehen.

Definition 3.2.4 (Variation). Seien $f : [0,T] \to \mathbb{R}^n$ und $\mathfrak{Z} = \{\tau_1, \ldots, \tau_m\} \subset [0,T]$ eine Zerlegung. Dann ist die *Variation von f bezüglich $\mathfrak{Z}$* auf dem Intervall $[0,T]$ gegeben durch

$$\mathsf{V}_0^T(\mathfrak{Z})\, f = \sum_{k=1}^{m-1} |f(\tau_{k+1}) - f(\tau_k)|.$$

Die *(totale) Variation von f* auf $[0,T]$ ist dann

$$\mathsf{V}_0^T f = \sup_{\mathfrak{Z}} \mathsf{V}_0^T(\mathfrak{Z}).$$

$\Diamond$

Die Variation ist keine Norm, denn von $\mathsf{V}_0^T f = 0$ ist der Schluss auf $f = 0$ nicht möglich. So besitzen z.B. alle konstanten Funktionen verschwindende Variation. Da aber $\mathsf{V}_0^T af = |a|\, \mathsf{V}_0^T f$ für alle $a \in \mathbb{R}$ und auch die Dreiecks-Ungleichung $\mathsf{V}_0^T(f+g) \leq \mathsf{V}_0^T f + \mathsf{V}_0^T g$ gilt, sprechen wir von einer Semi-Norm.

Definition 3.2.5 (Funktionen von beschränkter Variation). Eine Funktion $f : [0,T] \to \mathbb{R}^m$ heißt *von beschränkter Variation*, falls gilt

$$\mathsf{V}_0^T f < \infty.$$

Den Vektorraum aller Funktionen $f : [0,T] \to \mathbb{R}^m$ von beschränkter Variation bezeichnen wir mit $BV[0,T,\mathbb{R}^m]$. $\Diamond$

Bei der Definition der Variation übergingen wir einen formalen Aspekt. Wir benötigen Werte der Funktion an den konkreten Stellen der Zerlegung. Bei den Elementen der Räume $L_p[0,T,\mathbb{R}^n]$ handelt es sich allerdings um Funktionenklassen, d.h. das Element ändert sich nicht, wenn wir die Funktionswerte an höchstens abzählbar vielen Stellen durch beliebige Werte ersetzen. Die Variation dieser neuen Funktion kann durchaus einen anderen Wert annehmen oder sogar unendlich groß werden, obwohl wir die Klasse nicht verlassen. Das bedeutet, in einer Äquivalenzklasse liegen Funktionen von beschränkter und von unbeschränkter Variation. Darüber hinaus ist eine Identifizierung einer Klasse mit einer speziellen Funktion über die Eigenschaft der beschränkten Variation nicht möglich.

Die Verwendung der Variation ist also nur bei Funktionen möglich, bei denen jeder Funktionswert eindeutig definiert ist. Später betrachten wir lediglich die optimale Steuerung, also die Lösung von (StP). Aus diesem Grund aher fällt das Definitionsproblem weg, denn $\bar{u}$ ist nach Lemma 2.4.1 Lipschitz-stetig. An Hand der dort hergeleiteten Projektionsformel (2.16) erkennen wir die spezielle Struktur von $\bar{u}$. Die Funktionswerte sind eindeutig bestimmt, denn sie nehmen entweder den Wert $-\frac{1}{\nu}B^{\mathsf{T}}\bar{p}(t)$ oder den Wert einer der beiden Schranken a und b an. Auf Grund von

$$\bar{p} \in W_2^2[0,T,\mathbb{R}^n] \overset{\text{Satz } A.3.2}{\subset} C^1[0,T,\mathbb{R}^n]$$

sind die Funktionswerte von $\bar{u}$ und auch von $\dot{\bar{u}}$ an jeder Stelle des Intervalls eindeutig bestimmt. Daher ist der später häufig gebrauchte Ausdruck $\mathsf{V}_0^T \dot{\bar{u}}$ wohldefiniert.

Beispiel: Seien $f : [0,T] \to \mathbb{R}$ eine stetig differenzierbare Funktion und $a, b \in \mathbb{R}$ Konstanten mit $a < b$. Betrachten wir die Funktion $F : [0,T] \to \mathbb{R}^m$, definiert durch

$$F(t) := \begin{cases} a, & f(t) < a \\ b, & f(t) > b \\ f(t), & \text{sonst} \end{cases},$$

und überprüfen wir die Anzahl der Schnittstellen von f mit den konstanten Funktionen a und b, wobei nur diejenigen Stellen gemeint sind, an denen $f - a$ bzw. $f - b$ das Vorzeichen wechselt. Die Funktion F ist genau an diesen Stellen nicht differenzierbar im klassischen Sinne. Falls die Anzahl der Schnittstellen eine abzählbare Menge ist, so existiert dennoch eine schwache Ableitung, denn F ist Lipschitz-stetig und daher auch absolut stetig. Die schwache Ableitung stimmt demnach mit der punktweisen klassischen Ableitung überein. Darüber hinaus gilt

$$\dot{F}(t) = \begin{cases} \dot{f}(t), & f(t) \in (a,b) \\ 0, & \text{sonst} \end{cases}.$$

Die Funktion $\dot{F}$ ist somit stückweise stetig und besitzt abzählbar viele Sprungstellen. Da wir f als stetig differenzierbar vorausgesetzt hatten, gilt auch $|\dot{F}(t)| < \infty$ für fast alle $t \in [0,T]$ und somit $F \in W^1_\infty[0,T,\mathbb{R}]$. Ist die Summe der Beträge über alle Sprunghöhen sogar endlich, so folgt $F \in BV[0,T,\mathbb{R}]$. ◊

Abschließend beweisen wir noch zwei Aussagen über das Verhalten der Variation einer Funktion, die wir in späteren Abschätzungen häufig benötigen werden.

Lemma 3.2.6. *Seien $f \in BV[0,T,\mathbb{R}^m]$ und τ_0 eine beliebige Stelle in $[0,T]$. Dann gilt sowohl $f \in BV[0,\tau_0,\mathbb{R}^m]$ als auch $f \in BV[\tau_0,T,\mathbb{R}^m]$ und darüber hinaus*

$$\mathsf{V}_0^{\tau_0} f + \mathsf{V}_{\tau_0}^T f = \mathsf{V}_0^T f.$$

Beweis:

„≤“ Seien $\mathfrak{Z}_1 = \{\tau_1, \ldots, \tau_m\}$ eine Zerlegung von $[0,\tau_0]$ und $\mathfrak{Z}_2 = \{\tau_{m+1}, \ldots, \tau_k\}$ eine Zerlegung von $[\tau_0,T]$. Dann ist $\mathfrak{Z}_1 \cup \mathfrak{Z}_2$ eine Zerlegung vom gesamten Intervall $[0,T]$. Daraus folgt

$$\mathsf{V}_0^T(\mathfrak{Z}_1) f + \mathsf{V}_0^T(\mathfrak{Z}_2) f = \mathsf{V}_0^T(\mathfrak{Z}_1 \cup \mathfrak{Z}_2) f \le \mathsf{V}_0^T f.$$

„≥“ Sei $\mathfrak{Z} = \{\tau_1, \ldots, \tau_m\}$ eine Zerlegung von $[0,T]$. Dann ist $\mathfrak{Z} \cup \{\tau_0\}$ mit $\tau_l < \tau_0 < \tau_{l+1}$ eine Verfeinerung und da

$$|f(\tau_l) - f(\tau_{l+1})| \le |f(\tau_l) - f(\tau_0)| + |f(\tau_0) - f(\tau_{l+1})|$$

gilt, folgt schließlich

$$\begin{aligned}\mathsf{V}_0^T(\mathfrak{Z})f &= \sum_{j=1}^{m-1} |f(\tau_{j+1}) - f(\tau_j)| \\ &= \sum_{j=1}^{l-1} |f(\tau_{j+1}) - f(\tau_j)| + |f(\tau_0) - f(\tau_l)| \\ &\quad + \sum_{j=l+1}^{m-1} |f(\tau_{j+1}) - f(\tau_j)| + |f(\tau_{l+1}) - f(\tau_0)| \\ &\leq \mathsf{V}_0^{\tau_0} f + \mathsf{V}_{\tau_0}^T f.\end{aligned}$$

□

Korollar 3.2.7. *Seien $f \in BV[0,T,\mathbb{R}^m]$ und $0 = t_0 < t_1 < \ldots < t_{N-1} < t_N = T$ die oben eingeführten Stützstellen. Dann gilt*

$$\sum_{i=0}^{N-1} \mathsf{V}_{t_i}^{t_{i+1}} f = \mathsf{V}_0^T f. \tag{3.6}$$

Beweis: Wir verweisen auf das voraus gehende Lemma. Das Korollar folgt sofort durch Iteration. □

Bemerkung: Den Ausdruck $\mathsf{V}_{t_i}^{t_{i+1}}$ werden wir noch häufiger benutzen, daher verwenden wir ab sofort $\mathsf{V}_{T_i} := \mathsf{V}_{t_i}^{t_{i+1}}$. ◊

Korollar 3.2.8. *Sei $f \in W_1^1[0,T,\mathbb{R}^n]$. Dann gilt $f \in BV[0,T,\mathbb{R}^n]$ und*

$$\mathsf{V}_0^T f \leq \|\dot{f}\|_{L_1[0,T,\mathbb{R}^n]}. \tag{3.7}$$

Beweis: Auf Grund der Voraussetzung ist f stetig, daher ist die Variation wohldefiniert und es gilt mit den obigen Bezeichnungen und einer beliebigen Zerlegung $\mathfrak{Z} = \{\tau_0, \ldots, \tau_k\}$

$$\mathsf{V}_0^T(\mathfrak{Z})f = \sum_{i=0}^{k-1} |f(\tau_{i+1}) - f(\tau_i)| = \sum_{i=0}^{k-1} \Big| \int_{\tau_i}^{\tau_{i+1}} \dot{f}(t) \Big| \, dt \leq \|\dot{f}\|_{L_1[0,T,\mathbb{R}^n]}.$$

Da $\mathfrak{Z}$ beliebig ist, gehen wir links zum Supremum über und erhalten die Aussage. □

Lemma 3.2.9. *Sei $z \in W_{2,0}^2[0,T,\mathbb{R}^n]$ und $z(t_i)_{\mathsf{h}\bar{\mathsf{h}}} = \dfrac{z(t_{i+1}) - 2z(t_i) + z(t_{i-1})}{h^2}$ für die Indizes $i = 1, \ldots, N-1$, so gilt*

$$\|(z)_{\mathsf{h}\bar{\mathsf{h}}}\|_h = \left(h \sum_{i=1}^{N-1} \left| z(t_i)_{\mathsf{h}\bar{\mathsf{h}}} \right|^2 \right)^{\frac{1}{2}} \leq 2\|\ddot{z}\|_2.$$

Sei $u \in W^1_\infty[0,T,\mathbb{R}^m]$ *mit* $\dot u \in BV[0,T,\mathbb{R}^m]$ *und* $u(t_i)_{\mathsf{h}\bar{\mathsf{h}}}$ *für* $i = 1,\dots,N-1$ *wie eben, sowie* $u(t_0)_{\mathsf{h}\bar{\mathsf{h}}} = \dfrac{u(t_1)-u(t_0)}{h^2}$ *und* $u(t_N)_{\mathsf{h}\bar{\mathsf{h}}} = \dfrac{u(t_{N-1})-u(t_N)}{h^2}$*, so folgt*

$$\|(u)_{\mathsf{h}\bar{\mathsf{h}}}\|_h = \left(h \sum_{i=1}^{N-1} \left|u(t_i)_{\mathsf{h}\bar{\mathsf{h}}}\right|^2 \right)^{\frac12} \le 2\,(\mathsf{V}_0^T \dot u + \|\dot u\|_\infty)\, h^{-\frac12}.$$

Beweis: Auf Grund der Differenzierbarkeit von z schließen wir zunächst mit der Bezeichnung $S_i = (t_i + t_{i+1})/2$, $i = 0,\dots,N-1$, für den Mittelpunkt des Intervalls T_i

$$\begin{aligned}
\|(z)_{\mathsf{h}\bar{\mathsf{h}}}\|_h^2 &= h \sum_{i=1}^{N-1} \frac{|z(t_{i+1}) - 2z(t_i) + z(t_{i-1})|^2}{h^4} \\
&= \frac{1}{h^3} \sum_{i=1}^{N-1} \Big| \int_{-\frac{h}{2}}^{\frac{h}{2}} \dot z(S_i + t)\,dt - \int_{-\frac{h}{2}}^{\frac{h}{2}} \dot z(S_{i-1} + t)\,dt \Big|^2 \\
&\le \frac{1}{h^3} \sum_{i=1}^{N-1} \left(\int_{-\frac{h}{2}}^{\frac{h}{2}} |\dot z(S_i + t) - \dot z(S_{i-1} + t)|\,dt \right)^2 \\
&\le \frac{1}{h^3} \sum_{i=1}^{N-1} \left(\int_{-\frac{h}{2}}^{\frac{h}{2}} \int_{S_{i-1}}^{S_i} |\ddot z(s + t)|\,ds\,dt \right)^2 \\
&\le \frac{1}{h} \sum_{i=1}^{N-1} (\|\ddot z\|_{L_1(T_{i-1})} + \|\ddot z\|_{L_1(T_i)})^2 \\
&\overset{(A.2)}{\le} \sum_{i=1}^{N-1} (\|\ddot z\|_{L_2(T_{i-1})} + \|\ddot z\|_{L_2(T_i)})^2 \le 4\|\ddot z\|_2^2,
\end{aligned}$$

also

$$\|(z)_{\mathsf{h}\bar{\mathsf{h}}}\|_h \le 2\|\ddot z\|_2.$$

Weiterhin gilt

$$\begin{aligned}
h \sum_{i=1}^{N-1} \frac{|u(t_{i+1}) - 2u(t_i) + u(t_{i-1})|^2}{h^4} &\le \frac{1}{h^3} \sum_{i=1}^{N-1} \left(\int_{-\frac{h}{2}}^{\frac{h}{2}} |\dot u(S_i + t) - \dot u(S_{i-1} + t)|\,dt \right)^2 \\
&\le \frac{1}{h^3} \sum_{i=1}^{N-1} \left(\int_{-\frac{h}{2}}^{\frac{h}{2}} \mathsf{V}_{t_{i-1}}^{t_{i+1}} \dot u\,dt \right)^2 \\
&= \frac{1}{h} \sum_{i=1}^{N-1} \left(\mathsf{V}_{t_{i-1}}^{t_{i+1}} \dot u \right)^2 \overset{(3.6)}{\le} 4\big(\mathsf{V}_0^T \dot u\big)^2 \frac{1}{h},
\end{aligned}$$

und

$$h \frac{|u(t_1) - u(t_0)|^2}{h^4} \le \|\dot u\|_\infty^2 \frac{1}{h}, \qquad h \frac{|u(t_{N-1}) - u(t_N)|^2}{h^4} \le \|\dot u\|_\infty^2 \frac{1}{h}$$

also

$$\|(u)_{\mathsf{h}\bar{\mathsf{h}}}\|_h \leq 2\,(\mathsf{V}_0^T \dot{u} + \|\dot{u}\|_\infty)\, h^{-\frac{1}{2}}.$$

□

3.3 Diskretisierungsoperatoren und Interpolation

In den späteren Abschnitten untersuchen wir Approximationen der Funktionen im Steuerungsproblem (StP) durch Elemente der zuletzt eingeführten Räume. Das heißt, wir untersuchen die Eigenschaften stückweise konstanter und stückweise linearer Interpolationsfunktionen. Von besonderem Interesse sind die Abstände zwischen einer Funktion und ihrer Approximation in verschiedenen Normen.

Wir führen den Operator $P_1 : W_2^1[0,T,\mathbb{R}^n] \to V_h[0,T,\mathbb{R}^n]$ ein, der jeder Funktion z diejenige Funktion $P_1 z$ zuordnet mit $(P_1 z)(t_j) = z(t_j)$ für $j = 0,\ldots,N$. Das ist für $W_2^1[0,T,\mathbb{R}^n]$ sinnvoll, denn der Raum liegt eingebettet in $C[0,T,\mathbb{R}^n]$ und die Funktionswerte an isolierten Stellen sind daher eindeutig bestimmt. Die Funktion $P_1 z$ stellen wir als Linearkombination der Basiselemente $v_h^{(j)}$ dar

$$P_1 z = \sum_{j=0}^{N} z(t_j) v_h^{(j)}.$$

Es ist offensichtlich, dass $P_1 v_h = v_h$ für alle $v_h \in V_h[0,T,\mathbb{R}^n]$ gilt.

Neben diesem Diskretisierungsoperator auf den Raum $V_h[0,T,\mathbb{R}^n]$, benötigen wir eine Approximation durch stückweise konstante Funktionen. Wir definieren daher den Operator $P_0 : W_2^1[0,T,\mathbb{R}^m] \to U_h[0,T,\mathbb{R}^m]$ durch

$$(P_0 u)(t) = u(S_i) \qquad \forall t \in T_i,\ i = 0,\ldots,N-1$$

wobei S_i den Mittelpunkt des Intervalls T_i bezeichnet, also $S_i = (t_i + t_{i+1})/2$. Auch das Bild unter P_0 besitzt eine Darstellung in den Basiselementen des Bildraumes, es ist

$$P_0 u = \sum_{j=0}^{N-1} u(S_j) u_h^{(j)}.$$

Bevor wir weitere Aussagen beweisen, halten wir die so eben eingeführten Operatoren in einer Definition fest.

Definition 3.3.1 (Diskretisierungsoperatoren).

1. Der lineare Diskretisierungsoperator $P_1 : W_2^1[0,T,\mathbb{R}^n] \to V_h[0,T,\mathbb{R}^n]$ ordne jeder Funktion $z \in W_2^1[0,T,\mathbb{R}^n]$ die Funktion $P_1 z \in V_h[0,T,\mathbb{R}^n]$ zu, für die an den Punkten $t_0,\ldots,t_N$ gilt

$$z(t_i) = (P_1 z)(t_i) \qquad i = 0,\ldots,N.$$

Auf dem Intervall T_i besitzt $P_1 z$ die folgende Gestalt

$$(P_1 z)(t) = \frac{z(t_{i+1}) - z(t_i)}{h}(t - t_i) + z(t_i), \qquad i = 0, \ldots, N-1.$$

2. Sei $S_i = (t_{i+1} + t_i)/2$ der Mittelpunkt des Intervalls T_i. Der lineare Diskretisierungsoperator $P_0 : W_2^1[0, T, \mathbb{R}^m] \to U_h[0, T, \mathbb{R}^m]$ ordne jeder Funktion $u \in W_2^1[0, T, \mathbb{R}^m]$ diejenige Funktion $P_0 u \in U_h[0, T, \mathbb{R}^m]$ zu, für die gilt

$$(P_0 u)(t) = u(S_i), \qquad t \in T_i,\ i = 0, \ldots, N-1.$$

Auf dem Intervall T_i nimmt dann $P_0 u$ konstant den Wert $u(S_i)$ an. ◊

Bemerkungen:

(1) Die Operatoren P_1 und P_0 sind wohldefiniert. Seien r^1 und r^2 zwei Funktionen aus $V_h[0, T, \mathbb{R}^n]$ mit $r^1(t_i) = z(t_i) = r^2(t_i)$ für $i = 0, \ldots, N$. Dann besitzt die Funktion $r^1 - r^2$ in jedem Teilintervall T_i zwei Nullstellen, nämlich t_i und t_{i+1} und auf Grund der Linearität in T_i gilt $r^1 \equiv r^2$. Ein ähnliches Argument zeigt die Korrektheit der Definition von P_0.

(2) Der Operator P_1 bildet formal auf den Raum $V_h[0, T, \mathbb{R}^n]$ ab. Es ist offensichtlich, dass der Bildbereich von $W_{2,0}^1[0, T, \mathbb{R}^n]$ gleich dem Raum $V_{h,0}[0, T, \mathbb{R}^n]$ ist. Daher gehen wir beim Bild von P_1 ohne weiteren Kommentar von einer Funktion mit verschwindenden Randwerten aus, falls das Argument diese Eigenschaft besitzt. Die Basisfunktionen $v_h^{(0)}$ und $v_h^{(N)}$ benötigen wir nicht und lassen sie daher weg.

(3) Wir gewinnen aus $N + 1$ bzw. N Vektoren der Dimension n eine stückweise lineare bzw. konstante Funktion v_h bzw. u_h, indem wir die Vektoren auf jedem Teilintervall T_i komponentenweise linear verbinden bzw. konstant auf die Werte des Vektors setzen. Umgekehrt gewinnen wir aus jeder Funktion $v_h \in V_h[0, T, \mathbb{R}^n]$ bzw. $u_h \in U_h[0, T, \mathbb{R}^n]$ an jeder Stelle t_i bzw. S_i für $i = 0, \ldots, N$ bzw. für $i = 0, \ldots, N-1$ einen n-dimensionalen Vektor. Wir sprechen daher beim Bild von P_0 bzw. P_1 synonym von Vektor oder Funktion, wobei wir bei der Bezeichnung „Vektor“ anmerken, dass es sich dabei um eine Menge von Vektoren handelt, also um eine Matrix. Im eindimensionalen Fall vereinfacht sich dies wiederum, so dass wir für eine einheitliche Schreibweise immer auf die Bezeichnung „Vektor“ zurückgreifen. ◊

Nach Einführung der Diskretisierungsoperatoren P_0 und P_1 stellt sich sofort die Frage, welcher Fehler entsteht, wenn wir anstatt der Funktion ihre stückweise konstante bzw. stückweise lineare Interpolation verwenden. Der Abstand zwischen beiden Funktion, gemessen in verschiedenen Normen, spielt eine zentrale Rolle bei späteren Fehlerabschätzungen zwischen der exakten Lösung der Systemgleichung und ihrer näherungsweisen Lösung. Die Approximationsgüte hängt im Wesentlichen von drei Parametern

ab. So halten wir - rein qualitativ - fest, dass eine Erhöhung der Differenzierbarkeit der Ausgangsfunktion, den Fehler nur verringern kann. Eine bessere Approximation sollte ähnliche Resultate nach sich ziehen. Eine wichtige Frage dabei ist der Einfluss der zur Abstandsmessung herangezogenen Norm, d.h. wir untersuchen die Fehler

$$E_{k,p,j} = \|z - P_j z\|_p, \qquad z \in W_2^k[0,T,\mathbb{R}^n],\ j \in \{0,1\}.$$

Wir gehen systematisch vor und untersuchen zunächst stückweise konstante Approximation.

Lemma 3.3.2. *Sei $z \in W_p^1[0,T,\mathbb{R}^n]$ für $p \in \{1,2\}$. Dann gelten folgende Ungleichungen*

$$\|z - P_0 z\|_p \le \|\dot z\|_p\, h.$$

Fordern wir $z \in W_\infty^1[0,T,\mathbb{R}^n]$, so ist

$$\|z - P_0 z\|_\infty \le \frac{1}{2}\|\dot z\|_\infty\, h.$$

Beweis: Im Fall $p = 1$ gilt

$$\|z - P_0 z\|_1 = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |z(t) - z(S_i)|\, dt \le \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \int_{S_i}^{t} |\dot z(s)|\, ds\, dt \le \|\dot z\|_1\, h.$$

Weiter ist

$$\begin{aligned}\|z - P_0 z\|_2^2 &= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |z(t) - z(S_i)|^2\, dt = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big| \int_{S_i}^{t} \dot z(s)\, ds \Big|^2\, dt \\ &\le \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|\dot z\|_{L_1(T_i)}^2\, dt \le \sum_{i=0}^{N-1} h^2 \|\dot z\|_{L_2(T_i)}^2 = \|\dot z\|_2\, h^2.\end{aligned}$$

Auf Grund der Voraussetzung $\|\dot z\|_\infty < \infty$ ist z Lipschitz-stetig, denn für beliebige $t_1, t_2 \in [0,T,]$ gilt

$$|z(t_1) - z(t_2)| \le \Big| \int_{t_1}^{t_2} \dot z(v)\, dv \Big| \le \|\dot z\|_\infty |t_2 - t_1|.$$

Daraus folgt sofort die Aussage. □

Erhöhen wir die Differenzierbarkeit von z, also $z \in W_2^k[0,T,\mathbb{R}^n]$ mit $k > 1$, so erreichen wir dennoch keine bessere Approximation in den oben betrachteten Normen. Wir sind allerdings in der Lage den Abstand auch auf andere Weise zu messen und dort erhalten wir auch eine höhere Konvergenzordnung.

Lemma 3.3.3. *Sei $u \in W_\infty^1[0,T,\mathbb{R}^m]$ mit $\dot{u} \in BV[0,T,\mathbb{R}^m]$. Dann gilt folgende Abschätzung*

$$\Big| \int_0^T u(t) - (P_0 u)(t)\, dt \Big| \leq \frac{1}{4} \mathsf{V}_0^T \dot{u}\, h^2. \tag{3.8}$$

Sei $z \in W_2^2[0,T,\mathbb{R}^n]$, so gilt

$$\Big| \int_0^T z(t) - (P_0 z)(t)\, dt \Big| \leq \frac{1}{4} \|\ddot{z}\|_1\, h^2. \tag{3.9}$$

Beweis: Zunächst gilt auf Grund der Voraussetzungen mit

$$\begin{aligned}
\Big| \int_0^T u(t) - (P_0 u)(t)\, dt \Big| &= \Big| \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} u(t) - u(S_i)\, dt \Big| \\
&= \Big| \sum_{i=0}^{N-1} \int_0^{\frac{h}{2}} u(S_i + t) - 2u(S_i) + u(S_i - t)\, dt \Big| \\
&= \Big| \sum_{i=0}^{N-1} \int_0^{\frac{h}{2}} \Big[\int_0^t \dot{u}(S_i + s) - \dot{u}(S_i - t + s)\, ds \Big]\, dt \Big| \\
&\leq \sum_{i=0}^{N-1} \int_0^{\frac{h}{2}} \int_0^{\frac{h}{2}} |\dot{u}(S_i + s) - \dot{u}(S_i - t + s)|\, ds\, dt \\
&\leq \frac{h^2}{4} \sum_{i=0}^{N-1} \mathsf{V}_{T_i} \dot{u} \overset{(3.6)}{=} \frac{1}{4} \mathsf{V}_0^T \dot{u}\, h^2.
\end{aligned}$$

Für die zweite Abschätzung knüpfen wir direkt an die Herleitung im ersten Teil an

$$\begin{aligned}
\sum_{i=0}^{N-1} \Big| \int_{t_i}^{t_{i+1}} z(t) - (P_0 z)(t)\, dt \Big| &\leq \sum_{i=0}^{N-1} \int_0^{\frac{h}{2}} \int_0^{\frac{h}{2}} |\dot{z}(S_i + s) - \dot{z}(S_i - t + s)|\, ds\, dt \\
&\leq \frac{h^2}{4} \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |\ddot{z}(t)|\, dt \leq \frac{1}{4} \|\ddot{z}\|_1\, h^2.
\end{aligned}$$

Zum Abschluss halten wir noch ein Ergebnis für den lokalen Fehler fest. Es ist

$$\Big| \int_{t_i}^{t_{i+1}} z(t) - (P_0 z)(t)\, dt \Big| \leq \frac{1}{4} \|\ddot{z}\|_{L_1(T_i)}\, h^2 \leq \frac{1}{4} \|\ddot{z}\|_{L_2(T_i)}\, h^2 \sqrt{h}. \tag{3.10}$$

□

Jetzt erhöhen wir die Approximationsgüte und erhalten die erwarteten Verbesserungen, wie folgendes Lemma zeigt.

Lemma 3.3.4. *Sei z für $p \in \{1, 2, \infty\}$ ein Element des $W_p^2[0, T, \mathbb{R}^n]$. Dann gelten folgende Abschätzungen*

$$\|z - P_1 z\|_p \leq \|\ddot{z}\|_p \, h^2 = |z|_{2,p} \, h^2. \tag{3.11}$$

Beweis: Der Beweis der Abschätzung für $p = 2$ läuft zunächst über die Herleitung der geforderten Ungleichung für ein beliebiges Teilintervall T_i, $i = 0, \ldots, N-1$, und anschließender Summation über alle i, wie z.B. auch in D. Braess [7] (1997). Sei also $t \in T_i$, dann schreiben wir

$$z(t) - (P_1 z)(t) = \underbrace{z(t_i) - (P_1 z)(t_i)}_{0} + \int_{t_i}^{t} \dot{z}(s) - (\dot{P_1 z})(s) \, ds$$

und somit

$$|z(t) - (P_1 z)(t)| \leq \int_{t_i}^{t_{i+1}} |\dot{z}(s) - (\dot{P_1 z})(s)| \, ds = \|\dot{z} - (\dot{P_1 z})\|_{L_1(T_i)}.$$

Daraus folgt mit der Hölder'schen Ungleichung (siehe Anhang, Lemma A.3.1)

$$\|z - P_1 z\|_{L_2(T_i)} \leq \|\dot{z} - (\dot{P_1 z})\|_{L_1(T_i)} \sqrt{h} \leq \|\dot{z} - (\dot{P_1 z})\|_{L_2(T_i)} \, h.$$

Darüber hinaus besitzt $z - P_1 z$ in T_i zwei Nullstellen, nämlich genau die Intervallrandpunkte. Somit existiert nach dem Satz von Rolle ein $\xi \in T_i$ mit $(\dot{z} - (\dot{P_1 z}))(\xi) = 0$ und es ist

$$\dot{z}(t) - (\dot{P_1 z})(t) = \underbrace{(\dot{z} - (\dot{P_1 z}))(\xi)}_{0} + \int_{\xi}^{t} \ddot{z}(s) \, ds.$$

Auf Grund der Linearität der Funktion $P_1 z$ verschwindet deren zweite Ableitung und wir schließen wieder wie eben

$$|\dot{z}(t) - (\dot{P_1 z})(t)| \leq \int_{t_i}^{t_{i+1}} |\ddot{z}(s)| \, ds = \|\ddot{z}\|_{L_1(T_i)}$$

und

$$\|\dot{z} - (\dot{P_1 z})\|_{L_2(T_i)} \leq \|\ddot{z}\|_{L_1(T_i)} \sqrt{h} \leq \|\ddot{z}\|_{L_2(T_i)} \, h.$$

Zusammen mit der obigen Abschätzung erhalten wir

$$\|z - P_1 z\|_{L_2(T_i)} \leq \|\dot{z} - (\dot{P_1 z})\|_{L_2(T_i)} \, h \leq \|\ddot{z}\|_{L_2(T_i)} \, h^2.$$

Das Quadrat der Norm bezüglich des gesamten Gebiets $[0, T]$ berechnet sich als Summe der Quadrate der Normen über die Teilintervalle, daher gelangen wir zu

$$\|z - P_1 z\|_2 \leq \|\ddot{z}\|_2 \, h^2.$$

Im Fall $p = 1$ besteht die Möglichkeit eines kürzeren Beweises, daher betrachten wir ihn separat. Es ist

$$\begin{aligned}\int_{t_i}^{t_{i+1}} |z(t) - (P_1 z)(t)|\, dt &\le h \int_{t_i}^{t_{i+1}} \Big| \dot{z}(t) - \frac{z(t_{i+1}) - z(t_i)}{h} \Big|\, dt \\ &= \int_{t_i}^{t_{i+1}} |h\dot{z}(t) - (z(t_{i+1}) - z(t_i))|\, dt \\ &= \int_{t_i}^{t_{i+1}} \int_{t_i}^{t_{i+1}} |\dot{z}(t) - \dot{z}(s)|\, ds\, dt \\ &\le \int_{t_i}^{t_{i+1}} \int_{t_i}^{t_{i+1}} \int_{s}^{t} |\ddot{z}(v)|\, dv\, ds\, dt \\ &\le \|\ddot{z}\|_{L_1(T_i)}\, h^2.\end{aligned}$$

Daraus folgt durch Summation über alle $i = 0, \ldots, N-1$

$$\int_0^T |z(t) - (P_1 z)(t)|\, dt \le \|\ddot{z}\|_1\, h^2.$$

Zum Beweis der Ungleichung (3.11) für $p = \infty$ gehen wir ähnlich vor. Es ist für alle $t \in T_i$

$$|z(t) - (P_1 z)(t)| \le \int_{t_i}^{t_{i+1}} |\dot{z}(s) - (\dot{P_1 z})(s)|\, ds \le \|\dot{z} - (\dot{P_1 z})\|_{L_\infty(T_i)}\, h$$

und somit auch

$$\|z - (P_1 z)\|_{L_\infty(T_i)} \le \|\dot{z} - (\dot{P_1 z})\|_{L_\infty(T_i)}\, h.$$

Weiterhin gilt für alle $t \in T_i$

$$|\dot{z}(t) - (\dot{P_1 z})(t)| \le \int_{t_i}^{t_{i+1}} |\ddot{z}(s)|\, ds \le \|\ddot{z}\|_{L_\infty(T_i)}\, h,$$

da die zweite Ableitung von $P_1 z$ verschwindet. Auch hier folgt

$$\|\dot{z} - (\dot{P_1 z})\|_{L_\infty(T_i)} \le \|\ddot{z}\|_{L_\infty(T_i)}\, h.$$

Fassen wir das eben Angeführte zusammen, so erhalten wir

$$\|z - P_1 z\|_{L_\infty(T_i)} \le \|\dot{z} - (\dot{P_1 z})\|_{L_\infty(T_i)}\, h \le \|\ddot{z}\|_{L_\infty(T_i)}\, h^2 \le \|\ddot{z}\|_\infty\, h^2.$$

Da die letzte Abschätzung für alle Intervalle T_i gilt, folgt schließlich

$$\|z - P_1 z\|_\infty \le \|\ddot{z}\|_\infty\, h^2.$$

□

Im Verlauf des letzten Beweises gelang es uns nicht nur die Kernaussagen herzuleiten. Es ergab sich darüber hinaus ein „Nebenprodukt“, welches wir in folgendem Korollar festhalten.

Korollar 3.3.5. *Sei $z \in W_2^2[0,T,\mathbb{R}^n]$, dann gilt*

$$\|z - P_1 z\|_{1,2} \leq c_{2.6}\, \|\ddot{z}\|_2\, h. \tag{3.12}$$

Beweis: Aus dem ersten Teil des letzten Beweises übernehmen wir

$$\begin{aligned} \|z - P_1 z\|^2_{W_2^1(T_i)} &= \|z - P_1 z\|^2_{L_2(T_i)} + \|\dot{z} - (\dot{P_1 z})\|^2_{L_2(T_i)} \\ &\overset{(2.3)}{\leq} 2\max\{1,T^2\}\, \|\dot{z} - (\dot{P_1 z})\|^2_{L_2(T_i)} \\ &\leq 2\max\{1,T^2\}\, \|\ddot{z}\|^2_{L_2(T_i)}\, h^2. \end{aligned}$$

Das Ergebnis folgt durch Summation über alle $i = 0, \ldots, N-1$. □

Die Approximation durch stückweise lineare Interpolation lässt uns Freiheiten in Bezug auf die Wahl der Differenzierbarkeitseigenschaften. Wie folgendes Lemma zeigt, ist einer Verringerung der Forderungen an die betrachtete Funktion durchaus ohne Verlust der gewünschten Konvergenzordnung h^2 möglich. Im Gegensatz zu den Ergebnissen aus Lemma 3.3.4 hängt die Ordnung des Fehlers nun unter Umständen von der verwendeten Norm ab.

Lemma 3.3.6. *Sei $u \in W_\infty^1[0,T,\mathbb{R}^m]$ mit $\dot{u} \in BV[0,T,\mathbb{R}^m]$. Dann gelten folgende Abschätzungen*

$$\int_0^T |u(t) - (P_1 u)(t)|\, dt \leq \mathsf{V}_0^T \dot{u}\, h^2, \qquad \|u - P_1 u\|_2 \leq \mathsf{V}_0^T \dot{u}\, h^{\frac{3}{2}}. \tag{3.13}$$

Beweis: Wir nutzen die Ableitung von $u - P_1 u$ auf den Intervallen T_i, sowie die Identität $u(t_i) = (P_1 u)(t_i)$ für $i = 0, \ldots, N$. Es folgt dann

$$\begin{aligned} \int_0^T |u(t) - (P_1 u)(t)|\, dt &= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |u(t) - (P_1 u)(t)|\, dt \\ &\leq h \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big| \dot{u}(s) - \frac{u(t_{i+1}) - u(t_i)}{h} \Big|\, ds \\ &\leq \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big| \int_{t_i}^{t_{i+1}} \dot{u}(s) - \dot{u}(v)\, dv \Big|\, ds \\ &\leq \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \int_{t_i}^{t_{i+1}} \underbrace{|\dot{u}(s) - \dot{u}(v)|}_{\leq \mathsf{V}_{T_i} \dot{u}}\, dv\, ds \\ &\leq h^2 \sum_{i=0}^{N-1} \mathsf{V}_{T_i} \dot{u} \overset{(3.6)}{=} \mathsf{V}_0^T \dot{u}\, h^2. \end{aligned}$$

Der zweite Teil folgt auf ähnliche Weise durch

$$\begin{aligned}
\int_0^T |u(t) - (P_1 u)(t)|^2 \, dt &= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |u(t) - (P_1 u)(t)|^2 \, dt \\
&\leq h \sum_{i=0}^{N-1} \left(\int_{t_i}^{t_{i+1}} \left| \dot{u}(s) - \frac{u(t_{i+1}) - u(t_i)}{h} \right| ds \right)^2 \\
&\leq \frac{1}{h} \sum_{i=0}^{N-1} \left(\int_{t_i}^{t_{i+1}} \left| \int_{t_i}^{t_{i+1}} \dot{u}(s) - \dot{u}(v) \, dv \right| ds \right)^2 \\
&\leq \frac{1}{h} \sum_{i=0}^{N-1} (\mathsf{V}_{T_i} \dot{u})^2 \, h^4 \\
&\leq h^3 \sum_{i=0}^{N-1} \left(\mathsf{V}_{T_i} \dot{u} \right)^2 \overset{(3.6)}{=} \left(\mathsf{V}_0^T \dot{u} \right)^2 h^3.
\end{aligned}$$

□

Wir fassen an dieser Stelle alle in diesem Abschnitt bewiesenen Fehlerabschätzungen tabellarisch zusammen. Neben den verwendeten Normen führen wir zum Vergleich auch den Ausdruck $|\int_0^T z(t) - (P_j z)(t) \, dt|$ unter $p = 1^*$. Die Voraussetzungen spiegeln sich in der jeweils hergeleiteten Konstante wieder. Das bedeutet, sind die jeweiligen Ausdrücke definiert, so verhält sich der Fehler mit der angegebenen Ordnung.

	P_0		P_1	
p	Konstante	Konv.ord.	Konstante	Konv.ord.
1	$\|\dot{z}\|_1$	h	$\|\ddot{z}\|_1$	h^2
			$\frac{1}{2}\mathsf{V}_0^T \dot{u}$	h^2
2	$\|\dot{z}\|_2$	h	$\|\ddot{z}\|_2$	h^2
2			$\mathsf{V}_0^T \dot{u}$	$h^{\frac{3}{2}}$
∞	$\frac{1}{2}\|\dot{z}\|_\infty$	h	$\|\ddot{z}\|_\infty$	h^2
1^*	$\frac{1}{2}\|\ddot{z}\|_1$	h^2	$\|\ddot{z}\|_1$	h^2
	$\frac{1}{2}\mathsf{V}_0^T \dot{u}$	h^2		

Tabelle 3.1: Konvergenzordnung und Abschätzungskonstanten in Abhängigkeit von der Approximationsgüte.

Wie wir sehen, erhöht sich wie erwartet die Konvergenzordnung durch Verbesserung der Approximationsgüte. Die Werte für $p = 1^*$ nehmen eine Sonderrolle ein, denn in diesem Fall stellt der Ausdruck $\ell(z) = \int z(t) - (P_1 z)(t) \, dt$ ein lineares beschränktes Funktional dar und es gilt darüber hinaus $\ell(p) = 0$ für alle Polynome p vom Grad

höchstens 1. Diese Tatsache nutzt man auch allgemein für $f : \mathbb{R}^m \to \mathbb{R}^n$ und $z \in W_2^k[0,T,\mathbb{R}^n]$ aus. Wird z durch den Operator $\hat{z}$ mit $\hat{p} = 0$ für alle $p \in \mathcal{P}_k$ auf den Wert $\hat{z} \in \mathbb{R}$ abgebildet, so gilt

$$\|z - \hat{z}\|_2 \leq c\,|z|_{k+1,2}\,h^{k+1}$$

mit einer von h und z unabhängigen Konstante c. Für einen Beweis dieser Aussage verweisen wir auf Ciarlet [10] (1987, Kapitel 3.1). Nun erfüllt ℓ die eben genannten Voraussetzungen für $k = 1$ und daher folgt die quadratische Konvergenzordnung. Der Vorteil der Beweismethode aus Ciarlet [10] ist die Anwendungsmöglichkeit auch auf Funktionen mit mehrdimensionalem Definitionsbereich.

3.4 Diskrete Steuerungsprobleme

Für die Bearbeitung unseres Steuerungsproblems (StP) ist die eindeutige Lösbarkeit der Systemgleichung (1.4) notwendig. Den Beweis dafür liefert Satz 2.2.5. Wir schreiben daher die Zuordnungen $u \mapsto z$ bzw. $y \mapsto z$ durch die Operatoren $\mathcal{S} \circ \mathcal{T}$ bzw. $\mathcal{S}$ aus.

Bei der Diskretisierung von (1.4) ersetzen wir den Lösungsoperator $\mathcal{S}$ durch eine geeignete Approximation. Dazu stehen uns verschiedene Mittel zur Verfügung (vgl. Kapitel 4 und 5), die auf verschiedene Art und Weise die Funktion y auf eine Funktion z_h aus einem endlich-dimensionalen Unterraum des $W_2^1[0,T,\mathbb{R}^n]$ abbilden, d.h. wir fassen das diskrete Pendant zu $\mathcal{S}$ auch als Operator $\mathcal{S}_h$ zwischen Funktionenräumen auf. Es bleibt die Frage nach der Diskretisierung der Steuerung offen. Ob diese stattfindet oder nicht und falls ja, in welcher Form, legen wir später unabhängig von $\mathcal{S}_h$ fest. Für uns steht dann die Gleichung $z_h(u) = \mathcal{S}_h(Bu + e)$ zur weiteren Untersuchung bereit. Dabei kennzeichnen wir mit dem Index h, dass es sich um eine Funktion bzw. einen Operator im diskreten Fall handelt.

Auch der adjungierte Operator $\mathcal{S}^*$ ist linear und beschränkt. Wir diskretisieren ihn auf die gleiche Art und Weise wie eben, also

$$p_h^*(u) = \mathcal{S}_h^*(\mathcal{S}\mathcal{T}u - z_d) \qquad \text{oder} \qquad p_h(u) = \mathcal{S}_h^*(\mathcal{S}_h\mathcal{T}u - z_d).$$

Die Schreibweise p_h^* soll beim diskreten adjungierten Zustand verdeutlichen, ob das Argument von $\mathcal{S}_h^*$ eine kontinuierliche oder eine diskrete Funktion, verschoben um z_d, war. Beim Zustand ist dies nicht nötig, da wir nicht auf das Argument verzichten und erkennen daran, welche Form es besitzt. Die konkrete Form der diskreten Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$ hängt von der gewählten Diskretisierungsmethode ab. An dieser Stelle soll die Art der Diskretisierung nicht berücksichtigt werden, denn wir leiten zunächst allgemeine Konvergenzresultate her. Die Forderungen Stabilität und Konsistenz an die diskreten Operatoren bzw. an die Diskretisierungsmethode sind hier Voraussetzungen und müssen im konkreten Fall (mühsam) nachgerechnet werden. In den Kapiteln 4 und 5 stellen wir dann zwei unterschiedliche Diskretisierungsmethoden vor.

Mit den eben eingeführten Bezeichnungen erstellen wir Diskretisierungen des Steuerungsproblems (StP), indem wir die Operatoren $\mathcal{S}$ und $\mathcal{S}^*$ durch ihre Diskretisierungen $\mathcal{S}_h$ und $\mathcal{S}_h^*$ ersetzen. Weiterhin sei $\langle\cdot,\cdot\rangle_{\mathfrak{C}}$ ein Skalarprodukt und die diskreten Operatoren bezüglich dessen adjungiert. Dieses Skalarprodukt verwenden wir ebenfalls in der Definition der diskreten Probleme

$$(\mathrm{StP})_h \qquad \min_{u_h} J_h(u_h) = \min_{u_h} \frac{1}{2}\|\mathcal{S}_h\mathcal{T}u_h - z_d\|_{\mathfrak{C}}^2 + \frac{\nu}{2}\|u_h\|_{\mathfrak{C}}^2$$

unter den Kontrollrestriktionen

$$u_h \in U_h^{ad} \subset U^{ad}.$$

Die zulässige Menge U_h^{ad} ist eine geeignet gewählte, nichtleere, abgeschlossene und konvexe Teilmenge von U^{ad}. Die Gleichheit ist dabei ausdrücklich zugelassen. Die Lösung der Probleme $(\mathrm{StP})_h$ bezeichnen wir mit u_h und zeigen zunächst analog zu Lemma 2.4.1 einen Zusammenhang zwischen dieser Funktion und dem zugehörigen adjungierten Zustand $p_h(u_h) = \mathcal{S}_h^*(\mathcal{S}_h\mathcal{T}u_h - z_d)$.

Lemma 3.4.1. *Das Problem* $(\mathrm{StP})_h$ *besitzt unter den obigen Voraussetzungen sowohl für* $-\infty < a < b < \infty$ *als auch für* $a = -\infty$ *und/oder* $b = \infty$ *eine eindeutig bestimmte Lösung* $u_h \in U_h^{ad}$*. Die Bedingung*

$$\langle B^{\mathsf{T}} p_h(u_h) + \nu u_h, \zeta_h - u_h\rangle_{\mathfrak{C}} \geq 0 \qquad \forall \zeta_h \in U_h^{ad} \tag{3.14}$$

ist notwendig und hinreichend für die Optimalität der Steuerung u_h *mit dem zugehörigen adjungierten Zustand* $p_h(u_h)$*.*

Beweis: Wie im kontinuierlichen Fall (siehe Lemma 2.4.1) sehen wir dem Funktional J_h seine gleichmäßige Konvexität an. Für $-\infty < a$ und $b < \infty$ ist die zulässige Menge U_h^{ad} nichtleer, abgeschlossen, konvex und beschränkt. Somit schließen wir mit Alt [2], Satz 2.5.2, auf eine eindeutige Lösung $u_h \in U_h^{ad}$. Für unbeschränktes U_h^{ad} folgt die Aussage auf Grund der strikten Konvexität der Zielfunktion, denn sie impliziert $J_h(u_h) \to \infty$ falls $\|u_h\| \to \infty$.

Für die Lösung u_h von $(\mathrm{StP})_h$ lautet die notwendige Optimalitätsbedingung

$$\langle J_h'(u_h), \zeta_h - u_h\rangle_{\mathfrak{C}} \geq 0 \qquad \forall \zeta \in U_h^{ad}.$$

Berechnen wir also die Ableitung des Zielfunktionals an der Optimalstelle u_h. Es ist

$$\begin{aligned} J_h'(u_h)(h) &= \left(\frac{\partial \frac{1}{2}\|\mathcal{S}_h\mathcal{T}u_h - z_d\|_2^2}{\partial(\mathcal{S}_h\mathcal{T}u_h - z_d)} \frac{\partial(\mathcal{S}_h\mathcal{T}u_h - z_d)}{\partial(\mathcal{T}u_h)} \frac{\partial \mathcal{T}u_h}{\partial u_h} + \nu \frac{\partial \|u_h\|_2^2}{\partial u_h}\right)(h) \\ &= ((\mathcal{S}_h\mathcal{T}u_h - z_d) \circ \mathcal{S}_h \circ B + \nu u_h)(h) \\ &= \langle B^{\mathsf{T}}\mathcal{S}_h^*(\mathcal{S}_h\mathcal{T}u_h - z_d) + \nu u_h, h\rangle_{\mathfrak{C}} \end{aligned}$$

für beliebiges $h \in U_h^{ad}$ und wegen der Adjungiertheit der Operatoren folgt (3.14). Die Erweiterung auf unbeschränktes U_h^{ad} erfolgt analog dem kontinuierlichen Fall (vgl. Lemma 2.4.1). □

Kapitel 4

Methode der Finiten Elemente

4.1 Motivation

Ausgangspunkt für die Methode der Finiten Elemente ist die Variationsgleichung oder schwache Formulierung (2.1) der Systemgleichung (1.4). Für jede rechte Seite existiert eine eindeutig bestimmte Lösung, welche darüber hinaus äquivalent zur Lösung der Systemgleichung ist. Erster Ansatz für eine Diskretisierung des Problems (StP) ist die Berechnung einer Näherungslösung von (2.1). Ersetzen wir also darin die Bedingung „$\forall v \in W^1_{2,0}[0,T,\mathbb{R}^n]$“ durch „$\forall v_h \in V_{h,0}[0,T,\mathbb{R}^n]$“ mit einem geeignet gewählten endlich-dimensionalem Hilbert-Raum $V_{h,0}[0,T,\mathbb{R}^n]$. Diese Vorgehensweise wird auch *Ritz-Galerkin-Verfahren* genannt. So bekommen wir eine eindeutig bestimmte Lösung $z_h \in V_{h,0}[0,T,\mathbb{R}^n]$ mit dem gleichen Ziel wie im kontinuierliche Fall, die aus der Diskretisierung folgenden diskreten Steuerungsprobleme als allein von der Steuerung abhängig zu betrachten.

4.2 Diskretisierung und Stabilität

Wie im letzten Abschnitt bereits angedeutet, ist es unser Ziel, die Systemgleichung adäquat zu diskretisieren, um daraus eine Diskretisierung des Steuerungsproblems (StP) abzuleiten. Wir ersetzen dazu in der Variationsgleichung (2.1) die Bedingung „$\forall v \in W^1_{2,0}[0,T,\mathbb{R}^n]$“ durch „$\forall v_h \in V_{h,0}[0,T,\mathbb{R}^n]$“, woraus folgende Gleichung resultiert

$$\int_0^T \dot{z}_h(t)^\mathsf{T}\dot{v}_h(t) + z_h(t)^\mathsf{T} A v_h(t)\,dt = \int_0^T (Bu(t)+e(t))^\mathsf{T} v_h(t)\,dt \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n],$$

die wir als *diskrete* oder *diskretisierte Systemgleichung* bezeichnen. Unter Einbeziehung der in (2.2) eingeführten Bilinearform $a(\cdot,\cdot)$ formulieren wir folgendes Steuerungsproblem

$$(\widetilde{\mathrm{StP}})_h \qquad \min_{z_h,u_h} \widetilde{J}_h(z_h,u_h) = \min_{z_h,u_h} \frac{1}{2}\|z_h - z_d\|_2^2 + \frac{\nu}{2}\|u_h\|_2^2$$

unter den Nebenbedingungen

$$a(z_h, v_h) = \int_0^T (Bu_h(t) + e(t))^\mathsf{T} v_h(t)\, dt \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n]$$
$$z_h(0) = z_h(T) = 0$$

und den Kontrollrestriktionen

$$u_h \in U_h^{ad} \subset U^{ad},$$

mit einer abgeschlossenen und konvexen Menge U_h^{ad}, die sich nicht notwendig von U^{ad} unterscheidet. Das Vorgehen bei $(\widetilde{\mathrm{StP}})_h$ ist angelehnt an den kontinuierlichen Fall. So betrachten wir zunächst zur Lösung von (4.1) Existenz- und Eindeutigkeitsfragen und leiten Stabilitätsaussagen her.

Lemma 4.2.1. *Die diskrete Systemgleichung*

$$\int_0^T \dot{z}_h(t)^\mathsf{T} \dot{v}_h(t) + z_h(t)^\mathsf{T} A v_h(t)\, dt = \int_0^T y(t)^\mathsf{T} v_h(t)\, dt \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n] \tag{4.1}$$

besitzt für jedes $y \in L_2[0,T,\mathbb{R}^n]$ eine eindeutig bestimmte Lösung $z_h \in V_{h,0}[0,T,\mathbb{R}^n]$ und es gilt

$$\|z_h\|_2 \le T^2 \|y\|_2. \tag{4.2}$$

Beweis: Dem Beweis liegen die gleichen Überlegungen zu Grunde wie für Satz 2.2.5. Wir geben sie dennoch in etwas anderer Art wieder. Der Ausdruck $\ell : L_2[0,T,\mathbb{R}^n] \to \mathbb{R}$ mit $\ell(\cdot) = \int_0^T y(t)^\mathsf{T} \cdot (t)\, dt$ ist für $y \in L_2[0,T,\mathbb{R}^n]$ ein Skalarprodukt, also auch ein lineares beschränktes Funktional. Es ist bekanntlich $V_{h,0} \subset L_2 \cong (L_2)' \subset (V_{h,0})'$. Daraus folgt, dass die Einschränkung von ℓ auf $V_{h,0}[0,T,\mathbb{R}^n]$ ebenfalls ein lineares beschränktes Funktional ist und es auch in diesem Hilbert-Raum ein eindeutig bestimmtes Element gibt, welches als „fester“ Eintrag im $V_{h,0}$-Skalarprodukt fungiert

$$\exists! z_h \in V_{h,0}[0,T,\mathbb{R}^n] \,:\, \langle z_h, v_h \rangle_{V_{h,0}} = \ell(v_h) \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n].$$

Die symmetrische Bilinearform $a(\cdot,\cdot)$ ist offensichtlich gleichmäßig elliptisch, d.h. es gilt

$$a(z_h, z_h) \ge c\, \|z_h\|_{1,2}^2$$

mit einer von z und h unabhängigen Konstante $c > 0$. Daher ist sie ein Skalarprodukt auf $V_{h,0}[0,T,\mathbb{R}^n]$ und somit folgt die Eindeutigkeitsaussage für die diskrete Systemgleichung.

Für die Ungleichung (4.2) setzen wir in (4.1) $v_h = z_h$ und erhalten

$$\|\dot{z}_h\|_2^2 \le \|y\|_2 \|z_h\|_2,$$

da A positiv semi-definit ist. Mit der Poincaré'schen Ungleichung aus Lemma 2.2.1 und nach beidseitigem Kürzen von $\|z_h\|_2$ folgern wir weiter

$$\|z_h\|_2 \leq T^2 \|y\|_2.$$

Letzteres ist möglich, denn im Fall $z_h = 0$ wäre nichts zu zeigen gewesen. □

Bemerkung: An dieser Stelle versäumen wir nicht auf ein etwas anderes Beweisverfahren hinzuweisen. Die diskrete Systemgleichung (4.1) stellt - analog dem kontinuierlichen Fall - die notwendige Optimalitätsbedingung für das Optimierungsproblem

$$\min_{z_h} \int_0^T \dot{z}_h(t)^\mathsf{T} \dot{z}_h(t) + z_h(t)^\mathsf{T} A z_h(t) - y(t)^\mathsf{T} z_h(t)\, dt$$

mit $z_h \in V_{h,0}[0,T,\mathbb{R}^n]$ und $y \in L_2[0,T,\mathbb{R}^n]$ dar. Die Zielfunktion besitzt quadratische Form

$$a(z_h, z_h) - \ell(y)$$

mit dem beschränkten linearen Funktional $\ell(y) = \int_0^T y(t)^\mathsf{T} z_h(t)\, dt$. Die Bilinearform $a(\cdot,\cdot)$ ist nach Lemma 2.2.4 auf dem Raum $W_2^1[0,T,\mathbb{R}^n]$ elliptisch und demnach auch auf $V_{h,0}[0,T,\mathbb{R}^n]$. Daher besitzt f ein eindeutig bestimmtes Minimum in $V_{h,0}[0,T,\mathbb{R}^n]$ (siehe Alt [2], Satz 2.5.2) und (4.1) somit eine eindeutig bestimmte Lösung. ♢

Ähnlich dem kontinuierlichen Fall leiten wir auch hier das Stabilitätsresultat in anderen Normen her.

Korollar 4.2.2. *Unter den Voraussetzungen und mit den Bezeichnungen aus Lemma 4.2.1 gelten folgende Abschätzungen*

$$\begin{aligned} \|z_h\|_{1,2} &\leq c_{2.6}\, \|y\|_2 \\ \|z_h\|_\infty &\leq T^2\, \|y\|_\infty. \end{aligned} \tag{4.3}$$

Beweis: Wir übernehmen für diesen Fall den Beweis von Korollar 2.2.6. □

Nach der Diskretisierung der Systemgleichung folgen sofort Resultate für die adjungierte Gleichung.

Korollar 4.2.3. *Die diskrete adjungierte Gleichung*

$$\int_0^T \dot{p}_h(t)^\mathsf{T} \dot{v}_h(t) + p_h(t)^\mathsf{T} A v_h(t)\, dt = \int_0^T (z(t) - z_d(t))^\mathsf{T} v_h(t)\, dt \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n] \tag{4.4}$$

besitzt für jedes $z \in L_2[0,T,\mathbb{R}^n]$ genau eine Lösung $p_h \in V_{h,0}[0,T,\mathbb{R}^n]$ und es gilt

$$\|p_h\|_2 \leq T^2 \|z - z_d\|_2 \leq T^2 (\|z\|_2 + \|z_d\|_2).$$

Ist z die Lösung von (1.4) für $y \in L_\infty[0,T,\mathbb{R}^n]$, so gilt

$$\begin{aligned} \|p_h\|_{1,2} &\leq c_{2.6}\, \|z - z_d\|_2 \leq c_{2.6}^2\, (\|y\|_2 + \|z_d\|_2) \\ \|p_h\|_\infty &\leq T^{\frac{3}{2}}\, \|z_h - z_d\|_2 \leq T^2 c_{2.6}\, (\|y\|_\infty + \|z_d\|_\infty). \end{aligned} \tag{4.5}$$

Beweis: siehe Beweis zu Lemma 4.2.1 und Korollar 4.2.2. □

Bemerkung: Vergleichen wir die abschätzungen für z_h und p_h mit den im kontinuierlichen Fall bewiesenen Ungleichungen (2.5) bzw. (2.13), so stellen wir Übereinstimmung in Gestalt und sogar in der Abschätzungskonstante fest. ◊

Für die weiteren Betrachtungen benötigen wir analog zu Lemma 2.2.7 ein etwas anderes Ergebnis.

Lemma 4.2.4. *Seien $y \in L_2[0,T,\mathbb{R}^n]$ und $Y(t) = \int_0^t y(s)\,ds$ für $t \in [0,T]$. Die Funktion $z_h \in V_{h,0}[0,T,\mathbb{R}^n]$ sei ferner die Lösung der diskreten Systemgleichung (4.1) für y. Dann gilt*

$$\|z_h\|_\infty \le c_{2.7}\,\|Y\|_1. \tag{4.6}$$

Beweis: Wir betrachten zunächst für $A = 0$ die diskrete Systemgleichung (4.1) mit deren Lösung $z_h^0 \in V_{h,0}[0,T,\mathbb{R}^n]$, d.h.

$$\int_0^T (\dot z_h^0(t) + Y(t))^\mathsf{T} \dot v_h(t)\,dt = 0 \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n].$$

Setzen wir für das beliebige v_h nacheinander die Basisfunktionen $v_h^{(j)}$ ein, so erhalten wir für $j = 1,\dots,N-1$

$$-\frac{z_h^0(t_{j+1}) - 2z_h^0(t_j) + z_h^0(t_{j-1})}{h} + \frac{1}{h}\int_{t_{j-1}}^{t_j} Y(t)\,dt - \frac{1}{h}\int_{t_j}^{t_{j+1}} Y(t)\,dt = 0.$$

Mit den Bezeichnungen $\widetilde Y_j = \int_{t_j}^{t_{j+1}} Y(t)\,dt$ für $j = 0,\dots,N-1$ ergibt sich daraus

$$-\frac{z_h^0(t_{j+1}) - 2z_h^0(t_j) + z_h^0(t_{j-1})}{h} + \frac{\widetilde Y_j - \widetilde Y_{j-1}}{h} = 0, \qquad j = 1,\dots,N-1.$$

Die linke Seite ist die Ableitung der stückweise linearen Funktion w_h, definiert durch die Vektoren

$$w_h(t_j) = z_h^0(t_{j+1}) - z_h^0(t_j) + \widetilde Y_j, \qquad j = 0,\dots,N-1.$$

Da sie auf dem gesamten Intervall $[0,T]$ verschwindet, ist $w_h \equiv q$ mit einer Konstante q, also auch

$$z_h^0(t_{j+1}) - z_h^0(t_j) + \widetilde Y_j = q, \qquad j = 0,\dots,N-1.$$

Weiterhin gilt

$$0_n = z_h(T) = h\sum_{i=0}^{N-1} \frac{z_h^0(t_{i+1}) - z_h^0(t_i)}{h} = Nq - \sum_{i=0}^{N-1} \widetilde Y_i,$$

woraus

$$q = \frac{1}{N}\sum_{i=0}^{N-1} \widetilde{Y}_i = \frac{1}{N}\int_0^T Y(t)\,dt$$

folgt.

Schließlich erhalten wir auf Grund von $z_h(0) = 0_n$

$$\begin{aligned} z_h^0(t_j) &= h\sum_{i=0}^{j-1} \frac{z_h^0(t_{i+1}) - z_h^0(t_i)}{h} = jq - \sum_{i=0}^{j-1} \widetilde{Y}_i \\ &= \frac{j}{N}\sum_{i=0}^{N-1} \widetilde{Y}_i - \sum_{i=0}^{j-1} \widetilde{Y}_i \\ &= (\frac{j}{N} - 1)\sum_{i=0}^{j-1} \widetilde{Y}_i + \frac{j}{N}\sum_{i=j}^{N-1} \widetilde{Y}_i, \qquad j = 1, \ldots, N-1. \end{aligned}$$

Gehen wir zum Betrag über, so folgt

$$|z_h^0(t_j)| \le (1 - \frac{j}{N})\sum_{i=0}^{j-1} |\widetilde{Y}_i| + \frac{j}{N}\sum_{i=j}^{N-1} |\widetilde{Y}_i|, \qquad j = 1, \ldots, N-1,$$

also wegen $z_h^0 \in V_{h,0}[0, T, \mathbb{R}^n]$ auch

$$\|z_h^0\|_\infty \le \sum_{i=0}^{N-1} |\widetilde{Y}_i| = \sum_{i=0}^{N-1} \Big| \int_{t_i}^{t_{i+1}} Y(t)\,dt \Big| \le \int_0^T |Y(t)|\,dt.$$

Weiterhin ist $z_h - z_h^0$ die Lösung der Randwertaufgabe

$$\begin{aligned} a(z_h, v_h) &= -\int_0^T z_h^0(t)^\mathsf{T} A v_h(t)\,dt \qquad \forall v_h \in V_{h,0}[0, T, \mathbb{R}^n] \\ z_h(0) &= z_h(T) = 0_n. \end{aligned}$$

Damit folgt unter Beachtung von Korollar 4.2.2

$$\begin{aligned} \|z_h\|_\infty \le \|z_h - z_h^0\|_\infty + \|z_h^0\|_\infty &\overset{(4.3)}{\le} T^2\|A\|\,\|z_h^0\|_\infty + \|z_h^0\|_\infty \\ &\le (1 + T^2\|A\|)\,\|Y\|_1 \le c_{2.7}\,\|Y\|_1. \end{aligned}$$

□

Nachdem wir die eindeutige Lösbarkeit der diskreten Systemgleichung (4.1) und der diskreten adjungierten Gleichung (4.4) nachgewiesen haben, kommen wir auf die Schreibweisen aus Abschnitt 3.4 zurück und deklarieren die Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$ genauer. Demnach ist $\mathcal{S}_h : L_2[0, T, \mathbb{R}^n] \to V_{h,0}[0, T, \mathbb{R}^n]$ diejenige Abbildung, welche einer

Funktion $y \in L_2[0,T,\mathbb{R}^n]$ die eindeutig bestimmte Lösung von (4.1) zuweist. Eine analoge Definition ergibt sich für den Operator $\mathcal{S}_h^*$. Beide Abbildungen sind offensichtlich linear und nach Lemma 4.2.1 unabhängig von h nach oben beschränkt. Darüber hinaus sind sie adjungiert, was wir in folgendem Lemma nachweisen.

Lemma 4.2.5. *Die durch die Eindeutigkeitsaussage in Lemma 4.2.1 und Korollar 4.2.3 definierten Operatoren S_h und S_h^* sind auf dem Raum $L_2[0,T,\mathbb{R}^n]$ adjungiert, d.h.*

$$\langle w, \mathcal{S}_h v\rangle = \langle \mathcal{S}_h^* w, v\rangle \qquad \forall v, w \in L_2[0,T,\mathbb{R}^n].$$

Beweis: Wir erinnern zunächst an die Definition der beiden Operatoren. Sie sind formal gleich und werden nur für anschauliche Zwecke unterschieden. Daher genügt es die Selbstadjungiertheit von $\mathcal{S}_h$ zu zeigen. Mit den Bezeichnungen $z_h(v) = \mathcal{S}_h v$ und $z_h(w) = \mathcal{S}_h w$ setzen wir in der diskreten Systemgleichung (4.1) zunächst $v_h = z_h(w)$ und anschließend $v_h = z_h(v)$. Dann folgt auf Grund der Symmetrie der Bilinearform $a(\cdot,\cdot)$

$$\langle v, z_h(w)\rangle = a(z_h(v), z_h(w)) = a(z_h(w), z_h(v)) = \langle w, z_h(v)\rangle.$$

□

Mit dem Wissen über die eindeutige Lösbarkeit der diskreten Systemgleichung wandeln wir $(\widetilde{\mathrm{StP}})_h$ in ein Steuerungsproblem um, dessen Zielfunktional allein von der Steuerung u_h abhängt. Die Kontrollrestriktionen formulieren wir mit Hilfe der später noch genauer festzulegenden, nichtleeren, abgeschlossenen und konvexen Menge $U_h^{ad} \subset U^{ad}$. So ist dann

$$(\mathrm{StP})_h \qquad \min_{u_h} J_h(u_h) = \min_{u_h} \frac{1}{2}\|\mathcal{S}_h\mathcal{T}u_h - z_d\|_2^2 + \frac{\nu}{2}\|u_h\|_2^2, \qquad u_h \in U_h^{ad}.$$

Die Existenz und Eindeutigkeit einer Lösung von $(\mathrm{StP})_h$ folgt aus dem allgemeinen Resultat in Lemma 3.4.1. Dort setzen wir $\langle\cdot,\cdot\rangle_{\mathfrak{C}} = \langle\cdot,\cdot\rangle$ und erhalten die notwendige Optimalitätsbedingung

$$\langle J_h'(u_h), \zeta - u_h\rangle = \langle B^{\mathsf{T}}\mathcal{S}_h^*(\mathcal{S}_h(Bu_h + e) - z_d), \zeta - u_h\rangle \geq 0 \qquad \forall \zeta \in U_h^{ad}. \tag{4.7}$$

Wir haben dabei die Adjungiertheit der diskreten Operatoren bezüglich des diskreten Skalarprodukts ausgenutzt.

4.3 Konvergenz

Nach Betachtungen über die Existenz und Eindeutikeit einer Lösung der diskreten Systemgleichung wenden wir uns nun dem Fehler zu, der beim diskretisieren entsteht. Wir untersuchen den Fehler zwischen der exakten Lösung und ihrer Approximation sowohl punktweise als auch im quadratischen Mittel. Obwohl letzteres aus dem ersten Fall folgt, geben wir dennoch die Herleitung an, denn die Beweise sind mit wenig Aufwand auf Funktionen mit mehrdimensionalem Wertebereich erweiterbar.

4.3.1 Betrachtungen bezüglich der L_2-Norm

Als Zwischenschritt in einem der folgenden Beweise benötigen wir eine Aussage, die unter dem Namen *Céa-Lemma* bekannt ist. Wir zeigen sie auf Grund ihrer Bekanntheit als eigenständiges Ergebnis und erweitern die Aussage auf vektorwertige Funktionen.

Lemma 4.3.1. *Seien $z \in W^2_{2,0}[0,T,\mathbb{R}^n]$ bzw. $z_h \in V_{h,0}[0,T,\mathbb{R}^n]$ die Lösungen der Systemgleichung* (1.4) *bzw. der diskreten Systemgleichung* (4.1) *für die Funktion $y \in L_2[0,T,\mathbb{R}^n]$. Darüber hinaus seien die Konstanten, $c^u_{2.4} = \left(2\max\{1,T^2\}\right)^{-1}$ und $c^o_{2.4} = \max\{1,\|A\|\}$, aus Lemma 2.2.4 übernommen. Dann gilt*

$$\|z - z_h\|_{1,2} \leq \sqrt{c^o_{2.4} c^u_{2.4}} \inf_{v_h \in V_{h,0}} \|z - v_h\|_{1,2}. \tag{4.8}$$

Seien $p \in W^2_{2,0}[0,T,\mathbb{R}^n]$ und $p^*_h \in V_{h,0}[0,T,\mathbb{R}^n]$ die Lösungen der adjungierten Gleichung (2.10) *und ihrer Diskretisierung* (4.4) *für die Funktion $z \in L_2[0,T,\mathbb{R}^n]$, dann gilt*

$$\|p - p^*_h\|_{1,2} \leq \sqrt{c^o_{2.4} c^u_{2.4}} \inf_{v_h \in V_{h,0}} \|p - v_h\|_{1,2}.$$

Bemerkung: Die Abschätzungen sind sinnvoll, denn für die Konstante gilt

$$\sqrt{c^o_{2.4} c^u_{2.4}} = \sqrt{2\max\{1,\|A\|\}}\max\{1,T\} \geq 1.$$

$\Diamond$

Beweis: Betrachten wir die Variationsgleichung (2.1) und die diskrete Systemgleichung (4.1) jeweils für die Funktion y mit den zugehörigen Lösungen z und z_h

$$\begin{aligned} a(z,v) &= \int_0^T y(t)^\mathsf{T} v(t)\,dt && \forall v \in W^1_{2,0}[0,T,\mathbb{R}^n], \\ a(z_h,v_h) &= \int_0^T y(t)^\mathsf{T} v_h(t)\,dt && \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n]. \end{aligned}$$

Wir schränken die erste Gleichung auf den Raum $V_{h,0}[0,T,\mathbb{R}^n]$ ein, subtrahieren davon die zweite Gleichung und erhalten so

$$a(z - z_h, v_h) = 0 \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n].$$

Dann verwenden wir die Elliptizitätsabschätzung aus Lemma 2.2.4 und schreiben für eine beliebige Funktion $v_h \in V_{h,0}[0,T,\mathbb{R}^n]$

$$\begin{aligned} c^u_{2.4}\,\|z - z_h\|^2_{1,2} \leq a(z - z_h, z - z_h) &= a(z - z_h, z - v_h) + \underbrace{a(z - z_h, v_h - z_h)}_{=0} \\ &\leq c^o_{2.4}\,\|z - z_h\|_{1,2}\|z - v_h\|_{1,2}. \end{aligned}$$

Der Fall $\|z - z_h\|_{1,2} = 0$ ist trivial, also kürzen wir und erhalten

$$\|z - z_h\|_{1,2} \leq c_{2.4}^o c_{2.4}^u \|z - v_h\|_{1,2} \qquad \forall v_h \in V_{h,0}[0, T, \mathbb{R}^n].$$

Die linke Seite hängt nicht von v_h ab, somit gehen wir rechts zum Infimum über und erhalten die erste Aussage des Lemmas in einer etwas schwächeren Form.

Eine Verbesserung der Abschätzungskonstante ist möglich, denn im bisherigen Beweis nutzten wir die Symmetrie der Matrix A nicht aus. Das holen wir jetzt nach, denn wir erhalten auf diese Weise auch eine interessante Interpretation von z_h. Die Gleichung $a(z - z_h, v_h) = 0$ gilt für alle Elemente des Raumes $V_{h,0}[0, T, \mathbb{R}^n]$, daher ist z_h die Projektion von z auf diesen Unterraum bezüglich der Norm $\|\cdot\|_a = \sqrt{a(\cdot, \cdot)}$. Damit ist z_h das Element aus $V_{h,0}[0, T, \mathbb{R}^n]$ mit dem kleinsten Abstand zu z, gemessen in der Norm $\|\cdot\|_a$, also

$$\|z - z_h\|_a = \min_{v_h \in V_{h,0}} \|z - v_h\|_a.$$

Mit Lemma 2.2.4 folgt abschließend

$$\|z - z_h\|_{1,2} = \sqrt{c_{2.4}^o c_{2.4}^u} \min_{v_h \in V_{h,0}} \|z - v_h\|_{1,2}.$$

Der Beweis für die adjungierten Zustände läuft analog. □

Bemerkung: Tatsächlich veröffentlichte Céa [8] (1964) zunächst den Fall einer symmetrischen Bilinearform. Den nicht-symmetrischen Fall finden wir zuerst in Birkhoff, Schultz, Varga [6] (1968). ◊

Lemma 4.3.2. *Unter den Voraussetzungen aus Lemma 4.3.1 gelten folgende Fehlerabschätzungen*

$$\begin{aligned} \|z - z_h\|_2 &\leq c_{2.6}^2 c_{2.7}^2 c_{2.4} \|y\|_2 \, h^2 \\ \|p - p_h\|_2 &\leq c_{4.9} \left(\|y\|_2 + \|z_d\|_2\right) h^2. \end{aligned} \tag{4.9}$$

Die Konstante $c_{4.9} = c_{2.6}^3 c_{2.7}^2 c_{2.4}$, mit der Abkürzung $c_{2.4} = \sqrt{{c_{2.4}^o}^3 c_{2.4}^u}$, ist von h und y unabhängig.

Beweis: Wir benutzen ein Dualitätsargument, welches uns im $L_2[0, T, \mathbb{R}^n]$ als Hilbert-Raum zur Verfügung steht. In der Literatur ist dieses Verfahren als *Aubin-Nitsche-Lemma* bekannt. Wir verweisen auf den Anhang, Lemma A.3.7, und die Originalveröffentlichungen Aubin [5] (1967) und Nitsche [17] (1968) für allgemeinere Fassungen der Aussage.

Analog zu Lemma 4.3.1 erhalten wir zunächst die Gleichung

$$a(z - z_h, v_h) = 0 \qquad \forall v_h \in V_{h,0}[0, T, \mathbb{R}^n].$$

Die Funktion $z - z_h$ stammt aus dem Hilbert-Raum $L_2[0,T,\mathbb{R}^n]$, daher definieren wir ein lineares beschränktes Funktional $\ell(v) : L_2[0,T,\mathbb{R}^n] \to \mathbb{R}$ durch $\ell(v) = \langle z - z_h, v\rangle$ und es gilt mit $c_{2.4}^o = \max\{1, \|A\|\}$ aus Lemma 2.2.4 für alle $v_h \in V_{h,0}[0,T,\mathbb{R}^n]$

$$\begin{aligned}
\|z - z_h\|_2 = \|\ell\| &= \sup_{v\in L_2} \frac{\langle v, z - z_h\rangle}{\|v\|_2} \\
&= \sup_{v\in L_2} \frac{a(z(v), z - z_h)}{\|v\|_2} \\
&= \sup_{v\in L_2} \frac{a(z(v) - v_h, z - z_h)}{\|v\|_2} \\
&\le c_{2.4}^o \sup_{v\in L_2} \frac{\|z(v) - v_h\|_{1,2}\|z - z_h\|_{1,2}}{\|v\|_2} \\
&= c_{2.4}^o \, \|z - z_h\|_{1,2} \sup_{v\in L_2}\Big\{ \frac{\|z(v) - v_h\|_{1,2}}{\|v\|_2}\Big\},
\end{aligned}$$

wobei wir das Supremum stets nur über alle Funktionen aus dem Raum bilden, die nicht äquivalent zur Nullfunktion sind. Für das beliebige $v_h \in V_{h,0}[0,T,\mathbb{R}^n]$ setzen wir die Diskretisierung von $z(v)$ im Raum $V_{h,0}[0,T,\mathbb{R}^n]$ ein. Für den ersten Faktor verwenden wir Lemma 4.3.1 und schätzen das dort vorkommende Infimum über alle $v_h \in V_{h,0}[0,T,\mathbb{R}^n]$ ebenfalls mit der Diskretisierung von z nach oben ab. Es folgt

$$\begin{aligned}
\|z - z_h\|_2 &\le c_{2.4}^o \, \|z - z_h\|_{1,2} \sup_{v\in L_2}\Big\{ \frac{\|z(v) - v_h\|_{1,2}}{\|v\|_2}\Big\} \\
&\overset{(4.8)}{\le} c_{2.4} \, \|z - P_1 z\|_{1,2} \sup_{v\in L_2}\Big\{ \frac{\|z(v) - P_1 z(v)\|_{1,2}}{\|v\|_2}\Big\} \\
&\overset{(3.12)}{\le} c_{2.6}^2 c_{2.4} \, |z|_{2,2} \sup_{v\in L_2}\Big\{ h\frac{|z(v)|_{2,2}}{\|v\|_2}\Big\} h \\
&\overset{(2.7)}{\le} c_{2.6}^2 c_{2.4} \, |z|_{2,2} \sup_{v\in L_2} \frac{c_{2.7}\|v\|_2}{\|v\|_2} \, h^2 \\
&= c_{2.6}^2 c_{2.7} c_{2.4} \, |z|_{2,2} \, h^2 \\
&\le c_{2.6}^2 c_{2.7}^2 c_{2.4} \, \|y\|_2 \, h^2.
\end{aligned}$$

Der Beweis der zweiten Aussage läuft ähnlich ab, allerdings betrachten wir wie schon in Lemma 4.3.1 die Funktion p_h^* und definieren das Funktional $\ell : L_2[0,T,\mathbb{R}^n] \to \mathbb{R}$

durch $\ell(v) = \langle p - p_h^*, v\rangle$. Dann schreiben wir wieder für ein beliebiges $v_h \in V_{h,0}[0,T,\mathbb{R}^n]$

$$\begin{aligned}
\|p - p_h^*\|_2 = \|\ell\| &= \sup_{v\in L_2} \frac{\langle v, p - p_h^*\rangle}{\|v\|_2} \\
&= \sup_{v\in L_2} \frac{a(z(v), p - p_h^*)}{\|v\|_2} \\
&= \sup_{v\in L_2} \frac{a(z(v) - v_h, p - p_h^*)}{\|v\|_2} \\
&\le c_{2.4}^o \sup_{v\in L_2} \frac{\|z(v) - v_h\|_{1,2}\|p - p_h^*\|_{1,2}}{\|v\|_2} \\
&= c_{2.4}^o \|p - p_h^*\|_{1,2} \sup_{v\in L_2}\Big\{ \frac{\|z(v) - v_h\|_{1,2}}{\|v\|_2}\Big\}
\end{aligned}$$

und beachten wieder die Einschränkung bei der Bildung des Supremums. Jetzt schließen wir ebenfalls mit Lemma 4.3.1 und setzen jeweils für das beliebige v_h die Diskretisierungen von p bzw. z ein.

$$\begin{aligned}
\|p - p_h^*\|_2 &\le c_{2.4} \|p - P_1 p\|_{1,2} \sup_{v\in L_2}\Big\{ \frac{\|z(v) - P_1 z(v)\|_{1,2}}{\|v\|_2}\Big\} \\
&\overset{(3.12)}{\le} c_{2.6}^2 c_{2.4} |p|_{2,2} \sup_{v\in L_2}\Big\{ h\frac{|z(v)|_{2,2}}{\|v\|_2}\Big\} h \\
&\le c_{2.6}^2 c_{2.7} c_{2.4} |p|_{2,2} h^2 \\
&\le c_{2.6}^2 c_{2.7}^2 c_{2.4} \max\{1, T^2\} (\|y\|_2 + \|z_d\|_2) h^2.
\end{aligned}$$

Außerdem fassen wir $p_h^* - p_h \in V_{h,0}[0,T,\mathbb{R}^n]$ als Lösung der diskreten adjungierten Gleichung (4.4) zu $z - z_h \in L_2[0,T,\mathbb{R}^n]$ auf und schätzen mit Lemma 4.2.1 ab

$$\|p_h^* - p_h\|_2 \le T^2 \|z - z_h\|_2 \le T^2 c_{2.6}^2 c_{2.7}^2 c_{2.4} \|y\|_2 h^2.$$

Schließlich benutzen wir die Dreiecks-Ungleichung und erhalten

$$\begin{aligned}
\|p - p_h\|_2 &\le \|p - p_h^*\|_2 + \|p_h^* - p_h\|_2 \\
&\le c_{2.6}^2 c_{2.7}^2 c_{2.4} \max\{1, T^2\} \big(\|y\|_2 + \|z_d\|_2 + \|y\|_2\big) h^2 \\
&\le c_{2.6}^3 c_{2.7}^2 c_{2.4} (\|y\|_2 + \|z_d\|_2) h^2.
\end{aligned}$$

□

4.3.2 Betrachtungen bezüglich der L_∞-Norm

Bis jetzt standen Konvergenzergebnisse mit dem Abstandsmaß $\|\cdot\|_2$ im Mittelpunkt. Nun ist es an der Zeit auch Resultate in der Norm $\|\cdot\|_\infty$ herzuleiten. Dazu erinnern wir an die diskrete Systemgleichung (4.1), worin die Funktion y ein beliebiges Element

aus dem $L_2[0,T,\mathbb{R}^n]$ darstellt. Auf Grund der Linearität der Gleichung reicht es aus, die Identität für alle Basisfunktionen von $V_{h,0}[0,T,\mathbb{R}^n]$ zu fordern, also

$$\int_0^T \dot{z}_h(t)\dot{v}_h^{(j)}(t) + Az_h(t)v_h^{(j)}(t)\,dt = \int_0^T y(t)v_h^{(j)}(t)\,dt, \qquad j = 1,\ldots,N-1,$$

mit den Funktionen $v_h^{(j)}$ aus Abschnitt 3.1. Die verwendete Vektorschreibweise mit eindimensionaler Basisfunktion erklärten und rechtfertigten wir bereits bei deren Einführung in besagtem Abschnitt. Da z_h aus dem Raum $V_{h,0}[0,T,\mathbb{R}^n]$ stammt, existiert eine Darstellung dieser Funktion als Linearkombination der Basisfunktionen, also $z_h = \sum_{j=1}^{N-1} \beta_j v_h^{(j)}$. Es ist offensichtlich, dass $z_h(t_j) = \beta_j$ gilt und somit folgt für die linke Seite

$$\begin{aligned} a(z_h, v_h) &= \int_0^T \dot{z}_h(t)\dot{v}_h^{(j)}(t) + Az_h(t)v_h^{(j)}(t)\,dt \\ &= \int_0^T \left(\sum_{l=1}^{N-1} \beta_l \dot{v}_h^{(l)}(t)\right)\dot{v}_h^{(j)}(t) + A\left(\sum_{l=1}^{N-1} \beta_l v_h^{(l)}(t)\right)v_h^{(j)}(t)\,dt \\ &= -\frac{1}{h}(\beta_{j+1} - 2\beta_j + \beta_{j-1}) + \frac{h}{6}A(\beta_{j+1} + 4\beta_j + \beta_{j-1}). \end{aligned}$$

Für Details der Berechnung verweisen wir an dieser Stelle auf Formel (3.2) und deren Herleitung.

Nun setzen wir die exakte Lösung z ebenfalls auf der linken Seite der diskreten Systemgleichung ein. Es ist mit $z_j = z(t_j)$ für $j = 0,\ldots,N$

$$\begin{aligned} a(z, v_h^{(j)}) &= \int_0^T \dot{z}(t)\dot{v}_h^{(j)}(t) + Az(t)v_h^{(j)}(t)\,dt \\ &= \frac{1}{h}\int_{t_{j-1}}^{t_j} \dot{z}(t)\,dt + \frac{1}{h}\int_{t_j}^{t_{j+1}} \dot{z}(t)\,dt + \int_0^T Az(t)v_h^{(j)}(t)\,dt \\ &= -\frac{1}{h}(z_{j+1} - 2z_j + z_{j-1}) + \int_0^T Az(t)v_h^{(j)}(t)\,dt. \end{aligned}$$

Betrachten wir den Fehler $z - z_h$, so folgt (siehe auch Beweis zu Lemma 4.3.1)

$$a(z - z_h, v_h) = 0 \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n].$$

Also ergibt sich zusammen mit den bisherigen Betrachtungen und $\gamma_i = z(t_i) - z_h(t_i) = z_i - \beta_i$ für $i = 0,\ldots,N$ folgende Gleichungen

$$\begin{aligned} a(z - z_h, v_h^{(j)}) = -\frac{1}{h}(\gamma_{j+1} - 2\gamma_j + \gamma_{j-1}) &+ \int_0^T Az(t)v_h^{(j)}(t)\,dt \\ &- \frac{h}{6}A(\beta_{j+1} + 4\beta_j + \beta_{j-1}) = 0, \qquad j = 1,\ldots,N-1. \end{aligned}$$

Daraus folgt nach Definition der Vektoren $\psi_j \in \mathbb{R}^n$ durch

$$\psi_j = -\frac{1}{h}\int_0^T Az(t)v_h^{(j)}(t)\,dt + \frac{1}{6}A(z_{j+1} + 4z_j + z_{j-1}), \qquad j = 1, \ldots, N-1$$

und Beachtung der Randwertvorgaben an z und z_h das Gleichungssystem

$$\begin{aligned} -\frac{\gamma_{j+1} - 2\gamma_j + \gamma_{j-1}}{h^2} + \frac{1}{6}A(\gamma_{j+1} + 4\gamma_j + \gamma_{j-1}) = \psi_j, \qquad j = 1, \ldots, N-1 \\ \gamma_0 = \gamma_N = 0_n. \end{aligned} \tag{4.10}$$

Dass eine Lösung dieses Gleuichungssystems existiert ist auf Grund der Herleitung offensichtlich. Über deren Eindeutigkeit und den Zusammenhang zur rechten Seite gibt folgendes Lemma Auskunft.

Lemma 4.3.3. *Sei $\gamma = (\gamma_0^\mathsf{T}, \ldots, \gamma_N^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{(N+1)n}$ Lösung des Gleichungssystems* (4.10). *Falls mit einem $0 < c < 1$ die Bedingung $h \leq \sqrt{(1-c)\frac{6}{\|A\|}}$ erfüllt ist, so gelten die Ungleichungen*

$$\|\gamma\|_\infty \leq \sqrt{T}\|\gamma_\mathsf{h}\|_h \leq \frac{T}{c}\,h\sum_{j=1}^{N-1}|\psi_j|.$$

Beweis: Zunächst erinnern wir an die Schreibweisen $(\gamma_j)_\mathsf{h} = \frac{\gamma_{j+1}-\gamma_j}{h}$ und $(\gamma_j)_\mathsf{h} = \frac{\gamma_j - \gamma_{j-1}}{h}$ sowie ihre Hintereinanderausführung $(\gamma_j)_{\mathsf{h}\bar{\mathsf{h}}}$. Wir multiplizieren in (4.10) jede Gleichung mit dem jeweiligen γ_j und summieren über alle $j = 1, \ldots, N-1$, also

$$\sum_{j=1}^{N-1} -\gamma_j^\mathsf{T}(\gamma_j)_{\mathsf{h}\bar{\mathsf{h}}} + \frac{1}{6}(\gamma_{j+1} + 4\gamma_j + \gamma_{j-1})^\mathsf{T}A\gamma_j = \sum_{j=1}^{N-1}\psi_j^\mathsf{T}\gamma_j.$$

Nach dem Prinzip der partiellen Summation gilt für beliebige $\phi_i, \mu_i \in \mathbb{R}^n$, $i = 0, \ldots, N$, mit $\phi_0 = \mu_0 = \phi_N = \mu_N = 0_n$ die Gleichung

$$\sum_{i=1}^{N-1}\mu_i^\mathsf{T}(\phi_i)_{\mathsf{h}\bar{\mathsf{h}}} = \sum_{i=0}^{N-1}(\mu_i)_\mathsf{h}^\mathsf{T}(\phi_i)_\mathsf{h} = \sum_{i=1}^{N-1}\phi_i^\mathsf{T}(\mu_i)_{\mathsf{h}\bar{\mathsf{h}}}.$$

Demnach folgt wegen $\gamma_0 = \gamma_N = 0_n$

$$\begin{aligned} \sum_{j=1}^{N-1}\gamma_j^\mathsf{T}(\gamma_j)_{\mathsf{h}\bar{\mathsf{h}}} + \frac{1}{6}(\gamma_{j+1} - 2\gamma_j + \gamma_{j-1})^\mathsf{T}A\gamma_j + \gamma_j^\mathsf{T}A\gamma_j = \\ \sum_{j=0}^{N-1}(\gamma_j)_\mathsf{h}^\mathsf{T}(\gamma_j)_\mathsf{h} - \frac{h^2}{6}(\gamma_j)_\mathsf{h}^\mathsf{T}A(\gamma_j)_\mathsf{h} + \gamma_j^\mathsf{T}A\gamma_j = \sum_{j=1}^{N-1}\psi_j^\mathsf{T}\gamma_j \end{aligned}$$

oder mit Hilfe des diskreten Skalarprodukts $\langle \cdot, \cdot \rangle_h$ in der Form

$$\langle \gamma_\mathsf{h}, \gamma_\mathsf{h} \rangle_h - \frac{h^2}{6} \langle \gamma_\mathsf{h}, A\gamma_\mathsf{h} \rangle_h + \langle \gamma, A\gamma \rangle_h = \langle \psi, \gamma \rangle_h.$$

Daraus folgt

$$\langle \gamma_\mathsf{h}, (I_n - \frac{h^2}{6} A)\gamma_\mathsf{h} \rangle_h \leq \langle \psi, \gamma \rangle_h,$$

denn A ist positiv semi-definit. Wählen wir h klein genug, so ist $I_n - \frac{h^2}{6}A$ positiv definit und wir schätzen weiter ab

$$\lambda_{\min}(I_n - \frac{h^2}{6} A)\, \|\gamma_\mathsf{h}\|_h^2 \leq \|\gamma\|_\infty\, h \sum_{j=1}^{N-1} |\psi_j| \overset{(3.5)}{\leq} \sqrt{T} \|\gamma_\mathsf{h}\|_h\, h \sum_{j=1}^{N-1} |\psi_j|.$$

Es bleibt nun noch die Bestimmung des kleinsten Eigenwerts $\lambda_{\min}$ der Matrix $I - \frac{h^2}{6}A$. Nach dem Satz von Gerschgorin (siehe Anhang, Satz A.3.3) gilt für die Eigenwerte λ_i, $i = 1, \ldots, n$, die Abschätzungen

$$|\lambda_i - (1 - \frac{h^2}{6} A_{i,i})| \leq \frac{h^2}{6} \sum_{\substack{j=1 \\ j \neq i}}^{n} |A_{i,j}| \Longrightarrow 1 - \frac{h^2}{6}\|A\| \leq \lambda_i \leq \frac{h^2}{6}\|A\| + 1.$$

Wählen wir ein c mit $0 < c < 1$ und fordern vom kleinsten Eigenwert $c \leq \lambda_{\min}$, so folgt daraus die Forderung

$$h \leq \sqrt{(1-c)\frac{6}{\|A\|}}$$

an die Schrittweite h. Wählen wir die Schrittweite geeignet, so ist die Matrix $I - \frac{h^2}{6}A$ positiv definit und wir erhalten

$$\|\gamma_\mathsf{h}\|_h^2 \leq \frac{1}{c}\|\gamma\|_\infty\, h \sum_{j=1}^{N-1} |\psi_j| \overset{(3.5)}{\leq} \frac{\sqrt{T}}{c} \|\gamma_\mathsf{h}\|_h\, h \sum_{j=1}^{N-1} |\psi_j|.$$

Den Fall $\gamma_\mathsf{h} = 0$ schließen wir aus, denn auf Grund der Randwertvorgabe $\gamma_0 = 0_n$ folgte daraus $\gamma = 0$ und es wäre nichts zu zeigen gewesen. Daher ist das Kürzen des Terms $\|\gamma_\mathsf{h}\|_h$ auf beiden Seiten zulässig und es folgt für $h \leq \sqrt{(1-c)\frac{6}{\|A\|}}$

$$\|\gamma_\mathsf{h}\|_h \leq \frac{\sqrt{T}}{c}\, h \sum_{j=1}^{N-1} |\psi_j|.$$

Die zweite Ungleichung folgt mit $\gamma_0 = 0_n$ und Lemma 3.2.3, nämlich

$$\|\gamma\|_\infty \overset{(3.5)}{\leq} \sqrt{T}\|\gamma_\mathsf{h}\|_h \leq \frac{T}{c}\, h \sum_{j=1}^{N-1} |\psi_j|.$$

□

Nach Lemma 4.3.3 reicht es also aus, die Werte

$$|\psi_j| = \Big| - \frac{1}{h}\int_0^T z(t)^\mathsf{T} A v_h^{(j)}(t)\,dt + \frac{1}{6}A(z_{j+1} + 4z_j + z_{j-1})\Big|$$

abzuschätzen. Dazu betrachten wir zur exakten Lösung z ihre stückweise lineare Interpolierende $P_1 z$. Für $j = 1, \ldots, N-1$ gilt

$$\frac{1}{h}\int_0^T (P_1 z)(t)^\mathsf{T} A v_h^{(j)}(t)\,dt = \frac{1}{h}\int_0^T \sum_{l=1}^{N-1} A z_l v_h^{(l)}(t) v_h^{(j)}(t)\,dt = \frac{1}{6}A(z_{j+1} + 4z_j + z_{j-1}).$$

Damit folgt unter Beachtung von $|v_h^{(j)}(t)| \le 1$

$$\begin{aligned}
|\psi_j| &\le \frac{1}{h}\int_0^T |z(t)^\mathsf{T} A v_h^{(j)}(t) - (P_1 z)(t)^\mathsf{T} A v_h^{(j)}(t)|\,dt \\
&= \frac{1}{h}\int_{t_{j-1}}^{t_{j+1}} |\big(z(t) - (P_1 z)(t)\big)^\mathsf{T} A v_h^{(j)}(t)|\,dt \\
&\le \frac{1}{h}\int_{t_{j-1}}^{t_{j+1}} |A\big(z(t) - (P_1 z)(t)\big)|\,dt \\
&\overset{(3.11)}{\le} h\|A\|\,(\|\ddot{z}\|_{L_1(T_{j-1})} + \|\ddot{z}\|_{L_1(T_j)}), \qquad j = 1, \ldots, N-1,
\end{aligned}$$

also auch

$$\sum_{j=1}^{N-1} |\psi_j| \le h\|A\| \sum_{j=0}^{N-1} (\|\ddot{z}\|_{L_1(T_{j-1})} + \|\ddot{z}\|_{L_1(T_j)}) = 2\|A\|\,\|\ddot{z}\|_1\, h \le 2T\|A\|\,\|\ddot{z}\|_\infty\, h.$$

Mit Lemma 4.3.3 folgt dann für $h \le \sqrt{(1-c)\frac{6}{\|A\|}}$

$$\|P_1 z - z_h\|_\infty = \|\gamma\|_\infty \le \frac{2}{c}T^2\|A\|\,\|\ddot{z}\|_\infty\, h^2$$

Kommen wir nun zur Abschätzung des Fehlers zwischen z und z_h. Es ist für $z \in W_\infty^2[0, T, \mathbb{R}^n]$ nach Lemma 3.3.4 und Lemma 4.3.3 mit $h \le \sqrt{(1-c)\frac{6}{\|A\|}}$

$$\begin{aligned}
\|z - z_h\|_\infty &\le \|z - P_1 z\|_\infty + \|P_1 z - z_h\|_\infty \\
&\le \|\ddot{z}\|_\infty\, h^2 + \frac{T}{c}h\sum_{j=1}^{N-1} |\psi_j| \le \left(1 + 2\frac{T^2}{c}\|A\|\right)\|\ddot{z}\|_\infty\, h^2 \\
&\le \max\{1, \frac{2}{c}\}\, c_{2.7}\, \|\ddot{z}\|_\infty\, h^2.
\end{aligned}$$

Wir halten die eben hergeleiteten Ergebnisse in einem Satz fest.

Satz 4.3.4. *Seien $y \in L_\infty[0,T,\mathbb{R}^n]$ und z die Lösung der Systemgleichung* (1.4)*, sowie z_h die Lösung der diskreten Systemgleichung* (4.1) *für y. Weiterhin sei $h \leq \sqrt{\frac{2}{\|A\|}}$. Dann gelten*

$$\begin{aligned} \|P_1 z - z_h\|_\infty &\leq 3T^2\|A\|c_{2.7}\,\|y\|_\infty\,h^2 \\ \|z - z_h\|_\infty &\leq 3c_{2.7}^2\,\|y\|_\infty\,h^2. \end{aligned}$$

Seien weiterhin p die Lösung der adjungierten Gleichung (2.10) *und p_h die Lösung der diskreten adjungierten Gleichung* (4.4) *zu z_h. Dann gelten*

$$\|P_1 p - p_h\|_\infty \leq 3c_{2.6}^2 c_{2.7}^2\,(\|y\|_\infty + \|z_d\|_\infty)\,h^2 \tag{4.11}$$

$$\|p - p_h\|_\infty \leq 3c_{2.6}^2 c_{2.7}^2\,(\|y\|_\infty + \|z_d\|_\infty)\,h^2. \tag{4.12}$$

Beweis: Zunächst leiten wir

$$\|\ddot{z}\|_\infty \leq \|A\|\,\|z\|_\infty + \|y\|_\infty \overset{(2.8)}{\leq} (T^2\|A\| + 1)\,\|y\|_\infty \leq c_{2.7}\,\|y\|_\infty$$

aus der Systemgleichung (1.4) ab. Wir erhalten die beiden ersten Aussagen mit den Betrachtungen vor dem Satz und $c = \frac{2}{3}$.

Im zweiten Teil bezeichne p_h^* wieder die Lösung der diskreten adjungierten Gleichung zu z als rechter Seite. Daher gilt zunächst

$$\|p_h^* - p_h\|_\infty \leq T^2\|z - z_h\|_\infty \leq 3T^2 c_{2.7}^2\,\|y\|_\infty\,h^2,$$

Dann schließen wir analog zur Vorgehensweise eben

$$\begin{aligned} \|P_1 p - p_h\|_\infty &\leq \|P_1 p - p_h^*\|_\infty + \|p_h^* - p_h\|_\infty \\ &\leq 3T^2\|A\|c_{2.7}\,\|z - z_d\|_\infty\,h^2 + T^2\|z - z_h\|_\infty \\ &\leq \left(3T^2\|A\|c_{2.7}\max\{1,T^2\}\,(\|y\|_\infty + \|z_d\|_\infty) + 3T^2 c_{2.7}^2\|y\|_\infty\right)h^2 \\ &\leq \left(\frac{3}{2}T^2 c_{2.7}^2\,(\|y\|_\infty + \|z_d\|_\infty) + 3T^2 c_{2.7}^2\|y\|_\infty\right)h^2 \\ &\leq 3c_{2.6}^2 c_{2.7}^2\|z - z_h\|_\infty\,h^2. \end{aligned}$$

Weiterhin gilt

$$\begin{aligned} \|p - p_h^*\|_\infty &\leq \left(1 + 2\frac{T^2}{c}\|A\|\right)\|\ddot{p}\|_\infty\,h^2 \\ &\leq \left(1 + 2\frac{T^2}{c}\|A\|\right)(\|A\|\,\|p\|_\infty + \|z - z_d\|_\infty)\,h^2 \\ &\overset{c=\frac{2}{3}}{\leq} 3(1 + T^2\|A\|)^2\,\|z - z_d\|_\infty\,h^2 \\ &\leq 3c_{2.7}^2\max\{1,T^2\}\,(\|y\|_\infty + \|z_d\|_\infty)\,h^2. \end{aligned}$$

und mit Hilfe der Dreiecks-Ungleichung erhalten wir

$$\|p - p_h\|_\infty \leq 6 \max\{1, T^2\}\, c_{2.7}^2\, (\|y\|_\infty + \|z_d\|_\infty)\, h^2 = 3 c_{2.6}^2 c_{2.7}^2\, (\|y\|_\infty + \|z_d\|_\infty)\, h^2.$$

□

Bemerkung: Die Forderung $h^2 \leq \frac{2}{\|A\|} =: \kappa$ besitzt nur formalen Charakter. Betrachten wir den Fall $h^2 > \kappa$, so folgt

$$\|z - z_h\|_\infty \leq \|z\|_\infty + \|z_h\|_\infty \underset{(4.3)}{\overset{(2.8)}{\leq}} 2T^2 \|y\|_\infty \frac{\kappa}{\kappa} \leq \frac{2}{\kappa} T^2 \|y\|_\infty\, h^2.$$

Aus diesem Grund ist die obige Bedingung an die Schrittweite keine Einschränkung.

An dieser Stelle sei noch einmal die spezielle Wahl der Basisfunktionen $v_h^{(j)}$ des Raumes $V_{h,0}[0, T, \mathbb{R}^n]$ bemerkt. Ihre Eigenschaften spielen bei der Herleitung der Konvergenzordnung eine gewichtige Rolle. Bei Benutzung allgemeiner Basisfunktionen sind gewisse Forderungen unverzichtbar. So darf sich die Menge der Stellen $t \in [0, T]$ mit $v_h^{(j)}(t) \neq 0$ höchstens über zwei der Teilintervalle T_i erstrecken. Dadurch verschwinden alle Skalarprodukte $\langle v_h^{(j)}, v_h^{(k)} \rangle$ für $|j - k| > 1$ und die Darstellung des Fehlers vor Lemma 4.3.3 wird möglich. ◊

4.4 Zusammenfassung

In den beiden vorangegangenen Abschnitten stand die Untersuchung von Eigenschaften der Finiten Elemente Methode im Zentrum des Interesses. Wir geben an dieser Stelle eine Zusammenfassung der eruierten Resultate.

In Lemma 4.2.1 konnten wir zeigen, dass die diskrete Systemgleichung bzw. die diskrete adjungierte Gleichung für jede Funktion aus dem Raum $L_2[0, T, \mathbb{R}^n]$ eine eindeutig bestimmte Lösung z_h bzw. p_h besitzt. Eine Formulierung der beiden Gleichungen mit Hilfe der linearen und beschränkten Operatoren $\mathcal{S}_h, \mathcal{S}_h^* : L_2[0, T, \mathbb{R}^n] \to L_2[0, T, \mathbb{R}^n]$ ist somit möglich, also

$$z_h(u) = \mathcal{S}_h \mathcal{T} u, \qquad p_h(u) = \mathcal{S}_h^*(\mathcal{S}_h \mathcal{T} u - z_d),$$

wobei die $\mathcal{T}$ für die Transformation $\mathcal{T} u = Bu + e$ steht. Wir unterscheiden für den diskreten adjungierten Zustand zwei Funktionen. Neben $p_h(u)$, bei dessen Berechnung die Diskretisierung in beiden Gleichungen berücksichtigt wird, steht

$$p_h^*(u) = \mathcal{S}_h^*(\mathcal{S} \mathcal{T} u - z_d)$$

für den diskreten adjungierten Zustand, bei dem wir nur in der adjungierten Gleichung diskretisieren.

Für die Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$ konnten wir in Lemma 4.2.1 und den nachfolgenden Korollaren ihre Beschränktheit in verschiedenen Normen nachweisen, wobei die oberen Schranken für die Operatoren übereinstimmen. So ist

$$\|S_h\|_2 \leq T^2, \qquad \|S_h^*\|_2 \leq T^2, \qquad \|S_h\|_\infty \leq T^2, \qquad \|S_h^*\|_\infty \leq T^2.$$

Neben der Beschränktheit der diskreten Differentialoperatoren wissen wir bereits um die Konvergenz der diskreten Lösung gegen die exakte Lösung für Systemgleichung und adjungierte Gleichung. Wir erinnern an Lemma 4.3.2 sowie an Satz 4.3.4 mit den Abschätzungen

$$\begin{aligned}
\|z(u) - z_h(u)\|_2 &\leq c_{2.6}^2 c_{2.7}^2 c_{2.4} \|\mathcal{T}u\|_2 \, h^2, \\
\|p(u) - p_h(u)\|_2 &\leq c_{4.9} \left(\|\mathcal{T}u\|_2 + \|z_d\|_2\right) h^2 \\
\|z(u) - z_h(u)\|_\infty &\leq 3c_{2.7} \|\mathcal{T}u\|_\infty \, h^2, \\
\|p(u) - p_h(u)\|_\infty &\leq 3c_{2.6}^2 c_{2.7} \left(\|\mathcal{T}u\|_\infty + \|z_d\|_\infty\right) h^2.
\end{aligned}$$

Damit sind wir im Besitz aller notwendigen Informationen für die Betrachtung von diskreten Steuerungsproblemen und ihrer Eigenschaften hinsichtlich der Approximation der Lösung von (StP).

4.5 Hauptergebnisse

Nach dem Beweis der eindeutigen Lösbarkeit von Systemgleichung im kontinuierlichen und diskreten Fall und der Herleitung der hinreichenden und notwendigen Bedingungen für die Optimalität einer Steuerung untersuchen wir nun, wie sich der Fehler bei der Diskretisierung der Steuerung auswirkt. Zunächst stellen wir Ergebnisse aus der Literatur zusammen. Das Problem (StP) finden wir in diversen Veröffentlichungen über die Diskretisierung von optimalen Steuerungsproblemen mit Hilfe Finiter Elemente wieder. So gibt Hinze [14] (2005) Fehlerabschätzungen für eine Semi-Diskretisierung an, wobei nur die Differentialoperatoren durch finite Operatoren approximiert werden. Die Steuerung bleibt - *a priori* - unverändert ein Element des $L_2[0,T]$. Dagegen diskretisieren Meyer/Rösch [16] (2004) die Steuerung stückweise konstant. Diese Resultate sollen zunächst vorgestellt werden. Wir „übersetzen“ die Ausführungen in unseren Kontext und erweitern sie auf vektorwertige Funktionen unter Verwendung der oben eingeführten Bezeichnungen.

4.5.1 Nicht-Diskretisierung der Steuerung

Um das Problem der Diskretisierung der Steuerung zu umgehen, schlägt Hinze [14] (2005) ein Verfahren vor, bei dem lediglich die Operatoren $\mathcal{S}$ und $\mathcal{S}^*$ durch die diskreten Operatoren $\mathcal{S}_h, \mathcal{S}_h^* : L_2[0,T,\mathbb{R}^n] \to V_{h,0}[0,T,\mathbb{R}^n]$ ersetzt werden. Die zu behandelnden Steuerungsprobleme besitzen dann die Form

$$(\mathrm{StP})_{\frac{h}{2}} \qquad \min_u J_{\frac{h}{2}}(\mathcal{S}_h\mathcal{T}u, u) = \min_u \frac{1}{2}\|\mathcal{S}_h\mathcal{T}u - z_d\|_2^2 + \frac{\nu}{2}\|u\|_2^2, \qquad u \in U^{ad}.$$

Dies nennen wir eine Semi-Diskretisierung des Problems (StP), daher auch die Verwendung des Subscripts $\frac{h}{2}$.

Satz 4.5.1 ([14], Theoreme 2.4 und 3.3). *Seien $\bar{u}$ und $\bar{u}_h$ die Lösungen der Probleme* (StP) *und* $(\mathrm{StP})_{\frac{h}{2}}$. *Dann gilt für genügend kleines h*

$$\|\bar{u} - \bar{u}_h\|_2 \leq c\,(\|\bar{u}\|_2 + \|z_d\|_2)\,h^2$$

mit einer von h unabhängigen Konstante $c > 0$.

In der anschließenden Bemerkung vergleicht der Autor dieses Resultat mit den Ergebnissen bei voller Diskretisierung, d.h. falls die Probleme $(\mathrm{StP})_h$ nur über einem endlich-dimensionalem Teilraum ablaufen. Der Autor gibt an, dass bei stückweise konstanter Approximation der Steuerung nur lineare Konvergenz gezeigt werden kann. Bei stückweise linearer Approximation erhöht sich die Konvergenzordnung auf $\frac{3}{2}$, was allerdings nur in numerischen Experimenten beobachtet wurde. Für einen Beweis der letzten Aussage verweisen wir auf Satz 4.5.12. Eine Verbesserung der Konvergenzordnung ist sowohl für stückweise konstante als auch für stückweise lineare Approximation ausgeschlossen. Betrachten wir die beste Approximation u_h^* der optimalen Steuerung $\bar{u}$ bzgl. der L_2-Norm in $U_h[0,T,\mathbb{R}^m]$, d.h.

$$\|\bar{u} - u_h^*\|_2 = \inf_{u_h \in U_h} \|\bar{u} - u_h\|_2,$$

so erhalten wir zunächst

$$u_h^*(t) = \frac{1}{h}\int_{t_i}^{t_{i+1}} \bar{u}(s)\,ds, \qquad \forall t \in T_i,\ i = 0,\dots,N-1$$

und darauf aufbauend für den Fehler die Abschätzung

$$\begin{aligned}
\|\bar{u} - u_h^*\|_2^2 &= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big|\bar{u}(t) - \frac{1}{h}\int_{t_i}^{t_{i+1}} \bar{u}(s)\,ds\Big|^2 dt \\
&= \frac{1}{h^2}\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big|\int_{t_i}^{t_{i+1}} \bar{u}(t) - \bar{u}(s)\,ds\Big|^2 dt \\
&\leq \frac{1}{h^2}\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big(\int_{t_i}^{t_{i+1}} \|\dot{\bar{u}}\|_\infty |t-s|\,ds\Big)^2 dt \\
&\leq \frac{1}{h^2}\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \big(\|\dot{\bar{u}}\|_\infty\, h^2\big)^2 dt \\
&\leq \sum_{i=0}^{N-1} \|\dot{\bar{u}}\|_\infty^2\, h^3 \leq T\|\dot{\bar{u}}\|_\infty^2\, h^2,
\end{aligned}$$

also

$$\|\bar{u} - u_h^*\|_2 \leq \sqrt{T}\|\dot{\bar{u}}\|_\infty h.$$

Es ist demnach nicht zu erwarten, dass die Lösung unseres Optimierungsproblems mit stückweise konstanter Approximation der Steuerung eine höhere Konvergenzordnung als 1 besitzt. Für den Fall $u_h^* \in V_h[0,T,\mathbb{R}^m]$ zeigt eine ähnliche Rechnung die Optimalität der Konvergenzordnung $h^{\frac{3}{2}}$.

Für den Fehler in der L_∞-Norm zitieren wir als Ausgangspunkt wieder Hinze [14].

Satz 4.5.2 ([14], Theorem 3.6). *Seien $\bar{u}$ und $\bar{u}_h$ die Lösungen der Probleme* (StP) *und* $(\mathrm{StP})_{\frac{h}{2}}$*. Dann gilt für genügend kleines h*

$$\|\bar{u} - \bar{u}_h\|_\infty \leq c\,(\|p(\bar{u}) - p_h(\bar{u})\|_\infty + c\,h^2)$$

mit einer von h unabhängigen Konstante $c > 0$.

Die Erweiterung dieser Aussage folgt mit Satz 4.3.4.

Satz 4.5.3. *Seien $\bar{u}$ und $\bar{u}_h$ die Lösungen der Probleme* (StP) *und* $(\mathrm{StP})_{\frac{h}{2}}$*. Dann gilt für genügend kleine Schrittweite*

$$\|\bar{u} - \bar{u}_h\|_\infty \leq c\,h^2$$

mit einer von h unabhängigen Konstante $c > 0$.

Beweis: Satz 4.3.4 begründet die Aussage unter Verwendung des zuletzt zitierten Resultats. □

Somit ist für beide Normen die quadratische Konvergenzordnung der Lösungen der diskreten Probleme gegen die exakte Lösung bewiesen, falls wir auf eine Diskretisierung der Steuerung verzichten. Die dadurch entstehenden Probleme $(\mathrm{StP})_{\frac{h}{2}}$ erfordern allerdings einen schwer abzuschätzenden numerischen Mehraufwand gegenüber den Problemen $(\mathrm{StP})_h$ in den folgenden Abschnitten mit expliziter Diskretisierung der Steuerung.

4.5.2 Konstante Approximation der Steuerung

Konvergenz im quadratischen Mittel

Im Gegensatz zu den im letzten Abschnitt vorgestellten Ergebnissen diskretisieren Meyer/Rösch [16] (2004) ebenfalls die Steuerung und untersuchen somit das endlichdimensionale Steuerungsproblem

$$(\mathrm{StP})_h \qquad \min_{u_h} J_h(\mathcal{S}_h\mathcal{T}u, u) = \min_{u} \frac{1}{2}\|\mathcal{S}_h\mathcal{T}u_h - z_d\|_2^2 + \frac{\nu}{2}\|u_h\|_2^2, \qquad u_h \in U_h^{ad},$$

mit $U_h^{ad} = U^{ad} \cap U_h[0,T,\mathbb{R}^m]$. Betrachten wir in dieser Situation den Fehler zwischen exakter und diskreter Lösung, so erhalten wir weder bei stückweise konstanter noch bei stückweise linearer Approximation quadratische Konvergenz. Daher benutzen die Autoren für ein zweidimensionales Problem, d.h. $u : \mathbb{R}^2 \to \mathbb{R}$ und $z, p : \mathbb{R}^2 \to \mathbb{R}$, und stückweise konstante Approximation der Steuerung die Funktion

$$\tilde{u} = \Pi_{[\mathrm{a,b}]}\left(-\frac{1}{\nu} B^{\mathsf{T}} p_h(u_h)\right),$$

wobei $u_h \in U_h^{ad}$ die Lösung von $(\mathrm{StP})_h$ ist. Diese Formel ist aus den Optimalitätsbedingungen für das diskrete Problem abgeleitet und $\tilde{u}$ besitzt stärkere Glattheitseigenschaften als die diskreten Steuerungen. Die Funktion $\tilde{u}$ ist wegen der Lipschitz-Stetigkeit von $p_h(u_h)$ ebenfalls Lipschitz-stetig, denn der stetige Operator $\Pi_{[\mathrm{a,b}]}$ ändert diese Eigenschaft nicht. Wir geben an dieser Stelle den Weg von Meyer/Rösch [16] wieder.

Voraussetzung: Das Steuerungsproblem (StP) besitze eine optimale Steuerung mit der Eigenschaft:

$$|\{t \in [0,T] : \nexists\, \dot{\bar{u}}(t)\}| = K < \infty. \tag{$\mathrm{V_{con}}$}$$

Die Konstante K sei über diese Voraussetzung definiert, sie wird in den folgenden Fehlerabschätzung wieder erscheinen. Die Voraussetzung unterteilt die Intervalle T_i demnach in zwei Gruppen. Die Menge K_1 umfasst diejenigen Intervalle, auf denen $\Pi_{[\mathrm{a,b}]}$ die Identität ist und somit $\bar{u} \in W_2^2[t_i, t_{i+1}, \mathbb{R}^m]$ gilt. Mit K_2 dagegen benennen wir die Menge der Intervalle, wo $\bar{u}$ nicht zweimal differenzierbar ist, sondern nur Lipschitz-stetig. Das sind jene Intervalle, auf denen $-\frac{1}{\nu} B^{\mathsf{T}} p(\bar{u})$ einen Schnittpunkt mit einer der Geraden a oder b besitzt.

Lemma 4.5.4. *Bezeichne $\bar{u}$ die Lösung von* (StP) *und die Voraussetzung* ($\mathrm{V_{con}}$) *sei erfüllt. Dann gilt*

$$\|z_h(\bar{u}) - z_h(P_0\bar{u})\|_2 \le c\, \|p(\bar{u})\|_{2,2}\, h^2,$$

mit einer von $\bar{u}$ und h unabhängigen Konstante c.

Beweis: Für die Erweiterung des Beweises von Meyer/Rösch [16] auf Funktionen mit mehrdimensionalem Wertebereich und die genaue Deklaration der Abschätzungskonstante verweisen wir auf den Anhang, Abschnitt A.2. □

Lemma 4.5.5. *Sei $\bar{u}$ die Lösung des Problems* (StP) *mit $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$. Dann gilt*

$$\|z_h(\bar{u}) - z_h(P_0\bar{u})\|_\infty \le c_{4.13}\, (\mathsf{V}_0^T \dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty)\, h^2, \tag{4.13}$$

mit $c_{4.13} = \frac{T}{4}\|B\|\, c_{2.7}$ unabhängig von $\bar{u}$ und h.

Beweis: Zunächst ist $z_h(\bar{u}) - z_h(P_0\bar{u}) \in V_{h,0}[0,T,\mathbb{R}^n]$ die Lösung von (4.1) für $\bar{u} - P_0\bar{u}$, d.h. wir schließen mit Lemma 4.2.4 und der Funktion $Y(t) = \int_0^t B(\bar{u} - P_0\bar{u})(s)\, ds$ auf

$$\|z_h(\bar{u}) - z_h(P_0\bar{u})\|_\infty \le c_{2.7}\, \|Y\|_1 = c_{2.7} \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |Y(t)|\, dt.$$

Für Y gilt per Definition $Y(t) = Y(t_i) + \int_{t_i}^{t} B(\bar{u} - P_0\bar{u})(s)\,ds$ für alle $t \in T_i$ und $i = 0, \dots, N-1$. Daher erhalten wir

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |Y(t)|\,dt = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big| Y(t_i) + \int_{t_i}^{t} B(\bar{u} - P_0\bar{u})(s)\,ds \Big|\,dt.$$

Für eine bessere Übersicht untersuchen wir die rechte Seite getrennt. Zum Einen ist wegen $Y(0) = 0_n$

$$\begin{aligned}
\int_{t_i}^{t_{i+1}} |Y(t_i)|\,ds &= h\Big| \sum_{j=0}^{i-1} Y(t_{j+1}) - Y(t_j) \Big| = h\Big| \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \dot{Y}(t)\,dt \Big| \\
&\le h\|B\| \Big| \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \bar{u}(t) - \bar{u}(S_j)\,dt \Big| \\
&\le h\|B\| \sum_{j=0}^{i-1} \Big| \int_0^{\frac{h}{2}} \bar{u}(S_j + t) - 2\bar{u}(S_j) + \bar{u}(S_j - t)\,dt \Big| \\
&\le h\|B\| \sum_{j=0}^{N-1} \int_0^{\frac{h}{2}} \Big| \int_0^t \dot{\bar{u}}(S_j + s)\,ds - \int_0^t \dot{\bar{u}}(S_j - t + s)\,ds \Big|\,dt \\
&\le h\|B\| \sum_{j=0}^{N-1} \int_0^{\frac{h}{2}} \int_0^t \big| \dot{\bar{u}}(S_j + s) - \dot{\bar{u}}(S_j - t + s) \big|\,ds\,dt \\
&\le h\|B\| \sum_{j=0}^{N-1} \int_0^{\frac{h}{2}} \int_0^{\frac{h}{2}} \mathsf{V}_{T_j} \dot{\bar{u}}\,ds\,dt \\
&\le \frac{h^3}{4}\|B\| \sum_{j=0}^{N-1} \mathsf{V}_{T_j} \dot{\bar{u}} \overset{(3.6)}{=} \frac{h^3}{4} \mathsf{V}_0^T \dot{\bar{u}},
\end{aligned}$$

und somit

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |Y(t)|\,dt \le \frac{T}{4}\|B\|\,\mathsf{V}_0^T \dot{\bar{u}}\,h^2.$$

Andererseits gilt

$$\begin{aligned}\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big|\int_{t_i}^{t}(\bar u - P_0\bar u)(s)\,ds\Big|\,dt &\le \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\int_{t_i}^{t_{i+1}}|(\bar u - P_0\bar u)(s)|\,ds\,dt\\ &= h\sum_{i=0}^{N-1}\int_{-\frac{h}{2}}^{\frac{h}{2}}|\bar u(S_i+s)-\bar u(S_i)|\,ds\\ &\le h\|\dot{\bar u}\|_\infty\sum_{i=0}^{N-1}\int_{-\frac{h}{2}}^{\frac{h}{2}}|s|\,ds\\ &= \frac{T}{4}\|\dot{\bar u}\|_\infty h^2.\end{aligned}$$

Nach Abschätzung beider Summanden ergibt sich folgendes Resultat

$$\int_0^T |Y(t)|\,dt \le \frac{T}{4}\|B\|\,(\mathsf{V}_0^T\dot{\bar u}+\|\dot{\bar u}\|_\infty)\,h^2 \le \frac{T}{4}\|B\|\,(\mathsf{V}_0^T\dot{\bar u}+\|\dot{\bar u}\|_\infty)\,h^2.$$

□

Bemerkung: Lemma 4.5.4 ist eine Erweiterung der von Meyer/Rösch [16] (2004) veröffentlichten Herleitung auf Funktionen mit mehrdimensionalem Wertebereich. Die Autoren stellen die stärkere Forderung ($\mathrm{V}_{\mathrm{con}}$) an die optimale Steuerung. Unsere Bedingung $\dot{\bar u}\in BV[0,T,\mathbb{R}^m]$ stellt tatsächlich eine schwächere Forderung dar. Betrachten wir z.B. eine Funktion deren Ableitung abzählbar viele Sprüngen auf $[0,T]$ mit Sprunghöhe h_i für $i=1,2,\ldots$ besitzt, d.h. der Wert von $\dot f$ soll vom Niveau h_i auf Null springen, den Wert für eine Menge vom Maß größer als Null den Wert beibehalten und dann auf h_{i+1} springen. Dann gilt $\mathsf{V}_0^T\dot f = 2\sum_{i=1}^\infty h_i$. Konvergiert die Reihe $(h_i)_{i=1}^\infty$, so ist die Variation von $\dot f$ auf $[0,T]$ beschränkt. Das bedeutet, es existiert eine Funktion F mit $\dot F = f$ und $K=\infty$, deren Ableitung dennoch von beschränkter Variation ist. ◊

Korollar 4.5.6. *Mit den obigen Bezeichnungen gelten für die adjungierten Zustände folgende Abschätzungen*

$$\|p_h(\bar u)-p_h(P_0\bar u)\|_2 \le T^2\sqrt{T}c_{4.13}\,(\mathsf{V}_0^T\dot{\bar u}+\|\dot{\bar u}\|_\infty)\,h^2$$

und

$$\|p(\bar u)-p_h(P_0\bar u)\|_2 \le \sqrt{T}c_{4.14}\,C_{4.14}^{\bar u}\,h^2 \tag{4.14}$$

mit der von $\bar u$ und h unabhängigen Konstante $c_{4.14}=\max\{c_{4.9},T^2c_{4.13}\}$ und der Abkürzung $C_{4.14}^{\bar u}=\|\mathcal{T}\bar u\|_\infty+\|z_d\|_\infty+\mathsf{V}_0^T\dot{\bar u}+\|\dot{\bar u}\|_\infty$.

Beweis: Der erste Teil folgt leicht aus dem zuletzt durchgeführten Beweis:

$$\|p_h(\bar u)-p_h(P_0\bar u)\|_2 \le T^2\|z_h(\bar u)-z_h(P_0\bar u)\|_2 \le T^2\sqrt{T}c_{4.13}\,(\mathsf{V}_0^T\dot{\bar u}+\|\dot{\bar u}\|_\infty)\,h^2.$$

Daher gilt mit Hilfe der Dreiecks-Ungleichung

$$\begin{aligned}\|p(\bar{u}) - p_h(P_0\bar{u})\|_2 &\leq \|p(\bar{u}) - p_h(\bar{u})\|_2 + \|p_h(\bar{u}) - p_h(P_0\bar{u})\|_2 \\ &\overset{(4.9)}{\leq} c_{4.9}\left(\|\mathcal{T}\bar{u}\|_2 + \|z_d\|_2\right)h^2 + \|p_h(\bar{u}) - p_h(P_0\bar{u})\|_2 \\ &\leq \left(c_{4.9}\left(\|\mathcal{T}\bar{u}\|_2 + \|z_d\|_2\right) + T^2\sqrt{T}c_{4.13}\left(\mathsf{V}_0^T\dot{\bar{u}} + \|\ddot{\bar{u}}\|_\infty\right)\right)h^2 \\ &\leq \sqrt{T}c_{4.14}\left(\|\mathcal{T}\bar{u}\|_\infty + \|z_d\|_\infty + \mathsf{V}_0^T\dot{\bar{u}} + \|\ddot{\bar{u}}\|_\infty\right)h^2.\end{aligned}$$

□

Jetzt benutzen wir die bisherigen Ergebnisse für die Untersuchung des Abstands zwischen der diskreten Lösung und der Interpolation der optimalen Steuerung im Raum $U_h[0,T,\mathbb{R}^m]$. Auch hier ist die Voraussetzung ($\mathrm{V}_{\mathrm{con}}$) in unserem Fall nicht nötig. Wir ersetzen sie durch die etwas schwächere Forderung $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$ und erweitern den Beweis von Meyer/Rösch [16] (2004) auf unser vektorwertiges Problem.

Satz 4.5.7. *Seien $\bar{u}$ die Lösung von* (StP) *und $u_h \in U_h[0,T,\mathbb{R}^m]$ die Lösung des diskreten Optimierungsproblems* $(\mathrm{StP})_h$. *Dann gilt unter der Voraussetzung $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$ die folgende Abschätzung*

$$\|u_h - P_0\bar{u}\|_2 \leq \sqrt{T}c_{4.15}\,C_{4.14}^{\bar{u}}\,h^2 \tag{4.15}$$

mit $c_{4.15} = \frac{\|B\|}{\nu}(c_{4.14} + \frac{1}{8}c_{2.6}c_{2.7})$, einer von $\bar{u}$ und h unabhängigen Konstante.

Diese Aussage bedeutet, dass die Konvergenz der Lösungen der diskreten Probleme gegen die optimale Steuerung in den Intervallmittelpunkten schon quadratische Ordnung besitzt.

Beweis: In Lemma 2.4.1 und im Verlauf dieses Kapitels konnten wir zeigen, dass die Bedingungen

$$\langle B^{\mathsf{T}}\bar{p} + \nu\bar{u}, u - \bar{u}\rangle \geq 0 \qquad \forall u \in U^{ad} \tag{2.17}$$

$$\langle B^{\mathsf{T}}p_h(u_h) + \nu u_h, \zeta_h - u_h\rangle \geq 0 \qquad \forall \zeta_h \in U_h^{ad} \tag{4.7}$$

notwendig und hinreichend für die Optimalität der eindeutig bestimmten Lösungen der Probleme (StP) und $(\mathrm{StP})_h$ sind. Nach Lemma 2.4.3 besitzt die Variationsungleichung (2.17) auch punktweise Gültigkeit, d.h.

$$\left(B^{\mathsf{T}}\bar{p}(t) + \nu\bar{u}(t)\right)^{\mathsf{T}}\left(u - \bar{u}(t)\right) \geq 0 \qquad \forall' t \in [0,T],\ \forall u \in U$$

mit $U = \{u \in \mathbb{R}^m : a \leq u \leq b\}$. Da $\bar{u}$, $\bar{p}$ und u_h an den Stellen S_i stetig sind, setzen wir speziell $u = u_h(S_i)$ und erhalten für $i = 0,\ldots,N-1$

$$\begin{aligned}&\left(B^{\mathsf{T}}\bar{p}(S_i) + \nu\bar{u}(S_i)\right)^{\mathsf{T}}\left(u_h(S_i) - \bar{u}(S_i)\right) = \\ &\qquad\qquad \left(B^{\mathsf{T}}\bar{p}(S_i) + \nu(P_0\bar{u})(S_i)\right)^{\mathsf{T}}\left(u_h(S_i) - (P_0\bar{u})(S_i)\right) \geq 0.\end{aligned}$$

Durch Integration über T_i und Addition aller Ungleichungen erhalten wir

$$\langle B^{\mathsf{T}}(P_0\bar{p}) + \nu(P_0\bar{u}), u_h - P_0\bar{u}\rangle \geq 0.$$

Darüber hinaus testen wir die diskrete Optimalitätsbedingung (4.7) mit der Funktion $P_0\bar{u}$ und erhalten so

$$\langle B^{\mathsf{T}} p_h(u_h) + \nu u_h, P_0\bar{u} - u_h\rangle \geq 0.$$

Addieren wir diese zwei Ungleichungen, so folgt

$$\langle B^{\mathsf{T}}(P_0\bar{p} - p_h(u_h)) + \nu(P_0\bar{u} - u_h), u_h - P_0\bar{u}\rangle \geq 0,$$

was äquivalent ist zu

$$\nu\|u_h - P_0\bar{u}\|_2^2 \leq \langle P_0\bar{p} - p_h(u_h), B(u_h - P_0\bar{u})\rangle.$$

Wir bearbeiten nun die rechte Seite der letzten Ungleichung weiter und spalten das Skalarprodukt auf

$$\begin{aligned}\langle P_0\bar{p} - p_h(u_h), B(u_h - P_0\bar{u})\rangle &= \langle p_h(P_0\bar{u}) - p_h(u_h), B(u_h - P_0\bar{u})\rangle \\ &+ \langle \bar{p} - p_h(P_0\bar{u}), B(u_h - P_0\bar{u})\rangle \\ &+ \langle P_0\bar{p} - \bar{p}, B(u_h - P_0\bar{u})\rangle.\end{aligned}$$

Für den ersten Summand beachten wir die Adjungiertheit der diskreten Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$ auf dem $L_2[0,T,\mathbb{R}^n]$ (siehe Lemma 4.2.5) und schreiben

$$\langle p_h(P_0\bar{u}) - p_h(u_h), B(u_h - P_0\bar{u})\rangle = \langle z_h(P_0\bar{u}) - z_h(u_h), z_h(u_h) - z_h(P_0\bar{u})\rangle \leq 0.$$

Den zweiten Term schätzen wir mit der Cauchy-Schwarz-Ungleichung und mit Korollar 4.5.6 wie folgt ab

$$\begin{aligned}\langle \bar{p} - p_h(P_0\bar{u}), B(u_h - P_0\bar{u})\rangle &\leq \|B\|\,\|\bar{p} - p_h(P_0\bar{u})\|_2\|u_h - P_0\bar{u}\|_2 \\ &\overset{(4.14)}{\leq} \sqrt{T}c_{4.14}\|B\|\,\|u_h - P_0\bar{u}\|_2\, C_{4.14}^{\bar{u}}\, h^2.\end{aligned}$$

Der letzte Summand stellt eine Formel zur numerischen Integration dar und wir erhalten mit Lemma 3.3.2 und dem Wissen um die Konstanz von u_h und $P_0\bar{u}$ auf jedem

Intervall die Abschätzung

$$
\begin{aligned}
&\langle P_0\bar{p} - \bar{p}, B(u_h - P_0\bar{u})\rangle \\
&\quad = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \big(u_h(t) - (P_0\bar{u})(t)\big)^{\mathsf{T}} B^{\mathsf{T}} \big((P_0\bar{p})(t) - \bar{p}(t)\big)\, dt \\
&\quad = \sum_{i=0}^{N-1} \big(u_h(S_i) - (P_0\bar{u})(S_i)\big)^{\mathsf{T}} B^{\mathsf{T}} \left(\int_{t_i}^{t_{i+1}} \bar{p}(S_i) - \bar{p}(t)\, dt \right) \\
&\quad \overset{(3.10)}{\leq} \frac{\|B\|}{4} \sum_{i=0}^{N-1} \Big(|u_h(S_i) - (P_0\bar{u})(S_i)| \sqrt{h} \Big)\, |\bar{p}|_{W_2^2(T_i)}\, h^2 \\
&\quad \overset{(A.1)}{\leq} \frac{\|B\|}{4} \underbrace{\left(\sum_{i=0}^{N-1} \big|u_h(S_i) - (P_0\bar{u})(S_i)\big|^2 h \right)^{\frac{1}{2}}}_{=\|u_h - P_0\bar{u}\|_2} \underbrace{\left(\sum_{i=0}^{N-1} |\bar{p}|^2_{W_2^2(T_i)} \right)^{\frac{1}{2}}}_{=|\bar{p}|_{2,2}} h^2 \\
&\quad \overset{(2.13)}{\leq} \frac{\|B\|}{4} c_{2.7}\, \|u_h - P_0\bar{u}\|_2 \|z(\bar{u}) - z_d\|_2\, h^2 \\
&\quad \leq \frac{\|B\|}{8} c_{2.6}^2 c_{2.7}\, (\|\mathcal{T}\bar{u}\|_2 + \|z_d\|_2)\, \|u_h - P_0\bar{u}\|_2\, h^2.
\end{aligned}
$$

Nach dem Einsetzen der beiden positiven Summanden, und beidseitigem Kürzen des Terms $\|u_h - P_0\bar{u}\|_2$ erhalten wir

$$\nu \|u_h - P_0\bar{u}\|_2 \leq \sqrt{T}(c_{4.14} + \frac{1}{8} c_{2.6}^2 c_{2.7})\, \|B\|\, C_{4.14}^{\bar{u}}\, h^2.$$

Der letzte Schritt ist möglich, denn im Fall $u_h - P_0\bar{u} = 0$ wäre die Aussage trivialerweise erfüllt. □

Nun greifen wir die Gedanken vom Anfang des Abschnittes wieder auf und untersuchen die Approximationsgüte der Funktion $\tilde{u} = \Pi_{[a,b]}\big(-\frac{1}{\nu} B^{\mathsf{T}} p_h(u_h)\big)$ bezüglich $\bar{u}$.

Satz 4.5.8. *Seien $\bar{u}$ die Lösung von* (StP) *mit $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$ und $u_h \in U_h[0,T,\mathbb{R}^m]$ die Lösung des diskreten Steuerungsproblems* $(\mathrm{StP})_h$*. Die Funktion $\tilde{u}$ sei ferner definiert durch $\tilde{u} = \Pi_{[a,b]}\big(-\frac{1}{\nu} B^{\mathsf{T}} p_h(u_h)\big)$. Dann gilt*

$$\|\bar{u} - \tilde{u}\|_2 \leq c^{\Pi} \sqrt{T}(c_{4.14} + T^4 \|B\|\, c_{4.15})\, C_{4.14}^{\bar{u}}\, h^2,$$

wobei $c^{\Pi} = \frac{1}{\nu}\|B\|$ ist.

Beweis: Die Aussagen von Lemma 4.2.1 und Satz 4.5.7 implizieren

$$\|p_h(P_0\bar{u}) - p_h(u_h)\|_2 \leq T^4 \|B\|\, \|P_0\bar{u} - u_h\|_2 \leq T^4 \sqrt{T} \|B\|\, c_{4.15}\, C_{4.14}^{\bar{u}}\, h^2.$$

Von Korollar 4.5.6 übernehmen wir die Abshcätzung (4.14)

$$\|\bar{p} - p_h(P_0\bar{u})\|_2 \leq \sqrt{T} c_{4.14}\, C_{4.14}^{\bar{u}}\, h^2$$

und mit Hilfe der Dreiecks-Ungleichung erhalten wir

$$\|\bar{p} - p_h(u_h)\|_2 \leq \sqrt{T}(c_{4.14} + T^4\|B\|\, c_{4.15})\, C_{4.14}^{\bar{u}}\, h^2.$$

Der Operator $\Pi_{[\mathrm{a,b}]} : C[0,T,\mathbb{R}^m] \to U^{ad}$ ist als Projektion auf eine abgeschlossene und konvexe Menge Lipschitz-stetig mit Lipschitz-Konstante 1, daher gilt zunächst

$$\|\bar{u} - \tilde{u}\|_2 = \|(-\frac{1}{\nu}B^{\mathsf{T}}\bar{p}) - (-\frac{1}{\nu}B^{\mathsf{T}}p_h(u_h))\|_2 \leq \frac{\|B\|}{\nu}\|\bar{p} - p_h(u_h)\|_2$$

und daraus ableitend

$$\|\bar{u} - \tilde{u}\|_2 \leq c^{\Pi}\sqrt{T}(c_{4.14} + T^4\|B\|\, c_{4.15})\, C_{4.14}^{\bar{u}}\, h^2.$$

□

Quadratische Konvergenzordnung in der L_∞-Norm

Der im vergangenen Abschnitt durchgeführte Nachweis der quadratischen Konvergenz in der L_2-Norm beruhte zu einem Großteil auf der Darstellbarkeit dieser Norm durch ein Skalarprodukt (vgl. Lemmata 4.3.1 und 4.3.2). Schlagen wir einen ähnlichen Weg wie Hinze[14] (2005) oder Meyer/Rösch [16] (2005) ein und verwenden die L_∞-Norm, so verringert sich die Konvergenzordnung. Wir sind allerdings bestrebt auch in dieser Norm die obigen Resultate zu erhalten, denn eine punktweise Abschätzung gibt wesentlich mehr Informationen über die tatsächliche Lösung preis. Dabei bleiben wir vorerst bei stückweise konstanter Approximation der Steuerung. Die entscheidende Stelle für die Verringerung der Konvergenzgeschwindigkeit ist der Abstand der exakten und diskreten Lösung der adjungierten Gleichung. In Satz 4.3.4 konnten wir eine Abschätzung des Fehlers zwischen der Lösung der Systemgleichung und der Lösung der diskreten Systemgleichung herleiten. Dieses Ergebnis spielt im weiteren Verlauf ein wichtige Rolle. Wir weisen an dieser Stelle noch einmal auf die Einschränkung hinsichtlich der Basisfunktionen hin (vgl. den zweiten Teil der Bemerkung unter Satz 4.3.4).

Satz 4.5.9. *Seien $\bar{u}$ die Lösung des Problems* (StP) *mit $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$ und die Funktion $\tilde{u}$ gegeben durch $\tilde{u} = \Pi_{[\mathrm{a,b}]}\big(-\frac{1}{\nu}B^{\mathsf{T}}p_h(u_h)\big)$ mit der Lösung $u_h \in U_h[0,T,\mathbb{R}^m]$ der diskreten Probleme* $(\mathrm{StP})_h$. *Dann gilt unter der Voraussetzung $h \leq \sqrt{\frac{2}{\|A\|}}$ die Abschätzung*

$$\|\bar{u} - \tilde{u}\|_\infty \leq c\, C_{4.14}^{\bar{u}}\, h^2,$$

mit einer von $\bar{u}$ und h unabhängigen Konstante c.

Beweis: Wir betrachten zunächst wieder die adjungierten Zustände und schließen mit Hilfe der Dreiecks-Ungleichung

$$\|p(\bar{u}) - p_h(u_h)\|_\infty \leq \|\bar{p} - p_h(\bar{u})\|_\infty + \|p_h(\bar{u}) - p_h(P_0\bar{u})\|_\infty + \|p_h(P_0\bar{u}) - p_h(u_h)\|_\infty.$$

Für den ersten Term kennen wir aus Satz 4.3.4 die Abschätzung

$$\|p(\bar{u}) - p_h(\bar{u})\|_\infty \leq 3c_{2.6}^2 c_{2.7} \left(\|\mathcal{T}\bar{u}\|_\infty + \|z_d\|_\infty\right) h^2$$

und für den zweiten Summanden übernehmen wir aus Lemma 4.5.5

$$\|p_h(\bar{u}) - p_h(P_0\bar{u})\|_\infty \leq T^2 \|z_h(\bar{u}) - z_h(P_0\bar{u})\|_\infty \leq T^2 c_{4.13} \left(\mathsf{V}_0^T \dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty\right) h^2.$$

Die Abschätzung des dritten Summanden gestaltet sich etwas komplizierter. Es ist

$$\begin{aligned} \|p_h(P_0\bar{u}) - p_h(u_h)\|_\infty &\overset{(4.5)}{\leq} T\sqrt{T}\, \|z_h(P_0\bar{u}) - z_h(u_h)\|_2 \\ &\leq T^3\sqrt{T}\|B\|\, \|P_0\bar{u} - u_h\|_2\, h^2 \overset{(4.15)}{\leq} T^4\|B\|\, c_{4.15}\, C_{4.14}^{\bar{u}}\, h^2. \end{aligned}$$

Somit erhalten wir Dank der Lipschitz-Stetigkeit des Operators $\Pi_{[a,b]}$

$$\|\bar{u} - \tilde{u}\|_\infty \leq c^\Pi \|p(\bar{u}) - p_h(u_h)\|_\infty \leq c^\Pi (3c_{2.6}^2 c_{2.7} + T^2 c_{4.13} + T^4\|B\|\, c_{4.15})\, C_{4.14}^{\bar{u}}\, h^2.$$

□

4.5.3 Lineare Approximation der Steuerung

In den bisherigen Betrachtungen über die Diskretisierung mittels der Finite Elemente Methode approximierten wir die Steuerung stückweise konstant und erweiterten in Satz 4.5.7 ein wichtiges Resultat aus Meyer/Rösch [16] (2004). Der Beweis dieses Lemmas stützt sich an einer entscheidenden Stelle auf die spezielle Form der Diskretisierung. Sie nutzen die punktweise Gültigkeit der Optimalitätsbedingung für (StP) zu diskreten Aussagen und leiteten daraus wiederum Abschätzungen in der L_2-Norm her. Bei einer Veränderung der Diskretisierung geht dieser Vorteil womöglich verloren, allerdings wissen wir aus Lemma 2.4.1 von der optimalen Steuerung $\bar{u} \in W_\infty^1[0,T,\mathbb{R}^n]$, d.h. $\bar{u}$ ist eine stetige Funktion. Daher liegt es nahe die Approximation zu verbessern. Wir verwenden demnach Funktionen aus dem Raum $V_h[0,T,\mathbb{R}^n]$. An der eindeutigen Lösbarkeit der diskreten Systemgleichung ändert dieses Vorgehen nichts. Es gilt die Aussage, dass für jedes $u_h \in V_h[0,T,\mathbb{R}^n]$ eine eindeutig bestimmte Lösung $z_h \in V_{h,0}[0,T,\mathbb{R}^n]$ der diskreten Systemgleichung (4.1) existiert verbunden mit der stetigen Abhängigkeit von der rechten Seite. Die diskreten Steuerungsprobleme lauten dann

$$(\mathrm{StP})_h \qquad \min_{u_h} \frac{1}{2}\|\mathcal{S}_h\mathcal{T}u_h - z_d\|_2^2 + \frac{\nu}{2}\|u_h\|_2^2, \qquad u_h \in U_h^{ad},$$

wobei $U_h^{ad} = U^{ad} \cap V_h[0,T,\mathbb{R}^m]$.

Als Startpunkt erinnern wir an Lemma 3.3.6. Dort betrachteten wir den Fehler, der bei der Interpolation der optimalen Steuerung $\bar{u}$ durch stückweise lineare Funktionen entsteht. Dieses Resultat benutzen wir nun, um ein Analogon zu Lemma 4.5.5 zu beweisen.

Lemma 4.5.10. *Sei $\bar{u}$ die Lösung von* (StP) *mit $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$. Dann gilt*

$$\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_\infty \leq T\|B\|\,\mathsf{V}_0^T\dot{\bar{u}}\,h^2. \tag{4.16}$$

Beweis: Zunächst ist $z_h(\bar{u}) - z_h(P_1\bar{u}) \in V_{h,0}[0,T,\mathbb{R}^n]$ die Lösung von (4.1) für $y = B(\bar{u} - P_1\bar{u})$. Setzen wir dort speziell $v_h = z_h(\bar{u}) - z_h(P_1\bar{u})$, so schließen wir auf

$$\|\dot{z}_h(\bar{u}) - \dot{z}_h(P_1\bar{u})\|_2^2 \leq \|z_h(\bar{u}) - z_h(P_1\bar{u})\|_\infty\|B\| \int_0^T |\bar{u}(t) - (P_1\bar{u})(t)|\,dt,$$

denn A ist positiv semi-definit. Daher folgt mit Lemma 3.3.6

$$\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_\infty^2 \overset{(A.4)}{\leq} T\|\dot{z}_h(\bar{u}) - \dot{z}_h(P_1\bar{u})\|_2^2 \overset{(3.13)}{\leq} T\|B\|\,\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_\infty\,\mathsf{V}_0^T\dot{\bar{u}}\,h^2.$$

Im Fall $\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_\infty = 0$ wäre nichts zu zeigen gewesen, daher kürzen wir den Term auf beiden Seiten und vollenden den Beweis. □

Bemerkung: Zunächst halten wir eine Erweiterung der letzten Abschätzung fest. Es gilt

$$\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_2 \leq \sqrt{T}\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_\infty \leq T^{\frac{3}{2}}\|B\|\,\mathsf{V}_0^T\dot{\bar{u}}\,h^2.$$

Weiterhin betonen wir an dieser Stelle einen wichtigen Baustein im Beweis von Lemma 4.5.10. Wir verwenden auf Grund von $z_h(\bar{u}) - z_h(P_1\bar{u}) \in V_{h,0}[0,T,\mathbb{R}^n]$ und der daraus folgenden Stetigkeit den Sobolew'schen Einbettungssatz (siehe Anhang, Satz A.3.2) inklusive der Abschätzung

$$\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_\infty \leq \sqrt{T}\|\dot{z}_h(\bar{u}) - \dot{z}_h(P_1\bar{u})\|_2.$$

Notwendig für dieses Vorgehen ist aber der eindimensionale Definitionsbereich der Funktionen. Sind $z_h(\bar{u})$ und $z_h(P_1\bar{u})$ auf einem Gebiet $\Omega \subset \mathbb{R}^2$ mit $|\Omega| > 0$ definiert, so trifft die Aussage von Satz A.3.2 nicht mehr auf ihre Differenz zu. Es wäre dann nur noch

$$\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_\infty \leq c\,\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_{2,2}$$

möglich, allerdings ist $z_h(\bar{u}) - z_h(P_1\bar{u})$ auf Grund der stückweisen Linearität kein Element von $W_{2,0}^2[0,T,\mathbb{R}^n]$ und die Ungleichung ist daher nicht gültig. ◊

Satz 4.5.11. *Sei $\bar{u}$ die Lösung von* (StP) *und es gelte $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$. Weiter sei $u_h \in V_h[0,T,\mathbb{R}^m]$ die Lösung des diskreten Steuerungsproblems* $(\mathrm{StP})_h$*. Dann gilt für hinreichend kleines h*

$$\|u_h - P_1\bar{u}\|_2 \leq \frac{1}{3}(\mathsf{V}_0^T\dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty)\,h^{\frac{3}{2}}. \tag{4.17}$$

Beweis: Ausgangspunkt für den Beweis sind auch hier die notwendigen Optimalitätsbedingungen für $\bar{u}$ bzw. u_h aus Lemma 2.4.1 und Abschnitt 4.2. Es ist

$$\langle B^\mathsf{T}\bar{p} + \nu\bar{u}, u - \bar{u}\rangle \geq 0 \qquad \forall u \in U^{ad} \tag{2.17}$$

$$\langle B^\mathsf{T} p_h(u_h) + \nu u_h, \zeta_h - u_h\rangle \geq 0 \qquad \forall \zeta_h \in U_h^{ad}. \tag{4.7}$$

Da die Ungleichung (2.17) für alle $u \in U^{ad}$ gilt, schließen wir mit Lemma 2.4.3 auf die punktweise Gültigkeit, also

$$\big(B^\mathsf{T}\bar{p}(t) + \nu\bar{u}(t)\big)^\mathsf{T}\big(u - \bar{u}(t)\big) \geq 0 \qquad \forall' t \in [0,T], \forall u \in U$$

mit $U = \{u \in R^m : a \leq u \leq b\}$. Wir betrachten für $i = 0, \ldots, N$ die Spezialfälle $u = u_h(t_i)$ und folgern für die Stützstellen

$$\big(B^\mathsf{T}\bar{p}(t_i) + \nu\bar{u}(t_i)\big)^\mathsf{T}\big(u_h(t_i) - \bar{u}(t_i)\big) \geq 0, \qquad i = 0, \ldots, N.$$

An dieser Stelle betrachten wir allgemein zwei Funktionen $f, g \in V_h[0,T,\mathbb{R}^n]$ und die sie darstellenden Vektoren $\mathfrak{f}, \mathfrak{g} \in \mathbb{R}^{(N+1)n}$, also $\mathfrak{f}_i = f(t_i)$ und $\mathfrak{g}_i = g(t_i)$ für $i = 0, \ldots, N$. Falls $\mathfrak{f}_i^\mathsf{T}\mathfrak{g}_i \geq 0$ für $i = 0, \ldots, N$ gilt, so ist auch

$$0 \leq \frac{h}{2}\mathfrak{f}_0^\mathsf{T}\mathfrak{g}_0 + h\sum_{i=1}^{N-1}\mathfrak{f}_i^\mathsf{T}\mathfrak{g}_i + \frac{h}{2}\mathfrak{f}_N^\mathsf{T}\mathfrak{g}_N = \mathfrak{f}^\mathsf{T}\underbrace{\frac{h}{6}\begin{pmatrix} 3I & & & & \\ & 6I & & & \\ & & \ddots & & \\ & & & 6I & \\ & & & & 3I \end{pmatrix}}_{=:\mathfrak{H}}\mathfrak{g}.$$

Das L_2-Skalarprodukt der beiden Funktionen stellen wir ebenfalls als Produkt ihrer Vektoren im $\mathbb{R}^{(N+1)n}$ mit Hilfe der positiv definiten Matrix $\mathfrak{L}$ dar (vgl. (3.1)), daraus folgt

$$\langle f, g\rangle_2 = \mathfrak{f}^\mathsf{T}\mathfrak{L}\mathfrak{g} = \underbrace{\mathfrak{f}^\mathsf{T}\mathfrak{H}\mathfrak{g}}_{\geq 0} + \mathfrak{f}^\mathsf{T}\frac{h}{6}\begin{pmatrix} -I & I & & & \\ I & -2I & I & & \\ & & \ddots & & \\ & & I & -2I & I \\ & & & I & -I \end{pmatrix}\mathfrak{g},$$

Mit den Abkürzungen

$$\begin{aligned} (\mathfrak{f}_0)_{\mathfrak{h}\bar{\mathfrak{h}}} &= \frac{\mathfrak{f}_1 - \mathfrak{f}_0}{h^2}, \\ (\mathfrak{f}_i)_{\mathfrak{h}\bar{\mathfrak{h}}} &= \frac{\mathfrak{f}_{i+1} - 2\mathfrak{f}_i + \mathfrak{f}_{i-1}}{h^2}, \qquad i = 1, \ldots, N-1, \\ (\mathfrak{f}_N)_{\mathfrak{h}\bar{\mathfrak{h}}} &= \frac{\mathfrak{f}_{N-1} - \mathfrak{f}_N}{h^2} \end{aligned}$$

erhalten wir

$$0 \leq \langle f, g\rangle_2 - \frac{h^2}{6}\langle (\mathfrak{f})_{\mathfrak{h}_{\bar{\mathfrak{h}}}}, \mathfrak{g}\rangle_h.$$

Setzen wir nun $f = B^\mathsf{T}(P_1\bar{p}) + \nu(P_1\bar{u})$ und $g = u_h - P_1\bar{u}$, sowie $\mathfrak{f}_i = B^\mathsf{T}\bar{p}(t_i) + \nu(P_1\bar{u})(t_i)$ und $\mathfrak{g}_i = u_h(t_i) - (P_1\bar{u})(t_i)$ für $i = 0, \ldots, N$. Damit erhalten wir

$$\langle B^\mathsf{T}(P_1\bar{p}) + \nu(P_1\bar{u}), u_h - P_1\bar{u}\rangle - \frac{h^2}{6}\langle (B^\mathsf{T}(P_1\bar{p}) + \nu(P_1\bar{u}))_{\mathfrak{h}_{\bar{\mathfrak{h}}}}, u_h - P_1\bar{u}\rangle_h \geq 0.$$

Addieren wir diese Ungleichung und (4.7), so gelangen wir unter Verwendung von $\bar{u}(t_i) = (P_1\bar{u})(t_i)$ und $\bar{p}(t_i) = (P_1\bar{p})(t_i)$ für $i = 0, \ldots, N$ zu

$$\langle B^\mathsf{T}(P_1\bar{p} - p_h(u_h)) + \nu(P_1\bar{u} - u_h), u_h - P_1\bar{u}\rangle - \frac{h^2}{6}\langle (B^\mathsf{T}\bar{p} + \nu\bar{u})_{\mathfrak{h}_{\bar{\mathfrak{h}}}}, u_h - P_1\bar{u}\rangle_h \geq 0.$$

Das ist äquivalent zu

$$\begin{aligned}
\nu\|u_h - P_1\bar{u}\|_2^2 &\leq \langle P_1\bar{p} - p_h(u_h), B(u_h - P_1\bar{u})\rangle - \frac{h^2}{6}\langle (B^\mathsf{T}\bar{p} + \nu\bar{u})_{\mathfrak{h}_{\bar{\mathfrak{h}}}}, u_h - P_1\bar{u}\rangle_h \\
&= \langle P_1\bar{p} - p_h(P_1\bar{u}), B(u_h - P_1\bar{u})\rangle + \langle p_h(P_1\bar{u}) - p_h(u_h), B(u_h - P_1\bar{u})\rangle \\
&\quad - \frac{h^2}{6}\langle (B^\mathsf{T}\bar{p} + \nu\bar{u})_{\mathfrak{h}_{\bar{\mathfrak{h}}}}, u_h - P_1\bar{u}\rangle_h
\end{aligned}$$

und wir schätzen die entstandenen Skalarprodukte getrennt voneinander ab. Zunächst benutzen wir die Cauchy-Schwarz-Ungleichung und erhalten

$$\langle P_1\bar{p} - p_h(P_1\bar{u}), B(u_h - P_1\bar{u})\rangle \leq \sqrt{T}\|B\|\,\|P_1\bar{p} - p_h(P_1\bar{u})\|_\infty \|u_h - P_1\bar{u}\|_2$$

Weiterhin gilt

$$\begin{aligned}
\|P_1\bar{p} - p_h(P_1\bar{u})\|_\infty &\leq \|P_1\bar{p} - p_h(\bar{u})\|_\infty + \|p_h(\bar{u}) - p_h(P_1\bar{u})\|_\infty \\
&\overset{(4.5)}{\leq} \|P_1\bar{p} - p_h(\bar{u})\|_\infty + T^2\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_\infty \\
&\overset{(4.16)}{\leq} \left(\|P_1\bar{p} - p_h(\bar{u})\|_\infty + T^3\|B\|\,\mathsf{V}_0^T\dot{\bar{u}}\right) h^2 \\
&\overset{(4.12)}{\leq} \left(3c_{2.6}^2c_{2.7}^2\left(\|\mathcal{T}\bar{u}\|_\infty + \|z_d\|_\infty\right) + T^3\|B\|\,\mathsf{V}_0^T\dot{\bar{u}}\right) h^2.
\end{aligned}$$

Für den zweiten Ausdruck verwenden wir Lemma 4.2.5. Demnach sind die diskreten Operatoren bezüglich des L_2-Skalarprodukts adjungiert und wir schließen

$$\begin{aligned}
\langle p_h(P_1\bar{u}) - p_h(u_h), B(u_h - P_1\bar{u})\rangle &= \langle z_h(P_1\bar{u}) - z_h(u_h), z_h(u_h) - z_h(P_1\bar{u})\rangle \\
&= -\|z_h(P_1\bar{u}) - z_h(u_h)\|_2^2 \leq 0.
\end{aligned}$$

Beim letzten Summanden schließen wir nach Anwendung der Cauchy-Schwarz-Ungleichung mit Lemma 3.2.9 auf

$$\begin{aligned}
h^2\|(B^\mathsf{T}p(\bar{u}) + \nu\bar{u})_{\mathfrak{h}_{\bar{\mathfrak{h}}}}\|_h &\leq \left(\|B\|\,\|p(\bar{u})_{\mathfrak{h}_{\bar{\mathfrak{h}}}}\|_h + \nu\|(\bar{u})_{\mathfrak{h}_{\bar{\mathfrak{h}}}}\|_h\right) h^2 \\
&\leq 2\|B\|\,\|\ddot{p}(\bar{u})\|_2\, h^2 + 2\nu\left(\mathsf{V}_0^T\dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty\right) h^{\frac{3}{2}}.
\end{aligned}$$

Wir erhalten schließlich durch Addition der positiven Terme und beidseitigem Kürzen von $\|u_h - P_1\bar{u}\|_2$ zu

$$\|u_h - P_1\bar{u}\|_2 \leq +\sqrt{T}\frac{\|B\|}{\nu}\Big[\frac{1}{3}\|\ddot{p}(\bar{u})\|_\infty + 3c_{2.6}^2 c_{2.7}^2\left(\|\mathcal{T}\bar{u}\|_\infty + \|z_d\|_\infty\right) + T^3\|B\|\,\mathsf{V}_0^T\dot{\bar{u}}\Big]\,h^2 + \frac{1}{3}(\mathsf{V}_0^T\dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty)\,h^{\frac{3}{2}}$$

und es folgt für hinreichend kleine Schrittweite

$$\|u_h - P_1\bar{u}\|_2 \leq \frac{1}{3}(\mathsf{V}_0^T\dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty)\,h^{\frac{3}{2}}.$$

□

Damit ergeben sich automatisch weitere Ergebnisse.

Satz 4.5.12. *Seien $\bar{u}$ die Lösung von* (StP) *mit $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$ und $u_h \in V_h[0,T,\mathbb{R}^m]$ die Lösung des diskreten Steuerungsproblems* $(\mathrm{StP})_h$*. Dann gilt*

$$\|\bar{u} - u_h\|_2 \leq \frac{1}{3}\,(4\mathsf{V}_0^T\dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty)\,h^{\frac{3}{2}}.$$

Beweis: Die Aussage folgt mit der Dreiecks-Ungleichung aus dem letzten Satz und der Ungleichung (3.13). □

Für eine Fehlerabschätzung in der L_∞-Norm definieren wir wieder die Funktion $\tilde{u} = \Pi_{[a,b]}\big(-\frac{1}{\nu}B^\mathsf{T}p_h(u_h)\big)$ und gehen den schon bei stückweise konstanter Approximation der Steuerung vorgestellten Weg.

Satz 4.5.13. *Seien $\bar{u}$ die Lösung von* (StP) *mit $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$ und $u_h \in V_h[0,T,\mathbb{R}^m]$ die Lösung von* $(\mathrm{StP})_h$*. Dann gilt mit der Funktion $\tilde{u} = \Pi_{[a,b]}\big(-\frac{1}{\nu}B^\mathsf{T}p_h(u_h)\big)$ für hinreichend kleine Schrittweite*

$$\|\bar{u} - \tilde{u}\|_\infty \leq \frac{\|B\|^2}{3\nu}T^3\sqrt{T}\,(\mathsf{V}_0^T\dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty)\,h^{\frac{3}{2}}.$$

Beweis: Es ist

$$\|p_h(\bar{u}) - p_h(P_1\bar{u})\|_\infty \leq T^2\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_\infty \overset{(4.16)}{\leq} T^3\|B\|\,\mathsf{V}_0^T\dot{\bar{u}}\,h^2$$

und

$$\begin{aligned}\|p_h(P_1\bar{u}) - p_h(u_h)\|_\infty &\leq T^2\|z_h(P_1\bar{u}) - z_h(u_h)\|_\infty \\ &\leq T^3\sqrt{T}\|B\|\,\|P_1\bar{u} - u_h\|_2 \overset{(4.17)}{\leq} T^3\sqrt{T}\frac{\|B\|}{3}\,(\mathsf{V}_0^T\dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty)\,h^{\frac{3}{2}}.\end{aligned}$$

Weiterhin gilt laut Satz 4.3.4

$$\|\bar{p} - p_h(\bar{u})\|_\infty \leq 3c_{2.6}^2 c_{2.7}^2 \left(\|\mathcal{T}\bar{u}\|_\infty + \|z_d\|_\infty\right) h^2.$$

Demnach erhalten wir auf Grund der Lipschitz-Stetigkeit des Operators $\Pi_{[a,b]}$

$$\|\bar{u} - \tilde{u}\|_\infty \leq \frac{\|B\|}{\nu}\|\bar{p} - p_h(u_h)\|_\infty \leq \frac{\|B\|}{\nu}\Big(T^3\|B\|\,\mathsf{V}_0^T\dot{\bar{u}} + 3c_{2.6}^2 c_{2.7}^2 \left(\|\mathcal{T}\bar{u}\|_\infty + \|z_d\|_\infty\right)\Big) h^2 + T^3\sqrt{T}\frac{\|B\|^2}{3\nu}\left(\mathsf{V}_0^T\dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty\right) h^{\frac{3}{2}}.$$

Für hinreichend kleine Schrittweite folgt dann

$$\|\bar{u} - \tilde{u}\|_\infty \leq \frac{\|B\|^2}{3\nu}T^3\sqrt{T}\left(\mathsf{V}_0^T\dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty\right) h^{\frac{3}{2}}.$$

□

Damit gelingt es uns für stückweise lineare Approximation der Steuerung Konvergenz der Ordnung $\frac{3}{2}$ zu zeigen. Wie in Satz 4.5.12 bewiesen, erreichen wir diese Konvergenzgeschwindigkeit auch bei Betrachtung des Abstands von exakter und diskreter Lösung, allerdings nur im quadratischen Mittel. Durch die Definition von $\tilde{u}$ erhalten wir eine für das Ausgangsproblem zulässige Steuerung mit der gleichen Konvergenzordnung bei punktweiser Betrachtung des Fehlers. Die numerischen Ergebnisse weisen auf quadratische Konvergenzordnung hin, allerdings scheitert die oben durchgeführte Beweismethode (im Gegensatz zur stückweise konstanten Diskretisierung - vgl. Satz 4.5.7), da die diskrete Optimalitätsbedingung nicht auf ein endlich-dimensionales Skalarprodukt mit Diagonalmatrix zurückgeführt werden kann. Aus der punktweisen Gültigkeit der Optimalitätsbedingung für das kontinuierliche Problem erhalten wir allerdings eine solche Diagonalmatrix und damit entsteht ein Fehlerterm, welcher schließlich die niedrigere Konvergenzordnung nach sich zieht.

4.6 Numerische Durchführung

4.6.1 Bestimmung der diskreten Operatoren

Nach den theoretischen Vorarbeiten mit dem Nachweis der quadratischen Konvergenz steht nun die Beschreibung der praktischen Durchführung im Mittelpunkt des Interesses. Dazu bestimmen wir im ersten Schritt die genaue Gestalt der diskreten Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$ als Abbildungen zwischen den endlich-dimensionalen Funktionenräumen der diskreten Steuerungen und der diskreten Zustände.

Das Ziel einer Darstellung der Funktion z_h als Funktion von u_h erreichen wir über mittels Umweg über die äquivalenten Vektoren β und α (vgl. die Definition der diskreten Räume in Abschnitt 3.1). Die diskrete Systemgleichung (4.1) muss für alle

$v_h \in V_{h,0}[0,T,\mathbb{R}^n]$ erfüllt sein, dazu reicht es auf Grund der Linearität aus, die Gültigkeit für alle Basiselemente von $V_{h,0}[0,T,\mathbb{R}^n]$ zu fordern. Wir erhalten durch folgenden Ansatz

$$\int_0^T \dot{z}_h(t)\dot{v}_h^{(j)}(t) + Az_h(t)v_h^{(j)}(t)\,dt = \int_0^T (Bu_h(t)+e(t))v_h^{(j)}(t)\,dt, \quad j = 1,\dots,N-1$$

ein endlich-dimensionales Gleichungssystem. Die einzelnen Summanden bearbeiten wir für eine bessere Übersicht getrennt. Für die Behandlung der Ableitungen der $v_h^{(j)}$ erinnern wir zunächst an ihre Definition

$$v_h^{(j)}(t) = \begin{cases} \dfrac{t-t_{j-1}}{h}, & t \in T_{j-1} \\ \dfrac{t_{j+1}-t}{h}, & t \in T_j \\ 0, & \text{sonst} \end{cases}, \qquad j = 1,\dots,N-1.$$

Daraus folgern wir

$$\dot{v}_h^{(j)}(t) = \begin{cases} 1/h, & t \in (t_{j-1},t_j) \\ -1/h, & t \in (t_j,t_{j+1}) \\ 0, & \text{sonst} \end{cases}, \qquad j = 1,\dots,N-1.$$

Die Berechnung eines Funktionswerts von $z_h = \sum_{j=1}^{N-1} \beta_j v_h^{(j)}$ erfolgt für ein beliebiges $t \in T_i$ und $i = 0,\dots,N-1$ durch

$$z_h(t) = \beta_i v_h^{(i)}(t) + \beta_{i+1} v_h^{(i+1)}(t) = \frac{\beta_{i+1}-\beta_i}{h}(t-t_i) + \beta_i.$$

Für die Ableitung gilt dann folgendes

$$\dot{z}_h(t) = \frac{\beta_{i+1}-\beta_i}{h}, \qquad \forall t \in (t_i,t_{i+1}), \quad i = 0,\dots,N-1.$$

Dabei ist zu beachten, dass wir die Ableitungen von v_h und z_h punktweise verstehen, d.h. ihre Definition schließt die Punkte t_i nicht ein, da die Funktionen dort im klassischen Sinn nicht differenzierbar sind. Bei der Integration über das Intervall $[0,T]$ ist diese Tatsache nicht von Relevanz, da der Wert des Integrals durch Veränderungen des Integranden auf einer Nullmenge nicht beeinflusst wird. Die Funktionen $\dot{v}_h$ und $\dot{z}_h$ sind also nur für fast alle $t \in [0,T]$ definiert, denn die Menge $\{t_0,\dots,t_N\}$ bildet eine Nullmenge in $[0,T]$. Falls wir dennoch einmal den Wert an einer solchen Teilintervallgrenze benötigen, so fassen wir $\dot{v}_h(t_i)$ als Grenzwert von der jeweils betrachteten Seite auf.

Bevor wir die verschiedenen Summanden untersuchen, verabreden wir noch eine verkürzende Schreibweise. Auf Grund der Randwertbedingungen in der Definition von

$V_{h,0}[0,T,\mathbb{R}^n]$ benötigen wir keine Basisfunktionen $v_h^{(0)}$ und $v_h^{(N)}$ und demzufolge auch keine Koeffizienten β_0 und β_N. Wir vereinbaren für sie im Folgenden den Wert 0_n.

Schon in den Untersuchungen vor Lemma 4.3.3 betrachteten wir die Veränderungen der linken Seite der diskreten Systemgleichung beim Einsetzen der Basisfunktionen $v_h^{(j)}$. Wir erhalten

$$-\frac{1}{h}(\beta_{j+1}-2\beta_j+\beta_{j-1})+\frac{h}{6}A(\beta_{j+1}+4\beta_j+\beta_{j-1}), \qquad j=1,\dots,N-1,$$

was wir in Matrixschreibweise darstellen durch

$$\left[\underbrace{\frac{1}{h}\begin{pmatrix} 2I & -I & & & \\ -I & 2I & -I & & \\ & & \ddots & & \\ & & -I & 2I & -I \\ & & & -I & 2I \end{pmatrix}+\frac{h}{6}\widetilde{A}\begin{pmatrix} 4I & I & & & \\ I & 4I & I & & \\ & & \ddots & & \\ & & I & 4I & I \\ & & & I & 4I \end{pmatrix}}_{=:\mathfrak{R}^{-1}}\right]\begin{pmatrix}\beta_1\\ \vdots\\ \beta_{N-1}\end{pmatrix}$$

mit

$$\widetilde{A}=\begin{pmatrix} A & & \\ & \ddots & \\ & & A \end{pmatrix}.$$

Die Herleitung der rechten Seite unterscheiden wir nach stückweise konstanter und stückweise linearer Approximation der Steuerung. Es ist zunächst für $u_h \in U_h[0,T,\mathbb{R}^m]$ und $j=1,\dots,N-1$

$$\int_0^T (Bu_h(t))v_h^{(j)}(t)\,dt=\int_0^T\sum_{l=0}^{N-1}\alpha_l\,u_h^{(l)}(t)v_h^{(j)}(t)\,dt=\frac{h}{2}B(\alpha_{j-1}+\alpha_j).$$

In Matrixschreibweise zusammengefasst lautet die rechte Seite wie folgt

$$\underbrace{\frac{h}{2}\widetilde{B}\begin{pmatrix} I & I & & & \\ & I & I & & \\ & & \ddots & \ddots & \\ & & & I & I \\ & & & & I & I \end{pmatrix}}_{\mathfrak{C}_0}\begin{pmatrix}\alpha_0\\ \vdots\\ \alpha_{N-1}\end{pmatrix} \quad \text{mit} \quad \widetilde{B}=\begin{pmatrix} B & & \\ & \ddots & \\ & & B \end{pmatrix}.$$

Im Fall $u_h \in V_h[0,T,\mathbb{R}^m]$ erfolgt die Berechnung der rechten Seite ähnlich der linken, und zwar ist für $j=1,\dots,N-1$

$$\int_0^T (Bu_h(t))v_h^{(j)}(t)\,dt=\int_0^T\sum_{l=0}^{N-1}\alpha_l\,v_h^{(l)}(t)v_h^{(j)}(t)\,dt=\frac{h}{6}B(\alpha_{j+1}+4\alpha_j+\alpha_{j-1}).$$

Auch hier drücken wir die Gleichungen mit Hilfe von Matrizen aus

$$\underbrace{\frac{h}{6}\widetilde{B}\begin{pmatrix} I & 4I & I & & & & \\ & I & 4I & I & & & \\ & & & \ddots & & & \\ & & & I & 4I & I & \\ & & & & I & 4I & I \end{pmatrix}}_{=:\mathfrak{C}_1}\begin{pmatrix}\alpha_0\\ \vdots \\ \alpha_N\end{pmatrix}.$$

Nun kommen wir zum additiven Anteil e, den wir ebenfalls zu diskretisieren haben. Definieren wir

$$e_j := \int_0^T e(t)v_h^{(j)}(t)\,dt, \qquad j = 1,\dots,N-1$$

und setzen die Werte zusammen, so erhalten wir daraus den Vektor

$$e_h = (e_1^\mathsf{T},\dots,e_{N-1}^\mathsf{T})^\mathsf{T}.$$

Die Integrale sind nur für eine kleine Klasse von Funktionen analytisch lösbar. Bei komplizierteren Ausdrücken wenden wir Verfahren der numerischen Integration an.

Jetzt sind wir in der Lage die Koeffizienten des Zustandes z_h als Funktion der Koeffizienten der diskreten Steuerung u_h darzustellen, und zwar gilt mit $j \in \{1,2\}$

$$(\mathfrak{N}^{-1})\beta = \mathfrak{C}_j\alpha + e_h \Longleftrightarrow \beta = \mathfrak{N}(\mathfrak{C}_j\alpha + e_h), \tag{4.18}$$

wobei die Gestalt der Matrix $\mathfrak{A}$ als Diskretisierung des Operators $z \mapsto -\ddot{z}$ für unsere Problemstellung fest bleibt. Dagegen sind die Matrizen $\mathfrak{B}$, $\mathfrak{C}_0$ und $\mathfrak{C}_1$ von den Problemparametern A und B abhängig.

4.6.2 Ein Beispielproblem

Für ein Beispiel setzen wir die Parameter auf konkrete Werte. Seien $A = B = I_n$ und $z_d = d \in \mathbb{R}^n$, also konstant. Die Funktion e sei gegeben durch $e(t) = -2 + t^2 - t - \min\{-\frac{t^2-t}{\nu}, b\}$ (vgl. Hinze [14] (2005)). Das kontinuierliche Problem lautet dann folgender Maßen:

$$\min_u \frac{1}{2}\int_0^T |z(u)(t) - d|^2 + \nu|u|^2\,dt, \qquad u \in U^{ad}$$

unter den Nebenbedingungen:

$$\begin{aligned} -\ddot{z}(t) + z(t) &= u(t) + e(t) \qquad \forall' t \in [0,T] \\ z(0) &= z(T) = 0. \end{aligned}$$

Die diskreten Optimierungsprobleme ergeben sich zu

$$\min_{u_h} \frac{1}{2} \int_0^T |z_h(u_h)(t) - d|^2 + \nu |u_h(t)|^2 \, dt, \qquad u_h \in U_h^{ad}$$

mit

$$U_h^{ad} = U^{ad} \cap U_h[0, T, \mathbb{R}^m] \qquad \text{bzw.} \qquad U_h^{ad} = U^{ad} \cap V_h[0, T, \mathbb{R}^m]$$

und unter den Nebenbedingungen

$$\int_0^T \dot{z}_h(t)^\mathsf{T} \dot{v}_h(t) + z_h(t)^\mathsf{T} v_h(t) \, dt = \int_0^T (u_h(t) + e(t))^\mathsf{T} v_h(t) \, dt \qquad \forall v_h \in V_{h,0}[0, T, \mathbb{R}^n].$$

Die Systemgleichung in Vektorschreibweise mit den wie oben verwendeten Bezeichnungen β und α übernehmen wir aus (4.18)

$$\beta = \mathfrak{N}(\mathfrak{C}_j \alpha + e_h), \qquad j \in \{1, 2\},$$

wobei j durch die Approximationsart der Steuerung festgelegt wird. Wir leiten nun die Zielfunktion in den Variablen α her und bearbeiten zunächst den Term

$$\|z_h(u_h) - z_d\|_2^2 = \|z_h(u_h)\|_2^2 + \|z_d\|_2^2 - 2\langle z_h(u_h), z_d \rangle.$$

Den ersten Teil übernehmen wir aus Formel (3.2)

$$\begin{aligned} \|z_h(u_h)\|_2^2 = \beta^\mathsf{T} \mathfrak{L}_0 \beta &= (\mathfrak{C}_j \alpha + e_h)^\mathsf{T} \mathfrak{N}^\mathsf{T} \mathfrak{L}_0 \mathfrak{N} (\mathfrak{C}_j \alpha + e_h) \\ &= \left[\alpha^\mathsf{T} (\mathfrak{N}\mathfrak{C}_j)^\mathsf{T} \mathfrak{L}_0 \mathfrak{N} \mathfrak{C}_j \alpha + 2\big((\mathfrak{N}\mathfrak{C}_j)^\mathsf{T} \mathfrak{L}_0 \mathfrak{N} e_h\big)^\mathsf{T} \alpha + e_h^\mathsf{T} \mathfrak{N} \mathfrak{L}_0 \mathfrak{N} e_h \right]. \end{aligned}$$

Das Skalarprodukt vereinfachen wir mit Hilfe des Vektors $D = (d^\mathsf{T}, \ldots, d^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{(N-1)n}$

$$\begin{aligned} \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} z_h(t)^\mathsf{T} z_d(t) \, dt &= \sum_{i=0}^{N-1} \left[(\beta_i)_\mathsf{h}^\mathsf{T} d \int_{t_i}^{t_{i+1}} (t - t_i) \, dt + h \beta_j^\mathsf{T} d \right] \\ &= h \sum_{i=0}^{N-1} \frac{(\beta_{i+1} + \beta_i)^\mathsf{T}}{2} d \\ \beta_0 = \beta_N = 0_n \qquad &= h \sum_{i=1}^{N-1} \beta_i^\mathsf{T} d = h \beta^\mathsf{T} D \\ &= h (\mathfrak{C}_j \alpha + e_h)^\mathsf{T} \mathfrak{N}^\mathsf{T} D \\ &= h (\mathfrak{C}_j^\mathsf{T} \mathfrak{N}^\mathsf{T} D)^\mathsf{T} \alpha + h D^\mathsf{T} \mathfrak{N} e_h. \end{aligned}$$

Also ergibt sich für den ersten Term in der Zielfunktion

$$\begin{aligned}\frac{1}{2}\|z_h(u_h)-z_d\|_2^2 &= \frac{1}{2}\|z_h(u_h)\|_2^2 - \int_0^T z_h(t)^\mathsf{T} z_d(t)\,dt + \frac{1}{2}\|z_d\|_2^2\\ &= \frac{1}{2}\beta^\mathsf{T}\mathfrak{L}_0\beta - h\beta^\mathsf{T}D + \frac{1}{2}\|z_d\|_2^2\\ &= \frac{1}{2}\alpha^\mathsf{T}(\mathfrak{N}\mathfrak{C}_j)^\mathsf{T}\mathfrak{L}_0\mathfrak{N}\mathfrak{C}_j\alpha + \underbrace{\left[(\mathfrak{N}\mathfrak{C}_j)^\mathsf{T}\mathfrak{L}_0\mathfrak{N}e_h - h(\mathfrak{N}\mathfrak{C}_j)^\mathsf{T}D\right]^\mathsf{T}}_{=:f_j}\alpha\\ &\quad + \underbrace{\frac{1}{2}e_h^\mathsf{T}\mathfrak{N}\mathfrak{L}_0\mathfrak{N}e_h - hD^\mathsf{T}\mathfrak{N}e_h + \frac{T}{2}|d|^2}_{=:c_j}.\end{aligned}$$

Der Term $\frac{\nu}{2}\|u_h\|_2^2$ wandelt sich bei stückweise konstanter Approximation der Steuerung zu $\frac{h\nu}{2}\,\alpha^\mathsf{T}\alpha$ und bei stückweise linearer Approximation zu $\frac{\nu}{2}\,\alpha^\mathsf{T}\mathfrak{L}\alpha$. Die Einträge von α gehen also quadratisch in J ein und sowohl f als auch c werden nicht verändert. So erhalten wir für die Zielfunktion folgenden Ausdruck

$$J_h(\alpha) = \frac{1}{2}\alpha^\mathsf{T}\underbrace{\left((\mathfrak{N}\mathfrak{C}_0)^\mathsf{T}\mathfrak{L}_0\mathfrak{N}\mathfrak{C}_0 + h\nu I_{(N+1)n}\right)}_{=:H_\nu}\alpha + f_0^\mathsf{T}\alpha + c_0$$

bzw.

$$J_h(\alpha) = \frac{1}{2}\alpha^\mathsf{T}\underbrace{\left((\mathfrak{N}\mathfrak{C}_1)^\mathsf{T}\mathfrak{L}_0\mathfrak{N}\mathfrak{C}_1 + \nu\mathfrak{L}\right)}_{=:H_\nu}\alpha + f_1^\mathsf{T}\alpha + c_1.$$

Die Hesse-Matrix H_ν ist offensichtlich symmetrisch. Da der zweite Summand positiv definit und der erste positiv semi-definit ist, folgt die positive Definitheit für die Summe. Damit existiert für das diskrete Problem ein eindeutig bestimmtes Minimum.

Wählen wir im Fall $n = 1$ die Parameter $\nu = 0.1$, $T = 1$, $d = 2$, $a = -\infty$ und $b = 2.5(\sqrt{2}-1)$, so ist die optimale Steuerung für unsere Problemstellung gleich der Funktion $\bar{u}(t) = \min\{-\frac{t^2-t}{\nu}, b\}$. Es ergibt sich der optimale Zustand $z(\bar{u})(t) = t - t^2$. Der optimale adjungierte Zustand ist in diesem Beispiel identisch zu $z(\bar{u})$. Beide Aussagen prüfen wir leicht durch Einsetzen in die jeweilige Gleichung nach. Die Schnittpunkte von $\bar{u}$ und der oberen Schranke b sind an irrationalen Stellen, daher werden sie wegen $N \in \mathbb{N}$ niemals zu Stützstellen.

In den Tabellen 4.1 und 4.2 sind im Fall stückweise konstanter Approximation der Steuerung bezüglich verschiedener Schrittweiten die berechneten Werte für den Zielfunktionswert, die Norm von u_h sowie einige Fehlerterme. Zunächst führen wir zum Vergleich den maximalen Abstand der beiden Lösungen an. Darüber hinaus finden wir den Fehler der neu definierten Steuerung $\tilde{u} = \Pi_{[\mathrm{a,b}]}\left(-\frac{1}{\nu}B^\mathsf{T}p_h(u_h)\right)$ in der L_2-Norm und in der L_∞-Norm, sowie letzteren dividiert durch das Quadrat der Schrittweite. An

h	$J_h(u_h)$	$\|u_h\|_2$
1/3	2.6535	1.0355
1/4	2.5621	0.9881
1/6	2.4654	0.9360
1/10	2.4017	0.9478
1/20	2.3710	0.9543
1/50	2.3618	0.9537
1/100	2.3604	0.9541

Tabelle 4.1: Daten der diskreten Steuerungsprobleme bei stückweise konstanter Approximation.

h	$\|\bar{u} - u_h\|_\infty$	$\|\bar{u} - \tilde{u}\|_2$	$\|\bar{u} - \tilde{u}\|_\infty$	$\|\bar{u} - \tilde{u}\|_\infty / h^2$
1/3	1.0355	0.0904	0.2532	2.2787
1/4	0.9382	0.0565	0.1550	2.4797
1/6	0.6952	0.0266	0.0686	2.4712
1/10	0.4502	0.0082	0.0248	2.4843
1/20	0.2375	0.0021	0.0062	2.4953
1/50	0.0980	$3.5331 \cdot 10^{-4}$	$9.9840 \cdot 10^{-4}$	2.4960
1/100	0.0495	$8.8269 \cdot 10^{-5}$	$2.4985 \cdot 10^{-4}$	2.4985

Tabelle 4.2: Fehler der Steuerungen bei stückweise konstanter Approximation.

Hand der Konstanz der Werte in der letzten Spalte erhalten wir eine Bestätigung für unsere theoretischen Resultate.

In den Abbildung 4.1 bzw. 4.2 sehen wir den L_2- bzw. L_∞-Fehler gegen die Schrittweite h abgetragen und erkennen jeweils deutlich die quadratische Abnahme mit sich verringernder Schrittweite.

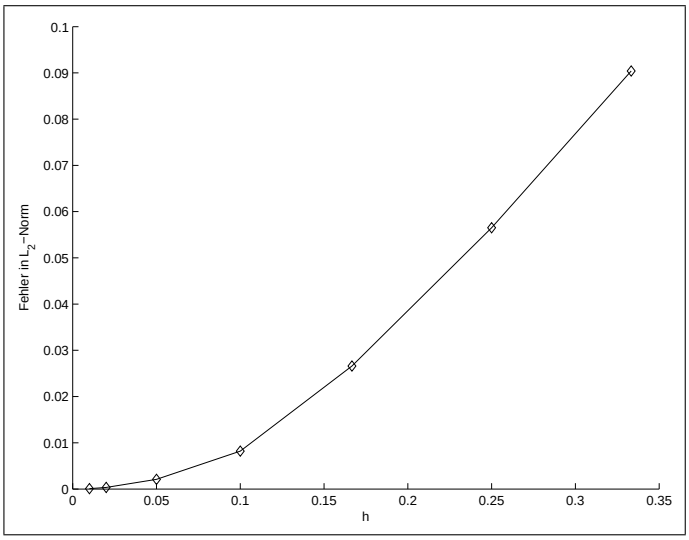

Abbildung 4.1: L_2-Fehler bei Finiten Elementen und stückweise konstanter Approximation

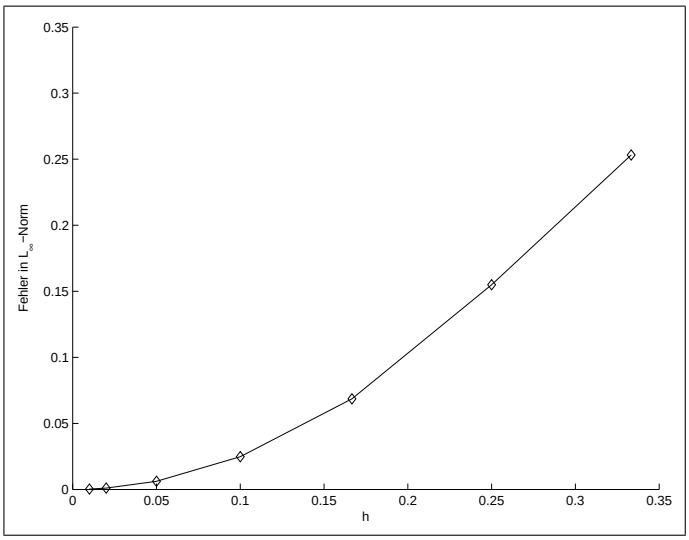

Abbildung 4.2: L_∞-Fehler bei Finiten Elementen und stückweise konstanter Approximation

Auch für den Fall stückweise linearer Approximation geben wir die Ergebnisse zunächst in den Tabellen 4.3 und 4.4 wieder und tragen anschließend in den Abbildungen 4.3 und 4.4 die Fehler gegen die Schrittweite ab. Wir erkennen in beiden Fällen deutlich die quadratische Abnahme mit sich verringernder Schrittweite.

h	$J_h(u_h)$	$\|u_h\|_2$
1/3	2.6514	0.7718
1/4	2.5592	0.8455
1/6	2.4632	0.9133
1/10	2.4011	0.9640
1/20	2.3709	0.9541
1/50	2.3618	0.9538
1/100	2.3604	0.9540

Tabelle 4.3: Daten der diskreten Steuerungsprobleme bei stückweise linearer Approximation.

h	$\|\bar{u} - u_h\|_\infty$	$\|\bar{u} - \tilde{u}\|_2$	$\|\bar{u} - \tilde{u}\|_\infty$	$\|\bar{u} - \tilde{u}\|_\infty / h^2$
1/3	0.5964	0.0896	0.2509	2.2581
1/4	0.4216	0.0561	0.1539	2.4624
1/6	0.2533	0.0267	0.0688	2.4768
1/10	0.1355	0.0083	0.0250	2.5000
1/20	0.0645	0.0021	0.0062	2.4800
1/50	0.0171	$3.5530 \cdot 10^{-4}$	$9.9918 \cdot 10^{-4}$	2.4979
1/100	0.0129	$8.8435 \cdot 10^{-5}$	$2.4988 \cdot 10^{-4}$	2.4988

Tabelle 4.4: Fehler der Steuerungen bei stückweise linearer Approximation.

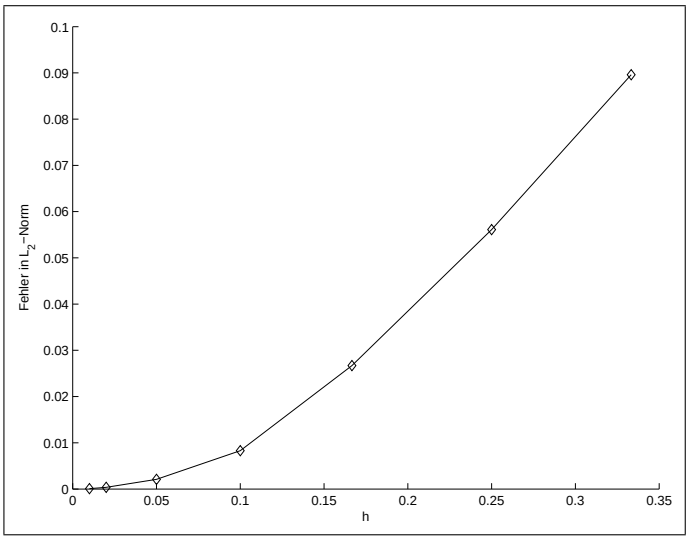

Abbildung 4.3: L_2-Fehler bei Finiten Elementen und stückweise linearer Approximation

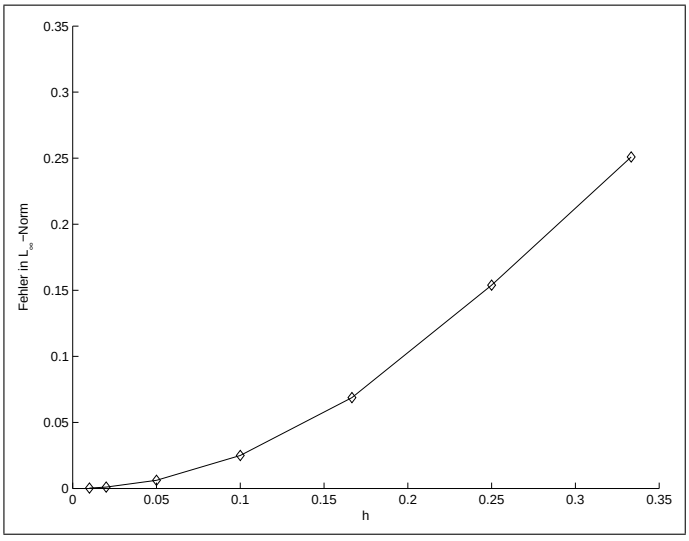

Abbildung 4.4: L_∞-Fehler bei Finiten Elementen und stückweise linearer Approximation

4.7 Veränderte Diskretisierung

In den bisherigen Betrachtungen der Finite Elemente Methode verwendeten wir das *Ritz-Galerkin-Verfahren* und diskretisierten die Systemgleichung (1.4) durch das Lösen der zugehörigen Variationsgleichung in einem endlich-dimensionalem Unterraum des $W^1_{2,0}[0,T,\mathbb{R}^n]$. In der Bemerkung nach Satz 2.2.5 notierten wir kurz einen anderen Beweis für die Eindeutigkeit der Lösung für jede rechte Seite. Die dort vorgestellte Idee greifen wir nun auf und untersuchen eine etwas veränderte Diskretisierung. Wir betrachten wieder das Funktional $I(z) = a(z,z) - \langle y, z\rangle$ und die Minimierungsaufgabe

$$\min_{z\in W^1_{2,0}} a(z,z) - \langle y,z\rangle = \min_{z\in W^1_{2,0}} \int_0^T \dot z(t)^\mathsf{T}\dot z(t) + \big(Az(t) - y(t)\big)^\mathsf{T} z(t)\,dt.$$

Die Systemgleichung (1.4) ist bekanntlich die notwendige und hinreichende Optimalitätsbedingung für die Lösung $z \in W^1_{2,0}[0,T,\mathbb{R}^n]$. Nun approximieren wir das Integral durch die Trapezregel auf dem Gitter $\{t_i = ih, i = 0,\dots,N\}$, also

$$I_h(z) = h\sum_{i=1}^{N-1} \dot z(t_i)^\mathsf{T}\dot z(t_i) + z(t_i)^\mathsf{T} A z(t_i) - y(t_i)^\mathsf{T} z(t_i)$$

und die Minimierungsaufgabe durch $\min\limits_{z\in W^1_{2,0}} I_h(z)$. Jetzt gehen wir ähnlich dem *Ritz-Galerkin-Verfahren* aus Abschnitt 4.2 vor und betrachten zunächst die notwendige Optimalitätsbedingung für die Lösung z

$$h\sum_{i=1}^{N-1} \dot z(t_i)^\mathsf{T}\dot v(t_i) + z(t_i)^\mathsf{T} A v(t_i) - y(t_i)^\mathsf{T} v(t_i) = 0 \qquad \forall v \in W^1_{2,0}[0,T,\mathbb{R}^n].$$

Die Vervollständigung der Diskretisierung erfolgt, indem wir in der notwendigen Optimalitätsbedingung die Forderung „$\forall v \in W^1_{2,0}[0,T,\mathbb{R}^n]$“ durch „$\forall v_h \in V_{h,0}[0,T,\mathbb{R}^n]$“ ersetzen. Dabei ist zu beachten, dass die Ableitung einer Funktion aus dem Raum $V_h[0,T,\mathbb{R}^n]$ an den Stützstellen nicht definiert ist. Daher verwenden wir für den ersten Teil auf der linken Seite die Schreibweise $\dot z(t) = (z(t_i))_\mathsf{h}$ für $t \in T_i$

$$h\sum_{i=0}^{N-1} \big(z_h(t_i)\big)_\mathsf{h}^\mathsf{T} \big(v_h(t_i)\big)_\mathsf{h} + z_h(t_i)^\mathsf{T} A v_h(t_i) = \sum_{i=1}^{N-1} y(t_i)^\mathsf{T} v_h(t_i) \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n].$$

Definieren wir in Anlehnung an $a(\cdot,\cdot)$ aus (2.2) die diskrete Bilinearform $a_h(\cdot,\cdot)$ durch

$$a_h(z_h, v_h) = \sum_{i=0}^{N-1} \big(z_h(t_i)\big)_\mathsf{h}^\mathsf{T} \big(v_h(t_i)\big)_\mathsf{h} + z_h(t_i)^\mathsf{T} A v_h(t_i), \tag{4.19}$$

so erhalten wir eine andere Diskretisierung der Systemgleichung (1.4)

$$a_h(z_h, v_h) = \langle Bu_h + e, v_h\rangle_h \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n] \tag{4.20}$$

und auch ein anderes diskretes Steuerungsproblem

$$(\widetilde{\mathrm{StP}})_h \qquad \min_{z_h,u_h} \widetilde{J}_h(z_h,u_h) = \min_{z_h,u_h} \frac{1}{2}\|z_h - z_d\|_2^2 + \frac{\nu}{2}\|u_h\|_2^2$$

unter den Nebenbedingungen

$$\begin{aligned} a_h(z_h,v_h) &= \langle Bu_h + e, v_h\rangle_h \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n] \\ z_h(0) &= z_h(T) = 0 \end{aligned}$$

und den Kontrollrestriktionen

$$a \le u_h(t) \le b \qquad \forall t \in [0,T].$$

Die Minimierung soll dabei über einen geeignet gewählten endlich-dimensionalen Unterraum des $L_2[0,T,\mathbb{R}^n]$ laufen, wir wählen dafür den $V_h[0,T,\mathbb{R}^m]$, approximieren die Steuerungen also stückweise linear.

Die in diesem Abschnitt vorgestellte Diskretisierungsmethode ähnelt den Finiten Elementen. Durch den Wegfall der Integrale besitzt sie einen großen Vorteil bei der späteren numerischen Umsetzung, allerdings ist die Voraussetzung der Stetigkeit an die rechte Seite der diskreten Systemgleichung unverzichtbar. Bei den später folgenden Fehlerabschätzungen sind nur Eigenschaften der optimalen Steuerung von Bedeutung, und von dieser wiesen wir die Stetigkeit bereits in Lemma 2.4.1 nach, so dass einer Anwendung auf unser Problem nichts im Wege steht. Auch in diesem Fall verwenden wir die Basisfunktionen aus Abschnitt 3.1 als Testfunktionen in der diskreten Systemgleichung. Setzen wir die $v_h^{(j)}$ für $j = 1,\dots,N-1$ nacheinander ein, so erhalten wir das Gleichungssystem

$$\begin{aligned} -\frac{\beta_{i+1} - 2\beta_i + \beta_{i-1}}{h^2} + A\beta_i &= (Bu+e)(t_i), \qquad i = 1,\dots,N-1 \\ \beta_0 = \beta_N &= 0_n. \end{aligned}$$

Wir verschieben die nähere Untersuchung dieser Diskretisierungsmethode in das folgende Kapitel, denn wir haben auf diesem Weg eine sehr anschauliche Motivation für das Verfahren der *Finiten Differenzen* oder auch *Differenzenverfahren* gefunden.

Kapitel 5

Differenzenverfahren

5.1 Motivation

Wie schon im vergangenen Kapitel greifen wir auch mit dieser Diskretisierungsmethode an der Systemgleichung (1.4) an. Wir ersetzen die Differentialgleichung durch ein Differenzenschema bzw. wir approximieren die Differentialquotienten durch Differenzenquotienten. Ausgangspunkt ist die Taylor-Reihe

$$z(t+h) = z(t) + h\dot{z}(t) + \frac{h^2}{2}\ddot{z}(t) + R(h),$$

wobei $|R(h)|h^{-2} \to 0$ für $h \to 0$ gilt. Mit Hilfe der Schrittweite $0 < h = T/N$ unterteilen wir das Intervall $[0, T]$ äquidistant durch die Stützstellen $0 = t_0 < t_1 < \ldots < t_N = T$ mit $t_i = ih$ für $i = 0, \ldots, N$. Bezeichnen wir die Werte an den Diskretisierungsstellen mit $z_i = z(t_i)$, so folgt

$$\begin{aligned} z_{i+1} &= z_i + h\dot{z}(t_i) + \frac{h^2}{2}\ddot{z}(t_i) + R_+(h) \\ z_{i-1} &= z_i - h\dot{z}(t_i) + \frac{h^2}{2}\ddot{z}(t_i) + R_-(h). \end{aligned}$$

Durch Addition der beiden Gleichungen und Vernachlässigung der Restterme, gelangen wir zu

$$\frac{z_{i+1} - 2z_i + z_{i-1}}{h^2} \approx \ddot{z}(t_i), \qquad i = 1, \ldots, N-1$$

als Approximation der zweiten Ableitung im Punkt $z(t_i)$ durch einen Differenzenquotienten. Die Divisionen und Ableitungen sind dabei jeweils komponentenweise aufzufassen. In den Fällen $i = 1$ und $i = N-1$ beachten wir zusätzlich die vorgegebenen Werte $z_0 = z_N = 0_n$. Die Systemgleichung (1.4) wird demzufolge diskretisiert durch das System von Differenzengleichungen

$$-\frac{z_{i+1} - 2z_i + z_{i-1}}{h^2} + Az_i = y(t_i) = y_i, \qquad i = 1, \ldots, N-1.$$

5.2 Diskretisierung und Stabilität

Im einleitenden Abschnitt gaben wir eine Motivationen für die Approximation der Systemgleichung (1.4) durch das Differenzenschema

$$\begin{aligned} -\frac{\beta_{i+1} - 2\beta_i + \beta_{i-1}}{h^2} + A\beta_i = y_i, \qquad & i = 1, \dots, N-1 \\ \beta_0 = \beta_N = 0_n. & \end{aligned} \tag{5.1}$$

Wir ersetzen demnach die kontinuierliche Bedingung an die gesamte Funktion z durch Bedingungen an den Stützstellen. Aus den Vektoren $\beta = (\beta_0^\mathsf{T}, \dots, \beta_N^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{(N+1)n}$ definieren wir anschließend die diskrete Lösung als stückweise lineare Verbindung der Punkte. Das bedeutet, die diskrete Lösung $z_h \in V_{h,0}[0, T, \mathbb{R}^n]$ erfüllt die Bedingungen

$$z_h(t_i) = \beta_i, \qquad i = 0, \dots, N.$$

Nach diesen Vorbetrachtungen folgt nun ebenfalls eine Diskretisierung des Zielfunktionals und wir stellen das endlich-dimensionale Steuerungsproblem auf mit

$$(\widetilde{\mathrm{StP}})_h \qquad \min_{z_h, u_h} \widetilde{J}_h(z_h, u_h) = \min_{z_h, u_h} \frac{1}{2}\|z_h - z_d\|_h^2 + \frac{\nu}{2}\|u_h\|_h^2$$

unter den Nebenbedingungen

$$\begin{aligned} z_h(t)\Big|_{T_i} &= (t - t_i)\frac{\beta_{i+1} - \beta_i}{h} + \beta_i, && i = 0, \dots, N-1 \\ (Bu_h + e)(t_i) &= -\frac{\beta_{i+1} - 2\beta_i + \beta_{i-1}}{h^2} + A\beta_i, && i = 1, \dots, N-1 \\ \beta_0 &= \beta_N = 0_n \end{aligned}$$

und den Kontrollrestriktionen

$$u_h \in U_h^{ad} \subset U^{ad}$$

mit einer nichtleeren, abgeschlossenen und konvexen Menge U_h^{ad}. Da das Differenzenschema (5.1) die Werte der rechten Seite an den Stützstellen benötigt, greifen wir zumeist auf stückweise lineare Approximation der Steuerung zurück, also

$$U_h^{ad} = V_h[0, T, \mathbb{R}^m] \cap U^{ad}.$$

Bemerkung: Wie im kontinuierlichen Fall (vgl. Abschnitt 2.5) ist auch hier die Untersuchung von allgemeinen Randwertbedingungen $\beta_0 = z_0$ und $\beta_N = z_T$ mit $z_0, z_T \in \mathbb{R}^n$ möglich. In diesem Fall wäre der erste Schritt eine äquivalente Umformulierung des Differenzenschemas, indem wir die Vektoren $v_i = \frac{i}{N}(z_T - z_0) + z_0$ für $i = 0, \dots, N$ definieren. Angenommen $\beta - v$ wäre Lösung des Differenzenschemas (5.1) mit $y_i + Av_i$ als rechte Seite und den Randwerten $(\beta - v)_0 = 0_n$ und $(\beta - v)_N = 0_n$, dann ist β Lösung von (5.1) mit den Randwerten $\beta_0 = z_0$ und $\beta_N = z_T$. Die Begründung liegt in $v_{i+1} - 2v_i + v_{i-1} = 0_n$ für $i = 1, \dots, N-1$. Verbunden mit $(\beta - v)_0 = z_0 \iff \beta_0 = 0_n$ bzw. $(\beta - v)_N = z_T \iff \beta_N = 0_n$ ergibt sich die Behauptung. $\lozenge$

Es ist auch an dieser Stelle unser Ziel, das Ausgangsproblem $(\widetilde{\mathrm{StP}})_h$ in ein unrestringiertes Problem umzuwandeln. Wir benötigen dazu Aussagen über den Zusammenhang zwischen dem diskreten Zustand z_h und der diskreten Steuerung u_h. Im kontinuierlichen Fall zeigt Satz 2.2.5 die eindeutige Lösbarkeit der Systemgleichung im Raum $W^2_{2,0}[0,T,\mathbb{R}^n]$ für jedes $y \in L_2[0,T,\mathbb{R}^n]$. Ein ähnliches Resultat für (5.1) und $y \in C[0,T,\mathbb{R}^n]$ abzuleiten, stellt unser erstes Ziel dar.

Dazu erinnern wir an Abschnitt 3.2 mit der Einführung der diskreten Norm $\|\cdot\|_h$ sowie an die Schreibweisen $(\cdot)_{\mathsf{h}}$ und $(\cdot)_{\bar{\mathsf{h}}}$ für die vorwärts bzw. rückwärts gerichtete Euler-Approximation der ersten Ableitung auf dem Intervall rechts bzw. links von der Stelle t_i, d.h. $(z_i)_{\mathsf{h}} = \frac{z_{i+1}-z_i}{h}$, $(z_i)_{\bar{\mathsf{h}}} = \frac{z_i - z_{i-1}}{h}$ sowie $(\cdot)_{\bar{\mathsf{h}}\mathsf{h}} = \frac{z_{i+1}-2z_i+z_{i-1}}{h^2}$, für die jeweilig sinnvollen Indizes. Wir gehen vom Differenzenschema (5.1) aus und bilden in jeder Gleichung das Skalarprodukt mit dem jeweiligen β_i, also

$$-\beta_i^{\mathsf{T}}(\beta_i)_{\bar{\mathsf{h}}\mathsf{h}} + \beta_i^{\mathsf{T}} A\beta_i = y_i^{\mathsf{T}}\beta_i, \qquad i = 1,\dots,N-1.$$

Durch Summation über alle $i = 1,\dots,N-1$ ergibt sich

$$\sum_{i=1}^{N-1}\left[-\beta_i^{\mathsf{T}}(\beta_i)_{\bar{\mathsf{h}}\mathsf{h}} + \beta_i^{\mathsf{T}} A\beta_i\right] = \sum_{i=1}^{N-1} y_i^{\mathsf{T}}\beta_i.$$

Für die linke Seite wenden wir das Prinzip der partiellen Summation an, d.h. für beliebige $\phi_i, \mu_i \in \mathbb{R}^n$, $i = 0,\dots,N$, mit $\phi_0 = \mu_0 = \phi_N = \mu_N = 0_n$ gilt

$$\sum_{i=1}^{N-1} \mu_i^{\mathsf{T}}(\phi_i)_{\mathsf{h}\bar{\mathsf{h}}} = \sum_{i=0}^{N-1}(\mu_i)_{\mathsf{h}}^{\mathsf{T}}(\phi_i)_{\mathsf{h}} = \sum_{i=1}^{N-1}\phi_i^{\mathsf{T}}(\mu_i)_{\mathsf{h}\bar{\mathsf{h}}}.$$

Daher schließen wir mit $\beta_0 = \beta_N = 0_n$ auf

$$-\sum_{i=1}^{N-1}\beta_i^{\mathsf{T}}(\beta_i)_{\bar{\mathsf{h}}\mathsf{h}} = -\sum_{i=0}^{N-1}\frac{1}{h}\beta_i^{\mathsf{T}}((\beta_{i+1})_{\mathsf{h}} - \beta_i)_{\mathsf{h}}) = \sum_{i=0}^{N-1}\frac{(\beta_{i+1}-\beta_i)^{\mathsf{T}}}{h}(\beta_{i+1})_{\bar{\mathsf{h}}} = \sum_{i=0}^{N-1}(\beta_i)_{\mathsf{h}}^{\mathsf{T}}(\beta_i)_{\mathsf{h}}.$$

Es folgt weiter

$$\sum_{i=0}^{N-1}|(\beta_i)_{\mathsf{h}}|^2 = \sum_{i=0}^{N-1}(\beta_i)_{\mathsf{h}}^{\mathsf{T}}(\beta_i)_{\mathsf{h}} \le -\sum_{i=1}^{N-1}\beta_i^{\mathsf{T}}(\beta_i)_{\mathsf{h}\bar{\mathsf{h}}} + \underbrace{\beta_i^{\mathsf{T}} A\beta_i}_{0\le} = \sum_{i=1}^{N-1} y_i^{\mathsf{T}}\beta_i.$$

Nach beidseitiger Multiplikation mit $h > 0$ und Anwendung der Cauchy-Schwarz-Ungleichung erhalten wir

$$\|\dot{z}_h\|_h^2 = h\sum_{i=0}^{N-1}|(\beta_i)_{\mathsf{h}}|^2 \le h\sum_{i=1}^{N-1} y_i^{\mathsf{T}}\beta_i = \langle y, z_h\rangle_h \le \|y\|_h\|z_h\|_h.$$

Daraus folgt mit der diskreten Poincaré'schen Ungleichung (3.4)

$$\|z_h\|_h^2 \leq \frac{T^2}{2}\|\dot{z}_h\|_h^2 \leq \frac{T^2}{2}\|y\|_h\|z_h\|_h,$$

und somit

$$\|z_h\|_h \leq \frac{T^2}{2}\|y\|_h.$$

Die eben hergeleitete Stabilitätsaussage bezüglich der Norm $\|\cdot\|_h$ halten wir in einem Satz fest.

Satz 5.2.1. *Seien $y \in C[0,T,\mathbb{R}^n]$ und $y_i = y(t_i) \in \mathbb{R}^n$. Dann besitzt die diskrete Systemgleichung*

$$\begin{aligned} z_h(t)\big|_{T_i} &= (t-t_i)\frac{\beta_{i+1}-\beta_i}{h} + \beta_i, && i = 0,\dots,N-1 \\ y(t_i) &= -\frac{\beta_{i+1}-2\beta_i+\beta_{i-1}}{h^2} + A\beta_i, && i = 1,\dots,N-1 \\ \beta_0 &= \beta_N = 0_n \end{aligned} \tag{5.2}$$

eine eindeutig bestimmte Lösung $z_h \in V_{h,0}[0,T,\mathbb{R}^n]$ und es gelten die Ungleichungen

$$\|z_h\|_h \leq \frac{T^2}{2}\|P_1 y\|_h. \tag{5.3a}$$

$$\|z_h\|_{1,2} \leq c_{2.6}\,\|P_1 y\|_h \tag{5.3b}$$

$$\|z_h\|_\infty \leq \sqrt{\frac{T}{2}}T\|P_1 y\|_h. \tag{5.3c}$$

Bemerkung: Da $(P_1 y)(t_i) = y(t_i)$ für alle $i = 0,\dots,N$ gilt und die Norm $\|\cdot\|_h$ nur die Werte an den Stützstellen benutzt, folgt sofort $\|P_1 y\|_h = \|y\|_h$ und somit auch $\|y\|_h \leq \sqrt{T}\|y\|_\infty$. Weiterhin ist für jedes $y \in C[0,T,\mathbb{R}^n]$ die Gleichung $z_h(y) = z_h(P_1 y)$ offensichtlich. ♢

Beweis: Für den Beweis der ersten Abschätzung verweisen wir auf die Herleitung vor dem Lemma. Weiterhin ist die Bilinearform

$$a_h(\cdot,\cdot) = h\sum_{i=0}^{N-1}(z_h(t_i))_{\mathsf{h}}^{\mathsf{T}}(v_h(t_i))_{\mathsf{h}} + z_h(t_i)^{\mathsf{T}}v_h(t_i)$$

aus (4.19) ein Skalarprodukt auf $V_{h,0}[0,T,\mathbb{R}^n]$ mit

$$a_h(z_h,z_h) \geq (c_{2.6})^{-2}\,\|z_h\|_{1,2}^2, \qquad \forall z_h \in V_{h,0}[0,T,\mathbb{R}^n],$$

d.h. sie ist gleichmäßig elliptisch. Der Beweis dieser Elliptizitätsaussage lehnt sich an die Gedanken aus Lemma 2.2.4 an

$$\begin{aligned} a_h(z_h, z_h) &= h\sum_{i=0}^{N}(z_h(t_i))_{\mathsf{h}}^{\mathsf{T}}(z_h(t_i))_{\mathsf{h}} + \underbrace{z_h(t_i)^{\mathsf{T}}Az_h(t_i)}_{\geq 0} \geq \|z_h\|_h^2 \\ &\geq \|z_h\|_2^2 \overset{(3.4)}{\geq} \frac{1}{2}\|\dot{z}_h\|_2^2 + \frac{1}{2}\frac{1}{T^2}\|z_h\|_2^2 \geq \frac{1}{2}\min\{1, T^{-2}\}\,\|z_h\|_{1,2}^2. \end{aligned}$$

Mit den gleichen Argumenten wie in Satz 2.2.5 folgt die Existenz und Eindeutigkeit der Lösung und darüber hinaus (5.3b).

Für die letzte Ungleichung erinnern wir an Lemma 3.2.3. Es ist wegen $z_h(t_0) = 0_n$

$$\|z_h\|_\infty \overset{(3.5)}{\leq} \sqrt{T}\|\dot{z}_h\|_h \leq \sqrt{\frac{T}{2}}T\|y\|_h.$$

□

Korollar 5.2.2. *Ersetzen wir in Satz 5.2.1 die Funktion z_h durch p_h und die Funktion y durch $z - z_d$, so erhalten die Aussagen auch für die diskrete adjungierte Gleichung*

$$\begin{aligned} p_h(t)\Big|_{T_i} &= (t - t_i)\frac{\zeta_{i+1} - \zeta_i}{h} + \zeta_i, && i = 0, \ldots, N-1 \\ (z - z_d)(t_i) &= -\frac{\zeta_{i+1} - 2\zeta_i + \zeta_{i-1}}{h^2} + A\zeta_i, && i = 1, \ldots, N-1 \\ \zeta_0 &= \zeta_N = 0_n \end{aligned} \tag{5.4}$$

Gültigkeit, d.h.

$$\begin{aligned} \|p_h\|_h &\leq \frac{T^2}{2}\|P_1(z - z_d)\|_h, \\ \|p_h\|_{1,2} &\leq c_{2.6}\,\|P_1(z - z_d)\|_h, \\ \|p_h\|_\infty &\leq \frac{T}{\sqrt{2}}\|P_1(z - z_d)\|_h. \end{aligned}$$

Beweis: Analog zu Satz 5.2.1. □

Bemerkung: Wie wir in diesem Abschnitt zeigen konnten, gibt es für die mit Hilfe des Differenzenschemas (5.1) diskretisierte Systemgleichung (1.4) eine eindeutig bestimmte Lösung für jede stetige rechte Seite. Wir konstruieren ferner unter Einbeziehung der Randwertvorgaben aus den Vektoren β_i, $i = 0, \ldots, N$, eine stückweise lineare Funktion z_h, die wir dann als Lösung der diskreten Systemgleichung (5.2) auffassen. Das bedeutet, wir sind in der Lage die Zuordnung $y \mapsto z_h$ als Operator $S_h : C[0, T, \mathbb{R}^n] \to V_{h,0}[0, T, \mathbb{R}^n]$ bzw. die Zuordnung $z - z_d \mapsto p_h$ als Operator

$S_h^* : C[0,T,\mathbb{R}^n] \to V_{h,0}[0,T,\mathbb{R}^n]$ zu definieren. Die Operatoren sind auf Grund der Form der diskreten Systemgleichung rein formal gleich, wir unterschieden sie dennoch aus Verständnisgründen. Die Linearität der Gleichungen überträgt sich auch auf die Abbildungen und wie Satz 5.2.1 und Korollar 5.2.2 zeigen, sind sie auch beschränkt in den Normen $\|\cdot\|_h$ und $\|\cdot\|_\infty$. ◊

Nach den Eindeutigkeitsaussagen aus Satz 5.2.1 und Korollar 5.2.2 sowie der damit verbundenen Definition der diskreten Operatoren sind wir in der Lage, eine wichtige Eigenschaft zu beweisen.

Lemma 5.2.3. *Die durch Satz 5.2.1 und Korollar 5.2.2 definierten Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$ sind auf dem Raum $V_h[0,T,\mathbb{R}^n]$ bezüglich des Skalarprodukts $\langle\cdot,\cdot\rangle_h$ adjungiert, d.h.*

$$\langle \mathcal{S}_h v, w\rangle_h = \langle v, \mathcal{S}_h^* w\rangle_h \qquad \forall v, w \in V_h[0,T,\mathbb{R}^n].$$

Beweis: Schon in der voraus gehenden Bemerkung hielten wir fest, dass die Operatoren identisch sind und wir sie nur aus Verständnisgründen unterscheiden. Daher reicht es aus, die Selbstadjungiertheit von $\mathcal{S}_h$ zu zeigen. Ausgehend vom Gleichungssystem

$$((\mathcal{S}_h v)_i)_{\mathsf{h}\bar{\mathsf{h}}} + A(\mathcal{S}_h v)_i = v_i, \qquad i = 1,\dots,N-1$$

multiplizieren wir die jeweilige Gleichung mit $(\mathcal{S}_h w)_i$, also

$$(\mathcal{S}_h w)_i^{\mathsf{T}}((\mathcal{S}_h v)_i)_{\mathsf{h}\bar{\mathsf{h}}} + (\mathcal{S}_h w)_i^{\mathsf{T}} A(\mathcal{S}_h v)_i = (\mathcal{S}_h w)_i^{\mathsf{T}} v_i, \qquad i = 1,\dots,N-1.$$

Anschließende Multiplikation mit $h > 0$ und Summation über alle Indizes ergibt

$$\langle \mathcal{S}_h w, (\mathcal{S}_h v)_{\mathsf{h}\bar{\mathsf{h}}}\rangle_h + \langle \mathcal{S}_h w, A(\mathcal{S}_h v)\rangle_h = \langle \mathcal{S}_h w, v\rangle_h.$$

Für den ersten Term wenden wir zweimal das Verfahren der partiellen Summation an und erhalten

$$\langle (\mathcal{S}_h w)_{\mathsf{h}\bar{\mathsf{h}}}, \mathcal{S}_h v\rangle_h + \langle \mathcal{S}_h w, A(\mathcal{S}_h v)\rangle_h = \langle \mathcal{S}_h w, v\rangle_h.$$

Analog multiplizieren wir in

$$((\mathcal{S}_h w)_i)_{\mathsf{h}\bar{\mathsf{h}}} + A(\mathcal{S}_h w)_i = w_i, \qquad i = 1,\dots,N-1$$

jede Gleichung jeweils mit $(\mathcal{S}_h v)_i$ und $h > 0$, summieren über alle Indizes und erhalten

$$\langle (\mathcal{S}_h w)_{\mathsf{h}\bar{\mathsf{h}}}, \mathcal{S}_h v\rangle_h + \langle \mathcal{S}_h w, A(\mathcal{S}_h v)\rangle_h = \langle w, \mathcal{S}_h v\rangle_h.$$

□

Mit dem Wissen um die Beziehung $z_h = \mathcal{S}_h(Bu_h + e)$ schreiben wir das Problem $\widetilde{(\mathrm{StP})}_h$ um, denn u_h bleibt als einzige freie Variable. So ist dann

$$(\mathrm{StP})_h \qquad \min_{u_h} J_h(u_h) = \min_{u_h} \frac{1}{2}\|\mathcal{S}_h\mathcal{T}u_h - z_d\|_h^2 + \frac{\nu}{2}\|u_h\|_h^2, \qquad u_h \in U_h^{ad}.$$

Die Existenz und Eindeutigkeit einer Lösung von $(\mathrm{StP})_h$ folgt aus dem allgemeinen Resultat in Lemma 3.4.1. Dort setzen wir $\langle\cdot,\cdot\rangle_{\mathfrak{C}} = \langle\cdot,\cdot\rangle_h$ bzw. $\mathfrak{C} = I_{(N+1)n}$ und erhalten die notwendige Optimalitätsbedingung

$$\langle J_h'(u_h), \zeta_h - u_h\rangle_h = \langle B^{\mathsf{T}}\mathcal{S}_h^*(\mathcal{S}_h(Bu_h + e) - z_d), \zeta_h - u_h\rangle_h \geq 0 \qquad \forall \zeta_h \in U_h^{ad}. \quad (5.5)$$

Wir haben dabei die Adjungiertheit der diskreten Operatoren bezüglich des diskreten Skalarprodukts ausgenutzt. Im Vergleich zur Diskretisierung mit Finiten Elementen betrachten wir nun ein anderes Zielfunktional und erhalten daraus eine veränderte Optimalitätsbedingung, die den Vorgaben durch die Definition der Operatoren angepasst ist.

5.3 Konvergenz

Um Aussagen über den Fehler bei der Diskretisierung der Systemgleichung (1.4) durch das Differenzenschema (5.1) zu treffen, betrachten wir die Differenz der beiden Lösungen. Die kontinuierliche Lösung $z \in W^2_{2,0}[0,T,\mathbb{R}^n]$ reduzieren wir dabei zunächst auf den Vektor der Werte an den Stützstellen, betrachten also ihre Diskretisierung $P_1 z$ im Raum $V_{h,0}[0,T,\mathbb{R}^n]$. Verwenden wir wieder die Bezeichnungen $\beta_i = z_h(t_i)$ für $i = 0,\dots,N$, so ist die Differenz $P_1 z - \beta =: \gamma = (\gamma_0^{\mathsf{T}},\dots,\gamma_N^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{(N+1)n}$ ist offensichtlich die Lösung des Differenzenschemas

$$\begin{aligned} -\frac{\gamma_{i+1} - 2\gamma_i + \gamma_{i-1}}{h^2} + A\gamma_i &= \psi_i, \qquad i = 1,\dots,N-1 \\ \gamma_0 = \gamma_N &= 0_n, \end{aligned} \quad (5.6)$$

mit

$$\psi_i = -(z_i)_{\bar{\mathsf{h}}\mathsf{h}} + Az_i - y_i, \qquad i = 1,\dots,N-1.$$

Betrachten wir die Systemgleichung (1.4), integriert zwischen den Grenzen $S_{i-1} = t_i - h/2$ und $S_i = t_i + h/2$ für $i = 1,\dots,N-1$, d.h.

$$\int_{-\frac{h}{2}}^{\frac{h}{2}} -\ddot{z}(t_i+t) + Az(t_i+t)\,dt = \dot{z}_{i-\frac{1}{2}} - \dot{z}_{i+\frac{1}{2}} + \int_{-\frac{h}{2}}^{\frac{h}{2}} Az(t_i+t)\,dt = \int_{-\frac{h}{2}}^{\frac{h}{2}} y(t_i+t)\,dt,$$

so folgt nach Division durch h

$$\underbrace{\frac{\dot{z}_{i+\frac{1}{2}} - \dot{z}_{i-\frac{1}{2}}}{h}}_{(\dot{z}_{i-\frac{1}{2}})_{\mathsf{h}}} - \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} Az(t_i+t)\,dt + \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} y(t_i+t)\,dt = 0, \qquad i = 1,\dots,N-1.$$

Diese „Null“ setzen wir in die Gleichung für ψ_i ein und erhalten für $i = 1,\dots,N-1$

$$\psi_i = \left(\dot{z}_{i-\frac{1}{2}} - (z_i)_{\bar{\mathsf{h}}}\right)_{\mathsf{h}} + Az_i - \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} Az(t_i+t)\,dt - y_i + \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} y(t_i+t)\,dt. \quad (5.7)$$

Die Summe teilen wir durch $\psi_i = (\psi_i^0)_{\mathsf{h}} + \psi_i^1$ auf und vergeben neue Bezeichnungen

$$\begin{aligned}
\psi_i^0 &:= \dot{z}_{i-\frac{1}{2}} - (z_i)_{\bar{\mathsf{h}}}\,, && i = 1, \ldots, N,\\
\psi_i^1 &:= Az_i - \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} Az(t_i + t)\,dt - y_i + \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} y(t_i + t)\,dt, && i = 1, \ldots, N-1.
\end{aligned}$$

Bei der Definition von ψ^0 ist Folgendes zu beachten. Der Vektor besteht im Gegensatz zu ψ aus N Komponenten mit den Indizes $i = 1, \ldots, N$.

Wir spalten also die rechte Seite des Differenzenschemas (5.6) in eine Summe von Vektoren auf, wobei der erste Summand die Ableitung einer stückweise linearen Funktion darstellt und der zweite Summand eine solche Funktion selbst. Diese Möglichkeit nimmt später für Aussagen über die Konvergenzgeschwindigkeit eine wichtige Rolle ein.

In Lemma 5.3.2 werden wir zeigen, dass es für die Betrachtung des Abstandes zwischen exakter und diskreter Lösung ausreicht, die Normen der Vektoren ψ^0 und ψ^1 abzuschätzen. Wir beginnen mit $\psi_i^0 = \dot{z}_{i-\frac{1}{2}} - (z_i)_{\bar{\mathsf{h}}}$. Diese Größe ist nur von der Lösung der kontinuierlichen Systemgleichung z abhängig und wir nutzen unser Wissen um deren Differenzierbarkeitseigenschaften aus. Für $i = 1, \ldots, N$ ist dann

$$\begin{aligned}
\dot{z}_{i-\frac{1}{2}} - \frac{z_i - z_{i-1}}{h} &= \dot{z}_{i-\frac{1}{2}} - \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} \dot{z}(S_i + t)\,dt\\
&= -\frac{1}{h}\int_0^{\frac{h}{2}} \dot{z}(S_{i-1} + t) - 2\dot{z}(S_{i-1}) + \dot{z}(S_{i-1} - t)\,dt\\
&= -\frac{1}{h}\int_0^{\frac{h}{2}}\int_0^t \ddot{z}(S_{i-1} + s) - \ddot{z}(S_{i-1} - t + s)\,ds\,dt.
\end{aligned}$$

Durch Übergang zum Betrag auf beiden Seiten gelangen wir zu

$$|\psi_i^0| = \frac{1}{h}\int_0^{\frac{h}{2}}\int_0^{\frac{h}{2}} |\ddot{z}(S_{i-1} + s) - \ddot{z}(S_{i-1} - t + s)|\,ds\,dt \le \frac{h}{4}\,\mathsf{V}_{T_i}\ddot{z}.$$

Die Untersuchung von $|\psi_i^1|$ zerlegen wir in zwei Teile und betrachten zuerst die

Abschätzung

$$\begin{aligned}\Big|Az_i - \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} Az(t_i+t)\,dt\Big| &= \frac{1}{h}\Big|\int_{-\frac{h}{2}}^{\frac{h}{2}} A(z_i - z(t_i+t))\,dt\Big| \\ &\le \frac{1}{h}\|A\|\int_0^{\frac{h}{2}} |z(t_i+t) - 2z_i + z(t_i-t)|\,dt \\ &= \frac{1}{h}\|A\|\int_0^{\frac{h}{2}} \Big|\int_0^t \dot z(t_i+s) - \dot z(t_i-t+s)\,ds\Big|\,dt \\ &\le \frac{1}{h}\|A\|\int_0^{\frac{h}{2}}\int_0^{\frac{h}{2}} \mathsf{V}_{S_{i-1}}^{S_i}\dot z\,ds\,dt \\ &= \frac{h}{4}\|A\|\,\mathsf{V}_{S_{i-1}}^{S_i}\dot z \overset{(3.7)}{\le} \|\ddot z\|_{L_1(S_{i-1},S_i)}.\end{aligned}$$

Nun kommen wir zum zweiten Summanden und schließen analog

$$\begin{aligned}\Big|y_i - \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} y(t_i+t)\,dt\Big| &= \frac{1}{h}\Big|\int_{-\frac{h}{2}}^{\frac{h}{2}} y_i - y(t_i+t)\,dt\Big| \\ &\le \frac{1}{h}\Big|\int_0^{\frac{h}{2}}\int_0^t \dot y(t_i+s) - \dot y(t_i-t+s)\,ds\,dt\Big| \\ &\le \frac{h}{4}\,\mathsf{V}_{S_{i-1}}^{S_i}\dot y.\end{aligned}$$

Abschließend fassen wir die hergeleiteten Abschätzungen in einem Lemma zusammen.

Lemma 5.3.1. *Sei $\psi = (\psi_1^{\mathsf{T}}, \ldots, \psi_{N-1}^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{(N-1)n}$ der Vektor aus (5.7). Dann gibt es eine Darstellung $\psi_i = (\psi_i^0)_{\mathsf{h}} + \psi_i^1$ und es gelten für die beiden Summanden die Abschätzungen*

$$\begin{aligned}|\psi_i^0| &\le \frac{h}{4}\,\mathsf{V}_{T_i}\ddot z, && i = 1,\ldots,N \\ |\psi_i^1| &\le \frac{h}{4}\Big(\|A\|\,\|\ddot z\|_{L_1(S_{i-1},S_i)} + \mathsf{V}_{S_{i-1}}^{S_i}\dot y\Big), && i = 1,\ldots,N-1.\end{aligned}$$

Beweis: Ebenda. □

Diese Vorarbeit besaß ihren Grund, denn Samarskii [19] (1984) sowie Sendov/Popov [20] (1988) beinhalten unter anderem ein Konvergenzresultat für die numerische Lösung von skalaren Randwertproblemen. Wir erweitern die Aussage auf Funktionen mit mehrdimensionalem Wertebereich.

Lemma 5.3.2. *Sei $\gamma = (\gamma_0^{\mathsf{T}}, \ldots, \gamma_N^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{(N+1)n}$ die Lösung des Differenzenschemas (5.6) und ψ sei zerlegbar in eine Summe der Form $\psi = \psi^0{}_{\mathsf{h}} + \psi^1$. Dann gilt mit $\mu_1 = 0_n$ und $\mu_i = \sum_{k=1}^{i-1} h\psi_k^1$ für $i = 2,\ldots,N$ die Abschätzung*

$$\|\gamma\|_\infty \le c_{5.8}\left(\sum_{i=1}^N h|\psi_i^0| + \sum_{i=2}^N h|\mu_i|\right) \tag{5.8}$$

mit $c_{5.8} = 1 + \frac{T^2}{\sqrt{2}}\|A\|$.

Beweis: Zunächst stellen wir $\mu_{\mathsf{h}} = \psi^1$ fest, so dass es ausreicht, das System

$$\begin{aligned} -(\gamma_i)_{\bar{\mathsf{h}}\mathsf{h}} + A\gamma_i &= (\psi_i^0)_{\mathsf{h}}, \qquad i = 1, \ldots, N-1 \\ \gamma_0 = \gamma_N &= 0_n \end{aligned}$$

zu behandeln.

Betrachten wir die Gleichungen

$$-w_{\bar{\mathsf{h}}\mathsf{h}} = \psi^0{}_{\mathsf{h}}, \qquad w_0 = w_N = 0_n.$$

Daraus folgt $-(w_{\bar{\mathsf{h}}} + \psi^0)_{\mathsf{h}} = 0_n$ und somit $w_{\bar{\mathsf{h}}} + \psi^0 = \text{const} =: q$. Jetzt berechnen wir die einzelnen Vektoren in w durch eine rekursive Formel, denn es ist

$$\frac{w_k - w_{k-1}}{h} + \psi_k^0 = q \Longleftrightarrow w_k = w_{k-1} - h\psi_k^0 + hq, \qquad k = 1, \ldots, N.$$

Summieren wir diese Gleichungen über den Index k auf, so folgt

$$w_i = \sum_{k=1}^{i} w_k = w_0 - \sum_{k=1}^{i} h\psi_k^0 + ihq.$$

Für $i = N$ und unter Beachtung der Randwertbedingungen $w_0 = w_N = 0_n$ erhalten wir

$$0_n = -h\sum_{k=1}^{N} \psi_k^0 + Nhq \Longleftrightarrow q = \frac{1}{N}\sum_{k=1}^{N} \psi_k^0.$$

Setzen wir oben ein, so folgt

$$w_i = -h\sum_{k=1}^{i} \psi_k^0 + \frac{ih}{N}\sum_{k=1}^{N} \psi_k^0 = -\left(1 - \frac{i}{N}\right) h\sum_{k=1}^{i} \psi_k^0 + \frac{ih}{N}\sum_{k=i+1}^{N} \psi_k^0.$$

Nach dem Bilden des Betrags auf beiden Seiten erhalten wir

$$|w_i| \leq \left(1 - \frac{i}{N}\right) h\sum_{k=1}^{i} |\psi_k^0| + \frac{ih}{N}\sum_{k=i+1}^{N} |\psi_k^0| \leq h\sum_{k=1}^{N} |\psi_k^0|, \qquad i = 1, \ldots, N-1.$$

Die rechte Seite ist unabhängig vom Index i, so dass wir links zum Maximum übergehen und erhalten

$$\|w\|_\infty \leq h\sum_{k=1}^{N} |\psi_k^0|.$$

Aus der Definition von w folgen für den Vektor $\gamma - w$ die Gleichungen

$$\begin{aligned} -((\gamma - w)_i)_{\bar{\mathsf{h}}\mathsf{h}} + A(\gamma - w)_i &= -Aw_i, \qquad i = 1, \ldots, N-1 \\ (\gamma - w)_0 = (\gamma - w)_N &= 0_n \end{aligned}$$

und mit Satz 5.2.1 schließen wir auf

$$\|\gamma - w\|_\infty \overset{(5.3c)}{\leq} \frac{T^2}{\sqrt{2}}\|A\|\,\|w\|_\infty.$$

Davon ausgehend folgt als letzter Schritt

$$\|\gamma\|_\infty \leq \|\gamma - w\|_\infty + \|w\|_\infty \leq (1 + \frac{T^2}{\sqrt{2}}\|A\|)\,\|w\|_\infty \leq c_{5.8}\,h\sum_{k=1}^{N}|\psi_k^0|.$$

□

Bemerkung: Die Aussage des Lemmas bewiesen wir bereits für den kontinuierlichen Fall in Lemma 2.2.7 und bei der Diskretisierung mittels Finiter Elemente in Lemma 4.2.4. ◊

Lemma 5.3.3. *Seien $z \in W^2_{2,0}[0,T,\mathbb{R}^n]$ die Lösung der Systemgleichung (1.4) zur Funktion $y \in W^1_\infty[0,T,\mathbb{R}^n]$ mit $\dot y \in BV[0,T,\mathbb{R}^n]$ und $\beta = (\beta_0^\mathsf{T},\ldots,\beta_N^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{(N+1)n}$ die Lösung des Differenzenschemas (5.1). Dann gilt für den Vektor $(\gamma_0^\mathsf{T},\ldots,\gamma_N^\mathsf{T})^\mathsf{T} = \gamma = P_1 z - \beta$*

(1) *γ ist Lösung des Systems (5.6) mit $\psi_i = (z_i)_{\bar{\mathsf{h}}\mathsf{h}} + Az_i - y_i$*

(2) *Weiterhin gilt*

$$\|\gamma\|_\infty \leq c_{5.9}\,(\|y\|_{1,\infty} + \mathsf{V}_0^T\dot y)\,h^2. \tag{5.9}$$

mit der Konstante $c_{5.9} = \frac{T}{2}c_{5.8}c^o_{2.4}c_{2.6}c_{2.7}$.

Beweis: Teil (1) ist offensichtlich. Für die Abschätzung in Teil (2) erinnern wir an die Definitionen von ψ_i^0 und ψ_i^1 mit $\psi_i = (\psi_i^0)_\mathsf{h} + \psi_i^1$ durch

$$\begin{aligned}
\psi_i^0 &= \dot z_{i-\frac12} - (z_i)_{\bar{\mathsf{h}}}, && i = 1,\ldots,N\\
\psi_i^1 &= Az_i - \int_{-\frac{h}{2}}^{\frac{h}{2}} Az(t_i+t)\,dt - y_i + \int_{-\frac{h}{2}}^{\frac{h}{2}} y(t_i+t)\,dt, && i = 1,\ldots,N-1
\end{aligned}$$

und an die Ungleichungen aus Lemma 5.3.1

$$\begin{aligned}
|\psi_i^0| &\leq \frac{h}{4}\,\mathsf{V}_{T_i}\ddot z\\
|\psi_i^1| &\leq \frac{h}{4}\big(\|A\|\,\|\ddot z\|_{L_1(S_{i-1},S_i)} + \mathsf{V}_{S_{i-1}}^{S_i}\dot y\big).
\end{aligned}$$

Es ist

$$\begin{aligned}
&\sum_{i=1}^{N} h|\psi_i^0| + \sum_{i=2}^{N} h \sum_{k=1}^{i-1} h|\psi_k^1| \\
&\leq \sum_{i=1}^{N} \frac{h^2}{4} \mathsf{V}_{T_i} \ddot{z} + \sum_{i=2}^{N} h \left[\sum_{k=1}^{i-1} \frac{h^2}{4} \Big(\|A\| \, \|\ddot{z}\|_{L_1(S_{k-1},S_k)} + \mathsf{V}_{S_{k-1}}^{S_k} \dot{y} \Big) \right] \\
&\leq \frac{h^2}{4} \mathsf{V}_0^T \ddot{z} + T \frac{h^2}{4} \big(\|A\| \, \|\ddot{z}\|_1 + \mathsf{V}_0^T \dot{y} \big) \\
&\leq \frac{1}{4} \big(\mathsf{V}_0^T \ddot{z} + T^{\frac{3}{2}} \|A\| \, \|\ddot{z}\|_2 + T \, \mathsf{V}_0^T \dot{y} \big) \, h^2.
\end{aligned}$$

Mit Lemma 5.3.2 schließen wir auf

$$\|\gamma\|_\infty \leq \frac{1}{4} c_{5.8} \left(\mathsf{V}_0^T \ddot{z} + T^{\frac{3}{2}} \|A\| \, \|\ddot{z}\|_2 + T \, \mathsf{V}_0^T \dot{y} \right) h^2.$$

Aus der Systemgleichung (1.4) erhalten wir

$$\mathsf{V}_0^T \ddot{z} \leq \|A\| \, \mathsf{V}_0^T z + \mathsf{V}_0^T y \overset{(3.7)}{\leq} \|A\| \, \|\dot{z}\|_1 + \|\dot{y}\|_1 \leq T^{\frac{3}{2}} \|A\| \, \|y\|_2 + \sqrt{T} \|\dot{y}\|_2$$

und somit

$$\begin{aligned}
\mathsf{V}_0^T \ddot{z} + T^{\frac{3}{2}} \|A\| \, \|\ddot{z}\|_2 + T \, \mathsf{V}_0^T \dot{y} &\leq \sqrt{T} \Big[T \|A\| \, \|y\|_2 + \|\dot{y}\|_2 + T \|A\| \, c_{2.7} \, \|y\|_2 \Big] + T \, \mathsf{V}_0^T \dot{y} \\
&\leq 2\sqrt{T} \max\{1, T\} \, c_{2.4}^o c_{2.7} \, (\|y\|_2 + \|\dot{y}\|_2) + T \, \mathsf{V}_0^T \dot{y} \\
&\leq 2\sqrt{T} c_{2.4}^o c_{2.6} c_{2.7} \, \|y\|_{1,2} + T \, \mathsf{V}_0^T \dot{y}.
\end{aligned}$$

Abschließend ergibt sich

$$\|\gamma\|_\infty \leq \frac{T}{2} c_{5.8} c_{2.4}^o c_{2.6} c_{2.7} \, (\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y}) \, h^2.$$

□

Bemerkung: Diese Fehlerabschätzung betrifft die Konvergenz im Raum der stückweise linearen Funktionen, daher sprechen wir dabei von *diskreter Konvergenz* mit Ordnung 2. Nach Erweiterung auf kontinuierliche Funktionen nennen wir das Ergebnis dann *Konvergenz* der Ordnung 2. Diesen Schritt vollziehen wir im folgenden Satz. ◊

Satz 5.3.4. *Seien $z \in W_{2,0}^2[0,T,\mathbb{R}^n]$ die Lösung der Systemgleichung (1.4) für $y \in W_\infty^1[0,T,\mathbb{R}^n]$ mit $\dot{y} \in BV[0,T,\mathbb{R}^n]$ und $z_h \in V_{h,0}[0,T,\mathbb{R}^n]$ die Lösung der diskreten Systemgleichung (5.2) mit $P_1 y$ als rechter Seite. Dann gilt*

$$\|z - z_h\|_\infty \leq \max\{c_{2.7}, c_{5.9}\} \, (\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y}) \, h^2. \tag{5.10}$$

Seien $p \in W^2_{2,0}[0,T,\mathbb{R}^n]$ die Lösung der adjungierten Gleichung (2.10) *und p_h die Lösung der diskreten adjungierten Gleichung* (5.4) *mit $P_1(z-z_d)$ als rechter Seite. Dann gelten*

$$\|P_1 p - p_h\|_\infty \le c_{2.6}^2 c_{5.9} \left(\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y} + \|z_d\|_{2,\infty}\right) h^2 \tag{5.11}$$

$$\|p - p_h\|_\infty \le 2c_{2.6}^2 \max\{c_{2.7}, c_{5.9}\} \left(\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y} + \|z_d\|_{2,\infty}\right) h^2. \tag{5.12}$$

Beweis: Teil I: Wir knüpfen direkt an Lemma 5.3.3 an und verwenden die Dreiecks-Ungleichung und Lemma 3.3.4, nämlich

$$\begin{aligned}
\|z - z_h\|_\infty &\le \|z - P_1 z\|_\infty + \|P_1 z - z_h\|_\infty \\
&\le \left(\|\ddot{z}\|_\infty + c_{5.9} \left(\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y}\right) h^2 \right. \\
&\le \left(\|A\| \, \|z\|_\infty + \|y\|_\infty + c_{5.9} \left(\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y}\right) h^2 \right. \\
&\le \left(1 + T^2 \|A\|\right) \|y\|_\infty + c_{5.9} \left(\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y}\right) h^2 \\
&\le \max\{c_{2.7}, c_{5.9}\} \left(\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y}\right) h^2.
\end{aligned}$$

Teil II: Der Beweis für die adjungierten Zustände verläuft ähnlich, aber mit einer zusätzlichen Bemerkung. Wir verwenden zunächst die Dreiecks-Ungleichung mit der Bezeichnung $p_h^* = \mathcal{S}_h^*(\mathcal{S}y - z_d)$ für den diskreten adjungierten Zustand zur Funktion z

$$\|P_1 p - p_h\|_\infty \le \|P_1 p - p_h^*\|_\infty + \|p_h^* - p_h\|_\infty$$

und übertragen die Erkenntnisse des ersten Teils auf die adjungierte Gleichung und ihre Diskretisierung Dabei nutzen wir die stärkeren Differenzierbarkeitseigenschaften von $z - z_d$ aus und erhalten so

$$\begin{aligned}
\|P_1 p - p_h^*\|_\infty \quad &\le \quad \frac{1}{4} c_{5.8} \left(\mathsf{V}_0^T \ddot{p} + T^{\frac{3}{2}} \|A\| \, \|\ddot{p}\|_2 + T \, \mathsf{V}_0^T (\dot{z} + \dot{z}_d)\right) \\
&\overset{(3.7)}{\le} \quad \frac{1}{4} c_{5.8} \left(\|A\| \, \|\dot{p}\|_1 + \|\dot{z} - \dot{z}_d\|_1 + T^{\frac{3}{2}} \|A\| \, \|\ddot{p}\|_2 + T \|\ddot{z} - \ddot{z}_d\|_1\right) \\
&\le \quad \frac{\sqrt{T}}{4} c_{5.8} c_{2.4}^o \left(\|\dot{p}\|_2 + \|\dot{z} - \dot{z}_d\|_2 + T\|\ddot{p}\|_2 + T\|\ddot{z} - \ddot{z}_d\|_2\right) \\
&\le \quad \frac{\sqrt{2T}}{4} \max\{1, T\} \, c_{5.8} c_{2.4}^o \left(\|p\|_{2,2} + \|\dot{z} - \dot{z}_d\|_2 + \|\ddot{z} - \ddot{z}_d\|_2\right) \\
&\overset{(2.14)}{\le} \quad \frac{\sqrt{T}}{4} \, c_{5.8} c_{2.4}^o c_{2.6} c_{2.7} \left(\sqrt{2}\|z - z_d\|_2 + \|\dot{z} - \dot{z}_d\|_2 + \|\ddot{z} - \ddot{z}_d\|_2\right) \\
&\overset{(A.1)}{\le} \quad \frac{\sqrt{T}}{2} c_{5.8} c_{2.4}^o c_{2.6} c_{2.7} \, \|z - z_d\|_{2,2} \\
&\le \quad \sqrt{\frac{T}{2}} c_{5.8} c_{2.4}^o c_{2.6} c_{2.7}^2 \left(\|y\|_2 + \|z_d\|_{2,2}\right) \\
&\le \quad \sqrt{2} c_{5.9} \left(\|y\|_\infty + \|z_d\|_{2,\infty}\right).
\end{aligned}$$

Den zweiten Summanden bearbeiten wir mit Hilfe der Stabilitätsaussage aus Satz 5.2.1 und Lemma 5.3.3

$$\|p_h^* - p_h\|_\infty \overset{(5.3c)}{\leq} \frac{T^2}{2}\|P_1 z - z_h\|_\infty \leq \frac{T^2}{2} c_{5.9} \left(\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y}\right) h^2.$$

Wir erhalten somit

$$\|P_1 p - p_h\|_\infty \leq c_{2.6}^2 c_{5.9} \left(\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y} + \|z_d\|_{2,\infty}\right) h^2.$$

Abschließend folgt

$$\begin{aligned}
\|p - p_h\|_\infty &\leq \|p - P_1 p\|_\infty + \|P_1 p - p_h\|_\infty \\
&\leq \|\ddot{p}\|_\infty h^2 + c_{2.6}^2 c_{5.9} \left(\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y} + \|z_d\|_{2,\infty}\right) h^2 \\
&\leq c_{2.6}^2 c_{2.7} \left(\|y\|_\infty + \|z_d\|_\infty\right) + c_{2.6}^2 c_{5.9} \left(\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y} + \|z_d\|_{2,\infty}\right) h^2 \\
&\leq 2 c_{2.6}^2 \max\{c_{2.7}, c_{5.9}\} \left(\|y\|_{1,\infty} + \mathsf{V}_0^T \dot{y} + \|z_d\|_{2,\infty}\right) h^2.
\end{aligned}$$

□

5.4 Zusammenfassung

In den beiden vorangegangenen Abschnitten untersuchten wir die Eigenschaften der Finiten Differenzen. Wir geben an dieser Stelle eine Zusammenfassung der eruierten Resultate.

In Satz 5.2.1 und Korollar 5.2.2 zeigten wir die eindeutige Lösbarkeit der diskretisierten Systemgleichung (5.2) und der diskretisierten adjungierten Gleichung (5.4) für jede rechte Seite aus dem Raum $C[0,T,\mathbb{R}^n]$. Die Forderung der Stetigkeit begründet sich in der Nutzung von Werten an konkreten Stellen innerhalb des Intervalls. Das gibt uns die Möglichkeit beide Gleichungen mit Hilfe der linearen und beschränkten Operatoren $\mathcal{S}_h, \mathcal{S}_h^* : C[0,T,\mathbb{R}^n] \to V_{h,0}[0,T,\mathbb{R}^n]$ auch in der Form

$$z_h(u) = \mathcal{S}_h \mathcal{T} u, \qquad p_h(u) = \mathcal{S}_h^*(\mathcal{S}_h \mathcal{T} u - z_d)$$

zu schreiben. Wir unterscheiden für den diskreten adjungierten Zustand zwei Funktionen. Neben $p_h(u)$, bei dessen Berechnung wir beide Gleichungen diskretisieren, steht

$$p_h^*(u) = \mathcal{S}_h^*(\mathcal{S}\mathcal{T} u - z_d)$$

für den diskreten adjungierten Zustand, bei dessen Berechnung die Diskretisierung nur in der adjungierten Gleichungen erfolgt. Die Schreibweise mit dem Argument u verdeutlicht die Abhängigkeit von der Steuerung.

Lemma 5.2.3 zeigt die Selbstadjungiertheit von $\mathcal{S}_h$ bezüglich dem h-Skalarprodukt. Die Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$ sind daher per Definition gleich, wir unterscheiden sie dennoch aus Verständnisgründen. In Satz 5.2.1 wiesen wir ihre Beschränktheit bezüglich

diverser Normen nach. Die oberen Schranken $\frac{T^2}{2}$ bzw. $\frac{T^2}{\sqrt{2}}$ gelten für $\mathcal{S}_h$ in der h-Norm bzw. in der L_∞-Norm.

Neben der Beschränktheit der diskreten Operatoren wissen wir bereits um die Konvergenz der diskreten Lösung gegen die exakte Lösung für Systemgleichung und adjungierte Gleichung. Wir erinnern an Satz 5.3.4 mit den Abschätzungen

$$\begin{aligned}\|z(u) - z_h(u)\|_\infty &\leq \max\{c_{2.7}, c_{5.9}\}\,(\|\mathcal{T}u\|_{1,\infty} + \mathsf{V}_0^T(\dot{\mathcal{T}}u))\,h^2\\ \|p(u) - p_h(u)\|_\infty &\leq 2c_{2.6}^2 \max\{c_{2.7}, c_{5.9}\}\,(\|\mathcal{T}u\|_{1,\infty} + \mathsf{V}_0^T(\dot{\mathcal{T}}u) + \|z_d\|_{2,\infty})\,h^2.\end{aligned}$$

Damit sind wir im Besitz aller notwendigen Informationen für die Betrachtung von diskreten Steuerungsproblemen und ihrer Eigenschaften hinsichtlich der Approximation der Lösung von (StP).

5.5 Hauptergebnisse

In diesem Abschnitt sollen nun die Vorbetrachtungen zum Hauptresultat bezüglich der Konvergenzgeschwindigkeit der Funktion $\tilde{u} = \Pi_{[\mathrm{a,b}]}\big(-\frac{1}{\nu}B^\mathsf{T}p_h(u_h)\big)$ gegen die exakte optimale Steuerung zusammenfließen. Dabei verzichten wir im Gegensatz zu Kapitel 4 auf eine Unterteilung nach der Approximationsart der Steuerung. Wie schon aus der Aufstellung des diskreten Problems $(\mathrm{StP})_h$ hervorgeht, diskretisieren wir die Steuerung grundsätzlich stückweise linear. Eine Vorgehensweise die nahe liegt, denn im Differenzenschema (5.1) benötigen wir exakt die Werte an den Stützstellen. Alternative Möglichkeiten der Approximation der Steuerung untersuchen wir im Anschluss.

Zunächst nutzen wir die Aussagen über die Stabilität und die Konvergenz, um den Abstand zwischen der Lösung des diskreten Steuerungsproblems $(\mathrm{StP})_h$ und optimalen Steuerung herzuleiten. Als Abstandsmaß verwenden wir die Norm $\|\cdot\|_h$ und zeigen somit diskrete Konvergenz. Wir wenden uns also der Größe $\|u_h - P_1\bar{u}\|_h$ zu.

Satz 5.5.1. *Seien $\bar{u}$ die Lösung von* (StP) *mit $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$ und $u_h \in V_h[0,T,\mathbb{R}^m]$ die Lösung des diskreten Steuerungsproblems* $(\mathrm{StP})_h$*. Dann gilt mit der Norm $\|v_h\|_h = \sqrt{h\sum_{i=0}^N |v_h(t_i)|^2}$*

$$\|u_h - P_1\bar{u}\|_h \leq \sqrt{T}\frac{\|B\|}{\nu}c_{2.6}^2 c_{5.9}\,C_{5.13}^{\bar{u}}\,h^2, \tag{5.13}$$

wobei die Abkürzung $C_{5.13}^{\bar{u}} = \|\mathcal{T}\bar{u}\|_{1,\infty} + \mathsf{V}_0^T(\dot{\mathcal{T}}\bar{u}) + \|z_d\|_{2,\infty}$ zur Verringerung des Notationsaufwands an dieser Stelle definiert sei.

Beweis: Ausgangspunkt für den Beweis sind die notwendigen Optimalitätsbedingungen für $\bar{u}$ bzw. u_h, die wir in Lemma 2.4.1 bzw. in diesem Kapitel herleiten. Es ist

$$\langle B^\mathsf{T}\bar{p} + \nu\bar{u}, u - \bar{u}\rangle \geq 0 \qquad \forall u \in U^{ad} \tag{2.17}$$

$$\langle B^\mathsf{T}p_h(u_h) + \nu u_h, \zeta_h - u_h\rangle_h \geq 0 \qquad \forall \zeta_h \in U_h^{ad}. \tag{5.5}$$

Da die Ungleichung (2.17) für alle $u \in U^{ad}$ gilt, schließen wir mit Lemma 2.4.3 auf die punktweise Gültigkeit, d.h.

$$\big(B^\mathsf{T}\bar{p}(t) + \nu\bar{u}(t)\big)^\mathsf{T}\big(u - \bar{u}(t)\big) \geq 0 \qquad \forall' t \in [0,T], \forall u \in U$$

mit $U = \{u \in R^m : a \leq u \leq b\}$. Wir betrachten für $i = 0, \ldots, N$ die Spezialfälle $u = u_h(t_i)$ und schließen unter Beachtung von $(P_1\bar{u})(t_i) = \bar{u}(t_i)$, $i = 0, \ldots, N$, für die Stützstellen auf

$$\begin{aligned}\big(B^\mathsf{T}\bar{p}(t_i) + \nu\bar{u}(t_i)\big)^\mathsf{T}&\big(u_h(t_i) - \bar{u}(t_i)\big)\\ &= \big(B^\mathsf{T}\bar{p}(t_i) + \nu(P_1\bar{u})(t_i)\big)^\mathsf{T}\big(u_h(t_i) - (P_1\bar{u})(t_i)\big) \geq 0, \qquad i = 0, \ldots, N\end{aligned}$$

und addieren über alle Indizes zu

$$\langle B^\mathsf{T}\bar{p} + \nu P_1\bar{u}, u_h - P_1\bar{u}\rangle_h \geq 0.$$

Weiterhin betrachten wir in Gleichung (5.5) den Spezialfall $\zeta_h = P_1\bar{u}$

$$\langle B^\mathsf{T}p_h(u_h) + \nu u_h, P_1\bar{u} - u_h\rangle_h \geq 0$$

und durch Addition der letzten beiden Ungleichungen erhalten wir

$$\langle B^\mathsf{T}(\bar{p} - p_h(u_h)) + \nu(P_1\bar{u} - u_h), u_h - P_1\bar{u}\rangle_h \geq 0.$$

Das ist äquivalent zu

$$\begin{aligned}\nu\|u_h - P_1\bar{u}\|_h^2 &\leq \langle\bar{p} - p_h(u_h), B(u_h - P_1\bar{u})\rangle_h\\ &= \langle\bar{p} - p_h(P_1\bar{u}), B(u_h - P_1\bar{u})\rangle_h + \langle p_h(P_1\bar{u}) - p_h(u_h), B(u_h - P_1\bar{u})\rangle_h.\end{aligned}$$

Die entstandenen Skalarprodukte schätzen wir nun getrennt voneinander ab.

Für den zweiten Summanden erinnern wir an die Darstellung der diskreten Systemgleichung (5.2) und der diskreten adjungierten Gleichung (5.4) mit Hilfe der Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$, nämlich $z_h(u) = \mathcal{S}_h\mathcal{T}u$ und $p_h(u) = \mathcal{S}_h^*(\mathcal{S}_h\mathcal{T}u - z_d)$. Weiterhin wissen wir aus Lemma 5.2.3, dass die Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$ bezüglich des Skalarprodukts $\langle\cdot,\cdot\rangle_h$ adjungiert sind. Daher folgt auf Grund der Linearität der diskreten Operatoren

$$\begin{aligned}\langle p_h(P_1\bar{u}) - p_h(u_h), B(u_h - P_1\bar{u})\rangle_h &= \langle\mathcal{S}_h^*\mathcal{S}_hB(P_1\bar{u} - u_h), B(u_h - P_1\bar{u})\rangle_h\\ &= \langle\mathcal{S}_hB(P_1\bar{u} - u_h), \mathcal{S}_hB(u_h - P_1\bar{u})\rangle_h\\ &= -\|\mathcal{S}_hB(P_1\bar{u} - u_h)\|_h^2 \leq 0.\end{aligned}$$

Somit fahren wir mit dem ersten Summanden fort und beachten $p_h(\bar{u}) = p_h(P_1\bar{u})$, was aus der Form des Differneznschemas (5.1) folgt. Demnach ist

$$\begin{aligned}\langle\bar{p} - p_h(P_1\bar{u}), B(u_h - P_1\bar{u})\rangle_h \quad &\leq \quad \|P_1\bar{p} - p_h(\bar{u})\|_h\|B\|\,\|P_1\bar{u} - u_h\|_h\\ &\leq \quad \sqrt{T}\|P_1\bar{p} - p_h(\bar{u})\|_\infty\|B\|\,\|P_1\bar{u} - u_h\|_h\\ &\overset{(5.11)}{\leq} \quad \sqrt{T}\|B\|\,c_{2.6}^2c_{5.9}\,C_{5.13}^{\bar{u}}\,h^2\,\|P_1\bar{u} - u_h\|_h.\end{aligned}$$

Damit folgt

$$\|u_h - P_1\bar{u}\|_h^2 \leq \sqrt{T}\frac{\|B\|}{\nu} c_{2.6}^2 c_{5.9}\, C_{5.13}^{\bar{u}}\, h^2\, \|P_1\bar{u} - u_h\|_h.$$

In dieser Ungleichung dürfen wir durch $\|P_1\bar{u} - u_h\|_h$ dividieren, da im Fall $u_h = \bar{u}$ nichts zu zeigen wäre, also

$$\|P_1\bar{u} - u_h\|_h \leq \sqrt{T}\frac{\|B\|}{\nu} c_{2.6}^2 c_{5.9}\, C_{5.13}^{\bar{u}}\, h^2.$$

□

Die Aussage des letzten Satzes bedeutet, dass die Lösungen der diskreten Probleme an den Stützstellen bereits mit der Ordnung 2 gegen die exakte Lösung konvergieren. Daher sprechen wir von *diskreter Konvergenz*. Im Folgenden ist es das Ziel, diese Konvergenzaussage auf das gesamte Intervall auszudehnen. Die Dreiecks-Ungleichung liefert sofort

$$\|\bar{u} - u_h\|_2 \leq \|\bar{u} - P_1\bar{u}\|_2 + \|P_1\bar{u} - u_h\|_2 \overset{(3.13)}{\leq} \mathsf{V}_0^T\dot{\bar{u}}\, h^{\frac{3}{2}} + \|P_1\bar{u} - u_h\|_h.$$

Für genügend kleine Schrittweite dominiert der erste Term die Summe, daher erhalten wir insgesamt die Konvergenzordnung $\frac{3}{2}$. Eine Erweiterung dieser Ergebnisse formulieren wir in folgendem Satz.

Satz 5.5.2. *Seien $\bar{u}$ die Lösung von* (StP) *mit $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$ und $u_h \in V_h[0,T,\mathbb{R}^m]$ die Lösung des diskreten Problems* $(\mathrm{StP})_h$*. Dann gilt für $\tilde{u} = \Pi_{[\mathrm{a,b}]}\big(-\frac{1}{\nu}B^{\mathsf{T}}p_h(u_h)\big)$ die Abschätzung*

$$\|\bar{u} - \tilde{u}\|_\infty \leq c^{\Pi} c_{2.6}^2 \max\{c_{2.7}, c_{5.9}\}\left(2 + \frac{T^4}{2}\frac{\|B\|^2}{\nu}\right) C_{5.13}^{\bar{u}}\, h^2.$$

Die Abkürzung c^{Π} steht dabei für $\frac{1}{\nu}\|B\|$.

Beweis: Es ist nach Satz 5.3.4

$$\|p(\bar{u}) - p_h(\bar{u})\|_\infty \leq 2c_{2.6}^2 \max\{c_{2.7}, c_{5.9}\}\, C_{5.13}^{\bar{u}}\, h^2.$$

Weiterhin gilt

$$\begin{aligned}\|p_h(\bar{u}) - p_h(u_h)\|_\infty &\overset{(5.3c)}{\leq} \frac{T^2}{\sqrt{2}}\|z_h(\bar{u}) - z_h(u_h)\|_\infty = \frac{T^2}{\sqrt{2}}\|z_h(P_1\bar{u}) - z_h(u_h)\|_\infty \\ &\overset{(5.3a)}{\leq} \frac{T^3}{2}\sqrt{T}\|B\|\,\|P_1\bar{u} - u_h\|_h \overset{(5.13)}{\leq} \frac{T^4}{2}\frac{\|B\|^2}{\nu} c_{2.6}^2 c_{5.9}\, C_{5.13}^{\bar{u}}\, h^2.\end{aligned}$$

Der Operator $\Pi_{[\mathrm{a,b}]}$ ist Lipschitz-stetig mit Konstante 1, daher folgt unter Verwendung der Dreiecks-Ungleichung

$$\|\bar{u} - \tilde{u}\|_\infty \leq c^{\Pi}\|p(\bar{u}) - p_h(u_h)\|_\infty \leq c^{\Pi} c_{2.6}^2 \max\{c_{2.7}, c_{5.9}\}\left(2 + \frac{T^4}{2}\frac{\|B\|^2}{\nu}\right) C_{5.13}^{\bar{u}}\, h^2.$$

□

5.6 Numerische Durchführung

5.6.1 Bestimmung der diskreten Operatoren und Probleme

Wie bereits bei den obigen theoretischen Untersuchung benutzt, fassen wir die Steuerung u_h auch als Vektor α mit und den Zustand z_h als Vektor β auf. Die Randwerte sind mit $\beta_0 = \beta_N = 0_n$ vorgegeben. Für die Diskretisierung mittels des Differenzenschemas (5.1) zeigten wir in Satz 5.2.1 die eindeutige Lösbarkeit der diskreten Systemgleichung (5.2). Das bedeutet, $\beta = \beta(\alpha)$ ist eine Funktion der diskreten Steuerung α. Diese Abbildung leiten wir nun mit Hilfe der Vektoren her.

Dazu erinnern wir an das Differenzenschema (5.1) und stellen das Gleichungssystem als Matrizengleichung dar. Es ist

$$\underbrace{\left[\frac{1}{h^2}\begin{pmatrix} 2I & -I & & & \\ -I & 2I & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & 2I & -I \\ & & & -I & 2I \end{pmatrix} + \begin{pmatrix} A & & & & \\ & A & & & \\ & & \ddots & & \\ & & & A & \\ & & & & A \end{pmatrix}\right]}_{=:\mathfrak{N}^{-1}} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{N-2} \\ \beta_{N-1} \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} 0 & I & 0 & & & \\ & 0 & I & 0 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 0 & I & 0 \\ & & & & 0 & I & 0 \end{pmatrix}}_{\mathfrak{C}} \left[\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{N-1} \\ \alpha_N \end{pmatrix} + \begin{pmatrix} e_0 \\ e_1 \\ \vdots \\ e_{N-1} \\ e_N \end{pmatrix}\right].$$

Wir benötigen die Matrix $\mathfrak{C} \in \mathbb{R}^{(N-1)n \times (N+1)n}$, da alle Komponenten von α später in den Optimierungsprozess einbezogen werden, aber für die Berechnung von β die Werte an den Randstellen nicht von Bedeutung sind. Auf Grund der Randwertvorgaben lassen wir β_0 und β_N sowohl in der Systemgleichung als auch bei der Zielfunktion außen vor. So erhalten wir mit $e_h = (e(t_0)^\mathsf{T}, \ldots, e(t_N)^\mathsf{T})^\mathsf{T}$ eine andere Formulierung der diskreten Systemgleichung, nämlich

$$\beta = \mathfrak{N}\mathfrak{C}(\alpha + e_h).$$

Mit Hilfe dieser vektoriellen Darstellung leiten wir nun aus dem Problem $(\mathrm{StP})_h$ eine andere Formulierung her. Dazu berechnen wir die Terme in der Zielfunktion getrennt. Es ist zunächst

$$\|z_h(u_h) - z_d\|_h^2 = \sum_{i=0}^{N} \beta_i^\mathsf{T}\beta_i + |z_d(t_i)|^2 - 2\beta_i^\mathsf{T} z_d(t_i).$$

Definieren wir $D = (z_d(t_1)^\mathsf{T}, \ldots, z_d(t_{N-1})^\mathsf{T})^\mathsf{T}$ und $\mathfrak{H} = (\mathfrak{N}\mathfrak{C})^\mathsf{T}\mathfrak{N}\mathfrak{C}$, so folgt sofort

$$\begin{aligned}\|z_h(u_h) - z_d\|_h^2 &= \beta^\mathsf{T}\beta - 2\beta^\mathsf{T}D + \|z_d\|_h^2 \\ &= (\alpha + e_h)^\mathsf{T}\mathfrak{H}(\alpha + e_h) - 2(\alpha + e_h)^\mathsf{T}\mathfrak{C}^\mathsf{T}\mathfrak{N}D + \|z_d\|_h^2 \\ &= \alpha^\mathsf{T}\mathfrak{H}\alpha + 2(\mathfrak{H}e_h - \mathfrak{C}^\mathsf{T}\mathfrak{N}D)^\mathsf{T}\alpha + (\mathfrak{H}e_h - 2\mathfrak{C}^\mathsf{T}\mathfrak{N}D)^\mathsf{T}e_h + \|z_d\|_h^2.\end{aligned}$$

Der Term $\frac{\nu}{2}\|u_h\|_h^2$ wandelt sich zu $\frac{\nu}{2}\alpha^\mathsf{T}\alpha$. Die Einträge von α gehen also quadratisch in J_h ein und sowohl f als auch c werden nicht verändert. So erhalten wir für die Zielfunktion folgenden Ausdruck

$$J_h(\alpha) = \frac{1}{2}\alpha^\mathsf{T}\underbrace{(\mathfrak{H} + \nu I_{(N+1)n})}_{=:H_\nu}\alpha + \underbrace{(\mathfrak{H}e_h - h\mathfrak{C}^\mathsf{T}\mathfrak{N}D)^\mathsf{T}}_{f}\alpha + \underbrace{\Big[\frac{1}{2}\mathfrak{H}e_h - h\mathfrak{C}^\mathsf{T}\mathfrak{N}D\Big]^\mathsf{T}e_h + \frac{1}{2}\|z_d\|_h^2}_{c}.$$

Die Hesse-Matrix H_ν ist offensichtlich symmetrisch und darüber hinaus positiv definit, denn der zweite Summand ist wegen $\nu > 0$ positiv definit und der erste Summand positiv semi-definit. Damit besitzt das diskrete Problem eine eindeutig bestimmte Lösung.

5.6.2 Ein Beispielproblem

Wählen wir nun im Fall $n = 1$ konkrete Werte für die Parameter. Seien $A = B = 1$, $\nu = 0.1$, $T = 1$, $z_d \equiv 2$, $a = -\infty$, $b = 2.5(\sqrt{2} - 1) \approx 1.0355$ und $e(t) = -2 + t^2 - t - \min\{-\frac{t^2-t}{\nu}, b\}$ (vgl. Hinze [14] (2005)), so stellt sich uns das Problem

$$\min_u \frac{1}{2}\|z(u) - 2\|_h^2 + \frac{\nu}{2}\|u\|_h^2, \qquad u \in U^{ad}.$$

unter den Nebenbedingungen:

$$\begin{aligned}&-\ddot{z}(t) + z(t) = u(t) + e(t), \qquad \forall' t \in [0, T] \\ &z(0) = z(T) = 0.\end{aligned}$$

Die diskreten Optimierungsprobleme lauten

$$\min_{u_h} \frac{1}{2}\|z_h(u_h) - d\|_h^2 + \frac{\nu}{2}\|u_h\|_h^2, \qquad u_h \in U_h^{ad}$$

oder

$$\min_\alpha \frac{1}{2}\alpha^\mathsf{T}H_\nu\alpha + f^\mathsf{T}\alpha + c, \qquad \alpha_i \le b,\ i = 0, \ldots, N.$$

Die optimale Steuerung ist gleich der Funktion $\bar{u}(t) = \min\{-\frac{t^2-t}{\nu}, b\}$. Es ergibt sich der optimale Zustand $z(\bar{u})(t) = t - t^2$ und der optimale adjungierte Zustand ist in diesem Beispiel identisch zu $z(\bar{u})$. Beide Aussagen prüfen wir leicht durch Einsetzen in die jeweilige Gleichung nach. Die Schnittpunkte von $\bar{u}$ und der oberen Schranke b sind an irrationalen Stellen, daher werden sie wegen $N \in \mathbb{N}$ niemals zu Stützstellen.

In den Tabellen 5.1 und 5.2 sind die Werte für den minimalen Zielfunktionswert, die Norm der Lösung des diskreten Optimierungsproblems, sowie der Abstand zwischen exakter und per Diskretisierung berechneter optimaler Steuerung bezüglich verschiedener Schrittweiten aufgelistet. Daneben finden wir den Abstand der exakten Steuerung und der neu definierten Funktion $\tilde{u}$, sowie in der letzten Spalte den Quotient aus dem Fehler und dem Quadrat der Schrittweite. An Hand der Konstanz dieses Wertes erhalten wir auch die numerische Bestätigung unserer theoretischen Fehlerabschätzungen.

h	$J_h(u_h)$	$\|u_h\|_2$
1/3	2.3399	0.7718
1/4	2.3607	0.8455
1/6	2.3767	0.9133
1/10	2.3835	0.9291
1/20	2.3858	0.9410
1/50	2.3866	0.9450
1/100	2.3867	0.9458

Tabelle 5.1: Daten der diskreten Steuerungsprobleme.

h	$\|\bar{u} - u_h\|_\infty$	$\|\bar{u} - \tilde{u}\|_2$	$\|\bar{u} - \tilde{u}\|_\infty$	$\|\bar{u} - \tilde{u}\|_\infty / h^2$
1/3	0.6696	0.0904	0.2531	2.7800
1/4	0.5485	0.0567	0.1556	2.4900
1/6	0.3062	0.0269	0.0694	2.5000
1/10	0.1709	0.0082	0.0249	2.4862
1/20	0.1348	0.0021	0.0062	2.4826
1/50	0.0843	$3.125 \cdot 10^{-4}$	$9.818 \cdot 10^{-4}$	2.4545
1/100	0.0840	$5.591 \cdot 10^{-5}$	$2.415 \cdot 10^{-4}$	2.4148

Tabelle 5.2: Fehler der diskreten optimalen Steuerungen bei verschiedenen Schrittweiten.

Zur Verdeutlichung des Inhalts von Tabelle 5.2 seien für beide Fehlermaße der Verlauf des Fehlers in Abhängigkeit von der Schrittweite in den Abbildungen 5.1 und 5.2 dargestellt. Wir erkennen in beiden Fällen die quadratische Abnahme des Fehlers bei sich verringernder Schrittweite.

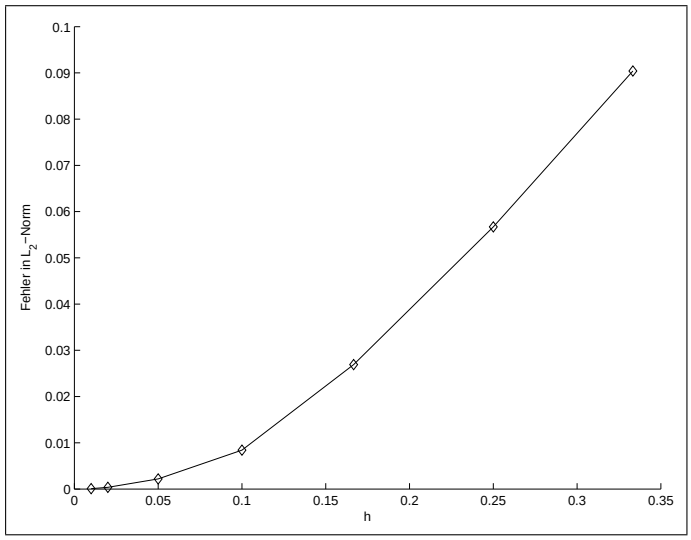

Abbildung 5.1: L_2-Fehler für das Differenzenverfahren

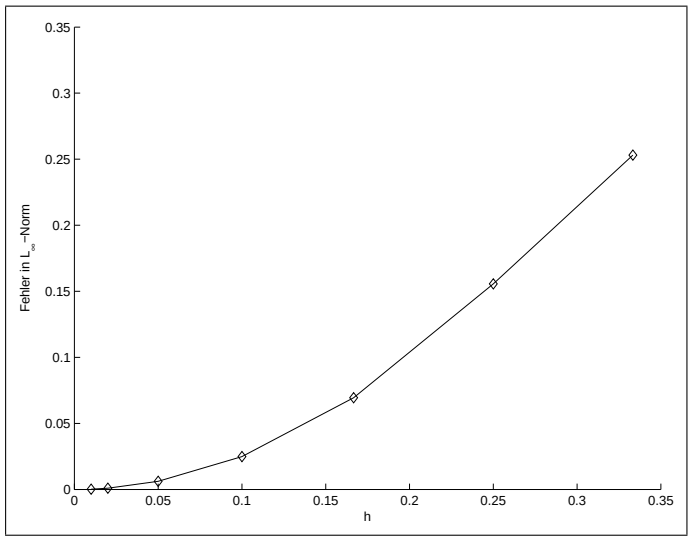

Abbildung 5.2: L_∞-Fehler für das Differenzenverfahren

5.7 Veränderte Diskretisierung

Vergleichen wir die zu Beginn dieses Kapitels untersuchte Methode der Finiten Differenzen mit den Finiten Elementen aus Kapitel 4, so fällt als erster Unterschied die schärferen Voraussetzung für die Konvergenz bei der Lösung der Systemgleichung auf. Anstelle der Beschränktheit (vgl. Lemma 4.3.2) fordern wir hier die beschränkte Variation der Ableitung der rechten Seite (vgl. Satz 5.3.4) um quadratische Konvergenzordnung zu erhalten. Dieser Umstand stellt - bis auf die komplexen Abschätzungskonstanten - keine Hürde dar, denn die Forderung der beschränkten Variation ist unverzichtbar. Wir benötigen sie später bei den Betrachtungen zur Approximation der Steuerung. Dennoch untersuchen wir eine Möglichkeit der Diskretisierung, diese Einschränkung (vorerst) zu umgehen. Dazu definieren wir die rechte Seite der diskreten Systemgleichung bzw. des Differenzenschemas (5.1) über einen Integralausdruck. So sei die Systemgleichung (1.4) diskretisiert durch

$$\begin{aligned} z_h(t)\Big|_{T_i} &= (t-t_i)\frac{\beta_{i+1}-\beta_i}{h}+\beta_i, && i=0,\dots,N-1 \\ -(\beta_i)_{\mathsf{h}\bar{\mathsf{h}}}+A\beta_i &= \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}}(\mathcal{T}u)(t_i+t)\,dt, && i=1,\dots,N-1 \\ \beta_0 &= \beta_N = 0_n \end{aligned} \tag{5.14}$$

mit $\mathcal{T}u = Bu + e$ und die adjungierte Gleichung (2.10) durch

$$\begin{aligned} p_h(t)\Big|_{T_i} &= (t-t_i)\frac{\zeta_{i+1}-\zeta_i}{h}+\zeta_i, && i=0,\dots,N-1 \\ -(\zeta_i)_{\mathsf{h}\bar{\mathsf{h}}}+A\zeta_i &= \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}}(z_h-z_d)(t_i+t)\,dt, && i=1,\dots,N-1 \\ \zeta_0 &= \zeta_N = 0_n. \end{aligned} \tag{5.15}$$

Die Stabilität zeigen wir wieder mit der Energiemethode aus Abschnitt 5.2. Weil A positiv semi-definit ist, folgt

$$\|\dot{z}_h\|_h^2 \le h\sum_{i=1}^{N-1}\beta_i^{\mathsf{T}}\left(\frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}}(\mathcal{T}u)(t_i+t)\,dt\right) \le \|z_h\|_\infty\|\mathcal{T}u\|_1.$$

Mit (3.5) folgt dann für $z_h \neq 0$

$$\|z_h\|_\infty \le T\|\mathcal{T}u\|_1 \le T^{\frac{3}{2}}\|\mathcal{T}u\|_2 \le T^2\|\mathcal{T}u\|_\infty. \tag{5.16}$$

Im Fall $z_h = 0$ ist die Ungleichung trivialerweise erfüllt. Damit sichern wir die Eindeutigkeit der Lösungen von (5.14) bzw. (5.15) und somit auch die Darstellung durch die linearen beschränkten Operatoren

$$z_h(u) = \mathcal{S}_h(Bu+e), \qquad p_h(u) = \mathcal{S}_h^*(\mathcal{S}_h(Bu+e)-z_d).$$

Durch die veränderte Diskretisierung gilt im Gegensatz zum obigen Differenzenverfahren im Allgemeinen nicht $z_h(u) = z_h(P_1 u)$. Das bedeutet, wir benötigen ein analoges Resultat zu Lemma 4.5.10. Dazu nutzen wir die erste Ungleichung der eben hergeleiteten Stabilitätsungleichung und schließen

$$\|z_h(\bar{u}) - z_h(P_1\bar{u})\|_\infty \le T\|B\|\,\|\bar{u} - P_1\bar{u}\|_1 \overset{(3.13)}{\le} T\|B\|\,\mathsf{V}_0^T\dot{\bar{u}}\,h^2.$$

Bei der Untersuchung der Konvergenz verwenden wir im Unterschied zur obigen Vorgehensweise die Abschätzung

$$|\psi_i^1| = \Big|Az_i - \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} Az(t_i+t)\,dt\Big| \le \frac{h}{4}\|A\|\,\|\ddot{z}\|_{L_1(S_{i-1},S_i)}, \qquad i = 1,\dots,N-1.$$

So erhalten wir auf Grund der Vereinfachung schon bei der Definition von ψ_i^1 die Abschätzungen (vgl. Satz 5.3.4)

$$\begin{aligned}
\|P_1 z(u) - z_h(u)\|_\infty &\le c_{5.9}\,\|\mathcal{T}u\|_{1,\infty}\,h^2 \\
\|P_1 p(u) - p_h(u)\|_\infty &\le c_{2.6}^2 c_{5.9}\,(\|\mathcal{T}u\|_{1,\infty} + \|z_d\|_{1,\infty})\,h^2 \qquad (5.17)\\
\|p(u) - p_h(u)\|_\infty &\le 2c_{2.6}^2\max\{c_{2.7}, c_{5.9}\}\,(\|\mathcal{T}u\|_{1,\infty} + \|z_d\|_{1,\infty})\,h^2. \qquad (5.18)
\end{aligned}$$

Die Voraussetzung $\dot{u} \in BV[0,T,\mathbb{R}^m]$ brauchen wir an dieser Stelle (noch) nicht.

Betrachten wir eine stückweise lineare Approximation der Steuerung, also $u_h \in V_h[0,T,\mathbb{R}^m]$, so erhalten wir auf Grund der Linearität von u_h auf den Teilintervallen mit $u_h(t_i) = \alpha_i$ für $i = 0,\dots,N$

$$\begin{aligned}
\int_0^{\frac{h}{2}} u_h(t_i+t)\,dt &= \frac{h}{8}(\alpha_1 + 3\alpha_0)\\
\int_{-\frac{h}{2}}^{\frac{h}{2}} u_h(t_i+t)\,dt &= \frac{h}{8}(\alpha_{i+1} + 6\alpha_i + \alpha_{i-1}), \qquad i = 1,\dots,N-1\\
\int_{-\frac{h}{2}}^{0} u_h(t_i+t)\,dt &= \frac{h}{8}(3\alpha_N + \alpha_{N-1})
\end{aligned}$$

Mit Kenntnis dieser Darstellung bilden wir auf dem Raum $V_h[0,T,\mathbb{R}^n]$ das Skalarprodukt

$$\langle v_h, w_h\rangle_M = \begin{pmatrix} v_h(t_0)\\ \vdots\\ v_h(t_N)\end{pmatrix}^{\mathsf{T}} \underbrace{\frac{h}{8}\begin{pmatrix} 3 & 1 & & & \\ 1 & 6 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 6 & 1\\ & & & 1 & 3\end{pmatrix}}_{=:M} \begin{pmatrix} w_h(t_0)\\ \vdots\\ w_h(t_N)\end{pmatrix} \qquad \forall v_h, w_h \in V_h[0,T,\mathbb{R}^n]$$

und definieren die Norm $\|\cdot\|_M = \sqrt{\langle\cdot,\cdot\rangle_M}$. Jetzt stellen wir ein diskretes Steuerungsproblem auf, nämlich

$$(\mathrm{StP})_h^M \qquad \min_{u_h} \frac{1}{2}\|\mathcal{S}_h\mathcal{T}u_h - z_d\|_M^2 + \frac{\nu}{2}\|u_h\|_M^2, \qquad u_h \in U_h^{ad},$$

mit $U_h^{ad} = V_h[0,T,\mathbb{R}^m] \cap U^{ad}$. Zur Begründung für die Wahl der Norm $\|\cdot\|_M$ betrachten wir die diskreten Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$. Sie sind auf dem Raum $V_{h,0}[0,T,\mathbb{R}^n]$ bezüglich des Skalarprodukts $\langle\cdot,\cdot\rangle_M$ adjungiert, denn es gilt

$$\langle z_h(v_h), w_h\rangle_M = \langle v_h, z_h(w_h)\rangle_M.$$

Das sehen wir leicht ein, denn die rechte Seite von (5.14) ist für stückweise lineare Funktionen darstellbar als Multiplikation des äquivalenten Vektors mit der Matrix M. Die notwendige Optimalitätsbedingung für die Lösung u_h von $(\mathrm{StP})_h^M$ lautet (wieder mit Alt [2] (2005), Satz 4.2.2) wie folgt

$$\langle B^\mathsf{T} p_h(u_h) + \nu u_h, \zeta_h - u_h\rangle_M \geq 0 \qquad \forall \zeta_h \in U_h^{ad}.$$

Nun bereiten wir unsere Hauptergebnisse vor und untersuchen Skalarprodukte zweier Funktionen u_h und p_h aus dem $V_h[0,T,\mathbb{R}^n]$ und deren Darstellung mit Hilfe der äquivalenten Vektoren α und ζ. Es ist

$$\langle u_h, p_h\rangle_M = \alpha^\mathsf{T} M \zeta.$$

Erinnern wir an die Matrix $\mathfrak{H}$ aus Satz 4.5.11 mit

$$\langle u_h, p_h\rangle_{\mathfrak{H}} = \alpha^\mathsf{T}\mathfrak{H}\zeta = \alpha^\mathsf{T}\frac{h}{8}\begin{pmatrix} 4I & & & & \\ & 8I & & & \\ & & \ddots & & \\ & & & 8I & \\ & & & & 4I \end{pmatrix}\zeta,$$

so folgt

$$\begin{aligned}\langle u_h, p_h\rangle_M - \langle u_h, p_h\rangle_{\mathfrak{H}} &= \alpha^\mathsf{T}\frac{h}{8}\begin{pmatrix} -I & I & & & & \\ I & -2I & I & & & \\ & & \ddots & & & \\ & & & I & -2I & I \\ & & & & I & -I \end{pmatrix}\zeta \\ &= \frac{h^2}{8}\langle\alpha, \zeta_{\mathsf{h}\bar{\mathsf{h}}}\rangle_h \leq \frac{h^2}{8}\|u_h\|_h\|(p_h)_{\mathsf{h}\bar{\mathsf{h}}}\|_h.\end{aligned}$$

Diese Vorarbeiten werden wir beim Beweis des folgenden Satz benötigen.

Satz 5.7.1. *Seien $\bar{u}$ die Lösung von* (StP) *mit $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$ und $u_h \in V_h[0,T,\mathbb{R}^n]$ die Lösung des diskreten Steuerungsproblems* $(\mathrm{StP})_h^M$. *Dann gilt*

$$\|u_h - P_1\bar{u}\|_2 \le \frac{1}{2}(\|\dot{\bar{u}}\|_\infty + \mathsf{V}_0^T\dot{\bar{u}})\, h^{\frac{3}{2}}. \tag{5.19}$$

Beweis: Wir beginnen analog zu Satz 5.5.1 mit den Optimalitätsbedingungen des kontinuierlichen Problems (StP). Dort setzen wir zunächst $u = u_h(t_i)$ für $i = 0,\dots,N$ und auf Grund ihrer punktweisen Gültigkeit (vgl. Lemma 2.4.3) folgern wir für die Stützstellen t_i und $i = 0,\dots,N$

$$\big(B^{\mathsf{T}}\bar{p}(t_i) + \nu\bar{u}(t_i)\big)^{\mathsf{T}}\big(u_h(t_i) - \bar{u}(t_i)\big) \ge 0.$$

Multiplizieren wir nun die Ungleichungen für $i = 0$ und $i = N$ mit dem Faktor $\frac{1}{2}$ und addieren über alle Indizes, so erhalten wir unter Beachtung von $\bar{u}(t_i) = (P_1\bar{u})(t_i)$ für $i = 0,\dots,N$

$$\begin{aligned} 0 &\le \langle B^{\mathsf{T}}\bar{p} + \nu\bar{u}, u_h - \bar{u}\rangle_{\mathfrak{H}} = \langle B^{\mathsf{T}}\bar{p} + \nu\bar{u}, u_h - P_1\bar{u}\rangle_{\mathfrak{H}} \\ &= \langle B^{\mathsf{T}}\bar{p} + \nu\bar{u}, u_h - P_1\bar{u}\rangle_M - \frac{h^2}{8}\langle (B^{\mathsf{T}}\bar{p} - \nu\bar{u})_{\mathsf{h}\bar{\mathsf{h}}}, u_h - P_1\bar{u}\rangle_h. \end{aligned}$$

Greifen wir an dieser Stelle die Optimalitätsbedingung für das Problem $(\mathrm{StP})_h^M$ mit $\zeta_h = P_1\bar{u}$ auf

$$\langle B^{\mathsf{T}}p_h(u_h) + \nu u_h, P_1\bar{u} - u_h\rangle_M \ge 0.$$

Addieren wir die beiden Ungleichungen, so folgt

$$\nu\|u_h - P_1\bar{u}\|_M \le \langle P_1\bar{p} - p_h(u_h), B(u_h - P_1\bar{u})\rangle_M + \frac{h^2}{8}\langle (B^{\mathsf{T}}\bar{p} - \nu\bar{u})_{\mathsf{h}\bar{\mathsf{h}}}, P_1\bar{u} - u_h\rangle_h.$$

Der zweite Summand ist nach Lemma 3.2.9 nach oben beschränkt, denn es gilt

$$\begin{aligned} h^2\langle (B^{\mathsf{T}}\bar{p} - \nu\bar{u})_{\mathsf{h}\bar{\mathsf{h}}}, P_1\bar{u} - u_h\rangle_h &\le \|u_h - P_1\bar{u}\|_h \|(B^{\mathsf{T}}\bar{p} - \nu\bar{u})_{\mathsf{h}\bar{\mathsf{h}}}\|_h\, h^2 \\ &\le 2\|u_h - P_1\bar{u}\|_h \Big(\|B\|\, \|\ddot{p}(\bar{u})\|_2\, h^2 + \nu(\mathsf{V}_0^T\dot{\bar{u}} + \|\dot{\bar{u}}\|_\infty)\, h^{\frac{3}{2}} \Big). \end{aligned}$$

Das vordere Skalarprodukt auf der rechten Seite spalten wir auf

$$\begin{aligned} \langle P_1\bar{p} - p_h(u_h), B(u_h - P_1\bar{u})\rangle_M &= \langle p_h(P_1\bar{u}) - p_h(u_h), B(u_h - P_1\bar{u})\rangle_M \\ &\quad + \langle p_h(\bar{u}) - p_h(P_1\bar{u}), B(u_h - P_1\bar{u})\rangle_M \\ &\quad + \langle P_1\bar{p} - p_h(\bar{u}), B(u_h - P_1\bar{u})\rangle_M \end{aligned}$$

und schätzen die entstandenen Terme getrennt voneinander ab.

Für den letzten Summanden erinnern wir an die Darstellung der Lösung des Differenzenschemas (5.14) mit Hilfe der Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$. Diese sind bezüglich des M-Skalarprodukts adjungiert und somit folgt

$$\begin{aligned}\langle p_h(P_1\bar{u}) - p_h(u_h), B(u_h - P_1\bar{u})\rangle_M &= \langle \mathcal{S}_h^*\mathcal{S}_h B(P_1\bar{u} - u_h), B(u_h - P_1\bar{u})\rangle_M \\ &= \langle \mathcal{S}_h B(P_1\bar{u} - u_h), \mathcal{S}_h B(u_h - P_1\bar{u})\rangle_M \\ &= -\|\mathcal{S}_h B(P_1\bar{u} - u_h)\|_M^2 \leq 0.\end{aligned}$$

Somit fahren wir mit den anderen Summanden fort

$$\begin{aligned}\langle p_h(\bar{u}) - p_h(P_1\bar{u}), B(u_h - P_1\bar{u})\rangle_M &\leq \|p_h(\bar{u}) - p_h(P_1\bar{u})\|_M \|B\| \, \|u_h - P_1\bar{u}\|_M \\ &\leq \sqrt{T}\|p_h(\bar{u}) - p_h(P_1\bar{u})\|_\infty \|B\| \, \|u_h - P_1\bar{u}\|_M \\ &\overset{(5.16)}{\leq} T^3\sqrt{T}\|\bar{u} - P_1\bar{u}\|_1 \|B\| \, \|u_h - P_1\bar{u}\|_M \\ &\overset{(3.13)}{\leq} T^3\sqrt{T}\|B\| \, \mathsf{V}_0^T \dot{\bar{u}} \, \|u_h - P_1\bar{u}\|_M \, h^2.\end{aligned}$$

Außerdem gilt

$$\begin{aligned}&\langle P_1\bar{p} - p_h(\bar{u}), B(u_h - P_1\bar{u})\rangle_M \\ &\quad\leq \|P_1\bar{p} - p_h(\bar{u})\|_M \|B\| \, \|u_h - P_1\bar{u}\|_M \\ &\quad\leq \sqrt{T}\|P_1\bar{p} - p_h(\bar{u})\|_\infty \|B\| \, \|u_h - P_1\bar{u}\|_M \\ &\quad\overset{(5.17)}{\leq} \sqrt{T}\|B\| \, c_{2.6}^2 c_{5.9} \left(\|\mathcal{T}\bar{u}\|_{1,\infty} + \|z_d\|_{1,\infty}\right) h^2 \, \|u_h - P_1\bar{u}\|_M.\end{aligned}$$

Damit folgt

$$\begin{aligned}\nu\|u_h - P_1\bar{u}\|_M^2 \leq \Big(&T^3\sqrt{T}\|B\| \, \mathsf{V}_0^T \dot{\bar{u}} \, h^2 + \sqrt{T}\|B\| \, c_{2.6}^2 c_{5.9} \left(\|\mathcal{T}\bar{u}\|_{1,\infty} + \|z_d\|_{1,\infty}\right) h^2 \\ &+ \frac{\|B\|}{4}\|\ddot{p}(\bar{u})\|_2 \, h^2 + \frac{\nu}{4}(\|\dot{\bar{u}}\|_\infty + \mathsf{V}_0^T \dot{\bar{u}}) \, h^{\frac{3}{2}}\Big) \, \|u_h - P_1\bar{u}\|_M.\end{aligned}$$

In dieser Ungleichung dürfen wir durch $\|u_h - P_1\bar{u}\|_M$ dividieren, da im Fall $u_h = P_1\bar{u}$ nichts zu zeigen wäre. Demnach gilt für hinreichend kleine Schrittweite

$$\|u_h - P_1\bar{u}\|_M \leq \frac{1}{4}(\|\dot{\bar{u}}\|_\infty + \mathsf{V}_0^T \dot{\bar{u}}) \, h^{\frac{3}{2}}.$$

Die Eigenwerte der Matrix M liegen nach dem Satz von Gerschgorin (siehe Anhang, Satz A.3.3) im Intervall $[\frac{h}{4}, h]$, daher folgt mit Lemma 3.2.1

$$\|u_h - P_1\bar{u}\|_2 \leq \|u_h - P_1\bar{u}\|_h \leq 2\|u_h - P_1\bar{u}\|_M$$

und somit die Aussage. □

Auch bei dieser Diskretisierungsmethode zeigten wir zunächst die diskrete Konvergenz der beiden Lösungen von (StP) und (StP)$_h$. Davon ausgehend leiten wir sofort eine Abschätzung für den Abstand über dem gesamten Intervall her.

Satz 5.7.2. *Seien $\bar{u}$ die Lösung von* (StP) *mit $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$ und $u_h \in V_h[0,T,\mathbb{R}^m]$ die Lösung des diskreten Steuerungsproblems* $(\mathrm{StP})_h^M$*. Dann gilt für hinreichend kleine Schrittweite*

$$\|\bar{u} - u_h\|_2 \leq (\|\dot{\bar{u}}\|_\infty + \mathsf{V}_0^T \dot{\bar{u}})\, h^{\frac{3}{2}}.$$

Beweis: Wir verwenden die Dreiecks-Ungleichung, das eben hergeleitete Resultat sowie Ungleichung (3.13) und erhalten direkt die Aussage. □

Als letzten Schritt prüfen wir die Möglichkeit einer Verbesserung der Approximation durch Definition und Verwendung der Funktion $\tilde{u} = \Pi_{[\mathrm{a,b}]}\big(-\frac{1}{\nu} B^{\mathsf{T}} p_h(u_h)\big)$, wie bereits in den obigen Fällen erreicht.

Satz 5.7.3. *Seien $\bar{u}$ die Lösung von* (StP) *mit $\dot{\bar{u}} \in BV[0,T,\mathbb{R}^m]$ und $u_h \in V_h[0,T,\mathbb{R}^m]$ die Lösung des diskreten Problems* $(\mathrm{StP})_h^M$*. Dann gilt für $\tilde{u} = \Pi_{[\mathrm{a,b}]}\big(-\frac{1}{\nu} B^{\mathsf{T}} p_h(u_h)\big)$ und hinreichend kleine Schrittweite die Abschätzung*

$$\|\bar{u} - \tilde{u}\|_\infty \leq T^3 \frac{\sqrt{T}}{2\nu} \|B\|\, (\|\dot{\bar{u}}\|_\infty + \mathsf{V}_0^T \dot{\bar{u}})\, h^{\frac{3}{2}}.$$

Beweis: Es ist

$$\|\bar{p} - p_h(\bar{u})\|_\infty \overset{(5.18)}{\leq} 2c_{2.6}^2 \max\{c_{2.7}, c_{5.9}\}\, (\|\mathcal{T}\bar{u}\|_{1,\infty} + \|z_d\|_{1,\infty})\, h^2.$$

Darüber hinaus gilt

$$\|p_h(\bar{u}) - p_h(P_1\bar{u})\|_\infty \overset{(5.16)}{\leq} T^3 \|B\|\, \|\bar{u} - P_1\bar{u}\|_1 \leq T^3 \|B\|\, \mathsf{V}_0^T \dot{\bar{u}}\, h^2.$$

Weiterhin wissen wir für hinreichend kleine Schrittweite die Abschätzung

$$\|p_h(P_1\bar{u}) - p_h(u_h)\|_\infty \leq T^3 \sqrt{T} \|P_1\bar{u} - u_h\|_2 \overset{(5.19)}{\leq} T^3 \frac{\sqrt{T}}{2}\, (\|\dot{\bar{u}}\|_\infty + \mathsf{V}_0^T \dot{\bar{u}})\, h^{\frac{3}{2}}.$$

Damit folgt unter Verwendung der Dreiecks-Ungleichung und der Lipschitz-Stetigkeit des Operators $\Pi_{[\mathrm{a,b}]}$ für hinreichend kleine Schrittweite

$$\|\bar{u} - \tilde{u}\|_\infty \leq \frac{1}{\nu} \|B\|\, \|\bar{p} - p_h(u_h)\|_\infty \leq T^3 \frac{\sqrt{T}}{2\nu} \|B\|\, (\|\dot{\bar{u}}\|_\infty + \mathsf{V}_0^T \dot{\bar{u}})\, h^{\frac{3}{2}}.$$

□

Numerische Durchführung

Wieder benutzen wir die Darstellung der Funktionen u_h und z_h durch die äquivalenten Vektoren $\alpha = (\alpha_0^{\mathsf{T}}, \ldots, \alpha_N^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{(N+1)n}$ mit $\alpha_i = u_h(t_i)$ und $\beta = (\beta_1^{\mathsf{T}}, \ldots, \beta_{N-1}^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{(N+1)n}$ mit $\beta_i = z_h(t_i)$. Wir erinnern an das Differenzenschema (5.14) und stellen

das Gleichungssystem als Matrizengleichung dar. Es ist mit $e_i = \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} e(t_i + t)\,dt$ für $i = 1, \ldots, N-1$

$$\underbrace{\left[\frac{1}{h^2}\begin{pmatrix} 2I & -I & & & \\ -I & 2I & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & 2I & -I \\ & & & -I & 2I \end{pmatrix} + \begin{pmatrix} A & & & \\ & \ddots & & \\ & & & A \end{pmatrix}\right]}_{=:\mathfrak{N}^{-1}} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{N-2} \\ \beta_{N-1} \end{pmatrix}$$

$$= \underbrace{\frac{1}{8}\begin{pmatrix} I & 6I & I & & \\ & \ddots & \ddots & \ddots & \\ & & I & 6I & I \end{pmatrix}\begin{pmatrix} B & & & \\ & \ddots & & \\ & & & B \end{pmatrix}}_{\mathfrak{C}} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{N-1} \\ \alpha_N \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_{N-1} \end{pmatrix}.$$

Auf Grund der Randwertvorgaben lassen wir β_0 und β_N sowohl in der Systemgleichung als auch bei der Zielfunktion außen vor und erhalten mit $e_h = (e_1^\mathsf{T}, \ldots, e_{N-1}^\mathsf{T})^\mathsf{T}$ eine andere Formulierung der diskreten Systemgleichung, nämlich

$$\beta = \mathfrak{N}(\mathfrak{C}\alpha + e_h).$$

Mit Hilfe dieser vektoriellen Darstellung leiten wir nun aus der Problemstellung $(\mathrm{StP})_h^M$ eine andere Formulierung her. Dazu berechnen wir die Terme in der Zielfunktion getrennt und definieren $D = (z_d(t_1)^\mathsf{T}, \ldots, z_d(t_{N-1})^\mathsf{T})^\mathsf{T}$. Dann folgt

$$\begin{aligned} &\|z_h(u_h) - z_d\|_M^2 \\ &= \beta^\mathsf{T} M\beta - 2\beta^\mathsf{T} MD + \|z_d\|_M^2 \\ &= (\mathfrak{C}\alpha + e_h)^\mathsf{T}\mathfrak{N}M\mathfrak{N}(\mathfrak{C}\alpha + e_h) - 2(\mathfrak{C}\alpha + e_h)^\mathsf{T}\mathfrak{N}MD + \|z_d\|_M^2 \\ &= \alpha^\mathsf{T}(\mathfrak{N}\mathfrak{C})^\mathsf{T} M\mathfrak{N}\mathfrak{C}\alpha + 2(\mathfrak{C}^\mathsf{T}\mathfrak{N}M\mathfrak{N}e_h - \mathfrak{C}^\mathsf{T}\mathfrak{N}MD)^\mathsf{T}\alpha \\ &\quad + (\mathfrak{N}M\mathfrak{N}e_h - 2\mathfrak{N}MD)^\mathsf{T} e_h + \|z_d\|_M^2. \end{aligned}$$

Der Term $\frac{\nu}{2}\|u_h\|_M^2$ wandelt sich zu $\frac{\nu}{2}\alpha^\mathsf{T} M\alpha$. Die Einträge von α gehen also quadratisch in J_h ein und sowohl f als auch c werden nicht verändert. So erhalten wir für die Zielfunktion folgenden Ausdruck

$$J_h(\alpha) = \frac{1}{2}\alpha^\mathsf{T}\underbrace{\left((\mathfrak{N}\mathfrak{C})^\mathsf{T} M\mathfrak{N}\mathfrak{C} + \nu M\right)}_{=:H_\nu}\alpha + \underbrace{\left(\mathfrak{C}^\mathsf{T}\mathfrak{N}M\mathfrak{N}e_h - \mathfrak{C}^\mathsf{T}\mathfrak{N}D\right)^\mathsf{T}}_{f}\alpha + \underbrace{\left[\frac{1}{2}\mathfrak{N}M\mathfrak{N}e_h - \mathfrak{N}MD\right]^\mathsf{T} e_h + \frac{1}{2}\|z_h\|_M^2}_{c}.$$

Die Hesse-Matrix H_ν ist offensichtlich symmetrisch und darüber hinaus positiv definit, denn der zweite Summand ist positiv definit und der erste Summand positiv semidefinit. Damit besitzt das diskrete Problem wegen $\nu > 0$ eine eindeutig bestimmte Lösung.

Beispielproblem

Auch hier untersuchen wir die theoretischen Ergebnisse an einem numerischen Beispiel. Dazu erinnern wir an das Ausgangsproblem (StP) und dessen hier betrachtete Diskretisierung $(\mathrm{StP})_h^M$. Die Parameter setzen wir wieder auf $n = 1$, $A = B = 1$, $\nu = 0.1$, $T = 1$, $d = 2$, $a = -\infty$, $b = 2.5(\sqrt{2}-1) \approx 1.0355$ und $e(t) = -2+t^2-t-\min\{-\frac{t^2-t}{\nu}, b\}$ (vgl. Hinze [14] (2005)). Die optimale Steuerung für unsere Problemstellung ist gleich der Funktion $\bar{u}(t) = \min\{-\frac{t^2-t}{\nu}, b\}$. Es ergibt sich der optimale Zustand $z(\bar{u})(t) = t-t^2$ und der optimale adjungierte Zustand ist in diesem Beispiel identisch zu $z(\bar{u})$. Beide Aussagen prüfen wir leicht durch Einsetzen in die jeweilige Gleichung nach. Die Schnittpunkte von $\bar{u}$ und der oberen Schranke b sind an irrationalen Stellen, daher werden sie wegen $N \in \mathbb{N}$ niemals zu Stützstellen.

In den Tabellen 5.3 und 5.4 sind die Werte für den minimalen Zielfunktionswert, die Norm der Lösung des diskreten Optimierungsproblems, sowie der Abstand zwischen exakter und der diskreten optimalen Steuerung bezüglich verschiedener Schrittweiten aufgelistet. Daneben finden wir den Abstand der exakten Steuerung und der neu definierten Funktion $\tilde{u}$, sowie in der letzten Spalte den Quotient des Fehlers und dem Quadrat der Schrittweite. An Hand der Konstanz dieses Wertes erhalten wir auch die numerische Bestätigung unserer theoretischen Fehlerabschätzungen.

h	$J_h(u_h)$	$\|u_h\|_2$
1/3	2.5351	0.9281
1/4	2.4862	0.8941
1/6	2.4420	0.9169
1/10	2.4138	0.9606
1/20	2.4004	0.9566
1/50	2.3964	0.9543
1/100	2.3957	0.9541

Tabelle 5.3: Daten der diskreten Steuerungsprobleme.

Zur Verdeutlichung des Inhalts von Tabelle 5.4 seien für $\|\bar{u}-u_h\|_\infty$ und $\|\bar{u}-\tilde{u}\|_\infty$ der Verlauf des Fehlers in Abhängigkeit von der Schrittweite in den Abbildungen 5.3 und 5.4 dargestellt. Wir erkennen in beiden Fällen die quadratische Abnahme des Fehlers bei sich verringernder Schrittweite, d.h. numerisch erhalten wir eine um die Ordnung $\frac{1}{2}$ höhere Konvergenzgeschwindigkeit als bei den theoretischen Untersuchungen.

h	$\|\bar{u} - u_h\|_\infty$	$\|\bar{u} - \tilde{u}\|_2$	$\|\bar{u} - \tilde{u}\|_\infty$	$\|\bar{u} - \tilde{u}\|_\infty / h^2$
1/3	0.5579	0.0757	0.2122	1.9102
1/4	0.3073	0.0465	0.1240	1.9847
1/6	0.0530	0.0218	0.0567	2.0401
1/10	0.1355	0.0069	0.0218	2.1843
1/20	0.0683	0.0018	0.0058	2.3379
1/50	0.0279	$2.925 \cdot 10^{-4}$	$9.719 \cdot 10^{-4}$	2.4298
1/100	0.0140	$7.331 \cdot 10^{-5}$	$2.465 \cdot 10^{-4}$	2.4652

Tabelle 5.4: Fehler der diskreten optimalen Steuerungen bei verschiedenen Schrittweiten.

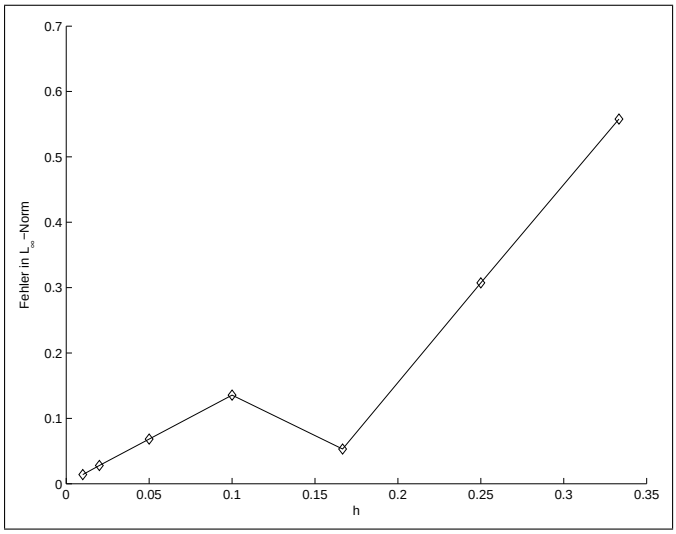

Abbildung 5.3: $\|\bar{u} - u_h\|_\infty$ für das alternative Differenzenverfahren

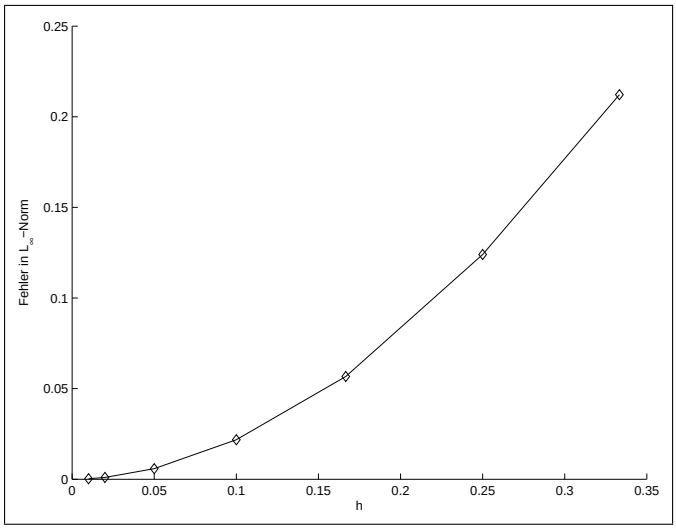

Abbildung 5.4: $\|\bar{u} - \tilde{u}\|_\infty$ für das alternative Differenzenverfahren

Kapitel 6

Zusammenfassung

Nach der theoretischen Untersuchung des Steuerungsproblems (StP), sowie der Diskretisierung durch zwei verschiedene Methoden fassen wir abschließend die erreichten Ergebnisse zusammen. Die theoretischen Aussagen zur Stabilität und Konvergenz im Ausgangsproblem und bei den Diskretisierungen lassen wir an dieser Stelle außen vor und beschränken uns auf die Resultate bezüglich der Konvergenzgeschwindigkeit. In Tabelle 6.1 sind die Voraussetzungen an die optimale Steuerung für die Konvergenzordnung 2 bzw. $\frac{3}{2}$ aufgeführt. Dabei seien zunächst die Hilfsresultate auf dem Weg zur Abschätzung der Fehler $\|\bar{u} - u_h\|_2$ bzw. $\|\bar{u} - \tilde{u}\|_\infty$ wiedergegeben.

Resultat	FEM con	FEM lin	Diff	alt. Diff
$\mathcal{S}^*\mathcal{S} - \mathcal{S}_h^*\mathcal{S}_h$	$\|\bar{u}\|_\infty < \infty$	$\|\bar{u}\|_\infty < \infty$	$\mathsf{V}_0^T\dot{\bar{u}} < \infty$	$\|\bar{u}\|_{1,\infty} < \infty$
$p_h(\bar{u}) - p_h(P_k\bar{u})$	$\mathsf{V}_0^T\dot{\bar{u}} < \infty$	$\mathsf{V}_0^T\dot{\bar{u}} < \infty$	0	$\mathsf{V}_0^T\dot{\bar{u}} < \infty$

Tabelle 6.1: Zusammenfassung der Voraussetzungen für die Konvergenz der Hilfsresultate.

Nun widmen wir uns der Approximation der optimalen Steuerung $\bar{u}$. In den Tabellen 6.2 und 6.3 sind die Ergebnisse mit den jeweiligen Voraussetzungen für die beiden Diskretisierungsmethoden getrennt zusammengefasst. Es wird deutlich, welchen Stellenwert die Voraussetzung der beschränkten Variation an die optimale Steuerung besitzt. Bei der näherungsweisen Lösung der Systemgleichung kann sie unter Umständen außen vor gelassen werden, bei der Betrachtung von diskreten Steuerungen ist sie allerdings unverzichtbar.

Resultat	**FEM konstant**		**FEM linear**	
$\|\bar{u} - u_h\|_2$	$\|\bar{u}\|_\infty < \infty$	h	$\mathsf{V}_0^T\dot{\bar{u}} < \infty$	$h^{\frac{3}{2}}$
$\|\bar{u} - \tilde{u}\|_2$	$\mathsf{V}_0^T\dot{\bar{u}} < \infty$	h^2		
$\|\bar{u} - \tilde{u}\|_\infty$	$\mathsf{V}_0^T\dot{\bar{u}} < \infty$	h^2		

Tabelle 6.2: Zusammenfassung der Voraussetzungen für die Hauptergebnisse bei den Finite Elementen.

Resultat	**Differenzenverf.**		**Alternatives Diff.**	
$\|\bar{u} - u_h\|_h$	$\mathsf{V}_0^T \dot{\bar{u}} < \infty$	h^2	$\mathsf{V}_0^T \dot{\bar{u}} < \infty$	$h^{\frac{3}{2}}$
$\|\bar{u} - \tilde{u}\|_2$	$\mathsf{V}_0^T \dot{\bar{u}} < \infty$	h^2		
$\|\bar{u} - \tilde{u}\|_\infty$	$\mathsf{V}_0^T \dot{\bar{u}} < \infty$	h^2		

Tabelle 6.3: Zusammenfassung der Voraussetzungen für die Hauptergebnisse bei den Finiten Differenzen.

Bei den vorgestellten Methoden war es möglich die Konvergenzgeschwindigkeit für den Abstand von exakter und diskreter Lösung zu bestimmen. Darüber hinaus erhielten wir durch Definition einer neuen, für das Ausgangsproblem zulässigen Steuerung einerseits punktweise Abschätzungen und andererseits unter Umständen eine Erhöhung der Konvergenzordnung. Bei den Finiten Elementen fällt der Verlust der quadratischen Konvergenzordnung mit Erhöhung der Approximationsgüte von stückweise konstant zu stückweise linear auf. Ebenso ist das Differenzenverfahren aus Abschnitt 5.2 mit der Approximation der rechten Seite in der Systemgleichung durch die konkreten Werte an den Stützstellen seiner Alternative aus Abschnitt 5.7 mit Verwendung eines Integralmittels überlegen. Die Erkenntnisse über den Zusammenhang zwischen Diskretisierung der Systemgleichung, dem Aufstellen des diskreten Steuerungsproblems und die Konvergenzordnung für die Steuerungen, die wir aus den Diskretisierungsmethoden ableiten konnten, verwenden wir nun für allgemeine Konvergenzresultate. Die konkrete Art und Weise der Diskretisierung ujnd ihre Eigenschaften behandeln wir dabei nicht.

Ausgangspunkt ist die Systemgleichung (1.4) in der Form

$$Lz = \ell(\mathcal{T}u),$$

mit dem auf $W^1_{2,0}[0,T,\mathbb{R}^n]$ linearen, symmetrischen und elliptischen Differentialoperator $Lz = -\ddot{z} + Az$ und einem linearen stetigen Funktional $\ell \in L_2[0,T,\mathbb{R}^n]$. Neben einer eindeutig bestimmten Lösung z erhalten wir zusammen mit einer Diskretisierungsvorschrift auch eine diskrete Lösung. Dabei ist zu beachten, dass wir auf beiden Seiten durchaus verschiedene Diskretisierungsmethoden anwenden können. Setzen wir eine stabile Diskretisierung voraus, so erhalten wir eine eindeutig bestimmte diskrete Lösung z_h im endlich-dimensionalen Raum V_h, die wir mit dem Vektor β identifizieren. Die Zuordnungen $u \mapsto z$ und $u \mapsto z_h$ drücken wir daher mit Hilfe der Operatoren $\mathcal{S} \circ \mathcal{T}$ und $\mathcal{S}_h \circ \mathcal{T}$ aus, wobei $\mathcal{T}u = Bu + e$ ist. Neben der Diskretisierung der Systemgleichung legen wir eine Approximation der rechten Seite y fest, die Funktionen mögen aus dem endlich dimensionalem Raum U_h stammen und wir charakterisieren sie durch den Vektor ψ. Dann ist $\mathcal{S}_h\big|_{U_h}$ ein finiter Operator und wir erhalten durch Anwendung der Diskretisierungsvorschriften eine endliches Gleichungssystem

$$\beta = \mathfrak{N}\mathfrak{C}\psi.$$

An dieser Stelle ist eine Fallunterscheidung notwendig, denn die Matrix $\mathfrak{C}$ hängt wie schon bemerkt von der Diskretisierung der rechten Seite ab. Aus der Definition der

Diskretisierung der rechten Seite gewinnen wir ein Skalarprodukt, welches wir durch eine symmetrische und positiv definite Matrix ausdrücken können. Wir bezeichnen diese Matrix dann ebenfalls mit $\mathfrak{C}$. Definieren wir nun das Skalarprodukt und die zugehörige Norm

$$\langle \cdot, \cdot \rangle_{\mathfrak{C}} = \langle \mathfrak{C} \cdot, \cdot \rangle = \langle \cdot, \mathfrak{C} \cdot \rangle, \qquad \| \cdot \|_{\mathfrak{C}} = \sqrt{\langle \cdot, \cdot \rangle_{\mathfrak{C}}}.$$

Setzen wir für die Diskretisierung die Erhaltung der Symmetrieeigenschaft von L voraus, so ist die Matrix $\mathfrak{N}$ symmetrisch. Daraus folgt die Selbstadjungiertheit bzw. Symmetrie der Abbildung $\mathcal{S}_h$ mit $\mathcal{S}_h \psi = \beta$ bezüglich des Skalarprodukts $\langle \cdot, \cdot \rangle_{\mathfrak{C}}$. Genauer gilt

$$\langle \psi_1, \beta_2 \rangle_{\mathfrak{C}} = \langle \mathfrak{C} \psi_1, \mathfrak{N} \mathfrak{C} \psi_2 \rangle = \langle (\mathfrak{N} \mathfrak{C} \psi_1), \mathfrak{C} \psi_2 \rangle = \langle \beta_1, \psi_2 \rangle_{\mathfrak{C}}.$$

Daher folgt $\mathcal{S}_h = \mathcal{S}_h^*$, wobei $\mathcal{S}_h^*$ der diskrete adjungierte Operator, also Diskretisierung der adjungierten Gleichung (2.10) ist. Ferner erhalten wir eine Diskretisierung des Problems (StP) durch

$$(\mathrm{StP})_h^{\mathfrak{C}} \qquad \min_{u_h} \frac{1}{2} \| \mathcal{S}_h (B u_h + e) - z_d \|_{\mathfrak{C}}^2 + \frac{\nu}{2} \| u_h \|_{\mathfrak{C}}^2, \qquad u_h \in U_h^{ad} = U_h \cap U^{ad}.$$

Wegen der Adjungiertheit von $\mathcal{S}_h$ und $\mathcal{S}_h^*$ erhalten wir als notwendige Optimalitätsbedingung für die Lösung $u_h \in U_h^{ad}$

$$\langle B^{\mathsf{T}} \mathcal{S}_h^* (\mathcal{S}_h \mathcal{T} u_h - z_d) + \nu u_h, \xi_h - u_h \rangle_{\mathfrak{C}} \geq 0 \qquad \forall \xi_h \in U_h^{ad}. \tag{6.1}$$

Bei einer anderen Wahl der Norm im diskreten Problem $(\mathrm{StP})_h^{\mathfrak{C}}$ wäre eine Ausnutzung dieser Beziehung zwischen den diskreten Operatoren und die damit verbundene Aufstellung der Optimalitätsbedingungen in dieser Form nicht möglich. Daher besteht ein wichtiger Zusammenhang zwischen der Diskretisierung der Systemgleichung und dem daraus resultierenden diskreten Steuerungsproblem.

Beispiel: Bei den Finiten Elementen mit stückweise konstante Approximation der Steuerung ist $\mathfrak{C}$ die Einheitsmatrix und im Fall stückweise linearer Approximation besitzt die Matrix $\mathfrak{C}$ eine symmetrische tridiagonale Struktur mit den Vektoren $\frac{h}{6}(1, \ldots, 1)$ auf der Nebendiagonale und $\frac{h}{6}(2, 4, \ldots, 4, 2)$ auf der Hauptdiagonale. Wir verwenden in beiden Fällen die Norm $\| \cdot \|_2$ zur Definition des diskreten Steuerungsproblems.

Für die Methode der Finiten Differenzen ergibt sich wieder $\mathfrak{C} = I_{(N+1)n}$ und bei der Betrachtung des alternativen Differenzenverfahrens definiert sich $\mathfrak{C}$ als Tridiagonalmatrix durch die Vektoren $\frac{h}{8}(1, \ldots, 1)$ auf den Nebendiagonalen und $\frac{h}{8}(3, 6, \ldots, 6, 3)$ auf der Hauptdiagonale. $\Diamond$

Für die Untersuchung der Steuerungen beginnen wir mit dem Abstand zwischen der Interpolation der exakten Lösung im Raum U_h und der diskreten Lösung u_h. Wie in Lemma 2.4.3 gezeigt, gilt die Optimalitätsbedingung für das kontinuierliche Problem auch punktweise, d.h.

$$\big(B^{\mathsf{T}} \bar{p}(t) + \nu \bar{u}(t)\big)^{\mathsf{T}} \big(u - \bar{u}(t)\big) \geq 0 \qquad \forall' t \in [0, T], \forall u \in U$$

mit $U = \{u \in R^m : a \le u \le b\}$. Wir betrachten diese Ungleichung mit $u = u_h(t_j)$ an allen Stellen t_j, an denen $\bar{u}$ mit ihrer Interpolation $D_h\bar{u} \in U_h$ übereinstimmt. Ferner versehen wir jede Ungleichung mit einem Faktor und addieren über alle Punkte. Daraus folgt

$$(B^\mathsf{T}\bar{\zeta} + \nu\bar{\alpha})^\mathsf{T} D(\alpha - \bar{\alpha}) \ge 0,$$

mit einer Diagonalmatrix D und den Vektoren $\bar{\zeta}$, $\bar{\alpha}$ und α, die die jeweiligen Funktionen darstellen durch $\bar{\zeta}_j = \bar{p}(t_j)$ für alle Indizes j sowie $\bar{\alpha}$ und α analog. Aus der Optimalitätsbedingung für das diskrete Problem (6.1) erhalten wir mit $\xi_h = D_h\bar{u}$ die Ungleichung

$$(B^\mathsf{T}\zeta + \nu\alpha)^\mathsf{T}\mathfrak{C}(\bar{\alpha} - \alpha) \ge 0.$$

Addieren wir beide, so folgt

$$\nu\|\alpha - \bar{\alpha}\|_\mathfrak{C}^2 \le (B^\mathsf{T}(\bar{\zeta} - \zeta))^\mathsf{T}\mathfrak{C}(\alpha - \bar{\alpha}) - (B^\mathsf{T}\bar{\zeta} + \nu\bar{\alpha})^\mathsf{T}(D - \mathfrak{C})(\alpha - \bar{\alpha}).$$

Wir betrachten nun wieder die dahinter stehenden Funktionen und spalten den ersten Summanden auf

$$\begin{aligned}\nu\|u_h - D_h\bar{u}\|_\mathfrak{C}^2 \le\ & \langle(\mathcal{S}^*\mathcal{S} - \mathcal{S}_h^*\mathcal{S}_h)B\bar{u}, B(u_h - D_h\bar{u})\rangle_\mathfrak{C} \\ & + \langle\mathcal{S}_h^*\mathcal{S}_h B(\bar{u} - D_h\bar{u}), B(u_h - D_h\bar{u})\rangle_\mathfrak{C} \\ & + \langle\mathcal{S}_h^*\mathcal{S}_h B(D_h\bar{u} - u_h), B(u_h - D_h\bar{u})\rangle_\mathfrak{C} \\ & - \langle(B^\mathsf{T}\bar{\zeta} + \nu\bar{\alpha})^\mathsf{T}(D - \mathfrak{C}), \mathfrak{C}^{-1}(\alpha - \bar{\alpha})\rangle_\mathfrak{C}.\end{aligned}$$

Der dritte Term nimmt im weiteren Verlauf eine Sonderrolle ein. Wir nutzen die Adjungiertheit der diskreten Operatoren bezüglich des Skalarprodukts aus und schließen

$$\langle\mathcal{S}_h^*\mathcal{S}_h B(D_h\bar{u} - u_h), B(u_h - D_h\bar{u})\rangle_\mathfrak{C} = \langle\mathcal{S}_h B(D_h\bar{u} - u_h), \mathcal{S}_h B(u_h - D_h\bar{u})\rangle_\mathfrak{C} \le 0.$$

Dann folgt mit der Cauchy-Schwarz-Ungleichung und Division durch $\|u_h - D_h\bar{u}\|_\mathfrak{C}$

$$\begin{aligned}\nu\|u_h - D_h\bar{u}\|_\mathfrak{C} \le\ & \|B\|\,\|(\mathcal{S}^*\mathcal{S} - \mathcal{S}_h^*\mathcal{S}_h)B\bar{u}\|_\mathfrak{C} \\ & + \|B\|\,\|\mathcal{S}_h^*\mathcal{S}_h B(\bar{u} - D_h\bar{u})\|_\mathfrak{C} + \|\mathfrak{C}^{-1}\|\,|(B^\mathsf{T}\zeta + \nu\alpha)^\mathsf{T}(D - \mathfrak{C})|.\end{aligned}$$

Somit erhalten wir mit der Konvergenzordnung P_k, d.h. $\|(\mathcal{S}^*\mathcal{S} - \mathcal{S}_h^*\mathcal{S}_h)y\| \le c\,h^{P_k}$, und mit $\|\bar{u} - D_h\bar{u}\|_\mathfrak{C} \le c\,h^{P_i}$ die Abschätzung

$$\nu\|u_h - D_h\bar{u}\|_\mathfrak{C} \le c\,h^{P_k} + \|\mathcal{S}_h^*\|\,\|\mathcal{S}_h\|\,h^{P_i} + \|\mathfrak{C}^{-1}\|\,|(B^\mathsf{T}\bar{\zeta} + \nu\bar{\alpha})^\mathsf{T}(D - \mathfrak{C})|.$$

Für den Gesamtfehler bei der Diskretisierung folgt dann

$$\|\bar{u} - u_h\|_\mathfrak{C} \le \|\bar{u} - D_h\bar{u}\|_\mathfrak{C} + \|D_h\bar{u} - u_h\|_\mathfrak{C} \le c\,h^{P_i} + c\,h^{P_k} + \|\mathfrak{C}^{-1}\|\,|(B^\mathsf{T}\bar{\zeta} + \nu\bar{\alpha})^\mathsf{T}(D - \mathfrak{C})|.$$

Demnach erhalten wir für jede auf U_h zu $\|\cdot\|_\mathfrak{C}$ äquivalenten Norm mit von h unabhängigen Abschätzungskonstanten die Fehlerschranke

$$\|\bar{u} - u_h\| \le c\,h^{P_i} + c\,h^{P_k} + c\,|(B^\mathsf{T}\bar{\zeta} + \nu\bar{\alpha})^\mathsf{T}(D - \mathfrak{C})|.$$

In Abhängigkeit von der konkreten Gestalt der Matrix $D-\mathfrak{C}$ ist es möglich, den letzten Summanden in Abhängigkeit der Schrittweite abzuschätzen, also bleibt

$$\|\bar{u} - u_h\| \leq c\, h^{\min\{P_i, P_k, P_{D-\mathfrak{C}}\}}.$$

Beispiel: Bei den Finiten Elementen mit stückweise konstanter Approximation gilt $\|\bar{u} - D_h\bar{u}\|_2 = \|\bar{u} - P_0\bar{u}\|_2 \leq c\, h$ und $\mathfrak{C}$ bekanntlich die Einheitsmatrix, daher folgt

$$\|\bar{u} - u_h\|_2 \leq c\, h.$$

Erhöhen wir auf stückweise lineare Approximation, so gilt $\|\bar{u} - D_h\bar{u}\|_2 \leq c\, h^{\frac{3}{2}}$. Weiterhin ist $\mathfrak{C}$ keine Diagonalmatrix, so dass der letzte Summand nicht verschwindet. Allerdings ist es möglich

$$|(B^{\mathsf{T}}\bar{\zeta} + \nu\bar{\alpha})^{\mathsf{T}}(D - \mathfrak{C})| \leq c\, h^{\frac{3}{2}}$$

zun zeigen. Insgesamt ergibt sich somit

$$\|\bar{u} - u_h\|_2 \leq c\, h^{\frac{3}{2}}.$$

Bei den Finiten Differenzen ergibt sich auf die gleiche Weise wie eben

$$\|\bar{u} - u_h\|_2 \leq c\, h^{\frac{3}{2}}.$$

Bei der alternativen Vorgehensweise ist $\mathfrak{C}$ wiederum keine Diagonalmatrix. Allerdings gelingt wie eben die Herleitung der Abschätzung

$$|(B^{\mathsf{T}}\bar{\zeta} + \nu\bar{\alpha})^{\mathsf{T}}(D - \mathfrak{C})| \leq c\, h^{\frac{3}{2}}$$

und somit ebenfalls Konvergenz der Ordnung $\frac{3}{2}$. ◊

Wie im Verlauf der bisherigen Zusammenfassung deutlich wurde, erreichen wir die Konvergenzordnung $\frac{3}{2}$ im quadratischen Mittel. Für punktweise Abschätzungen und eine eventuelle Erhöhung der Konvergenzordnung ist dieser Zugang nicht geeignet. Um dennoch Resultate dieser Art abzuleiten, definieren wir zunächst die für das Ausgangsproblem zulässige Steuerung

$$\tilde{u} = \Pi_{[\mathrm{a,b}]}\left(-\frac{1}{\nu}B^{\mathsf{T}}\mathcal{S}_h^*(\mathcal{S}_h\mathcal{T}u_h - z_d)\right).$$

Betrachten wir nun den Abstand $\|\bar{u} - \tilde{u}\|$ in einer beliebigen Norm, so nutzen wir im ersten Schritt die Lipschitz-Stetigkeit des Projektionsoperators $\Pi_{[\mathrm{a,b}]}$ aus und erhalten wegen der Linearität der Operatoren

$$\|\bar{u} - \tilde{u}\| \leq \frac{\|B\|}{\nu}\|\mathcal{S}^*(\mathcal{S}\mathcal{T}\bar{u} - z_d) - \mathcal{S}_h^*(\mathcal{S}_h\mathcal{T}u_h - z_d)\| = \frac{\|B\|}{\nu}\|\mathcal{S}^*\mathcal{S}B\bar{u} - \mathcal{S}_h^*\mathcal{S}_hBu_h\|.$$

Den letzten Ausdruck spalten wir wieder auf und verwenden die bisherigen Ergebnisse. Es ist auf Grund der Konvergenz bei der Lösung der Systemgleichung

$$\|(\mathcal{S}^*\mathcal{S} - \mathcal{S}_h^*\mathcal{S}_h)B\bar{u}\|_\infty \leq c\, h^{P_k}.$$

Weiter erhalten wir

$$\|\mathcal{S}_h^*\mathcal{S}_h B(\bar{u} - D_h\bar{u})\|_\infty \leq \|\mathcal{S}_h^*\|\,\|\mathcal{S}_h\|\,\|B\|\,h^{P_i},$$

sowie als letzten Schritt

$$\|\mathcal{S}_h^*\mathcal{S}_h B(D_h\bar{u} - u_h)\|_\infty \leq \|\mathcal{S}_h^*\|\,\|\mathcal{S}_h\|\,\|B\|\,h^{\min\{P_k,P_i,P_{D-\mathfrak{C}}\}}.$$

Somit ergibt auch insgesamt die Abschätzung

$$\|\bar{u} - \tilde{u}\|_\infty \leq c\, h^{\min\{P_k,P_i,P_{D-\mathfrak{C}}\}}.$$

Beispiel: Als Beispiele betrachten wir wieder die Diskretisierung von (StP) mit Hilfe der vorgestellten Methoden. Besondere Beachtung erhält dabei die Tatsache, dass sich bei der Untersuchung von $\|\mathcal{S}_h(\bar{u} - D_h\bar{u})\|$ die Konvergenzordnung im Vergleich zum Abstand $\|\bar{u} - D_h\bar{u}\|$ erhöht. Wir erhalten in allen betrachteten Fällen $P_i = 2$. Die Konvergenz bei der Diskretisierung der Systemgleichung ist für jede Methode einzeln herzuleiten. Wir zeigten jeweils $P_k = 2$ in der L_∞-Norm, woraus auch alle anderen Normen folgen. Schließlich bleibt die Bestimmung von $P_{D-\mathfrak{C}}$. In unseren Fällen ist es möglich die Matrix D so zu wählen, dass die Differenz $D - \mathfrak{C}$ die

$$\begin{pmatrix} -1 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & & \ddots & & & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -1 \end{pmatrix}$$

annimmt. An dieser Stelle verweisen wir auf Lemma 3.2.9 und halten für unsere Betrachtungen

$$|(B^\mathsf{T}\bar{\zeta} + \nu\bar{\alpha})^\mathsf{T}(D - \mathfrak{C})| \leq c\, h^{\frac{3}{2}}$$

fest. Im Fall der Finiten Elemente mit stückweise konstante Approximation der Steuerung bzw. der Finiten Differenzen ist sogar $D - \mathfrak{C} = 0$, so dass dieser Ausdruck wegfällt. Damit folgt

$$\|\bar{u} - \tilde{u}\|_\infty \leq c\, h^2.$$

Approximieren wir bei den Finiten Elementen die Steuerung stückweise linear, oder verändern wir die Methode der Finiten Differenzen in der oben erwähnten Weise, so geht diese Struktur verloren und wir erhalten

$$\|\bar{u} - \tilde{u}\|_\infty \leq c\, h^{\frac{3}{2}}.$$

Eine Erhöhung der Konvergenzgeschwindigkeit ist also nicht in jedem Fall möglich, allerdings erreichen wir mit Hilfe der Funktion $\tilde{u}$ auch punktweise Konvergenz der Ordnung $\frac{3}{2}$. $\Diamond$

Kapitel 7

Anwendung

7.1 Problemstellung

Zum Abschluss spannen wir den Bogen zurück zum Ausgangsbespiel in Abschnitt 1.1. Das dort vorgestellte Problem der stationären Wärmeverteilung in einem Stab der Länge T lösen wir nun numerisch. Das Ziel ist eine Energie optimale Aufheizung eines Stabes, unter Berücksichtigung eines zusätzlichen Wärmeaustrags und konstanter Temperatur an beiden Enden. Zur Vereinfachung, aber ohne Einschränkung der Allgemeinheit, wählen wir $T = 1$. Die gewünschte Temperaturverteilung ist eine symmetrische Kurve in der Form einer Glocke und den äußeren Einfluss modellieren wir mit Hilfe der Sinus-Funktion

$$z_d(t) = e^{\frac{1}{4}}(e^{(t-t^2)} - 1), \qquad e(t) = -\frac{1}{4}\sin(\pi t).$$

Eine graphische Darstellung der beiden Funktionen gaben wir bereits in Abschnitt 1.1. Zusammen dem Operator $\mathcal{S}$, der jeder Steuerung die Lösung von

$$-\ddot{z}(t) = u(t) + e(t), \qquad z(0) = z(1) = 0$$

zuordnet und der zulässigen Menge

$$U^{ad} = \Big\{u \in L_2[0, T, \mathbb{R}] : -1 \leq u(t) \leq 1, \, \forall' t \in [0, 1]\Big\},$$

stellt sich unser Steuerungsproblem in folgender Form dar

$$\text{(StP)} \qquad \min_u \frac{1}{2}\int_0^1 |(\mathcal{S}u)(t) - z_d(t)|^2 + \nu|u(t)|^2 \, dt, \qquad u \in U^{ad}.$$

7.2 Ergebnisse

Wir gehen bei der Präsentation der Ergebnisse in der gleichen Reihenfolge wie im theoretischen Teil vor. Zunächst betrachten wir die Finite Elemente Methode mit stückweise

konstanter Approximation der Steuerung und der Wahl von 0.05 für die Schrittweite. Abbildung 7.1 zeigt die diskrete Lösung u_h, sowie den zugehörigen diskreten Zustand und den diskreten adjungierten Zustand. In Abbildung 7.2 sind anschließend die Funktionen $-\frac{1}{\nu}p_h(u_h)$ (unterbrochene Linie) und $\tilde{u} = \Pi_{[a,b]}\big(-\frac{1}{\nu}p_h(u_h)\big)$ (durchgehende Linie) eingezeichnet.

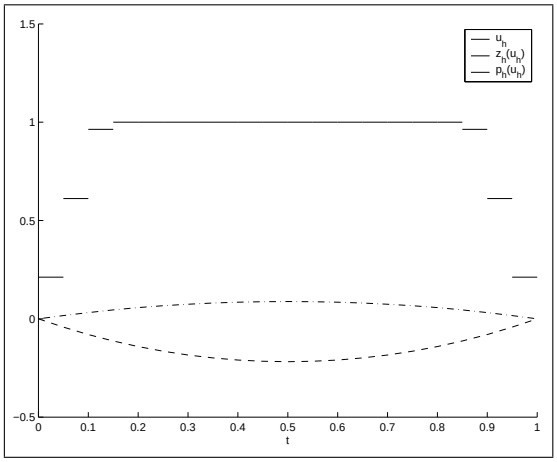

Abbildung 7.1: Die Funktionen u_h, $z_h(u_h)$ sowie $p_h(u_h)$ für die Finite Elemente mit stückweise konstanter Approximation

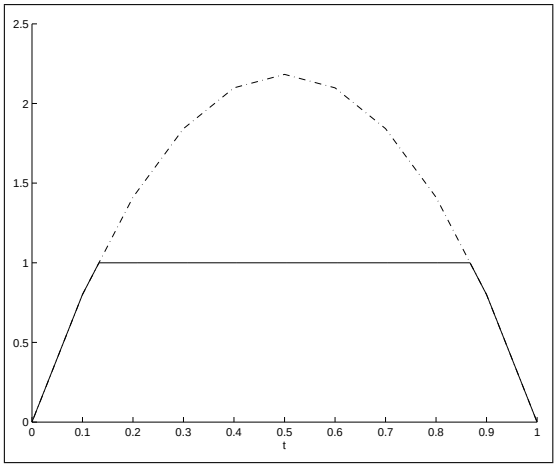

Abbildung 7.2: Die Funktionen $-\frac{1}{\nu}p_h(u_h)$ und $\tilde{u}$ für die Finite Elemente mit stückweise konstanter Approximation

Verbessern wir die Approximation der Steuerung und betrachten die Ergebnisse für 0.1 als Schrittweite. Wieder zeigt Abbildung 7.3 zunächst die diskrete Lösung und die zugehörigen Zustände. In Abbildung 7.4 sehen wir anschließend die im letzten Schritt entstehenden Funktionen.

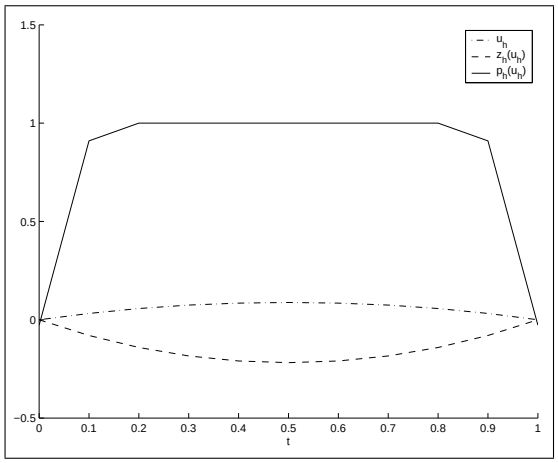

Abbildung 7.3: Die Funktionen u_h, $z_h(u_h)$ sowie $p_h(u_h)$ für die Finite Elemente mit stückweise linearer Approximation

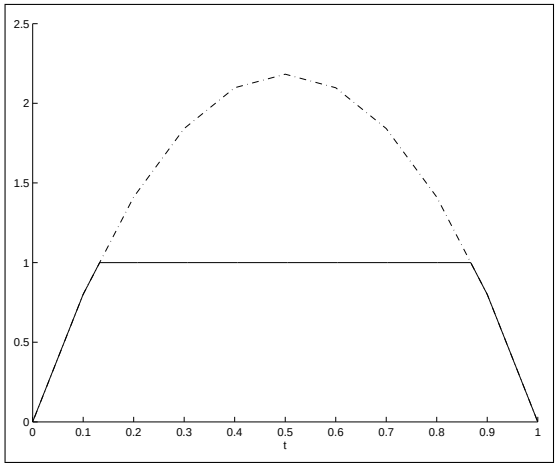

Abbildung 7.4: Die Funktionen $-\frac{1}{\nu}p_h(u_h)$ und $\tilde{u}$ für die Finite Elemente mit stückweise linearer Approximation

Beim Differenzenverfahren erhalten wir schließlich folgende Lösungen für 0.1 als Schrittweite. Wie eben enthält Abbildung 7.5 diem Darstellung der diskreten Lösung sowie der zugehörigen Zustände. Abbildung 7.6 zeigt im Anschluss die Transformation des diskreten adjungierten Zustand und deren Projektion auf das zulässige Intervall.

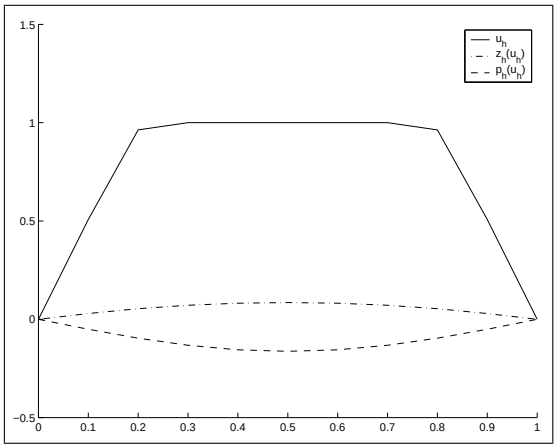

Abbildung 7.5: Die Funktionen u_h, $z_h(u_h)$ sowie $p_h(u_h)$ für die Finiten Differenzen

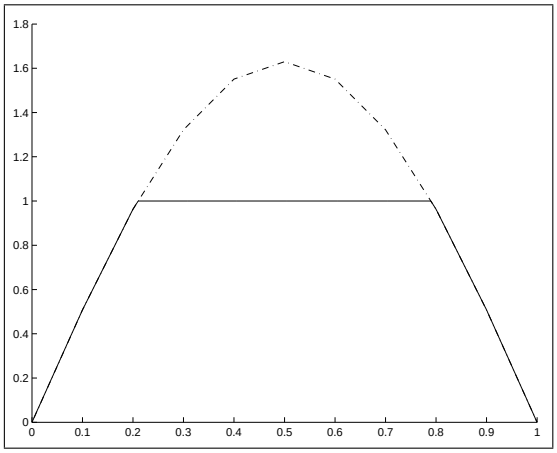

Abbildung 7.6: Die Funktionen $-\frac{1}{\nu}p_h(u_h)$ und $\tilde{u}$ für die Finiten Differenzen

Zum Abschluss geben wir die Ergebnisse beim alternativen Differenzenverfahren ebenfalls mit der Schrittweite 0.1 an.

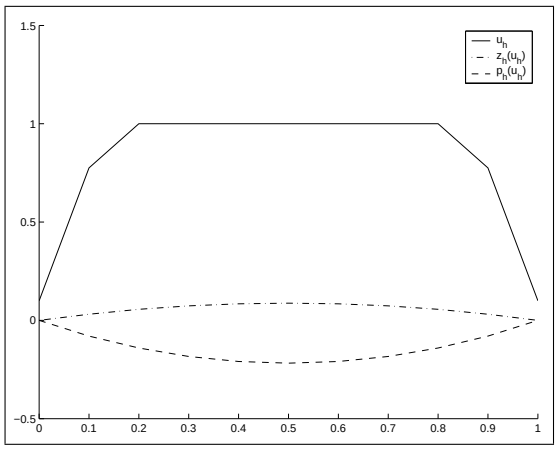

Abbildung 7.7: Die Funktionen u_h, $z_h(u_h)$ sowie $p_h(u_h)$ für die Alternative der Finiten Differenzen

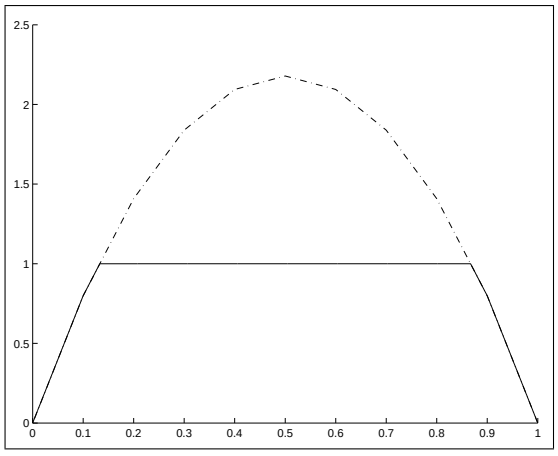

Abbildung 7.8: Die Funktionen $-\frac{1}{\nu}p_h(u_h)$ und $\tilde{u}$ für die Alternative der Finiten Differenzen

A

Anhang

A.1 Das formale *Lagrange-Prinzip*

In Abschnitt 2.3 führten wir die adjungierte Gleichung zur Systemgleichung per Definition ein, da der Lösungsoperator $\mathcal{S}$ selbstadjungiert ist. Eine weitere Möglichkeit die adjungierte Gleichung herzuleiten, ist der Weg über die *Lagrange-Funktion*. Ausgehend von diesem Ansatz fällt es uns nicht schwer, das bekannte *Pontryagin'sche Maximum-Prinzip* aufzustellen. Dieses Vorgehen ist auch bei allgemeineren, nicht-linearen Differentialoperatoren anwendbar und wird in der anwendungsorientierten Literatur sehr häufig zitiert (vgl. z.B. Feichtinger/Hartl [12] (1986)), aber auch in theoretischen Arbeiten verwendet (vgl. Dontchev, Hager, Veliov [11] (2000)). Aus diesem Grund stellen wir es an diesem Punkt vor.

Wir verwandeln zunächst die Zustandsgleichung als Differentialgleichung 2. Ordnung in ein System von Gleichungen 1. Ordnung. Dazu seien $z_1, z_2 \in W_2^1[0,T,\mathbb{R}^n]$ mit $\dot{z}_1 + Az_2 = y$ und $\dot{z}_2 = -z_1$. So ergibt sich mit $\hat{z} = (z_1^\mathsf{T}, z_2^\mathsf{T})^\mathsf{T} \in W_2^1[0,T,\mathbb{R}^{2n}]$:

$$\dot{\hat{z}} + \begin{pmatrix} 0 & A \\ I & 0 \end{pmatrix} \hat{z} = \begin{pmatrix} y \\ 0 \end{pmatrix} \Longleftrightarrow \dot{\hat{z}} + \hat{A}\hat{z} = \hat{y}$$

und

$$z_2(0) = z_2(T) = 0.$$

Da das eine äquivalente Schreibweise zur Systemgleichung (1.4) ist, wissen wir schon um die eindeutige Lösbarkeit. Daraus folgern wir ein anderes zu minimierendes Funktional, nämlich

$$\min_{\hat{z},u} \hat{J}(\hat{z},u) = \frac{1}{2}\|z_2(u) - z_d\|_2^2 + \frac{\nu}{2}\|u\|_2^2.$$

Für die Herleitung der adjungierten Gleichung definieren wir zunächst die Räume

$$X = W_2^1[0,T,\mathbb{R}^{2n}] \times L_2[0,T,\mathbb{R}^m], \qquad Y = L_2[0,T,\mathbb{R}^{2n}] \times \mathbb{R}^n \times \mathbb{R}^n$$

sowie mit einem $p^* \in Y'$ die *Lagrange-Funktion* $\mathcal{L} : X \times Y' \to \mathbb{R}$ durch

$$\mathcal{L}(\hat{z}, u, p^*) = \hat{J}(\hat{z}, u) + p^* \begin{pmatrix} \hat{y} - \hat{A}\hat{z} - \dot{\hat{z}} \\ z_2(0) \\ z_2(T) \end{pmatrix}.$$

Die Bezeichnung Y' für den Dualraum von Y führten wir bereits in Abschnitt 1.2 ein. Da Y ein Hilbert-Raum ist, besitzt jedes Funktional auf Y eine Darstellung als Skalarprodukt mit einem eindeutig identifizierten Element aus Y, also gilt mit $(p, \mu, \kappa) \in Y$

$$p^*(v, w, q) = \int_0^T p(t)^\mathsf{T} v(t)\, dt + \mu^\mathsf{T} w + \kappa^\mathsf{T} q \qquad \forall (v, w, q) \in Y.$$

Demnach ist

$$\mathcal{L}(\hat{z}, u, p^*) = \hat{J}(\hat{z}, u) + \int_0^T p(t)^\mathsf{T} (\hat{y}(t) - \hat{A}\hat{z}(t) - \dot{\hat{z}}(t))\, dt + \mu^\mathsf{T} z_2(0) + \kappa^\mathsf{T} z_2(T).$$

Für ein beliebiges Element $(\hat{z}^0, u^0) \in X$ berechnen wir die Fréchet-Ableitung der Lagrange-Funktion an dieser Stelle. Nach der Regel für die Differentation von Abbildungen zwischen Banach-Räumen ist

$$\mathcal{L}'(\hat{z}^0, u^0, p^*)(\hat{z}, u) = \mathcal{L}_{\hat{z}}(\hat{z}^0, u^0, p^*)(\hat{z}) + \mathcal{L}_u(\hat{z}^0, u^0, p^*)(u),$$

wobei $\mathcal{L}_{\hat{z}}$ bzw. $\mathcal{L}_u$ die partielle Ableitung nach der jeweiligen Komponente bezeichnet. Die beiden Summanden ergeben sich zu

$$\mathcal{L}_{\hat{z}}(\hat{z}^0, u^0, p^*)(\hat{z}) = \hat{J}_{\hat{z}}(\hat{z}^0, u^0)(\hat{z}) + \int_0^T p(t)^\mathsf{T} (-\hat{A}\hat{z}(t) - \dot{\hat{z}}(t))\, dt + \mu^\mathsf{T} z_2(0) + \kappa^\mathsf{T} z_2(T)$$

und

$$\mathcal{L}_u(\hat{z}^0, u^0, p^*)(u) = \hat{J}_u(\hat{z}^0, u^0)(u) + \int_0^T p(t)^\mathsf{T} \begin{pmatrix} Bu(t) \\ 0 \end{pmatrix} dt.$$

Damit ergibt sich für die Fréchet-Ableitung der Lagrange-Funktion der Ausdruck

$$\begin{aligned} \mathcal{L}'(\hat{z}^0, u^0, p^*)(\hat{z}, u) = &\int_0^T (z_2^0(t) - z_d(t))^\mathsf{T} z_2(t) - p(t)^\mathsf{T} (\hat{A}\hat{z}(t) + \dot{\hat{z}}(t))\, dt \\ &+ \mu^\mathsf{T} z_2(0) + \kappa^\mathsf{T} z_2(T) \\ &+ \nu \int_0^T u^0(t)^\mathsf{T} u(t) + p(t)^\mathsf{T} \begin{pmatrix} Bu(t) \\ 0 \end{pmatrix} dt \qquad \forall (\hat{z}, u) \in X. \end{aligned}$$

Für die Optimalität des Elements $(\hat{z}^0, u^0) \in X$ bezüglich der Funktion $\mathcal{L}$ ist

$$\mathcal{L}'(\hat{z}^0, u^0, p^*)(\hat{z}, u) = 0 \qquad \forall (\hat{z}, u) \in X$$

notwendig. Da die vorherige Gleichung für alle $(\hat{z}, u) \in X$ gilt, nehmen wir $u(t) = 0$ für fast alle $t \in [0, T]$ an und schließen auf

$$\int_0^T (z_2^0(t) - z_d(t))^\mathsf{T} z_2(t) - p(t)^\mathsf{T}(\hat{A}\hat{z}(t) + \dot{\hat{z}}(t))\,dt + \mu^\mathsf{T} z_2(0) + \kappa^\mathsf{T} z_2(T) = 0 \qquad \forall \hat{z} \in W_2^1[0, T, \mathbb{R}^{2n}].$$

Nun sind die Voraussetzungen des Variationslemmas 2.1.1 erfüllt, denn es gilt

$$\int_0^T \begin{pmatrix} 0 \\ z_2^0(t) - z_d(t) \end{pmatrix}^\mathsf{T} \hat{z}(t) - p(t)^\mathsf{T}\hat{A}\hat{z}(t) - p(t)^\mathsf{T}\dot{\hat{z}}(t)\,dt = 0 \qquad \forall \hat{z} \in W_{2,0}^1[0, T, \mathbb{R}^{2n}].$$

Wir folgern daher $p \in W_2^1[0, T, \mathbb{R}^{2n}]$ und für fast alle $t \in [0, T]$

$$-\dot{p}(t) = \begin{pmatrix} 0 \\ z_2^0(t) - z_d(t) \end{pmatrix} - \hat{A}^\mathsf{T} p(t) = \begin{pmatrix} 0 \\ z_2^0(t) - z_d(t) \end{pmatrix} - \begin{pmatrix} 0 & I \\ A^\mathsf{T} & 0 \end{pmatrix} p(t)$$

sowie mit $p = (p_1^\mathsf{T}, p_2^\mathsf{T})^\mathsf{T}$

$$\begin{aligned} -\dot{p}_1 &= -p_2 \\ -\dot{p}_2 &= z_2^0 - z_d - p_1^\mathsf{T} A. \end{aligned}$$

Damit gelangen wir durch die vorausgesetzte Symmetrie von A zu

$$-\ddot{p}_1(t) + Ap_1(t) = z_2^0(t) - z_d(t) \qquad \forall' t \in [0, T].$$

Die Randbedingungen leiten wir wieder mit Hilfe des Hauptsatzes der Differential- und Integralrechnung her. Es gilt laut der Optimalitätsbedingung mit speziell $u \equiv 0$

$$\begin{aligned} 0 &= \int_0^T (\hat{z}_2^0(t) - z_d(t))^\mathsf{T} \hat{z}_2(t) - p(t)^\mathsf{T}(\dot{\hat{z}}(t) + \hat{A}\hat{z}(t))\,dt \\ &= -\int_0^T p(t)^\mathsf{T}\dot{\hat{z}}(t)\,dt + \int_0^T \left(\begin{pmatrix} 0 \\ z_2^0(t) - z_d(t) \end{pmatrix} - \hat{A}^\mathsf{T} p(t) \right)^\mathsf{T} \hat{z}(t)\,dt \\ &= -\int_0^T p(t)^\mathsf{T}\dot{\hat{z}}(t)\,dt - \int_0^T \dot{p}(t)^\mathsf{T}\hat{z}(t)\,dt \\ &= -p(t)^\mathsf{T}\hat{z}(t)\Big|_0^T \\ &= p(0)^\mathsf{T}\hat{z}(0) - p(T)^\mathsf{T}\hat{z}(T) \qquad \forall \hat{z} \in W_2^1[0, T, \mathbb{R}^{2n}]. \end{aligned}$$

Wir beziehen nun die Randbedingungen für $\hat{z}$ (bzw. für z_2) mit ein und erhalten so

$$0 = p_1(0)^\mathsf{T} z_1(0) + p_1(T)^\mathsf{T} z_1(T) \qquad \forall z_1 \in W_2^1[0, T, \mathbb{R}^n].$$

Da die Gleichung für alle $z_1 \in W_2^1[0,T,\mathbb{R}^n]$ gilt, ergeben sich die Forderungen $p_1(0) = p_1(T) = 0_n$. Somit erhalten wir als adjungierte Gleichung

$$\begin{aligned} -\ddot{p}(t) + Ap(t) &= z(t) - z_d(t) \qquad \forall' t \in [0,T] \\ p(0) &= p(T) = 0_n. \end{aligned} \tag{2.10}$$

Bemerkung: Mit Hilfe der *Hamilton*-Funktion $H : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times [0,T] \to \mathbb{R}$ und

$$H(\hat{z}(t), u(t), p(t), t) = f(\hat{z}(t), u(t), t) + p(t)^{\mathsf{T}} h(\hat{z}(t), u(t), t)$$

schreiben wir die Lagrange-Funktion in der Form

$$\mathcal{L}(\hat{z}, u, p^*) = \int_0^T H(\hat{z}(t), u(t), p(t), t) - p(t)^{\mathsf{T}} \dot{\hat{z}}(t)\, dt + \mu^{\mathsf{T}} z_2(0) + \kappa^{\mathsf{T}} z_2(T).$$

Somit ergeben sich die Gleichungen

$$-\dot{p}(t) = H_z(\hat{z}(u^0), u^0, p(t), t) \qquad \forall' t \in [0,T]$$

und

$$J'(u^0)(u) = \int_0^T H_u(\hat{z}(u^0), u^0, p(t), t)^{\mathsf{T}} u(t)\, dt \iff J'(u^0) = H_u(\hat{z}(u^0), u^0, p).$$

$\Diamond$

Nach der Herleitung der adjungierten Gleichung mit Hilfe des formalen Lagrange-Prinzips zitieren wir eines der bekanntesten Resultate bei der Betrachtung von Problemen der optimalen Steuerung.

Lemma A.1.1 (Pontryagin'sches Maximum-Prinzip). *Gegeben sei das Steuerungsproblem*

$$J(u) = \int_0^T f(z(u)(t), u(t), t)\, dt = F(z(u), u), \qquad u \in L_\infty[0,T,\mathbb{R}^m]$$

mit den als differenzierbar angenommenen Funktionalen

$$F : W_\infty^1[0,T,\mathbb{R}^n] \times L_\infty[0,T,\mathbb{R}^m] \to \mathbb{R}, \qquad f : \mathbb{R}^n \times \mathbb{R}^m \times [0,T] \to \mathbb{R}.$$

Die Abhängigkeit der Funktion $z \in W_2^1[0,T,\mathbb{R}^n]$ von u sei mit einem $z_0 \in \mathbb{R}^n$ und einer Funktion $h : \mathbb{R}^n \times \mathbb{R}^m \times [0,T] \to \mathbb{R}^n$ gegeben durch

$$\begin{aligned} \dot{z}(t) &= h(z(t), u(t), t) \qquad \forall' t \in [0,T] \\ z(0) &= z_0. \end{aligned}$$

Die Hamilton-Funktion sei definiert durch

$$H(z(t),u(t),p(t),t) = f(z(t),u(t),t) + p(t)^{\mathsf{T}} h(z(t),u(t),t), \qquad \forall t \in [0,T].$$

Dann existiert für die optimale Steuerung $\bar{u} \in L_\infty[0,T,\mathbb{R}^m]$ *genau ein* $p \in W_2^1[0,T,\mathbb{R}^n]$ *mit*

$$\begin{aligned} -\dot{p}(t)^{\mathsf{T}} &= f_z(z(\bar{u})(t),\bar{u}(t),t) + p(t)^T h_z(z(\bar{u})(t),\bar{u}(t),t) \qquad \forall' t \in [0,T] \\ p(T) &= 0_n \end{aligned}$$

und es gilt

$$H(z(\bar{u})(t),\bar{u}(t),p(t),t) = \min_{u\in\mathbb{R}^m} H(z(u)(t),u,p(t),t) \qquad \forall' t \in [0,T].$$

Beweis: Siehe Pontryagin et al. [18] (1962). □

A.2 Beweis von Lemma 4.5.4

Beweis: Wir erinnern zunächst an die Definition der diskreten Operatoren $\mathcal{S}_h$ und $\mathcal{S}_h^*$, d.h. $z_h(u) = \mathcal{S}_h(Bu+e)$ und $p_h(u) = \mathcal{S}_h^*(\mathcal{S}_h(Bu+e) - z_d)$. Damit gelten folgende Umformungen für alle $u \in L_2[0,T,\mathbb{R}^m]$, also insbesondere auch für $\bar{u}$

$$\begin{aligned} \|z_h(\bar{u}) - z_h(P_0\bar{u})\|_2^2 &= \langle z_h(\bar{u}) - z_h(P_0\bar{u}), z_h(\bar{u}) - z_h(P_0\bar{u})\rangle \\ &= \langle \mathcal{S}_h B(\bar{u} - P_0\bar{u}), \mathcal{S}_h B(\bar{u} - P_0\bar{u})\rangle \\ \text{Lemma 4.2.5} \quad &= \langle \mathcal{S}_h^*\mathcal{S}_h B(\bar{u} - P_0\bar{u}), B(\bar{u} - P_0\bar{u})\rangle \\ &= \int_0^T \big(\bar{u}(t) - (P_0\bar{u})(t)\big)^{\mathsf{T}} B^{\mathsf{T}} \big(p_h(\bar{u})(t) - p_h(P_0\bar{u})(t)\big)\, dt. \end{aligned}$$

Das Integral auf der rechten Seite spalten wir in zwei Summanden auf, definiert durch die Mengen K_1 und K_2 aus Voraussetzung ($\mathrm{V_{con}}$)

$$\begin{aligned} \|z_h(\bar{u}) - z_h(P_0\bar{u})\|_2^2 = &\int_{K_1} \big(p_h(\bar{u})(t) - p_h(P_0\bar{u})(t)\big)^{\mathsf{T}} B\big(\bar{u}(t) - (P_0\bar{u})(t)\big)\, dt \\ &+ \int_{K_2} \big(p_h(\bar{u})(t) - p_h(P_0\bar{u})(t)\big)^{\mathsf{T}} B\big(\bar{u}(t) - (P_0\bar{u})(t)\big)\, dt. \end{aligned}$$

Für die Bearbeitung des Integrals über die Menge K_1 benötigen wir weitere Umformungen. Sei $v_h \in V_{h,0}[0,T,\mathbb{R}^n]$ beliebig gewählt. Dann gilt auf Grund der Linearität von v_h auf den Intervallen T_i, $i = 0,\ldots,N-1$, folgendes

$$\int_{t_i}^{t_{i+1}} v_h(t)^{\mathsf{T}} B(P_0\bar{u})(t)\, dt = \int_{t_i}^{t_{i+1}} v_h(t)^{\mathsf{T}} B\bar{u}(S_i)\, dt = \int_{t_i}^{t_{i+1}} v_h(S_i)^{\mathsf{T}} B\bar{u}(S_i)\, dt.$$

Damit ist es möglich für die Funktion $(B\bar{u})^\mathsf{T} v_h$ Lemma 3.3.2 anzuwenden, und zwar ist

$$\begin{aligned}\Big| \int_0^T v_h(t)^\mathsf{T} B(\bar{u}(t) - (P_0\bar{u})(t))\, dt\Big| &\leq \sum_{i=0}^{N-1} \Big| \int_{t_i}^{t_{i+1}} v_h(t)^\mathsf{T} B\big(\bar{u}(t) - \bar{u}(S_i)\big)\, dt\Big| \\ &\overset{(3.9)}{\leq} \sqrt{T} h^2 \left(\sum_{i=0}^{N-1} |(B\bar{u})^\mathsf{T} v_h|^2_{W_2^2(T_i)} \right)^{\frac{1}{2}}.\end{aligned}$$

Wir teilen jedes der Intervalle T_i in einen aktiven Teil, dort wo die optimale Steuerung konstant die Werte a oder b annimmt, und einen inaktiven Teil $I_i \subset T_i$, in dessen Gebiet $\bar{u} = -\frac{1}{\nu} B^\mathsf{T} \bar{p}$ gilt. Daher schätzen wir ab

$$|(B\bar{u})^\mathsf{T} v_h|^2_{W_2^2(T_i)} = |(B\bar{u})^\mathsf{T} v_h|^2_{W_2^2(I_i)} = \frac{|\bar{p}^\mathsf{T} BB^\mathsf{T} v_h|_{W_2^2(I_i)}}{\nu} \leq \frac{|\bar{p}^\mathsf{T} BB^\mathsf{T} v_h|_{W_2^2(T_i)}}{\nu}.$$

Somit folgt

$$\begin{aligned}\Big| \int_0^T \big(\bar{u}(t) - (P_0\bar{u})(t)\big)^\mathsf{T} B^\mathsf{T} v_h(t)\, dt\Big| &\leq \frac{\sqrt{T}}{\nu} \left(\sum_{i=0}^{N-1} |p(\bar{u})^\mathsf{T} BB^\mathsf{T} v_h|^2_{W_2^2(T_i)} \right)^{\frac{1}{2}} h^2 \\ &= \frac{\sqrt{T}}{\nu} \|p(\bar{u})^\mathsf{T} BB^\mathsf{T} v_h\|_{2,2}\, h^2 \\ &\leq \frac{\sqrt{T}}{\nu} \|BB^\mathsf{T}\|_\infty \|p(\bar{u})\|_{2,2} \|v_h\|_{1,2}\, h^2.\end{aligned}$$

Hierin verwenden wir für v_h die spezielle Funktion $p_h(\bar{u}) - p_h(P_0\bar{u})$ und erhalten

$$\begin{aligned}&\Big| \int_0^T \big(\bar{u}(t) - (P_0\bar{u})(t)\big)^\mathsf{T} B^\mathsf{T} \big(p_h(\bar{u})(t) - p_h(P_0\bar{u})(t)\big)\, dt\Big| \\ &\leq \frac{\sqrt{T}}{\nu} \|BB^\mathsf{T}\|_\infty \|p(\bar{u})\|_{2,2} \|p_h(\bar{u}) - p_h(P_0\bar{u})\|_{1,2}\, h^2 \\ &\leq c_{2.6} \sqrt{T} \frac{\|B\|^2}{\nu} \|p(\bar{u})\|_{2,2} \|z_h(\bar{u}) - z_h(P_0\bar{u})\|_2\, h^2.\end{aligned}$$

Weiterhin gilt wegen der Optimalitätsbedingung (2.16)

$$\|\dot{\bar{u}}\|_\infty \leq \frac{\|B\|}{\nu} \|\dot{p}(\bar{u})\|_\infty \leq \sqrt{T} \frac{\|B\|}{\nu} \|p(\bar{u})\|_{2,2}.$$

Die zweite Abschätzung ist gültig, denn $\dot{p}(\bar{u})$ besitzt auf Grund des Satzes von Rolle eine Nullstelle in $[0, T]$ und wir verfahren wie im Beispiel nach Satz A.3.2. Somit folgt wegen der Lipschitz-Stetigkeit von $\bar{u}$ für $t \in T_i \in K_1$ die Ungleichung

$$|\bar{u}(t) - (P_0\bar{u})(t)| = |\bar{u}(t) - \bar{u}(S_i)| \leq \|\dot{\bar{u}}\|_\infty |t - S_i| \leq \sqrt{T} \frac{\|B\|}{\nu} \|p(\bar{u})\|_{2,2}\, \frac{h}{2}.$$

Daraus resultiert dann unter Beachtung von Voraussetzung (V_{con}), d.h. $|K_2| \leq K < \infty$

$$\begin{aligned}
&\Big| \int_{K_2} \big(p_h(\bar{u})(t) - p_h(P_0\bar{u})(t)\big)^\mathsf{T} B\big(\bar{u}(t) - (P_0\bar{u})(t)\big)\, dt\Big| \\
&\leq \sum_{i\in K_1} \int_{t_i}^{t_{i+1}} |p_h(\bar{u})(t) - p_h(P_0\bar{u})(t)|\|B\|\, |\bar{u}(t) - (P_0\bar{u})(t)|\, dt \\
&\leq \sum_{i\in K_1} \sqrt{T}\frac{\|B\|^2}{\nu}\|p(\bar{u})\|_{2,2}\frac{h}{2}\|p_h(\bar{u}) - p_h(P_0\bar{u})\|_\infty \int_{t_i}^{t_{i+1}} dt \\
&\leq \sqrt{T}\frac{K}{2}\frac{\|B\|^2}{\nu}\|p(\bar{u})\|_{2,2}\|p_h(\bar{u}) - p_h(P_0\bar{u})\|_\infty\, h^2 \\
&\leq T\frac{K}{2}\frac{\|B\|^2}{\nu}\|p(\bar{u})\|_{2,2}\|p_h(\bar{u}) - p_h(P_0\bar{u})\|_{1,2}\, h^2 \\
&\leq T c_{2.6}\frac{K}{2}\frac{\|B\|^2}{\nu}\|p(\bar{u})\|_{2,2}\|z_h(\bar{u}) - z_h(P_0\bar{u})\|_2\, h^2.
\end{aligned}$$

Zusammenfassend erhalten wir

$$\|z_h(\bar{u}) - z_h(P_0\bar{u})\|_2^2 \leq (T\frac{K}{2} + \sqrt{T})\frac{\|B\|^2}{\nu}\, c_{2.6}\|p(\bar{u})\|_{2,2}\|z_h(\bar{u}) - z_h(P_0\bar{u})\|_2\, h^2$$

und da wir $\|z_h(\bar{u}) - z_h(P_0\bar{u})\|_2 = 0$ ausschließen können auch

$$\|z_h(\bar{u}) - z_h(P_0\bar{u})\|_2 \leq (T\frac{K}{2} + \sqrt{T})\frac{\|B\|^2}{\nu}\, c_{2.6}\|p(\bar{u})\|_{2,2}\, h^2,$$

was die Aussage des Lemmas gewesen war. □

A.3 Hilfsresultate

A.3.1 Funktionalanalytische Ergebnisse

Lemma A.3.1 (Hölder'sche Ungleichungen).

(1) *Die Vektoren* $\xi = (\xi_1^\mathsf{T}, \ldots, \xi_n^\mathsf{T})^\mathsf{T}$ *und* $\eta = (\eta_1^\mathsf{T}, \ldots, \eta_n^\mathsf{T})^\mathsf{T}$ *seien gegeben mit* $\xi_k, \eta_k \in \mathbb{R}^n$. *Dann gilt für* $1 < p, q < \infty$ *mit* $1 = \frac{1}{p} + \frac{1}{q}$,

$$\left|\sum_{k=1}^n \xi_k^\mathsf{T}\eta_k\right| \leq \left(\sum_{k=1}^n |\xi_k|^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^n |\eta_k|^q\right)^{\frac{1}{q}}. \tag{A.1}$$

(2) *Seien* $x \in L_p[a, b, \mathbb{R}^n]$ *und* $y \in L_q[a, b, \mathbb{R}^n]$ *für* $1 < p, q < \infty$ *mit* $1 = \frac{1}{p} + \frac{1}{q}$. *Dann ist* $x^\mathsf{T} y \in L_1[a, b, \mathbb{R}]$ *und es gilt*

$$\left|\int_a^b x(t)^\mathsf{T} y(t)\, dt\right| \leq \left(\int_a^b |x(t)|^p\, dt\right)^{\frac{1}{p}} \left(\int_a^b |y(t)|^q\, dt\right)^{\frac{1}{q}}. \tag{A.2}$$

Beweis: Für den Fall $n = 1$ verweisen wir auf Triebel [22] (1972, Kapitel 1). Der Übergang zu beliebigem $n \in \mathbb{N}$ ist offensichtlich. □

Satz A.3.2 (Sobolew'scher Einbettungssatz). *Sei $1 \leq p < \infty$ und $k, l \in \mathbb{N}$. Falls $k > l + \frac{1}{p}$ gilt, so ist der Einbettungsoperator $E_{k,p,l} : W_p^k[0,T,\mathbb{R}^n] \to C^l[0,T,\mathbb{R}^n]$, mit $E_{k,p,l}f = f$ für $f \in W_p^k[0,T,\mathbb{R}^n]$, stetig und es gilt für $j = 0,\ldots,l$ die Ungleichung*

$$\|\mathscr{D}^j f\|_\infty \leq c_{k,p,l}\,\|f\|_{W_p^k}$$

mit einer von f unabhängigen Konstante $c_{k,p,l} > 0$.

Das bedeutet, dass in jeder Äquivalenzklasse von $W_p^k[0,T,\mathbb{R}^n]$ eine stetige Funktion enthalten ist.

Beweis: Eine vollständige Abhandlung verschiedener Einbettungsmöglichkeiten finden wir in Adams [1] (1975, Kapitel V). □

Beispiel: An dieser Stelle beweisen wir exemplarisch einige Ungleichungen direkt und bestimmen die Abschätzungskonstanten exakt. Wir wählen dazu $p = 2$, $k = 1$, $l = 0$ und $z \in W_2^1[0,T,\mathbb{R}^n]$. Dann ist auf Grund der absoluten Stetigkeit von z die Anwendung des Mittelwertsatzes gerechtfertigt und wir wählen die Stelle $t_0 \in [0,T]$ mit $|z(t_0)| = \min\limits_{t\in[0,T]} |z(t)|$. Das Minimum wird auf GRund der Stetigkeit auf dem kompakten Intervall $[0,T]$ angenommen. Dann folgt

$$z(t) = z(t_0) + \int_{t_0}^{t} \dot{z}(t)\,dt$$

und weiter

$$\begin{aligned}|z(t)| &\leq |z(t_0)| + \Big|\int_{t_0}^{t} \dot{z}(t)\,dt\Big| \leq \frac{1}{T}\int_0^T |z(t)|\,dt + \int_{t_0}^{t} |\dot{z}(t)|\,dt \\ &\leq \frac{1}{T}\|z\|_1 + \|\dot{z}\|_1 \leq \frac{1}{\sqrt{T}}\|z\|_2 + \sqrt{T}\|\dot{z}\|_2 \leq \sqrt{2}\max\{\frac{1}{\sqrt{T}}, \sqrt{T}\}\,\|z\|_{1,2}.\end{aligned}$$

Daraus folgt dann abschließend

$$\|z\|_\infty \leq \sqrt{2}\max\{\frac{1}{\sqrt{T}}, \sqrt{T}\}\,\|z\|_{1,2}. \tag{A.3}$$

Besitzt z in $[0,T]$ eine Nullstelle t_0, so vereinfacht sich die eben durchgeführte Schlussweise. Es ist dann

$$\|z\|_\infty \leq \sqrt{T}\|\dot{z}\|_2 \leq \sqrt{T}\|z\|_{1,2}. \tag{A.4}$$

Sei nun $z \in W_{2,0}^2[0,T,\mathbb{R}^n]$ und die Größe $\|\dot{z}\|_\infty$ von Interesse. Da die beiden Randpunkte Nullstellen von z sind, existiert nach dem Satz von Rolle mindestens ein $t_0 \in [0,T]$ mit $\dot{z}(t_0) = 0_n$, so dass wir wie eben abschätzen

$$\|\dot{z}\|_\infty \leq \sqrt{T}\|\dot{z}\|_{1,2} \leq \sqrt{T}\|z\|_{2,2}.$$

Damit folgt sofort

$$\|z\|_{C^1} = \max\{\|z\|_\infty, \|\dot{z}\|_\infty\} \leq \sqrt{T}\|z\|_{2,2}.$$

$\Diamond$

Satz A.3.3 (Satz von Gerschgorin). *Sei A eine reelle symmetrische $n \times n$-Matrix mit den Einträgen A_{ik} und $1 \leq i, k \leq n$. Dann gilt für die Menge ihrer Eigenwerte $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$*

$$\Lambda \subset \bigcup_{i=1}^{n} \Big\{ \lambda \in \mathbb{R} : |\lambda - A_{ii}| \leq \sum_{\substack{k=1 \\ k \neq i}}^{n} |A_{ik}| \Big\}.$$

Beweis: Dazu verweisen wir auf Stoer/Bulirsch [21] (2000). □

Satz A.3.4 (Charakterisierung der Projektion). *Sei $U^{ad} \subset U$ eine abgeschlossene, konvexe Teilmenge des Hilbert-Raumes $[U, \langle \cdot, \cdot \rangle]$ und $\bar{u} \in U$. Dann gilt*

$$\|u - \bar{u}\| = \inf_{y \in U^{ad}} \|y - \bar{u}\| \Longleftrightarrow \langle u - \bar{u}, y - \bar{u} \rangle \leq 0 \qquad \forall y \in U^{ad}.$$

Beweis: [Werner [24] (2003)]

„$\Leftarrow$“ Es ist für alle $y \in U^{ad}$

$$\|u - y\|^2 = \|u - \bar{u} + \bar{u} - y\|^2 = \|u - \bar{u}\|^2 + 2\langle u - \bar{u}, \bar{u} - y \rangle + \|\bar{u} - y\|^2 \geq \|u - \bar{u}\|^2.$$

Damit folgt, da $y \in U^{ad}$ beliebig ist,

$$\|u - \bar{u}\| = \inf_{y \in U^{ad}} \|u - y\|.$$

„$\Rightarrow$“ Zu $t > (0, 1)$ setze $y_t = (1-t)\bar{u} + ty$ mit $y \in U^{ad}$. Dann ist für beliebiges $y \in U^{ad}$

$$\begin{aligned} \|u - \bar{u}\|^2 \leq \|u - y_t\|^2 &= \langle u - \bar{u} + t(\bar{u} - y), u - \bar{u} + t(\bar{u} - y) \rangle \\ &= \|u - \bar{u}\|^2 + 2\langle u - \bar{u}, t(\bar{u} - y) \rangle + t^2 \|\bar{u} - y\|^2 \end{aligned}$$

Daraus folgt auf Grund $0 < t < 1$

$$\langle u - \bar{u}, y - \bar{u} \rangle \leq \frac{t}{2} \|\bar{u} - y\|^2$$

und für $t \to 0$ erhalten wir

$$\langle u - \bar{u}, y - \bar{u} \rangle \leq 0 \qquad \forall y \in U^{ad}.$$

□

Satz A.3.5 (Fréchet-Riesz). *Sei X ein Hilbert-Raum mit dem Skalarprodukt $\langle \cdot, \cdot \rangle$ und dem Dualraum X'. Dann wird für jedes feste $z \in X$ durch $\ell(x) = \langle z, x \rangle$ ein stetiges lineares Funktional auf X definiert. Umgekehrt existiert zu jedem stetigen linearen Funktional ℓ auf X genau ein Element $z \in X$, so dass $\ell(x) = \langle z, x \rangle$ für alle $x \in X$ gilt. Darüber hinaus gilt $\|\ell\|_{X'} = \|z\|_X$.*

Beweis: Wir verweisen wieder auf Werner [24] (2003). □

A.3.2 Resultate zur Methode der Finiten-Elemente

Lemma A.3.6 (Lax-Milgram-Lemma). *Sei V ein Hilbert-Raum, $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ eine stetige Bilinearform und $\ell : V \to \mathbb{R}$ ein stetiges lineares Funktional. Ist $a(\cdot,\cdot)$ darüber hinaus gleichmäßig elliptisch auf V, d.h. gilt mit einer von den Argumenten unabhängigen Konstante $c > 0$*

$$a(z,z) \geq c \, \|z\|_V^2, \qquad \forall v \in V$$

so besitzt das Variationsproblem

$$a(z,v) = \ell(v) \qquad \forall v \in V$$

genau eine Lösung $z \in V$.

Beweis: Einen knappen, aber dennoch leicht verständlichen Beweis finden wir in Ciarlet [9] (1978). □

Bemerkung: Im Gegensatz zu Satz 2.2.5 wird hier nicht die Symmetrie der Bilinearform $a(\cdot,\cdot)$ vorausgesetzt. ♢

Beispiel: Für ein Beispiel setzen wir $H = L_2[0,T,\mathbb{R}^n]$, $V = W_{2,0}^1[0,T,\mathbb{R}^n]$ und $V_h = V_{h,0}[0,T,\mathbb{R}^n]$. Dann besitzt die Gleichung

$$a(z,v) = \int_0^T y(t)^\mathsf{T} v(t) \, dt \qquad \forall v \in W_{2,0}^1[0,T,\mathbb{R}^n].$$

genau eine Lösung $z \in W_2^1[0,T,\mathbb{R}^n]$ für alle $y \in L_2[0,T,\mathbb{R}^n]$. Die Gleichung

$$a(z_h,v_h) = \int_0^T y(t)^\mathsf{T} v_h(t) \, dt \qquad \forall v_h \in V_{h,0}[0,T,\mathbb{R}^n]$$

besitzt ebenfalls eine eindeutige Lösung z_h, in diesem Fall aus dem Raum $V_{h,0}[0,T,\mathbb{R}^n]$. ♢

Lemma A.3.7 (Aubin-Nitsche-Lemma). *Sei H ein Hilbert-Raum ausgestattet mit dem Skalarprodukt $\langle\cdot,\cdot\rangle$ und der zugehörigen Norm $\|\cdot\|$. Seien weiterhin V und V_h zwei lineare Unterräume von H, für die gelte $V_h \subset V \subset H$ und $\dim V_h < \infty$. Der Raum V werde mit der Norm $\|\cdot\|_V$ zum Hilbert-Raum und die Einbettung $V \hookrightarrow H$ sei stetig. Betrachten wir für eine stetige und auf V gleichmäßig elliptische Bilinearform $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ und eine Funktion $u \in H$ die beiden Variationsgleichungen*

$$\begin{aligned} a(z,v) &= \langle u,v\rangle \qquad \forall v \in V \\ a(z,v_h) &= \langle u,v_h\rangle \qquad \forall v_h \in V_h, \end{aligned}$$

mit den eindeutig bestimmten Lösungen z und z_h.

Dann gilt

$$\|z - z_h\| \leq c\|z - z_h\|_V \sup_{g \in H} \Big\{ \frac{1}{\|g\|} \inf_{v_h \in V_h} \|z(g) - v_h\|_V \Big\},$$

wobei $z(g)$ die eindeutig bestimmte Lösung der Gleichung

$$a(z(g), v) = \langle g, v \rangle \qquad \forall v \in V$$

zu $g \in H$ ist.

Beweis: Dazu verweisen wir auf Braess [7] (1997, 7.6). □

Literaturverzeichnis

[1] R. A. Adams, *Sobolev Spaces*, 1975, Academic Press, New York

[2] W. Alt, *Optimale Steuerung*, 2005, Vorlesungsskript, Universität Jena

[3] W. Alt, *Nichtlineare Optimierung*, 2003, Vieweg, Heidelberg

[4] K. Atkinson, W. Han, *Theoretical Numerical Analysis, A Functional Analysis Framework*, 2001, Texts in Applied Mathematics 39, Springer, Berlin bibitem[17] QSS A. Quarteroni, R. Sacco, F. Scaleri, *Numerische Mathematik 2*, 2002, Springer, Berlin

[5] J. P. Aubin, *Behaviour of the error of the approximate solution of boundary value problems for linear elliptic operators by Galerkin's and finite difference methods*, 1967, Ann. scoula Norm. Sup. Pisa, 21, 599–637

[6] G. Birkhoff, M. H. Schultz, R. S. Varga, *Piecewise Hermite interpolation in one and two variables with apllications to partial differential equations*, 1968, Numerical Mathematics, 11, 232–256

[7] D. Braess, *Finite Elemente*, 1997, Springer, Berlin

[8] J. Céa, *Approximation variationelle des problémes aux limites*, 1964, Ann. Inst. Fourier, 14, 345–444

[9] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, 1987, 2. Auflage, North-Holland, Amsterdam

[10] P. G. Ciarlet, *Basic Error Estimates for Elliptic Problems*, 1991, in: Ciarlet, Lions, *Handbook of Numerical Analysis*, North-Holland, 17–293

[11] A. L. Dontchev, W. W. Hager, V. M. Veliov, *Second-order Runge-Kutta Approximations in Control Constrained Optimal Control*, 2000, SIAM Journal of Numerical Analysis, 38, 1, 202–226

[12] G. Feichtinger, R. Hartl, *Optimale Kontrolle ökonomischer Prozesse: Anwendungen des Maximumprinzips in den Wirtschaftswissenschaften*, 1986, de Gruyter, Berlin

[13] C. Großmann, H.-G. Roos, *Numerische Behandlung partieller Differentialgleichungen*, 2005, 3. Auflage, Teubner, Wiesbaden

[14] M. Hinze, *A Variational Discretization Concept in Control Constrained Optimization: The Linear-Quadratic Case*, 2005, Computational Optimization and Applications, 30, 45–61

[15] P. Knabner, L. Angermann, *Numerical Methods for Elliptic and Parabolic Partial Differential Equations*, 2003, Texts in Applied Mathematics 44, Springer, Berlin

[16] C. Meyer, A. Rösch, *Superconvergence Properties of Optimal Control Problems*, 2004, SIAM Journal of Control and Optimization, 43, 3, 970–985

[17] J. A. Nitsche, *Ein Kriterium für die Quasioptimalität des Ritzschen Verfahrens*, 1968, Numerische Mathematik, 11, 346–348

[18] L. S. Pontryagin, V. G. Boltyankii, R. V. Gamkrelidze, E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, 1962, Wiley-Interscience, New York

[19] A. A. Samarski, *Theorie der Differenzenverfahren*, 1984, Geest & Portig, Leipzig

[20] B. Sendov, V. A. Popov, *The averaged Moduli of Smoothness*, 1988, Wiley-Interscience

[21] J. Stoer, R. Bulirsch, *Numerische Mathematik 2*, 4. Auflage, 2000, Springer, Berlin

[22] H. Triebel, *Höhere Analysis*, 1972, VEB Deutscher Verlag der Wissenschaften, Berlin

[23] A. Tveito, R. Winter, *Einführung in partielle Differentialgleichungen*, 2002, Springer, Berlin

[24] D. Werner, *Funktionalanalysis*, 5. Auflage, 2004, Springer, Berlin

[25] F. Tröltzsch, *Optimale Steuerung partieller Differentialgleichungen*, 2005, Vieweg, Heidelberg

Selbständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Jena, 24.04.2006 Nils Bräutigam