

Zur Lösung optimaler Steuerungsprobleme:
Diskretisierung, Konvergenz, Anwendung

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor rerum naturalium (dr. rer. nat.)
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von

Herr Dipl. Math. Nzali Appolinaire
geboren am 20.12.1963 in Abong-Mbang (Kamerun)

Präsident der Humboldt-Universität zu Berlin:

Prof. Dr. Jürgen Mlynek

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:

Prof. Dr. Bodo Krause

Gutachter:

1. Prof. Dr. sc. nat. Werner Römisch
2. Prof. Dr. Fredi Tröltzsch
3. PD. Dr. Ursula Felgenhauer

eingereicht am: 01.09.2000

Tag der mündlichen Prüfung: 12.10.2001

Abstract

The focal point of this work is the investigation of a class of discretization methods for nonlinear optimal control problems governed by ordinary differential equations with control restrictions, as well as the implementation of some numerical experiments. The theoretical investigations are based on a coupled parameterization-discretization pattern, a piecewise linear parameterization representation of the control, and the application of a Runge-Kutta method for the integration of the differential state equation. The order of convergence of the solution is obtained with the help of regularity conditions and the second order optimality conditions. Furthermore, we also present in this paper a possibility of the numerical computation of the gradients via numerical differentiation. Finally some numerical results are given and their relationship to the theoretical convergence results are discussed.

Keywords:

Optimal control, Discretization, Runge-Kutta scheme, Rate of convergence, Numerical results

Zusammenfassung

Schwerpunkt dieser Arbeit ist die Untersuchung einer Klasse von Diskretisierungsmethoden für nichtlineare optimale Steuerungsprobleme mit gewöhnlichen Differentialgleichungen und Steuerungsbeschränkung sowie die Durchführung von numerischen Experimente. Die theoretischen Untersuchungen basieren auf einem gekoppeltes Parametrisierungs-Diskretisierungsschema auf stückweise polinomialen Ansatz für die Steuerung und einem Runge-Kutta-Verfahren zur Integration der Zustands-Differentialgleichung. Die Konvergenzordnung der Lösung wird unter Regularitätsbedingung und Optimalitätsbedingung 2. Ordnung ermittelt. Außerdem wird eine Möglichkeit zur numerischen Berechnung der Gradienten über interne numerische Differentiation erläutert. Schließlich werden einige numerische Resultate gegeben und die Abhängigkeiten zu den theoretischen Konvergenzresultate diskutiert.

Schlagerwörter:

Optimale Steuerungsprobleme, Diskretisierung, Runge-Kutta-Verfahren, Konvergenzordnung, Numerische Resultate

Widmung

Dieses Manual ist meine Mutter: Madeleine Kembou mein Vater: David Tchouangue und meine Geschwistern gewidmet.

Vorwort

Danksagung

Die vorliegende Arbeit entstand während meines Promotionsstudiums am Institut für Mathematik der Humboldt-Universität zu Berlin.

Mein herzlicher Dank gilt dem Betreuer Prof. Dr. sc. nat. Werner Römisch, für seine wohlwollende und stete Unterstützung, wertvollen Anregungen sowie fachlich kritischen Hinweise.

Ebenso bin ich dem Hochschulerneuerungsprogramm (HEP) sowie der Deutsche Forschungsgemeinschaft (in Deutsche Forschungsgemeinschaft-Projekt unter der Leitung von PD. Dr. René Henrion) für die finanzielle Unterstützung herzlich dankbar.

Viele kritische Hinweise in Darstellung von Dr. Kengne Jean-Marie und sprachlich kritische Hinweise von Dr. Matthias Gelbrich haben dazu beigetragen, die Lesebarkeit des Textes zu erhöhen.

Ich bin Frau Heidemarie Lange (Sekretariat) und allen Kolleginnen und Kollegen der Forschungsgruppe unter der Leitung von Prof. Dr. sc. nat. Werner Römisch für die vielen Diskussionen zu Dank verpflichtet.

Der TRAD.EST GmbH danke ich für ihre stete Hilfsbereitschaft, insbesondere bei Problemen mit deutschen Behörden während des Studiums.

Nicht zuletzt möchte ich mich bei der Regierung der Republik Kamerun bedanken, die mir mein Auslandsstudium ermöglicht hat.

Berlin, im Oktober 2001

Appolinaire Nzali

Inhaltsverzeichnis

Abbildungsverzeichnis	vii
Tabellenverzeichnis	viii
Symbole	ix
1 Einleitung	1
1.1 Optimale Steuerungsprobleme	1
1.2 Untersuchung von optimalen Steuerungsproblemen	1
1.3 Motivation und Ziel der Untersuchung	2
1.4 Inhaltliche Struktur	5
2 Problemstellung	7
2.1 Formulierung eines Steuerungsproblems	7
2.2 Umformulierung anderer Problemstellungen	8
3 Diskretisierung	10
3.1 Numerische Optimierungsmethoden	10
3.2 Optimierungsvariable	11
3.3 Parametrisierung des Steuerungsraumes	12
3.3.1 Ansatz der Parametrisierung der Steuerung	13
3.3.2 Andere Parametrisierungsmethoden der Steuerung	14
3.4 Überführung des Steuerungsproblems in ein Optimierungsproblem	17
3.5 Notwendige stetige Optimalitätsbedingungen	18
3.5.1 Phänomen der Zweinorm-Diskrepanz bei Steuerungsproblemen	19
3.5.2 Hinreichende Regularitätsbedingungen eines Punktes	20
3.5.3 Stetiges Minimum-Prinzip	21
3.6 Diskretisierung des Problems	25

3.6.1	Diskrete Aufgabe	25
3.6.2	Minimum-Prinzip der diskreten Aufgabe	26
3.7	Gradientenberechnung diskretisierter Funktionen	27
3.7.1	Die Methode der internen numerischen Differentiation	28
3.7.2	Gradientenberechnung bei Runge-Kutta-Verfahren	29
4	Konvergenzuntersuchung	34
4.1	Allgemeine Approximationstheorie	34
4.2	Parametrisierte Aufgaben	38
4.2.1	Parametrisierte verallgemeinerte Gleichung	38
4.2.2	Approximierende verallgemeinerte Gleichung	39
4.3	Konvergenzsatz	44
4.3.1	Normenvergleich	47
4.3.2	Abschätzung des Diskretisierungsfehlers	50
4.3.3	Lipschitz-Stetigkeit der Lösung gewöhnlicher Differentialgleichungen bzgl. Parametern	55
4.4	Anwendung des Konvergenzsatzes	59
4.4.1	Runge-Kutta-Verfahren	59
4.4.2	Stabile diskrete Koerzitivitätsbedingung	61
5	Optimierungsverfahren	64
5.1	Grundlegende Idee des SQP-Verfahrens	65
5.2	QP-Teilproblem	65
5.3	Bewertungsfunktion	66
5.4	Approximation der Hesse-Matrix	67
5.5	Abbruchkriterien	68
6	Implementierung des Algorithmus	69
6.1	Beschreibung des Algorithmus	69
6.1.1	Gradientenberechnung bei explizitem Runge-Kutta-Verfahren	70
6.1.2	Implizite Berechnung des Zustandes	72
6.1.3	Interne Routinen	73
6.2	Beschreibung der SNOPT-Routine	74
6.2.1	Lösungsprozeß der SNOPT-Routine	75
6.2.2	Erfassung der Jacobi-Matrix	76
6.2.3	Behandlung von nichtzulässigen Restriktionen durch SNOPT	78

7	Testrechnung	79
7.1	Benutzung der RINDOPT-Routine	79
7.1.1	Beschreibung von Abkürzungen und Arbeitsspeicher	79
7.1.2	Numerische Auswertung der Konvergenzrate	80
7.1.3	Erläuterungsbeispiel	81
7.1.4	Nutzer-Quellcode für RINDOPT	87
7.2	Anwendungsbeispiele	92
7.2.1	Container-Kran-Problem	92
7.2.2	Flug eines Hängegleiters bei Aufwind	96
7.2.3	Tunneldiodenoszillator	100
7.2.3.1	Variante 1	102
7.2.3.2	Variante 2	103
8	Abschlußbemerkungen	106
	Literaturverzeichnis	109

Abbildungsverzeichnis

6.1	RINDOPT-Algorithmus	70
6.2	OBJFUN-Routine	72
7.1	Optimale Steuerung, ohne Zustandsrestriktion	82
7.2	Optimaler Zustand, ohne Zustandsrestriktion	82
7.3	Diskrete optimale Steuerung, ohne Zustandsrestriktion	84
7.4	Diskrete Zustandskomponenten, ohne Zustandsrestriktion	84
7.5	Diskrete optimale Steuerung, mit Zustandsrestriktion	86
7.6	Diskrete Zustandskomponenten, mit Zustandsrestriktion	86
7.7	Diskrete optimale Steuerung, willkürliches Steuerungsgitter	87
7.8	Diskrete Zustandskomponenten, beliebiges Steuerungsgitter	87
7.9	Container-Kran-Problem	92
7.10	Container-Kran-Problem, Steuerungskomponenten	96
7.11	Container-Kran-Problem, Zustandskomponenten	97
7.12	Hängegleiter, optimale Steuerung	99
7.13	Hängegleiter, Zustandskomponenten	99
7.14	Tunneldiodenoszillator	100
7.15	Tunneldiode, optimale Steuerung, Variante 1	102
7.16	Tunneldiode, Zustandskomponenten, Variante 1	103
7.17	Tunneldiode, optimale Steuerung, Variante 2	104
7.18	Tunneldiode, Zustandskomponenten, Variante 2	104

Tabellenverzeichnis

6.1	Koeffizienten des Runge-Kutta-Merson-Verfahrens	69
7.1	Bahnoptimierung, ohne Zustandsrestriktion	83
7.2	Bahnoptimierung, Konvergenzrate I	84
7.3	Bahnoptimierung, mit Zustandsrestriktion	85
7.4	Bahnoptimierung, beliebiges Steuerungsgitter	85
7.5	Container-Kran-Problem, Bezeichnungen	93
7.6	CKP, Numerische Resultate, Fall 1	95
7.7	Daten und Funktionen des Hängegleiter-Problems	98
7.8	Hängegleiter, Numerische Ergebnisse, $x_{A,0} = 250$ [m]	98
7.9	Hängegleiter, Numerische Ergebnisse, $x_{A,0} = 350$ [m]	100
7.10	Hängegleiter, Numerische Ergebnisse aus [120]	100
7.11	Tunneldiodenoszillator, Bezeichnungen	101
7.12	Tunneldiode, Numerische Resultate, Variante 1	102
7.13	Tunneldiode, Numerische Resultate, Variante 2	103
7.14	Tunneldiode, Konvergenzrate II	105

Symbole

f.ü.	fast überall,
\emptyset	leere Menge,
\mathbb{N}	$\{0, 1, 2, \dots\}$ (Menge der nicht negativen ganzen Zahlen),
\mathbb{R}	Raum der reellen Zahlen,
\mathbb{R}_+	$:= [0, \infty)$,
\mathbb{R}^n	Euklidischer n-dimensionaler Raum,
$ \cdot $	absoluter Wert,
$\ \cdot\ $	allgemeine Bezeichnung für eine Norm,
$\frac{\partial f}{\partial x}$	Fréchet-Ableitung der Funktion f nach x mit $f : \mathbb{R} \rightarrow \mathbb{R}$
$ A - B $	die einseitige Definition des Abstands der Menge A zur Menge B definiert auf Seite 35,
∇	Differentialoperator, definiert durch $\nabla f(x) = (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})^\top$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (Falls es nicht eindeutig ist, nach welchem Argument differenziert wird, dann erscheint dies als Index bei dem Operator, z.B. ∇_x),
∇^2	zweiter Differentialoperator, definiert durch $\nabla(\nabla f(x))$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (Falls es nicht eindeutig ist, nach welchem Argument differenziert wird, dann erscheint dies als Index bei dem Operator, z.B. $\nabla_{x,x}^2$),
2^{Ω_p}	Potenzmenge von Ω_p (Menge aller Teilmengen von Ω_p),
$\text{int}(A)$	Inneres der Menge A (Menge aller inneren Punkte von A),
$\text{cl}(A)$	Abschließung der Menge A ,
$d(A, B)$	$\inf_{a \in A, b \in B} \ a - b\ $,

(x, y)	Skalarprodukt definiert durch $(x, y) = \int_0^1 x(t)y(t)dt$ für stetige Funktionen x und y auf dem Intervall $[0, 1]$,
$\langle x, y \rangle$	Kreuzprodukt definiert durch $\langle x, y \rangle = x^\top y = \sum_{l=1}^n x_l y_l$ für $x, y \in \mathbb{R}^n$ mit $x = (x_1, \dots, x_n)^\top, y = (y_1, \dots, y_n)^\top$,
U_a	Menge der zulässigen Steuerungsfunktionen in \mathbb{R}^r ,
$C([0, 1]; \mathbb{R}^n)$	Raum der auf dem Intervall $[0, 1]$ stetigen Funktionen, mit der Norm $\ x\ _C = \max\{ x(t) : t \in [0, 1]\}$,
$C^k([0, 1]; \mathbb{R}^n)$	Raum der auf dem Intervall $[0, 1]$ k -mal stetig differenzierbaren Funktionen, $k \in \{1, \dots\}$,
$\overset{1}{V}_0$	Menge aller Funktionen von beschränkten Variationen auf dem Segment $[0, 1]$,
$\overset{1}{V}_0(u)$	die Variation der Funktion u auf dem Intervall $[0, 1]$,
$\mathcal{L}^e([0, 1], \mathbb{R}^n)$	Lebesgue Raum (Menge aller Funktionen $y : [0, 1] \rightarrow \mathbb{R}^n$, wobei $ y(t) ^e$ für $1 \leq e < \infty$ integrierbar ist und mit der Norm: $\ y\ _{\mathcal{L}^e} = \left(\int_0^1 y(t) ^e dt \right)^{1/e},$
$\mathcal{L}^\infty([0, 1], \mathbb{R}^r)$	Menge der wesentlich beschränkten Funktionen $u : [0, 1] \rightarrow \mathbb{R}^r$ mit der Norm $\ u\ _{\mathcal{L}^\infty} = \sup_{[0,1]} \text{ess } u(t) , (= \max_{[0,1]} u(t) $ für stetige Funktionen),
$\mathcal{W}^{k,e}([0, 1], \mathbb{R}^n)$	Sobolev-Raum (Menge aller Funktionen $y : [0, 1] \rightarrow \mathbb{R}^n$, wobei y und ihre j -te Ableitung $y^{(j)}, 0 \leq j \leq k$ in der Menge $\mathcal{L}^e([0, 1], \mathbb{R}^n)$ liegen und mit der Norm $\ y\ _{\mathcal{W}^{k,e}} = \sum_{j=0}^k \ y^{(j)}\ _{\mathcal{L}^e}$),
$\mathcal{W}^{1,\infty}([0, 1], \mathbb{R}^n)$	Menge der Lipschitz-stetigen Funktionen $y : [0, 1] \rightarrow \mathbb{R}^n$ mit der Norm $\ y\ _{\mathcal{W}^{1,\infty}} = \ y\ _{\mathcal{L}^\infty} + \ y^{(1)}\ _{\mathcal{L}^\infty}$,
\mathcal{X}	$:= \mathcal{W}^{1,\infty}([0, 1], \mathbb{R}^n)$ Raum der Zustandsfunktionen,
\mathcal{U}	$:= \mathcal{L}^\infty([0, 1]; \mathbb{R}^r)$ Raum der Steuerungsfunktionen,
\mathcal{X}^* bzw. \mathcal{Y}^*	Dualer Raum von \mathcal{X} bzw. \mathcal{Y} ,
P^+	Polarkegel (d.h. für $P \in \mathcal{X}, P^+ = \{p \in \mathcal{X}^* : p(a) \geq 0, \forall a \in P\}$),
$\prod_{j=-k}^{M-1} U$	das Cartesische Produkt von $M + k$ Kopien von $U (= U^{M+k})$,
\mathcal{I}_h	Steuerungsgitter: $\{\tau_0, \tau_1, \tau_2, \dots, \tau_{M-1}, \tau_M\}, \tau_0 := 0 < \tau_1 < \tau_2 < \dots < \tau_{M-1} < \tau_M := 1$,

h	$:= \max_{1 \leq i \leq M} (\tau_i - \tau_{i-1}),$
\mathcal{G}_h	Zustandsgitter: $\{t_0, t_1, \dots, t_N\}$, $t_0 := 0 < t_1 < \dots < t_N := 1$ mit $\mathcal{I}_h \subseteq \mathcal{G}_h$,
\mathcal{IP}_μ	Menge der Polynome vom Grad strikt kleiner als μ ,
$\mathcal{IP}_{\mu, \Psi_h, \eta}$	Stückweise Polynome der Ordnung μ mit Unstetigkeitsstellen (Bruchstellen) ς_i und der Glattheit η_i in ς_i mit $\varsigma_i < \varsigma_{i+1}$, $i = 1, \dots, K-1$, $\Psi_h := \{\varsigma_0, \varsigma_1, \dots, \varsigma_{K-1}, \varsigma_K\}$, $K \in \mathbb{N} \setminus \{0\}$ (d.h. die Menge aller Funktionen, welche auf dem Teilintervall $(\varsigma_i, \varsigma_{i+1})$ mit einem Element von \mathcal{IP}_μ übereinstimmt und den j -ten Ableitungen für $j < \eta_i$ im Punkt ς_i stetig ist),
$\Delta_h^k u(t)$	die k -te Differenz („ k th difference“) der Funktion u mit der Schrittweite h im Punkt t definiert auf Seite 15,
$\omega_k(u, t; \gamma)$	das lokale Stetigkeitsmodul der Funktion u von der Ordnung k im Punkt $t \in [0, 1]$ definiert auf Seite 15,
$\tau_k(u; \gamma)$	das mittlere Stetigkeitsmodul der Funktion u von der Ordnung k gegeben auf Seite 15,
$l^e(N, \mathbb{R}^n)$	Raum der stückweisen konstanten Funktionen, auf ein Gitter \mathcal{G}_h ausgestattet mit der Norm $\ y_h\ _{l^e} := \sum_{l=1}^N h_l \ y_{h,l}\ ^e < \infty$, wobei y_h eine stückweise Erweiterung der Folge $\{y_{h,l}\}$ auf $[t_{l-1}, t_l]$, $l = 1, \dots, N$ ist,
$l^\infty(N, \mathbb{R}^n)$	Raum der stückweisen konstanten Funktionen, auf ein Gitter \mathcal{G}_h ausgestattet mit der Norm $\ y_h\ _{l^\infty} := \max\{\ y_{h,l}\ , l \in \{1, \dots, N\}\} < \infty$, wobei y_h eine stückweise Erweiterung der Folge $\{y_{h,l}\}$ auf $[t_{l-1}, t_l]$, $l = 1, \dots, N$ ist,
$O(b_n)$	Für zwei Folgen a_n und b_n reeller Zahlen mit $n \in \mathbb{N}$ schreiben wir $a_n = O(b_n)$, falls folgende Äquivalenz gilt:
	$a_n = O(b_n) \iff \frac{a_n}{b_n} \leq c$
	für alle n und ein geeignetes c ,
$o(b_n)$	Für zwei Folgen a_n und b_n reeller Zahlen mit $n \in \mathbb{N}$ schreiben wir $a_n = o(b_n)$, falls folgende Äquivalenz gilt:
	$a_n = o(b_n) \iff \lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0.$

Kapitel 1

Einleitung

1.1 Optimale Steuerungsprobleme

Optimale Steuerungsprobleme bilden eine wichtige und interessante Gruppe von Optimierungsproblemen. Ihre Besonderheit im Vergleich zu den einfachen Optimierungsproblemen ist die Existenz eines dynamischen Systems in der Beschreibung des Problems, welches in einem Funktionenraum (unendlich-dimensionalen Raum) formuliert ist. In der Natur finden optimale Steuerungsprobleme unter anderem in folgenden Bereichen Anwendung: Förder- und Transportanlagen (z.B. Container-Kran-Problem [94, 48]), Elektrotechnik (z.B. Tunnelndiodenoszillator [75, 63]) und Luft- und Raumfahrt (z.B. Aufstieg eines Huperschall-Raumtransporters [22]).

1.2 Untersuchung von optimalen Steuerungsproblemen

Generell sind optimale Steuerungsprobleme komplexe Optimierungsaufgaben und daher nur numerisch lösbar. Die kennzeichnenden Merkmale der Komplexität solcher Aufgaben sind u.a. die hohe Anzahl von nichtlinearen Gleichungs- und Ungleichungsrestriktionen, die zahlreichen Steuerungs- oder Zustandsbeschränkungen und das dynamische System von komplizierten Prozessen. Die numerische Lösung von optimalen Steuerungsproblemen basiert auf dem Ersatz des ursprünglichen Problems durch ein angenähertes Problem mittels Approximationsverfahren und der Lösung des angenäherten Problems. Einige der wichtigsten Ziele der Anwendung von Approximationsverfahren sind die Konvergenz und die Konvergenzgeschwindigkeit der Lösungen der diskreten Aufgaben zur Lösung der ursprünglichen Aufgabe. Hierbei stellt die Stabilität der gesuchten Lösung die fundamentale Voraussetzung dar.

Die Lösung von Steuerungsproblemen zieht die numerische Lösung des dynamischen Systems mit sich. Zur numerischen Lösung von Differentialgleichungen bietet die Literatur zahlreiche Methoden an. Zu nennen wären das mehrfache Schießverfahren („multiple shooting technique“) für die Lösung von nichtlinearen Randwertaufgaben [13, 16, 59]), das finite Differenzen-Verfahren [118], das Integrationsverfahren. Diese Arbeit widmet sich unter anderem der Untersuchung eines allgemeinen Einschrittverfahrens.

Bereits in den siebziger Jahren wurde die Konvergenz der endlich-dimensionalen Approximation von optimalen Steuerungsproblemen mittels Integrationsverfahren untersucht. Siehe beispielsweise die Arbeiten *Cullum* [26], *Budak, Berkovich, Solov'eva* [15], *Daniel* [27], *Mordukhovich* [81].

Malanowski [70] behandelt optimale Steuerungsprobleme mit konvexer Zielfunktion und linearem (bezüglich der Steuerung) dynamischen System. *Dontchev* [30] führt eine Studie für konvexe Probleme an. In diesen älteren Arbeiten stelle man die Anwendung der *Euler*-Methode für die Diskretisierung der Differentialgleichung und die Betrachtung des gleichen Gitters für die Steuerung und den Zustand fest.

Alt [3] stellt eine Studie über die Behandlung von nichtlinearen, beschränkten Steuerungsproblemen unter Anwendung der *Euler*-Methode für die Diskretisierung der Differentialgleichung dar und erzielt dabei eine Konvergenzrate mit der Ordnung $\frac{1}{2}$.

Es gibt neben diesen Autoren einige, die sich mit Integrationsverfahren höherer Ordnung beschäftigt haben.

In einer früheren Arbeit studierte *Hager* [51] ein optimales Steuerungsproblem ohne Beschränkung und stellte das Verhältnis zwischen den stetigen dualen Variablen und den entsprechenden Lagrange-Multiplikatoren des diskreten Steuerungsproblems fest. Es wurde beobachtet, daß ein Integrationsverfahren der Ordnung q für Differentialgleichungen nicht immer zu einen Näherungswert der Ordnung q in der optimalen Steuerung führte.

1.3 Motivation und Ziel der Untersuchung

Die Erforschung von Bedingungen für die bestmögliche Konvergenz ist heute noch Gegenstand wissenschaftlicher Untersuchungen. Dabei spielen die verschiedenen Diskretisierungsansätze eine besondere Rolle. Siehe unter anderem das Kollokationsverfahren [105, 119, 10, 54], das Ritzsche Projektionsverfahren [40, 39, 38], das Runge-Kutta-Verfahren bei der Abschätzung des Optimalwertes [117, 118, 106] bzw. der Optimallösung [35, 53, 52], das *Adams-Bashforth*-Verfahren und *Adams-Multon*-Verfahren [25], das *Euler*-Verfahren [74, 31].

Betrachten wir Arbeiten über die Approximation der optimalen Lösung mittels Integrationsverfahren, so stellen wir folgendes fest:

Dontchev und *Hager* [33, 32] stellen eine allgemeine Theorie zur Untersuchung der Lipschitzstetigkeit von parametrisierten Optimierungsproblemen vor. Als eine der Anwendungsbeispiele betrachten die Autoren die Diskretisierung optimaler Steuerungsprobleme mit gewöhnlichen Differentialgleichungen ohne Zustandsbeschränkung und erreichen mit einer optimalen Steuerungsfunktion von beschränkter Variation eine Konvergenzrate mit der Ordnung Eins. Beim Beweis der Stabilität der Optimalitätsbedingungen (vgl. [32], Lemma 11) wird implizit die Steuerungsfunktion als stetig vorausgesetzt.

Malanowski, *Büskens* und *Maurer* [75] betrachten ein Problem mit Zustandsbeschränkung und erzielen das gleiche Ergebnis wie in [32] mit einer stetigen optimalen Steuerung. Die ganze Untersuchung basiert auf einer allgemeinen Approximationstheorie von *Malanowski*. In beiden Arbeiten [32, 75] werden die Steuerung und der Zustand erneut auf demselben Gitter diskretisiert, die Euler-Methode für die Diskretisierung des dynamischen Systems angewendet und eine stückweise konstante Approximation für die Steuerung betrachtet.

Häufig werden für die numerische Lösung von Differentialgleichungen Verfahren mit höherer Konvergenzordnung für die Diskretisierung der Differentialgleichung verwendet. Es wird beispielsweise die Runge-Kutta Diskretisierung in den Arbeiten [104, 109, 119] angewendet. Jedoch wird oft bei der Konvergenzuntersuchung ([31, 32, 75, 72, 26, 15, 27, 81]) nur die *Euler*-Methode angewendet.

Konvergenzresultate für Runge-Kutta-Diskretisierung von optimalen Steuerungsproblemen treten in einer minimalen Anzahl in der Literatur auf.

Schwartz und *Polak* [106] analysieren die Konsistenz der expliziten Runge-Kutta-Approximation. In dieser Arbeit [106] wird eine Konvergenz für die globalen Lösungen des diskreten Problems zur globalen Lösung des stetigen Problems nachgewiesen. Außerdem basiert die Analyse auf Runge-Kutta-Verfahren, deren Koeffizienten alle positiv sind.

Velio analysiert in [117] die Konvergenz des Optimalwertes. Er stellt einen Näherungswert der Ordnung $O(h^2)$ für Steuerungsprobleme mit Beschränkungen und linearer Differentialgleichungen her, ohne eine Voraussetzung über die Regularität der optimalen Steuerung anzunehmen. Dieses Resultat wird später in [118] auf Systeme ausgedehnt, die in Bezug auf Zustandvariable nichtlinear sind.

In [53, 52] werden von *Hager* Bedingungen für die Koeffizienten eines Runge-Kutta-Integrationsverfahrens berechnet, die eine erteilte Ordnung der Genauigkeit in der optimalen Steuerung für Ordnungen bis zu vier sicherstellen. Insbesondere die Arbeit [53] konzentriert

sich auf Runge-Kutta-Diskretisierungen, dessen Koeffizienten in der letzten Berechnungsphase alle positiv sind.

Die Analyse in [53] nutzt die Baumstrukturen („tree-based expansions“ [20]) und die Ordnungsvoraussetzungen, die von Butcher in [20] für gewöhnliche Differentialgleichungen entwickelt wurden, und eine Transformation der notwendigen Optimalitätsbedingungen erster Ordnung für das diskrete Steuerungsproblem.

Diese Transformation führt zu einer Runge-Kutta-Diskretisierung für die adjungierte Gleichung des optimalen Steuerungsproblems, die häufig zu der ursprünglichen Runge-Kutta-Diskretisierung der Zustandsgleichung unterschiedlich ist. Die Diskrepanz zwischen der primären und der dualen Diskretisierung führt zu zusätzlichen Bedingungen, die die Koeffizienten der Runge-Kutta-Diskretisierung erfüllen müssen, um die Dritte- oder Vierte-Ordnungsgenauigkeit im Kontext der optimalen Steuerung zu erzielen. Unter Umständen beobachten wir, daß die Fehlerabschätzung der diskreten Steuerung in (14) [53] nicht die Fehlerabschätzung der diskreten Steuerung in der Lösung des diskreten Problems ausdrückt (siehe Remark 2.2 [53]).

In [35] wird gezeigt, daß jede mögliche Runge-Kutta-Diskretisierung zweiter Ordnung für die Differentialgleichungen durch eine passende Deutung der diskreten Steuerung einen Näherungswert in der optimalen Steuerung von der Ordnung zwei erbringt. Hier wird analog wie in [53] und unter folgender Einschränkung vorgegangen. Die Autoren verringern die Dimension des diskreten Steuerungsraumes, indem sie erfordern, daß die Zwischenschritte der Steuerungskomponenten identisch sind, wenn die Komponenten des dazugehörigen Teilvektors gleich sind (siehe Seite 7 [35] für mehr Details).

Gegenstand dieser Arbeit ist die Untersuchung eines Integrationsverfahrens mit höherer Ordnung zur numerischen Behandlung von optimalen Steuerungsproblemen mit gewöhnlichen Differentialgleichungen. Weiterhin wird ein Algorithmus zur numerischen Lösung solcher Aufgaben unter Berücksichtigung des untersuchten Verfahrens implementiert. Die Ziele sind dabei die Vorstellung einer Diskretisierung mit Verfahren höherer Ordnung, die Herleitung von Voraussetzungen für die Regularitätsbedingungen sowie Optimalitätsbedingungen diskretisierter Aufgaben, die ausführliche Vorstellung einer effektiven Methode für die Berechnung der Ableitungen, die Weiterentwicklung der Ergebnisse aus [32] zur Lipschitzkonvergenz von diskretisierten Aufgaben und eine bessere Annäherung an die Praxis durch die Betrachtung des gleichen Verfahrens bei der Konvergenzuntersuchung und bei der Implementierung des Verfahrens. Dazu betrachten wir:

- zwei Zerlegungen $\mathcal{I}_h, \mathcal{G}_h$ des Intervalls $[0, 1]$ (vgl. Kapitel 3),

- eine allgemeine Parametrisierung der Steuerung,
- ein Integrationsverfahren mit höherer Konsistenzordnung,
- die Methode der internen numerischen Differentiation.

Wir erwarten durch diesen Einsatz:

- eine höhere Konvergenzrate für die Lösung des approximierten Steuerungsproblems,
- eine bessere Verfolgung von Steuerungsfunktionen,
- ein effektives SQP-Verfahren (durch die genaue Berechnung der Gradienten mittels der Methode der internen numerischen Differentiation).

1.4 Inhaltliche Struktur

Inhaltlich ist die Arbeit wie folgt gegliedert:

Das Kapitel 2 beschreibt die Klasse von Aufgaben, die in dieser Arbeit betrachtet wird. Zusätzlich werden Umformulierungsmöglichkeiten für andere Problemstellungen beschrieben.

Das Hauptanliegen im Kapitel 3 ist die Vorstellung einer Diskretisierung von Steuerungsproblemen mit zwei Zerlegungen und einem allgemeinen Einschrittverfahren. Das Kapitel beginnt mit einer kurzen Einführung über die verschiedenen numerischen Optimierungsmethoden: direkte und indirekte Methode. Es werden einige Vor- und Nachteile aufgelistet und die Ergänzung beider Methoden herauskristallisiert.

Es werden die unterschiedlichen Betrachtungsweisen bezüglich der Variablen aus der Literatur vorgestellt und verglichen. Anschließend wird die in dieser Arbeit betrachtete Optimierungsvariable detailliert beschrieben. Zwei Zerlegungen werden eingeführt: ein Steuerungsgitter und ein Zustandsgitter. Bezüglich dem Steuerungsgitter wird ein Ansatzraum für die Steuerung definiert, die Mengen der zulässigen Steuerungen sowie die ursprüngliche Aufgabe entsprechend erneut formuliert. Für die entstandene Aufgabe werden die Voraussetzungen für die Regularität eines Punktes getroffen, die notwendigen Optimalitätsbedingungen erster Ordnung und die Voraussetzungen für die notwendigen Optimalitätsbedingungen zweiter Ordnung eingeführt. Ansätze und Hinweise über die Methode der Parametrisierung der Steuerung werden gegeben. Auf dem Zustandsgitter wird ein allgemeines Einschrittverfahren eingeführt und eine diskretisierte Aufgabe formuliert. Für die diskrete Aufgabe werden die Regularitätsbedingungen eines Punktes beschrieben. Das Kapitel endet mit der Vorstellung

der Methode der internen numerischen Differentiation für die Berechnung der Gradienten diskretisierter Funktionen im Fall eines allgemeinen Einschrittverfahrens. Speziell für das Runge-Kutta-Verfahren werden die Implementierungsschritte für die Berechnung der Gradienten angegeben.

Im Kapitel 4 wird eine Approximationstheorie vorgestellt. Diese Theorie basiert auf der Störungstheorie zur Untersuchung von Optimierungsproblemen in Banach-Räumen und orientiert sich an der Vorgehensweise von *Dontchev* und *Hager* [33, 32]. Es folgen die Formulierung von verallgemeinerten Gleichungen, die Formulierung und der Beweis des Konvergenzsatzes unter Anwendung der vorgestellten Theorie. Anschließend wird der Konvergenzsatz angewendet. Dabei werden sämtliche Voraussetzungen für Runge-Kutta-Diskretisierung verifiziert.

Das Kapitel 5 stellt einen kurzen Überblick über ein Verfahren für die numerische Lösung endlich-dimensionaler Optimierungsaufgaben dar. Dabei geht es um eine Beschreibung des Verfahrens der sukzessiven quadratischen Optimierung (SQP-Verfahren): die Struktur des SQP-Verfahrens, das quadratische Teilproblem, die Bewertungsfunktion, die Approximation der Hesse-Matrix, die Abbruchkriterien.

Die Implementierung eines Algorithmus für die numerische Lösung der diskreten Aufgabe wird im Kapitel 6 beschrieben. In diesem Algorithmus ist die Diskretisierung des Problems mittels explizitem Runge-Kutta-Verfahren für den Zustand und spezieller stückweise linearer Approximation (Abschnitt 3.3.1) für die Steuerung durchgeführt. Der entwickelte Algorithmus wird mit RINDOPT bezeichnet. Die Koeffizienten des eingesetzten Runge-Kutta-Verfahrens werden bestimmt, der Grundaufbau des RINDOPT-Algorithmus wird skizziert. Anschließend werden einige wichtige Komponenten des RINDOPT-Algorithmus beschrieben.

Zum Nachweis der Funktionsfähigkeit des entwickelten Algorithmus werden im Kapitel 7 ein Erläuterungsbeispiel sowie die Testrechnungen für ausgewählte Probleme im Bereich der Luft- und Raumfahrt, Elektrotechnik und Förder- und Transportanlage ausgeführt.

Die im Rahmen dieser Arbeit erreichten Ergebnisse werden im Kapitel 8 zusammengefaßt. Hinweise zu weiterführenden Arbeiten werden ebenfalls hier erläutert.

Kapitel 2

Problemstellung

2.1 Formulierung eines Steuerungsproblems

Im diesem Abschnitt wird auf die zugrundeliegende Klasse von Problemen dieser Arbeit näher eingegangen.

Es seien g_0 eine Funktion von \mathbb{R}^{n+d_0} in \mathbb{R} , f eine beschränkte und Lipschitz-stetige Funktion von \mathbb{R}^{n+r+d_0} in \mathbb{R}^n , U_a eine nichtleere, konvexe und abgeschlossene Menge in \mathbb{R}^r , V_a eine nichtleere konvexe und abgeschlossene Menge in \mathbb{R}^{d_0} .

Wir formulieren ein Steuerungsproblem (\mathbf{P}_*) durch:

$$\min_{x,u,v} g_0(x(1), v) \quad \text{bzgl.} \quad (2.1)$$

$$\dot{x}(t) = f(x(t), u(t), v), \quad 0 \leq t \leq 1, \quad x(0) = x_0(v), \quad (2.2)$$

$$u(t) \in U_a, \quad \text{f.ü. } t \in [0, 1], \quad v \in V_a. \quad (2.3)$$

wobei der Zustand $x(t)$ aus dem Raum $\mathcal{W}^{1,\infty}([0, 1], \mathbb{R}^n)$ ist, die Steuerung $u(t)$ zu dem Raum $\mathcal{L}^\infty([0, 1], \mathbb{R}^r)$ gehört, die Differentialform $\dot{x}(t)$ aus dem Raum $\mathcal{L}^\infty([0, 1], \mathbb{R}^n)$ ist, der Parameter v zu dem Raum \mathbb{R}^{d_0} gehört und die Startbedingung des Zustandes $x_0(v)$ eine stetig differenzierbare Funktion von \mathbb{R}^{d_0} in \mathbb{R}^n ist. Der Parameter v ist von $x(t)$ und $u(t)$ für $t \in [0, 1]$ unabhängig. Wichtig an dieser Stelle ist zu vermerken, daß in der Aufgabe (2.1)-(2.3) die Zielfunktion $g_0(x(1), v)$ als eine explizite Funktion der Steuerung u zu betrachten ist. Der Zustand x ist als implizite Funktion von u zu betrachten und nach dem dynamischen System (2.2) zu berechnen. Für die bevorstehende Untersuchung setzen wir folgendes voraus:

A 1 : *Es existiere eine Lösung (x^*, u^*, v^*) der Aufgabe (\mathbf{P}_*).*

Bemerkung 1 Der Parameter v darf nicht als Störungsparameter des optimalen Steuerungsproblems (vgl. Maurer, Augustin [4]) betrachtet werden. Der Parameter v in dieser Arbeit ist eine Optimierungsvariable und kann beispielsweise bei Problemen mit freier Anfangs- bzw. Endzeit diese Randdaten beinhalten (siehe Abschnitt 2.2).

2.2 Umformulierung anderer Problemstellungen

Optimale Steuerungsprobleme mit Zielfunktionen in Integralform können durch Hinzufügen einer neuen Zustandsvariablen in die Form von (2.1) umgewandelt werden. Es sei g eine integrierbare Funktion. Bei der Minimierung einer Zielfunktion der Gestalt:

$$\min_{x,u,v} \int_0^1 g(x(t), u(t), v) dt,$$

definieren wir mit Hilfe der Differentialgleichung:

$$\dot{x}_{n+1}(t) = g(\bar{x}(t), u(t), v), \quad x_{n+1}(0) = 0$$

eine zusätzliche Zustandsvariable x_{n+1} , wobei $\bar{x} = (x, x_{n+1})$. Damit geht die Minimierungsaufgabe in die Form:

$$\min_{x,u,v} \bar{x}_{n+1}(1)$$

über.

Mit den Einsätzen $\dot{x}_{n+1}(t) = 1$, $x_{n+1}(t) = t$, und $x_{n+1}(0) = 0$ verwandeln wir nichtautonome Differentialgleichungen in die autonome Form (2.2).

Aufgaben mit freier Anfangs- bzw. Endzeit werden durch die Einführung einer neuen Zeitfunktion in Aufgaben mit fester Anfangs- bzw. Endzeit überführt. Es sei t_o bzw. t_e die freie Anfangs- bzw. Endzeit, und s_o bzw. s_e die fest gegebene Anfangs- bzw. Endzeit und eine Zeitfunktion gegeben durch:

$$t = t(s) := t_o + (s - s_o) \frac{t_e - t_o}{s_e - s_o}, \quad s \in [s_o, s_e].$$

Die ursprünglichen freien Anfangs- und Endzeiten können bei der Untersuchung in die Parameterliste von v aufgenommen werden. Es entstehen die folgenden neuen Optimierungsvariablen:

$$\tilde{x}(s) := x(t(s)), \quad \tilde{u}(s) := u(t(s)), \quad \tilde{v} = (v, t_o, t_e).$$

Durch die Zeittransformation ändert sich die Differentialgleichung (2.2) folgendermaßen um:

$$\dot{\tilde{x}}(s) = \frac{t_e - t_o}{s_e - s_o} f(\tilde{x}(s), \tilde{u}(s), \tilde{v}), \quad s \in [s_o, s_e], \quad \tilde{x}(s_o) = x(t_o) = x_0(\tilde{v}).$$

Für tiefergehende Betrachtungen über optimale Steuerungsprobleme sei auf [112, 1, 21] verwiesen.

Bemerkung 2 : *Die Überführung von Optimierungsproblemen in die Form von (2.1)-(2.3) hat den Vorteil, daß im Fall einer Diskretisierung die gleiche Diskretisierungsmethode für die Zielfunktion und für die Beschränkungen eingesetzt wird.*

Kapitel 3

Diskretisierung

3.1 Numerische Optimierungsmethoden

Auf Grund der Komplexität von optimalen Steuerungsproblemen werden diese Aufgaben gewöhnlich nur numerisch gelöst. Die Methoden für die numerische Lösung von optimalen Steuerungsproblemen werden in der Literatur in zwei Gruppen unterteilt: die direkte und die indirekte Methode [93]. Bei der indirekten Methode wird das Steuerungsproblem als nicht-lineare Randwertaufgabe betrachtet, welche aus den notwendigen Optimalitätsbedingungen der ursprünglichen Aufgabe entstanden sind. Die direkte Methode dagegen behandelt das Steuerungsproblem sofort als Optimierungsproblem in einem funktionalen Raum.

Durch die Anwendung des mehrfachen Schießverfahrens („multiple shooting technique“) für die Lösung von nichtlinearen Randwertaufgaben hat sich die indirekte Methode als sehr effektiv erwiesen (vgl. [13, 16, 59]). Der Nachteil bei der indirekten Methode ist die Aufforderung bei jedem Optimierungsproblem einiger Kenntnisse über das Verhalten des Problems (z.B. die adjungierte Variablen, die Hamilton-Funktion, das Maximum-Prinzip). Außerdem verlangt die Konvergenz des mehrfachen Schießverfahrens eine gute Startlösung, die jedoch vor dem Start unbekannt ist. Die direkte Methode benötigt keine Information über die adjungierten Variablen und erfordert minimale Kenntnisse über das Optimierungsproblem und demzufolge kaum Vorbereitungen.

In dieser Arbeit wird die direkte Methode betrachtet. Diese kann als Vorbereitung für die indirekte Methode zur Berechnung des Startwertes angesehen werden (vgl. [121]). Die direkte Methode ist eine Zusammensetzung aus einer endlich-dimensionalen Approximation und einem Optimierungsalgorithmus.

Die endlich-dimensionale Approximation verzweigt sich in eine endlich-dimensionale Approximation der Steuerung (falls die Steuerung als einzige Optimierungsvariable zu betrachten ist) und eine endlich-dimensionale Approximation der Steuerung und des Zustandes (falls die Steuerung und der Zustand als Optimierungsvariable zu betrachten sind). Bevor wir zur eigentlichen Approximation des originalen Problems (2.1)-(2.3) kommen, wird zuerst ein Vergleich beider Betrachtungsweisen der Optimierungsvariablen dargestellt.

3.2 Optimierungsvariable

Ein Weg für die endlich-dimensionale Approximation der Steuerung ist die Methode der Elimination des Zustandes (vgl. [93, 58]). Dieser Weg setzt die Existenz eines Transformationsoperators voraus, welcher die Zustandsvariable nach der Steuerungsvariable ausdrückt und anschließend in die Anfangswertaufgabe (2.2) einsetzt. Es entsteht ein Gleichungssystem, das diskretisiert wird und das nur von der Steuerungsvariable abhängig ist.

Ein anderer Weg für die endlich-dimensionale Approximation der Steuerung ist die a priori Parametrisierung der Steuerung (vgl. [48, 112, 66, 95]). Der Zustand, betrachtet als Funktion der Steuerung, wird dann rekursiv durch Integrationsverfahren oder Kollokationsverfahren berechnet. Dadurch verschwindet die Zustandsvariable aus der Liste der zu optimierenden Variablen. Es entsteht eine Optimierungsaufgabe bezüglich der Steuerung.

Durch die endlich-dimensionale Approximation der Steuerung entsteht ein diskretes Problem mit geringer Anzahl von Optimierungsvariablen, mit vollbesetzter Jacobi- und Hesse-Matrix. Mit dieser Methode wurde bereits gute Resultate erzielt (vgl. [19]).

Die Methode der Approximation der Steuerung und des Zustandes betrachtet beide Variablen als Optimierungsvariablen (vgl. [3, 25, 26, 32, 37, 70]).

Diese Methode erzeugt zwar ein großdimensioniertes Optimierungsproblem, hat aber den Vorteil, daß die Jacobi- und Hesse-Matrizen im allgemeinen schwachbesetzt sind. Durch die Anwendung von Methoden wie z.B. der Methode des strukturierten SQP-Verfahrens [108] oder die Ausnutzung der Schwachbesetztheit der Matrizen mit der „direct transcription“-Methode [7, 8], sind ebenfalls gute Resultate erzielt worden.

In den kommenden Abschnitten wird eine Diskretisierung durchgeführt. Dabei wird die Steuerung durch die Methode der Parametrisierung approximiert und der Zustand mit Hilfe eines allgemeinen Einschrittverfahrens angenähert.

3.3 Parametrisierung des Steuerungsraumes

Die Methode der Parametrisierung ermöglicht die Darstellung der Steuerungsfunktion auf einer endlichen Anzahl von Teilintervallen bei gegebenen Basisfunktionen (u.a. konstante Funktionen, Polynome). Die Basisfunktionen hängen ihrerseits von einer endlichen Menge von Parametern (z.B. Randwerten der Steuerung, Steuerungsgitter) ab.

Es sei $L \in \mathbb{N} \setminus \{0\}$ gegeben. In dieser Arbeit beruht die Parametrisierung der Steuerung auf dem Ersetzen der Menge der zulässigen Steuerungen:

$$\{u \in \mathcal{U} : u(t) \in U_a \quad \text{f.ü.} \quad t \in [0, 1]\}, \quad (3.1)$$

durch eine Menge gegeben durch:

$$\{u(\cdot, w) \in \mathcal{U} : u(t, w) \in U_a, t \in [0, 1], w \in \mathbb{R}^L\}. \quad (3.2)$$

Wir bezeichnen mit W die Teilmenge von \mathbb{R}^L , gebildet aus allen Parametern w , die die Menge (3.2) beschreiben. Das heißt

$$W := \{w \in \mathbb{R}^L : u(t, w) \in U_a \quad t \in [0, 1]\}. \quad (3.3)$$

Durch die Parametrisierung des Steuerungsraumes ist jedes Element $u(\cdot, w)$ durch den endlich-dimensionalen Parameter $w \in W$ eindeutig bestimmt.

Für die Elemente $u(\cdot, w_1)$ und $u(\cdot, w_2)$ mit $w_1, w_2 \in W$ folgt aus der Konvexitätseigenschaft von U_a die Zugehörigkeit der Verbindungsstrecke ¹ $\overline{u(t, w_1)u(t, w_2)}$ von Elementen $u(t, w_1)$ und $u(t, w_2)$ in die Menge U_a . Es bleibt offen, Aussagen über die Zusammenhänge zwischen der Verbindungsstrecke $\overline{w_1 w_2}$ der Elemente w_1 und w_2 und der Menge W zu treffen, solange die Funktion $u(\cdot, w)$ nicht eindeutig definiert ist.

Ansatzraum

Es sei $\mathcal{U}_w \subset \mathcal{L}^\infty([0, 1], \mathbb{R}^r)$ die Menge aller Elemente, die sich nach dem Modell in (3.2) mit $w \in \mathbb{R}^L$ darstellen lassen. Der Raum \mathcal{U}_w , ausgestattet mit der Norm des übergeordneten Raumes \mathcal{U} , ist wiederum linear normiert.

Im Vergleich zu den allgemeinen Optimierungsproblemen haben optimale Steuerungsprobleme die besondere Charakteristik, daß die Steuerungsvariablen Sprünge aufweisen

¹ Für zwei Elemente w_1, w_2 aus der Menge W heißt die Menge:

$$\overline{w_1 w_2} := \{w = \alpha w_1 + \beta w_2 : \alpha, \beta \text{ reell, } \alpha \geq 0, \beta \geq 0, \alpha + \beta = 1\}$$

Verbindungsstrecke der Elemente w_1 und w_2 .

können. Um solche Aufgaben sinnvoll behandeln zu können, ist es notwendig, bei der Parametrisierung des Steuerungsraumes, die eventuell auftretenden Sprünge der Steuerung mit zu berücksichtigen. Ein möglicher Methode zur Bewältigung dieses Problems ist die Einführung von zwei Gittern bei der Diskretisierung, dem Steuerungs- und dem Zustandsgitter.

Es sei $M \in \mathbb{N} \setminus \{0\}$ gegeben. Wir betrachten eine feste beliebige Zerlegung des Intervalls $I = [0, 1]$, gegeben durch:

$$\mathcal{I}_h := \{\tau_0, \tau_1, \tau_2, \dots, \tau_{M-1}, \tau_M\}, \quad \tau_0 := 0 < \tau_1 < \tau_2 < \dots < \tau_{M-1} < \tau_M := 1. \quad (3.4)$$

wobei der Index h durch:

$$h := \max_{1 \leq i \leq M} (\tau_i - \tau_{i-1}), \quad (3.5)$$

definiert ist. Der Index h wird im Abschnitt 4.2 ab Seite 38 als Parameter betrachtet. Die Menge \mathcal{I}_h bezeichnet das **Steuerungsgitter** auf dem Intervall I .

3.3.1 Ansatz der Parametrisierung der Steuerung

Dieser Abschnitt dient zur Formulierung des zugrundeliegenden Parameters dieser Arbeit.

Es seien die Zahl M , das Intervall I , das Gitter \mathcal{I}_h und die Schrittweite h gegeben wie in der Definition von (3.4) und (3.5). Wir betrachten eine Menge W gegeben durch:

$$W := \prod_{j=1}^M (U_a \times U_a).$$

Die numerischen Experimente dieser Arbeit verwenden einen Parameter w mit der Darstellung:

$$w := (u_1^+, u_1^-, u_2^+, u_2^-, \dots, u_M^+, u_M^-) \in W. \quad (3.6)$$

Dabei bezeichnet u_j^+ bzw. u_j^- den rechten Grenzwert im Punkt τ_{j-1} bzw. linken Grenzwert im Punkt τ_j ($j = 1, \dots, M$) der Funktion $u(\cdot, w)$, mit $u(\tau_j, w) = u_j^+$, $j = 0, \dots, M-1$ und $u(\tau_M, w) = u_M^-$. Für die Approximation $u(\cdot, w) \in \mathcal{U}$ betrachten wir eine spezielle stückweise linear stetige Funktion der Gestalt:

$$u(t, w) := \sum_{j=1}^M \frac{1}{\tau_j - \tau_{j-1}} (u_j^-(t - \tau_{j-1}) + u_j^+(\tau_j - t)) \chi_j(t), \quad t \in [0, 1], \quad (3.7)$$

wobei χ_j , $j = 1, \dots, M$, die charakteristischen Funktionen auf den Teilintervallen $[\tau_{j-1}, \tau_j)$, $j = 1, \dots, M-1$ und $[\tau_{M-1}, \tau_M]$ sind. Die Menge W ist als Produkt von endlich vielen konvexen Mengen ebenfalls eine konvexe Menge.

Bemerkung 3 Die Besonderheit für die Approximation der Steuerung durch die Formel (3.7) ist die Flexibilität der stückweise linearen Approximation. Sie ermöglicht die Widerspiegelung von sprunghaften Funktionsverläufen ebenso wie die Nachbildung von stetigen Funktionsanteilen. Zum Beispiel, wenn es bekannt ist, daß die gesuchte Funktion auf einigen Teilintervallen einfache Gestalt hat und auf anderen Teilintervallen eine komplizierte Gestalt hat, dann kann die Funktion auf den Teilintervallen separat mit Funktionen $u(\cdot, w)$ aus $\mathbb{P}_{\mu, \mathcal{I}_h, \eta}$ besser approximiert werden.

3.3.2 Andere Parametrisierungsmethoden der Steuerung

Parametrisierung durch stückweise konstante Funktionen

Es seien erneut M , I , \mathcal{I}_h und h gegeben als (3.4) und (3.5). Wir betrachten eine Menge W und die Funktion $u(\cdot, w)$ gegeben durch:

$$W := \prod_{j=0}^M U_a, \quad L = (M+1)r, \quad \text{und } w := (u_0, \dots, u_M),$$

$$u(t, w) := \sum_{j=1}^M u_j \chi_j(t), \quad t \in [0, 1]. \quad (3.8)$$

Die charakteristischen Funktionen χ_j , $j = 1, \dots, M$ sind bereits im Abschnitt 3.3.1 beschrieben. Die Funktion $u(\cdot, w)$, beschrieben durch (3.8), ist eine stückweise konstante Funktion auf dem Steuerungsgitter \mathcal{I}_h und die Menge W ist konvex.

Parametrisierung durch B-Spline

Ein anderer möglicher Ansatz für die Diskretisierung der Steuerung durch die Methode der Parametrisierung besteht in der Approximation der Steuerung durch B-Spline [110, 111]. Es seien M , $\mu \in \mathbb{N} \setminus \{0\}$. Wir nehmen an: es existieren 2μ zusätzliche Gitterpunkte $\tau_{-\mu} < \tau_{-\mu+1} < \dots < \tau_0$ und $\tau_M < \tau_{M+1} < \dots < \tau_{M+\mu}$. Wir bezeichnen mit $B_{\mu, j}$ den j -ten B-Spline der Ordnung μ gegeben durch:

$$B_{\mu, j}(t) := (\tau_{j+\mu+1} - \tau_j) [\tau_j, \tau_{j+1}, \dots, \tau_{j+\mu+1}] [\cdot - t]_+^\mu, \quad t \in \mathbb{R}, \quad j = -\mu, \dots, M-1,$$

wobei $[\tau_j, \tau_{j+1}, \dots, \tau_{j+\mu+1}]$ die $(\mu+1)$ -te dividierte Differenz („divided difference“) [28] ist und $[x]_+ := \max\{x, 0\}$.

Der Parameter w und die stetige Approximation der Steuerung $u(\cdot, w)$ werden folgendermaßen definiert:

$$w := (u_{-\mu}, \dots, u_{M-1}) \in W, \quad W \subset \mathbb{R}^{(M+\mu)r},$$

$$u(\cdot, w) := \sum_{j=-\mu}^{M-1} u_j B_{\mu, j}(\cdot). \quad (3.9)$$

Die Ordnung μ wird je nach Glätte der Steuerungsfunktion entsprechend festgelegt. Der Einsatz von B-Spline führt zu einem großen Optimierungsproblem. B-Spline bilden eine Basis für den Raum $\mathcal{P}_{\mu, \mathcal{I}_h, \eta}$ und können mittels der Rekursionsformel ([28], Satz 6c) effektiv berechnet werden. Für jedes $w \in W$ folgt aus der Nichtnegativität der B-Spline und der Eigenschaft $\sum_{j=-\mu}^{M-1} B_{\mu, j}(t) = 1$ die Zugehörigkeit von $u(t, w)$ zu U_a (vgl. [28], Proposition 6b).

Hinweise und weitere Beschreibungen über die Methode der Parametrisierung der Steuerung sind auch in den Arbeiten [5, 6, 116, 112, 48, 49, 66, 95]) nachzulesen.

Durch die Parametrisierung des Steuerungsraumes geht die Aufgabe (\mathbf{P}_*) , definiert auf dem Raum $\mathcal{X} \times \mathcal{U} \times \mathbb{R}^{d_0}$, in die Aufgabe (\mathbf{P}_w) definiert auf dem Raum $\mathcal{X} \times \mathbb{R}^L \times \mathbb{R}^{d_0}$ über, mit der Darstellung:

$$\min_{x, w, v} g_0(x(1), v) \quad \text{bzgl.} \quad (3.10)$$

$$\dot{x}(t) = f(x(t), u(t, w), v), \quad 0 \leq t \leq 1, \quad x(0) = x_0(v), \quad (3.11)$$

$$u(t, w) \in U_a, \quad t \in [0, 1], \quad v \in V_a, \quad (3.12)$$

wobei $x \in \mathcal{X}$ und $u(\cdot, w) \in \mathcal{U}_w$.

Parametrisierungsforderung

Es sei die k -te Differenz („ k th difference“) der Funktion u mit der Schrittweite h im Punkt t definiert durch:

$$\Delta_h^k u(t) := \sum_{i=0}^k (-1)^{i+k} \binom{k}{i} u(t + ih), \quad \Delta_h^1 u(t) := \Delta_h u(t) = u(t + h) - u(t),$$

wobei

$$\binom{k}{i} = \frac{k!}{i!(k-i)!}, \quad i = 0, \dots, k$$

die Binomialkoeffizienten sind.

Es sei $\omega_k(u, t; \gamma)$ das lokale Stetigkeitsmodul der Funktion u von der Ordnung k im Punkt $t \in [0, 1]$ definiert durch:

$$\omega_k(u, t; \gamma) := \sup\{|\Delta_h^k u(s)| : s, s + kh \in [t - \frac{k\gamma}{2}, t + \frac{k\gamma}{2}] \cap [0, 1]\} \text{ für } \gamma \in [0, \frac{1}{k}],$$

und $\tau_k(u; \gamma)$ das mittlere Stetigkeitsmodul der Funktion u von der Ordnung k gegeben durch:

$$\tau_k(u; \gamma) := \|\omega_k(u, \cdot; \gamma)\|_{\mathcal{L}^e} = \left[\int_0^1 (\omega_k(u, t; \gamma))^e dt \right]^{\frac{1}{e}}, \quad \gamma \in [0, \frac{1}{k}].$$

Es seien $d(A, B)_{\mathcal{L}^1} := \inf_{a \in A, b \in B} \|a - b\|_{\mathcal{L}^1}$, $\overset{1}{V}(u)$ die Variation ² der Funktion u auf dem Intervall $[0, 1]$. Wir nehmen an:

A 2 : Es existiere eine Funktion $\epsilon : [0, 1] \rightarrow \mathbb{R}_+$, mit $\lim_{h \rightarrow 0} \epsilon(h) = 0$, so daß

$$d(u^*, \mathcal{U}_w)_{\mathcal{L}^1} \leq \epsilon(h). \quad (3.13)$$

Die Ungleichung (3.13) setzt voraus, daß es ein Element $w^o \in W$ mit $u(\cdot, w^o) \in \mathcal{U}_w$ gibt, das gen

ügend nahe an der gesuchten optimalen Lösung liegt. Die Menge V_a ist von der Parametrisierung unabhängig und bleibt während des ganzen Parametrisierungsprozesses unverändert. Damit kann v^o der entsprechende Parameter zu w^o gleich v^* gesetzt werden.

Realisierbarkeit von (3.13)

Die Frage der Realisierbarkeit von (3.13) besteht in der Untersuchung für welche Funktionen die Bedingung (3.13) erfüllt werden kann.

Für u^* aus der Menge der beschränkten meßbaren Funktionen auf dem Intervall $[0, 1]$ ist mit Hilfe von Spline-Funktionen (siehe Abschnitt 3.3.2) ersten Grades eine Abschätzung der Art:

$$d(u^*, \mathcal{U}_w)_{\mathcal{L}^1} = \tau_2(u^*; h)$$

erreichbar (siehe Theorem 4.4 [107]).

Falls zusätzlich die Ableitung ∇u^* existiert und eine Funktion von beschränkter Variation ist, dann folgt aus den Eigenschaften (4); Seite 8 und (7); Seite 10 in [107] die Abschätzung:

$$\tau_2(u^*; h) \leq h\tau_1(\nabla u^*; h) \leq h^2 \overset{1}{V}(\nabla u^*).$$

Unter der Voraussetzung, daß die k -te Ableitung $\nabla^{(k)} u^*$ der Funktion u^* existiert und von beschränkter Variation sei, kann mit Hilfe von Spline höherer Ordnung die Konvergenzgeschwindigkeit von (3.13) verbessert werden. Es folgen aus Korollar 4.7 und 4.11 in [107] folgende Abschätzungen:

$$\begin{aligned} d(u^*, \mathcal{U}_w)_{\mathcal{L}^1} &= O(h^2 \overset{1}{V}(\nabla^{(2)} u^*)), \quad k = 2, \quad \text{für quadratische Spline,} \\ d(u^*, \mathcal{U}_w)_{\mathcal{L}^1} &= O(h^3 \overset{1}{V}(\nabla^{(3)} u^*)), \quad k = 3, \quad \text{für kubische Spline.} \end{aligned}$$

² Die Variation (oder die vollständige Variation) der Funktion u auf dem Intervall $[0, 1]$ ist die obere Grenze der Summe $\sum_{k=0}^M \|u(\tau_{k+1}) - u(\tau_k)\|$ für eine beliebige Zahlenfolge $\{\tau_k\}$ mit $\tau_0 := 0 < \tau_1 < \tau_2 < \dots < \tau_{M-1} < \tau_M := 1$.

3.4 Überführung des Steuerungsproblems in ein Optimierungsproblem

Optimale Steuerungsprobleme sind Optimierungsprobleme in Banach-Räumen. Wir wollen in diesem Abschnitt das Problem (P_w) als Optimierungsproblem im Banach-Raum formulieren. Außerdem werden hinreichende Regularitätsbedingungen eines Punktes untersucht, das Phänomen der Zweinorm-Diskrepanz erläutert und das Minimum-Prinzip formuliert. Wir betrachten zwei Banach-Räume \mathcal{Z} und \mathcal{Y} , gegeben durch:

$$\begin{aligned}\mathcal{Z} &= \mathcal{X} \times \mathbb{R}^L \times \mathbb{R}^{d_o} \text{ mit } \mathcal{X} := \mathcal{W}^{1,\infty}([0, 1], \mathbb{R}^n), \\ \mathcal{Y} &= \mathcal{L}^\infty([0, 1], \mathbb{R}^n) \times \mathbb{R}^n,\end{aligned}$$

und versehen sie mit den Normen:

$$\begin{aligned}\|(x, w, v)\|_{\mathcal{Z}} &= \|x\|_{\mathcal{X}} + \|w\|_{\mathbb{R}^L} + \|v\|_{\mathbb{R}^{d_o}}, \\ \|(a, b)\|_{\mathcal{Y}} &= \|a\|_{\mathcal{L}^\infty} + \|b\|_{\mathbb{R}^n},\end{aligned}$$

wobei $(x, w, v) \in \mathcal{Z}$ und $(a, b) \in \mathcal{Y}$ gilt.

Es seien $\mathcal{K} := \{0\}$, die Menge W konvex und abgeschlossen und die Menge \mathcal{C} der zulässigen Tupel (x, w, v) , definiert durch:

$$\mathcal{C} := \{(x, w, v) \in \mathcal{X} \times \mathbb{R}^L \times \mathbb{R}^{d_o} \text{ mit } x \in \mathcal{X}, w \in W, v \in V_a\}.$$

Wir definieren im Punkt $z = (x, w, v)$ die Abbildungen F von \mathcal{Z} in \mathbb{R} und G von \mathcal{Z} in \mathcal{Y} durch:

$$\begin{aligned}F(z) &:= g_o(x(1), v), \\ G(z) &:= (\dot{x}(\cdot) - f(x(\cdot), u(\cdot, w), v), x(0) - x_0(v))\end{aligned}$$

mit $G(z) = 0$. Nach den obigen Vereinbarungen kann das Steuerungsproblem (P_w) in das folgende Optimierungsproblem überführt werden:

$$\min F(z), \quad z \in \mathcal{C}, \quad G(z) \in \mathcal{K}. \quad (3.14)$$

Um Zugang zu den Resultaten der Optimierungstheorie zu bekommen, nehmen wir an: Die Menge \mathcal{K} sei ein konvexer abgeschlossener Kegel in \mathcal{Y} . Für die Aufgabe (3.10)-(3.12) ist die Konvexität und die Abgeschlossenheit des Kegels leicht zu überprüfen, da der Kegel \mathcal{K} aus einem einzigen Element 0 besteht. Diese Voraussetzungen sind wichtig, denn allein die Verletzung der Abgeschlossenheit des Kegels kann schon die Stabilität der Lösungsmenge gefährden (vgl. Beispiel in [65]).

3.5 Notwendige stetige Optimalitätsbedingungen

In endlich-dimensionalen Räumen folgt aus den Beweistechniken der notwendigen und hinreichenden Optimalitätsbedingungen, daß sie sich nur durch das Ungleichungszeichen \geq (für die notwendige Bedingung) und $>$ (für die hinreichende Bedingung) unterscheiden (vgl. [76]). Grund dafür ist folgendes: Im Beweis wird die Kompaktheit der Einheitskugel im endlich-dimensionalen Raum ausgenutzt. Somit haben die notwendigen und hinreichenden Optimalitätsaussagen im endlich-dimensionalen Räumen die gleichen Voraussetzungen und die gleichen Schlußfolgerungen.

In unendlich-dimensionalen Räumen ist eine analoge Untersuchung der notwendigen und hinreichenden Bedingungen ähnlich zum endlich-dimensionalen Fall unmöglich. Allein für reflexive³ unendlich-dimensionale Räume ist die Einheitskugel schwach kompakt⁴ und ihr Rand in einer schwachen Topologie⁵ im allgemeinen nicht abgeschlossen. Daher ist die obige Art und Weise des Übergangs von notwendig zu hinreichend im unendlich-dimensionalen Raum nicht möglich (vgl. Gegenbeispiel in [79], Seite 105). In der Literatur werden an dieser Stelle zusätzliche Bedingungen gestellt. In [62], Seite 249-279, beschränken sich die Autoren auf Aufgaben mit Gleichungsrestriktionen und geben eine andere Beweistechnik (direkte Beweismethode) für die hinreichende Bedingung zweiter Ordnung an. Diese Arbeit betrachtet eine Aufgabe mit Gleichungsrestriktionen. Daher ist der Beweis der folgenden notwendigen stetigen Optimalitätsbedingungen auf den Beweis in [62] zurückzuführen.

Im folgenden wollen wir den Index bei den Normen in einigen Fällen weglassen, falls dies zu keiner Verwechslung führt. Wir setzen folgende Differenzierbarkeitsbedingungen voraus:

³ Es sei X ein normierter Raum. X ist reflexiv, falls $X = (X^*)^*$ gilt.

⁴ Eine Folge $\{x^n\}$ in einem normierten Raum konvergiert schwach gegen \bar{x} , falls $\lim_{n \rightarrow \infty} \|x^+(x^n - \bar{x})\| = 0$ für alle $x^+ \in X^*$ gilt.

Eine Menge M aus dem normierten Raum X heißt schwach kompakt, falls für jede Folge $\{x^n\} \subset M$, eine gegen \bar{x} schwach konvergierte Teilfolge besitzt, mit $\bar{x} \in M$.

⁵ Eine Menge M aus dem normierten Raum X heißt schwach abgeschlossen, falls für jede Folge $\{x^n\} \subset M$, die schwach gegen \bar{x} konvergiert, $\bar{x} \in M$ gilt.

A 3 : Es existieren eine positive Zahl δ_* und eine konvexe und abgeschlossene Menge $\mathcal{C} \subset \mathcal{X} \times \mathbb{R}^L \times \mathbb{R}^{d_0}$ mit den Eigenschaften:

- (a) $z^o := (x^o, w^o, v^o) \in \mathcal{C}$ und ist Lösung von (\mathbf{P}_w) ,
- (b) die Funktion f und g_0 sind zweimal stetig Fréchet-differenzierbar,
- (c) $d(z^o, \partial\mathcal{C}) \geq \delta_*$,

wobei $\partial\mathcal{C}$ den Rand von \mathcal{C} bezeichnet.

3.5.1 Phänomen der Zweinorm-Diskrepanz bei Steuerungsproblemen

Bei vielen Optimierungsproblemen (z.B. optimalen Steuerungsproblemen, Variationsgleichungen) können die notwendigen Bedingungen erster und zweiter Ordnung für die Existenz eines Optimums in der Norm, mit welcher die Aufgabe definiert ist, nicht immer erfüllt werden. Grund dafür ist die Zweinorm-Diskrepanz. Zur Erläuterung dieses Phänomens orientieren wir uns an der Vorgehensweise in [76]. Wir bezeichnen mit g eine der Funktionen F, G aus der Aufgabe (3.14) oder die Funktion \mathcal{L} aus (3.21). Es sei ein Punkt $z = (x, u(\cdot, w), v)$ gestört von $\varepsilon = (\varepsilon_x, u(\cdot, \varepsilon_w), \varepsilon_v)$. Wir betrachten im Punkt $z + \varepsilon$ folgende Taylorentwicklungen:

$$\begin{aligned} g(z + \varepsilon) &= g(z) + \nabla_z g(z)\varepsilon + r_1(\varepsilon; g), \\ g(z + \varepsilon) &= g(z) + \nabla_z g(z)\varepsilon + \varepsilon^\top \nabla_{zz}^2 g(z)\varepsilon + r_2(\varepsilon; g), \end{aligned}$$

wobei $r_i(\varepsilon; g)$, $i = 1, 2$, die Restglieder der Taylorentwicklungen sind. Die Funktion g ist in \mathcal{Y} differenzierbar, und es gilt:

$$\|r_i(\varepsilon; g)\| = o(\|u(\cdot, \varepsilon_w)\|_{\mathcal{L}^\infty} + \|\varepsilon_v\|)^i, \quad i = 1, 2.$$

Andererseits gelten die notwendigen Bedingungen für die Existenz eines Optimums nur in der schwachen \mathcal{L}^p -Norm ($p=1,2$), wo die Gleichungen :

$$\|r_i(\varepsilon; g)\| = o(\|u(\cdot, \varepsilon_w)\|_{\mathcal{L}^p} + \|\varepsilon_v\|)^i, \quad i = 1, 2,$$

nicht immer erfüllt sind.

Das Problem der Zweinorm-Diskrepanz wird in der Literatur meistens durch die Einführung von zwei Konvergenzen (vgl. [31]) oder durch die Einführung von zwei Normen (vgl. [71, 76, 77, 61]) gelöst.

In [71] werden Einbettungsvoraussetzungen (vgl. Gleichung (2.2) in [71]) eingeführt, Lemma 2.1 in [71] bewiesen und die Grenzwerte (3.5a)-(3.5c) und (3.7a)-(3.7c) in [71] vorausgesetzt.

In [76] werden für die Lösung dieses Problems drei Grenzwertvoraussetzungen mit Hilfe von zwei Normen (vgl. Seite 166 Bedingungen **C1 - C3** in [76]) eingeführt. In [61] wird das Problem durch die Einführung von zwei Normen bewältigt.

In der vorliegenden Arbeit wird das Problem der Zweinorm-Diskrepanz durch die Einführung von zwei Normen und den Vergleich beider Normen (vgl. Abschnitt 4.3.1 ab Seite 47) gelöst.

3.5.2 Hinreichende Regularitätsbedingungen eines Punktes

Wir wollen im folgenden, Regularitätsbedingungen für die Existenz der Lagrange-Multiplikatoren (vgl. Theorem 3.1 in [125]) definieren und bezüglich der Aufgabe (\mathbf{P}_w) die Regularität eines Punktes diskutieren.

Es sei $\bar{z} := (\bar{x}, \bar{w}, \bar{v})$ ein Element aus \mathcal{C} . Wir definieren die Normalkegelabbildung $\mathcal{N}_c[\mathcal{C}]$ von der Menge \mathcal{C} in die Menge $\mathcal{C}^* = \mathcal{X}^* \times \mathbb{R}^L \times \mathbb{R}^{d_0}$ durch:

$$\mathcal{N}_c[\mathcal{C}](\bar{z}) := \begin{cases} \left\{ \begin{array}{l} y = (y^x, y^w, y^v) \in \mathcal{X}^* \times \mathbb{R}^L \times \mathbb{R}^{d_0} : \\ y^x(x - \bar{x}) + \langle y^w, w - \bar{w} \rangle + \langle y^v, v - \bar{v} \rangle \geq 0, \\ \forall z \in \mathcal{C} \end{array} \right\} & : \bar{z} \in \mathcal{C}, \\ \emptyset & : \bar{z} \notin \mathcal{C}. \end{cases}$$

Definition 1 Der Punkt $z^o := (x^o, w^o, v^o)$ heißt regulär, falls für alle $a \in \mathcal{C}$, das einzige Element a , das die Gleichung

$$-a^\top \nabla_z G(z^o) \in \mathcal{N}_c[\mathcal{C}](z^o) \quad (3.15)$$

genügt, ist $a = 0$.

Eine ähnliche Formulierung der Regularität eines Punktes (sog. „constraint qualification“) ist in [91], 6.15 Korollar zu finden.

Im allgemeinen sind die Restriktionen bei Optimierungsproblemen durch endlich viele Gleichungen und Ungleichungen beschrieben. Falls dies für das Problem (\mathbf{P}_w) zutrifft und die Funktionen, die die endlich vielen Gleichungen und Ungleichungen beschreiben, stetig Fréchet-differenzierbar sind, dann ist die Gleichung (3.15) ein Ausdruck der „Mangasarian-Fromowitz Constraint Qualification“, das heißt:

- die Gradienten der Gleichungsrestriktionen im Punkt (x^o, w^o, v^o) sind linear unabhängig,

- Es existiere ein Vektor $b \in \mathcal{X}^* \times \mathbb{R}^{L+d_o}$ so daß:
 - das Skalarprodukt der Ableitung jeder aktiven Ungleichungsrestriktion im Punkt z^o mit b eine strikt negative (bei negativen Ungleichungen) oder eine strikt positive (bei positiven Ungleichungen) Zahl ergibt,
 - das Skalarprodukt der Ableitung jeder Gleichungsrestriktion im Punkt z^o mit b gleich Null ist.

Hinreichende Bedingungen für die Regularität eines Punktes finden wir in [125, 88]. Für die Aufgabe (3.14) ist die Definition 1 äquivalent zu:

$$0 \in \text{int}\{\nabla_z G(z^o)(\mathcal{C} - z^o)\}. \quad (3.16)$$

Die Bedingung (3.16) ist eine relativ starke Voraussetzung, denn in dem Raum, der in Anwendung kommt, besitzt der Kegel \mathcal{K} im allgemeinen ein leeres Inneres (vgl. Theorem 2.1 in [124]).

Nach Lemma 2.3 in [79] ist die Voraussetzung (3.16) zu folgenden geschwächten Voraussetzung äquivalent:

$$\nabla_z G(z^o)(\mathcal{C} - z^o) = \mathcal{Y}. \quad (3.17)$$

Die Gleichung (3.17) ist ein Ausdruck der Surjektivität des Operators $\nabla_z G(z^o)$.

Für die Aufgabe (\mathbf{P}_w) ist jeder zulässige Punkt regulär (vgl. Exemple 9.2 in [47]), und damit ist die Surjektivitätsbedingung (3.17) erfüllt. Zusätzlich folgt aus der Regularität des Punktes z^o die Beschränktheit des entsprechenden Lagrange-Multiplikator (vgl. Theorem 4.1 in [125]).

Die Regularitätsbedingung eines Punktes tritt in der Literatur verschieden auf. Eine allgemeine Formulierung ist zuerst bei *Robinson* [90] gegeben. *Zowe* und *Kurcyusz* [125] geben hinreichende Bedingungen und benutzen dabei die Gleichheit von Mengen. *Penot* [85] und *Tröltzsch* [114] bringen relativ einfach überprüfbare notwendige Bedingungen ein, wenn in den betrachteten Mengen innere Punkte vorliegen. *Maurer* [76] stellt eine Variante für Probleme mit Gleichungs- und Ungleichungsrestriktionen dar.

3.5.3 Stetiges Minimum-Prinzip

In der Aufgabe (3.10) – (3.12) folgt die Abhängigkeit der Funktion f (rechte Seite der Differentialgleichung) vom Parameter w aus der Parametrisierung der Steuerung. Für die Wohldefiniertheit des Problems soll die parametrisierte Steuerung, über die Stetigkeit hinaus, weitergehende Differenzierbarkeiten aufweisen.

A 4 (Wohldefiniertheit des Problems) *Wir nehmen an:*

$$u(t, \cdot) \in C^2(W, \mathbb{R}^r), \quad \text{für jedes } t \text{ in } [0, 1], \quad (3.18)$$

$$\nabla_w u(t, \cdot) \in \mathcal{L}^\infty(W), \quad \text{für jedes } t \text{ in } [0, 1], \quad (3.19)$$

$$u(\cdot, w) \in IP_{\mu, \mathcal{I}_h, \eta}, \quad \mu \geq 0, \quad \text{für jedes } w \text{ in } W. \quad (3.20)$$

Die Voraussetzung (3.18) bedeutet, daß die parametrisierte Steuerungsfunktionen nach dem Parameter w zweimal differenzierbar sei. Dies ermöglicht die Formulierung von Voraussetzungen für die Optimalitätsbedingungen. Für die Fehlerabschätzung der Steuerung im Raum $\mathcal{L}^\infty([0, 1], \mathbb{R}^r)$ findet später ab Seite 50 die Voraussetzung (3.19) Anwendung.

Die Voraussetzung (3.20) spielt eine grundlegende Rolle für die Konvergenz des Integrationsverfahrens. Besonders an den Grenzübergängen auf dem Steuerungsgitter \mathcal{I}_h hängt die Glattheit des Zustandes von dem Verhalten der parametrisierten Steuerung $u(\cdot, w)$ ab.

Die in dieser Arbeit angesetzte Approximation der Steuerung (siehe (3.7)) sowie die anderen vorgeschlagenen Approximationen (siehe (3.8) und (3.9)) erfüllen die Bedingung (3.18). Bei (3.20) soll der Parameter μ in $IP_{\mu, \mathcal{I}_h, \eta}$ entsprechend gewählt werden.

Lagrange-Funktion der Aufgabe (\mathbf{P}_w)

Es sei $\lambda \in \mathcal{L}^\infty([0, 1], \mathbb{R}^n)$. Die Lagrange-Funktion \mathcal{L}^w der Aufgabe (\mathbf{P}_w) wird gebildet durch:

$$\mathcal{L}^w(x, w, v, \lambda) = g_0(x(1), v) - \lambda^\top (\dot{x} - f(x, u(\cdot, w), v)). \quad (3.21)$$

Es seien die Matrizen $\check{\mathbf{a}}, \check{\mathbf{p}}, \check{\mathbf{q}}, \check{\mathbf{w}}, \check{\mathbf{r}}, \check{\mathbf{s}}, \check{\mathbf{v}}, \check{\mathbf{l}}, \check{\mathbf{i}}, \check{\mathbf{j}}, \check{\mathbf{k}}, \check{\mathbf{x}}, \mathbf{H}$ definiert durch:

$$\begin{aligned}
\check{\mathbf{a}}[t] &:= \nabla_x f(x, u(\cdot, w), v)(t)|_{(x^o, w^o, v^o)}, \\
\check{\mathbf{p}}[t] &:= \nabla_u f(x, u(\cdot, w), v) \nabla_w u(\cdot, w)(t)|_{(x^o, w^o, v^o)}, \\
\check{\mathbf{q}}[t] &:= \nabla_v f(x, u(\cdot, w), v)(t)|_{(x^o, w^o, v^o)}, \\
\check{\mathbf{w}}[t] &:= [\langle \nabla_w u(\cdot, w), \lambda^\top \nabla_{uu}^2 f(x, u(\cdot, w), v) \nabla_w u(\cdot, w) \rangle \\
&\quad + \langle \nabla_u f(x, u(\cdot, w), v), \nabla_{w,w}^2 u(\cdot, w) \rangle](t)|_{(x^o, w^o, v^o, \lambda^o)} \\
\check{\mathbf{r}}[t] &:= \lambda^\top \nabla_{u,v}^2 f(x, u(\cdot, w), v) \nabla_w u(\cdot, w)(t)|_{(x^o, w^o, v^o, \lambda^o)}, \\
\check{\mathbf{s}}[t] &:= \lambda^\top \nabla_{v,u}^2 f(x, u(\cdot, w), v) \nabla_w u(\cdot, w)(t)|_{(x^o, w^o, v^o, \lambda^o)}, \\
\check{\mathbf{v}}[t] &:= \nabla_{v,v}^2 g_0(x^o(1), v^o) + \lambda^\top \nabla_{v,v}^2 f(x, u(\cdot, w), v)(t)|_{(x^o, w^o, v^o, \lambda^o)}, \\
\check{\mathbf{x}}[t] &:= \nabla_{x,x}^2 \mathcal{L}^w(x, w, v, \lambda)|_{(x^o, w^o, v^o, \lambda^o)} \\
&= \nabla_{x,x}^2 g_0(x^o(1), v^o) + \langle \lambda, \nabla_{x,x}^2 f(x(t), u(t, w), v) \rangle|_{(x^o, w^o, v^o, \lambda^o)}, \\
\check{\mathbf{j}}[t] &:= \nabla_{x,v}^2 \mathcal{L}^w(x, w, v, \lambda)|_{(x^o, w^o, v^o, \lambda^o)} \\
&= \nabla_{x,v}^2 g_0(x^o(1), v^o) + \langle \lambda, \nabla_{x,v}^2 f(x(t), u(t, w), v) \rangle|_{(x^o, w^o, v^o, \lambda^o)}, \\
\check{\mathbf{l}}[t] &:= \nabla_{v,x}^2 \mathcal{L}^w(x, w, v, \lambda)|_{(x^o, w^o, v^o, \lambda^o)} \\
&= \nabla_{v,x}^2 g_0(x^o(1), v^o) + \langle \lambda, \nabla_{v,x}^2 f(x(t), u(t, w), v) \rangle|_{(x^o, w^o, v^o, \lambda^o)}, \\
\check{\mathbf{k}}[t] &:= \nabla_{w,x}^2 \mathcal{L}^w(x, w, v, \lambda)|_{(x^o, w^o, v^o, \lambda^o)} \\
&= \langle \lambda, \nabla_{u,x}^2 f(x(t), u(t, w), v) \nabla_w u(t, w) \rangle|_{(x^o, w^o, v^o, \lambda^o)}, \\
\check{\mathbf{i}}[t] &:= \nabla_{x,w}^2 \mathcal{L}^w(x, w, v, \lambda)|_{(x^o, w^o, v^o, \lambda^o)} \\
&= \langle \lambda, \nabla_{x,u}^2 f(x(t), u(t, w), v) \nabla_w u(t, w) \rangle|_{(x^o, w^o, v^o, \lambda^o)}, \\
\mathbf{H}[t] &:= \nabla_{(x,w,v),(x,w,v)}^2 \mathcal{L}^w(x, w, v, \lambda)|_{(x^o, w^o, v^o, \lambda^o)},
\end{aligned} \tag{3.22}$$

mit

$$\dot{x}^o(t) = f(x^o(t), u(t, w^o), v^o), \quad x^o(0) = x_0(v^o), \tag{3.23}$$

$$\dot{\lambda}^o(t) = -(\lambda^o(t))^\top \nabla_x f(x^o(t), u(t, w^o), v^o), \quad \lambda^o(1) = [\nabla_x g_0(x^o(1), v^o)]^\top, \tag{3.24}$$

wobei der entsprechende Lagrange-Multiplikator λ^o , von dem es angenommen wird, daß er existiert.

Die Berechnung der Operatoren $\nabla_{(w,v)} \mathcal{L}^w|_{(x^o, w^o, v^o, \lambda^o)}$ und $\mathbf{H}[t]$ ergibt die Gleichungen:

$$\nabla_{(x,w,v)} \mathcal{L}^w|_{(x^o, w^o, v^o, \lambda^o)} = \begin{pmatrix} \nabla_x g_0(x^o(1), v^o) + (\lambda^o)^\top(t) \check{\mathbf{a}}[t] \\ (\lambda^o)^\top(t) \check{\mathbf{p}}[t] \\ \nabla_v g_0(x^o(1), v^o) + (\lambda^o)^\top(t) \check{\mathbf{q}}[t] \end{pmatrix}^\top,$$

$$\mathbf{H}[t] = \begin{pmatrix} \check{\mathbf{x}}[t] & \check{\mathbf{i}}[t] & \check{\mathbf{j}}[t] \\ \check{\mathbf{k}}[t] & \check{\mathbf{w}}[t] & \check{\mathbf{k}}[t] \\ \check{\mathbf{l}}[t] & \check{\mathbf{s}}[t] & \check{\mathbf{v}}[t] \end{pmatrix}.$$

Es seien die Mengen \mathcal{Y} und \mathcal{K} gegeben als im Abschnitt 3.4 auf Seite 17. Die Beschränkungen der Aufgabe (\mathbf{P}_w) beschreiben eine Kegel \mathcal{K} mit $\mathcal{K} = \{0\}$. Es sei \mathcal{K}^+ eine Teilmenge des dualen Raumes \mathcal{Y}^* gegeben durch:

$$\mathcal{K}^+ := \{\lambda \in \mathcal{Y}^* : \langle \lambda, y \rangle \geq 0, \forall y \in \mathcal{K}\} = \mathcal{Y}^*.$$

Theorem 1 (Notwendige Optimalitätsbedingungen erster Ordnung) :

Es sei $z^o = (x^o, w^o, v^o)$ ein regulärer Punkt und es gelten die Voraussetzungen der Annahmen **A3-A4**. Dann existiert ein Element $\lambda^o \in \mathcal{K}^+$, so daß

$$- \begin{pmatrix} \nabla_x g_0(x^o(1), v^o) + \langle \lambda^o(t), \check{\mathbf{a}}[t] \rangle \\ \langle \lambda^o(t), \check{\mathbf{p}}[t] \rangle \\ \nabla_v g_0(x^o(1), v^o) + \langle \lambda^o(t), \check{\mathbf{q}}[t] \rangle \end{pmatrix}^\top \in \mathcal{N}_c[\mathcal{C}](z^o). \quad (3.25)$$

Für den Beweis des Theorems 1 sei auf § 2.4.2 in [62] verwiesen.

Das System gebildet aus (3.23), (3.24) und (3.25) wird in der Steuerungstheorie als kanonisches System bezeichnet. Die Existenz von λ^o wird durch die Regularität von z^o gewährleistet (vgl. [67, 69]). Andere ausführliche Beweise des Minimumprinzips sind u.a. in [47], Seite 83–89 oder [1], § 6.2.2 zu finden. Im Falle der linearen Aufgaben optimaler Steuerung verweisen wir auf [14], § 2.2.2 – § 2.2.5.

Andere Formulierungen und Anwendungen der notwendigen Optimalitätsbedingungen sind beispielsweise in [57, 73, 96, 89, 79] vorhanden.

Koerzitivitätsbedingung

A 5 Es existiere eine positive Konstante α mit der Eigenschaft:

$$\begin{pmatrix} x \\ w \\ v \end{pmatrix}^\top \mathbf{H}[t] \begin{pmatrix} x \\ w \\ v \end{pmatrix} \geq \alpha (\|x\|_{\mathcal{X}}^2 + \|w\|_{\mathbb{R}^L}^2 + \|v\|_{\mathbb{R}^{d_o}}^2), \quad (3.26)$$

für alle $t \in [0, 1]$, $x = x_1 - x_2$, mit $x_1, x_2 \in \mathcal{W}^{1,2}([0, 1], \mathbb{R}^n)$, $w = w_1 - w_2$, mit $w_1, w_2 \in W$, $v = v_1 - v_2$, mit $v_1, v_2 \in V_a$ wobei die entsprechenden parametern w und v und der zugehörige Zustand $x \in \mathcal{W}^{1,2}([0, 1], \mathbb{R}^n)$ folgenden Gleichungen genügen:

$$\dot{x}(t) = \check{\mathbf{a}}[t]x(t) + \check{\mathbf{p}}[t]w + \check{\mathbf{q}}[t]v, \quad x(0) = 0. \quad (3.27)$$

Die Koerzitivitätsbedingung bringt nicht nur die hinreichenden Optimalitätsbedingungen zum Ausdruck, sondern auch Lipschitz-Abhängigkeit der Lösung und der Lagrange-Multiplikatoren in Bezug auf Parameter (siehe [32, 33, 34]).

3.6 Diskretisierung des Problems

3.6.1 Diskrete Aufgabe

Es sei $N \in \mathbb{N} \setminus \{0\}$ gegeben, h definiert durch die Gleichung (3.5), das Steuerungsgitter, beschrieben durch (3.4). Wir betrachten ein festes **Zustandsgitter** \mathcal{G}_h aus dem Intervall $I = [0, 1]$ und die lokalen Schrittweiten h_l gegeben durch:

$$\mathcal{G}_h := \{t_0, t_1, \dots, t_N\}, \quad t_0 := 0 < t_1 < \dots < t_N := 1, \quad (3.28)$$

$$h_l := t_l - t_{l-1}, \quad l = 1, \dots, N. \quad (3.29)$$

Für die Auswahl des Gitters \mathcal{G}_h sind folgende technische Voraussetzungen notwendig:

A 6 : Das Zustandsgitter \mathcal{G}_h sei als eine Verfeinerung des Steuerungsgitters \mathcal{I}_h zu wählen (d.h. $\tau_j \in \{t_0, t_1, \dots, t_N\}$, $j = 1, \dots, M$) und die lokalen Schrittweiten sollen so gewählt werden, daß die folgenden Bedingungen erfüllt werden:

$$h_{\min} := \frac{c_1}{N} \leq h_l \leq \frac{c_2}{N} =: h_{\max}, \quad l = 1, \dots, N, \quad (3.30)$$

wobei c_1 und c_2 geeignete positive Konstanten sind.

Die Bedingung (3.30) dient zur Vermeidung von numerischer Singularität. Das heißt, die lokalen Schrittweiten h_l , $l = 1, \dots, N$ sollen die kleinste beliebig vorgeschriebene Schrittweite h_{\min} nicht unterschreiten bzw. die größte beliebig vorgeschriebene Schrittweite h_{\max} nicht überschreiten.

Zur Formulierung des diskreten Problems soll eine Numerik für gewöhnliche Differentialgleichungen zur Verfügung gestellt werden (vgl. [55, 50]). Dabei wird der Zustand $x \in \mathcal{X}$ nur in den diskreten Zeitpunkten $t_l \in [0, 1]$, $l = 0, 1, \dots, N$, betrachtet. Speziell für die Differentialgleichung (2.2) wird zuerst die Steuerung $u(\cdot)$ durch die parametrisierte Steuerung $u(\cdot, w)$ ersetzt, und anschließend wird ein allgemeines nichtlineares Einschrittverfahren der Form:

$$x_l = x_{l-1} + h_l \varphi_l(x_l, x_{l-1}, w, v), \quad x_0 = x_0(v), \quad l = 1, \dots, N, \quad (3.31)$$

angewendet. Die Funktionen φ_l , $l = 0, 1, \dots, N$ charakterisieren das Integrationsverfahren und werden oft als Verfahrensfunktionen bezeichnet.

Wir formulieren die diskrete Aufgabe (\mathbf{P}_{wd}):

$$\min_{x, w, v} g_0(x_N, v) \quad \text{bzgl.} \quad (3.32)$$

$$x_l = x_{l-1} + h_l \varphi_l(x_l, x_{l-1}, w, v), \quad l = 1, \dots, N, \quad x_0 = x_0(v) \quad (3.33)$$

$$x \in \mathbb{R}^{(N+1)n}, \quad w \in W, \quad v \in V_a. \quad (3.34)$$

Zusätzlich zu den Parametern w und v betrachten wir die maximale Schrittweite auf dem Steuerungsgitter h als Parameter. Der Parameter h ist für die Störung der Aufgabe (\mathbf{P}_{wd}) zuständig.

3.6.2 Minimum-Prinzip der diskreten Aufgabe

Es seien die Gitter \mathcal{I}_h und \mathcal{G}_h gegeben und (x^h, w^h, v^h) mit $x \in \mathbb{R}^{(N+1)n}$, ein zulässiger Punkt der Aufgabe (\mathbf{P}_{wd}). Wir bezeichnen mit \mathcal{C}^h die Menge definiert durch

$$\mathcal{C}^h := \mathbb{R}^{(N+1)n} \times W \times V_a.$$

Regularitätsbedingungen eines Punktes

Es seien $G_{h,l} : \mathbb{R}^{(N+1)n} \times \mathbb{R}^L \times \mathbb{R}^{d_o} \rightarrow \mathbb{R}^n$, $l = 0, \dots, N$ eine Folge von Funktionen definiert durch:

$$\begin{aligned} G_{h,0}(x, w, v) &:= x_0 - x_0(v), \\ G_{h,l}(x, w, v) &:= x_l - x_{l-1} - h_l \varphi_l(x_l, x_{l-1}, w, v), \quad l = 1, \dots, N. \end{aligned}$$

Definition 2 : Der Punkt $z^h = (x^h, w^h, v^h)$ heißt regulär, falls für alle $a = (a_0, \dots, a_N)$ mit $a_l \geq 0$ und $a_l \in \mathbb{R}^{(N+1)n+L+d_o}$, $l = 0, \dots, N$, das einzige Element a , das die Gleichung

$$-\sum_{l=0}^N a_l^\top \nabla_{(x,w,v)} G_{h,l}(z^h) \in \mathcal{N}_c[\mathcal{C}^h](z^h) \quad (3.35)$$

genügt, ist $a = 0$.

Diskrete Lagrange-Funktion

Es sei $\lambda = (\lambda_0, \dots, \lambda_N) \in \mathbb{R}^{(N+1)n}$, $x = (x_0, \dots, x_N) \in \mathbb{R}^{(N+1)n}$. Die diskrete Lagrange-Funktion \mathcal{L}^d für die Aufgabe (\mathbf{P}_{wd}) wird gebildet durch:

$$\mathcal{L}^d(x, w, v, \lambda) = g_0(x_N, v) - \sum_{l=1}^N \lambda_l^\top (x_l - x_{l-1} - h_l \varphi_l(x_l, x_{l-1}, w, v)). \quad (3.36)$$

Die Formulierung von Optimalitätsbedingungen des diskretisierten Problems (\mathbf{P}_{wd}) erfordern zusätzliche Differenzierbarkeitsvoraussetzungen an die Verfahrensfunktion φ_l , die in folgender Annahme formuliert wird:

A 7 : Die Verfahrensfunktionen φ_l , $l = 1, \dots, N$ seien Lipschitz-stetig in allen Argumenten, zweimal stetig differenzierbar in allen Argumenten.

Die Lipschitz-Stetigkeit der Funktionen φ_l , $l = 1, \dots, N$ gewährleistet die Stabilität des Einschrittverfahrens (vgl. [123], 4.2.8 Satz). Es seien $\mathcal{K}_h = \{0\}$, $\mathcal{K}_h^+ = \{\lambda \in \mathbb{R}^{(N+1)n} : \langle \lambda, k \rangle \geq 0, \forall k \in \mathcal{K}_h\} = \mathbb{R}^{(N+1)n}$.

Theorem 2 *Es seien (x^h, w^h, v^h) ein regulärer Punkt und es gelten die Voraussetzungen von Annahmen **A3** – **A7**. Dann existiert einen Vektor $\lambda^h = (\lambda_1^h, \dots, \lambda_N^h) \in \mathcal{K}_h^+$, so daß folgendes System erfüllt ist:*

$$-\nabla_{(w,v)} \left(g_0(x_N, v) + \sum_{l=1}^N \lambda_l^\top (h_l \varphi_l(x_l, x_{l-1}, w, v)) \right) \Big|_{p^h} \in \mathcal{N}_c[W \times V_a](w^h, v^h), \quad (3.37)$$

mit $p^h = (x^h, w^h, v^h, \lambda^h)$ wobei der zugehörige Zustand x^h und der entsprechende Lagrange-Multiplikator λ^h folgenden Gleichungen genügen

$$\left. \begin{aligned} x_l - x_{l-1} &= h_l \varphi_l(x_l, x_{l-1}, w, v), \quad x_0 = x_0(v) \\ l &= 1, \dots, N, \end{aligned} \right\} \quad (3.38)$$

$$\left. \begin{aligned} \lambda_l - \lambda_{l+1} &= h_l [\nabla_{x_l} \varphi_l(x_l, x_{l-1}, w, v)]^\top \lambda_l \\ &\quad + h_{l+1} [\nabla_{x_l} \varphi_{l+1}(x_{l+1}, x_l, w, v)]^\top \lambda_{l+1}, \\ l &= N-1, \dots, 0, \\ \lambda_N &= \nabla_x g_0(x_N, v)^\top + h_N (\nabla_{x_N} \varphi_N(x_N, x_{N-1}, w, v))^\top \lambda_N. \end{aligned} \right\} \quad (3.39)$$

Das Theorem 2 ist ein Analogon von (1.1) in [125] bzw. Theorem 2.1 in [76] in endlich-dimensionalen Räumen, denn die Aufgabe (\mathbf{P}_{wd}) bzw. die zugehörigen Mengen haben eine ähnliche Gestalt wie die Aufgabe (\mathbf{P}) bzw. die zugehörigen Mengen in [125, 76].

Der nächste Abschnitt stellt ein Verfahren für die Berechnung von Gradienten der Zielfunktion diskretisierter optimaler Steuerungsprobleme dar. Dabei spielt das Gleichungssystem (3.39) eine besondere Rolle.

3.7 Gradientenberechnung diskretisierter Funktionen

Die numerische Lösung von nichtlinearen optimalen Steuerungsproblemen durch Standard-Methoden erfordert eine sorgfältige Berechnung der Ableitungen der Zielfunktion und der Restriktionen. Für die Berechnung von Ableitungen existiert eine Vielzahl von Methoden aus der Literatur. Zu nennen sind beispielsweise die numerische Berechnung der Gradienten durch finite Differenzen (vgl. [13, 66]), die numerische Approximation von stetigen Gradienten durch Integration der dualen Gleichung (vgl. [49, 112]) oder die interne numerische Differentiation (vgl. [5, 37, 92]).

Diese Arbeit betrachtet die verallgemeinerten Resultate der internen numerischen Differentiation von [5] für die Berechnung der Gradienten. Der Grund dafür ist die einmalige numerische Integration der Differentialgleichung (3.33), die zu einer drastischen Senkung der Rechenzeit führt (vgl. [11, 12]).

Nun wollen wir das Verfahren der internen numerischen Differentiation für unsere Problem kurz beschreiben und anschließend die Gradientenberechnung beim Runge-Kutta-Verfahren ausführlich vorstellen.

3.7.1 Die Methode der internen numerischen Differentiation

Es seien die Zielfunktion $g_0(x_N, v)$ als eine explizite Funktion des Parameters $w \in \mathbb{R}^L$ (vgl. Definition 3.6 auf Seite 13) und des Parameters $v \in \mathbb{R}^{d_o}$ betrachtet sowie die diskreten Zustandsvariablen x_l , $l = 1, \dots, N$ durch das diskrete dynamische System (3.33) gegeben. Es seien $l(j) \in \{0, \dots, N\}$ mit $t_{l(j)} = \tau_j$, $j = 0, \dots, M$, $\varphi_l = \varphi_l(x_l, x_{l-1}, w, v)$. Wir betrachten eine Funktion F gegeben durch:

$$F(z) = g_0(x_N, v), \quad \text{mit } z = (w, v).$$

Lemma 1 *Es seien g_0 , φ_l , und h_l , betrachtet als Funktionen von z , stetig differenzierbar für alle $l \in \{1, \dots, N\}$. Falls die partiellen Ableitungen $\nabla_{z_k} x_l$, $l = 1, \dots, N$ für ein $k \in \{1, \dots, L + d_o\}$ existieren, dann gilt:*

$$\nabla_{w_j} F = \sum_{l=l(j-1)+1}^{l(j)} [h_l \lambda_l^\top \nabla_{w_j} \varphi_l(x_l, x_{l-1}, w, v)] + \nabla_{w_j} g_0(x_N, v), \quad j = 1, \dots, L \quad (3.40)$$

$$\nabla_{v_j} F = \sum_{l=1}^N [h_l \lambda_l^\top \nabla_{v_j} \varphi_l(x_l, x_{l-1}, w, v)] + \nabla_{v_j} g_0(x_N, v), \quad 1 \leq j \leq d_o, \quad (3.41)$$

mit

$$\lambda_N = \nabla_x g_0(x_N, v)^\top + h_N (\nabla_{x_N} \varphi_N)^\top \lambda_N, \quad (3.42)$$

$$\lambda_l - \lambda_{l+1} = h_l [\nabla_{x_l} \varphi_l]^\top \lambda_l + h_{l+1} [\nabla_{x_l} \varphi_{l+1}]^\top \lambda_{l+1}, \quad (3.43)$$

$$l = N - 1, \dots, 0.$$

Die Behauptungen von Lemma 1 folgen aus Theorem 3.1 in [5], mit den Ansätzen $\alpha_{l0} = 1$, $\alpha_{l1} = -1$, $k_l = 1$ für alle $l = 1, \dots, L + d_o$ und $\alpha_{l,j} = 0$ für alle anderen Koeffizienten bei der Anwendung des Einschrittverfahrens.

Der Parameter v ist ein freier Parameter. Somit können auch die inneren Steuerungsgitterpunkte (**nur für numerische Zwecke**) als Komponenten des Parameters v betrachtet werden.

Für $v_k = \tau_j$, $1 \leq j \leq M - 1$ gilt:

$$\nabla_{\tau_j} F = \sum_{l=l(j-1)+1}^{l(j+1)} \lambda_l^\top h_l \nabla_{\tau_j} \varphi_l + \lambda_{l(j)}^\top \varphi_{l(j)} - \lambda_{l(j)+1}^\top \varphi_{l(j)+1} + \nabla_{\tau_j} g_0(x_N, v) \quad (3.44)$$

In den Teilintervallen $[t_{l(0)}, t_{l(j-1)}]$ und $(t_{l(j+1)}, t_{l(M)}]$ ist die Funktion φ_l von Parametern τ_j , $j = 1, \dots, M - 1$ unabhängig. Damit ist die erste und letzte Summe folgender Gleichung gleich Null:

$$\begin{aligned} \nabla_{\tau_j} F &= \sum_{l=1}^N \lambda_l^\top \left(h_l \nabla_{\tau_j} \varphi_l + \nabla_{\tau_j} h_l \cdot \varphi_l \right) + \nabla_{\tau_j} g_0(x_N, v) \\ &= \sum_{l=1}^{l(j-1)} \lambda_l^\top h_l \nabla_{\tau_j} \varphi_l + \sum_{l=l(j-1)+1}^{l(j)} \lambda_l^\top h_l \nabla_{\tau_j} \varphi_l + \sum_{l=l(j)+1}^{l(j+1)} \lambda_l^\top h_l \nabla_{\tau_j} \varphi_l \\ &\quad + \sum_{l=l(j+1)+1}^N \lambda_l^\top h_l \nabla_{\tau_j} \varphi_l + \lambda_{l(j)}^\top \varphi_{l(j)} - \lambda_{l(j)+1}^\top \varphi_{l(j)+1} + \nabla_{\tau_j} g_0(x_N, v). \end{aligned}$$

Somit folgt die Gleichung (3.44).

Eine tiefgehende Gradientenberechnung im Falle des Runge-Kutta-Verfahrens und des im Abschnitt 3.3.1 angesetzten Parameters wird im folgenden dargestellt.

3.7.2 Gradientenberechnung bei Runge-Kutta-Verfahren

Runge-Kutta-Verfahren sind effektive Verfahren für die numerische Lösung von gewöhnlichen Differentialgleichungen. Es sei $q \in \mathbb{N} \setminus \{0\}$. Im Fall eines impliziten q -stufigen Runge-Kutta-Verfahrens bekommt die Verfahrensfunktion die Gestalt (vgl. [123, 55, 20, 50]):

$$\varphi_l(x_l, x_{l-1}, w, v) := \sum_{i=1}^q b_i K_{li}, \quad (3.45)$$

$$K_{li} = f\left(x_{l-1} + h_l \sum_{j=1}^q \alpha_{ij} K_{lj}, u(t_{l-1} + c_i h_l, w), v\right), \quad i = 1, \dots, q, \quad (3.46)$$

mit $\sum_{i=1}^q b_i = 1$, $\sum_{j=1}^q \alpha_{ij} = c_i$, $i = 1, \dots, q$, wobei $(\alpha_{i,j})$, $b = (b_1, \dots, b_q)$, $c = (c_1, \dots, c_q)$ die Parameter des Runge-Kutta-Schemas sind.

Bemerkung 4 Durch die explizite Beschreibung der Verfahrensfunktion φ_l in den Gleichungen (3.45) – (3.46) ist die Erfüllbarkeit von **A7** ohne zusätzliche Voraussetzung keine Trivialität. Grund dazu ist die Gleichung (3.46), die eine implizite Funktion für die Berechnung der

Komponenten K_{li} , $i = 1, \dots, q$, und dies für jedes $l = 1, \dots, N$ andeutet. Für die Lösung des Systems (3.46) werden der implizite Funktionssatz und die Bedingung h klein benötigt.

Aufgrund der Definition von φ_l in den Gleichungen (3.45)–(3.46) und der Voraussetzungen von Annahmen **A3**–**A4** und für hinreichende kleine Schrittweite h (siehe Bemerkung 4) erfüllt die Verfahrensfunktion bei Runge-Kutta-Verfahren die Voraussetzungen von Annahme **A7**. Die Erfüllbarkeit der Bedingungen von Annahme **A4** folgt ihrerseits von der Definition der Steuerung in (3.7).

Für die Berechnung der Gradienten sind folgende Schritte erforderlich:

Eingangsdaten

$$L = 2Mr, \quad (u_1^+, u_1^-, \dots, u_j^+, u_j^-, \dots, u_M^+, u_M^-, v_1, \dots, v_{d_o})^\top \in \mathbb{R}^{2Mr+d_o}.$$

Erster Schritt

Berechnung der Lösung des diskreten dynamischen Systems (3.33)–(3.34) für $u(t) = u(t, w)$. Das heißt die Berechnung von

$$x_1, x_2, \dots, x_N.$$

Zweiter Schritt

Berechnung der Lösung des diskreten dualen Systems:

$$\lambda_N = [\nabla_x g_0(x_N, v)]^\top + h_N [\nabla_{x_N} \varphi_N]^\top \lambda_N, \quad (3.47)$$

$$\lambda_l = \lambda_{l+1} + \sum_{i=1}^q b_i \left(h_l (Y_{l,i}^a)^\top \lambda_l + h_{l+1} (Y_{l+1,i}^b)^\top \lambda_{l+1} \right), \quad (3.48)$$

$$l = N - 1, \dots, 0,$$

mit

$$Y_{l,i}^a = \nabla_{x_l} K_{l,i} = [\nabla_x f|_{l,i}] \cdot \left[h_l \sum_{j=1}^q \alpha_{ij} Y_{l,j}^a \right], \quad i = 1, \dots, q, \quad (3.49)$$

$$Y_{l+1,i}^b = \nabla_{x_l} K_{l+1,i} = [\nabla_x f|_{l+1,i}] \cdot \left[I + h_{l+1} \sum_{j=1}^q \alpha_{ij} Y_{l+1,j}^b \right], \quad i = 1, \dots, q, \quad (3.50)$$

wobei für $l = 1, \dots, N$; $i = 1, \dots, q$ gilt:

$$[\nabla_x f|_{l,i}] = \nabla_x f(x_{l-1} + h_l \sum_{j=1}^q \alpha_{ij} K_{l,j}, u(t_{l-1} + c_i h_l, w), v),$$

$$[\nabla_u f|_{l,i}] = \nabla_u f(x_{l-1} + h_l \sum_{j=1}^q \alpha_{ij} K_{l,j}, u(t_{l-1} + c_i h_l, w), v).$$

Dritter Schritt

Berechnung der Ableitung von K_{li} , $l = 0, \dots, N-1$, $i = 1, \dots, q$, nach z_k , $k = 1, \dots, 2Mr + d_o$.

Fall $k \leq 2Mr$, $t = t_{l-1} + c_i h_l$:

$$\begin{aligned}\nabla_{z_k} K_{li} &= \nabla_{z_k} f(x_{l-1} + h_l \sum_{j=1}^q \alpha_{ij} K_{lj}, u(t, w), v) \\ &= [\nabla_x f|_{l,i}] \cdot [h_l \sum_{j=1}^q \alpha_{ij} \nabla_{z_k} K_{lj}] + [\nabla_u f|_{l,i}] \nabla_{z_k} u(t, w).\end{aligned}$$

Für $t \in [\tau_{j-1}, \tau_j)$, $z_k \in \{u_j^+, u_j^-\}$, $j = 1, \dots, M-1$, gilt:

$$\begin{aligned}u(t, w) &= \frac{1}{\tau_j - \tau_{j-1}} [u_j^-(t - \tau_{j-1}) + u_j^+(\tau_j - t)], \\ \nabla_{z_k} u(t, w) &= \nabla_t u(t, w) \cdot \nabla_{z_k} t + \nabla_w u(t, w) \cdot \nabla_{z_k} w, \\ \nabla_{z_k} t &= 0, \text{ für } z_k \in \{u_j^+, u_j^-\}, j = 1, \dots, M, \\ \nabla_z u(t, w) &= (0, \dots, 0, \frac{\tau_j - t}{\tau_j - \tau_{j-1}}, \frac{t - \tau_{j-1}}{\tau_j - \tau_{j-1}}, 0, \dots, 0), \\ \nabla_{z_k} w &= 1_k,\end{aligned}$$

wobei 1_k der k -te Einheitsvektor in \mathbb{R}^{2Mr+d_o} ist. Bei $j = M$ betrachten wir das Intervall $[\tau_{M-1}, \tau_M]$.

Fall $k > 2Mr$, $t = t_{l-1} + c_i h_l$:

$$\begin{aligned}\nabla_{v_k} K_{li} &= \nabla_{v_k} f(x_{l-1} + h_l \sum_{j=1}^q \alpha_{ij} K_{lj}, u(t, w), v) \\ &= [\nabla_x f|_{l,i}] [h_l \sum_{j=1}^q \alpha_{ij} \nabla_{v_k} K_{lj}] + [\nabla_v f|_{l,i}] \nabla_{v_k} v.\end{aligned}$$

Im impliziten Fall bilden die Vektoren $\nabla_{z_k} K_{li}$, $i = 1, \dots, q$ ein lineares Gleichungssystem. Beim expliziten Fall werden die Vektoren $\nabla_{z_k} K_{li}$, $i = 1, \dots, q$ rekursiv berechnet.

Vierter Schritt

Berechnung der Gradienten (3.40), (3.41) und (3.44).

Bemerkung 5 Für die Bestimmung der diskreten Vektoren $\lambda_l \in \mathbb{R}^n$ sollen die Matrizen $Y_{l,i}^a$ und $Y_{l+1,i}^b$ für jedes $i = 1, \dots, q$ bestimmt werden. Beim impliziten Runge-Kutta-Verfahren führen diese Berechnungen zu der Lösung eines linearen Gleichungssystems mit $2n^2$ Unbekannten für jedes $i = 1, \dots, q$. Beim expliziten Runge-Kutta-Verfahren verschwinden die Matrizen $Y_{l,i}^a$, $i = 1, \dots, q$. Die Matrizen $Y_{l+1,i}^b$, $i = 1, \dots, q$ werden rekursiv berechnet.

Falls die inneren Steuerungsgitterpunkte als Komponenten des Parameters v betrachtet werden, dann gelten für die Berechnung der Ableitungen K_{li} , $l = 0, \dots, N-1$, $i = 1, \dots, q$, nach z_k , $k \in 2Mr+1, \dots, 2Mr+d_o$ mit $z_k = \tau_j$, $j = 1, \dots, M-1$ folgende Fallunterscheidungen:

$$\begin{aligned} \nabla_{z_k} K_{li} &= \nabla_{z_k} f(x_{l-1} + h_l \sum_{k=1}^q \alpha_{ik} K_{lk}, u(t, w), v) \\ &= [\nabla_x f|_{li}] \left[[\nabla_{z_k} h_l] \sum_{k=1}^q \alpha_{ik} K_{lk} + h_l \sum_{k=1}^q \alpha_{ik} \nabla_{z_k} K_{lk} \right] + [\nabla_u f|_{li}] \nabla_{z_k} u(t, w). \end{aligned}$$

Für $t \in [\tau_{j-1}, \tau_j)$:

$$\begin{aligned} u(t, w) &= \frac{1}{\tau_j - \tau_{j-1}} [u_j^-(t - \tau_{j-1}) + u_j^+(\tau_j - t)], \\ \nabla_{z_k} u(t, w) &= \nabla_t u \cdot \nabla_{z_k} t + \nabla_w u \cdot \nabla_{z_k} w, \\ \nabla_t u &= \frac{u_j^- - u_j^+}{\tau_j - \tau_{j-1}}, \\ \nabla_{z_k} t &= \begin{cases} c_i, t = t_{l(j)-1} + c_i(\tau_j - t_{l(j)-1}) & \text{d.h. } t \in [t_{l(j)-1}, \tau_j) \\ 0, & \text{sonst.} \end{cases}, \\ \nabla_{z_k} u &= (0, \dots, 0, \frac{(t - \tau_{j-1})(u_j^+ - u_j^-)}{(\tau_j - \tau_{j-1})^2}, 0, \dots, 0), \text{ mit } z_k = \tau_j, \\ \nabla_{z_k} w &= 1_k \end{aligned}$$

Bei $t = t_{l(j)-1} + c_i(\tau_j - t_{l(j)-1})$:

$$\nabla_{\tau_j} h_l = \begin{cases} 1, t = t_{l(j)-1} + c_i(\tau_j - t_{l(j)-1}) & \text{d.h. } t \in [t_{l(j)-1}, \tau_j) \\ 0, & \text{sonst.} \end{cases},$$

Für $t \in [\tau_j, \tau_{j+1})$:

$$\begin{aligned} u(t, w) &= \frac{1}{\tau_{j+1} - \tau_j} [u_{j+1}^-(t - \tau_j) + u_{j+1}^+(\tau_{j+1} - t)], \\ \nabla_{z_k} u(t, w) &= \nabla_t u \cdot \nabla_{z_k} t + \nabla_w u \cdot \nabla_{z_k} w, \\ \nabla_t u &= \frac{u_{j+1}^- - u_{j+1}^+}{\tau_{j+1} - \tau_j}, \\ \nabla_{\tau_j} t &= \begin{cases} 1 - c_i, t = \tau_j + c_i(t_{l(j)+1} - \tau_j) & \text{d.h. } t \in [\tau_j, t_{l(j)+1}) \\ 0, & \text{sonst.} \end{cases}, \\ \nabla_{z_k} u &= (0, \dots, 0, \frac{(\tau_{j+1} - t)(u_{j+1}^+ - u_{j+1}^-)}{(\tau_{j+1} - \tau_j)^2}, 0, \dots, 0), \text{ mit } z_k = \tau_j, \\ \nabla_{z_k} w &= 1_k. \end{aligned}$$

Bei $t = \tau_j + c_i(t_{l(j)+1} - \tau_j)$:

$$\nabla_{\tau} h_l = \begin{cases} -1, & t = \tau_j + c_i(t_{l(j)+1} - \tau_j) \quad \text{d.h. } t \in [\tau_j, t_{l(j)+1}) \\ 0, & \text{sonst.} \end{cases}$$

Bei der Diskretisierung mit Kollokationsverfahren werden die Berechnungen der Gradienten ungenau⁶ (vgl. [6]). Sie haben aber den Vorteil, daß diese Berechnungen relativ einfach sind. Zur Verbesserung der Gradientenberechnung werden Optimierungsverfahren angewendet, in dem die diskreten Zustände zur Liste der Optimierungsvariablen gehören.

Die vorliegende Arbeit betrachtet die Steuerung als einzige Optimierungsvariable. Dadurch hat das Optimierungsverfahren keinen direkten Einfluß auf die diskreten Zustände. Somit liegt die genaue Gradientenberechnung nur am angewandten Verfahren der internen numerischen Differentiation an.

⁶ Die Ausdrücke für die Berechnung der Gradienten beinhalten ungenaue diskrete Zustände. Dadurch entsteht die Ungenauigkeit bei der Berechnung von Gradienten.

Kapitel 4

Konvergenzuntersuchung

Nach der Formulierung einer parametrisierten stetigen Aufgabe und dem Vorschlag einer Diskretisierung möchten wir in diesem Kapitel eine allgemeine Approximationstheorie vorstellen und anschließend diese Theorie für die Konvergenzuntersuchung der Lösungen der diskreten Aufgaben gegen eine Lösung der stetigen Aufgabe anwenden.

4.1 Allgemeine Approximationstheorie

In der Literatur existieren verschiedene Ansätze für die Approximationstheorien. Zu nennen seien: Eine Störungstheorie zur Untersuchung von Optimierungsproblemen in Banach-Räumen (vgl. *Alt* [2, 3]) und eine Störungstheorie basierend auf der Untersuchung von verallgemeinerten Gleichungen (Notwendige Optimalitätsbedingungen von Optimierungsproblemen), (vgl. *Dontchev* und *Hager* [33, 32], *Malanowski* [75, 72]). Wir beschreiben an dieser Stelle kurz eine Approximationstheorie und orientieren uns dabei an der Vorgehensweise in [32].

Es seien ein metrischer Raum P (Parameterraum), p ein beliebiges Element aus P . In Verbindung mit p seien ein Banach-Raum \mathcal{Z}_p , ein normierter Raum \mathcal{Y}_p und eine Teilmenge Ω_p von \mathcal{Z}_p gegeben. Es sei \mathcal{T}_p eine Abbildung von Ω_p in \mathcal{Y}_p , \mathcal{F}_p eine Abbildung von Ω_p in der Potenzmenge $2^{\mathcal{Y}_p}$ und $\|\cdot\|_p$ die Norm im entsprechenden parametrisierten Raum.

Wir betrachten ein festes Element p_o aus P . Ausgangspunkt für die Approximationstheorie ist die Lösung folgender verallgemeinerter Gleichung:

$$\text{Finde } z \in \Omega_{p_o} \quad : \quad \mathcal{T}_{p_o}(z) \in \mathcal{F}_{p_o}(z), \quad (4.1)$$

wobei die entsprechenden Mengen \mathcal{Z}_{p_o} , \mathcal{Y}_{p_o} , Ω_{p_o} als ungestörte Mengen zu betrachten sind, und die entsprechenden Operatoren \mathcal{T}_{p_o} , \mathcal{F}_{p_o} für ungestörte Operatoren zu halten sind.

In den meisten Fällen ist die Aufgabe (4.1) nicht einfach zu behandeln, insbesondere dann, wenn die Abbildung \mathcal{T}_{p_0} nicht genügend glatt ist. Für die Ermittlung einer Lösung von (4.1) betrachten wir eine Familie parameterabhängiger Aufgaben, welche durch folgende verallgemeinerte Gleichungen definiert sind:

$$\text{Finde } z \in \Omega_p \quad : \quad \mathcal{T}_p(z) \in \mathcal{F}_p(z). \quad (4.2)$$

Unsere Ziel ist, die Abhängigkeit der optimalen Lösung der Aufgabe (4.2) vom Parameter p zu erforschen. Hauptsächlich geht es dabei um das Verhalten der Lösung von (4.2) für p in einer nahe gelegenen Umgebung von p_0 . Wir nehmen folgendes an:

Es existiere eine optimale Lösung z_{p_0} der Aufgabe (4.1).

Um Informationen zur Lösung von (4.2) für verschiedene Werte des Parameters p zu gewinnen, betrachten wir folgende Familie von Hilfsaufgaben:

$$\text{Finde } z \in \Omega_p \quad : \quad \Upsilon_p(z) + y \in \mathcal{F}_p(z), \quad (4.3)$$

wobei Υ_p den Raum \mathcal{Z}_p im Raum \mathcal{Y}_p abbildet und der Parameter y ein Element von \mathcal{Y}_p ist. Die Mengen \mathcal{Z}_p , \mathcal{Y}_p bzw. Ω_p resultieren aus einer Störung der Mengen \mathcal{Z}_{p_0} , \mathcal{Y}_{p_0} bzw. Ω_{p_0} . Der Operator \mathcal{T}_p bzw. \mathcal{F}_p entsteht aus einer Störung des Operators \mathcal{T}_{p_0} bzw. \mathcal{F}_{p_0} .

Nun interessieren uns zwei Fragen. Erstens ist die Existenz einer Lösung der Aufgabe (4.2) zu überprüfen und zweitens wollen wir einen Weg für die Approximation der Lösung von (4.1) finden. In [32] zeigen die Autoren folgendes: Falls \mathcal{T}_p durch Υ_p „gut“ approximiert wird und die Lösungsmenge von (4.3) bezüglich y gleichmäßig in p Lipschitz-stetig ist, dann besitzt die Lösungsmenge von (4.2) äquivalente Eigenschaften bezüglich p . Das für diese Arbeit hinreichende Resultat wollen wir in folgendem formulieren.

Es seien z_p ein Referenzpunkt und $B_\beta(z)$ eine abgeschlossene Kugel um den Punkt z und mit Radius β . Wir definieren mit Hilfe der Operatoren Υ_p und \mathcal{T}_p die Parameter D_β , Δ_β und δ :

$$\begin{aligned} D_\beta(p) &= \sup_{\substack{y, z \in B_\beta(z_p) \cap \Omega_p \\ y \neq z}} \frac{\|\mathcal{T}_p(z) - \mathcal{T}_p(y) - \Upsilon_p(z) + \Upsilon_p(y)\|_p}{\|z - y\|_p}, \\ \Delta_\beta(p) &= \Delta_\beta(p, z_p) \quad \text{wobei} \quad \Delta_\beta(p, x) = \bigcup_{z \in B_\beta(x) \cap \Omega_p} \{\mathcal{T}_p(z) - \Upsilon_p(z)\}, \\ \delta(p) &= \|\mathcal{T}_p(z) - \Upsilon_p(z) - y_p\|_p. \end{aligned} \quad (4.4)$$

Es seien A und B zwei Teilmengen von \mathcal{Z}_p . Wir bezeichnen mit $\|A - B\|$ die einseitige Definition des Abstands der Menge A zur Menge B definiert durch:

$$\|A - B\| := \sup_{b \in B} \inf_{a \in A} \|a - b\|.$$

Lemma 2 ([32], Corollary 1) *Es sei ψ_p eine Abbildung von einer Umgebung von y_p in die Menge 2^{Ω_p} mit folgenden Eigenschaften :*

- $z_p \in \psi_p(y_p)$, $\psi_p(y)$ ist eine abgeschlossene, nichtleere Menge aus Lösungen von (4.3) für alle $y \in \Delta_\theta(p)$, wobei $\theta > 0$, und
- es existieren positive Konstanten γ und α so, daß gilt :

$$\|\psi_p(y_1) \cap B_\alpha(z_p) - \psi_p(y_2)\| \leq \gamma \|y_1 - y_2\|_p, \quad y_1, y_2 \in \Delta_\theta(p) \cup \{y_p\}. \quad (4.5)$$

Falls $D_\beta(p)$ und $\delta(p)$ gegen Null konvergieren, wenn β gegen Null geht und p gegen p_0 geht, dann besitzt die Inklusion (4.2) für alle $\gamma^+ > \gamma$ und für alle p hinreichend nahe an p_0 eine Lösung z so, daß

$$\|z_p - z\|_p \leq \gamma^+ \|\mathcal{T}_p(z) - \Upsilon_p(z) - y_p\|_p, \quad (4.6)$$

gilt.

Das Lemma 2 vermittelt Bedingungen für die Existenz einer Lösung von (4.2) für verschiedene Werte von p und bietet gleichzeitig eine Möglichkeit für die Approximation einer Lösung von (4.1) an.

Die Aussagen über die Konvergenz von $D_\beta(p)$ und $\delta(p)$ gegen Null, wenn β bzw. p gegen Null bzw. p_0 geht, garantieren die Approximation der Aufgabe (4.2) durch die Hilfsaufgabe (4.3). Dies bedeutet für glatte Aufgaben (d.h. Aufgaben, bei denen die auftretenden Operatoren hinreichend glatt sind), daß die Hilfsaufgabe durch ihre Linearisierung ersetzt werden kann und somit die Konvergenzbedingungen $D_\beta(p) \rightarrow 0$ und $\delta(p) \rightarrow 0$ für $\beta \rightarrow 0$ und $p \rightarrow p_0$ automatisch erfüllt sind. Die Ungleichung (4.5) bedeutet, daß die Abbildung ψ_p „pseudo“-Lipschitz ist. Wegen der einseitigen Definition des Abstandes ist die „pseudo“-Lipschitzstetigkeit schwächer als die übliche Lipschitzeigenschaft.

Bemerkung 6 : Falls $\delta := \delta(p) = 0$ gilt, dann folgt aus der Definition von $\delta(p)$ und der Ungleichung (4.6) in Lemma 2 die Gleichung

$$\mathcal{T}_p(z) = \Upsilon_p(z) - y_p \quad \forall y_p \in \Delta = \Delta_\beta(p).$$

Damit ist z_p eine Lösung von (4.1). Außerdem gilt die Gleichheit in (4.6). Für $\delta > 0$ soll der Parameter p so nah an p_0 gewählt werden, daß die Ungleichungen

$$\sigma \geq \beta, \quad \gamma D_\beta(p) < 1, \quad \alpha \geq \beta \quad \text{und} \quad \gamma^+ > \frac{\gamma}{1 - \gamma D_\beta(p)} \quad (4.7)$$

gelten.

Die Abbildung ψ_p in Lemma 2 soll eine kontrahierende ¹ Abbildung sein. Die Forderungen in (4.7) ermöglichen die Anwendung von Theorem 1 in [32] für den Beweis von Lemma 2, denn in Theorem 1 in [32] werden kontrahierende Abbildungen eingesetzt.

¹ Die Abbildung ψ_p heißt auf einer Umgebung $U(y_h)$ von y_p kontrahierend, wenn es eine Konstante α , ($0 < \alpha < 1$) existiert, so daß

$$\|\psi_p(y_1) - \psi_p(y_2)\| \leq \alpha \|y_1 - y_2\| \text{ für alle } y_1, y_2 \in U(y_p)$$

4.2 Parametrisierte Aufgaben

Hauptanliegen dieses Abschnittes ist die Anwendung der vorgestellten Approximationstheorie.

4.2.1 Parametrisierte verallgemeinerte Gleichung

Es seien der Parameter p vom Abschnitt 4.1 mit der maximalen Schrittweite auf dem Steuerungsgitter h identifiziert und die Räume \mathcal{Z}_h und \mathcal{Y}_h und die Menge $\Omega_h \subset \mathcal{Z}_h$ folgendermassen definiert:

$$\begin{aligned}\mathcal{Z}_h &:= l^\infty(N, \mathbb{R}^n) \times \mathbb{R}^L \times \mathbb{R}^{d_o} \times l^\infty(N, \mathbb{R}^n), \\ \mathcal{Y}_h &:= \mathbb{R}^{L+d_o} \times l^1(N, \mathbb{R}^n) \times l^1(N, \mathbb{R}^n), \\ \Omega_h &:= \{(w, v, x, \lambda) \in \mathcal{Z}_h, w \in W, v \in V_a, x \in l^\infty(N, \mathbb{R}^n), \lambda \in l^\infty(N, \mathbb{R}^n)\}.\end{aligned}$$

Es seien $z = (x, w, v, \lambda) \in \mathcal{Z}_h$ und $y = (y_1, y_2, y_3) \in \mathcal{Y}_h$. Wir betrachten in den diskreten Räumen die Normen:

$$\begin{aligned}\|z\|_{\mathcal{Z}_h} &:= \|x\|_{l^\infty} + \|w\|_{\mathbb{R}^L} + \|v\|_{\mathbb{R}^{d_o}} + \|\lambda\|_{l^\infty}, \\ \|y\|_{\mathcal{Y}_h} &:= \|y_3\|_{l^1} + \|y_1\|_{\mathbb{R}^{L+d_o}} + \|y_2\|_{l^1}.\end{aligned}$$

Es seien:

- $z^o = (x^o, w^o, v^o, \lambda^o)$ der Referenzpunkt, wobei (x^o, w^o, v^o) eine lokale optimale Lösung der diskreten Aufgabe (3.10) - (3.12) ist und λ^o der zugehörige Lagrange-Multiplikator zu (x^o, w^o, v^o) , von dem es angenommen wird, daß er existiert,
- $z^h = (x^h, w^h, v^h, \lambda^h)$ ein Punkt, wobei (x^h, w^h, v^h) eine lokale optimale Lösung der diskreten Aufgabe (3.32) - (3.34) ist und λ^h der zugehörige Lagrange-Multiplikator zu (w^h, v^h) , von dem es angenommen wird, daß er existiert,
- φ_l bzw. φ_l^o die Abkürzung von $\varphi_l(x_l, x_{l-1}, w, v)$ bzw. $\varphi_l(x_l^o, x_{l-1}^o, w^o, v^o)$, mit $x_l^o = x^o(t_l)$, $l = 0, \dots, N$,
- $B_\beta(\bar{z}) := \{z \in \mathcal{Z}_h : \|z - \bar{z}\|_{\mathcal{Z}_h} \leq \beta\}$, eine abgeschlossene Kugel um \bar{z} mit Radius β .

Wir betrachten eine Abbildung \mathcal{T}^h , definiert von Ω_h in \mathcal{Y}_h und gegeben durch:

$$\mathcal{T}^h(z) := \begin{bmatrix} \mathcal{T}_{h,1}(z) \\ \mathcal{T}_{h,2}(z) \\ \mathcal{T}_{h,3}(z) \end{bmatrix},$$

mit

$$\mathcal{T}_{h,1}(z) := \nabla_{(w,v)} \left(g_0(x_N, v) + \sum_{l=1}^N \lambda_l^\top (h_l \varphi_l(x_l, x_{l-1}, w, v)) \right).$$

Der Operator $\mathcal{T}_{h,2}$ umfaßt das diskrete primäre Gleichungssystem. Der Operator $\mathcal{T}_{h,2}$ besitzt $N + 1$ Komponenten jeweils mit der Darstellung:

$$\left. \begin{aligned} \mathcal{T}_{l,2}(z) &:= (\mathcal{T}_{h,2})_l(z) = x_l - x_{l-1} - h_l \varphi_l(x_l, x_{l-1}, w, v), \quad l = 1, \dots, N, \\ \mathcal{T}_{0,2}(z) &:= x_0 - x_0(v). \end{aligned} \right\} \quad (4.8)$$

Der letzte Teiloperator $\mathcal{T}_{h,3}$ stellt die diskreten dualen Komponenten dar. Der Teiloperator $\mathcal{T}_{h,3}$ kann in $N + 1$ Komponenten auseinander genommen werden. Die l -te Komponente $\mathcal{T}_{l,3} = (\mathcal{T}_{h,3})_l$ hat die Darstellung:

$$\left. \begin{aligned} \mathcal{T}_{N,3}(z) &:= \nabla_{x_N} g_0(x_N, v) - \lambda_N + h_N \lambda_N^\top \nabla_{x_N} \varphi_N(x_N, x_{N-1}, w, v) \\ \mathcal{T}_{l,3}(z) &:= \nabla_{x_l} \left(g_0(x_N, v) - \sum_{j=1}^N \lambda_j^\top (x_j - x_{j-1} - h_j \varphi_j(x_j, x_{j-1}, w, v)) \right) \\ &= \lambda_{l+1} - \lambda_l + h_l \lambda_l^\top \nabla_{x_l} \varphi_l + h_{l+1} \lambda_{l+1}^\top \nabla_{x_l} \varphi_{l+1}, \\ & \quad l = N - 1, \dots, 0. \end{aligned} \right\} \quad (4.9)$$

Wir betrachten für \mathcal{F}^h eine Abbildung von Ω_h in $2^{\mathcal{Y}_h}$, definiert durch:

$$\mathcal{F}^h(z) := \left[\mathcal{N}_c[W \times V_a](w^h, v^h), \underbrace{0, \dots, 0}_{(N+1)n}, \underbrace{0, \dots, 0}_{(N+1)n} \right]^\top,$$

Die Voraussetzungen für die Optimalitätsbedingungen erster Ordnung der Aufgabe (3.32)-(3.34) und die zugehörigen primären und dualen Systeme werden im folgenden als **parametrisierte verallgemeinerte Gleichung** formuliert:

$$\text{Finde } z \in \Omega_h \quad : \quad \mathcal{T}^h(z) \in \mathcal{F}^h(z). \quad (4.10)$$

Die Inklusionen (4.2) und (4.10) haben eine ähnliche Struktur.

4.2.2 Approximierende verallgemeinerte Gleichung

Für die Formulierung der approximierenden verallgemeinerten Gleichung (vgl. (4.3)) wird zuerst ein Operator Υ_h derart konstruiert, daß Υ_h eine „sehr gute“ Approximation von \mathcal{T}_h darstellt. Die Grundgedanken für die Konstruktion von Υ_h sind die folgenden. Die Inklusion (4.10) enthält die notwendigen Optimalitätsbedingungen erster Ordnung der diskreten Aufgabe (3.32)-(3.34). Es wird eine Familie von Aufgaben konstruiert, die eine passende

wobei

$$\begin{aligned}
\check{\mathbf{x}}_{0,0} &:= \lambda_1^\top h_1 \nabla_{x_0, x_0}^2 \varphi_1, \\
\check{\mathbf{x}}_{0,1} &:= \lambda_1^\top h_1 \nabla_{x_0, x_1}^2 \varphi_1, \\
\check{\mathbf{x}}_{1,0} &:= \lambda_1^\top h_1 \nabla_{x_1, x_0}^2 \varphi_1, \\
\check{\mathbf{x}}_{1,1} &:= \lambda_1^\top h_1 \nabla_{x_1, x_1}^2 \varphi_1 + \lambda_2^\top h_2 \nabla_{x_1, x_1}^2 \varphi_2, \\
\check{\mathbf{x}}_{1,2} &:= \lambda_2^\top h_2 \nabla_{x_1, x_2}^2 \varphi_2, \\
\check{\mathbf{x}}_{N-2, N-2} &:= \lambda_{N-2}^\top h_{N-2} \nabla_{x_{N-2}, x_{N-3}}^2 \varphi_{N-2}, \\
\check{\mathbf{x}}_{N-2, N-1} &:= \lambda_{N-2}^\top h_{N-2} \nabla_{x_{N-2}, x_{N-2}}^2 \varphi_{N-2} + \lambda_{N-1}^\top h_{N-1} \nabla_{x_{N-2}, x_{N-2}}^2 \varphi_{N-1}, \\
\check{\mathbf{x}}_{N-2, N} &:= \lambda_{N-1}^\top h_{N-1} \nabla_{x_{N-2}, x_{N-1}}^2 \varphi_{N-1}, \\
\check{\mathbf{x}}_{N-1, N-2} &:= \lambda_{N-1}^\top h_{N-1} \nabla_{x_{N-1}, x_{N-2}}^2 \varphi_{N-1}, \\
\check{\mathbf{x}}_{N-1, N-1} &:= \lambda_{N-1}^\top h_{N-1} \nabla_{x_{N-1}, x_{N-1}}^2 \varphi_{N-1} + \lambda_N^\top h_N \nabla_{x_{N-1}, x_{N-1}}^2 \varphi_N, \\
\check{\mathbf{x}}_{N-1, N} &:= \lambda_N^\top h_N \nabla_{x_{N-1}, x_N}^2 \varphi_N, \\
\check{\mathbf{x}}_{N, N-1} &:= \lambda_N^\top h_N \nabla_{x_{N-1}, x_N}^2 \varphi_N, \\
\check{\mathbf{x}}_{N, N} &:= \nabla_{x_N, x_N}^2 g_o(x_N, v) + \lambda_N^\top h_N \nabla_{x_N, x_N}^2 \varphi_N,
\end{aligned}$$

$$\check{\mathbf{i}}^d = \begin{pmatrix} \lambda_1^\top h_1 \nabla_{x_0, w}^2 \varphi_1 \\ \lambda_1^\top h_1 \nabla_{x_1, w}^2 \varphi_1 + \lambda_2^\top h_2 \nabla_{x_1, w}^2 \varphi_2 \\ \lambda_2^\top h_2 \nabla_{x_2, w}^2 \varphi_2 + \lambda_3^\top h_3 \nabla_{x_2, w}^2 \varphi_3 \\ \vdots \\ \lambda_{N-2}^\top h_{N-2} \nabla_{x_{N-2}, w}^2 \varphi_{N-2} + \lambda_{N-1}^\top h_{N-1} \nabla_{x_{N-2}, w}^2 \varphi_{N-1} \\ \lambda_{N-1}^\top h_{N-1} \nabla_{x_{N-1}, w}^2 \varphi_{N-1} + \lambda_N^\top h_N \nabla_{x_{N-1}, w}^2 \varphi_N \\ \lambda_N^\top h_N \nabla_{x_N, w}^2 \varphi_N \end{pmatrix},$$

$$\check{\mathbf{j}}^d = \begin{pmatrix} \lambda_1^\top h_1 \nabla_{x_0, v}^2 \varphi_1 \\ \lambda_1^\top h_1 \nabla_{x_1, v}^2 \varphi_1 + \lambda_2^\top h_2 \nabla_{x_1, v}^2 \varphi_2 \\ \lambda_2^\top h_2 \nabla_{x_2, v}^2 \varphi_2 + \lambda_3^\top h_3 \nabla_{x_2, v}^2 \varphi_3 \\ \vdots \\ \lambda_{N-2}^\top h_{N-2} \nabla_{x_{N-2}, v}^2 \varphi_{N-2} + \lambda_{N-1}^\top h_{N-1} \nabla_{x_{N-2}, v}^2 \varphi_{N-1} \\ \lambda_{N-1}^\top h_{N-1} \nabla_{x_{N-1}, v}^2 \varphi_{N-1} + \lambda_N^\top h_N \nabla_{x_{N-1}, v}^2 \varphi_N \\ \nabla_{x_N, v}^2 g_o(x_N, v) + \lambda_N^\top h_N \nabla_{x_N, v}^2 \varphi_N \end{pmatrix},$$

$$\check{\mathbf{k}}^d = \begin{pmatrix} \lambda_1^\top h_1 \nabla_{w,x_0}^2 \varphi_1 \\ \lambda_1^\top h_1 \nabla_{w,x_1}^2 \varphi_1 + \lambda_2^\top h_2 \nabla_{w,x_1}^2 \varphi_2 \\ \lambda_2^\top h_2 \nabla_{w,x_2}^2 \varphi_2 + \lambda_3^\top h_3 \nabla_{w,x_2}^2 \varphi_3 \\ \vdots \\ \lambda_{N-2}^\top h_{N-2} \nabla_{w,x_{N-2}}^2 \varphi_{N-2} + \lambda_{N-1}^\top h_{N-1} \nabla_{w,x_{N-2}}^2 \varphi_{N-1} \\ \lambda_{N-1}^\top h_{N-1} \nabla_{w,x_{N-1}}^2 \varphi_{N-1} + \lambda_N^\top h_N \nabla_{w,x_{N-1}}^2 \varphi_N \\ \lambda_N^\top h_N \nabla_{w,x_N}^2 \varphi_N \end{pmatrix},$$

$$\check{\mathbf{l}}^d = \begin{pmatrix} \lambda_1^\top h_1 \nabla_{v,x_0}^2 \varphi_1 \\ \lambda_1^\top h_1 \nabla_{v,x_1}^2 \varphi_1 + \lambda_2^\top h_2 \nabla_{v,x_1}^2 \varphi_2 \\ \lambda_2^\top h_2 \nabla_{v,x_2}^2 \varphi_2 + \lambda_3^\top h_3 \nabla_{v,x_2}^2 \varphi_3 \\ \vdots \\ \lambda_{N-2}^\top h_{N-2} \nabla_{v,x_{N-2}}^2 \varphi_{N-2} + \lambda_{N-1}^\top h_{N-1} \nabla_{v,x_{N-2}}^2 \varphi_{N-1} \\ \lambda_{N-1}^\top h_{N-1} \nabla_{v,x_{N-1}}^2 \varphi_{N-1} + \lambda_N^\top h_N \nabla_{v,x_{N-1}}^2 \varphi_N \\ \nabla_{v,x_N}^2 g_0(x_N, v) + \lambda_N^\top h_N \nabla_{v,x_N}^2 \varphi_N \end{pmatrix},$$

$$\check{\mathbf{w}}^d := \sum_{l=1}^N h_l \check{\mathbf{w}}_l, \quad \check{\mathbf{w}}_l := \nabla_{w,w}^2 \lambda_l^\top \varphi_l |_{(w^o, v^o, x_l^o, x_{l-1}^o, \lambda^o)},$$

$$\check{\mathbf{r}}^d := \sum_{l=1}^N h_l \check{\mathbf{r}}_l, \quad \check{\mathbf{r}}_l := \nabla_{w,v}^2 \lambda_l^\top \varphi_l |_{(w^o, v^o, x_l^o, x_{l-1}^o, \lambda^o)},$$

$$\check{\mathbf{s}}^d := \sum_{l=1}^N h_l \check{\mathbf{s}}_l, \quad \check{\mathbf{s}}_l := \nabla_{v,w}^2 \lambda_l^\top \varphi_l |_{(w^o, v^o, x_l^o, x_{l-1}^o, \lambda^o)},$$

$$\check{\mathbf{v}}^d := \nabla_{v,v}^2 g_0(x^o(1), v^o) + \sum_{l=1}^N h_l \check{\mathbf{v}}_l, \quad \check{\mathbf{v}}_l := \nabla_{v,v}^2 \lambda_l^\top \varphi_l |_{(w^o, v^o, x_l^o, x_{l-1}^o, \lambda^o)},$$

$$\mathbf{H}^d = \begin{pmatrix} \check{\mathbf{x}}^d & \check{\mathbf{l}}^d & \check{\mathbf{j}}^d \\ \check{\mathbf{k}}^d & \check{\mathbf{w}}^d & \check{\mathbf{r}}^d \\ \check{\mathbf{l}}^d & \check{\mathbf{s}}^d & \check{\mathbf{v}}^d \end{pmatrix}.$$

Wir betrachten folgende modifizierte Lagrange-Funktion:

$$\mathcal{L}^{dm}(x, w, v, \lambda) := \frac{1}{2} \begin{pmatrix} x \\ w \\ v \end{pmatrix}^\top \mathbf{H}^d \begin{pmatrix} x \\ w \\ v \end{pmatrix} - \sum_{l=1}^N \lambda_l^\top \{x_l - x_{l-1} - h_l [\check{\mathbf{a}}_{l,1} x_l + \check{\mathbf{a}}_{l,2} x_{l-1} + \check{\mathbf{p}}_l w + \check{\mathbf{q}}_l v]\}.$$

Die Ausrechnung der Ableitungen von \mathcal{L}^{dm} ergibt:

$$\mathcal{M}_{h,1}(z) := \nabla_{(w,v)} \mathcal{L}^{dm}(w, v, \lambda) = \left(\nabla_w \mathcal{L}^{dm}(w, v, \lambda), \nabla_v \mathcal{L}^{dm}(w, v, \lambda) \right)$$

mit

$$\begin{aligned} \nabla_w \mathcal{L}^{dm}(w, v, \lambda) &:= \frac{1}{2} [x^\top \check{\mathbf{i}}^d + \check{\mathbf{k}}^d x + \check{\mathbf{r}}^d v + v \check{\mathbf{s}}^d + 2\check{\mathbf{w}}^d w] + \sum_{l=1}^N h_l(\lambda_l)^\top \check{\mathbf{p}}_l \\ \nabla_v \mathcal{L}^{dm}(w, v, \lambda) &:= \frac{1}{2} [x^\top \check{\mathbf{j}}^d + \check{\mathbf{l}}^d x + w \check{\mathbf{r}}^d + \check{\mathbf{s}}^d w + 2\check{\mathbf{v}}^d v] + \sum_{l=1}^N h_l(\lambda_l)^\top \check{\mathbf{q}}_l \end{aligned}$$

Wir definieren die l -te Komponente von $\mathcal{M}_{h,2}$ bzw. $\mathcal{M}_{h,3}$ durch:

$$\left. \begin{aligned} (\mathcal{M}_{h,2})_l(z) &:= x_l - x_{l-1} - h_l [\check{\mathbf{a}}_{l,1} x_l + \check{\mathbf{a}}_{l,2} x_{l-1} + \check{\mathbf{p}}_l w + \check{\mathbf{q}}_l v], \\ l &= 1, \dots, N, \\ \mathcal{M}_{0,2}(z) &:= x_0 - x_0(v). \end{aligned} \right\} \quad (4.12)$$

bzw.

$$\left. \begin{aligned} \mathcal{M}_{N,3}(z) &:= \nabla_{x_N} g_0(x_N, v) - \lambda_N + h_N \lambda_N^\top \nabla_{x_N} \varphi_N(x_N, x_{N-1}, w, v) \\ (\mathcal{M}_{h,3})_l(z) &:= \lambda_{l+1} - \lambda_l + h_l(\lambda_l)^\top \check{\mathbf{a}}_{l,1} + h_{l+1}(\lambda_{l+1})^\top \check{\mathbf{a}}_{l+1,2}, \\ l &= N-2, \dots, 0. \end{aligned} \right\} \quad (4.13)$$

Die Operatoren $\mathcal{M}_{h,1}$, $\mathcal{M}_{h,2}$, $\mathcal{M}_{h,3}$ fassen wir in den Operator \mathcal{M} zusammen:

$$\mathcal{M}(z) := \begin{bmatrix} \mathcal{M}_{h,1}(z) \\ \mathcal{M}_{h,2}(z) \\ \mathcal{M}_{h,3}(z) \end{bmatrix}.$$

Mit Hilfe von \mathcal{M} wird der Hilfsoperator Υ^h wie folgt definiert:

$$\Upsilon^h(z) = \mathcal{M}(z) - \mathcal{M}(z^o).$$

Für y_h betrachten wir ein Element aus \mathcal{Y}_h definiert durch:

$$y_h := \left(y_1^h, \dots, y_{L+d_0}^h, \underbrace{0, \dots, 0}_{(N+1)n}, \underbrace{0, \dots, 0}_{(N+1)n} \right)^\top,$$

wobei die einzelnen Komponenten von y_h wie folgt beschrieben sind:

$$\begin{aligned} y_j^h &:= \nabla_w \left(\sum_{l=1}^N \bar{\lambda}_l^\top (h_l \varphi_l^o) \right) 1_j^L, \quad j = 1, \dots, L, \\ y_{L+j}^h &:= \nabla_v \left(g_0(x^o(1), v^o) + \sum_{l=1}^N \lambda_{l-1}^\top (h_l \varphi_l^o) \right) 1_j^{d_0}, \quad j = 1, \dots, d_0, \quad \text{mit,} \end{aligned}$$

wobei 1_j^L bzw. $1_j^{d_0}$ der j -te Einheitsvektor in \mathbb{R}^L bzw. \mathbb{R}^{d_0} ist.

Schließlich formulieren wir folgende **approximierende verallgemeinerte Gleichung**:

$$\text{Finde } z \in \Omega_h \quad : \quad \Upsilon^h(z) + y_h \in \mathcal{F}^h(z). \quad (4.14)$$

4.3 Konvergenzsatz

Eines der Ziele dieser Arbeit ist die Verbesserung der Konvergenz durch den Einsatz von Integrationsverfahren höherer Ordnung. Die Anwendung solcher Verfahren erfordert zusätzliche Bedingungen, welche im folgenden zu erklären sind. Während die erste Annahme Voraussetzungen für die Optimalitätsbedingungen zweiter Ordnung bereitstellt, setzt die zweite Annahme gewisse Abschätzungen der Ableitungen der Verfahrensfunktion voraus. Beide Annahmen werden bei der Abschätzung des Diskretisierungsfehlers benötigt.

Vor der Erzeugung des Steuerungsgitters \mathcal{I}_h sind weder die Anzahl der möglichen auftretenden Sprungstellen der optimalen Steuerung u^* noch ihre genaue Lage bekannt. Für die gesuchte Funktion $u(\cdot, w^o)$, die eine gute Approximation von u^* auf dem Raum \mathcal{U}_w darstellt (siehe Ungleichung (3.13)), sind die Teilintervalle, auf denen die Funktion $u(\cdot, w^o)$ stückweise stetig ist, durch die Vorgabe des Gitters \mathcal{I}_h bekannt. Damit das Einschrittverfahren eine Konsistenzordnung von mindestens Eins besitzen kann, ist neben der geeigneten Bestimmung von φ_l im allgemeinen erforderlich, daß die Funktion f aus (3.11) über die Stetigkeit hinaus gewisse Differenzierbarkeitseigenschaften auf dem ganzen Segment $[0, 1]$ besitzt. Dies hat zur Folge, daß auch die Lösung $u(\cdot, w^o)$ weitergehende Differenzierbarkeiten als die Stetigkeit aufweist. Die Definitionen (3.7), (3.8) und (3.9) bieten je nach der gewünschten Glattheitsvoraussetzungen und den dazugehörigen Restriktionen drei Ansatzmöglichkeiten für die Definition der Funktion $u(\cdot, w^o)$ mit Hilfe von Funktionen aus dem Raum $\mathcal{P}_{\mu, \mathcal{I}_h, \eta}$. Somit ist die parametrisierte Funktion $u(\cdot, w)$ eine Zusammensetzung von N Polynomen vom Grad μ , die $(\mu - 1)$ -mal Fréchet-differenzierbar sind. Die Funktion $u(\cdot, w^o)$ selbst und ihre $(\mu - 1)$ -ten Ableitungen besitzen an den Knoten τ_k , $k = 1, \dots, M - 1$ keine Sprungstellen.

A 8 (Voraussetzung für diskrete Optimalitätsbedingungen zweiter Ordnung) : *Es existiere ein Skalar $\alpha > 0$, unabhängig von N mit der Eigenschaft:*

$$\begin{pmatrix} x \\ w \\ v \end{pmatrix}^\top \mathbf{H}^d \begin{pmatrix} x \\ w \\ v \end{pmatrix} \geq \alpha (\|x\|^2 + \|w\|^2 + \|v\|^2), \quad (4.15)$$

$x = x_1 - x_2$, mit $x_1, x_2 \in l^{1,2}(N, \mathbb{R}^n)$, $w = w_1 - w_2$, mit $w_1, w_2 \in W$, $v = v_1 - v_2$, mit $v_1, v_2 \in V_a$ und wobei die Folge der zugehörigen diskreten Zustände $\{x_l\}_{l=0}^N$, die folgenden Gleichungen genügen:

$$x_l = x_{l-1} + h_l(\check{\mathbf{a}}_{l,1}x_l + \check{\mathbf{a}}_{l,2}x_{l-1} + \check{\mathbf{p}}_l w + \check{\mathbf{q}}_l v), \quad x_0 = 0, \quad l = 1, \dots, N. \quad (4.16)$$

Wegen der Lipschitz-Stetigkeit der Verfahrensfunktionen φ_l , $l = 1, \dots, N$ (siehe **A7**) ist das Einschrittverfahren (3.31) stabil. Damit die Konvergenz des Approximationsverfahrens gesichert ist, sollen auch Konsistenzbedingungen vorhanden sein.

Definition 3 : Das Integrationsverfahren (3.31) heißt gleichmäßig konsistent („uniformly consistent“, vgl. [25]) von der Ordnung $q \in \mathbb{N} \setminus \{0\}$, falls es eine Konstante ρ existiert, die vom Gitter \mathcal{I}_h unabhängig ist, so daß

$$\|\gamma_l^x(x_p, w, v)\| \leq \rho \cdot h^q, \quad \forall (w, v) \in W \times V_a, \quad l = 1, \dots, N, \quad (4.17)$$

$$\gamma_l^x(x_p, w, v) := \frac{1}{h_l}(x_p(t_l) - x_p(t_{l-1})) - \varphi_l(x_p(t_l), x_p(t_{l-1}), w, v), \quad (4.18)$$

wobei x_p die Lösung des Anfangswertproblems (3.11) für das Parameterpaar (w, v) bezeichnet.

Wir nehmen folgendes an:

A9 : Es existiere zwei positive Konstanten c und κ mit $0 \leq \kappa \leq 1$, so daß für die Lösung (x^o, w^o, v^o) der Aufgabe (\mathbf{P}_w) und die Lösung λ^o von (3.24) sowie λ^h Lösung des diskreten dualen Systems (3.39), folgende Voraussetzungen erfüllen:

(a): das Integrationsverfahren (3.31) sei gleichmäßig konsistent von Ordnung q mit $q \geq 1$ bezüglich des Gitters \mathcal{G}_h ,

$$\begin{aligned} \text{(b): } & \left\| \int_{t_{l-1}}^{t_l} (\nabla_{x_l} \varphi_l(x_l, x_{l-1}, w, v) - \kappa \nabla_x f(x(t), u(t, w), v)) \Big|_{(x^o, w^o, v^o)} dt \right\| \leq c h_l, \\ & \left\| \int_{t_{l-1}}^{t_l} (\nabla_{x_{l-1}} \varphi_l(x_l, x_{l-1}, w, v) - (1 - \kappa) \nabla_x f(x(t), u(t, w), v)) \Big|_{(x^o, w^o, v^o)} dt \right\| \leq c h_l, \end{aligned}$$

$$\text{(c): } \left\| \int_{t_{l-1}}^{t_l} (\nabla_w \varphi_l(x_l, x_{l-1}, w, v) - \nabla_u f(x(t), u(t, w), v) \nabla_w u(t, w)) \Big|_{(x^o, w^o, v^o)} dt \right\| \leq c h_l,$$

$$\text{(d): } \left\| \int_{t_{l-1}}^{t_l} (\nabla_v \varphi_l(x_l, x_{l-1}, w, v) - \nabla_v f(x(t), u(t, w), v)) \Big|_{(x^o, w^o, v^o)} dt \right\| \leq c h_l,$$

$$\text{(e): } \max_{l \in \{1, \dots, N\}} \|\lambda^o(t_l) - \lambda_l^h\| = O(h).$$

Die Bedingungen **A9(b) – (d)** beziehen sich auf die Teilintervalle $[\tau_{k-1}, \tau_k)$, $k = 1, \dots, M$. Die gleichmäßige konsistente Eigenschaft **A9(a)** ist auf dem Segment $[0, 1]$ definiert.

Das in dieser Arbeit eingesetzte diskrete duale Gleichungssystem (3.39) resultiert aus dem Differenzieren der diskreten Lagrange-Funktion (3.36). Folgt man den Beweisweg für die Bestimmung der Ordnung beim Integrationsverfahren, so kann man folgendes beobachten.

Die Ordnung der Lösung des Integrationsverfahrens (3.31) ist nicht immer gleich der Ordnung einer Näherungslösung des diskreten dualen Gleichungssystems (3.39). Dies begründet die Voraussetzung **A9 (e)**. Für ein allgemeines explizites Runge-Kutta-Verfahren wollen wir im folgenden Lemma die stets Erfüllbarkeit der Voraussetzung **A9(e)** nachweisen.

Lemma 3 *Es seien (x^o, w^o, v^o) bzw. (x^h, w^h, v^h) ein regulärer Punkt. Es gelten die Voraussetzungen von Annahmen **A3, A4, A6, A7** und die Beschränktheit von $[\nabla_{x_l} \varphi_{l+1}]$. Es seien λ^o Lösung des dualen Systems (3.23) mit $\lambda^o \in \mathcal{W}^{1,\infty}([0, 1], \mathbb{R}^n)$, (x^h, w^h, v^h) eine Lösung von (\mathbf{P}_{wd}) , λ^h Lösung des diskreten dualen Systems*

$$\lambda_N = \nabla_x g_0(x_N, v)^\top, \quad \lambda_l - \lambda_{l+1} = h_{l+1} [\nabla_{x_l} \varphi_{l+1}]^\top \lambda_{l+1}, \quad l = N-1, \dots, 0, \quad (4.19)$$

mit

$$\varphi_l(x_{l-1}, w, v) := \sum_{i=1}^q b_i K_{li}, \quad K_{li} = f(x_{l-1} + h_l \sum_{j=1}^{i-1} \alpha_{ij} K_{lj}, u(t_{l-1} + c_i h_l, w), v).$$

Dann gilt die Konvergenzaussage:

$$\max_{1 \leq l \leq N} \|\lambda^o(t_l) - \lambda_l^h\| = O(h). \quad (4.20)$$

Beweis:

Das System (4.19) ist ein lineares Gleichungssystem bezüglich λ . Damit ist die Verfahrensfunktion des Einschrittverfahrens (4.19) Lipschitz-stetig. Somit folgt aus 4.2.8 Satz in [123] die Stabilität des Einschrittverfahrens (4.19).

Zum Beweis der Konsistenz betrachten wir den folgenden lokalen Verfahrensfehler:

$$r(t_{l+1}, h_{l+1}, \lambda^o) := \frac{1}{h_{l+1}} (\lambda^o(t_l) - \lambda^o(t_{l+1})) - [\nabla_{x_l} \varphi_{l+1}(x^o(t_{l+1}), x^o(t_l), w^o, v^o)]^\top \lambda^o(t_{l+1})$$

Wegen der Lipschitz-Stetigkeitsvoraussetzungen in **A3(b)**, die Zugehörigkeit von $u(\cdot, w^o)$ zu $\mathcal{P}_{\mu, \mathcal{I}_h, \eta}$ (siehe (3.20)) und $\lambda^o \in \mathcal{W}^{1,\infty}([0, 1], \mathbb{R}^n)$ existiere eine allgemeine positive Konstante c , so daß folgende Abschätzungskette gilt:

$$\begin{aligned} & \|r(t_{l+1}, h_{l+1}, \lambda^o)\| \\ &= \left\| \frac{1}{h_{l+1}} \int_{t_l}^{t_{l+1}} \{\nabla_x f(x^o(t), u(t, w^o), v^o)^\top \lambda^o(t) \right. \\ & \quad \left. - [\nabla_{x_l} \sum_{i=1}^q b_i f(x^o(t_l) + h_{l+1} \sum_{j=1}^{i-1} \alpha_{ij} K_{l+1,j}, u(t_l + c_i h_{l+1}, w^o), v^o)]^\top \lambda^o(t_{l+1})\} dt \right\| \\ &\leq \frac{c}{h_{l+1}} \int_{t_l}^{t_{l+1}} \sum_{i=1}^q b_i \left\{ \|x^o(t) - x^o(t_l) + h_{l+1} \sum_{j=1}^{i-1} \alpha_{ij} K_{l+1,j}\| \right. \end{aligned}$$

$$\begin{aligned}
& + \|u(t, w^o) - u(t_l + c_i h_{l+1}, w^o)\| + \|\lambda^o(t) - \lambda^o(t_{l+1})\| \} dt \\
\leq & c \left\{ \max_{l \in \{1, \dots, N\}} \|x^o(t) - x^o(t_l)\| + h_{l+1} \max_{i \in \{1, \dots, q\}} \sum_{j=1}^{i-1} \alpha_{ij} \|K_{l+1, j}\| \right. \\
& \left. + \max_{\substack{i \in \{1, \dots, q\} \\ l \in \{1, \dots, N\}}} \|u(t, w^o) - u(t_l + c_i h_{l+1}, w^o)\| + \max_{l \in \{1, \dots, N\}} \|\lambda^o(t) - \lambda^o(t_{l+1})\| \right\} \\
\leq & c \cdot h.
\end{aligned}$$

Damit ist die Voraussetzung **A9(e)** für ein allgemeines explizites Runge-Kutta-Verfahren ständig erfüllt \triangleleft

Wir wollen zwei Ungleichungen (4.25) und (4.26) beweisen. Diese Ungleichungen sind für die Untersuchungen in dieser Arbeit von erheblicher Bedeutung. Sie werden später im Beweis des Konvergenzsatzes für den Übergang von der schwachen Norm $\|\cdot\|_{l^2}$ zur starken Norm $\|\cdot\|_{l^\infty}$ ausgenutzt.

4.3.1 Normenvergleich

Dieser Abschnitt dient dem Vergleich von zwei Normen (siehe Definitionen (4.21), (4.22)). Es seien $N, M \in \mathbb{N} \setminus \{0\}$, $[\tau_{l(k)}, \tau_{l(k+1)})$, $k = 1, \dots, M$ Teilintervalle auf $[0, 1]$, $l^e(N, \mathbb{R}^n)$, $1 \leq e \leq \infty$ ein Raum aus stückweise stetigen Funktionen auf dem Gitter \mathcal{G}_h . Der Raum $l^e(N, \mathbb{R}^n)$ wird mit folgenden Normen versehen:

$$(\|x\|_{l^e})^e := \sum_{l=1}^N h_l \|x_l\|^e, \quad 1 \leq e < \infty, \quad (4.21)$$

$$\|x\|_{l^\infty} := \max_{0 \leq l \leq N} \{\|x_l\|\}, \quad (4.22)$$

wobei $x = \{x_l\}_{l=0}^N$. Die Steuerungsgitterpunkte τ_k , $k = 0, 1, \dots, M$ sind im Abschnitt 3.3.1 eingeführt, die Indizes $l(j)$, $j = 0, 1, \dots, M + 1$ im Abschnitt 3.7 definiert. Einfachheitshalber betrachten wir ein äquidistantes Zustandsgitter (d.h. $h_l = h$, $l = 1, \dots, N$).

Lemma 4 *Es sei $\{x_l\}_{l=0}^N$ eine Lösung des diskreten Gleichungssystems:*

$$x_l = x_{l-1} + h_l (\check{\mathbf{a}}_{l,1} x_l + \check{\mathbf{a}}_{l,2} x_{l-1} + \check{\mathbf{p}}_l w + \check{\mathbf{q}}_l v), \quad l = 1, \dots, N, \quad x_0 = 0, \quad (4.23)$$

(vgl. Gleichungssystem (4.16) in der Annahme **A8**) für $(w, v) = (\bar{w}, \bar{v})$ und $\{\lambda_l\}_{l=0}^N$ eine Lösungsfolge der diskreten dualen Gleichungen:

$$\begin{aligned}
\lambda_N &= \nabla_{x_N} g_0(x_N, v) + h_N [\nabla_{x_N} \varphi_N(x_N, x_{N-1}, w, v)]^\top \lambda_N \\
\lambda_l &= \lambda_{l+1} + h_{l+1} (\lambda_l^\top \check{\mathbf{a}}_{l,1} + \lambda_{l+1}^\top \check{\mathbf{a}}_{l+1,2}), \quad l = N-1, \dots, 1,
\end{aligned} \quad (4.24)$$

für $(w, v) = (\bar{w}, \bar{v})$ und $x = \{x_l\}_{l=0}^N$ Lösung von (4.23). Dann existieren zwei positive Konstanten c_3 und c_4 , welche den Ungleichungen:

$$\|x_l\| \leq c_3(\|\bar{w}\| + \|\bar{v}\|) \quad (4.25)$$

$$\|\lambda_l\| \leq c_4(\|\bar{w}\| + \|\bar{v}\|), \quad (4.26)$$

genügen.

Beweis:

Es seien $\{x_l\}_{l=0}^N$ ein Lösung von (4.23) für $w = \bar{w}$ und $v = \bar{v}$, und \bar{x} die Lösung von

$$\dot{\bar{x}}(t) = \check{\mathbf{a}}[t]\bar{x}(t) + \check{\mathbf{p}}[t]\bar{w} + \check{\mathbf{q}}[t]\bar{v}, \quad \bar{x}(0) = 0, \quad (4.27)$$

wobei die Matrizen $\check{\mathbf{a}}$, $\check{\mathbf{p}}$, und $\check{\mathbf{q}}$ im Abschnitt 3.3 definiert sind. Es sei y^h eine stückweise konstante Erweiterung der Folge $\{y_l\}$ auf $[t_{l-1}, t_l]$, $l = 1, \dots, N$ mit $y^h(t) = y_{l-1} := \bar{x}(t_{l-1})$ für $t \in [t_{l-1}, t_l]$, $l = 1, \dots, N-1$ und $y^h(t) = y_{N-1} := \bar{x}(t_{N-1})$ für $t \in [t_{N-1}, t_N]$. Wir betrachten eine Konstante κ als gegeben in der Annahme **A9**. Die Integration von (4.27) auf dem Intervall $[t_{l-1}, t_l]$ ergibt:

$$y_l = y_{l-1} + h_l(\check{\mathbf{a}}_{l,1}y_l + \check{\mathbf{a}}_{l,2}y_{l-1} + \check{\mathbf{p}}_l\bar{w} + \check{\mathbf{q}}_l\bar{v}) + \delta_{x,l} \quad (4.28)$$

mit $\delta_{x,l}$ gegeben durch:

$$\begin{aligned} \delta_{x,l} &:= -h_l\delta\check{\mathbf{a}}_{l,1}y_l - h_l\delta\check{\mathbf{a}}_{l,2}y_{l-1} - \delta y_l - h_l\delta\check{\mathbf{p}}_l\bar{w} - h_l\delta\check{\mathbf{q}}_l\bar{v}, \\ \delta\check{\mathbf{a}}_{l,1} &:= \check{\mathbf{a}}_{l,1} - \frac{\kappa}{h_l} \int_{t_{l-1}}^{t_l} \check{\mathbf{a}}[t]dt, \\ \delta\check{\mathbf{a}}_{l,2} &:= \check{\mathbf{a}}_{l,2} - \frac{(1-\kappa)}{h_l} \int_{t_{l-1}}^{t_l} \check{\mathbf{a}}[t]dt, \\ \delta y_l &:= \int_{t_{l-1}}^{t_l} \check{\mathbf{a}}[t](\kappa y_l + (1-\kappa)y_{l-1} - \bar{x}(t))dt, \\ \delta\check{\mathbf{p}}_l &:= \check{\mathbf{p}}_l - \frac{1}{h_l} \int_{t_{l-1}}^{t_l} \check{\mathbf{p}}[t]dt, \\ \delta\check{\mathbf{q}}_l &:= \check{\mathbf{q}}_l - \frac{1}{h_l} \int_{t_{l-1}}^{t_l} \check{\mathbf{q}}[t]dt. \end{aligned}$$

Aus den Annahmen **A7** und **A9(b)** existiert eine positive Konstante c_0 , so das folgende Abschätzung von $\|\delta\check{\mathbf{a}}_{l,1}\|$ und $\|\delta\check{\mathbf{a}}_{l,2}\|$ genügt:

$$\begin{aligned} \|\delta\check{\mathbf{a}}_{l,1}\| &\leq \|\check{\mathbf{a}}_{l,1} - \frac{\kappa}{h_l} \int_{t_{l-1}}^{t_l} \check{\mathbf{a}}[t]dt\| \leq c_0 h_l, \quad l = 1, \dots, N' \\ \|\delta\check{\mathbf{a}}_{l,2}\| &\leq \|\check{\mathbf{a}}_{l,2} - \frac{(1-\kappa)}{h_l} \int_{t_{l-1}}^{t_l} \check{\mathbf{a}}[t]dt\| \leq c_0 h_l, \quad l = 1, \dots, N. \end{aligned}$$

Analog werden die Terme $\delta\check{\mathbf{p}}_l, \delta\check{\mathbf{q}}_l, l = 1, \dots, N$ mit Hilfe der Annahmen **A7** und **A9(c)-(d)** von oben abgeschätzt.

Bei immer klein werdenden Schrittweiten $h_l, l = 1, \dots, N$ ist die Norm

$\|\bar{x}(t) - (\kappa y_l + (1 - \kappa)y_{l-1})\|$ mit Hilfe einer geeigneten Konstante c von oben durch $c h_l$ abschätzbar. Somit existiert eine positive Konstante c_1 , so daß

$$\left\| \int_{t_{l-1}}^{t_l} \check{\mathbf{a}}[t](\bar{x}(t) - (\kappa y_l + (1 - \kappa)y_{l-1}))dt \right\| \leq c_1 h_l^2, \quad l = 1, \dots, N. \quad (4.29)$$

Die Konstante c_1 folgt aus der Stetigkeit von $\check{\mathbf{a}}[t], \bar{x}$ dem Teilintervall $[t_{l-1}, t_l]$.

Das Subtrahieren der diskreten Differenzgleichung (4.23) von (4.28) ergibt die Abschätzung:

$$\|x_l - y_l\| \leq c_5 \sum_{l=1}^N \|\delta_{x,l}\|, \quad (4.30)$$

mit einer passenden Konstante c_5 . Nach der Definition von $\delta_{x,l}$ existiert eine Konstante c_6 , so daß

$$\|\delta_{x,l}\| \leq c_6 h_l (h_l \|y_l\| + h_l \|\bar{w}\| + h_l \|\bar{v}\|), \quad l = 1, \dots, N. \quad (4.31)$$

Aus der Differentialgleichung (4.27) folgt die Ungleichungskette:

$$\|y^h\|_{l^2} \leq \|y^h\|_{l^\infty} \leq \|\bar{x}\|_{L^\infty} \leq c_7 (\|\bar{w}\| + \|\bar{v}\|), \quad (4.32)$$

wobei c_7 eine von h unabhängige und geeignete Konstante für alle h genügend klein ist. Die Ungleichung (4.32) folgt aus folgender Überlegung:

Die Matrix-Funktionen $\check{\mathbf{a}}[t], \check{\mathbf{p}}[t]$ und $\check{\mathbf{q}}[t]$ sind auf den Teilintervallen $[\tau_{k-1}, \tau_k] \subset [0, 1], k = 1, \dots, M$ stetig. Für jede Lösung \bar{x}, \bar{w} und \bar{v} der Differentialgleichung (4.27) folgt aus § I, Theorem 11.1 in [55] die Abschätzung:

$$\begin{aligned} \|\bar{x}(t)\| &\leq \exp \left[\left(\int_{\tau_{k-1}}^{\tau_k} \|\check{\mathbf{a}}[\tau]\| d\tau \right) \int_{\tau_{k-1}}^t \left(\exp \left[- \int_{\tau_{k-1}}^{\tau} \|\check{\mathbf{a}}[s]\| ds \right) \|\check{\mathbf{p}}[\tau]\bar{w} + \check{\mathbf{q}}[\tau]\bar{v}\| \right) d\tau \\ &\leq c_8 \left(\int_{\tau_{k-1}}^t (\|\bar{w}\| + \|\bar{v}\|) ds \right), \quad t \in [\tau_{k-1}, \tau_k], \end{aligned} \quad (4.33)$$

wobei \exp die Exponentialfunktion bezeichnet und c_8 eine passende positive Konstante ist. Es folgt somit folgende Abschätzung auf dem ganzen Intervall $[0, 1]$

$$\|\bar{x}\|_{L^\infty} \leq \max_{\substack{t \in [\tau_{k-1}, \tau_k] \\ k \in \{1, \dots, M\}}} \|\bar{x}(t)\| \leq c_9 (\|\bar{w}\| + \|\bar{v}\|).$$

Die Abschätzung $\|y^h\|_{l^2} \leq \|y^h\|_{l^\infty}$ folgt aus der Äquivalenz von Normen in endlich-dimensionalen Räumen (vgl. [122], § 10, III. Hilfssatz).

Es folgt aus den Ungleichungen (4.30) – (4.32) und der Ungleichungskette (3.30) die Abschätzung:

$$\begin{aligned} \|x_l - y_l\| &\leq 2c_{10} h \sum_{l=1}^N h_l (\|\bar{w}\| + \|y_l\| + \|\bar{v}\|) = 2c_{10} h (\|y\|_{l^1} + \|\bar{w}\| + \|\bar{v}\|) \\ &\leq c_{11} h (\|\bar{w}\| + \|\bar{v}\|). \end{aligned} \quad (4.34)$$

Die Gleichung (4.34) verdeutlicht eine „gute“ Approximation von \bar{x} durch die Folge $\{x_l\}_{l=0}^N$. Wegen der Ungleichung (4.34) genügt auch die Lösungsfolge von (4.23) für hinreichend große N einer ähnlichen Ungleichung. Das heißt, es gilt die Ungleichung:

$$\|x_l\| \leq c_3 (\|\bar{w}\| + \|\bar{v}\|). \quad (4.35)$$

Es seien $\{\lambda_l\}_{l=N}^1$ ein Lösung von (4.24), $\bar{\lambda}$ eine Lösung von

$$\dot{\bar{\lambda}}(t) = -(\lambda(t))^\top \check{\mathbf{a}}[t], \quad \bar{\lambda}(1) = \lambda_N \quad (4.36)$$

Wir betrachten y^h diesmal als eine stückweise konstante Erweiterung der Folge $\{y_l\}$ auf $[t_{l-1}, t_l]$, $l = 1, \dots, N$ mit $y^h(t_l) = y_l := \bar{\lambda}(t_l)$. Wir betrachten

$$y_l = y_{l+1} + h_{l+1} (\lambda_l^\top \check{\mathbf{a}}_{l,1} + \lambda_{l+1}^\top \check{\mathbf{a}}_{l+1,2}), \quad l = N-1, \dots, 1,$$

und definieren die Terme (analog wie in der Gleichung (4.28)) $\delta_{\lambda,l}$, $\delta \check{\mathbf{a}}_{l,1}$, $\delta \check{\mathbf{a}}_{l,2}$, δy_l , $l = 1, \dots, N$.

Schließlich verfolgen wir die Beweisschritte wie bei dem Beweis von (4.25). Es folgt die Abschätzung (4.26) \Rightarrow

4.3.2 Abschätzung des Diskretisierungsfehlers

Lemma 5 : *Es seien (x^o, w^o, v^o) eine reguläre Lösung der Aufgabe (\mathbf{P}_w) , λ^o der zugehörige Lagrange-Multiplikator mit $\lambda^o \in \mathcal{W}^{1,\infty}([0, 1], \mathbb{R}^n)$, $l(k) \in \{0, \dots, N\}$ der Index von Zustandsgitterpunkten mit der Eigenschaft $t_{l(k)} = \tau_k$, $k = 0, \dots, M$ sowie M_k , $k = 1, \dots, M$ die Anzahl der Unterteilung des Teilintervalls $[\tau_{k-1}, \tau_k]$. Es gelten die Voraussetzungen der Annahmen **A3** - **A9**. Dann existieren für jedes genügend große N ein lokales Minimum (x^h, w^h, v^h) der Aufgabe (3.32)-(3.34) sowie λ^h der diskrete zugehörige Lagrange-Multiplikator und es gilt folgende Fehlerabschätzung zur Lösung der parametrisierten Aufgabe (3.10)-(3.12):*

$$\max_{l(k-1) \leq l \leq l(k)} |u(t_l, w^o) - u_l^h| = O\left(\frac{h}{M_k}\right), \quad (4.37)$$

$$\max_{l(k-1) \leq l \leq l(k)} |x^o(t_l) - x_l^h| = O\left(\frac{h}{M_k}\right), \quad (4.38)$$

$$\max_{l(k-1) \leq l \leq l(k)} |\lambda^o(t_l) - \lambda_l^h| = O\left(\frac{h}{M_k}\right), \quad (4.39)$$

für $k = 0, \dots, M$.

Beweis:

Für den Beweis wenden wir das Lemma 2 an. Es seien $[\tau_{k-1}, \tau_k]$, $k = 1, \dots, M$ Teilintervalle von $[0, 1]$ mit $\tau_{k-1}, \tau_k \in \mathcal{I}_h$ und M_k die Anzahl der Unterteilungen von $[\tau_{k-1}, \tau_k]$. Es seien $z^a = (x^a, w^a, v^a, \lambda^a)$ und $z^b = (x^b, w^b, v^b, \lambda^b)$ zwei Elemente aus $B_\beta(z^o)$, wobei (w^a, v^a) durch die Optimierung bestimmt werden, x^a bzw. λ^a der zugehörige Zustand bzw. Lagrange-Multiplikator ist. Dabei wird die Existenz von λ^a angenommen. Die Komponenten w^b, v^b, x^b, λ^b werden analog definiert. Wir definieren entsprechend in Lemma 2 die Parameter D_β , Δ_β und δ durch:

$$D_\beta(h) := \sup_{\substack{z^a, z^b \in B_\beta(z^o) \cap \Omega_h \\ z^a \neq z^b}} \frac{\|\mathcal{T}^h(z^b) - \mathcal{T}^h(z^a) - \Upsilon^h(z^b) + \Upsilon^h(z^a)\|_{\mathcal{Y}_h}}{\|z^b - z^a\|_{\mathcal{Z}_h}} \quad (4.40)$$

$$\Delta_\beta(h) := \Delta_\beta(h, z^o) \quad \text{mit} \quad \Delta_\beta(h, x) = \bigcup_{z \in B_\beta(x) \cap \Omega_h} \{\mathcal{T}^h(z) - \Upsilon^h(z)\}, \quad (4.41)$$

$$\delta(h) := \|\mathcal{T}^h(z) - \Upsilon^h(z) - y_h\|_{\mathcal{Y}_h}. \quad (4.42)$$

Konvergenz von $D_\beta(h)$

Der Zähler auf der rechten Seite von (4.40) kann wie folgt umgeschrieben werden

$$\|E^h(z^b) - E^h(z^a)\|_{\mathcal{Y}_h}$$

wobei die Funktion $E^h(z^b)$ den Fehler bei der Linearisierung von $\mathcal{T}^h(z^b)$ im Punkt z^o ist, d.h.:

$$E^h(z^b) = \mathcal{T}^h(z^b) - \mathcal{M}(z^b) + \mathcal{M}(z^o).$$

Für hinreichend kleine β existiere eine Konstante c , die der Abschätzung:

$$D_\beta(h) \leq c \beta \frac{O(z^b - z^a)}{\|z^b - z^a\|_{\mathcal{Z}_h}}$$

genügt. Damit konvergiert $D_\beta(h)$ gegen Null, falls h und β gegen Null konvergieren.

Konvergenz von $\delta(h)$ im Punkt z^o

Zu zeigen ist die Konvergenz von $\delta(h)$ gegen Null, falls β und h gegen Null konvergieren. Außerdem interessieren wir uns für die Konvergenzgeschwindigkeit von $\delta(h)$.

$$\|\mathcal{T}^h(z^o) - y_h\|_{\mathcal{Y}_h} = \|\mathcal{T}_{h,1}(z^o) - y_{h,1}\|_{\mathbb{R}^{L+d_o}} + \|\mathcal{T}_{h,2}(z^o) - y_{h,2}\|_{l^1} + \|\mathcal{T}_{h,3}(z^o) - y_{h,3}\|_{l^1}$$

mit $y_{h,1} \in \mathbb{R}^{L+d_o}$, $y_{h,2} \in l^1(N, \mathbb{R}^n)$, $y_{h,3} \in l^1(N, \mathbb{R}^n)$.

Für den ersten Teiloperator $\mathcal{T}_{h,1}(z)$ gilt:

$$\|\mathcal{T}_{h,1}(z^o) - y_{h,1}\|_{\mathbb{R}^{L+d_o}} = \|\nabla_{w,v} \left(\sum_{l=1}^N h_l ((\lambda^o)_l^\top - (\lambda^o)_{l-1}^\top) (\varphi_l(x_l^o, x_{l-1}^o, w^o, v^o)) \right)\|.$$

Aus $\lambda^o \in \mathcal{W}^{1,\infty}([0, 1], \mathbb{R}^n)$ gilt die Abschätzung:

$$\|(\lambda^o)_l^\top - (\lambda^o)_{l-1}^\top\| \leq c_\lambda h_l,$$

mit c_λ eine passende Konstante. Mit Hilfe der Ungleichung (3.30) von Annahme **A6** ergibt sich die Abschätzung:

$$\|\mathcal{T}_{h,1}(z^o) - y_{h,1}\|_{\mathbb{R}^{L+d_o}} \leq c \left(\frac{h}{M_k} \right).$$

Der Term $\|\mathcal{T}_{h,2}(z^o) - y_{h,2}\|_{l^1}$ wird mit Hilfe der Annahmen **A9 (a)** und **A6** folgendermaßen abgeschätzt:

$$\begin{aligned} \|\mathcal{T}_{h,2}(z^o) - y_{h,2}\|_{l^1} &= \sum_{l=1}^N h_l \|x^o(t_l) - x^o(t_{l-1}) - h_l \varphi_l(x^o(t_l), x^o(t_{l-1}), w^o, v^o)\| \\ &\leq \sum_{l=1}^N \frac{c_2}{N} h_l \|\gamma_l^x(x^o, w^o, v^o)\| \\ &\leq c_2 \rho \cdot h^{q+1} \end{aligned}$$

mit einer passenden positiven Konstante c_2 . Aus den Annahmen **A9(e)** und **A6** folgen die Abschätzungen:

$$\begin{aligned} \|\mathcal{T}_{h,3}(z^o) - y_{h,3}\|_{l^1} &= \sum_{l=1}^N h_l \|h_l \lambda_l^{o,\top} \nabla_{x_l} \varphi_l^o + h_{l+1} \lambda_{l+1}^{o,\top} \nabla_{x_l} \varphi_{l+1}^o + (\lambda_l^o - \lambda_{l+1}^o)\| \\ &\leq \sum_{l=1}^N h \frac{c_2}{N} \\ &\leq c_2 \cdot h \end{aligned}$$

mit einer passenden positiven Konstante c_2 .

Wegen $q \geq 1$ konvergiert $\delta(h)$ gegen Null, falls β und h gegen Null konvergieren und es gilt die Abschätzung:

$$\delta(h) = O\left(\left(\frac{h}{M_k}\right)^{\min\{q,1\}}\right) = O\left(\frac{h}{M_k}\right). \quad (4.43)$$

Lipschitzeigenschaft der approximierend verallgemeinerten Aufgabe

Die vollständige Anwendung von Lemma 2 erfordert die Lipschitzeigenschaft für die Aufgabe (4.14) bezüglich $y \in \Delta_\beta(h) = \Delta_\beta(h, z^o)$. Das heißt, es soll eine Konstante γ existieren, die der Ungleichung (4.5) genügt.

Es sei der Parameter y beschrieben durch $y := ((y_w, y_v), y_x, y_\lambda) \in \mathcal{Y}_h$ mit $y_x := (y_{1,0}, \dots, y_{1,N})$ und $y_\lambda := (y_{3,0}, \dots, y_{3,N})$. Für den Beweis der Existenz dieser Konstante betrachten wir folgende Aufgabe:

$$\left\{ \begin{array}{l} \langle \mathcal{M}_{h,1}(z) + \begin{pmatrix} y_w \\ y_v \end{pmatrix}, \begin{pmatrix} w - w^h \\ v - v^h \end{pmatrix} \rangle \geq 0, \quad \forall (w, v) \in W \times V_a \\ \mathcal{M}_{h,2}(z) + y_x = 0, \\ \mathcal{M}_{h,3}(z) + y_\lambda = 0 \end{array} \right\}, \quad (4.44)$$

Nun untersuchen wir, wie die Lösung des Systems (4.44) vom Parameter $y := ((y_w, y_v), y_x, y_\lambda)$ gestört werden könnte. Das System (4.44) umfaßt die notwendigen Optimalitätsbedingungen des folgenden nichtlinearen Optimierungsproblems:

$$\begin{aligned} \min_{x,w,v} \frac{1}{2} \begin{pmatrix} x \\ w \\ v \end{pmatrix}^\top \mathbf{H}^d \begin{pmatrix} x \\ w \\ v \end{pmatrix} + \langle (y_x, y_w, y_v), \begin{pmatrix} x \\ w \\ v \end{pmatrix} \rangle \quad \text{bzgl.} \\ x_l - x_{l-1} = h_l(\check{\mathbf{a}}_{l,1}x_l + \check{\mathbf{a}}_{l,2}x_{l-1} + \check{\mathbf{p}}_l w + \check{\mathbf{q}}_l v) - y_{x,l}, \quad x_0 = x_0(v), \\ l = 1, \dots, N, \\ w \in W, \quad v \in V_a. \end{aligned} \quad (4.45)$$

Es seien $(x^k, w^k, v^k, \lambda^k)$, $k = 1, 2$, die Lösungen von (4.44) bezüglich der Parameter $((y_w^k, y_v^k), y_1^k, y_3^k)$, $k = 1, 2$. Nach Lemma 4 in [32] haben die Aufgaben (4.44) und (4.45) die gleiche Lösung genau dann, wenn die Koerzitivitätsbedingung (4.15) aus der Annahme **A8** gilt. Außerdem, es existiere eine positive Konstante c , so daß

$$\|w^1 - w^2\|_{\mathbb{R}^L} + \|v^1 - v^2\|_{\mathbb{R}^{d_0}} \leq c [\|(y_w, y_v)^1 - (y_w, y_v)^2\|_{\mathbb{R}^{L+d_0}} + \|y_x^1 - y_x^2\|_{l^2} + \|y_\lambda^1 - y_\lambda^2\|_{l^2}]. \quad (4.46)$$

Abschätzung der Lösung in \mathcal{L}^∞ -Norm

Die Ungleichung (4.46) ist für optimale Steuerungsprobleme nicht stark genug, da die Funktionen, die die Aufgabe beschreiben, im Raum \mathcal{L}^∞ definiert sind. Interessant wäre die Herleitung einer ähnlichen Ungleichung in die \mathcal{L}^∞ -Norm.

Die Funktion $u(t, \cdot)$ ist stetig Fréchet-differenzierbar nach dem Parameter $w \in W$ (siehe (3.18) in Annahme **A4**) mit beschränkter erster Ableitung (siehe (3.19) in Annahme **A4**). Für jede $w^1, w^2 \in W$ existiert ein w_t , daß zwischen w^1 und w^2 liegt mit der Eigenschaft:

$$\|u(t, w^1) - u(t, w^2)\|_{\mathcal{L}^\infty} \leq \sup_{t \in [0,1]} |\nabla_w u(t, w_t)| \|w^1 - w^2\|_{\mathbb{R}^L}, \quad t \in [0, 1].$$

Wegen (3.19) in Annahme A4 existiere eine positive Konstante c_1 mit der Eigenschaft

$$\|u(\cdot, w^1) - u(\cdot, w^2)\|_{\mathcal{L}^\infty} \leq c_1 \|w^1 - w^2\|_{\mathbb{R}^L} \quad (4.47)$$

Eine Kombination von (4.46) und (4.47) ergibt folgende Abschätzung in \mathcal{L}^∞ -Norm:

$$\|u(\cdot, w^1) - u(\cdot, w^2)\|_{\mathcal{L}^\infty} + \|v^1 - v^2\| \leq c_1 c [\|y_2^1 - y_2^2\|_{\mathbb{R}^{L+d_o}} + \|y_1^1 - y_1^2\|_{l^2} + \|y_3^1 - y_3^2\|_{l^2}]. \quad (4.48)$$

Anschließend folgen aus (4.25) und (4.26) die strengen Abschätzungen

$$\|x^1 - x^2\|_{\mathcal{L}^\infty} = \max_{1 \leq l < N} \{|x_l^1 - x_l^2|\} \leq c_5 (\|w^1 - w^2\|_{\mathbb{R}^L} + \|v^1 - v^2\|_{\mathbb{R}^{d_o}}), \quad (4.49)$$

$$\|\lambda^1 - \lambda^2\|_{\mathcal{L}^\infty} = \max_{1 \leq l < N} \{|\lambda_l^1 - \lambda_l^2|\} \leq c_6 (\|w^1 - w^2\|_{\mathbb{R}^L} + \|v^1 - v^2\|_{\mathbb{R}^{d_o}}), \quad (4.50)$$

mit passenden Konstanten c_5 und c_6 . Es folgt aus (4.46) und (4.48) die strenge Abschätzung:

$$\begin{aligned} & \|x^1 - x^2\|_{\mathcal{L}^\infty} + \|u(\cdot, w^1) - u(\cdot, w^2)\|_{\mathcal{L}^\infty} + \|\lambda^1 - \lambda^2\|_{\mathcal{L}^\infty} + \|v^1 - v^2\|_{\mathbb{R}^{d_o}} \\ & \leq \gamma [\|y_1^1 - y_1^2\|_{\mathbb{R}^{L+d_o}} + \|y_2^1 - y_2^2\|_{l^2} + \|y_3^1 - y_3^2\|_{l^2}]. \end{aligned} \quad (4.51)$$

Damit existiert eine Konstante γ , die der Ungleichung (4.5) genügt.

Lokale Optimalität von (x^h, w^h, v^h)

Wir betrachten folgende diskrete Aufgabe:

$$\min F(x, w, v), \quad (x, w, v) \in \mathcal{C} \quad \text{bzgl.} \quad C(x, w, v) = 0, \quad (4.52)$$

wobei die Funktionen F , C und die Menge \mathcal{C} wie folgt gegeben sind:

$$\begin{aligned} \mathcal{C} &= \mathbb{R}^{(N+1)n} \times W \times V_a, \\ F(x, w, v) &= \frac{1}{2} (x, w, v) \mathbf{H}^d (x, w, v)^\top \\ C(x, w, v) &= (C_0(x, w, v), \dots, C_N(x, w, v))^\top \text{ mit} \\ C_0(x, w, v) &= x_0 - x_0(v), \\ C_l(x, w, v) &= x_l - x_{l-1} - h_l (\check{\mathbf{a}}_{l,1} x_l + \check{\mathbf{a}}_{l,2} x_{l-1} + \check{\mathbf{p}}_l w + \check{\mathbf{q}}_l v), \quad l = 1, \dots, N. \end{aligned}$$

Die Funktionen F und C sind nach ihrer Herleitung zweimal stetig differenzierbare Funktionen in \mathcal{C} .

Es seien (x, w, v) ein zulässiger Punkt der Aufgabe (4.52), $\lambda = (\lambda_0, \dots, \lambda_N)$ ein Element aus $\mathbb{R}^{(N+1)n}$, \bar{L} die Lagrange-Funktion zur Aufgabe (4.52) gegeben durch:

$$\bar{L}(x, w, v, \lambda) = F(x, w, v) - \lambda^\top C(x, w, v),$$

und U_c eine abgeschlossene Kugel um den Punkt (x, w, v) mit dem Radius β und eine Menge U_c gegeben durch:

$$U_c := \mathcal{C} \cap B_\beta(x, w, v).$$

Die Menge U_c ist konvex, abgeschlossen und beschränkt in $\mathbb{R}^{(N+1)n} \times \mathbb{R}^L \times \mathbb{R}^{d_o}$ und somit kompakt. Damit existiert eine lokale Lösung $(\bar{x}, \bar{w}, \bar{v})$ für die Aufgabe (4.52) und es gilt die Gleichung:

$$\nabla_{(x,w,v)} \bar{L}(\bar{x}, \bar{w}, \bar{v}, \bar{\lambda}) \in \mathcal{N}_c[U_c](\bar{x}, \bar{w}, \bar{v}). \quad (4.53)$$

mit $\bar{\lambda}$ Lösung von:

$$\begin{aligned} \lambda_l &= \lambda_{l+1} + h_l \lambda_l^\top \check{\mathbf{a}}_{l,1} + h_{l+1} \lambda_{l+1}^\top \check{\mathbf{a}}_{l+1,2}, \quad l = N-1, \dots, 1, \\ \lambda_N &= \nabla_{x_N} g_0(x_N, \bar{v}) + h_N [\nabla_{x_N} \varphi_N(x_N, x_{N-1}, \bar{w}, \bar{v})]^\top \lambda_N. \end{aligned}$$

Die Gleichung (4.53) ist eine Linearisierung von (3.37)-(3.37) im Punkt (x^o, w^o, v^o) aus der Umgebung von U_c . Folglich, es existiert einen Punkt (x^h, w^h, v^h) , der die Voraussetzungen für die Optimalitätsbedingungen erster Ordnung (3.37) erfüllt. Die Voraussetzungen für die Optimalitätsbedingungen zweiter Ordnung in **A8** sind eine Folgerung von Annahme **A5** und der Stabilität der Bedingung zweiter Ordnung für kleine Störung an den Matrizen $\check{\mathbf{a}}_{l,1}$, $\check{\mathbf{a}}_{l,2}$, $\check{\mathbf{p}}_l$, $\check{\mathbf{q}}_l$, \mathbf{H}^d für $l = 1, \dots, N$ (siehe Gleichung (4.11)).

Im Abschnitt 4.4.2 werden Voraussetzungen zur Gewährleistung der Stabilität der Matrizen in (4.11) für das Runge-Kutta-Verfahren formuliert und bewiesen. Damit ist die Existenz und die lokale Optimalität von (w^h, v^h) bewiesen und folglich das Lemma 5 vollständig bewiesen. \clubsuit

4.3.3 Lipschitz-Stetigkeit der Lösung gewöhnlicher Differentialgleichungen bzgl. Parametern

Es sei f auf $\mathbb{R}^n \times \mathbb{R}^r \times \mathbb{R}^{d_o}$ stetig und existiere eine reelle positive Zahl L_f mit der Eigenschaft:

$$\begin{aligned} \|f(x_1, u_1, v_1) - f(x_2, u_2, v_2)\| &\leq L_f (\|x_1 - x_2\| + \|u_1 - u_2\| + \|v_1 - v_2\|), \\ \forall (x_i, u_i, v_i) &\in \mathbb{R}^n \times \mathbb{R}^r \times \mathbb{R}^{d_o}, \quad i = 1, 2. \end{aligned}$$

Die Aufgabe

$$\dot{x}(t) = f(x(t), u(t), v), \quad x(0) = x_0(v), \quad t \in [0, 1].$$

besitze eine eindeutige Lösung x^* für $(u, v) = (u^*, v^*)$ mit $x^* \in \mathcal{W}^{1,\infty}$. In Verbindung mit Annahme **A2** sei mit $u(\cdot, w^o)$ eine minimale Störung von u^* gegeben. Wir führen folgende gewöhnliche Differentialgleichung ein

$$\dot{x}(t) = f(x(t), u(t, w^o), v), \quad x(0) = x_0(v), \quad t \in [0, 1]. \quad (4.54)$$

Es sei λ^* eine Lösung der linearen Anfangswertaufgabe:

$$\dot{\lambda}(t) = -\lambda^\top(t) \nabla_x f(x^*(t), u^*(t), v^*), \quad \lambda(1) = [\nabla_x g_0(x^*(1), v^*)]^\top,$$

mit $\lambda^* \in \mathcal{W}^{1,\infty}$. λ^* ist der assoziierte Lagrange-Multiplikator zu (x^*, u^*, v^*) . Es sei x^o eine Lösung von (4.54). Wir betrachten folgende gestörte Aufgabe:

$$\dot{\lambda}(t) = -\lambda^\top(t) \nabla_x f(x^o(t), u(t, w^o), v^*), \quad \lambda(1) = [\nabla_x g_0(x^o(1), v^*)]^\top. \quad (4.55)$$

Lemma 6 : *Es sei die Funktion f in allen Argumenten Lipschitz-stetig. Es gilt die Annahmen A1 – A3. Dann besitze die Differentialgleichung (4.54) bzw. (4.55) eine eindeutige Lösung x^o bzw. λ^o und es gilt die Abschätzungen:*

$$\|x^* - x^o\|_{\mathcal{L}^\infty} \leq c_o \|u^* - u(\cdot, w^o)\|_{\mathcal{L}^1} \quad (4.56)$$

$$\|\lambda^* - \lambda^o\|_{\mathcal{L}^\infty} \leq c_o \|u^* - u(\cdot, w^o)\|_{\mathcal{L}^1}, \quad (4.57)$$

mit einer geeigneten positiven Konstante c_o .

Beweis

Für den Beweis wird Lemma 3.11 in [24] angewendet. Es sei f Lipschitz-stetig mit der Konstante L_f . Die gestörte Gleichung (4.54) können wir folgendermaßen umformulieren:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u^*(t), v^*) + g(t) \quad \text{mit} \\ g(t) &= f(x(t), u(t, w^o), v^*) - f(x(t), u^*(t), v^*). \end{aligned}$$

Die Ungleichung (3.11) in [24] liefert für $\varepsilon_0 = 0$, $\varepsilon_1 = 1$ die Abschätzung:

$$\begin{aligned} \|x^*(t) - x^o(t)\| &\leq \int_0^t \exp[(s-0)L_f] \cdot \|f(x^*(s), u(s, w^o), v^*) - f(x^*(s), u^*(s), v^*)\| ds \\ &\leq \max_{0 \leq t \leq 1} \exp[(t-0)L_f] \int_0^t \|u(s, w^o) - u^*(s)\| ds \\ &\leq \exp[L_f] \int_0^1 \|u(s, w^o) - u^*(s)\| ds. \end{aligned}$$

Somit existiert eine Konstante c_o , die der Ungleichung (4.56) genügt. Für den Beweis der Eindeutigkeit von x^o , nehmen wir an, die Gleichung (4.54) besitze zwei unterschiedliche Lösungen x_1^o und x_2^o . Mit Hilfe der Dreieckungleichung gilt die Ungleichungskette

$$\begin{aligned} \|x_1^o - x_2^o\| &\leq \|x_1^o - x^*\| + \|x^* - x_2^o\| \\ &\leq 2c_o \|u^* - u(\cdot, w^o)\|_{\mathcal{L}^1} \\ &\leq 2c_o \epsilon(h) \end{aligned}$$

mit $\lim_{h \rightarrow 0} \epsilon(h) = 0$ (siehe Annahme **A2**). Dies widerspricht die Unterschiedlichkeit beider Lösungen.

Für den Beweis von (4.57) betrachten wir die Gleichung

$$\begin{aligned}\dot{\lambda}(t) &= -(\lambda(t))^\top \nabla_x f(x^*(t), u^*(t), v^*) + g(t) \quad \text{mit} \\ g(t) &= -(\lambda(t))^\top \nabla_x f(x^o(t), u(t, w^o), v^*) + (\lambda(t))^\top \nabla_x f(x^*(t), u^*(t), v^*).\end{aligned}$$

Die Funktion $\nabla_x f$ ist nach **A3(b)** Lipschitz-stetig. Es sei L_{∇_x} die Lipschitz-Konstante von $\nabla_x f$. Mit Hilfe von (4.56) und [24], (3.11) gilt folgende Ungleichungskette:

$$\begin{aligned}\|\lambda^*(t) - \lambda^o(t)\| &\leq \int_0^t \exp[(s-0)L_{\nabla_x}] \cdot \|(\lambda(s))^\top \nabla_x f(x^o(s), u(s, w^o), v^*) \\ &\quad - (\lambda(s))^\top \nabla_x f(x^*(s), u^*(s), v^*)\| ds \\ &\leq \max_{0 \leq t \leq 1} \exp[(t-0)L_{\nabla_x}] \int_0^t \|x^o(s) - x^*(s)\| + \|u(s, w^o) - u^*(s)\| ds \\ &\leq c \exp[L_{\nabla_x}] \int_0^1 \|u(s, w^o) - u^*(s)\| ds,\end{aligned}$$

mit einer geeigneten Konstante c . Es folgt die Ungleichung (4.57). Die Eindeutigkeit von λ^o ist offensichtlich \Rightarrow

Formulierung des Hauptsatzes

Theorem 3 : Es seien f auf $\mathbb{R}^n \times \mathbb{R}^r \times \mathbb{R}^{d_o}$ Lipschitz-stetig, (x^*, u^*, v^*) eine optimale Lösung von (\mathbf{P}_*) . Es existiere $\lambda^* \in \mathcal{W}^{1,\infty}$ der zugehörige Lagrange-Multiplikator zu (x^*, u^*, v^*) . Es gelten die Voraussetzungen von Annahmen **A1** - **A9** und die Regularitätsbedingung (3.35). Dann existieren für jedes genügend große N ein lokales Minimum (x^h, w^h, v^h) der Aufgabe (\mathbf{P}_{wd}) sowie λ^h der diskrete zugehörige Lagrange-Multiplikator und es gilt folgende Fehlerabschätzung:

$$\|u^* - u(\cdot, w^h)\|_{\mathcal{L}^\infty} + \|v^* - v^h\| + \|x^* - x^h\|_{\mathcal{L}^\infty} + \|\lambda^* - \lambda^h\|_{\mathcal{L}^\infty} = O(h) + \epsilon(h), \quad (4.58)$$

wobei h die maximale Schrittweite auf dem Steuerungsgitter bezeichnet und die Funktion $\epsilon(\cdot)$ in **A2** definiert ist.

Beweis

Mit Hilfe der Parametrisierung des Steuerungsraumes (siehe Abschnitt 3.3) und der Lösbarkeit der daraus entstandenen Aufgabe (siehe Annahme **A3(a)**) sind wir in Lage, unter Anwendung der Dreiecksungleichung folgende Ungleichung zu schreiben:

$$\|u^* - u(\cdot, w^h)\|_{\mathcal{L}^1} \leq \|u^* - u(\cdot, w^o)\|_{\mathcal{L}^1} + \|u(\cdot, w^o) - u(\cdot, w^h)\|_{\mathcal{L}^\infty}.$$

Die Annahme **A2** liefert die Abschätzung

$$\|u^* - u(\cdot, w^o)\|_{\mathcal{L}^1} \leq \epsilon(h).$$

Aus Lemma 5, (4.37) gilt der Diskretisierungsfehler

$$\|u(\cdot, w^o) - u(\cdot, w^h)\|_{\mathcal{L}^\infty} = O(h).$$

Es folgt die Abschätzung

$$\|u^* - u(\cdot, w^h)\|_{\mathcal{L}^1} \leq O(h) + \epsilon(h). \quad (4.59)$$

Mit Hilfe von x^o und der Dreiecksungleichung ist die folgende Zerlegung möglich

$$\|x^* - x^h\|_{\mathcal{L}^\infty} \leq \|x^* - x^o\|_{\mathcal{L}^\infty} + \|x^o - x^h\|_{\mathcal{L}^\infty}.$$

Wegen Lemma 5, (4.38) und Lemma 6, (4.56) sowie die Ungleichung (4.59) gilt die Abschätzung

$$\|x^* - x^h\|_{\mathcal{L}^\infty} \leq O(h) + \epsilon(h). \quad (4.60)$$

Aus Lemma 5, (4.39); Lemma 6, (4.57) sowie die Ungleichungen (4.59) und (4.60) gilt die Abschätzung

$$\|\lambda^* - \lambda^h\|_{\mathcal{L}^\infty} \leq O(h) + \epsilon(h). \quad (4.61)$$

Damit ist der Theorem 3 vollständig bewiesen \blacktriangleleft

Bemerkung 7 Für Riemann-integrierbare Funktionen u^* gilt der folgende Spezialfall (vgl. Theorem 6 in [32]):

$$\mu = 0, \quad \epsilon(h) = O(h^1)_{\mathcal{L}^1} = \tau_1(u^*, h), \quad (4.62)$$

wobei μ die Ordnung in $\mathbb{P}_{\mu, \mathcal{I}_h, \eta}$ in der Annahme **A4** ist, der Index \mathcal{L}^1 im Ausdruck $O(h^1)_{\mathcal{L}^1}$ die Norm bezeichnet, mit welcher der Parametrisierungsfehler gerechnet wird.

Das Resultat von Theorem 6 in [32] gilt nur für Steuerungsfunktionen, die durch stückweise konstante Funktionen approximierbar sind. Andersfalls wäre beispielsweise die Wahl des Referenzpunktes ungeeignet.

Die Diskretisierung mit anderen Verfahren liefern auch erfolgreiche Ergebnisse. In [119] wird nur die Konvergenz des Kollokationsverfahrens gezeigt. Diese Diskretisierungsmethode erweist sich als sehr effektiv für die Lösung einiger grossen Problemen aus der Anwendung [105, 119, 10, 54].

Schließlich nennen wir die Arbeiten von *Felgenhauer* [39, 38]. In [39] wird mit Hilfe von Projektoren Konvergenz mit Rate k unter der Glattheitsvoraussetzungen $x \in C^{k+1}$, $u \in C^k$ erzielt. Durch die Freiheit bei der Auswahl von Projektionsräumen, können in Abhängigkeit der Glattheit, Approximationsverfahren höherer Ordnung (Basisfunktionen höherer Ordnung) eingesetzt werden. In [38] wird das Ritzsche Verfahren erster Ordnung hergeleitet und analysiert. Als Einsatzräume bei der Diskretisierung betrachtet die Autorin stückweise konstante Funktionen für die Steuerung, stückweise lineare Funktionen für den Zustand und ein gemeinsames Gitter für die Steuerungs- und Zustandsvariable. Die Anwendung der stabilen Optimalitätsbedingungen [75] ermöglichen hier eine Konvergenzrate Eins zu erreichen.

4.4 Anwendung des Konvergenzsatzes

4.4.1 Runge-Kutta-Verfahren

Das Ziel dieses Abschnittes besteht in der Überprüfung der Voraussetzungen **A1** - **A9** im Falle der Approximation des dynamischen Systems durch das Runge-Kutta-Verfahren. Wir betrachten erneut die Verfahrensfunktion, beschrieben durch ein explizites Runge-Kutta-Verfahren (vgl. Gleichungen (3.45), (3.46)):

$$\varphi_l(x_{l-1}, w, v) := \sum_{i=1}^q b_i K_{li},$$

$$K_{li} = f\left(x_{l-1} + h_l \sum_{j=1}^{i-1} \alpha_{ij} K_{lj}, u(t_{l-1} + c_i h_l, w), v\right), \quad i = 1, \dots, q.$$

Die Voraussetzungen **A1** - **A3** und **A5** beziehen sich auf die ursprüngliche Aufgabe sowie auf die parametrisierte Aufgabe und sind dadurch von jeder Diskretisierung unabhängig.

Für die angesetzte Methode der Parametrisierung der Steuerung (siehe Abschnitt 3.3.1) sowie die anderen vorgeschlagenen Methoden der Parametrisierung im Abschnitt 3.3.2 sind die Voraussetzung von Annahme **A4** vollständig erfüllt.

Die Annahme **A6** ist an die positiven Konstanten c_1, c_2 (siehe Ungleichungen (3.30)) gebunden. Diese Konstanten sollen so gewählt werden, daß keine numerische Singularität auftritt. Die Voraussetzung **A7** ist eine Folgerung von **A3**. Es bleiben die Annahmen **A8** und **A9** zu verifizieren. Die Verifizierung von **A8** ist das Thema des Abschnittes 4.4.2.

Überprüfung von Annahme A9

Wir betrachten $\kappa = 0$. Es sei L_f bzw. L_{∇_x} , L_{∇_u} und L_{∇_v} die entsprechende Lipschitz-Konstante von f bzw. $\nabla_x f$, $\nabla_u f$ und $\nabla_v f$. Gemäß der Definition der Verfahrensfunktion beim Runge-Kutta-Verfahren ergibt sich die Ungleichungskette:

$$\begin{aligned}
& \|\varphi_{l+1}^o - \frac{1}{h_{l+1}}(x_{l+1}^o - x_l^o)\| \\
& \leq \frac{1}{h_{l+1}} \int_{t_l}^{t_{l+1}} \left| \sum_{i=1}^q b_i K_{l+1,i} - f(x^o(t), u(t, w^o), v^o) \right| dt \\
& \leq \frac{1}{h_{l+1}} \int_{t_l}^{t_{l+1}} \left| \sum_{i=1}^q b_i f\left(x_l^o + h_{l+1} \sum_{j=1}^{i-1} \alpha_{ij} K_{l+1,j}, u(t_l + c_i h_{l+1}, w^o), v^o\right) - f(x^o(t), u(t, w^o), v^o) \right| dt \\
& \leq L_f \frac{\max_{1 \leq i \leq q} b_i}{h_{l+1}} \int_{t_l}^{t_{l+1}} (|x_l^o + h_{l+1} \sum_{j=1}^{i-1} \alpha_{ij} K_{l+1,j} - x^o(t)| + |u(t, w^o) - u(t, w^o)|) dt \\
& \leq L_f \max_{1 \leq i \leq q} b_i [h_{l+1}^q + h_{l+1}] \\
& \leq c h_{l+1}.
\end{aligned}$$

mit einer passenden positiven Konstante c . Damit ist die Voraussetzung **A9(a)** nachgewiesen. Aus der Voraussetzung von Annahme **A3** und der Eigenschaften der Parameter des Runge-Kutta-Verfahrens und die Betrachtung $\kappa = 0$ folgt die Abschätzungskette:

$$\begin{aligned}
& \left\| \int_{t_l}^{t_{l+1}} (\nabla_x \varphi_l^o|_{(x^o, w^o, v^o)} - \nabla_x f(x^o(t), u(t, w^o), v^o)) dt \right\| \\
& = \left\| \int_{t_l}^{t_{l+1}} \left(\sum_{i=1}^q b_i \nabla_x f\left(x_{l-1}^o + h_l \sum_{j=1}^q \alpha_{ij} K_{l,j}, u(t_{l-1} + c_i h_l, w^o), v\right) \right)_{(x^o, w^o, v^o)} \right. \\
& \quad \left. - \nabla_x f(x^o(t), u(t, w^o), v^o) \right\| dt \\
& \leq L_{\nabla_x} \frac{\max_{1 \leq i \leq q} b_i}{h_{l+1}} \int_{t_l}^{t_{l+1}} (|x_{l-1}^o + h_l \sum_{j=1}^q \alpha_{ij} K_{l,j} - x^o(t)| + |u(t, w^o) - u(t, w^o)|) dt \\
& \leq c h_{l+1},
\end{aligned}$$

mit einer passenden Konstante c . Die erste Gleichung von **A9(b)** verschwindet wegen dem expliziten Verfahren. Es folgt somit die Voraussetzung **A9(b)**.

Wegen Annahme **A3(b)** und **A4**, (3.19) gilt folgende Ungleichungskette

$$\left\| \int_{t_l}^{t_{l+1}} (\nabla_w \varphi_l^o|_{(x^o, w^o, v^o)} - \nabla_u f(x^o(t), u(t, w^o), v^o) \nabla_w u(t, w^o)) dt \right\|$$

$$\begin{aligned}
&\leq = \left\| \int_{t_l}^{t_{l+1}} \left(\sum_{i=1}^q b_i \nabla_u f(x_{l-1}^o + h_l \sum_{j=1}^{i-1} \alpha_{ij} K_{l,j}, u(t_{l-1} + c_i h_l, w), v) \nabla_w u(t_{l-1} + c_i h_l, w^o) \right. \right. \\
&\quad \left. \left. - \nabla_u f(x^o(t), u(t, w^o), v^o) \nabla_w u(t, w^o) \right) dt \right\| \\
&\leq L_{\nabla_u} \cdot \max_{t \in [t_l, t_{l+1}]} \|\nabla_w u(t, w)\| \cdot \max_{1 \leq i \leq q} b_i \cdot \int_{t_l}^{t_{l+1}} (\|x_{l-1}^o + h_l \sum_{j=1}^{i-1} \alpha_{ij} K_{l,j} - x^o(t)\| \\
&\quad + \|u(t_{l-1} + c_i h_l, w^o) - u(t, w^o)\|) dt \\
&\leq c h_{l+1},
\end{aligned}$$

mit einer passenden Konstante c . Es folgt somit die Voraussetzung **A9(c)**. Analog werden die Voraussetzungen **A9(d)** nachgewiesen.

Die Voraussetzung **A9(e)** ist in Lemma 3 für ein allgemeines explizites Runge-Kutta-Verfahren ausführlich bewiesen. Damit ist die Annahme **A9** vollständig überprüft.

4.4.2 Stabile diskrete Koerzitivitätsbedingung

Wir wollen in diesem Abschnitt Voraussetzungen finden, so daß sich die diskreten Optimalitätsbedingungen zweiter Ordnung von Annahme **A8** im Falle des impliziten Runge-Kutta-Verfahrens aus der stetigen Voraussetzungen für die Optimalitätsbedingungen zweiter Ordnung von Annahme **A5** schlußfolgern lassen.

Wir betrachten erneut die Verfahrensfunktion beim Runge-Kutta-Verfahren (vgl. (3.45), (3.46)):

$$\begin{aligned}
\varphi_l(x_l, x_{l-1}, w, v) &:= \sum_{i=1}^q b_i K_{li}, \\
K_{li} &= f(x_{l-1} + h_l \sum_{j=1}^q \alpha_{ij} K_{lj}, u(t_{l-1} + c_i h_l, w), v), \quad i = 1, \dots, q.
\end{aligned}$$

Die Gleichung für die Bestimmung der Komponenten K_{li} , $i = 1, \dots, q$, $l = 1, \dots, N$ stellt eine implizite Funktion dar. Für die Lösung dieses Systems wird der implizite Funktionsatz eingesetzt und h hinreichend klein angenommen.

Lemma 7 *Es seien h die maximale Schrittweite auf dem Steuerungsgitter, $e : [0, 1] \rightarrow \mathbb{R}_+$ eine Funktion mit $\lim_{t \rightarrow 0} e(t) = 0$, $w = w_1 - w_2$, mit $w_1, w_2 \in W$, und $v = v_1 - v_2$, mit $v_1, v_2 \in V$. Es gelten die Voraussetzungen der Annahmen **A3-A6**. Die Funktion $\nabla_{((u,v),(u,v))}^2 f(x, \cdot, \cdot)$ sei Lipschitz-stetig für alle $x \in \mathbb{R}^n$ und die Ableitung $\nabla_{w,w}^2 u(t, \cdot)$ sei fast überall beschränkt in $[0, 1]$. Dann gilt die Abschätzung*

$$\left\| (w, v) \int_0^1 (\mathbf{H}[t] - \mathbf{H}^d) dt (w, v)^\top \right\| \leq e(h) (\|(w, v)\|^2) \quad (4.63)$$

Beweis:

Es seien die $\check{w}[t]$ bzw. \check{w}_l definiert als in (3.22) auf Seite 23, f aus \mathbb{R}^n , und wir bezeichnen mit f_m , $\check{w}[t]^m$ bzw. \check{w}_l^m die m -Komponente von f , $\check{w}[t]$ bzw. \check{w}_l . Wir betrachten zuerst die quadratische Terme bezüglich der Variablen w in (3.26) und (4.15)

$$\|\bar{w}^\top \int_0^1 (\check{w}[t] - \sum_{l=1}^N h_l \check{w}_l) dt \bar{w}\| = \|\sum_{l=1}^N \bar{w}^\top \int_{t_{l-1}}^{t_l} \delta \check{w}_l dt \bar{w}\|, \quad (4.64)$$

mit

$$\begin{aligned} \delta \check{w}_l &= (\delta^1 \check{w}_l, \dots, \delta^n \check{w}_l) \quad \delta^m \check{w}_l = \check{w}[t]^m - \check{w}_l^m, \quad m \in \{1, \dots, n\} \\ \check{w}[t]^m &= [\langle \nabla_w u(\cdot, w), \lambda^\top \nabla_{uu}^2 f_m(x, u(\cdot, w), v) \nabla_w u(\cdot, w) \rangle \\ &\quad + \langle \lambda^\top \nabla_u f_m(x, u(\cdot, w), v), \nabla_{w,w}^2 u(\cdot, w) \rangle](t)|_{(w^o, v^o, x^o, \lambda^o)} \\ \check{w}_l^m &= \langle \nabla_w u(t_{l-1} + c_l h_l, w^o), \lambda_l^\top \sum_{i=1}^q b_i \nabla_{uu}^2 f_m(x_{l-1}^o + \\ &\quad h_l \sum_{j=1}^q \alpha_{ij} K_{lj}, u(t_{l-1} + c_l h_l, w^o), v^o) \nabla_w u(t_{l-1} + c_l h_l, w^o) \rangle \\ &\quad + \langle \lambda_l^\top \sum_{i=1}^q b_i [\nabla_u f_m(x_{l-1}^o + h_l \sum_{j=1}^q \alpha_{ij} K_{lj}, u(t_{l-1} + c_l h_l, w^o), \\ &\quad v^o), \nabla_{w,w}^2 u(t_{l-1} + c_l h_l, w^o)] \rangle]. \end{aligned}$$

Aus (4.64) folgt die Ungleichung

$$\|\sum_{l=1}^N \bar{w}^\top \int_{t_{l-1}}^{t_l} \delta \check{w}_l dt \bar{w}\| \leq \sum_{m=1}^n \sum_{l=1}^N \|\bar{w}^\top \int_{t_{l-1}}^{t_l} \delta^m \check{w}_l dt \bar{w}\|, \quad (4.65)$$

Der (m, l) -te Term der Summe (4.65) ergibt:

$$\begin{aligned} &\|\bar{w}^\top \int_{t_{l-1}}^{t_l} \delta^m \check{w}_l dt \cdot \bar{w}\| \\ &\leq \|w\|^2 \cdot \int_{t_{l-1}}^{t_l} \|\langle \nabla_w u(\cdot, w^o), (\lambda^o)^\top \nabla_{uu}^2 f_m(x^o, u(\cdot, w^o), v^o) \nabla_w u(\cdot, w^o) \rangle(t) - \\ &\quad \langle \nabla_w u(t_{l-1} + c_l h_l, w^o), (\lambda_l^o)^\top \sum_{i=1}^q b_i \nabla_{uu}^2 f_m(x_{l-1}^o + h_l \sum_{j=1}^q \alpha_{ij} K_{lj}, u(t_{l-1} + c_l h_l, \\ &\quad w^o), v^o) \nabla_w u(t_{l-1} + c_l h_l, w^o) \rangle\| dt \\ &\quad + \|w\|^2 \cdot \int_{t_{l-1}}^{t_l} \|\langle (\lambda^o)^\top \nabla_u f_m(x^o, u(\cdot, w^o), v^o), \nabla_{w,w}^2 u(\cdot, w^o) \rangle(t) - \\ &\quad \langle (\lambda_l^o)^\top \sum_{i=1}^q b_i \nabla_u f_m(x_{l-1}^o + h_l \sum_{j=1}^q \alpha_{ij} K_{lj}, u(t_{l-1} + c_l h_l, w^o), \\ &\quad v^o), \nabla_{w,w}^2 u(t_{l-1} + c_l h_l, w^o) \rangle\| dt. \end{aligned}$$

Es seien L_{∇^2} die Lipschitz-Konstante von $\nabla_{((u,v),(u,v))}^2 f(x, \cdot, \cdot)$, L_{∇_u} die Lipschitz-Konstante von $\nabla_u f(x, \cdot, v)$, L_λ die Lipschitz-Konstante von λ° . Es seien c_u und c_{u2} gegeben durch

$$c_u := \sup_{t \in [0,1]} \|\nabla_{w,w} u(t, w^\circ)\|, \quad c_{u2} := \sup_{t \in [0,1]} \|\nabla_{w,w}^2 u(t, w^\circ)\|.$$

Wegen f sei zweimal stetig differenzierbar, $\lambda^\circ \in \mathcal{W}^{1,\infty}$ und $\sum_{i=1}^q b_i = 1$ folgen die Ungleichungen:

$$\begin{aligned} & \left\| \bar{w}^\top \int_{t_{i-1}}^{t_i} \delta^m \check{\mathbf{w}}_l dt \cdot \bar{w} \right\| \\ & \leq \|w\|^2 \cdot \int_{t_{i-1}}^{t_i} (c_u^2 \{L_\lambda \|\lambda^{\circ\top}(t) - (\lambda_l^\circ)^\top\| \\ & \quad + \max_{i \in \{1, \dots, q\}} b_i [L_{\nabla^2} (\|x^\circ(t) - x_{i-1}^\circ - h_l \sum_{j=1}^q \alpha_{ij} K_{lj}\| + \|u(t, w^\circ) - u(t_{i-1} + c_i h_l, w^\circ)\|)]\} \\ & \quad + c_{u2} \{L_\lambda \|\lambda^{\circ\top}(t) - (\lambda_l^\circ)^\top\| \\ & \quad + \max_{i \in \{1, \dots, q\}} b_i [L_{\nabla_u} (\|x^\circ(t) - x_{i-1}^\circ - h_l \sum_{j=1}^q \alpha_{ij} K_{lj}\| + \|u(t, w^\circ) - u(t_{i-1} + c_i h_l, w^\circ)\|)]\}) dt \\ & \leq c h_l^2. \end{aligned}$$

mit c eine passende Konstante. Diese Abschätzung gilt für alle (m, l) Komponente mit $m \in \{1, \dots, n\}$, $l \in \{1, \dots, N\}$.

Aus der Ungleichung (3.30) von Annahme **A6** ergibt sich die Abschätzung:

$$\left\| \bar{w}^\top \int_{t_{i-1}}^{t_i} \delta \check{\mathbf{w}}_l dt \cdot \bar{w} \right\| \leq e(h) \cdot \|w\|^2. \quad (4.66)$$

Analog zu (4.66) werden folgende Ungleichungen bewiesen

$$\left\| \bar{v}^\top \int_{t_{i-1}}^{t_i} \delta \check{\mathbf{v}}_l dt \cdot \bar{v} \right\| \leq e(h) \cdot \|v\|^2. \quad (4.67)$$

$$\left\| \bar{w}^\top \int_{t_{i-1}}^{t_i} \delta \check{\mathbf{r}}_l dt \cdot \bar{v} \right\| \leq e(h) \cdot \|w\| \|v\|. \quad (4.68)$$

$$\left\| \bar{v}^\top \int_{t_{i-1}}^{t_i} \delta \check{\mathbf{s}}_l dt \cdot \bar{w} \right\| \leq e(h) \cdot \|v\| \|w\|. \quad (4.69)$$

Schließlich ergeben die Dreiecksungleichung und die Ungleichungen (4.66), (4.67), (4.68), (4.69) die Abschätzung (4.63). Damit ist die Ungleichung (4.63) gezeigt.

Kapitel 5

Optimierungsverfahren

Nach der Diskretisierung der ursprünglichen Aufgabe entsteht eine endlich-dimensionale Aufgabe, die mit angepaßten nichtlinearen Optimierungsverfahren (nonlinear programming methods, vgl. [44, 6, 97, 98, 100, 101, 103]) gelöst werden kann.

In der vorliegenden Arbeit wird ein sequentielles quadratisches Programmierungsverfahren¹ („sequential quadratic programming“) angewendet. Das SQP-Verfahren hat sich durch seine Zuverlässigkeit und seine Effektivität als eine der besten Methoden für die Lösung von Optimierungsproblemen mit stetigen und nichtlinearen Zielfunktionen und Beschränkungen erwiesen (vgl. [102, 6, 93, 108, 112, 7, 8]²). Im Vergleich mit anderen Optimierungsverfahren benötigt das SQP-Verfahren bemerkenswert wenig Auswertung der Problemfunktionen und Gradienten, und konvergiert zu einer Lösung unter sehr milden Bedingungen bezüglich des Problems (vgl. [93, 99, 98]). Weitere Arbeiten zum Thema SQP-Verfahren sind zum Beispiel [9, 113].

Wir beschreiben an dieser Stelle kurz das SQP-Verfahren und orientieren uns dabei an der Darstellung von *Gill, Murray* und *Saunders* [44]. Andere Beschreibungen des SQP-Verfahrens sind zum Beispiel in [42, 100, 46, 95] zu finden.

Es seien $F : \mathbb{R}^n \rightarrow \mathbb{R}$ eine lineare oder nichtlineare Funktion, $C : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine Vektorfunktion mit den Komponenten $C_i(x)$, $i = 1, \dots, m$. Die Funktionen $F(x)$ und $C_i(x)$

¹ Der Ausdruck: sequentiell quadratisch Programmierungsverfahren wird nachfolgen mit dem Begriff SQP-Verfahren abgekürzt

² In diesen Literatur werden numerische Resultate für die meisten Probleme in der CUTE-Testansammlung und für eine Anzahl von anderen Anwendungen, einschließlich der Flugbahnoptimierung in dem Luftfahrt gegeben

besitzen stetige zweite Ableitungen. Wir bezeichnen mit $g(x)$ bzw. $\mathcal{J}(x)$ den Gradienten von F bzw. die Jacobi-Matrix der Funktion C . Wir betrachten das Optimierungsproblem:

$$\min F(x) \quad x \in \mathbb{R}^n \quad \text{bzgl.} \quad C(x) \geq 0, \quad (5.1)$$

und nehmen folgendes voraus:

Es existiere ein Punkt (der Karush-Kuhn-Tucker-Punkt) (x^*, λ^*) , so daß die Optimalitätsbedingungen erster Ordnung erfüllt sind:

$$C(x^*) \geq 0, \quad \lambda^* \geq 0, \quad C(x^*)^\top \lambda^* = 0, \quad \mathcal{J}(x^*)^\top \lambda^* = g(x^*). \quad (5.2)$$

5.1 Grundlegende Idee des SQP-Verfahrens

Die grundlegende Idee des SQP-Verfahrens ist die Formulierung und die Lösung eines quadratischen Teilproblems (QP-Teilproblem) zur Bestimmung der Suchrichtung in jeder Iteration. Das QP-Teilproblem minimiert ein Modell aus einer modifizierten Lagrange-Funktion (vgl. [60, 87]) bezüglich der linearisierten Beschränkungen. Dabei wird eine gewisse Bewertungsfunktion (Gütefunktion) entlang jeder Suchrichtung verringert, um die Konvergenz von irgendeinem Ausgangspunkt sicherzustellen. Die Grundstruktur des SQP-Verfahrens basiert auf einer Major-Iteration und einer Minor-Iteration. Die Major-Iteration erzeugt eine Folge von Iterierten (x^k, λ^k) , die gegen (x^*, λ^*) konvergieren soll. Die Minor-Iteration löst mit Hilfe einer iterativen Prozedur das QP-Teilproblem.

5.2 QP-Teilproblem

Es sei (x^k, λ^k) die aktuelle Iterierte mit dem Index k . Wir betrachten die modifizierte Lagrange-Funktion:

$$\mathcal{L}(x, x^k, \lambda^k) = F(x) - (\lambda^k)^\top d_L(x, x^k), \quad \text{mit} \quad \begin{aligned} d_L(x, x^k) &= C(x) - C_L(x, x^k), \\ C_L(x, x^k) &= C(x^k) + \mathcal{J}(x^k)(x - x^k). \end{aligned}$$

Die erste und zweite Ableitung der modifizierten Lagrange-Funktion bezüglich x ergibt:

$$\nabla_x \mathcal{L}(x, x^k, \lambda^k) = g(x) - (\mathcal{J}(x) - \mathcal{J}(x^k))^\top \lambda^k, \quad (5.3)$$

$$\beta^k(x) = \nabla_{xx}^2 \mathcal{L}(x, x^k, \lambda^k) = \nabla_{xx}^2 F(x) - \sum_{i=1}^m (\lambda^k)_i \nabla_{xx}^2 C_i(x). \quad (5.4)$$

Wir stellen fest: die Hesse-Matrix $\beta_k(x)$ der modifizierten Lagrange-Funktion unterscheidet sich nicht von der Hesse-Matrix der üblichen Lagrange-Funktion. Außerdem haben die

modifizierte Lagrange-Funktion und die Zielfunktion die gleichen Funktionswerte und die gleiche erste Ableitung im Punkt $x = x^k$,

$$\mathcal{L}(x^k, x^k, \lambda^k) = F(x^k), \quad \nabla_x \mathcal{L}(x^k, x^k, \lambda^k) = g(x^k).$$

Es sei \mathcal{Q} die quadratische Approximation von \mathcal{L} im Punkt x^k , gegeben durch:

$$\mathcal{Q}(x, x^k, \lambda^k) = F(x^k) + g(x^k)(x - x^k) + \frac{1}{2}(x - x^k)^\top \beta^k(x)(x - x^k).$$

Wir betrachten ein QP-Teilproblem, beschrieben durch folgendes Problem:

$$\min \mathcal{Q}(x, x^k, \lambda^k), \quad x \in \mathbb{R}^n, \quad \text{bzgl.} \quad C_L(x, x_k) \geq 0. \quad (5.5)$$

Falls $(x^k, \lambda^k) = (x^*, \lambda^*)$ gilt, dann haben die Aufgabe (5.5) und das Optimierungsproblem (5.1) identische Optimalitätsbedingungen erster Ordnung. Aus der obigen Darstellung empfiehlt sich folgende Überlegung:

Wenn $\bar{\beta}^k(x_k)$ ein Näherungswert zu $\beta^k(x_k) = \nabla_{xx}^2 \mathcal{L}(x^k, x^k, \lambda^k)$ am Punkt (x^k, λ^k) ist, dann kann eine verbesserte Schätzung der Lösung durch den Punkt $(\hat{x}^k, \hat{\lambda}^k)$ gefunden werden. Dabei ist $(\hat{x}^k, \hat{\lambda}^k)$ eine Lösung des folgenden quadratischen Teilproblems

$$\min_{x \in \mathbb{R}^n} F(x^k) + g(x^k)(x - x^k) + \frac{1}{2}(x - x^k)^\top \bar{\beta}^k(x^k)(x - x^k) \text{ bzgl. } C_L(x, x_k) \geq 0. \quad (5.6)$$

Die Optimalitätsbedingungen des Teilproblems (5.6) ergeben:

$$\begin{aligned} g(x^k) + \bar{\beta}^k(x^k)(\hat{x}^k - x^k) &= \mathcal{J}(\hat{x}^k)^\top \hat{\lambda}^k, \\ C(x^k) + \mathcal{J}(x^k)(\hat{x}^k - x^k) &= \hat{s}^k, \quad (\hat{\lambda}^k)^\top \cdot \hat{s}^k = 0, \\ \hat{\lambda}^k &\geq 0, \quad \hat{s}^k \geq 0, \end{aligned} \quad (5.7)$$

wobei \hat{s}^k ein Vektor der Schlupfvariablen („slack variable“) für die linearisierten Beschränkungen ist. Das Tripel $(\hat{x}^k, \hat{\lambda}^k, \hat{s}^k)$ kann als Schätzung der Lösung (x^*, λ^*, s^*) angesehen werden, wo die nichtnegative Variable s^* der Gleichung $C(x^*) - s^* = 0$ genügt (d.h., s^* ist der Vektor, der die Ungleichung $C(x^*) \geq 0$ in (5.2) in die Gleichung $C(x^*) = s^*$ überführt).

5.3 Bewertungsfunktion

Gleich nach der Bestimmung von $(\hat{x}^k, \hat{\lambda}^k, \hat{s}^k)$ wird die neue Iterierte mittels der Liniensuchmethode mit einer gewissen Bewertungsfunktion bestimmt. Falls (x^k, λ^k, s^k) die aktuel-

le Iterierte von (x^*, λ^*, s^*) ist, dann bestimmt die Liniensuch-Methode eine Schrittweite α_k , ($0 < \alpha_k \leq 1$), so daß die neue Iterierte, durch:

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \\ s^{k+1} \end{pmatrix} := \begin{pmatrix} x^k \\ \lambda^k \\ s^k \end{pmatrix} + \alpha_k \begin{pmatrix} \hat{x}^k - x^k \\ \hat{\lambda}^k - \lambda^k \\ \hat{s}^k - s^k \end{pmatrix}$$

gegeben wird und den Abstieg der betrachteten Bewertungsfunktion entlang jeder Suchrichtung gewährleistet. Es sei ψ_k eine geeignete Straffunktion, (x^k, λ^k, s^k) die aktuelle Iterierte von (x^*, λ^*, s^*) gegeben. Wir betrachten die Bewertungsfunktion:

$$\Gamma(\alpha_k) := \psi_k \left[x^k + \alpha_k(\hat{x}^k - x^k), \lambda^k + \alpha_k(\hat{\lambda}^k - \lambda^k), s^k + \alpha_k(\hat{s}^k - s^k) \right].$$

Falls α_k ($0 < \alpha_k \leq 1$) den Abstieg der eindimensionalen Funktion $\Gamma(\cdot)$ verursacht, dann ist $(x^{k+1}, \lambda^{k+1}, s^{k+1})$ die neue Iterierte. In der NPSOL- bzw. SNOPT-Routine (siehe [43, 45]) oder NLPQL-Routine (siehe [101]) wird die Liniensuch-Methode für die Berechnung der Schrittweite verwendet. Für die Wahl der Straffunktion ψ_k verweisen wir auf [41, 56, 86] für nicht differenzierbare l_1 -Funktionen und auf [44, 99, 98] für differenzierbare erweiterte Lagrange-Funktionen. Wir verweisen auch auf [82] für einige Bedingungen zur Konvergenz der Liniensuch-Methode.

5.4 Approximation der Hesse-Matrix

Die Matrix $\beta_k(x)$ in (5.4) wird in jedem Iterationsschritt neu berechnet. Dies erfordert die Berechnung der zweiten Ableitungen, die sehr rechen- und zeitaufwendig sind. Zur Vermeidung der Berechnung der zweiten Ableitungen und zum Erreichen einer dennoch überlinearen Konvergenz des Verfahrens wird die Schätzung $\bar{\beta}_k$ der Hesse-Matrix β_k von \mathcal{L} meist nach der Aufdatierungsformel von *Broyden, Fletcher, Goldfarb und Shanno* (BFGS-Formel) erneuert (vgl. [93]). Die BFGS-quasi-Newton-Formel erzeugt eine Approximation der Hesse-Matrix von \mathcal{L} für die nächste Iteration $\bar{\beta}_{k+1}$. Dabei hängt die Berechnung von $\bar{\beta}_{k+1}$ nur von $\bar{\beta}_k$, von der Änderung q_k des Gradienten der Lagrange-Funktion bezüglich x und von der Änderung p_k in x zwischen zwei Integrationsschritten ab, das heißt:

$$q_k := \nabla_x \mathcal{L}(x^{k+1}, \gamma_k) - \nabla_x \mathcal{L}(x^k, \gamma_k), \quad (5.8)$$

$$p_k := x^{k+1} - x^k, \quad (5.9)$$

$$\bar{\beta}_{k+1} := \bar{\beta}_k + \frac{1}{q_k^\top p_k} q_k q_k^\top - \frac{1}{(\bar{\beta}_k p_k)^\top p_k} \bar{\beta}_k p_k (\bar{\beta}_k p_k)^\top. \quad (5.10)$$

Falls $\bar{\beta}_k$ positiv definit und $q_k p_k$ positiv ist, dann ist nach der Definition von q_k , p_k und $\bar{\beta}_k$ auch $\bar{\beta}_{k+1}$ positiv definit (vgl. [29]). Zur Eigenschaft positiv definit der Matrix $\bar{\beta}_k$ schlagen Gill, Murray und Saunders zwei Modifikationsmöglichkeiten von p_k und q_k in [44] vor.

5.5 Abbruchkriterien

Das SQP-Verfahren endet mit Erfolg, falls die Folge der Iterierten $\{(x^k, \lambda^k)\}$ konvergiert und der Grenzwert (x^*, λ^*) die Optimalitätsbedingungen (5.2) innerhalb einer bestimmten Toleranz erfüllt. Das heißt in der Praxis, für $\delta_x = \delta_P(1 + \|x^k\|)$ und $\delta_\lambda = \delta_D(1 + \|\lambda^k\|)$ gelten die Abschätzungen:

$$\begin{aligned} C_i(x^k) &\geq -\delta_x, & \lambda_i^k &\geq -\delta_\lambda, \\ C_i(x^k)\lambda_i^k &\leq \delta_\lambda, \\ |d_i| &\leq \delta_P, \end{aligned} \tag{5.11}$$

mit $d = (d_1, \dots, d_n) = g(x^k) - \mathcal{J}(x^k)^\top \lambda^k$, wobei δ_P und δ_D zwei gegebene kleine positive Konstanten sind (Genauigkeiten des Optimierungsverfahrens). Diese Abschätzungen können nicht erfüllt werden, falls das Original-Problem keinen zulässigen Punkt besitzt. In diesem Fall geht das SQP-Verfahren in einen elastischen Modus über und führt den gleichen Test für das Problem

$$\min F(x) + \rho e^\top v, \quad x \in \mathbb{R}^n, \quad \text{bzgl. } C(x) + v \geq 0, \quad v \geq 0, \tag{5.12}$$

mit $e = (1, \dots, 1)^\top$ aus. Im elastischen Modus kommen folgende Bedingungen:

$$0 \leq \lambda_i \leq \rho, \quad (C_i(x) + v_i)\lambda_i = 0, \quad v_i(\rho - \lambda_i) = 0 \tag{5.13}$$

zu den Optimalitätsbedingungen hinzu. Dabei ist die Bedingung $\|\lambda\| \leq \rho$ für eine mögliche Lösung entscheidend für den Übergang zum elastischen Modus, falls in der aktuellen Iteration die Norm $\|\lambda^k\|$ den Strafparameter ρ überschreitet.

Hinsichtlich der numerischen Lösung des Problems (5.6) bemerken wir, daß es für gewisse x^k und $\hat{\beta}_k$ keine zulässige Lösung geben muß, obwohl das Problem (5.1) lösbar ist. Grund dafür ist die inkonsistente linearisierte Beschränkung. Zur Aufhebung des Mangels wird eine zusätzliche Variable eingeführt. Dadurch entsteht ein $(n+1)$ -dimensionales quadratisches Teilproblem mit konsistenten Beschränkungen (vgl. [86, 99]).

Kapitel 6

Implementierung des Algorithmus

6.1 Beschreibung des Algorithmus

Wie in der Einleitung erwähnt, haben wir im diesem praktischen Teil dieser Arbeit einen Algorithmus für die numerische Lösung der diskreten Aufgabe implementiert. Der Algorithmus wird RINDOPT genannt mit folgender Bedeutung:

- R Runge-Kutta-Verfahren,
- IND interne numerische Differentiation,
- OPT Optimierung.

Speziell für diese Arbeit wird das Runge-Kutta-Merson-Verfahren mit den Koeffizienten in der Tabelle 6.1 eingesetzt und welches in der NAG-Routine D02PAF [68] implementiert sind.

Tabelle 6.1: Koeffizienten des Runge-Kutta-Merson-Verfahrens

C^0	$\alpha_{i,j}$					B^0
0	0	0	0	0	0	1/6
1/3	1/3	0	0	0	0	0
1/3	1/6	1/6	0	0	0	0
1/2	1/8	0	3/8	0	0	2/3
1	1/2	0	-3/2	2	0	1/6

Die interne numerische Differentiation wurde ausführlich im Abschnitt 3.7 dargestellt. Für die Optimierung wird die SQP-Routine SNOPT verwendet. Vor dem Start der Routine SNOPT muß der Anwender die Routinen objfun und confun bereitstellen. Der Grundaufbau für

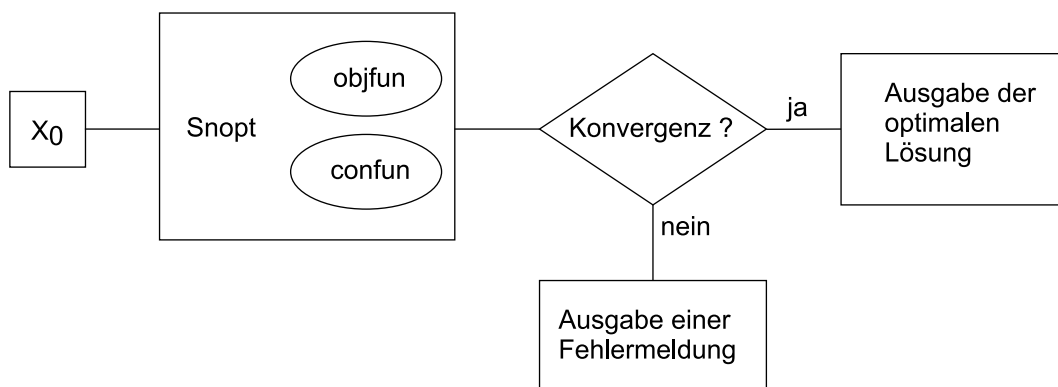


Abbildung 6.1: RINDOPT-Algorithmus

den RINDOPT-Algorithmus wird in der Abbildung 6.1 dargestellt. Es sei w_o die Startlösung und w^m die m -te Iterierte. Mit Hilfe der Routine `objfun` werden die Zielfunktion g_0 und ihre ersten partiellen Ableitungen

$$\nabla_{w_k} g_0(x_N), \quad k = 1, \dots, 2M,$$

an der aktuellen Iterierte w^m berechnet.

6.1.1 Gradientenberechnung bei explizitem Runge-Kutta-Verfahren

Die Berechnung der ersten Ableitung der Zielfunktion läuft über die Methode der internen numerischen Differentiation. Diese Methode erfordert die diskreten Zustandspunkte x_l , $l = 1, \dots, N$ (vgl. Gleichung (3.33)), die diskreten Vektoren λ_l , $l = 1, \dots, N$ (vgl. Gleichungen (3.47)-(3.48)), die Funktionswerte von φ_l (vgl. (3.45)) in jedem Zustandsgitterpunkt und die Ableitung von K_{li} , $l = 1, \dots, N$, $i = 1, \dots, q$ (vgl. (3.46)). Die Berechnung der Ableitung von g_0 nach der Komponente w_k mittels internen numerischen Differentiation ergibt für explizite Runge-Kutta-Verfahren folgende vereinfachte Formeln, die rekursiv zu berechnen sind.

Wir setzen für $k \leq 2M$, $w_k = u_j^*$ mit $* \in \{+, -\}$ und betrachten die Funktion

$$F(w) = g_0(x_N), \quad w \in W.$$

Für $w_k = u_j^*$, $k \leq 2M$, $* \in \{+, -\}$, $j = 1, \dots, M$,

$$\nabla_{u_j^*} F = \sum_{l=l(j)+1}^{l(j+1)} \lambda_l^\top h_l \nabla_{u_j^*} \varphi_l + \nabla_{u_j^*} g_0(x_N), \quad (6.1)$$

$$\varphi_l = \sum_{i=1}^q b_i K_{li}, \text{ mit } K_{li} = f(x_{l-1} + h_l \sum_{j=1}^{i-1} \alpha_{ij} K_{lj}, u(t_{l-1} + c_i h_l, w), v), \quad (6.2)$$

$$Y_{li} = \nabla_{x_{l-1}} K_{li} = \nabla_x f|_{li} \left(I + h_l \sum_{j=1}^{i-1} \alpha_{ij} Y_{lj} \right), \quad i = 1, \dots, q. \quad (6.3)$$

wobei für $l = 1, \dots, N$, $i = 1, \dots, q$,

$$\nabla_x f|_{li} = \nabla_x f(x_{l-1} + h_l \sum_{j=1}^{i-1} \alpha_{ij} K_{lj}, u(t_{l-1} + c_i h_l, w), v),$$

$$\nabla_u f|_{li} = \nabla_u f(x_{l-1} + h_l \sum_{j=1}^{i-1} \alpha_{ij} K_{lj}, u(t_{l-1} + c_i h_l, w), v).$$

Berechnung von λ_l ,

$$\lambda_N = [\nabla_x g_0(x_N)]^\top, \quad (6.4)$$

$$\lambda_{l-1} = \lambda_l + h_l \sum_{i=1}^q b_i (Y_{li}^\top \lambda_l), \quad l = N, \dots, 1, \quad (6.5)$$

Berechnung der Ableitung von φ_l , $l = 1, \dots, N$ nach w_k , $k = 1, \dots, 2M$

$$\nabla_{w_k} \varphi_l = \sum_{i=1}^q b_i \nabla_{w_k} K_{li}.$$

Fall $w_k = u_j^*$, $k \leq 2M$, $* \in \{+, -\}$, $j = 1, \dots, M$, $t = t_{l-1} + c_i h_l$

$$\nabla_{w_k} K_{li} = \nabla_x f|_{li} \cdot \left(h_l \sum_{j=1}^{i-1} \alpha_{ij} \nabla_{w_k} K_{lj} \right) + \nabla_u f|_{li} \cdot (\nabla_{w_k} u(t, w), v).$$

Der Aufbau der Routine `objfun` ist in der Abbildung 6.2 zu finden. Dabei spielen die Routinen `pxfeld`, `plfeld`, `pphi` die wesentliche Rolle.

`pxfeld` ist die Routine, mit deren Hilfe die diskreten Zustandspunkte berechnet werden,

`plfeld` berechnet die gesamten diskreten Vektoren λ_l , $l = 1, \dots, N$, und

`pphi` berechnet die diskreten Funktionen φ_l , $l = 1, \dots, N$.

Der Ablaufplan der Routine `pxfeld` wird im folgenden Algorithmus für die implizite Berechnung des diskreten Zustandes unter Berücksichtigung von zwei Zerlegungen ausführlich beschrieben.

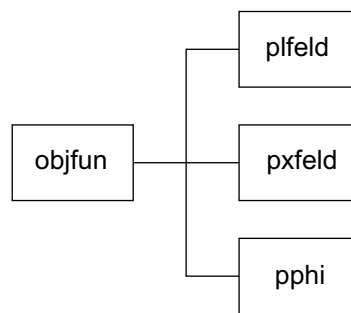


Abbildung 6.2: OBJFUN-Routine

6.1.2 Implizite Berechnung des Zustandes

Nach jeder Verbesserung der Optimierungsvariablen durch den Optimierungsalgorithmus hat der Parameter w einen neuen Wert, und somit folgt ein neuer Verlauf der Steuerung $u(\cdot, w)$. Diese neue Schätzung der Steuerung wird dann in das diskrete dynamische System eingesetzt, und danach folgt die numerische Neubestimmung des Zustandes.

Es seien

M-1 die Anzahl der internen Steuerungsgitterpunkte,

$t^0(j)$ die Steuerungsgitterpunkte mit $j = 1, \dots, M+1$, $t^0(1) = \tau_0 = 0$, $t^0(j) = \tau_{j-1}$,
 $j = 2, \dots, \tau_M$ und $t^0(M+1) = \tau_M = 1$,

U^0 ein Vektor mit der Eigenschaft $U^0(2j-1) = u_j^+$ bzw. $U^0(2j) = u_j^-$, $j = 1, \dots, M$ (siehe die Definition in (3.6)),

N+1 die Anzahl der Zustandsgitterpunkte,

t_k die Zustandsgitterpunkte mit $k = 1, \dots, N+1$,

$\{l(j)\}_{j=1}^{M+1}$ eine ganzzahlige Folge mit der Eigenschaft $t_{l(j)} = \tau_j$, $j = 1, \dots, M+1$, $t_{l(1)} = \tau_0$ und $t_{l(M+1)} = \tau_M$ (die Elemente der Folge $l(j)$ ermöglichen zu erkennen, welche Zustandsgitterpunkte Steuerungsgitterpunkte sind).

Es sei j der Index für das Intervall $[t^0(j), t^0(j+1)]$. Wir setzen für den Index j :

$$t_b = t^0(j),$$

$$t_e = t^0(j+1),$$

$$u_b = U^0(2j-1),$$

$$u_e = U^0(2j).$$

Mit diesen Ansätzen ist die Funktion $u(t, w)$ im Intervall $[t_b, t_e]$ eindeutig bestimmt. Der Anfangswert für das Integrationsverfahren im Intervall $[t_b, t_e]$ ist $x_{Start} = x_{l(j)}$.

Mit Hilfe eines expliziten Runge-Kutta-Verfahrens errechnen wir die Zustandswerte

$$x_{l(j)+1}, \dots, x_{l(j+1)} \quad (6.6)$$

in den Zustandsgitterpunkten

$$t_{l(j)+1}, \dots, t_{l(j+1)} = t^0(j+1). \quad (6.7)$$

Das Verfahren wird für $j = 1, \dots, M$ wiederholt durchgeführt.

Bei der Bestimmung diskreter Zustände (6.6) auf dem Teilintervall $[t^0(j), t^0(j+1)]$, $j = 1, \dots, M$ spielt die Routine D02PAF [68] für die numerische Lösung des dynamischen Systems die wesentliche Rolle. Dabei wird das Zustandsgitter (6.7) beim Iterationsverfahren automatisch erzeugt. Eine Vielzahl technischer Optionen für die Minderung des Fehlers und die Beschleunigung der Berechnungen wird in der Routine D02PAF zur Verfügung gestellt. Siehe beispielsweise Optionen für die Art der zu verwendenden Fehlertypsteuerung (gemischter Fehlertest, absoluter Fehlertest und relativer Fehlertest), für die kleinste bzw. größte mögliche Schrittweite, für die Spezifikation der Ausgangsschrittweite, für die Veränderung bestimmter Konstanten, die in der Wahl der folgenden Schrittweite verwendet werden und schließlich die Abbruchkriterien für die Unterbrechung endloser Berechnungen.

6.1.3 Interne Routinen

Die Routinen objfun und confun werden von folgenden Unterroutinen unterstützt:

f cn3 . f	Berechnung der rechten Seite des dynamischen Systems,
dpdgl . f	Berechnung der Ableitung der rechten Seite des dynamischen Systems nach dem Parameter v ,
inputfile	Übergabe der problemabhängigen Parameter,
pdklqt0lf . f	Berechnung von $\nabla_{\tau_{j+1}} F$ für $t \in (t_{l(k)}, t_{l(k+1)})$, (d.h. $t = t_l$, mit $l \in \{l(k), l(k+1)\}$),
pdklqt0rf . f	Berechnung von $\nabla_{\tau_{j+1}} F$ für $t \in (t_{l(k+1)}, t_{l(k+2)})$, (d.h. $t = t_l$, mit $l \in \{l(k+1), l(k+2)\}$),
pdklqu0 . f	Berechnung von $\nabla_{u_j^*} F$ für $* \in \{+, -\}$, $j = 1, \dots, M$,
pk1q . f	Berechnung von $K_{l,i}$, $i = 1, \dots, q$ für ein $l \in \{1, \dots, N\}$,
lambda . f	Berechnung von λ_l für ein $l \in \{1, \dots, N\}$,
pmaty . f	Berechnung von $Y_{l,i}^1$ und $Y_{l+1,i}^2$, $i = 1, \dots, q$ für ein $l \in \{0, \dots, N-1\}$,

psklq.f	Berechnung von $\sum_{j=1}^{i-1} \alpha_{ij} K_{lj}$, $i = 1, \dots, q$ für ein $l \in \{1, \dots, N\}$,
psmatu.f	Berechnung von $\sum_{j=1}^{i-1} \alpha_{ij} \nabla_{u_j^*} K_{lj}$, $i = 1, \dots, q$ für ein $l \in \{1, \dots, N\}$,
psmaty.f	Berechnung von $\sum_{j=1}^{i-1} \alpha_{ij} Y_{lj}$, $i = 1, \dots, q$ für ein $l \in \{1, \dots, N\}$.

Analog zu der Routine objfun wird die Routine confun aufgebaut. Die oben aufgelisteten Routinen werden von anderen Hilfsroutinen aus der NAG-Library[68] unterstützt. Im nächsten Abschnitt beschreiben wir kurz die SNOPT-Routine und orientieren uns dabei an [45, 44].

6.2 Beschreibung der SNOPT-Routine

Die Routine SNOPT ist eine Implementierung eines SQP-Verfahrens für die Lösung von Optimierungsproblemen mit wenigen Freiheitsgraden¹. Die Besonderheit von SNOPT ist seine Fähigkeit Probleme mit aufwändigen Funktionen- und Gradientenberechnungen lösen zu können (siehe Testbeispiele in [44]) und die Ausnutzung der Struktur von Problemen bei dessen Lösung (vgl. Abschnitt 6.2.2 Seite 76). Die Routine SNOPT ist eine Zusammensetzung von mehreren Routinen, welche in FORTRAN 77 geschrieben sind. Für große Probleme mit wenigen Freiheitsgraden existieren neben SNOPT Routine wie CONOPT [36], MINOS [83]. CONOPT betrachtet eine verallgemeinerte Methode der reduzierten Gradienten („generalized reduced gradient method“). MINOS verwendet eine „projected Lagrangian or sequential linearly constrained method“.

Für die Lösung von großen Problemen mit vielen Freiheitsgraden ist die Routine LANCELOT [23] empfehlenswert. Die Routine LANCELOT verwendet die „sequential augmented Lagrangian method“.

¹ Die Zahl Freiheitsgraden ergibt sich aus der Zahl der Variablen minus der Zahl aktiven constraints.

6.2.1 Lösungsprozeß der SNOPT-Routine

Es seien F eine glatte skalare Funktion, b^l und b^u zwei konstante Vektoren, C^L eine schwach besetzte konstante Matrix und $C^N(x)$ ein Vektor glatter nichtlinearer Funktionen $C_i^N(x)$. Wir betrachten die Aufgabe:

$$\min_{x \in \mathbb{R}^n} F(x) \quad \text{bzgl.} \quad b^l \leq \begin{pmatrix} x \\ C^N(x) \\ C^L x \end{pmatrix} \leq b^u, \quad (6.8)$$

F ist die Zielfunktion, b^l und b^u sind die unteren und oberen Schranken, $C^L x$ bzw. $C^N(x)$ ist der lineare bzw. nichtlineare Anteil der Beschränkungen.

Es sei m die Anzahl der globalen (linearen und nichtlinearen) Beschränkungen. Im Lösungsprozeß der SNOPT-Routine wird die Aufgabe zweimal transformiert. Zuerst wird das Problem in ein Problem niedriger Dimension transformiert, indem die Zielfunktion in einen linearen und einen nichtlinearen Anteil zerlegt wird. Dabei wird die Optimierungsvariable in lineare und nichtlineare Komponenten zerlegt. In SNOPT-Routine werden nur die nichtlinearen Komponenten optimiert. Danach werden die globalen Beschränkungen durch Einführung einer Schlupfvariablen $s = (s_1, \dots, s_m)^\top$ in ein Gleichungssystem überführt. Die Aufgabe erhält somit folgende Form:

$$\min_{x \in \mathbb{R}^n} F(x) \quad \text{bzgl.} \quad \begin{pmatrix} C^N(x) \\ C^L x \end{pmatrix} - \begin{pmatrix} s_N \\ s_L \end{pmatrix} = 0, \quad b^l \leq \begin{pmatrix} x \\ s \end{pmatrix} \leq b^u, \quad (6.9)$$

mit $s = (s_N, s_L)$, wobei s_N bzw. s_L nichtlineare bzw. lineare Schlupfvariable sind. Ähnlich wie beim SQP-Verfahren besitzt die SNOPT-Routine eine Minor- bzw. Major-Iteration. Die Major-Iteration generiert eine Folge von Iterierten x^k , die die lineare Beschränkung erfüllen. Diese Folge konvergiert gegen einen Punkt, der den ersten Optimalitätsbedingungen genügt. In jeder Iteration kommt für die Bestimmung der Suchrichtung zur nächsten Iterierten x^{k+1} die Lösung eines quadratischen Teilproblems (QP-Teilproblem) zur Anwendung. Das QP-Teilproblem minimiert eine quadratische Approximation der modifizierten Lagrange-Funktion $q(x)$ bezüglich der linearen Beschränkung $C^L x - s_L = 0$ und der linearisierten nichtlinearen Beschränkung:

$$\nabla_x C^N(x^k)x - x_S = -C^N(x^k) + \nabla_x C^N(x^k)x^k,$$

wobei $\nabla_x C^N(x^k)$ die Jacobi-Matrix von $C^N(x)$ im Punkt x^k und s die von b^l und b^u gebunden sind.

Es sei T eine $m \times n$ -Matrix und b ein n -dimensionaler Vektor, gegeben durch:

$$T = \begin{pmatrix} \nabla_x C^N(x^k) \\ C^L \end{pmatrix}, \quad b = \begin{pmatrix} -C^N(x^k) + \nabla_x C^N(x^k)x^k \\ 0 \end{pmatrix}. \quad (6.10)$$

Das QP-Teilproblem kann dann wie folgt geschrieben werden:

$$\min_{x,s} q(x), \quad x \in \mathbb{R}^n, \quad s \in \mathbb{R}^m, \quad \text{bzgl. } Tx - s = b, \quad b^l \leq \begin{pmatrix} x \\ s \end{pmatrix} \leq b^u.$$

Die Lösung des QP-Teilproblems ist selbst eine iterative Prozedur, nämlich die untere (Minor-) Iteration. In jeder unteren Iteration wird die lineare Gleichung $Tx - s = b$ gemäß:

$$T_B x_B + T_S x_S + T_N x_N = b$$

zerlegt, wobei die Grundmatrix („basic matrix“) T_B quadratisch und nicht singulär ist. Die Teilvektoren x_B , x_S bzw. x_N bezeichnen die „basic“, „superbasic“ bzw. „nonbasic“ Optimierungsvariablen. Bei einer Lösung des QP-Teilproblems liegen die „basic“ und die „superbasic“ Optimierungsvariablen zwischen den Schranken b^l und b^u , während die „nonbasic“ Optimierungsvariablen an der oberen oder unteren Schranken hängen. Die Optimierungsvariable x_S wird in jeder Iteration als unabhängige Variable behandelt. Diese unabhängige Optimierungsvariable wird während des Optimierungsverfahrens in alle möglichen Richtungen bewegt und besonders in die Richtung, die die Zielfunktion minimiert und der Gleichung $Tx - s = b$ genügt. Die Anzahl der „superbasic“ Optimierungsvariablen n_S bestimmt den Freiheitsgrad. Im allgemeinen ist n_S ein Maß dafür, wie nichtlinear ein Problem ist.

6.2.2 Erfassung der Jacobi-Matrix

Die Matrix T wird spaltenweise in der SNOPT-Routine mit Hilfe von drei Vektoren a , h_a und k_a eingegeben:

- $a(i)$ beinhaltet den Wert des i -ten nichtverschwindenden Elements in der Matrix T ,
- $h_a(i)$ gibt den Index derjenigen Zeile in der Matrix T an, welche das i -te nichtverschwindende Element der Matrix T enthält,
- $k_a(j)$ gibt den Index derjenigen Spalte in der Matrix T an, welche das i -te nichtverschwindende Element der Matrix T enthält,
- $k_a^{-1}(j)$ gibt den Index k des nichtverschwindenden Elements $a(k)$ mit der Eigenschaft $k_a(k) = j$ und $h_a(k) \leq h_a(l)$ für alle l mit $k_a(l) = j$ an.

6.2.3 Behandlung von nichtzulässigen Restriktionen durch SNOPT

Die SNOPT-Routine behandelt ausdrücklich die nichtzulässigen Restriktionen. Zunächst werden die nichtzulässigen linearen Restriktionen ermittelt, in dem folgendes Problem gelöst wird:

$$\min_{\alpha, \beta} e^\top (\alpha + \beta) \quad \text{bzgl.} \quad (6.11)$$

$$b^l \leq \begin{pmatrix} x \\ C^L x - \alpha + \beta \end{pmatrix} \leq b^u, \quad \alpha \geq 0, \beta \geq 0,$$

mit $e = (1, \dots, 1) \in \mathbb{R}^{m_n}$. Dabei wird die Eins-Norm der Überschreitungen aller linearen Restriktionen bezüglich der Schranken minimiert. Falls es immer noch keine zulässigen linearen Restriktionen gibt, dann endet die SNOPT-Routine, ohne die nichtlinearen Restriktionen zu betrachten.

Falls alle linearen Restriktionen erfüllt sind, dann bleiben sie für die restlichen Iterationen immer erfüllt. Dabei wird bei der Betrachtung der linearen Restriktionen ein Bereich bestimmt, in dem die Lösung gesucht wird.

Dann folgt die Betrachtung der nichtlinearen Restriktionen. Dabei wird das originale Problem in ein quadratisches Teilproblem transformiert, wobei die nichtlinearen Restriktionen linearisiert werden. Falls das Teilproblem nichtzulässig wird (d.h. die linearisierten Restriktionen die Randbedingungen nicht erfüllen oder der approximierte Lagrange-Multiplikator für die nichtlinearen Restriktionen größer ist (vgl. die Ungleichung in (5.13)), dann geht die SNOPT-Routine in einen Elastizitätsmodus über, und löst die folgende Aufgabe

$$\min_{x, \alpha, \beta} f(x) + \rho e^\top (\alpha + \beta), \quad \text{bzgl.} \quad (6.12)$$

$$b^l \leq \begin{pmatrix} x \\ C^N(x) - \alpha + \beta \\ C^L x \end{pmatrix} \leq b^u, \quad \alpha \geq 0, \beta \geq 0$$

mit einem nichtnegativen Strafterm ρ . Für sehr große ρ ist die Aufgabe äquivalent der Minimierung der Eins-Norm der Überschreitung der nichtlinearen Restriktionen.

Kapitel 7

Testrechnung

7.1 Benutzung der RINDOPT-Routine

Die RINDOPT-Routine ist die Implementierung des RINDOPT-Algorithmus in der Programmiersprache FORTRAN77.

7.1.1 Beschreibung von Abkürzungen und Arbeitsspeicher

Beschreibung von Abkürzungen

Zu den wichtigsten Abkürzungen der RINDOPT-Routine gehören folgende Zeichensätze:

N	Anzahl der Zustandsgitterpunkte,
mc	Anzahl der inneren Steuerungsgitterpunkte,
nf	Anzahl der Komponenten in der Funktion f (rechte Seite des dynamischen Systems),
nx	Dimension der Zustandsvariablen,
nu	Dimension der Steuerungsvariablen,
np	Anzahl des Optimierungsparameters v ,
Nq	Stufe des Runge-Kutta-Verfahrens,
U0	Vektor aus linken und rechten Grenzwerten der approximierten Steuerung $u(t,w)$ (Optimierungsvariable in RINDOPT),
t0	Vektor aus den Steuerungsgitterpunkten $(\tau_0, \dots, \tau_{M+1})$ (mögliche Optimierungsvariable in RINDOPT),

tfeld	Feld der Zustandsgitterpunkte,
lfeld	Indizes aller Steuerungsgitterpunkte auf dem Gitter \mathcal{G}_h (siehe Kapitel 3),
wxf	Feld der diskreten Zustände (Lösung des diskreten, dynamischen Systems),
wlf	Feld der diskreten Vektoren $(\lambda_l, l = 1, \dots, N)$ (Lösung des diskreten Systems (6.4)- (6.5)),
ltmax	Maximale Länge des Feldes tfeld,
TOLdgl	Genauigkeit für die numerische Lösung des dynamischen Systems,
tolopt	Genauigkeit für die Optimierung,
tobj	Zeit in Sekunden für die Berechnung der Zielfunktion und deren Gradienten,
tcon	Zeit in Sekunden für die Berechnung der Restriktionen und deren Gradienten,
topt	Zeit in Sekunden für die Optimierung,
optv	Optimalwert,
mci	Optionsparameter (ermöglicht eine restriktive Betrachtung der inneren Zustandsbeschränkungen),
istart	Optionsparameter (ermöglicht die Anwendung der Resultate aus der alten Ausführung von RINDOPT bei der neuen Ausführung von RINDOPT).

Arbeitsspeicher

Für ein Problem mit $n = 2(mc + 1) + d_0$ Optimierungsvariablen und m Gleichung- und Ungleichungsrestriktionen benötigt der Anwender von RINDOPT mindestens $500 * 8$ Byte für die „character“-Variablen, $10(n + m) * 4$ Byte für die „integer“-Variablen und $20(n + m) * 4$ Byte für die „real“-Variablen. Dies entspricht einen Speicherplatzbedarf von insgesamt $(500 * 8 + 120 * (n + m))$ Byte RAM.

7.1.2 Numerische Auswertung der Konvergenzrate

Es seien $h_i, i = 1, 2, \dots$ eine Folge von Zahlen gegeben durch die Formel (3.5) (maximale Schrittweite auf dem Steuerungsgitter) mit $h_{i+1} < h_i, i = 1, 2, \dots$. Zur Vereinfachung der Bezeichnungen betrachten wir für die Verfeinerung des Steuerungsgitters die Methode der Halbierung (d.h. $h_1, h_2 = h_1/2, h_3 = h_2/2, \dots$).

Es seien ρ_i, c_i zwei positive Zahlen, u^* eine exakte Lösung eines Problems und u^{h_i} die zugehörige Näherungslösung mit der Schrittweite h_i . Beim Übergang vom Gitter mit der

Schrittweite h_i zum Gitter mit der Schrittweite h_{i+1} läßt sich die numerische Konvergenzordnung des Verfahrens aus einer Ungleichung der Form (7.1) gewinnen (vgl. [18] §10.4.2, [75]).

$$\|u^{h_i} - u^*\|_{L^\infty} \leq c_i (h_i)^{\rho_i}. \quad (7.1)$$

Dabei soll der theoretische Wert $\rho_i = 1$ (siehe Theorem 3) mit dem numerischen Wert für die verschiedene Schrittweite verglichen werden. Es ergibt sich bei einer Gitterverfeinerung die Approximation:

$$\frac{\|u^{h_i} - u^*\|_{L^\infty}}{\|u^{h_{i+1}} - u^*\|_{L^\infty}} \cong 2^{\rho_i}. \quad (7.2)$$

Es folgen dann die Abschätzungen der Ordnung ρ_i und der Konstante c_i

$$\rho_i = \log_2 \left(\frac{\|u^{h_i} - u^*\|_{L^\infty}}{\|u^{h_{i+1}} - u^*\|_{L^\infty}} \right), \quad c_i = \frac{\|u^{h_i} - u^*\|_{L^\infty}}{(h_i)^{\rho_i}}. \quad (7.3)$$

Die nächsten Abschnitte berichten über ein Erläuterungsbeispiel und einige Testrechnungen. Diese Rechnungen werden auf einem Compute Server HP 9000 (780/J280) mit 180 MHz Frequenz und 768 MByte RAM unter HP-UX 10.20 durchgeführt.

7.1.3 Erläuterungsbeispiel

Wir betrachten eine akademische Bahnoptimierungsaufgabe als erste Testrechnung. Mit Hilfe dieses Beispiels wollen wir die Anwendung des implementierten Algorithmus RINDOPT veranschaulichen. Zuerst betrachten wir eine restriktive Aufgabe mit Steuerungsbeschränkung, und danach fügen wir Zustandsrestriktionen hinzu.

Es sei eine Kurve K_y beschrieben durch die Gleichung $y(t) = 2t(1-t)$, $t \in [0, 1]$. Das Ziel der Aufgabe ist es, die bestmögliche Approximation von K_y durch eine Zustandsfunktion x zu finden, so daß die Funktion x folgendem dynamischem System genügt:

$$\dot{x}(t) = u(t), \quad x(0) = 0, \quad |u(t)| \leq 1, \quad \text{f.ü. } t \in [0, 1]. \quad (7.4)$$

Das heißt, es ist eine Funktion u zu finden, so daß die Funktion y durch x am besten approximiert wird. Die Aufgabe hat folgende mathematische Darstellung:

$$\min_u \int_0^1 (x(t) - y(t))^2 dt \quad \text{bzgl.} \quad \left(\begin{array}{l} \dot{x}(t) = u(t), \quad x(0) = 0 \\ |u(t)| \leq 1 \end{array} \right). \quad (7.5)$$

Die optimale Lösung u^* von (7.5) ist bekannt. Sie ist eine Zusammensetzung aus einer Bang-Bang-Komponente ¹ und einem singulären Steuerungsanteil ². Die Steuerung u^* und der zugehörige Zustand x^* haben die Darstellungen:

$$u^*(t) := \begin{cases} 1 & : t \in [0, 1/2), \\ 2 - 4t & : t \in [1/2, 5/8], \\ -1 & : t \in (5/8, 1], \end{cases} \quad x^*(t) := \begin{cases} t & : t \in [0, 1/2), \\ y(t) & : t \in [1/2, 5/8], \\ y(5/8) - (t - 5/8) & : t \in (5/8, 1]. \end{cases} \quad (7.6)$$

Der Optimalwert von (7.5) beträgt 4.60612e-03 und die graphische Darstellung von (7.6) ergibt die Abbildungen 7.1 - 7.2.

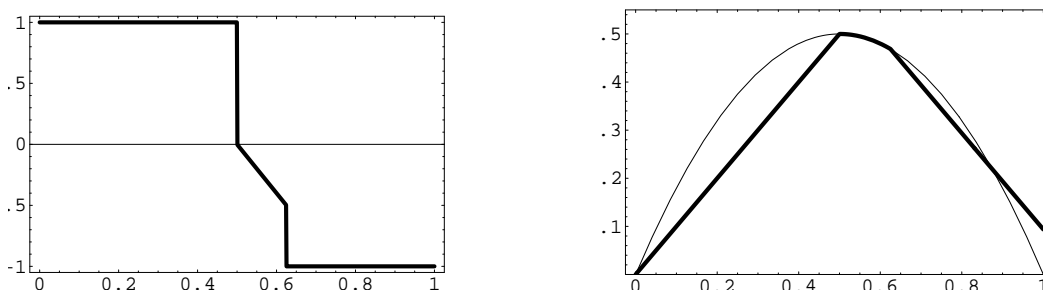


Abbildung 7.1: Optimale Steuerung, ohne Zustandsrestriktion **Abbildung 7.2:** Optimaler Zustand, ohne Zustandsrestriktion

Für die numerische Lösung von (7.5) wandeln wir die Aufgabe in eine Aufgabe vom Mayer-Typ um. Es entsteht durch die Einführung einer neuen Zustandsvariablen:

$$\begin{aligned} x_1(t) &= \int_0^t (x(s) - y(s))^2 ds, \quad x_1(0) = 0, \quad t \in [0, 1], \\ \dot{x}_2(t) &= u(t), \quad x_2(0) = 0, \end{aligned}$$

¹ Es seien u^* ein regulär lokales Minimum des Problems (7.5), x^* bzw. λ^* der entsprechende Zustand bzw. Lagrange-Multiplikator. Die Steuerung u^* erfülle das Bang-Bang-Prinzip, falls der Ausdruck

$$\nabla_u [(\lambda^*)^\top(t) f(x^*(t), u^*(t))] = 0$$

auf keinem Teilintervall mit positivem Maß verschwindet.

² Die Steuerung u^* heißt singulär, falls es für den zugehörigen Zustand x^* und den entsprechenden Lagrange-Multiplikator λ^* , folgende Gleichung gilt

$$\nabla_u [(\lambda^*)^\top(t) f(x^*(t), u^*(t))] = 0, \quad \text{für alle } t \in I_k,$$

wobei I_k ein Teilintervall aus $I = [0, 1]$ ist.

folgende Aufgabe

$$\min_u x_1(1) \quad \text{bzgl.} \quad \left(\begin{array}{l} \dot{x}_1(t) = (x_2(t) - y(t))^2, \quad x_1(0) = 0 \\ \dot{x}_2(t) = u(t), \quad x(0) = 0, \quad x_2(0) = 0 \\ |u(t)| \leq 1 \end{array} \right) \quad \text{f.ü. } t \in [0, 1] \quad (7.7)$$

Der in dieser Arbeit entwickelte Algorithmus RINDOPT erfordert vom Nutzer folgende Routinen:

`inputfile` Übergabe von Parametern,

`fcn.f` Berechnung der rechten Seite des dynamischen Systems,

`dudgl.f` Berechnung der Ableitung der rechten Seite des dynamischen Systems nach u ,

`dxdgl.f` Berechnung der Ableitung der rechten Seite des dynamischen Systems nach x ,

`objfun.f` Berechnung der Zielfunktion und deren Ableitungen,

`confun.f` Berechnung der Restriktionsfunktionen und deren Ableitungen.

Für die numerische Lösung der Aufgabe (7.7) ist der Quellcode für die oben genannten Routinen in FORTRAN 77 im Abschnitt 7.1.4 auf Seite 87 zu finden.

Die numerischen Ergebnisse für `TOLdgl=1.0e-10` und verschiedene Werte von `mc` sind in der Tabelle 7.1 dargestellt. Dabei ist der Nachweis der schnellen Konvergenz für die ersten Verfeinerungen deutlich in der Tabelle 7.2 abzulesen. Zusätzliche Verfeinerungen führen zu keiner Verbesserung. Mögliche Ursachen sind: die numerische Lösung liegt bereits sehr nah an der exakten Lösung oder die Formeln (7.3) für die Berechnung von Konstanten ρ_i und c_i sind nicht mehr angemessen. Die graphische Darstellung der numerischen Lösung ergibt für

Tabelle 7.1: Bahnoptimierung, ohne Zustandsrestriktion

mc	tobj	tconf	topt	tolopt	optv
1	.32	.00	.18	3.3E-09	5.143629E-03
3	.92	.00	.58	1.2E-09	4.664928E-03
7	4.35	.00	3.17	4.7E-09	4.606120E-03
15	13.61	.00	9.45	3.7E-08	4.606120E-03
31	43.69	.00	26.65	1.0E-08	4.606121E-03

`mc=20` die Abbildungen 7.3 - 7.4.

Tabelle 7.2: Bahnoptimierung, Konvergenzrate I

mc	i	$\ u^{h_i} - u^*\ _{L^\infty}$	ρ_i	c_i
1	1	.562300202	-/-	-/-
3	2	.391506407	.5223046	3.861282868
7	3	1.5164040E-03	8.012237	6.9+14
15	4	7.1978529E-02	-5.568840	8.0616E-16
31	5	-/-	-/-	-/-

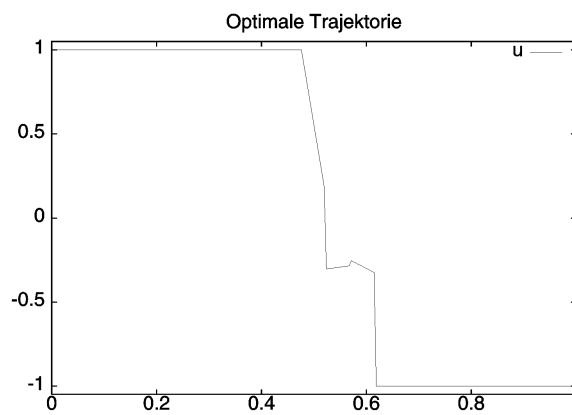
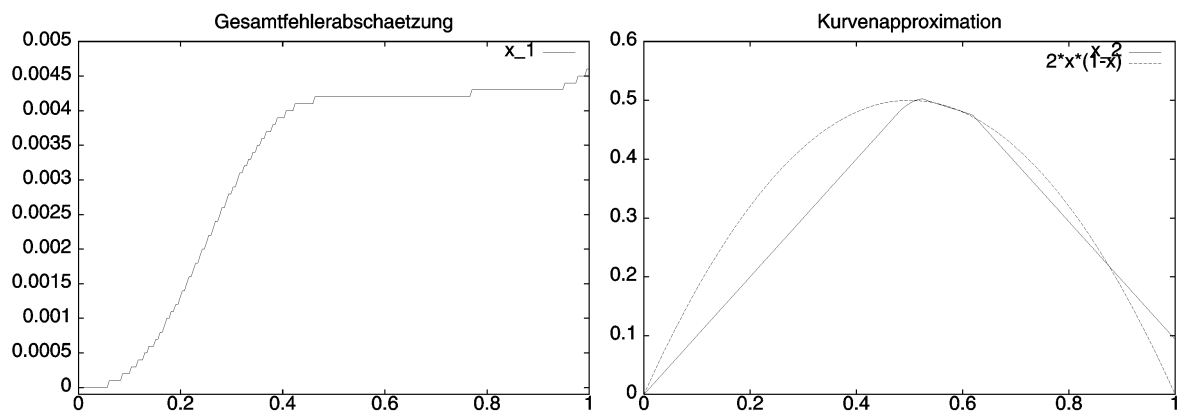
**Abbildung 7.3:** Diskrete optimale Steuerung, ohne Zustandsrestriktion**Abbildung 7.4:** Diskrete Zustandskomponenten, ohne Zustandsrestriktion

Tabelle 7.3: Bahnoptimierung, mit Zustandsrestriktion

mc	tobj	tconf	topt	tolopt	optv
3	6.63	17.28	23.11	1.2E-04	8.3510411155E-03

Tabelle 7.4: Bahnoptimierung, beliebiges Steuerungsgitter

mc	tobj	tconf	topt	tolopt	optv
3	2.40	4.83	6.89	7.9E-03	1.174224E-02
6	4.52	14.15	17.76	6.3E-03	9.3588035E-03
12	29.68	169.48	196.51	1.2E-03	1.002415E-02
13	10.38	56.69	63.80	1.3E-03	8.612884E-03

Wir betrachten erneut die Aufgabe (7.7) und fügen eine Zustandsrestriktion hinzu:

$$x_2(t) - 1 + t \leq 0, \text{ f.ü. } t \in [0, 1]. \quad (7.8)$$

Die optimale Lösung von (7.7)-(7.8) ist bekannt und hat die Gestalt:

$$u^*(t) := \begin{cases} 1 & : t \in [0, 1/2), \\ -1 & : t \in [1/2, 1]. \end{cases}, \quad x^*(t) := \begin{cases} t & : t \in [0, 1/2), \\ -t & : t \in [1/2, 1]. \end{cases} \quad (7.9)$$

Durch die Betrachtung der Zustandsrestriktion (7.8) verschwindet der singuläre Anteil von (7.6). Der Optimalwert von (7.7)-(7.8) beträgt 8.33333e-03. Der Quellcode für die Berechnung der Restriktion (7.8) in FORTRAN 77 ist in Abschnitt 7.1.4 auf Seite 87 zu finden.

Die numerische Lösung von (7.7)-(7.8) ergibt für die Optionen mc=3, istart=1, TOLdg1=1.0e-10 und das Steuerungsgitter

$$\mathcal{I}_h := \{0.0, 0.5, 0.6, 0.7, 1.0\}$$

das Resultat in der Tabelle 7.3.

Die graphische Darstellung der numerischen Lösung von (7.7)-(7.8) ergibt für mc=3 die Abbildungen 7.5 - 7.6.

Die Resultate von (7.7) - (7.8) werden für jede unglückliche Gitterverfeinerung nicht besser. Grund dafür ist der kritische Schalterpunkt $t = 0.5$. Zur Veranschaulichung geben wir die Tabelle 7.4 an.

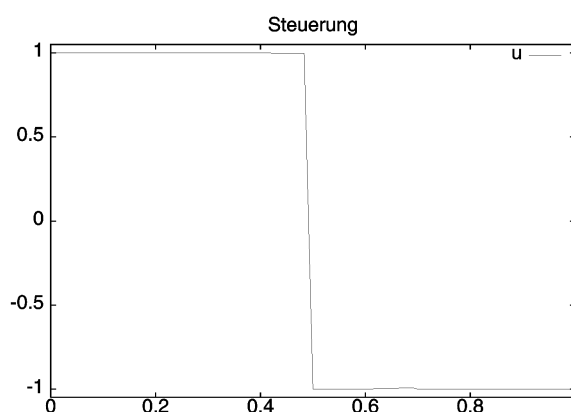


Abbildung 7.5: Diskrete optimale Steuerung, mit Zustandsrestriktion

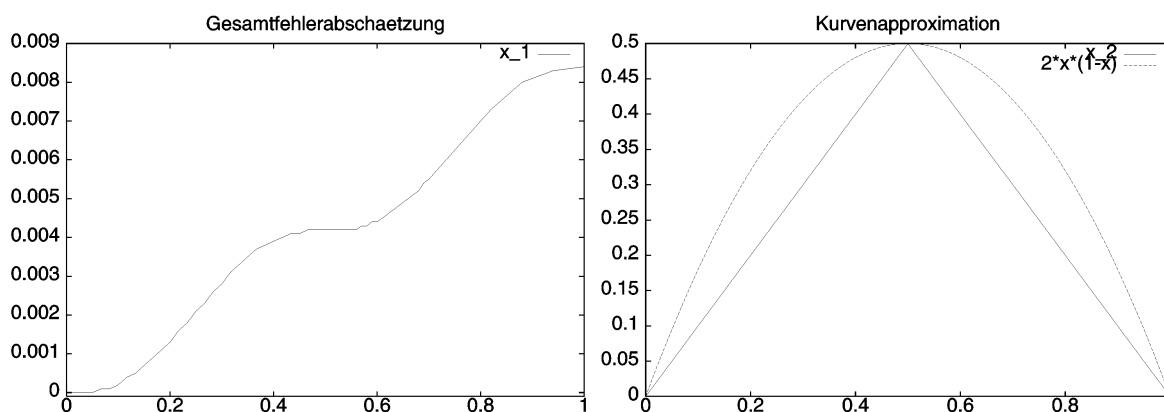


Abbildung 7.6: Diskrete Zustandskomponenten, mit Zustandsrestriktion

Die graphische Darstellung der numerischen Lösung von (7.7)-(7.8) bei willkürlichen Steuerungsgitter ergibt für $m_c=13$ die Abbildungen 7.7 - 7.8.

Es ist auch an der Tabelle 7.4 durch den Parametern t_{obj} , t_{conf} und t_{opt} zu erkennen, daß der Fall $m_c=12$ der RINDOPT-Routine größere Probleme als der Fall $m_c=13$ bereitet. Eine Erklärung dafür ist, daß im Fall $m_c=12$ der kritische Schaltpunkt $t = 0.5$ nicht zur Menge der Steuerungsgitterpunkte gehört.

Bemerkung 8 *Heuristisch gesehen kann der Parameter v in (2.1)-(2.3) einige Steuerungsgitterpunkte beinhalten (d.h. $v \in \{\tau_1, \tau_2, \dots, \tau_{M-1}\}$). Dadurch kann der Aufwand für manche Probleme erheblich reduziert werden. Für diesen Einsatz findet man in der Literatur keine passende Theorie.*

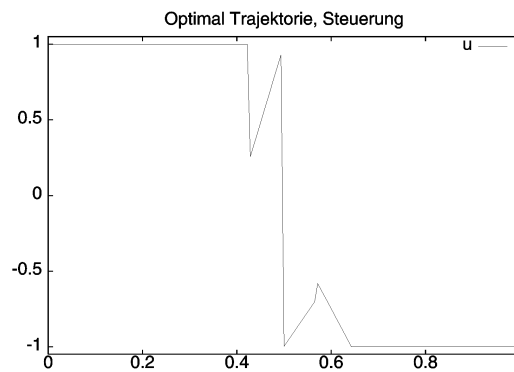


Abbildung 7.7: Diskrete optimale Steuerung, willkürliches Steuerungsgitter

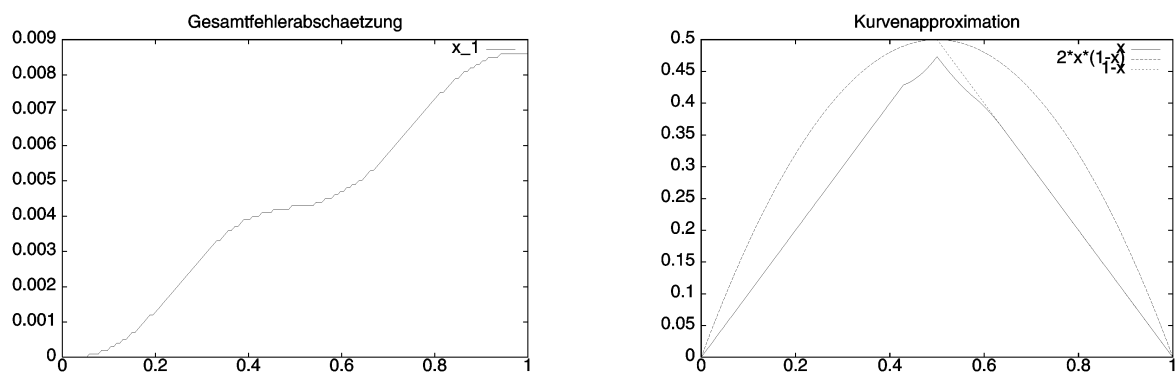


Abbildung 7.8: Diskrete Zustandskomponenten, beliebiges Steuerungsgitter

7.1.4 Nutzer-Quellcode für RINDOPT

inputfile

fcn.f

```

SUBROUTINE FCN(t1,x1,F)
c   Berechnung von  $f(x_1,u;v)$ .
c   Engabe: t1,x1
c       t1   diskrete Punkt
c       x1   Wert des Zustandes im Punkt t1
c   Ausgabe:  $F(i)$   $i=1,\dots,nx$ 
c   Letzte Aenderung: 25.1.98
c   Letzte Aenderung: 17.06.98
implicit          none
include           'inputfile3'
```

```

integer          j
double precision t1,x1(nx),F(nf)
double precision yt, ut

c  .computation of the function F
do 10 j=1,mc+1
if(t0(j).le.t1.and.t1.lt.t0(j+1)) then
    ut=(U0(2*j)*(t1-t0(j))+U0(2*j-1)*(t0(j+1)-t1))/(t0(j+1)-t0(j))
endif
10 continue
    if(t1.eq.t0(mc+2)) then
        ut = U0(2*(mc+1))
    endif
yt      = 2*t1*(1-t1)
F(1)    = (x1(2)-yt)**2
F(2)    = ut
return
end

BLOCK DATA inputfile3
c  unix version      03.07.97
c  Letzte Aenderung: 17.06.98
c  Aufruf von BLOCK DATA  inputfile
integer          mc, nf, nx, ltmax, nu, Nq, ndivise, istart
parameter        ( mc      = 3,
-                nf      = 2,
-                nx      = 2,
-                nu      = 1,
-                ltmax   = 1000,
-                Nq      = 5,
-                ndivise = 20,
-                istart  = 1 )
double precision TOLdgl
parameter        ( TOLdgl = 1.0e-10 )

```

```

c      .common parameter
      double precision      U0(2*(mc+1)), t0(mc+2),
+      tfeld(ltmax), wxf(nx,ltmax), wlf(nx,ltmax)
      integer               ltfeld(mc+2)
      common/inputfile/    U0, t0, tfeld, wxf, wlf, ltfeld
*
      data  t0(1),t0(mc+2)/0.0d0, 1.0d0/
      END

```

dudgl.f

```

      SUBROUTINE DUDGL(t1, x1, dfdu)
c      .Berechnung der Ableitung von f(x,u;v) nach u.
c      .Eingabe:  t1, nf
c      .Ausgabe:  dfdu(i) i=1,...,nf
      implicit          none
      include           'inputfile3'
      double precision  t1, x1(nx), dfdu(nf,nu)
*      .Executable statements
      dfdu(1,1) = 0.0d0
      dfdu(2,1) = 1.0d0
      return
      end

```

dxdgl.f

```

      SUBROUTINE DXDGL(t1, x1, dfdx)
c      .Berechnung der Ableitung von f(x,u;v) nach x.
c      .Eingabe:  nx
c      .Ausgabe:  dfdx(i,j) i=1,...,nf; j=1,...,nx
      implicit          none
      include           'inputfile'
      double precision  t1, x1(nx), dfdx(nf,nx), yt
*      .Executable statements
      yt               = 2*t1*(1-t1)
      dfdx(1,1) = 0.0d0

```

```

    dfdx(1,2) = 2*(x1(2)-yt)
    dfdx(2,1) = 0.0d0
    dfdx(2,2) = 0.0d0
  return
end

```

objfun.f

```

  subroutine objfun( mode, nnObj,
+                  x, fObj, gObj, nState,
+                  cu, lencu, iu, leniu, ru, lenru )
c  .see User's guide for snopt 5.3( A fortran package for large-scale
c  nonlinear programming) for more comment
  implicit          none
  include           'inputfile3'
  integer           mode,nnObj, nState, j
  double precision  x(nnObj), fObj, gObj(nnObj)
  integer           lenru, leniu, lencu
  double precision  ru(lenru)
  integer           iu(leniu)
  character*8       cu(lencu)
  external          FCN
*  .Executable statements
c  .update U0
    do 30 j = 1, 2*(mc+1)
      U0(j) = x(j)
30  continue
c  .solve the differential equation
    wxf(1,1) = 0.0d0
    wxf(2,1) = 0.0d0
    call pxfeld(FCN)
c  .solve the adjoint equation
    wlf(1,ltfeld(mc+2)) = 1.0d0
    wlf(2,ltfeld(mc+2)) = 0.0d0
    call plfeld(ltfeld(mc+2), FCN)
c  .compute the objective function

```

```

        fObj = wxf(1,ltfeld(mc+2))
c      .compute the gradients of the objective function
        call dobjfun( nnObj, x, gObj, nState )
*     end of objfun
        return
        end

```

confun.f

```

        subroutine confun( mode, nnCon, nnJac, neJac,
+           x, fCon, gCon, nState,
+           cu, lencu, iu, leniu, ru, lenru )
c      .see User's guide for snopt 5.3( A fortran package for large-scale
c      nonlinear programming) for more comment

        implicit      none
        include      'inputfile3'
        integer      mode,nnCon,nnJac,neJac,nState,j
        double precision  x(nnJac),fCon(nnCon),gCon(nnCon,nnJac)

        integer      lenru, leniu, lencu
        double precision  ru(lenru)
        integer      iu(leniu)
        character*8    cu(lencu)
        external      FCN
*     .Executable statements
c     .update U0
        do 10 j = 1, 2*(mc+1)
            U0(j) = x(j)
10    continue
c     .solve the differential equation
        wxf(1,1) = 0.0d0
        wxf(2,1) = 0.0d0
        call pxfeld(FCN)
c     .computes the constraints functions
        do 20 j = 1, mc+1

```

```

    fCon( j) = wxf(2,ltfeld(j+1)) + tfeld(ltfeld(j+1))
20  continue
c    .computes the gradient of the constraints functions
    call dconfun( nnCon, nnJac, gCon, nState )
*   end of confun
    return
    end

```

7.2 Anwendungsbeispiele

7.2.1 Container-Kran-Problem

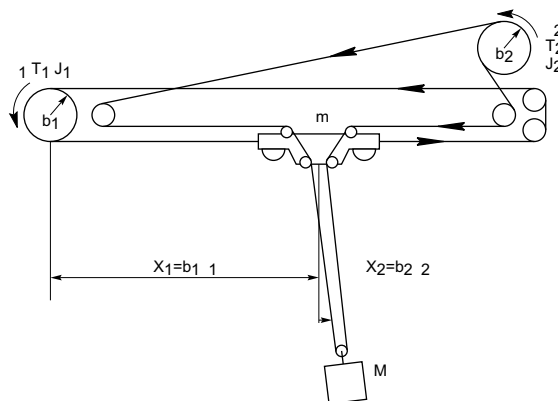


Abbildung 7.9: Container-Kran-Problem

Das Container-Kran-Problem (CKP) untersucht die Beladung und die Entladung von Schiffen an dem Hafen von Kobe in Japan. Dabei sollen während des Transfers (Beladung oder Entladung) von Schiffen die Schwingungen des Containers so schnell wie möglich minimiert werden. Wir betrachten den Container als eine Punktmasse und führen in der Tabelle 7.5 einige Bezeichnungen ein.

Wir betrachten θ_1, θ_2, ϕ als verallgemeinerte Koordinaten. Die Bewegungsgleichung des Container-Kran-Problems (vgl. [80]) ergibt folgendes Gleichungssystem:

$$\begin{aligned}
 [J_1 + (M + m)b_1^2]\ddot{\theta}_1 + Mb_1b_2\ddot{\theta}_2 \sin \phi + Mb_1b_2\dot{\theta}_2\dot{\phi} \cos \phi + \\
 Mb_1b_2(2\dot{\theta}_2\dot{\phi} \cos \phi - \dot{\theta}_2\dot{\phi}^2 \sin \phi) &= T_1 \\
 Mb_1b_2\ddot{\theta}_1 \sin \phi + (J_2 + Mb_2^2)\ddot{\theta}_2 - Mb_2^2\dot{\theta}_2\dot{\phi}^2 - Mgb_2 \cos \phi &= T_2 \\
 b_1\ddot{\theta}_1 \cos \phi + b_2\ddot{\theta}_2\dot{\phi} + 2b_2\dot{\theta}_2\dot{\phi} + g \sin \phi &= 0.
 \end{aligned} \tag{7.10}$$

Tabelle 7.5: Container-Kran-Problem, Bezeichnungen

θ_1	Rotationswinkel des Motors der Transportkarre
J_1	= 3.75 kg m ² , Trägheitsmoment des Motors der Transportkarre
b_1	= 1.44 × 10 ² m, Radius der Trommel des Motors der Transportkarre
θ_2	Rotationswinkel des Motors des Fahrstuhles
J_2	= 78.75 kg m ² , Trägheitsmoment des Motors des Fahrstuhles
b_2	= 1.22 × 10 ² m, Radius der Trommel des Motors des Fahrstuhles
ϕ	Schwingungswinkel der Punktmasse
m	= 6 t, Gesamtmasse des Fahrstuhles und der Kabel
M	= 42.5 t, Masse des Containers
T_1	Drehmoment, verursacht durch den Motor der Transportkarre
$T_{1,max}$	= 1.03 × 10 ³ Nm, maximales Drehmoment von T_1
T_2	Drehmoment, verursacht durch den Motor des Fahrstuhles
$T_{2,max}$	= 1.09 × 10 ⁴ Nm, maximales Drehmoment von T_2
$x_{4,max}$	= 2.5 ms ⁻¹ , maximale Geschwindigkeit der Transportkarre
$x_{5,max}$	= 1 ms ⁻¹ , maximale Geschwindigkeit des Fahrstuhles
g	= 9.81 ms ⁻² , Gravitationskonstante

Wir nehmen an: Der Schwingungswinkel der Punktmasse sei so klein, daß alle Terme mit $\phi^\alpha \phi^\beta$, ($\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta \geq 2$) zu vernachlässigen sind und die Approximationen $\cos \phi \cong 1$, $\sin \phi \cong \phi$ gelten. Es seien die Zustandsvariablen x_1, x_2, x_3 und die Steuerungsvariablen v_1, v_2 gegeben durch:

$$x_1 = b_1 \theta_1, \quad x_2 = b_2 \theta_2, \quad x_3 = \phi, \quad v_1 = \frac{b_1 T_1}{J_1 + m b_1^2}, \quad v_2 = \frac{b_2 (T_2 + M b_2 g)}{J_2 + M b_2^2}.$$

Es seien t_1 die Endzeit und w_1, w_2 zwei positive Konstanten. Nach der ausführlichen Definition der Zustands- und Steuerungsvariablen wollen wir nun das Container-Kran-Problem mathematisch formulieren:

$$\min_u \frac{1}{2} \int_0^{t_1} \{w_1 x_3^2(t) + w_2 x_6^2(t)\} dt \quad (7.11)$$

bezüglich des dynamischen Systems (erhalten aus (7.10))

$$\begin{aligned}
 \dot{x}_1 &= x_4 \\
 \dot{x}_2 &= x_5 \\
 \dot{x}_3 &= x_6 \\
 \dot{x}_4 &= v_1 - \delta_1 x_3 v_2 + \delta_1 g x_3 \\
 \dot{x}_5 &= -\delta_2 x_3 v_1 + v_2 \\
 \dot{x}_6 &= -\frac{1}{x_2} [v_1 - \delta_1 x_3 v_2 + (1 + \delta_1) g x_3 + 2x_5 x_6],
 \end{aligned} \tag{7.12}$$

wobei die Parameter δ_1 und δ_2 durch

$$\delta_1 = \frac{M b_1^2}{J_1 + m b_1}, \quad \delta_2 = \frac{M b_2^2}{J_2 + M b_2^2} \tag{7.13}$$

gegeben sind. Die Zustands- und Steuerungsvariablen erfüllen folgende Beschränkungen:

Initial- und Endbedingungen:

$$\begin{aligned}
 x(0) &= (0, l_2, 0, 0, -x_{5,max}, 0)^\top, \\
 x(t_1) &= (d_1, l_3, 0, x_{4,max}, 0, 0)^\top.
 \end{aligned} \tag{7.14}$$

Steuerungsbeschränkungen:

$$\begin{aligned}
 |v_1| &\leq v_{1,max}, & v_{1,max} &= b_1 T_{1,max} / (J_1 + m b_1^2), \\
 v_{2,min} \leq v_2 &\leq v_{2,max}, & \text{mit } v_{2,min} &= b_2 (T_{2,max} - M g b_2) / (J_2 + M b_2^2) < 0, \\
 & & v_{2,max} &= M g b_2^2 / (J_2 + M b_2^2) > 0.
 \end{aligned} \tag{7.15}$$

Zustandsbeschränkungen:

$$\begin{aligned}
 |x_4| &\leq x_{4,max}, \\
 |x_5| &\leq x_{5,max}.
 \end{aligned} \tag{7.16}$$

In [48] werden die Zustandsrestriktionen (7.16) als quadrierte Minimum-Funktion dargestellt. Die Endbedingungen in (7.14) werden quadriert. Danach werden die transformierte Restriktionen der Zielfunktion hinzuaddiert.

Goh und *Teo* [48] erzielen mit 20 Gitterpunkten den optimalen Wert 5.3996E-03. Die graphischen Darstellungen für die Zustandsvariablen stimmen mit der Abbildung 7.11 bis auf das Bild mit der Überschrift „CKP:Zustand x5“ übereint. Nachteil bei *Goh* und *Teo* ist der Glattheitsverlust, der durch die Minimum-Funktionen (vgl. [48] Seite 15, Gleichungen 52b, 52c) auftritt.

Sakawa und *Shindo* [94] stellen nur die graphische Lösung dar. Es sind gewisse Ähnlichkeiten mit einigen Bilder im Abbildung 7.11 festzustellen.

Tabelle 7.6: CKP, Numerische Resultate, Fall 1

mc	tobj	tconf	topt	istart	tolopt	optv
5	9.31	37.15	42.01	1	2.6E-06	5.360221E-03
10	21.48	71.31	80.81	1	2.6E-06	5.194515E-03
21	53.59	107.32	119.92	1	2.4E-06	5.161456E-03
43	167.27	149.52	162.60	1	4.6E-06	5.158099E-03
87	-/-	-/-	-/-	1	-/-	-/-

In dieser Arbeit betrachten wir ähnlich wie die Autoren in [48, 94] die Bewegung in der diagonalen Richtung. Es seien die Steuerungsvariablen auf folgender Weise umgewandelt

$$u_1 = v_1 - \delta_1 x_3 v_2, \quad u_2 = v_2 - \delta_2 x_3 v_1. \quad (7.17)$$

Das Gleichungssystem (7.12) bekommt somit eine neue Formulierung

$$\begin{aligned} \dot{x}_1 &= x_4 \\ \dot{x}_2 &= x_5 \\ \dot{x}_3 &= x_6 \\ \dot{x}_4 &= u_1 + \delta_1 g x_3 \\ \dot{x}_5 &= u_2 \\ \dot{x}_6 &= -\frac{1}{x_2} [u_1 + (1 + \delta_1) g x_3 + 2x_5 x_6]. \end{aligned} \quad (7.18)$$

Der Schwingungswinkel ist beschränkt durch:

$$|x_3| \leq \gamma, \quad (7.19)$$

wobei β eine geschätzte Zahl ist. Durch die Umwandlung in (7.17) entstehen neue Beschränkungen für die Steuerung:

$$\begin{aligned} |u_1| &\leq u_{1,max} & \text{mit} & \begin{aligned} u_{1,max} &= v_{1,max} - \delta_1 \gamma \max\{v_{2,max}, -v_{2,min}\} \\ u_{2,min} &= v_{2,min} + \delta_2 \gamma v_{1,max} \\ u_{2,max} &= v_{2,max} - \delta_2 \gamma v_{1,max}. \end{aligned} \\ u_{2,min} &\leq u_2 \leq u_{2,max} \end{aligned} \quad (7.20)$$

Die entwickelte RINDOPT-Routine wird in folgendem Fall angewendet: Problem : (7.11), (7.14), (7.18), (7.20),

Die numerischen Resultate für verschiedene Werte von mc sind in der Tabelle 7.6. Die graphische Darstellung ergibt für mc=20 die Abbildungen 7.10 - 7.11.

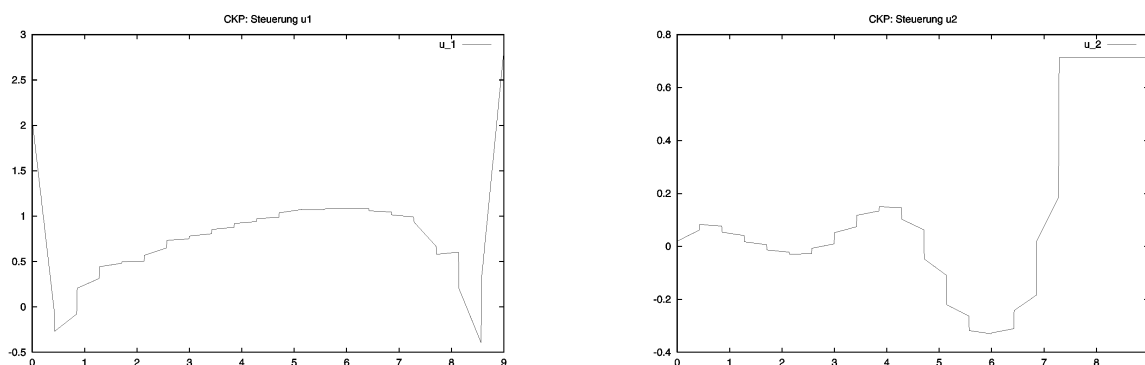


Abbildung 7.10: Container-Kran-Problem, Steuerungskomponenten

7.2.2 Flug eines Hängegleiters bei Aufwind

Wir betrachten einen Hängegleiter in der Vertikalebene (vgl. [17, 120, 121]). Hier soll der Pilot eines Hängegleiters aus einer gegebenen Startposition eine vorgeschriebene Endposition erreichen. Dabei soll der Flugkörper so gesteuert werden, daß die horizontale Reichweite maximiert wird. Die mathematische Formulierung des Problems in der Vertikalebene ergibt folgendes Problem:

Es seien x die horizontale Reichweite, y die Flughöhe, v_x die absolute horizontale Geschwindigkeitskomponente von x , v_y die absolute vertikale Geschwindigkeitskomponente von y , C_L der Auftriebsbeiwert. Wir betrachten (x, y, v_x, v_y) als Zustandsvariable und C_L als Steuerungsvariable. Das dynamische System der Aufgabe wird durch:

$$\left. \begin{aligned} \dot{x} &= v_x \\ \dot{y} &= v_y \\ \dot{v}_x &= (-L \sin(\eta) - D \cos(\eta))/m \\ \dot{v}_y &= (L \cos(\eta) - D \sin(\eta) - mg)/m \end{aligned} \right\} \quad (7.21)$$

beschrieben, und die Randbedingungen werden durch:

$$\left. \begin{aligned} x(0) &= 0[m], & x(t_f) &= \text{frei}, \\ y(0) &= 1000[m], & y(t_f) &= 900[m], \\ v_x(0) &= 13.23[m/s], & v_x(t_f) &= 13.23[m/s], \\ v_y(0) &= -1.288[m/s], & v_y(t_f) &= -1.288[m/s] \end{aligned} \right\} \quad (7.22)$$

gegeben. Die Daten und Funktionen des Hängegleiter-Problems sind in der Tabelle 7.7 dargestellt.

Der Aufwind ist von der Reichweite abhängig und ist durch die Funktion

$$u_A(x) = u_{A,max} \exp\left(-\left(\frac{x - x_{A,0}}{R}\right)^2\right) \left(1 - \left(\frac{x - x_{A,0}}{R}\right)^2\right) \quad (7.23)$$

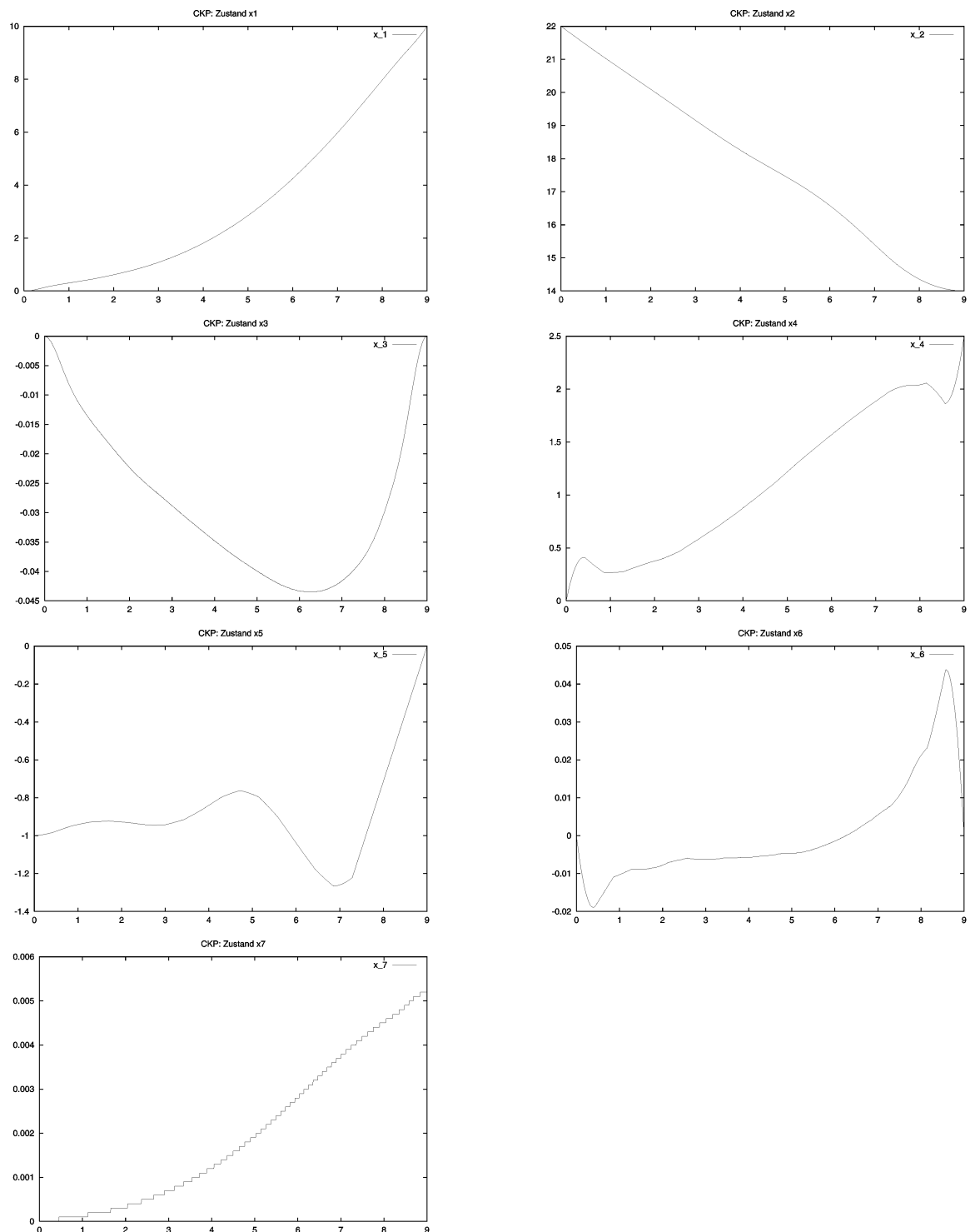


Abbildung 7.11: Container-Kran-Problem, Zustandskomponenten

Tabelle 7.7: Daten und Funktionen des Hängegleiter-Problems

η	Summe von Anstell- und Bahnneigungswinkel
g	= 9.80665 [ms ⁻²], Gravitationskonstante
m	= 100 [kg], Masse von Pilot und Hängegleiter
k	= 0.069662
C_{D0}	= 0.034
C_D	= $C_{D0} + kc_L^2$
ρ	= 1.13 [kg m ⁻³], Luftdichte in einer Höhe von 1000 m über dem Meeresspiegel
S	= 14 [m ²], Segelfläche
v_r	= $\sqrt{v_x^2 + (v_y - u_A(x))^2}$
$L(C_L, x, v_x, v_y)$	= $0.5C_L\rho S v_r^2$, Luftauftriebskraft
$D(C_L, x, v_x, v_y)$	= $0.5C_D(C_L)\rho S v_r^2$, Luftwiderstandskraft
R	= 100 [m]
$u_{A,max}$	= 2.5 [m/s], Maximale Aufwindgeschwindigkeit
$x_{A,0}$	= 350 [m] oder 250 [m]
$C_{L,max}$	= 1.4, Maximaler Auftriebsbeiwert

Tabelle 7.8: Hängegleiter, Numerische Ergebnisse, $x_{A,0} = 250$ [m]

mc	tobj	tconf	topt	istart	tolopt	optv
10	680.54	944.59	1611.06	1	-/-	-/-
20	239.75	666.26	893.76	1	2.0E-07	-1.246176E+03
50	422.21	1206.68	1597.59	2	5.6E-07	-1.247180E+03
80	748.52	2058.82	2757.29	1	9.2E-05	-1.247206E+03

gegeben. Der Winkel η ist die Summe von Anstell- und Bahnneigungswinkel:

$$\eta = \arctan\left(\frac{v_y - u_A(x)}{v_x}\right) \quad \text{mit} \quad \begin{aligned} \sin(\eta) &= (v_y - u_A(x))/v_r, \\ \cos(\eta) &= v_x/v_r. \end{aligned}$$

Für die numerischen Untersuchungen in dieser Arbeit setzen wir $t_f = 100$ und betrachten die Zielfunktion gegeben durch:

$$\min_{C_L} -x(t_f). \quad (7.24)$$

Die numerischen Ergebnisse für $x_{A,0} = 250$ [m] und verschiedene Werte von mc sind in der Tabelle 7.8 dargestellt. Die graphische Darstellung der numerischen Lösungen ergibt für die Optionen $x_{A,0} = 250$ [m] und mc=50 die Abbildungen 7.12 - 7.13. Für $x_{A,0} = 350$ [m] und

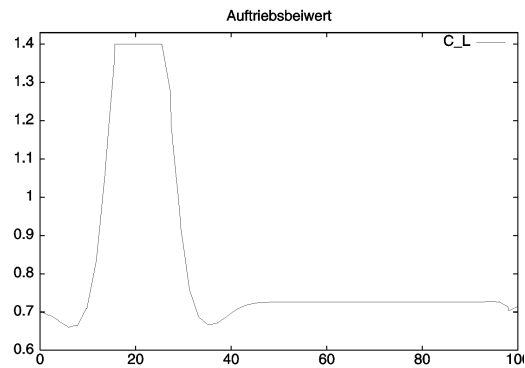


Abbildung 7.12: Hängegleiter, optimale Steuerung

verschiedene Werte von m_c liefert die RINDOPT-Routine die Ergebnisse in der Tabelle 7.9. Dieses Problem wurde in [17] durch die Kombination von direkten und indirekten Methode gelöst. Die Homotopie-Methode und das Mehrfachschießverfahren wurden angewendet. Zudem wurde der Zeithorizont (Zeitachse) in drei Phasen eingeteilt und folgende Ergebnisse erzielt: Optimale Zeit $t_f = 98.380$ [s], Maximale Reichweite $x(t_f) = 1247.60$ [m]. Die drei Phasen sind durch folgende Schaltpunkte der Steuerungsfunktion $t_{ein} = 23.301$ [s],

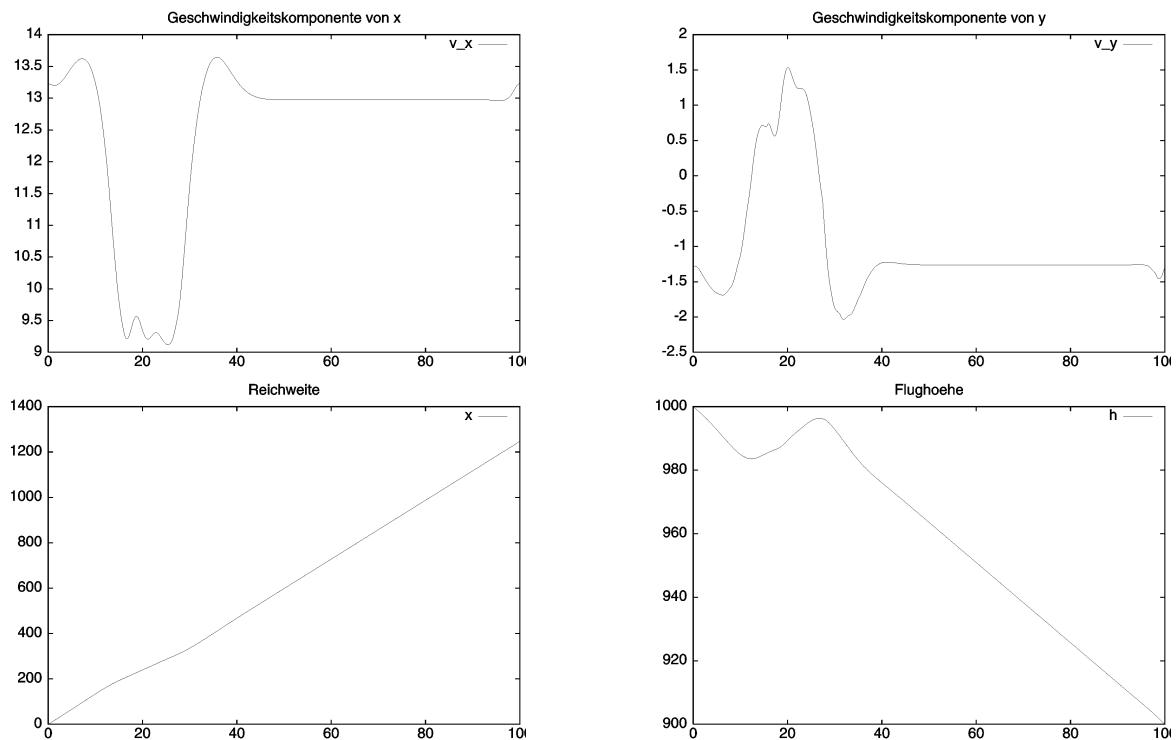


Abbildung 7.13: Hängegleiter, Zustandskomponenten

Tabelle 7.9: Hängegleiter, Numerische Ergebnisse, $x_{A,0} = 350$ [m]

mc	tobj	tconf	topt	istart	tolopt	optv
10	44.90	118.56	156.53	2	5.2E-07	-1.241665E+03
20	421.76	1153.61	1562.85	1	2.1E-07	-1.246494E+03
50	-/-	-/-	-/-	1	-/-	-/-
80	1012.48	2841.01	3802.84	1	9.9E-05	-1.246715E+03

Tabelle 7.10: Hängegleiter, Numerische Ergebnisse aus [120]

Phase	$x_{A,o}$ [m]	NG	optv	AERR	AEST	t_f	t_{ein}	t_{aus}
3	250	19/15/51	1248.030	0.00041	0.000437	98.437	15.777	25.70998
3	350	21/17/47	1247.602	0.00151	0.00176	98.380	23.319	33.24129

$t_{aus} = 33.250$ [s] gekennzeichnet.

In [120] wurde mit dem Algorithmus DIRCOL die Resultate aus Tabelle 7.10 erzielt. Dabei ist t_f die optimale Endzeit, t_{ein} der Aufsprungspunkt auf die Steuerungsbeschränkung, t_{aus} der Absprungspunkt auf die Steuerungsbeschränkung, AERR der absolute Fehler im Zielfunktion, AEST der geschätzte absolute Fehler im Zielfunktion, NG die Anzahl der Gitterpunkte bei DIRCOL. In beiden Fällen sind die durch RINDOPT erzielten Resultaten schlechter als bei DIRCOL, aber RINDOPT benötigt keinen guten Startwert und keine Homotopie-Transformation.

7.2.3 Tunneldiodenoszillator

Der Tunneldiodenoszillator stellt einen elektrischen Schaltkreis dar, deren Struktur in der Abbildung 7.14 gegeben ist. Die einfließenden Bezeichnungen der Abbildung 7.14 sind in

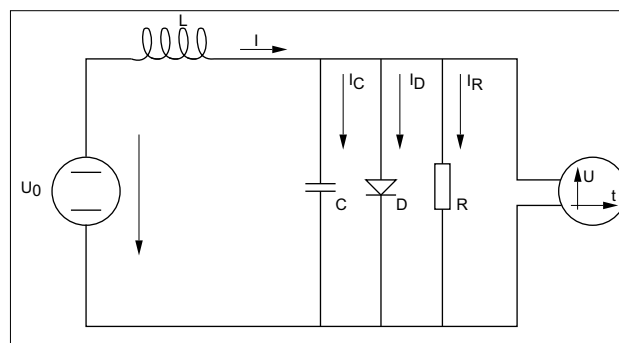
**Abbildung 7.14:** Tunneldiodenoszillator

Tabelle 7.11: Tunnelndiodenoszillator, Bezeichnungen

t	Zeit
L	Induktivität einer Spule
C	Kapazität eines Kondensators
R	Widerstand
I	Strom im System
U	Spannung
U_o	Spannung am Generator
D	Tunnelndiode

der Tabelle 7.11 erklärt. Es sei die Spannung I zum Zeitpunkt t als Zustandsvariable $x(t)$ und die Spannung am Generator $U_o(t)$ als Steuerung betrachtet. Das Ziel der Aufgabe ist es, die Gesamtspannung am Generator und den Gesamtstrom zu minimieren.

In [75, 18] wird das Tunnelndioden-Problem ausführlich behandelt. Dort wird eine optimale Lösung (x^*, u^*) folgendermassen konstruiert. Zuerst wird das Problem mittels der Euler-Methode mit sehr kleinen Schrittweite diskretisiert und gelöst. Daraus entsteht eine angenäherte Struktur der optimalen Steuerung. Danach wird das Mehrfachschießverfahren auf der Randwertaufgabe angewendet, welche aus der notwendigen Optimalitätsbedingungen der diskreten Aufgabe entstanden ist. Dabei wird die approximierte Steuerung als Startlösung benutzt. Die Ermittlung der genauen Lösung erfolgt durch die Anwendung der Routine BNDSCO in [84] und ergibt numerische Resultate mit zehn korrekten dezimalen Ziffern [75]. Die entstandene numerische Lösung erfüllt die Bedingungen A3-A5 (vgl. Bedingungen (II.1)-(II.6) von [75]).

Die mathematische Darstellung der Spannung am Generator ergibt folgende Bewegungsgleichung („The controlled Rayleigh equation“, vgl. [75, 63, 64, 115]):

$$\ddot{x}(t) = -x(t) + \dot{x}(t)(1.4 - 0.14\dot{x}(t)^2) + 4u(t) \quad (7.25)$$

Für die numerische Lösung des Problems führen wir folgende Zustandsvariablen ein:

$$\begin{aligned} x_1(t) &= x(t), \\ x_2(t) &= \dot{x}(t), \\ x_3(t) &= \int_0^t (u(\tau)^2 + x(\tau)^2) d\tau \end{aligned}$$

Tabelle 7.12: Tunneldiode, Numerische Resultate, Variante 1

mc	tobj	tconf	topt	tolopt	istart	optv
7	76.78	134.11	210.13	4.9E-06	1	4.473077E+01
15	14.53	21.00	32.05	3.9E-06	1	4.472521E+01
31	32.95	38.94	59.11	3.4E-06	1	4.472212E+01
63	76.73	44.45	66.36	3.1E-06	1	4.472101E+01
127	301.40	79.30	115.77	1.1E-06	1	4.472097E+01
256	1542.86	125.75	170.94	2.8E-07	3	4.472094E+01

und betrachten das Problem:

$$\begin{aligned}
 \min_{u \in \mathbb{R}} x_3(t_f) \quad \text{bzgl.} \quad & \begin{aligned}
 \dot{x}_1(t) &= x_1(t), \\
 \dot{x}_2(t) &= -x_1(t) + x_2(t)(1.4 - 0.14x_2(t)^2) + 4u(t), \\
 \dot{x}_3(t) &= u(t)^2 + x_2(t)^2, \\
 |u(t)| &\leq 1, \quad \text{f.ü. } t \in [0, t_f].
 \end{aligned}
 \end{aligned} \tag{7.26}$$

Die RINDOPT-Routine ergibt für die Varianten:

Variante 1: $t_f = 4.5$ und $x_1(t_t) = 0$,

Variante 2: $t_f = 2.5$,

die Ergebnissen in folgenden Tabellen und graphischen Darstellungen:

7.2.3.1 Variante 1

Die numerischen Resultate für verschiedene Werte von mc sind in der Tabelle 7.12 dargestellt. Die graphische Darstellung ergibt für mc=50 die Abbildungen 7.15 - 7.16.

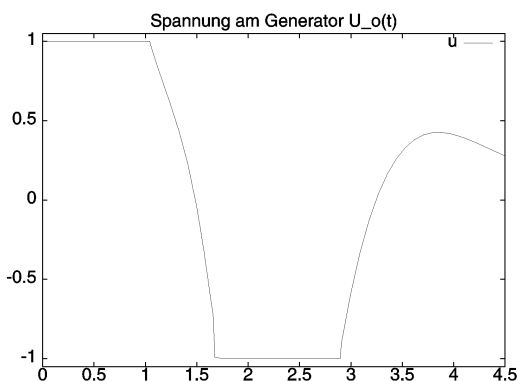


Abbildung 7.15: Tunneldiode, optimale Steuerung, Variante 1

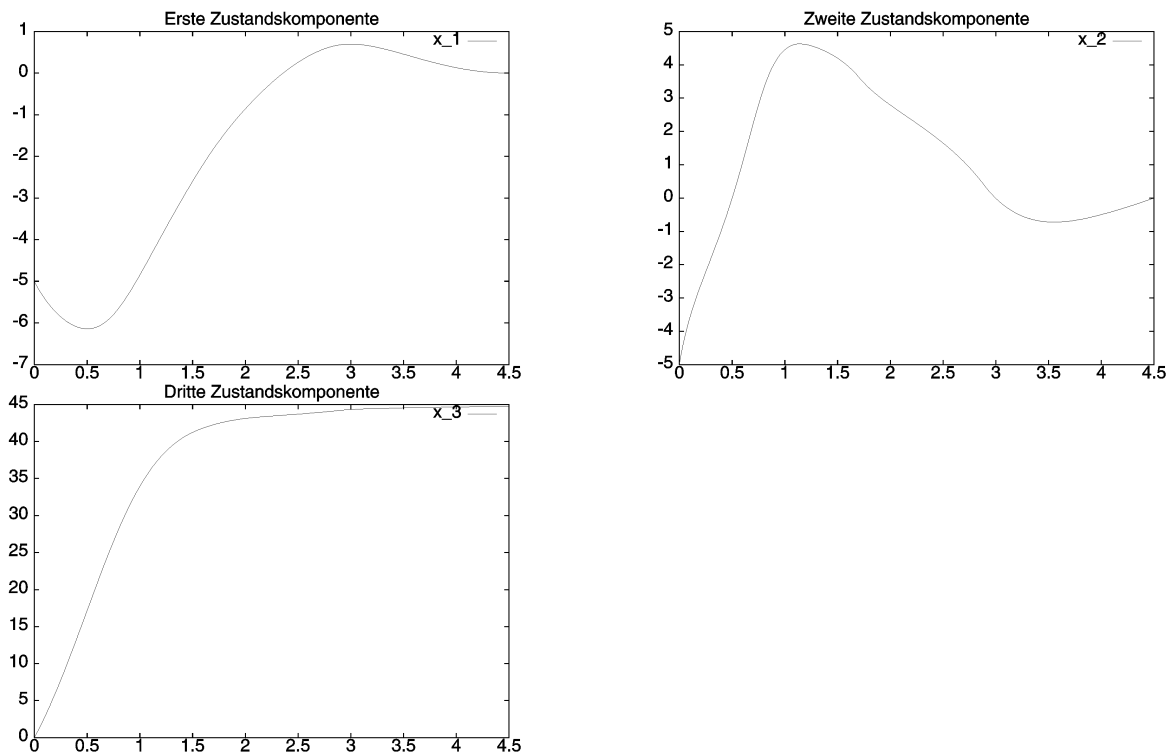


Abbildung 7.16: Tunneldiode, Zustandskomponenten, Variante 1

Tabelle 7.13: Tunneldiode, Numerische Resultate, Variante 2

mc	tobj	tconf	topt	tolopt	istart	optv
1	.32	.00	.22	6.4E-12	1	4.284172E+01
3	.91	.00	.64	3.6E-08	1	4.282389E+01
7	11.76	.00	10.95	3.7E-10	1	4.280879E+01
15	13.27	.00	10.31	2.7E-10	1	4.280784E+01
31	23.55	.00	11.40	2.7E-10	1	4.280744E+01
63	68.10	.00	14.78	1.1E-10	1	4.280744E+01
127	298.82	.00	38.02	9.3E-11	1	4.280744E+01

7.2.3.2 Variante 2

Die numerischen Resultate für verschiedene Werte von mc sind in der Tabelle 7.13 dargestellt. Die graphische Darstellung ergibt für mc=40 die Abbildungen 7.17 - 7.18.

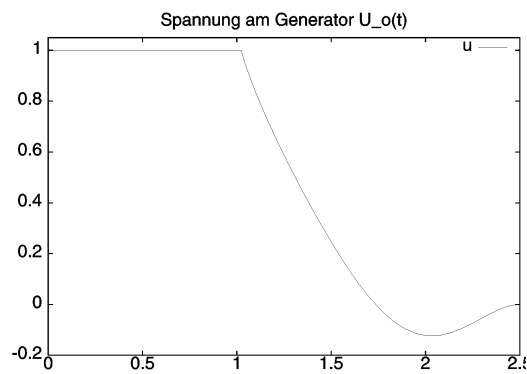


Abbildung 7.17: Tunneldiode, optimale Steuerung, Variante 2

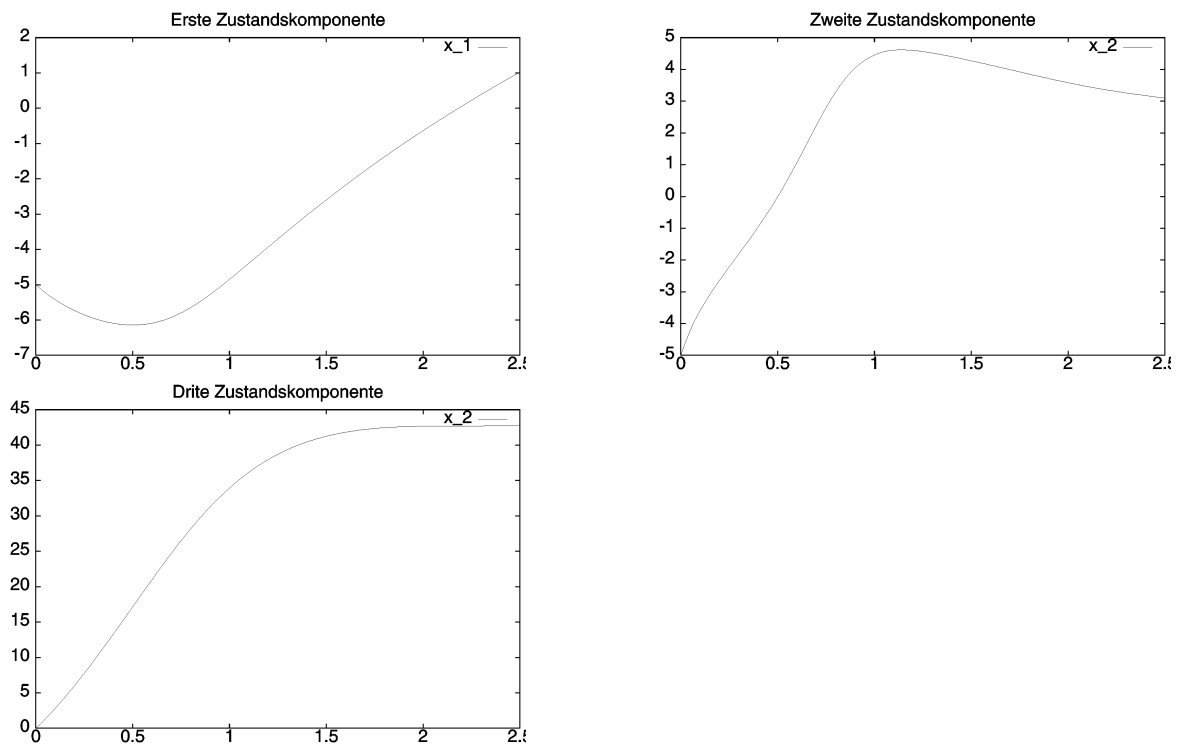


Abbildung 7.18: Tunneldiode, Zustandskomponenten, Variante 2

Tabelle 7.14: Tunnelodiode, Konvergenzrate II

mc	i	$\ u^{h_i} - u^*\ _{L^\infty}$	ρ_i	c_i
1	1	.26937	-	-
3	2	.31513	2.27662	173.70893
7	3	8.3606E-02	1.91428	63.60964
15	4	7.0345E-02	.24916	.19827
31	5	9.3878E-03	2.90559	12452.79352
63	6	7.0448E-03	.41422	7.00530E-02
127	7	6.9765E-08	16.62368	-/-

Kapitel 8

Abschlußbemerkungen

Das Anliegen dieser Arbeit war die Erforschung von Bedingungen für die bestmögliche Konvergenz bei der numerischen Lösung von optimalen Steuerungsproblemen mit gewöhnlichen Differentialgleichungen und die Entwicklung eines Algorithmus unter Berücksichtigung der vorgeschlagenen Diskretisierung.

Diese Arbeit bietet eine allgemeine und flexible Diskretisierung für eine Klasse von Steuerungsproblemen und einen Algorithmus RINDOPT für die numerische Lösung von Steuerungsproblemen an. Die Diskretisierung ist das Resultat der Kombination von zwei Zerlegungen (Steuerungsgitter und Zustandsgitter), der Parametrisierung der Steuerung und einem allgemeinen Einschrittverfahren höherer Ordnung.

Das allgemeine Einschrittverfahren läßt sich bequem mit zwei Zerlegungen kombinieren. Im Fall von bekannten Sprung- oder Knickstellen werden diese Stellen vom Verfahren eingehalten und das Integrationsverfahren wird ohne jeglichen Informationsverlust weiter fortgesetzt.

Die Grundlage der Konvergenzuntersuchungen ist eine allgemeine Approximationstheorie, basierend auf der Störungstheorie für verallgemeinerte Gleichungen, die sich am Konzept von *Dontchev* und *Hager* [33, 32] orientiert. Mit Hilfe der notwendigen Optimalitätsbedingungen werden parametrisierte verallgemeinerte Gleichungen formuliert. Im Abschnitt 4.2.2 ist ein Grundgedanke für die Konstruktion von approximierend verallgemeinerten Gleichungen dargestellt. Die Beweise für die Ungleichungen (4.25) und (4.26) sind im Abschnitt 4.3.1 detailliert ausgeführt. Diese Ungleichungen finden im Beweis des Konvergenzsatzes für die Aufhebung des Problems der Zweinorm-Diskrepanz Anwendung.

Der Gesamtfehler, entstanden aus der Parametrisierung und der Diskretisierung, wird im Theorem 3 formuliert. Dort wird folgende Abschätzung erzielt:

$$\|u^* - u(\cdot, w^h)\|_{\mathcal{L}^\infty} + \|v^* - v^h\| + \|x^* - x^h\|_{\mathcal{L}^\infty} + \|\lambda^* - \lambda^h\|_{\mathcal{L}^\infty} = O(h) + \epsilon(h), \quad (8.1)$$

wobei $\epsilon : [0, 1] \rightarrow \mathbb{R}_+$ mit $\lim_{h \rightarrow 0} \epsilon(h) = 0$. Die Gleichung (8.1) verdeutlicht die Möglichkeit einer besseren Abschätzung, die durch Anwendung zweier Zerlegungen angeboten wird. Durch die frei wählbaren Parameter M_k , $k = 1, \dots, M$ (siehe Lemma 5) und den Parameter v ist das Stabilitätsresultat (8.1) eine Verbesserung des Resultates aus [32, 38, 75]. Je nach dem Problem und der Stelle der Steuerungsgitterpunkte, können wir die Zahlen M_k , $k = 1, \dots, M$ groß oder klein wählen, um die Fehlerabschätzung so klein wie möglich zu halten. Es handelt sich hier um eine weiterentwickelte Form der Stabilitätsresultate aus [32].

Für Runge-Kutta-Verfahren ist die diskrete Koerzitivitätsbedingung eine Folge der stetigen Koerzitivitätsbedingung (siehe Abschnitt 4.4.2). Auch der Umgang mit den diskreten Matrizen in (4.11) bereitet keine Schwierigkeiten als dies der Fall in [32] ist.

Eine Verbesserung der Konvergenzrate ist stark von der Diskrepanz zwischen der primären und der dualen Diskretisierung belastet. Diese Diskrepanz führt zu zusätzlichen Bedingungen, die in Annahme A9 gestellt sind. Damit wird klar, eine Konvergenzrate besser als eins ist allein mit einem Integrationsverfahren der Ordnung q mit $q \geq 1$ schwer erreichbar.

Obwohl in [53] ein aussichtsreicher Weg für die Aufhebung der Belastung der Diskrepanz in der Diskretisierung angeboten wird, bleibt noch offen, dieses Problem für die tatsächliche diskrete optimale Steuerung zu untersuchen.

Eine weiterentwickelte Fassung für die Berechnung der Gradienten durch den Einsatz des impliziten Runge-Kutta-Verfahrens und der Methode der internen numerischen Differentiation wird im Abschnitt 3.7.2 vorgestellt. Einer der erheblichen Nachteile dieses Einsatzes ist der Zeitaufwand für die numerische Berechnung von diskreten Gradienten. Dies ist an verschiedenen Tabellen im Kapitel 7 deutlich zu erkennen. Der Vorteil dieses Verfahrens ist die genaue Berechnung der Gradienten.

Im Rahmen dieser Arbeit ist die RINDOPT-Routine für die numerische Lösung von Steuerungsproblemen implementiert worden. Die RINDOPT-Routine setzt sich zusammen aus dem Runge-Kutta-Verfahren, der internen numerischen Differentiation und dem neuen SQP-Routine SNOPT. Die aktuelle Version der RINDOPT-Routine verwendet das explizite Runge-Kutta-Verfahren. Durch den Einsatz von Verfahren höherer Ordnung können Differentialgleichungen mit hoher Genauigkeit gelöst werden. Das heißt, es werden in den Gleichungen (3.40) und (3.41) fast die exakten Werte des Zustandes in den entsprechenden Diskretisierungspunkten eingesetzt. Dies führt zu einer Steigerung der Genauigkeit von Gradienten diskretisierter Funktionen. Die genaue Berechnung der Gradienten, welche die Effektivität des SQP-Verfahrens erhöht und der Einsatz von zwei Zerlegungen, welche die Behandlung

von möglichen Sprung- oder Knickstellen verbessern, ermöglicht eine bessere Verfolgung von Trajektorien mit schlechten Verläufen.

Die RINDOPT-Routine benötigt keinen guten Startwert und konvergiert relativ schnell. Bei einer passenden Konstellation des Steuerungsgitters \mathcal{I}_h werden für eine gute numerische Lösung zirka nur zwei bis drei Verfeinerungen des Gitters \mathcal{I}_h erforderlich. Die Gründe dafür liegen in der genauen Berechnung der Gradienten und der richtigen Zerlegung. Zusätzliche Verfeinerungen des Steuerungsgitters führen zu keiner Verbesserung, da die resultierende numerische Lösung bereits nahe an der exakten Lösung liegt.

Die Schwierigkeit im Vorfeld des numerischen Lösungsprozesses besteht in der Bestimmung eines angepaßten Steuerungsgitters \mathcal{I}_h . Bei der numerischen Lösung von Steuerungsproblemen entsteht die Komplexität im allgemeinen an den auftretenden Knick- oder Sprungstellen der Steuerungsfunktion (vgl. Tabelle 7.4 und Abbildungen 7.7 - 7.8). In den Testrechnungen der vorliegenden Arbeit wurde dieses Problem durch Beobachtungen gelöst. Das Steuerungsgitter wird am Anfang beliebig gewählt und nach wiederholter Durchführung der RINDOPT-Routine problemabhängig gewählt. Eine systematischere Variante für die Ermittlung des richtigen Steuerungsgitters wäre die folgende:

1. Generierung eines beliebigen Gitters,
2. Betrachtung dieses Gitters als Optimierungsvariable.

Außerdem bietet die Methode der internen numerischen Differentiation eine Möglichkeit für die Gradientenberechnung von Funktionen an, die von diesen Optimierungsvariablen abhängig sind (siehe Gradientenberechnung ab Seite 32). Dadurch könnte der Aufwand für einige Probleme erheblich reduziert werden. Für diesen Einsatz findet man in der Literatur kaum eine passende Theorie.

Die aktuelle Version der RINDOPT-Routine ist für die Behandlung von allgemeinen Steuerungsproblemen mit steifen dynamischen Systemen noch nicht ausgereift. Die Erweiterung des Algorithmus durch Einbeziehung vom impliziten Runge-Kutta-Verfahren ist ein denkbarer Weg für numerische Lösung dieser Gruppe von Aufgaben. Dabei ist für die Berechnung von Vektoren λ_l , $l = 1, \dots, N$ die Bestimmung der Matrizen $Y_{l,i}^a$ und $Y_{l+1,i}^b$ für jedes $i = 1, \dots, q$ erforderlich. Beim impliziten Runge-Kutta-Verfahren führen diese Berechnungen zu der Lösung eines linearen Gleichungssystems mit $2n^2$ Unbekannten für jedes $i = 1, \dots, q$.

Literaturverzeichnis

- [1] N. U. Ahmed. *Elements of finite dimensional systems and control theory*. Longman Scientific and Technical, 1988.
- [2] W. Alt. Lipschitzian perturbations of infinite optimization problems. *In: AV Fiacco (ed) Mathematical Programming with data perturbations II, Lecture notes in pure and applied mathematics, Marcel Dekker, New York and Basel, 85, pages 7–21, 1983.*
- [3] W. Alt. On the approximation of the infinite optimization problems with an application to optimal control problems. *Applied Mathematics and Optimization, 12, pages 15–27, 1984.*
- [4] D. Augustin and H. Maurer. Stability and sensitivity analysis for different types of optimal control problems. In Maurer and Pickenhain [78]. DFG-Schwerpunktprogramm „Echzeitoptimierung großer Systeme“.
- [5] M. L. Baguer and W. Römisch. Computing gradients in parametrization-discretization schemes for constrained optimal control problems. *In: Florenzano et al.: Approximation and optimization in the Caribbean II; Verlag Peter Lang, Frankfurt am Main, pages 14–34, 1995.*
- [6] A. Barclay, P.E. Gill, and J. Ben Rosen. *SQP methods and thier application to numerical optimal control*. Report NA 97 - 3, Dept of Mathematics, University of California, San Diego, 1997.
- [7] J. T. Betts. Trajectory optimization unsing sparce sequential programming. *International Series of Numerical Mathematics, 111, Birkhäuser Verlag Basel, pages 115–128, 1993.*
- [8] J. T. Betts. Issues in the direct transcription of optimal control problems to sparce non-linear programms. *International Series of Numerical Mathematics, 115, Birkhäuser Verlag Basel, pages 3–17, 1994.*

- [9] J. T. Betts and P. D. Frank. A sparse nonlinear optimization algorithm. *Journal of Optimization Theory and Applications*, 82, pages 519–541, 1994.
- [10] J. T. Betts and W. P. Huffman. The application of sparse nonlinear programming to trajectory optimization. *J. of Guidance, Control, and Dynamics*, 14, pages 338–348, 1991.
- [11] H. G. Bock. Numerical treatment of the inverse problems in chemical reaction kinetics, in: Modelling of chemical reaction systems (K. H. Ebert, P. Deuffhard and W. Jäger, Eds.),. *Springer-Verlag Berlin*, pages 102–125, 1981.
- [12] H. G. Bock. Recent advances in parameteridentification techniques for o.d.e., in numerical treatment of the inverse problems in differential and integral equations (P. Deuffhard, E. Hairer, Eds.). *Birkhäuser, Boston*, pages 95–121, 1983.
- [13] H. G. Bock. *Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinear Differentialgleichungen*. Bonner Mathematische Schriften, 183, 1987.
- [14] V. G. Boltjanski. *Mathematische Methoden der optimalen Steuerung*. Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 2 edition, 1971.
- [15] B. M. Budak, E. M. Berkovich, and E. N. Solov'eva. Difference approximations in optimal control problems. *SIAM Journal on Control*, 7, pages 18–31, 1969.
- [16] R. Bulirsch. *Die Mehrzielmethode zur numerischen Lösung von nichtlinearen Randwertproblemen und Aufgaben der optimalen Steuerung*. Report of Carl-Cranz Gesellschaft, Oberpfaffenhofen, 1971.
- [17] R. Bulirsch, E. Nerz, H. J. Pesch, and O. von Stryk. Combining direct and indirect methods in optimal control: Range maximization of a Hang Glider. *International Series of Numerical Mathematics, Birkhäuser Verlag Basel*, 111, pages 273–288, 1993.
- [18] C. Büskens. *Optimierungsmethoden und Sensitivitätsanalyse für optimale Steuerprozesse mit Steuer- und Zustands-Beschränkung*. PhD thesis, Westfälische Wilhelms-Universität Münster, Fachbereich Mathematik und Informatik, 1998.
- [19] C. Büskens and H. Maurer. *Sensitivity analysis on real-time control of nonlinear optimal control systems*. in 12 th Conference on Variational Calculus, Optimal Control and Applications, R. Bulirsch, L. Bittner, W. Schmidt and K. Heier, eds, International Series of Numerical Mathematics, Basel, Birkhäuser, 1997.

- [20] J. C. Butcher. *The numerical analysis of ordinary differential equation: Runge-Kutta and general linear methods*. John Wiley & Sons Ltd, 1987.
- [21] L. Cesari. *Optimization: Theory and applications*. Springer-Verlag, New York, 1983.
- [22] K. Chudjei. *Optimale Steuerung des Aufstiegs eines zweistufigen Hyperschall-Raumtransporters*. PhD thesis, Mathematisches Institut der Technischen Universität München, 1994.
- [23] A. R. Conn, N. I. M. Gould, and P. L. Toint. *LANCELOT: a Fortran package for largescale nonlinear optimization (Release A), Lecture Notes in Computation Mathematics 17*,. Springer Verlag, Berlin, Heidelberg, New York, London, Paris and Tokio, (1992).
- [24] M. Crouzeix and A. L. Mignot. *Analyse numérique des équations différentielles*. Masson, Paris, 1984. ISBN: 2-225-77341-6.
- [25] L. E. Cuesta and M. Gelbrich. Convergence rates of discretizations for constrained nonlinear optimal control problems with c^1 data. In: *Florenzano et al.: Approximation and optimization in the Caribbean II; Verlag Peter Lang, Frankfurt am Main*, pages 164–199, 1995.
- [26] J. Cullum. Finite-dimensional approximations of state-constrained continuous optimal control problems. *SIAM Journal on Control*, 10, pages 649–670, 1972.
- [27] J. W. Daniel. On the approximate minimization of functionals. *Math. Comp.*, 23, pages 573–581, 1969.
- [28] C. de Boor. *Splinefunktionen*. Birkhäuser, Basel, 1990.
- [29] J. E. Dennis, Jr., and J. J. Moré. Quasi-newton methods, motivation and theory. *SIAM Review*, 19, pages 46–89, 1977.
- [30] A. L. Dontchev. Error estimates for a discrete approximation to constrained control problems. *SIAM Journal on Numerical Analysis*, 18, pages 500–514, 1981.
- [31] A. L. Dontchev. *Discrete approximations in optimal control*. in: *Nonsmooth analysis and geometric methods in deterministic optimal control*, (B. S. Mordukhovich, H. J. Sussmann, Eds), The IAM Volumes in Mathematics and its Applications, 78, Springer-Verlag, New-York, 1996.

- [32] A. L. Dontchev and W. W. Hager. Lipschitzian stability in nonlinear control and optimization. *SIAM Journal on Control and Optimization*, 31 (9), pages 569–603, 1993.
- [33] A. L. Dontchev and W. W. Hager. Lipschitzian stability for state constrained nonlinear optimal control. *SIAM Journal on Control and Optimization*, 36 (2), pages 696 – 718, 1998.
- [34] A. L. Dontchev, W. W. Hager, A. B. Poore, and B. Yang. Optimality, stability and convergence in nonlinear control. *Applied Mathematics and Optimization*, 31, pages 297–326, 1995.
- [35] A. L. Dontchev, W. W. Hager, and V. M. Veliov. Second-order Runge-Kutta approximation in control constrained optimal control, to appear in. *SIAM Journal on Numerical Analysis*, Revised October (1999). www.math.ufl.edu/hager.
- [36] A. Drug. CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems. *Mathematical Programming*, 31, pages 153–191, 1985.
- [37] Yu. G. Evtushenko. *Numerical Optimization Techniques*. Optimization software Inc., New York, (1985).
- [38] U. Felgenhauer. Diskretisierung von steuerungsproblemen unter stabilen optimalitätsbedingungen. Habilitation, Brandenburgische Technische Universität Cottbus, März 1998.
- [39] U. Felgenhauer. On the higher order methods for control problems with mixed inequality constraints. Technical report, Brandenburgische Technische Universität Cottbus, Reihe Mathematik M-01/1998, 1998.
- [40] U. Felgenhauer. On Ritz type discretizations for optimal control problems. In P. Kall, I. Lasiecka, and M. P. Polis, editors, Proc. 18 th IFIP - TC7 Conference on system modelling and Optimization. Pitman Research Notes, Mathem. Ser. 1998., (to appear).
- [41] R. Fletcher. *An l_1 penalty method for nonlinear constraints*, in *Numerical Optimization 1984* (P. T. Boggs, R. H. Byrd, and R. B. Schnabel, Eds., Philadelphia). SIAM, 1985.
- [42] R. Fletcher. *Practical methods of optimization*. Second edition, John Wiley, Chichester, 1987.

- [43] P. E. Gill, W. Murray, and M. A. Saunders. *User's guide for NPSOL (Version 4.0): a Fortran package for nonlinear programming*. Report SOL 86-2, Department of Operations Research, Stanford University, Stanford, CA, 1986.
- [44] P. E. Gill, W. Murray, and M. A. Saunders. *SNOPT : An SQP algorithm for large-scale constrained optimization*. Numerical Analysis Report 97-2, Department of Mathematics, University of California and Report SOL 97-0, Department of EESOR, Stanford University, 1997.
- [45] P. E. Gill, W. Murray, and M. A. Saunders. *User's guide for SNOPT 5.3: a Fortran package for large-scale nonlinear programming*. Department of Mathematics, University of California and Department of EESOR, Stanford University, Februar 1999.
- [46] P. E. Gill, W. Murray, and M. H. Wright. *Practical optimization*. Academic Press, London and New York, 1981.
- [47] I. V. Girsanov. *Lecture on mathematical theory of extreme problems*. Number 67 in Lecture notes in economic and mathematical systems. Springer, Berlin, 1972.
- [48] C. J. Goh and K. L. Teo. Control parametrisation: A unified approach to optimal control problems with general constraints. *Automatica*, 24, pages 3–18, 1988.
- [49] C. J. Goh and K. L. Teo. MISER, a FORTRAN program for solving optimal control problems. *Adv. Engineering Software* 10, pages 90–99, 1988.
- [50] R. D. Grigorieff. *Numerik für gewöhnliche Differentialgleichungen I*. B. G. Teubner, Stuttgart, 1972.
- [51] W. W. Hager. Rates of convergence for discrete approximations to unconstrained control problems. *SIAM Journal on Numerical Analysis*, 13, pages 449–472, 1976.
- [52] W. W. Hager. Runge-Kutta diskretisation of optimal control problem. Department of Mathematics, University of Florida, Gainesville, FL 32611, May (1999). www.math.ufl.edu/hager.
- [53] W. W. Hager. Runge-Kutta methods in optimal control and the transformed adjoint system. in *Numerische Mathematik*, Revised September (1999). www.math.ufl.edu/hager.

- [54] C. R. Hagraves and S. W. Paris. Direct trajectory optimization using nonlinear programming and collocation. *J. Guidance, Control, and Dynamics*, 10, pages 338–348, 1987.
- [55] E. Hairer, S.P. Norsett, and G. Wanner. *Solving Ordinary Differential Equations I*. Springer-Verlag Berlin Heidelberg, 1993.
- [56] S.-P. Han. A global convergence method for nonlinear programming. *Journal of Optimization Theory and Applications*, 22, pages 299–309, 1977.
- [57] R. F. Hartl, S. P. Sethi, and R. G. Vicson. A survey of the maximum principles for optimal control problems with state constraints. *SIAM Review*, 37, pages 181–218, 1995.
- [58] M. Heinkenschloss. Projected sequential quadratic programming methods. *SIAM J. Optimization*, 6(2):373–417, 1996.
- [59] P. Hiltmann. *Numerische Lösung von Mehrpunkt-Randwertproblemen und Aufgaben der optimalen Steuerung mit Steuerfunktionen über endlichdimensionalen Räumen*. DFG-Schwerpunktprogramm ”Anwendungsbezogene Optimierung und Steuerung”, Report No. 448, 1993.
- [60] G. Van Der Hoek. Asymptotic properties of reduction methods applying linearly equality constrained reduced problems. *Mathematical Programming Study*, 16, pages 162–189, 1982.
- [61] A. D. Ioffe. Necessary and sufficient conditions for a local minimum 3: second order conditions and augmented duality. *SIAM Journal and on Control and Optimization*, 17, pages 266–288, 1979.
- [62] A. D. Ioffe and V. M. Tichomirov. *Theorie der Extremalaufgaben*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1979.
- [63] D. H. Jacobson. New second-order and first-order algorithms for determining optimal control: A differential dynamic programming approach. *Journal of Optimization Theory and Applications*, 6, pages 411–440, 1968.
- [64] D. H. Jacobson and D. Q. Mayne. *Differential dynamic programming*. American Elsevier Publishing Company Inc., New York, 1970.

- [65] A. Kirsch. *Zur Störung von Optimierungsaufgaben unter besonderer Berücksichtigung von optimalen Steuerungsproblemen*. PhD thesis, Mathematisch-Naturwissenschaftlichen Fakultät der Georg-August-Universität zu Göttingen, (1978).
- [66] D. Kraft. Finite difference gradients versus error-quadrature gradients in the solution of parametrized optimal control problem. *Optimal Control Applications & Methods*, pages 191–199, 1981.
- [67] S. Kurcyusz. On the existence and nonexistence of lagrange multipliers in banach spaces. *Journal of Optimization Theory and Applications*, 20, pages 81–110, 1976.
- [68] NAG Ltd. *The NAG Fortran Library Manual, Mark 15*. NAG Ltd, Wilkinson House, Oxford, United Kingdom, 1991.
- [69] D. G. Luenberger. *Optimization by vector space methods*. John Wiley, New York, 1969.
- [70] K. Malanowski. Finite difference approximations to constrained optimal control problems. *Springer-Verlag, Berlin*, pages 243–254, 1981.
- [71] K. Malanowski. Two-norm approach in stability and sensitivity analysis of optimisation and optimal control problems. *Advances in Mathematical Sciences and Application*, 2 (2), pages 1397–443, 1993.
- [72] K. Malanowski. *Lagrange-Newton method for control-state and pure state constrained optimal control problems. Convergence of approximations to nonlinear optimal control problems*. System Research Institute * Polish Academy of Sciences, WP-2, 1995.
- [73] K. Malanowski. Sufficient optimality conditions for optimal control subject to state constraints. *SIAM Journal on Control and Optimization*, 35 (1), pages 205–227, 1997.
- [74] K. Malanowski. Convergence of the Euler approximation to control and state constrained optimal control. In Maurer and Pickenhain [78]. DFG-Schwerpunktprogramm „Echzeitoptimierung großer Systeme“.
- [75] K. Malanowski, C. Büskens, and H. Maurer. Convergence of approximations to nonlinear optimal control problems. *Mathematical Programming with Data Perturbations (A.V. Fiacco, ed.), Lecture Notes in Pure and Applied Mathematics, Marcel Dekker, Inc., New York*, 195:253–284, 1997.

- [76] H. Maurer. First and second order sufficient optimality conditions in mathematical programming and optimal control. *Mathematical programming study*, 14, pages 163–177, 1981.
- [77] H. Maurer. The two-norm approach for second order sufficient conditions in mathematical programming and optimal control. Universitaet Muenster, 1992.
- [78] H. Maurer and S. Pickenhain, editors. *Workshop „Stabilitat und Sensitivität von Optimierungs- und Steuerungsproblemen“*, Burg (Spreewald) Germany, April (1999). Deutsche Forschungsgemeinschaft. DFG-Schwerpunktprogramm „Echzeitoptimierung großer Systeme“.
- [79] H. Maurer and J. Zowe. First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Mathematical Programming*, 16, pages 98–110, 1979.
- [80] L. Meirovitch. *Analytical methods in vibrations*. Macmillan, New York, 1967.
- [81] B. Sh. Mordukhovich. On difference approximation of optimal control problems. *J. Appl. Math. Mech.*, 42, pages 452–461, 1978.
- [82] W. Murray and F. J. Prieto. *A second-derivative method for nonlinearly constrained optimization*. to appear, 1995.
- [83] B. A. Murthagh and M. A. Saunders. *MINOS 5.5 User's guide, Technical Report SOL 83-20R*. Departement of Operations Research, Stanford University, Stanford, CA 94305, Revised, 1998.
- [84] H. J. Oberle and W. Grimm. *BNDSCO: A program for numerical solution of optimal control problems*. Institute for flight systems dynamics. DLR Oberpfaffenhofen, Germany, Internal report No. 515-89/22, 1989.
- [85] J.-P. Penot. On regularity conditions in methematical programming. *Mathematical Programming*, 16, pages 98–110, 1982.
- [86] M. J. D. Powell. A fast algorithm for nonlinearly constrained optimization calculation. *Numerical Analysis*, ed. by G. A. Weston, *Lecture Notes in Mathematics*, 630, pages 144–157, 1978.
- [87] S. M. Robinson. A quadratically-convergent algorithm for general nonlinear programming problems. *Mathematical Programming*, 3, pages 145–156, 1972.

- [88] S. M. Robinson. Stability theory for systems of inequalities 1: linear systems. *SIAM Journal on Numerical Analysis*, 12, pages 754–769, 1975.
- [89] S. M. Robinson. First order conditions for general nonlinear optimization. *SIAM Journal of Applied Mathematics*, 30, pages 597–607, 1976.
- [90] S. M. Robinson. Stability theory for systems of inequalities, part II, differentiable nonlinear systems. *SIAM Journal on Numerical Analysis*, 13, pages 497–513, 1976.
- [91] R. T. Rockafellar and R. J. B. Wets. *Variational analysis*. Springer-Verlag Berlin Heidelberg, 1998.
- [92] W. Römisch. Remarks on the numerical treatment of optimal control problems. *Bericht des IX IKM Weimar, Heft 2*, pages 82–86, 1981.
- [93] E. W. Sachs. *Topics in modern computational methods for optimal control problems*. in: Modern methods of optimization, Proceedings, Bayreuth, FRG (W. Krabs, J. Zowe, Eds.), Lecture Notes in Economics and Mathematical Systems, 378, Springer-Verlag, Berlin, 1990.
- [94] Y. Sakawa and Y. Shindo. Optimal control of container cranes. *Automatica*, 18, pages 257–266, 1982.
- [95] R. W. H. Sargent and G. R. Sullivan. The development of an efficient optimal control package. in: *Optimization Techniques (J. Stoer Ed.)*, Lecture Notes in Control and Information Sciences Vol. 7, Springer-Verlag, Berlin, pages 158–168, 1978.
- [96] A. V. Sarychev. First and second order sufficient optimality condition for bang-bang controls. *SIAM Journal on Control and Optimization*, 35 (1), pages 315–340, 1997.
- [97] K. Schittkowski. The nonlinear programming method of wilson, han and powell with an augmented lagrangian type line search function. part 2: An efficient implementation with linear least squares subproblems. *Numer. Math.*, 38, pages 115–127, 1981.
- [98] K. Schittkowski. The nonlinear programming method of wilson, han and powell with an augmented lagrangian type line search function. part 1: Convergence analysis. *Numer. Math.*, 38, pages 83–114, 1981.
- [99] K. Schittkowski. On the convergence of a sequential quadratic programming method with augmented lagrangian line search function. *Math. Operationsforsch. u. Statist., Ser. Optimization Vol. 14 (2)*, pages 197–216, 1983.

- [100] K. Schittkowski. *Software for Mathematical Programming, NATO ASI Series Vol F15 Computational Mathematical Programming*. Springer-Verlag, 1985.
- [101] K. Schittkowski. *NLPQL : A fortran subroutine for solving constrained nonlinear programming problems*. Annals of Operations Research, 11, 1985/1986.
- [102] K. Schittkowski. *Lecture notes in Economics and Mathematical Systems, More test examples for nonlinear programming codes*. Springer-Verlag, 282, 1987.
- [103] K. Schittkowski. Solving constrained nonlinear least squares problems by a general purpose SQP-method. Trends in mathematical optimization. 4th French-German conference on optimization. Edited by K.H. Hoffmann u.a., Birkhäuser, pages 295–309, 1988.
- [104] V. H. Schulz. A direct PRSQP method for path planning of the satellite mounted robots. *ZAMM*, 76, pages 17–20, 1996.
- [105] V. H. Schulz, H.-G. Bock, and M. C. Steinbach. Exploiting invariants in the numerical solution of multipoint boundary value problems for dae. *submitted to SIAM J. Sci. Comput.*, (1996).
- [106] A. Schwartz and E. Polak. Consistent approximations for optimal control problems based on Runge-Kutta integration. *SIAM Journal on Control and Optimization*, 34, pages 1235–1269, 1996.
- [107] B. Sendov and V. A. Popov. *The Averaged moduli of smoothness*. John Wiley, New York, 1988.
- [108] M. Steinbach. *Fast SQP methods for large-scale optimal control problems*. PhD thesis, University of Heidelberg, June, 1995.
- [109] M. C. Steinbach. Structured interior point SQP methods in optimal control. *ZAMM*, 76, pages 17–20, 1996.
- [110] J. Stoer. *Numerische Mathematik I*. 7. Auflage, Springer-Verlag, Berlin, 1994.
- [111] J. Stoer and R. Bulirsch. *Introduction to numerical analysis*. Second Edition, Springer-Verlag, New York, 1993.
- [112] K. L. Teo, C. J. Goh, and K. H. Wong. *A unified computational approach to optimal control problems*. Longman, Harlow, 1991.

- [113] I. B. Tjoa and L. T. Biegler. A reduced SQP strategy for errors-in-variables estimation. *Comput. Chem. Eng.*, 16, pages 523–533, 1992.
- [114] F. Tröltzsch. *Optimality conditions for parabolic control problems and applications*. Teubner Verlag zur Mathematik, Band 62, Leipzig, 1984.
- [115] T. Tun and T. S. Dillon. Extensions of the differential dynamic programming method to include systems with state dependent control constraints and state variable inequality constraints. *Journal of Applied Science and Engineering*, A, 3, pages 171–192, 1978.
- [116] V.S. Vassiliadis, R.W.H. Sargent, and C.C. Pantelides. *Solution of a class of multistage dynamic optimization problems, Part 1 - Problems without path constraints*. submitted to Industrial and Engineering Chemistry Research, 1994.
- [117] V. Veliov. Second-order discret approximations to linear differential inclusions. *SIAM Journal on Numerical Analysis*, 29, pages 439–451, 1992.
- [118] V. Veliov. On the time discretization of control systems. *SIAM Journal on Control and Optimization*, 35 (5), pages 1470–1486, 1997.
- [119] O. von Stryk. Numerical solution of optimal control problems by direct collocation., in *Control Applications of Optimization*, R. Bulirsch, A. Miele, J. Stoer, and K. H. Well, eds., vol 111 of *International Ser. Numer. Math.*, Basel, Birkhäuser, pages 129–143, 1993.
- [120] O. von Stryk. *Numerische Lösung optimaler Steuerungsprobleme: Diskretisierung, Parameteroptimierung und Berechnung der adjungierten Variablen*. PhD thesis, Technische Universität München, (1994).
- [121] O. von Stryk and R. Bulirsch. Direct and indirect methods for trajectory optimization. *Annals of Operations Research*, 37, pages 357–373, 1992.
- [122] W. Walter. *Gewöhnliche Differentialgleichungen, 6. Auflage*. Springer-Verlag Berlin Inc., 1996.
- [123] H. Werner and H. Arndt. *Gewöhnliche Differentialgleichungen: Eine Einführung in Theorie und Praxis*. Springer-Verlag Berlin Heidelberg, 1986.
- [124] J. Zowe. A remark on a regularity for the mathematical problem in Banach spaces. *Journal of Optimization Theory and Applications*, 25, 1978.

- [125] J. Zowe and S. Kurcyusz. Regularity and stability for mathematical programming problem in Banach spaces. *Applied Mathematics and Optimization*, 5, pages 49–62, 1979.

Persönliche Daten:

Name: Appolinaire Nzali
geboren am: 20.12.1963, in Abong-Mbang (Kamerun)
Staatsangehörigkeit: Kamerun

Ausbildungsdaten:

Schulische Ausbildung

1967 - 1982 Primärschule und Grundschule in Kamerun , Abschluß: BEPC

1982 - 1986 Sekundärschule in Kamerun, Abschluß: Abitur

Studium

1986 - 1987 Deutsch-Sprachausbildung und Studienvorbereitung am Herder-Institut in Leipzig.

1987 - 1992 Mathematik-Studium an der Technischen Universität Chemnitz
Abschluß: Diplom Mathematiker

Praktikum während des Studiums

12/1990 - 6/1991 Betriebspraktikum,
Thema: Implementierung eines Algorithmus für die Lösung eines Staudammproblems und die graphische Darstellung der Lösung

Promotion

Seit März 1993 Promotion an der Humboldt-Universität zu Berlin

Praxiserfahrung

06/1999 - 08/2000 Wissenschaftliche Hilfsassistententätigkeit im Rahmen kraft des Deutschen Forschungsgemeinschaft-Projekt: Optimierung von Destillationskolonnen unter Wahrscheinlichkeitsrestriktionen

Seit 08/2000 Mitarbeiter bei Mannesmann Arcor AG & Co als Systemplaner CCB in der Abteilung Costomer Care und Billing in der Unternehmenszentrale in Escborn

Sprachkenntnisse Englisch (gut), Deutsch (sehr gut) Französisch (sehr gut)

Außeruniversitäres Engagement

Seit 6/1995 Mitbegründer und Vorsitzender von Exchange-Kamerun
Aufgabe: Sammlung von ausgesonderten Büchern für Universitäten in Kamerun

Selbständigkeitserklärung

Hiermit erkläre ich, daß ich die vorliegende Dissertationsschrift selbständig und ohne unerlaubte Hilfe angefertigt habe. Alle benutzten Referenzen sind im Literaturverzeichnis angegeben. Die wichtigsten erhaltenen Hilfstellungen von anderen wurden in Vorwort angedeutet.

Ich habe mich um einen Doktorgrad nur an der Humboldt-Universität zu Berlin beworben.

Die dem Verfahren zugrunde liegende Promotionsordnung ist mir bekannt.

Appolinaie Nzali