

# Beiträge zur Optimalen Steuerung partiell-differential algebraischer Gleichungen

Von der Universität Bayreuth  
zur Erlangung des Grades eines  
Doktors der Naturwissenschaften (Dr. rer. nat.)  
genehmigte Abhandlung

vorgelegt von

Armin Rund

geboren am 24. Juni 1980 in Bayreuth

1. Gutachter: Prof. Dr. Hans Josef Pesch  
Universität Bayreuth
2. Gutachter: Prof. Dr. Roland Herzog  
Technische Universität Chemnitz
3. Gutachter: Prof. Dr. Christian Meyer  
Technische Universität Dortmund

Tag der Einreichung: 21. Oktober 2011

Tag des Kolloquiums: 3. Februar 2012



# Inhaltsverzeichnis

Kurzfassung	vi
Abstract	viii
Vorwort	x
Einführung	1
<b>1 OC-ODE mit Zustandsbeschränkungen</b>	<b>3</b>
1.1 Notwendige Bedingungen aus dem Maximumprinzip	3
1.1.1 Problemstellung	4
1.1.2 Notwendige Bedingungen nach dem JLS-Ansatz	5
1.1.3 Ordnung einer Zustandsbeschränkung und der BDD-Ansatz	7
1.2 Notwendige Bedingungen aus dem Lagrange-Formalismus	10
1.2.1 Bedingung zur freien Endzeit	11
1.2.2 Analyse des Multiplikators $\mu(t)$ der Zustandsbeschränkung	12
1.2.3 Notwendige Bedingungen des aufgeteilten Problems analog zu BDD	13
1.3 Numerische Umsetzung über ein Mehrpunkt-RWP	16
1.3.1 Die Eliminationsmethode	16
1.3.2 Numerische Umsetzung der verschiedenen indirekten Verfahren	17
1.3.3 Direkte Verfahren	17
1.4 Zusammenfassung	18
<b>2 BDD-Ansätze bei elliptischen OC-PDE</b>	<b>19</b>
2.1 Problemstellung und notwendige Bedingungen	20
2.2 Neue notwendige Bedingungen nach dem BDD-Ansatz	23
2.2.1 Der Bryson-Denham-Dreyfus-Ansatz bei OC-PDE	24
2.2.2 Die Lagrange-Technik am aufgeteilten System	24
2.2.3 Vergleich der notwendigen Bedingungen	26
2.3 Das Mengen-Optimalsteuerungsproblem	28
2.4 Ordnung von Zustandsbeschränkungen	30
2.4.1 Eliminationsmethode	30
2.4.2 Ordnung von Zustandsbeschränkungen bei OC-PDE	30
2.4.3 Herleitung von Maß-Darstellungen	32
2.5 Zusammenfassung	33
<b>3 Das Rocketcar</b>	<b>35</b>
3.1 Einführung zu parabolischen OC-PDE	35
3.2 Das Modell und seine verschiedenen Formulierungen	36
3.2.1 Inhalt der einbezogenen Veröffentlichungen	40
3.3 Theoretische Behandlung der Problemstellung	41

3.4	Theoretische Behandlung als OC-PDAE Problem . . . . .	45
3.4.1	Notwendige Bedingungen . . . . .	45
3.4.2	Analyse von Randstücken . . . . .	49
3.4.3	Analyse von Berührungspunkten . . . . .	54
3.4.4	Höhere Regularität . . . . .	55
3.5	Direkte Verfahren . . . . .	59
3.5.1	Übersicht direkter Verfahren . . . . .	59
3.5.2	Direkte Verfahren für das Rocketcar-Problem . . . . .	59
3.6	Indirekte Verfahren . . . . .	60
3.6.1	Transformation auf ein Mehrpunkt-Anfangsrandwertproblem . . . . .	61
3.6.2	Lösung des Mehrpunkt-Anfangsrandwertproblems . . . . .	64
3.7	Numerische Experimente . . . . .	66
3.7.1	Verifikation der Ergebnisse direkter Verfahren . . . . .	67
3.7.2	Das Crank-Nicolson-Verfahren und Zustandsbeschränkungen . . . . .	70
3.8	Zusammenfassung . . . . .	71
<b>4</b>	<b>Optimale Steuerung von Brennstoffzellen</b> . . . . .	<b>73</b>
4.1	Hintergründe zu Brennstoffzellen . . . . .	73
4.1.1	Typen von Brennstoffzellen . . . . .	74
4.1.2	Die Schmelzkarbonat-Brennstoffzelle (MCFC) . . . . .	74
4.2	Das Modell und seine numerische Lösung . . . . .	75
4.2.1	Semidiskretisierung im Ort mit Differenzenverfahren . . . . .	78
4.2.2	Zeitdiskretisierung . . . . .	81
4.2.3	Anfangswerte . . . . .	82
4.2.4	Simulationsergebnisse . . . . .	82
4.3	Das Optimalsteuerungsproblem der MCFC . . . . .	82
4.3.1	Auswahl der Steuerungen . . . . .	83
4.3.2	Auswahl des Zielfunktionalis . . . . .	84
4.3.3	Nebenbedingungen . . . . .	84
4.3.4	Existenz von Lösungen . . . . .	85
4.4	Notwendige Bedingungen . . . . .	86
4.4.1	Herleitung der adjungierten Gleichungen der MCFC . . . . .	86
4.4.2	Herleitung der rechten Seiten . . . . .	92
4.4.3	Beitrag des Zielfunktionalis zu den adjungierten Gleichungen . . . . .	93
4.4.4	Gleichungen für die Steuerungen . . . . .	94
4.4.5	Elimination von Molenbrüchen . . . . .	98
4.4.6	Erweiterung auf 2D . . . . .	98
4.4.7	Zusammenfassung der notwendigen Bedingungen . . . . .	98
4.5	Direkte Verfahren der Optimalsteuerung . . . . .	100
4.6	Indirekte Verfahren der Optimalsteuerung . . . . .	101
4.6.1	Auswahl möglicher Verfahren . . . . .	101
4.6.2	Adjungierter Solver und Wahl des Zeitgitters . . . . .	105
4.6.3	Konstruktion von Testaufgaben . . . . .	107
4.7	Numerische Experimente . . . . .	108
4.7.1	Lösung von Testbeispielen . . . . .	109
4.7.2	Lösung des Anwendungsproblems . . . . .	115
4.7.3	Vergleich und Weiterentwicklung durch AD . . . . .	118
4.8	Zusammenfassung . . . . .	120

<b>5</b>	<b>Schluss</b>	<b>121</b>
<b>A</b>	<b>Zustandsgleichungen der MCFC</b>	<b>123</b>
A.1	Gaskanäle und Solid . . . . .	123
A.2	Stromdichten und Potentiale . . . . .	123
A.3	Reaktionsraten und Ausdrücke . . . . .	124
A.4	Brenn- und Mischkammer . . . . .	124
A.5	Konstanten . . . . .	126
A.6	Anfangs- und Randwerte . . . . .	127
<b>B</b>	<b>Adjungierte Gleichungen der MCFC</b>	<b>129</b>
B.1	Gaskanäle und Solid . . . . .	129
B.2	Stromdichten und Potentiale . . . . .	130
B.3	Partielle Ableitungen der Reaktionsraten . . . . .	131
B.4	Brenn- und Mischkammer . . . . .	133
B.5	Übergangs- und Randbedingungen . . . . .	133
	<b>Symbolverzeichnis</b>	<b>135</b>
	<b>Abkürzungsverzeichnis</b>	<b>137</b>
<b>C</b>	<b>Publikationen</b>	<b>147</b>
C.1	New Necessary Conditions for Distributed Optimal Control Problems of Linear Elliptic Equations with State Constraints . . . . .	149
C.2	On a Prototype Class of ODE-PDE State-constrained Optimal Control Problems . . . . .	175
C.3	On Some New Phenomena in State-constrained Optimal Control if ODE as well as PDE are Involved . . . . .	197
C.4	On a State-Constrained PDE Optimal Control Problem arising from ODE- PDE Optimal Control . . . . .	201
C.5	Optimal Control for a simplified 1D Fuel Cell Model . . . . .	205



# Kurzfassung

Diese Arbeit liefert Beiträge zur Optimalen Steuerung partiell-differential algebraischer Gleichungen. Insbesondere werden Zustandsbeschränkungen bei der Optimalen Steuerung gewöhnlicher und partieller Differentialgleichungen sowie gekoppelter Systeme untersucht. Die verschiedenen Konzepte dieser Gebiete werden verglichen, übertragen und eingeordnet.

Zentrale Ergebnisse sind die Übertragung der notwendigen Bedingungen nach Bryson, Denham und Dreyfus auf elliptische Optimalsteuerungsprobleme mit punktwisen Zustandsbeschränkungen, die Übertragung von Sprungbedingungen und Maßdarstellungen auf ein ODE-PDE beschränktes Optimalsteuerungsproblem mit Zustandsbeschränkungen bei niederdimensionalen aktiven Mengen, sowie die Entwicklung effizienter numerischer Methoden für komplexe Anwendungsprobleme.

Die Beiträge dieser Arbeit gliedern sich in vier Kapitel, deren Aspekte jeweils zusammengefasst werden:

Zunächst werden die Grundlagen aus der Optimalen Steuerung gewöhnlicher Differentialgleichungen mit Zustandsbeschränkungen wiederholt. Die beiden geläufigen notwendigen Bedingungen nach Jacobson, Lele und Speyer, sowie nach Bryson, Denham und Dreyfus (BDD-Ansatz) werden erläutert und in den Zusammenhang der Optimalen Steuerung partieller Differentialgleichungen gestellt. Dabei wird der Zusammenhang zwischen den Sprungbedingungen und dem Borel-Maß hergestellt.

In Kapitel 2 wird der BDD-Ansatz auf ein Optimalsteuerungsproblem einer elliptischen partiellen Differentialgleichung mit punktwisen Zustandsbeschränkungen und verteilten aktiven Mengen übertragen. Die Idee dieses BDD-Ansatzes ist es, die Zustandsbeschränkung auf der aktiven Menge äquivalent in eine Steuerungs-Zustandsbeschränkung oder ggf. eine reine Steuerungsbeschränkung zu transformieren. Dies erlaubt die Herleitung neuer notwendiger Bedingungen. Durch die Transformation der Zustandsbeschränkungen gewinnen die zugehörigen Lagrange-Multiplikatoren an Regularität. Man erhält aus den neuen notwendigen Bedingungen ein Randwertproblem auf verschiedenen Gebieten mit Übergangsbedingungen. Das Interface zwischen den verschiedenen Gebieten stellt eine Optimierungsvariable dar. Eine notwendige Bedingung am Interface wird mit Techniken der Shapeoptimierung hergeleitet.

Das Kapitel 3 behandelt Zustandsbeschränkungen bei gemischten ODE-PDE Problemen: Anhand eines zeitabhängigen Anwendungsproblems – des sogenannten Rocketcars – lässt sich eine vollständige Darstellung des Borel-Maßes auf niederdimensionalen aktiven Mengen angeben. In der Folge lassen sich Sprungbedingungen und weitgehende Regularitätsaussagen herleiten. Die explizite Massdarstellung ermöglicht weiterhin die Formulierung als Mehrpunkt-Anfangsrandwertproblem und den Einsatz angepasster Lösungsmethoden.

Kapitel 4 widmet sich schließlich einem komplexen Anwendungsproblem eines OC-PDAE: Ein Brennstoffzellenmodell stellt uns vor ein Optimalsteuerungsproblem eines Systems von partiell-differentiell algebraischen Gleichungen. Es werden notwendige Bedingungen hergeleitet und direkte sowie indirekte (adjungierten-basierte) Methoden der Optimalen Steuerung entwickelt und verglichen. Numerische Experimente bestätigen die Effizienz der vorgestellten Methoden. Insbesondere das indirekte Quasi-Newton-Verfahren erlaubt eine zeitadaptive optimale Steuerung der Brennstoffzellenanlage mit hoher Genauigkeit und unter geringer Rechenzeit.

# Abstract

The thesis is concerned with optimal control problems of coupled systems of partial differential algebraic equations. In order to investigate (pointwise) state constraints, a bridge is built from optimal control problems of ordinary differential equations (ODE) to optimal control problems of partial differential equations (PDE). Different concepts of both fields are discussed and applied to optimal control problems with coupled systems of equations.

Major contributions are the derivation of new necessary conditions for elliptic control problems with pointwise state constraints, the transfer of jump conditions to state-constrained ODE-PDE control problems via a structural analysis of the measures associated with state constraints with active sets of measure zero, and finally the development of efficient numerical methods for the solution of complicated optimal control problems from real-life applications.

The first chapter outlines the background of optimal control of ODE with state constraints. The two major sets of necessary conditions are discussed, the one of Jacobson, Lele and Speyer, as well as the one of Bryson, Denham and Dreyfus (BDD ansatz). In building the bridge from ODE to PDE control theory, connections between the jump conditions and the Borel measure are shown.

The second chapter transfers the BDD ansatz to an elliptic control problem with pointwise state constraints. The idea is to transform the state constraints on the active set equivalently into a mixed control-state constraint or even a pure control constraint. This yields new necessary conditions with more regular multipliers. The optimality system is treated as a boundary value problem on different domains with junction conditions. Therefore, the interface in between the different domains is an additional optimization variable. A necessary condition at the interface is derived by technics from shape optimization.

Chapter three is devoted to state constraints in mixed instationary ODE-PDE control problems. For the so-called rocketcar problem, an explicit formula for the Borel measure is given on active sets of measure zero. This allows the derivation of jump conditions and enhanced regularity theorems. Numerical results confirm the derived conditions.

Finally, an involved fuel cell model gives rise to an optimal control problem with partial differential algebraic equations and control constraints. The aim is to develop efficient methods for solving such problems. Therefore, necessary conditions are derived and direct as well as indirect (adjoint-based) methods of optimal control are designed and compared. Numerical studies confirm the efficiency of the methods. In particular, the indirect quasi-Newton method allows for a time adaptive optimal control of the fuel cell system with high accuracy and low computational effort.



# Vorwort

Der Autor hat als wissenschaftlicher Mitarbeiter am Lehrstuhl Ingenieurmathematik der Universität Bayreuth in verschiedenen Teams zusammen mit anderen Mitarbeitern mehrere Publikationen verfasst, welche die Dissertation aufgreift und fortführt.

Die Dissertation ist deshalb zum Teil als kumulative Arbeit angesetzt: Die Kapitel 2.1, 2.2, 2.3, 3.2, 3.5.2, 4.2, 4.3, 4.5 und Anhang A stellen eine ausführliche Zusammenfassung der Erkenntnisse aus den eingereichten Veröffentlichungen dar.

Die Dissertation wird monographisch fortgeführt durch alleinige Forschungen des Autors, die noch nicht veröffentlicht wurden. Dazu gehören die restlichen Abschnitte, insbesondere die Kapitel 3.3, 3.4, 3.6 und 3.7 des Rocketcars sowie die Kapitel 4.4, 4.6, 4.7 und Anhang B der Brennstoffzelle. Diese Gebiete werden entsprechend ausführlich dargestellt.

Der Leser möge dem Autor diese Inkonsequenz nachsehen.

Folgende Veröffentlichungen sind Teil dieser kumulativen Doktorarbeit. Abgestimmt mit den Co-Autoren können die prozentualen Anteile der Autoren jeweils festgelegt werden:

---

<b>Zu Kapitel 2:</b>				
[39]	M. Frey 35%	S. Bechmann 30%	<b>A. Rund</b> <b>30%</b>	H. J. Pesch 5%
<hr/>				
<b>Zu Kapitel 3:</b>				
[80]	H. J. Pesch 35%	<b>A. Rund</b> <b>35%</b>	W. von Wahl 25%	S. Wendl 5%
<hr/>				
[79]	H. J. Pesch 40%	<b>A. Rund</b> <b>40%</b>	W. von Wahl 10%	S. Wendl 10%
<hr/>				
[101]	S. Wendl 40%	H. J. Pesch 30%	<b>A. Rund</b> <b>30%</b>	
<hr/>				
<b>Zu Kapitel 4:</b>				
[90]	<b>A. Rund</b> <b>90%</b>	K. Chudej 10%		
<hr/>				

Es wurde nur ein Teil der eigenen Veröffentlichungen einbezogen. Eine vollständige Liste der Veröffentlichungen findet sich auf Seite 147.

## Danksagung

Zuerst möchte ich mich ganz herzlich bei Herrn Prof. Dr. Hans Josef Pesch bedanken für die hervorragenden Arbeitsbedingungen und die intensive Betreuung dieser Arbeit, für die gelassenen Freiräume, sowie für die fachliche und persönliche Unterstützung.

Herrn Prof. Dr. Kurt Chudej möchte ich dafür danken, dass er mir stets mit Rat und Tat zur Seite stand. Durch die enge Zusammenarbeit habe ich sehr viel profitiert, sei es in der Forschung oder in der Lehre.

Mein Dank gilt weiterhin meinen derzeitigen und ehemaligen Kollegen vom Lehrstuhl Ingenieurmathematik und meinen Co-Autoren für unzählige fruchtbare Gespräche. Insbesondere gilt mein Dank meinen Freunden und meiner Familie, welche mich während der Entstehung dieser Arbeit unterstützt und motiviert haben.

Bayreuth, den 21. Oktober 2011

Armin Rund

# Einführung

Vorgänge in Natur- und Ingenieurwissenschaften beschreibt man mit mathematischen Modellen, um daraus Vorhersagen treffen zu können. Aufbauend auf derartigen Simulationen will man Ziele verwirklichen, d.h. etwas optimal gestalten. Ein Satz von Größen soll optimal gewählt werden. Sind sowohl das Modell als auch die zu optimierenden Größen endlich-dimensional, erhält man ein lineares oder nichtlineares Optimierungsproblem (LP oder NLP).

Das Gebiet der Optimalen Steuerung setzt auf der nichtlinearen Optimierung auf. Anstelle endlich viele Parameter optimal einzustellen, fragt man bei der Optimalen Steuerung nach einer optimalen Funktion. Probleme dieser Art sind unendlich-dimensional, die Variablen leben in Funktionenräumen. Die Form der Nebenbedingungen legt die Bereiche innerhalb der Optimalen Steuerung fest: So unterscheiden wir die Gebiete der Optimalen Steuerung gewöhnlicher Differentialgleichungen (OC-ODE), sowie die Optimale Steuerung partieller Differentialgleichungen (OC-PDE).<sup>1</sup> Treten bei einem OC-PDE zusätzlich gewöhnliche oder algebraische Gleichungen als Nebenbedingungen auf, so sprechen wir von Optimaler Steuerung partiell-differential algebraischer Gleichungen (OC-PDAE).

Derartige Optimalsteuerungsprobleme mit gewöhnlichen und partiellen Differentialgleichungen treten häufig in Natur- und Ingenieurwissenschaften auf. Modelliert man gleichzeitig Effekte aus verschiedenen Bereichen wie z.B. Mechanik, Strömungsmechanik, Thermodynamik, Elektrodynamik oder chemischer Verfahrenstechnik, so steht man vor einem Optimalsteuerungsproblem mit einem gekoppelten System partiell-differential algebraischer Gleichungen. Zwei derartige Beispiele werden in dieser Arbeit behandelt: Ein Raketenauto mit Aufheizung und eine Brennstoffzellen-Anlage.

Aufgrund der gekoppelten Systemen von Nebenbedingungen mischen sich bei OC-PDAE häufig Effekte aus OC-ODE mit Effekten aus OC-PDE. Die vorliegende Arbeit setzt daher grundlegendes Wissen aus dem Bereich der OC-PDE voraus, wie es beispielsweise die Monografie von Tröltzsch [96] liefert. Grundwissen zur algorithmischen Behandlung im Funktionenraum und numerische Methoden findet man bei Hinze, Pinnau, Ulbrich und Ulbrich [59] sowie im Artikel von Herzog und Kunisch [52]. Nötiges Grundwissen aus OC-ODE wiederholen wir in Kap. 1.

Besonders anfordernd und aktuell sind Fragestellungen zu Zustandsbeschränkungen bei OC-PDAE, weshalb sich die Arbeit ausführlich mit diesem Themengebiet auseinandersetzt.

---

<sup>1</sup> Die englischen Abkürzungen sind gebräuchlich. ODE: Ordinary Differential Equations. PDE: Partial Differential Equations.

Das Gebiet der Optimalen Steuerung partieller Differentialgleichungen mit Zustandsbeschränkungen ist ein sehr anforderndes Spezialgebiet der Optimalen Steuerung. Eine Übersicht zum aktuellen Stand der Forschung anhand elliptischer PDE mit punktweisen Zustandsbeschränkungen liefern Hintermüller und Kunisch [55]. OC-PDE mit reinen, punktweisen Zustandsbeschränkungen sind deswegen schwierig zu behandeln, da die Lagrange-Multiplikatoren im allgemeinen nur reguläre Borel-Maße sind ([95], [22], [23]). Diese Maße erschweren nicht nur die Analyse, sondern insbesondere auch die numerischen Methoden, nicht zuletzt weil sie mit Ansatzfunktionen – wie den finiten Elementen – nicht genau beschrieben werden können.

Bevor Methoden für Zustandsbeschränkungen bei OC-PDE entwickelt wurden, gab es bereits beeindruckende Fortschritte auf dem Gebiet der Optimalen Steuerung gewöhnlicher Differentialgleichungen mit Zustandsbeschränkungen, welche auf der Zerlegung in die aktive und die inaktive Menge der Zustandsbeschränkung und deren struktureller Analyse beruhen. Der Artikel von Hartl, Sethi und Vickson [50] gibt einen umfassenden Überblick über die verschiedenen Systeme notwendiger Bedingungen. Die notwendigen Bedingungen werden anschließend als Mehrpunkt-Randwertproblem formuliert und mit der Mehrzielmethode oder dem Kollokationsverfahren gelöst. Eine Übersicht zur numerischen Lösung der notwendigen Bedingungen anhand diverser komplexer Anwendungsprobleme aus der Luft- und Raumfahrt enthält Pesch [76].

Das 1. Kapitel wiederholt die nötigen Grundideen aus OC-ODE mit Zustandsbeschränkungen. Zwei verschiedene Systeme notwendiger Bedingungen werden vorgestellt und eingeordnet. Beide werden ein zweites Mal über die formale Lagrange-Technik hergeleitet, um den Anschluss an OC-PDE zu schaffen. Eine genauere Analyse des Lagrange-Multiplikators der Zustandsbeschränkung liefert den Zusammenhang zwischen den Sprungbedingungen und dem Borel-Maß.

In Kapitel 2 übertragen wir den sogenannten BDD-Ansatz von OC-ODE auf OC-PDE anhand eines elliptischen Prototyp-Problems mit punktweisen Zustandsbeschränkungen und verteilten aktiven Mengen. Die Idee dieses BDD-Ansatzes ist es, die Zustandsbeschränkung auf der aktiven Menge äquivalent in eine Steuerungs-Zustandsbeschränkung oder ggf. eine reine Steuerungsbeschränkung zu transformieren.

Das Kapitel 3 behandelt Zustandsbeschränkungen bei gemischten ODE-PDE Problemen: Anhand eines zeitabhängigen Anwendungsproblems – des sogenannten Rocketcars – geben wir eine vollständige Darstellung des Borel-Maßes auf niederdimensionalen aktiven Mengen an. Dies erlaubt weitgehende Regularitätsaussagen, sowie eine Formulierung als Mehrpunkt-Anfangsrandwertproblem und den Einsatz angepasster Lösungsmethoden.

Kapitel 4 widmet sich schließlich einem komplexen Anwendungsproblem eines OC-PDAE: Ein Brennstoffzellenmodell stellt uns vor ein Optimalsteuerungsproblem eines Systems von partiell-differentiell algebraischen Gleichungen. Es werden notwendige Bedingungen hergeleitet und direkte sowie indirekte (adjungierten-basierte) Methoden der Optimalen Steuerung entwickelt und verglichen. Numerische Experimente bestätigen die Effizienz der vorgestellten Methoden. Insbesondere das indirekte Quasi-Newton-Verfahren erlaubt eine zeitadaptive optimale Steuerung der Brennstoffzellenanlage mit hoher Genauigkeit und unter geringer Rechenzeit.

# Kapitel 1

## OC-ODE mit Zustandsbeschränkungen

Optimalsteuerungsprobleme mit gewöhnlichen Differentialgleichungen wurden in der Vergangenheit sehr ausführlich untersucht. Dennoch sind Lehrbücher auf diesem Gebiet rar. Gute Anleitungen zum praktischen Aufstellen der notwendigen Bedingungen findet der Anwender im Buch von Bryson und Ho (Erstauflage 1969, [15]). Beweise suche man dagegen im Buch von Feichtinger und Hartl [36], welches mit Kapitel 6.2 auch eine gute Übersicht zu Zustandsbeschränkungen liefert.

Vergleicht man die Literatur zu notwendigen Bedingungen von OC-ODE und OC-PDE, so findet man aus erster Sicht heraus ein unterschiedliches Kalkül. Um den Zusammenhang zu schaffen, stellt dieses Kapitel daher in Abschnitt 1.1 zunächst die bekannten notwendigen Bedingungen von OC-ODE nach Pontryagins Maximumprinzip dar. Anschließend werden die notwendigen Bedingungen des OC-ODE im Abschnitt 1.2 mit etablierten Techniken aus OC-PDE hergeleitet. Ein Vergleich versorgt uns mit den nötigen Zusammenhängen. In den Kapiteln 2 und 3 werden wir diese Ergebnisse aufgreifen und Teile von ihnen auf OC-PDE übertragen.

### 1.1 Notwendige Bedingungen aus dem Maximumprinzip

Verschiedene notwendige Bedingungen zustandsbeschränkter OC-ODE nach dem Maximumprinzip findet der Leser z.B. im Übersichtsartikel von Hartl, Sethi und Vickson [50] und in [36, Kap. 6.2]. Prinzipiell unterscheiden sich die Ansätze in der Art, wie sie die Zustandsbeschränkung behandeln. Zwei herausragende Varianten zeichnen sich dabei ab: Die direkte Ankopplung der Zustandsbeschränkung nach Jacobson, Lele und Speyer [61] und die indirekte Ankopplung nach Bryson, Denham und Dreyfus [16]. Während die direkte Ankopplung naheliegend erscheint und die schärfsten notwendigen Bedingungen liefert, zeichnet sich die indirekte Ankopplung durch Praxisstärken gerade bei Anwendungsproblemen aus. So lösen z.B. Bulirsch, Montrone und Pesch aufbauend auf der indirekten Ankopplung in [19], [20] ein sehr anforderndes Problem des Landeabbruchmanövers eines Flugzeugs bei Fallwinden. Eine Anleitung zur praktischen Umsetzung anhand weiterer Anwendungsprobleme bietet Pesch [76]. Eine Einordnung der verschiedenen notwendigen Bedingungen und weitere Literaturstellen findet man in der Habilitationsschrift von Maurer [66] und dem zugehörigen Report [68].

Nach [15, S. 119] bietet sich neben den beiden Ansätzen noch eine dritte Möglichkeit, welche wir als Eliminationsmethode bezeichnen wollen: Die Reduktion der Variablen auf der aktiven Menge. Eine skalare Zustandsbeschränkung der Ordnung  $m$  erlaubt es dabei,  $m$  Zustände aus dem System zu entfernen (zusammen mit einer Steuerung). Wir sprechen die Eliminationsmethode in Abschnitt 1.3.1 an.

Wir werden die verschiedenen Ansätze anhand eines elementaren Beispiels diskutieren, welches wir jetzt formulieren werden.

### 1.1.1 Problemstellung

Gegeben sei folgendes Problem: Ein Rennauto soll in minimaler Zeit einen Kurs absolvieren. Der Einfachheit halber sei die Lenkung fest vorgegeben (das Auto fährt z.B. auf einer Schiene). Der Fahrer möchte die Beschleunigung  $u(t)$  optimal wählen, um das Rennen zu gewinnen. Allerdings führt die Beschleunigung zu einem entsprechenden Benzinverbrauch, welcher mit einem Kostenfaktor  $\lambda > 0$  berücksichtigt wird. Bei freier Bahn versucht der Fahrer das Beste herauszuholen. Allerdings darf er gewisse Objekte nicht überholen (z.B. ein Safety Car oder einen Teamkollegen aufgrund einer Stallorder). Dies wird mit einer Zustandsbeschränkung beschrieben.

Als einfachsten Fall betrachten wir den Kurs eines Dragsterrennens – also eine rein gerade Strecke. Die Fahrt wird beschrieben durch den Ort  $x_1(t)$  und die Geschwindigkeit  $x_2(t)$  des zur Punktmasse geschrumpften Autos. Wir erhalten das Problem des optimal zu steuernden Autos (vgl. auch Pesch [77, Übungsaufgabe 6]):

#### Definition 1.1.1 (Prototyp-Problem eines OC-ODE)

$$\min J(x_1, x_2, u, t_f) := t_f + \frac{\lambda}{2} \int_0^{t_f} u^2 dt \quad (1.1)$$

$$\text{s.t.} \quad \dot{x}_1 = x_2, \quad x_1 \leq x_{\max} \quad \text{auf } (0, t_f), \quad (1.2a)$$

$$\dot{x}_2 = u \quad \text{auf } (0, t_f), \quad (1.2b)$$

$$x_1(0) = x_0, \quad x_2(0) = \dot{x}_0, \quad x_1(t_f) = 0. \quad (1.2c)$$

Dabei seien  $\lambda, x_0, \dot{x}_0 \in \mathbb{R}$ ,  $\lambda > 0$ ,  $x_0 < x_{\max}(0)$  sowie  $x_{\max}(t) \in H^4(0, t_f)$  gegeben. Gesucht ist die Steuerung  $u \in L^2(0, t_f)$ .

Es handelt sich um ein Problem mit freier Endzeit  $t_f$ : Steuerung  $u(t)$  und Endzeit  $t_f$  sind optimal zu wählen. Die Größen  $\vec{x} := (x_1, x_2)$  sind durch die Nebenbedingungen festgelegt. Wir bezeichnen sie als Zustände, ihre Differentialgleichungen entsprechend als Zustandsgleichungen.

Hauptaugenmerk liegt auf der Zustandsbeschränkung  $x_1 \leq x_{\max}$ , da diese sowohl die Analyse als auch die numerische Umsetzung erschwert. Wir betrachten im Folgenden die verschiedenen notwendigen Bedingungen zustandsbeschränkter OC-ODE. Zur genaueren Aufschlüsselung der notwendigen Bedingungen benötigen wir den Begriff der aktiven Menge:

**Definition 1.1.2 (Aktive Menge der Zustandsbeschränkung)**

Die Menge

$$\mathcal{A} := \{t \in [0, t_f] : x_1(t) = x_{\max}(t)\}$$

bezeichnen wir als aktive Menge der Zustandsbeschränkung  $x_1 \leq x_{\max}$ .

Wir möchten die notwendigen Bedingungen aufschlüsseln in Bedingungen auf der aktiven und auf der inaktiven Menge – sowie Übergangsbedingungen dazwischen. Dafür müssen wir entartete Extremfälle ausschließen. Wir stellen daher eine Strukturannahme an die aktive Menge  $\mathcal{A}$  der Zustandsbeschränkung:

**Annahme 1.1.3** Die aktive Menge  $\mathcal{A}$  besteht aus endlich vielen Berührungspunkten  $t_c^j$  und Randstücken  $[t_{\text{on}}^i, t_{\text{off}}^i]$  d.h.

$$\mathcal{A} = \bigcup_{i=1}^n [t_{\text{on}}^i, t_{\text{off}}^i] \cup \bigcup_{j=1}^m t_c^j, \quad n \in \mathbb{N}, \quad m \in \mathbb{N},$$

wobei die Mengen  $[t_{\text{on}}^i, t_{\text{off}}^i]$ ,  $\{t_c^j\}$  alle paarweise disjunkt sind. Weiterhin gelte  $\{0, t_f\} \notin \mathcal{A}$ .

Unter  $\mathcal{I} := (0, t_f) \setminus \mathcal{A}$  verstehen wir entsprechend die inaktive Menge der Zustandsbeschränkung. Die Zeiten  $S := \{t_{\text{on}}^i, t_{\text{off}}^i, t_c^j \mid i = 1, \dots, n, j = 1, \dots, m\}$  bezeichnen wir als die Schaltzeiten. An diesen Zeiten werden Übergangsbedingungen auftreten in Form von Stetigkeits- oder Sprungbedingungen.

**1.1.2 Notwendige Bedingungen nach dem JLS-Ansatz**

Jacobson, Lele und Speyer schlagen in [61] vor, die Zustandsbeschränkung direkt an die Hamilton-Funktion anzukoppeln (JLS-Ansatz). Wir definieren daher zunächst die Hamilton-Funktion  $H$ . Dank  $t_f = \int_0^{t_f} 1 \, dt$  besteht das Zielfunktional im Beispiel nur aus einem sogenannten Lagrange-Anteil. Die Hamilton-Funktion setzt sich bekanntlich zusammen aus den Integranden des Lagrange-Anteils und den Produkten aus adjungierten Zuständen  $\vec{p} := (p_1, p_2)$  und den rechten Seiten der ODE [78]:

$$H(\vec{x}, u, \vec{p}) := \lambda_0 \left( 1 + \frac{\lambda}{2} u^2 \right) + p_2 u + p_1 x_2. \quad (1.3)$$

Im Falle des JLS-Ansatzes definiert man die Lagrange-Funktion  $L$  unter zusätzlicher Ankopplung der Zustandsbeschränkung mit ihrem Lagrange-Multiplikator  $\eta^{(0)}$ :

$$L(\vec{x}, u, \vec{p}, \eta^{(0)}, t) := H(\vec{x}, u, \vec{p}) + \eta^{(0)} (x_1 - x_{\max}). \quad (1.4)$$

Die notwendigen Bedingungen nach dem Maximumprinzip werden mit dieser Hamilton- bzw. Lagrange-Funktion dargestellt, siehe z.B. [50, Section 4] bzw. [15]. Sie lauten:

**Lemma 1.1.4 (Notwendige Bedingungen aus dem Maximumprinzip nach JLS, analog zu [50, S. 186f])**

Unter der Annahme 1.1.3 existieren  $\lambda_0 \geq 0$  und stückweise stetige Funktionen  $\vec{p}$  (genannt adjungierte Zustände) und  $\eta^{(0)}$  mit  $(\lambda_0, p_1(t), p_2(t), \eta^{(0)}(t), \varrho_s) \neq 0 \forall t$ , so dass eine optimale Lösung  $(\bar{x}_1, \bar{x}_2, \bar{u})$  des Problems (1.1), (1.2) neben den primalen Gleichungen (1.2) noch folgende Bedingungen für alle  $t$  erfüllt:<sup>1</sup>

$$\bar{u}(t) = \arg \min_{u \in \mathbb{R}} H(\bar{x}_1(t), \bar{x}_2(t), u, \vec{p}(t)), \quad (1.5a)$$

$$H_u(\bar{x}_1, \bar{x}_2, \bar{u}, \vec{p}) = \lambda_0 \lambda \bar{u} + p_2 = 0, \quad (1.5b)$$

$$-\dot{p}_1 = L_{x_1}(\bar{x}_1, \bar{x}_2, \bar{u}, \vec{p}, t) = \eta^{(0)}, \quad (1.5c)$$

$$-\dot{p}_2 = L_{x_2}(\bar{x}_1, \bar{x}_2, \bar{u}, \vec{p}, t) = p_1, \quad (1.5d)$$

$$\eta^{(0)} \geq 0, \quad 0 = \eta^{(0)} (\bar{x}_1 - x_{\max}). \quad (1.5e)$$

Weiterhin gelten die Endbedingung  $p_2(t_f) = 0$ , die Bedingung an die freie Endzeit  $H(t_f) = 0$  sowie folgende Sprung- bzw. Stetigkeitsbedingungen zur Schaltzeit  $t_s \in S$ :

$$p_1(t_s^-) = p_1(t_s^+) + \varrho_s, \quad (1.6a)$$

$$H(t_s^-) = H(t_s^+) - \varrho_s \dot{x}_{\max}(t_s), \quad (1.6b)$$

$$p_2(t_s^-) = p_2(t_s^+). \quad (1.6c)$$

Dabei ist  $p_1(t_s^+)$  der rechtsseitige Grenzwert,  $p_1(t_s^-)$  der linksseitige Grenzwert der stückweise stetigen Funktion  $p_1$  und  $H(t)$  die Abkürzung für  $H(\bar{x}_1(t), \bar{x}_2(t), \bar{u}(t), \vec{p}(t))$ . Für eine reguläre Hamilton-Funktion ist ferner die Steuerung  $u$  stetig (z.B. [36, Korollar 6.2]).

Unter einer Regularitätsbedingung (Constraint Qualification, CQ) kann man den abnormalen Fall  $\lambda_0 = 0$  ausschließen, siehe z.B. [67, Assumption 3.1]. Da hier der normale Fall diskutiert werden soll, sei im Folgenden eine CQ erfüllt. Durch Normierung der Lagrange-Multiplikatoren kann dann bekanntlich ohne Einschränkung  $\lambda_0 = 1$  gewählt werden.

Interessant sind die Sprungbedingungen  $[p_1]_{t_s^-}^{t_s^+} = \varrho_s$ : An jedem Schaltpunkt  $t_s$  müssen wir also eine Sprunghöhe  $\varrho_s \in \mathbb{R}$  bestimmen. Die Bedingungen (1.6b) führen wir näher aus, da wir sie in Kap. 3 brauchen werden:

$$\begin{aligned} 0 = H(t_s^+) - H(t_s^-) - \varrho_s \dot{x}_{\max}(t_s) &= \frac{\lambda}{2} (u^2(t_s^+) - u^2(t_s^-)) + \underbrace{p_1(t_s^+) - p_1(t_s^-)}_{=\varrho_s} x_2(t_s) \\ &+ p_2(t_s)(u(t_s^+) - u(t_s^-)) - \varrho_s \dot{x}_{\max}(t_s) \end{aligned} \quad (1.7)$$

Nun ist die Steuerung wegen (1.5b), (1.6c) oder wegen regulärer Hamilton-Funktion [36, Korollar 6.2] stetig in den Schaltzeiten. Daher folgt für  $\varrho_s \neq 0$

$$x_2(t_s) = \dot{x}_{\max}(t_s) \quad \forall t_s \in S. \quad (1.8)$$

Der Übergang von  $x_1$  auf die Zustandsbeschränkung ist also im Regelfall stetig differenzierbar.

<sup>1</sup> Die Bedingungen gelten stückweise auf den einzelnen Zusammenhangskomponenten von  $\mathcal{A}$  und  $\mathcal{I}$ . Wir bleiben hier vorerst an der Literatur. Später werden wir dies genauer aufschlüsseln.

**Bemerkung 1.1.5** Endbedingungen für adjungierte Zustände erhält man nur dort, wo Zustände keinen Endbedingungen unterliegen. Wegen  $x_1(t_f) = 0$  liegt z.B. für  $p_1(t_f)$  keine Endbedingung vor.

Zur numerischen Lösung dieser notwendigen Bedingungen transformiert man sie in ein Mehrpunkt-Randwertproblem und löst dieses mit einem passenden Verfahren, z.B. der Mehrzielmethode oder dem Kollokationsverfahren [76]. Am Ende des nächsten Abschnittes wird ein derartiges Mehrpunkt-Randwertproblem exemplarisch aufgestellt.

Bei den notwendigen Bedingungen nach JLS ist das Mehrpunkt-Randwertproblem unter Benutzung von  $x_1 = x_{\max}$  auf  $\mathcal{A}$  allerdings ein DAE-System mit einem relativ hohen Index. Im vorliegenden Beispiel berechnet sich der Differentiationsindex beispielsweise zu 5, bis auch ODE für  $u$  und  $\eta^{(0)}$  auf  $\dot{\mathcal{A}}$  vorliegen. Ein derart hoher Index schließt eine numerische Lösung nahezu aus. Daher sollte man diesen Weg nicht einschlagen, siehe Abschnitt 1.3.2.

### 1.1.3 Ordnung einer Zustandsbeschränkung und der BDD-Ansatz

Bryson, Denham und Dreyfus schlagen in [16] bei OC-ODE vor, anstatt  $(x_1 - x_{\max})$  eine spezielle Ableitung dieses Terms an die Lagrange-Funktion anzukoppeln. Sie wählen die Ableitung gleich der „Ordnung der Zustandsbeschränkung“. Auf der aktiven Menge kann die Zustandsbeschränkung dann wie eine gemischte Steuerungs-Zustandsbeschränkung behandelt werden. Dies birgt einige Vorteile. So stellt sich z.B. heraus, dass die zugehörigen Lagrange-Multiplikatoren dadurch regulärer sind als die des JLS-Ansatzes [66]. Weiterhin umgeht man obiges DAE-Problem, da das resultierende Mehrpunkt-Randwertproblem einen deutlich geringeren Differentiations-Index aufweist (i.d.R. den Index 1, [28, p. 81]), bzw. sich sogar auf eine reine ODE reduzieren lässt. Weiterhin wird der algebraische Charakter der Zustandsbeschränkung beseitigt, wodurch u.a. die numerischen Methoden deutlich profitieren.

Wir definieren daher zuerst den Begriff der Ordnung einer Zustandsbeschränkung analog zu [16], [15, S. 118], [66]:

**Definition 1.1.6 (Ordnung  $m$  einer Zustandsbeschränkung)**

Die Ordnung  $m$  einer skalaren Zustandsbeschränkung  $g(\vec{x}, t) \leq 0$  bezogen auf die Zustandsgleichungen  $\dot{\vec{x}}(t) = \vec{f}(\vec{x}, \vec{u}, t)$  wird wie folgt definiert:

1. Starte mit  $m := 0$  und  $S^{(0)} := g(\vec{x}, t)$ .
2. Wenn  $\frac{\partial S^{(m)}}{\partial \vec{u}} \neq \vec{0}$ , dann ist  $m$  die Ordnung.
3. Andernfalls setze

$$S^{(m+1)} := \frac{d}{dt} S^{(m)} = \nabla_{\vec{x}} S^{(m)} \cdot \vec{f}(\vec{x}, \vec{u}, t) + \frac{\partial S^{(m)}}{\partial t}.$$

4. Setze  $m := m + 1$  und fahre fort mit Schritt 2.

Wir differenzieren also fortlaufend die Schaltfunktion  $S^{(0)}$  und setzen die Zustandsgleichungen ein, bis eine Komponente der Steuerung explizit auftritt. Die Anzahl der Differentiationsschritte ist die Ordnung der Zustandsbeschränkung. Bei vektorwertigen Zu-

standsbeschränkungen definiert man entsprechend eine Ordnung für jede Komponente. Angewandt auf das Prototyp-Problem (1.1), (1.2) ergibt sich Ordnung  $m = 2$ :

$$S^{(0)} = x_1 - x_{\max}, \quad S^{(1)} = x_2 - \dot{x}_{\max}, \quad S^{(2)} = u - \ddot{x}_{\max}. \quad (1.9)$$

Exemplarisch betrachten wir den Fall eines einzigen Randstückes  $\mathcal{A} = [t_{\text{on}}, t_{\text{off}}]$ . Die Lagrange-Funktion des BDD-Ansatzes lautet dann

$$L_{\text{BDD}}(\vec{x}, u, \vec{p}^{(2)}, \eta^{(2)}, t) := H(\vec{x}, u, \vec{p}^{(2)}) + \eta^{(2)}(u - \ddot{x}_{\max}). \quad (1.10)$$

Damit auf dem Randstück  $S^{(2)} \equiv 0$  auch äquivalent zu  $S^{(0)} \equiv 0$  ist, müssen zwei Punktbedingungen erfüllt werden, welche hier per Konvention auf  $t_{\text{on}}$  gelegt werden. Dies sind  $S^{(0)}(t_{\text{on}}) = 0$  und  $S^{(1)}(t_{\text{on}}) = 0$ .

Die daraus erhaltenen notwendigen Bedingungen findet man z.B. in [50, S. 198] oder in [36]. Sie lauten: Neben den primalen Gleichungen (1.2) gibt es stückweise stetige Funktionen  $p_1^{(2)}, p_2^{(2)}, \eta^{(2)}$  mit

$$\bar{u}(t) = \arg \min_{u \in \mathbb{R}} H(\bar{x}_1(t), \bar{x}_2(t), u, p_1^{(2)}(t), p_2^{(2)}(t)), \quad (1.11a)$$

$$0 = L_{\text{BDD},u}(\bar{x}_1, \bar{x}_2, \bar{u}, p_1^{(2)}(t), p_2^{(2)}(t), t) = \lambda \bar{u} + p_2^{(2)} + \eta^{(2)}, \quad (1.11b)$$

$$-\dot{p}_1^{(2)} = L_{\text{BDD},x_1}(\bar{x}_1, \bar{x}_2, \bar{u}, p_1^{(2)}(t), p_2^{(2)}(t), t) = 0, \quad (1.11c)$$

$$-\dot{p}_2^{(2)} = L_{\text{BDD},x_2}(\bar{x}_1, \bar{x}_2, \bar{u}, p_1^{(2)}(t), p_2^{(2)}(t), t) = p_1^{(2)}, \quad (1.11d)$$

$$\eta^{(2)} \geq 0, \quad (1.11e)$$

für alle  $t \in (0, t_f)$  und  $p^{(2)}(t_f) = 0, H(t_f) = 0$ . Zusätzlich muss die Komplementaritätsbedingung erfüllt sein, welche sich aufschlüsselt in

$$\eta^{(2)} = 0 \text{ auf } \mathcal{I}, \quad x_1 < x_{\max} \text{ auf } \mathcal{I}, \quad (1.11f)$$

$$u = \ddot{x}_{\max} \text{ auf } \mathcal{A}, \quad x_1(t_{\text{on}}) = x_{\max}(t_{\text{on}}), \quad x_2(t_{\text{on}}) = \dot{x}_{\max}(t_{\text{on}}). \quad (1.11g)$$

In (1.11g) erkennen wir die beiden oben erwähnten Punktbedingungen. Sprünge sind nach der verwendeten Konvention nur an den Eintrittspunkten möglich ([15, Formeln (3.11.6), (3.13.4)-(3.13.6)]):

$$p_1^{(2)}(t_{\text{on}}^-) = p_1^{(2)}(t_{\text{on}}^+) + \varrho_1, \quad (1.11h)$$

$$p_2^{(2)}(t_{\text{on}}^-) = p_2^{(2)}(t_{\text{on}}^+) + \varrho_2, \quad (1.11i)$$

$$p_1^{(2)}(t_{\text{off}}^-) = p_1^{(2)}(t_{\text{off}}^+), \quad (1.11j)$$

$$p_2^{(2)}(t_{\text{off}}^-) = p_2^{(2)}(t_{\text{off}}^+), \quad (1.11k)$$

$$H(t_{\text{on}}^-) = H(t_{\text{on}}^+) - \varrho_1 \dot{x}_{\max}(t_{\text{on}}) - \varrho_2 \ddot{x}_{\max}(t_{\text{on}}), \quad (1.11l)$$

$$H(t_{\text{off}}^-) = H(t_{\text{off}}^+). \quad (1.11m)$$

Dabei sind  $\varrho_1, \varrho_2 \in \mathbb{R}, \varrho_1, \varrho_2 \geq 0$ . Die beiden letzten Bedingungen sind näher zu untersuchen. Zum einen findet man

$$0 = H(t_{\text{off}}^+) - H(t_{\text{off}}^-) = \frac{\lambda}{2} (u^2(t_{\text{off}}^+) - u^2(t_{\text{off}}^-)) + \underbrace{p_2^{(2)}(t_{\text{off}}^-)}_{=-\lambda u(t_{\text{off}}^-)} (u(t_{\text{off}}^+) - u(t_{\text{off}}^-)) \quad (1.12)$$

$$= \frac{\lambda}{2} (u(t_{\text{off}}^+) - u(t_{\text{off}}^-))^2 \iff u(t_{\text{off}}^+) = u(t_{\text{off}}^-), \quad (1.13)$$

da  $p_1^{(2)}$ ,  $x_2$  und  $p_2^{(2)}$  in  $t_{\text{off}}$  stetig sind. Im Ggs. zu dem JLS-Ansatz in (1.7) ist hier die Stetigkeit der Steuerung sogar äquivalent zur Stetigkeit der Hamilton-Funktion. Gleiches findet man etwas aufwändiger im Aufsprungpunkt  $t_{\text{on}}$ . Hier gilt

$$\begin{aligned} H(t_{\text{on}}^-) - H(t_{\text{on}}^+) &= \left[ \frac{\lambda}{2} u^2 + p_2^{(2)} u \right]_{t_{\text{on}}^+}^{t_{\text{on}}^-} + \underbrace{(p_1^{(2)}(t_{\text{on}}^-) - p_1^{(2)}(t_{\text{on}}^+))}_{=\varrho_1} x_2(t_{\text{on}}) \\ &\stackrel{(1.11b)}{=} \left[ \frac{\lambda}{2} u^2 \right]_{t_{\text{on}}^-}^{t_{\text{on}}^+} + \eta^{(2)}(t_{\text{on}}^+) u(t_{\text{on}}^+) + \varrho_1 x_2(t_{\text{on}}). \end{aligned} \quad (1.14)$$

Daher folgt aus (1.11l) entsprechend

$$\begin{aligned} 0 &= \left[ \frac{\lambda}{2} u^2 \right]_{t_{\text{on}}^-}^{t_{\text{on}}^+} + \eta^{(2)}(t_{\text{on}}^+) u(t_{\text{on}}^+) + \varrho_1 \underbrace{[x_2(t_{\text{on}}) - \dot{x}_{\text{max}}(t_{\text{on}})]}_{=0 \text{ nach (1.11g)}} - \varrho_2 \underbrace{\ddot{x}_{\text{max}}(t_{\text{on}})}_{=u(t_{\text{on}}^+)} \\ &= \left[ \frac{\lambda}{2} u^2 \right]_{t_{\text{on}}^-}^{t_{\text{on}}^+} + u(t_{\text{on}}^+) \underbrace{[\eta^{(2)}(t_{\text{on}}^+) - \varrho_2]}_{\stackrel{(1.11b, 1.11i)}{=} \lambda u(t_{\text{on}}^-) - \lambda u(t_{\text{on}}^+)} = -\frac{\lambda}{2} (u(t_{\text{on}}^+) - u(t_{\text{on}}^-))^2. \end{aligned} \quad (1.15)$$

Da die Optimalitätsbedingungen intervallweise als Mehrpunkt-Randwertproblem gelöst werden sollen, sind in den Schaltzeiten alle Größen doppelt zu belegen. Die Differentialgleichungen werden auf den Bereichen I:=  $(0, t_{\text{on}})$ , II:=  $(t_{\text{on}}, t_{\text{off}})$ , III:=  $(t_{\text{off}}, t_f)$  stückweise gelöst. Die Zustände erfüllen dann folgende Übergangsbedingungen:

Endbedingung I	AB II	AB III
$x_1(t_{\text{on}}^-) = x_{\text{max}}(t_{\text{on}})$ ,	$x_1(t_{\text{on}}^+) = x_1(t_{\text{on}}^-)$ ,	$x_1(t_{\text{off}}^+) = x_1(t_{\text{off}}^-)$ ,
$x_2(t_{\text{on}}^-) = \dot{x}_{\text{max}}(t_{\text{on}})$ ,	$x_2(t_{\text{on}}^+) = x_2(t_{\text{on}}^-)$ ,	$x_2(t_{\text{off}}^+) = x_2(t_{\text{off}}^-)$ ,

Wir zählen die Gleichungen und Unbekannten wie folgt: Zunächst eliminiert man die Steuerung durch  $u = -p_2/\lambda$  auf I, III bzw.  $u = \ddot{x}_{\text{max}}$  auf II, sowie den Lagrange-Multiplikator auf II durch (1.11b). Übrig bleiben auf allen drei Bereichen so viele ODE wie jeweilige Anfangsbedingungen. Zusätzlich treten in dem Mehrpunkt-Randwertproblem folgende 6 Unbekannte auf:  $t_{\text{on}}$ ,  $t_{\text{off}}$ ,  $\varrho_1$ ,  $\varrho_2$ ,  $p_1(t_f)$ ,  $t_f$ .

Betrachten wir die zusätzlichen Gleichungen: Neben der Stetigkeit der Steuerung in beiden Schaltzeiten zählt man die beiden Endbedingungen des Intervalls I,  $H(t_f) = 0$  und  $x_1(t_f) = 0$ : Es sind also 6 zusätzliche Bedingungen, welche obige 6 Konstanten festlegen. Das Mehrpunkt-Randwertproblem weist also genauso viele Gleichungen wie Unbekannte auf. In Kap. 3 wird darauf aufbauend ein Mehrpunkt-Anfangsrandwertproblem eines OC-PDAE aufgestellt.

Im Zuge des BDD-Ansatzes verliert man bekanntlich einige Vorzeichenbedingungen an die Multiplikatoren, welche hinterher ergänzt werden müssen, damit die notwendigen Bedingungen äquivalent zu denen der direkten Ankopplung werden ([50, Remark 5.3 und 6.2]).

Die Hauptmerkmale des BDD-Ansatzes sind die folgenden: Die notwendigen Bedingungen enthalten auf der aktiven Menge jeweils eine algebraisch explizite Formel für die Steuerung  $u$ , sowie für den Multiplikator  $\eta^{(2)}$  der abgeänderten Zustandsbeschränkung.

Dies hat numerisch Vorteile, da auf der aktiven Menge die Steuerung  $u = \ddot{x}_{\max}$  und der Multiplikator  $\eta^{(2)} = -p_2 - \lambda u$  direkt eliminiert werden können. Damit sind keine algebraischen Gleichungen mehr vorhanden, man braucht keinen DAE-Solver, sondern kommt mit einem stückweisen ODE-Solver aus. Die numerische Umsetzung der verschiedenen Methoden werden in Abschnitt 1.3 kurz angesprochen.

## 1.2 Notwendige Bedingungen nach der formalen Lagrange-Technik

Wir gehen im folgenden Abschnitt einen Schritt in Richtung OC-PDE, indem wir die formale Lagrange-Technik (siehe z.B. Tröltzsch [96, Kap. 2.10 und 4.7]) auf das OC-ODE anwenden. Wir leiten also obige notwendige Bedingungen ein zweites Mal her – diesmal mit Techniken aus OC-PDE.

Die im folgenden hergeleiteten KKT-Bedingungen (Karush-Kuhn-Tucker-Bedingungen) gelten nur unter einer CQ. Verschiedene CQ bei OC-PDE findet man in [96, Kap. 6]. Eine dieser CQ wird in Kapitel 2 angewandt. Bei der formalen Lagrange-Technik nimmt man die Erfülltheit einer CQ bzw. der Existenz der Lagrange-Multiplikatoren in den passenden Räumen an. Zusätzlich muss die Existenz einer optimalen Steuerung gezeigt werden.

Für eine feste Endzeit ist das vorliegende Problem konvex, so dass die Slater-Bedingung als CQ sowie die Existenz der optimalen Steuerung mit Standard-Sätzen beweisbar ist. Bei freier Endzeit handelt es sich dagegen um ein nicht-konvexes Problem, so dass sich beide Punkte deutlich aufwändig gestalten. Einen Existenzsatz für freie Endzeiten findet man z.B. in [50, Theorem 3.1]. Da der Fokus auf den notwendigen Bedingungen liegt, werden die Beweise hier nicht erbracht.

**Annahme 1.2.1** Das Problem (1.1), (1.2) besitze eine Lösung. Die Lagrange-Multiplikatoren  $p_1, p_2$  zu den ODE existieren und besitzen die nötige Regularität. Der Lagrange-Multiplikator  $\mu$  zu der Zustandsbeschränkung existiere mit  $\mu \in C([0, t_f])^*$ .

Die Regularität von  $\mu(t)$  wird diskutiert in Abschnitt 1.2.2, bzw. in Kapiteln 2, 3 im Rahmen von OC-PDE. Tröltzsch [96, Kap. 6] liefert, dass eine optimale Lösung zusammen mit den Multiplikatoren einen Sattelpunkt der Lagrange-Funktion  $\mathcal{L}(x_1, x_2, u, t_f, p_1, p_2, \mu)$  bildet mit

$$\begin{aligned} \mathcal{L} &:= J(x_1, x_2, u, t_f) - \int_0^{t_f} (\dot{x}_1 - x_2) p_1 dt - \int_0^{t_f} (\dot{x}_2 - u) p_2 dt + \int_0^{t_f} (x_1 - x_{\max}) d\mu(t) \\ &= \int_0^{t_f} 1 + \frac{\lambda}{2} u^2 dt + \int_0^{t_f} x_1 \dot{p}_1 + x_2 p_1 dt + \int_0^{t_f} x_2 \dot{p}_2 + u p_2 dt - \left[ x_1 p_1 + x_2 p_2 \right]_0^{t_f} \\ &\quad + \int_0^{t_f} (x_1 - x_{\max}) d\mu(t). \end{aligned} \tag{1.16}$$

Die notwendigen Bedingungen für ein Optimum  $(\bar{x}_1, \bar{x}_2, \bar{u})$  enthalten die Zustandsgleichungen (1.2) sowie

$$\begin{aligned} -\dot{p}_1 &= \mu, & \mu(\bar{x}_1 - x_{\max}) &= 0 & \text{auf } (0, t_f), \\ -\dot{p}_2 &= p_1, & \mu &\geq 0 & \text{auf } (0, t_f), \\ 0 &= \lambda \bar{u} + p_2 & & & \text{auf } (0, t_f), \\ p_2(t_f) &= 0. & & & \end{aligned} \quad (1.17)$$

Wie die Gleichung  $-\dot{p}_1 = \mu$  für ein Maß  $\mu \in \mathcal{M}([0, t_f])$  zu verstehen ist, wird in Abschnitt 1.2.2 erklärt.

Die Bedingungen (1.2), (1.17) sowie die noch folgende Bedingung zur freien Endzeit (1.20) bilden zusammen die notwendigen Bedingungen 1. Ordnung. Sie werden auch als die KKT-Bedingungen (Karush-Kuhn-Tucker-Bedingungen) oder das KKT-System bezeichnet. Ihre Lösungen heißen entsprechend KKT-Punkte; sie sind die Kandidaten für die optimale Lösung. Adjungierten-basierte numerische Methoden der Optimalen Steuerung setzen i.d.R. an der Lösung der KKT-Bedingungen an.

### 1.2.1 Bedingung zur freien Endzeit

Zusätzlich benötigen wir die Bedingung der freien Endzeit (Transversalitätsbedingung). Ausgedrückt mit der Hamilton-Funktion lautete sie

$$0 = H(t_f) = 1 + \frac{\lambda}{2} u^2(t_f) + \underbrace{p_2(t_f)}_{=0} u(t_f) + p_1(t_f) x_2(t_f). \quad (1.18)$$

Für  $p_2(t_f)$  setzt man weiterhin die bereits bekannte Endbedingung ein. Die Gleichung (1.18) ist eine skalare Gleichung, die entsprechend die Information zur Bestimmung der freien Endzeit liefert. Wir erhalten diese Beziehung auch durch Differentiation der integralen Lagrange-Funktion nach der Endzeit  $t_f$ . Allerdings ist hier im Blick auf zeitabhängige Endbedingungen ein allgemeinerer Weg zu wählen: Die Endbedingungen sind mit eigenen Lagrange-Multiplikatoren an die Lagrange-Funktion anzukoppeln. Daher definieren wir mit  $\alpha \in \mathbb{R}$

$$\mathcal{L}^{(2)} := \mathcal{L} + x_1(t_f) \cdot \alpha. \quad (1.19)$$

Man erhält weiterhin alle Bedingungen, sowie zusätzlich aus der Ableitung nach dem Zustand  $x_1(t)$  die Formel  $\alpha = p_1(t_f)$ . Die Ableitung nach  $t_f$  ergibt für diese Lagrange-Funktion die passende Transversalitätsbedingung:

$$\begin{aligned} 0 &= D_{t_f} \mathcal{L}^{(2)} = 1 + \frac{\lambda}{2} u^2(t_f) + \underbrace{\left[ x_1 \dot{p}_1 + x_2 p_1 + x_2 \dot{p}_2 + u p_2 \right]_{t=t_f}}_{\text{Integranden an } t_f} \\ &\quad - \underbrace{\left[ \dot{x}_1 p_1 + x_1 \dot{p}_1 + x_2 \dot{p}_2 + \dot{x}_2 p_2 \right]_{t=t_f}}_{\text{Produktregel für die Endwerte}} + \alpha \dot{x}_1(t_f) \\ &= 1 + \frac{\lambda}{2} u^2(t_f) + \left[ x_2 p_1 + u \underbrace{p_2}_{=0} - \dot{x}_2 \underbrace{p_2}_{=0} \right]_{t=t_f} + \dot{x}_1(t_f) \underbrace{(\alpha - p_1(t_f))}_{=0}. \end{aligned} \quad (1.20)$$

Die Zustandsbeschränkung spielt keine Rolle, da in Annahme 1.1.3 auch  $t_f \notin \mathcal{A}$  angenommen war. Man beachte, dass die Differentialgleichungen hier nicht verwendet werden. Zuletzt werden die Endbedingungen der adjungierten Zustände berücksichtigt. Diese Methodik wird in Kapitel 3 aufgegriffen, um eine Endzeitbedingung eines gekoppelten ODE-PDE Optimalsteuerungsproblems mit freier Endzeit zu bestimmen.

**Bemerkung 1.2.2** Arbeitet man in der Theorie bereits mit der Transformation auf feste Endzeit ( $t = \tau \cdot t_f$ ,  $\tau \in [0, 1]$ ), so ist die Endzeitbedingung über die integrale Lagrange-Funktion mit Bedacht herzuleiten: Bei Bilden der totalen Ableitung nach  $t_f$  muss die Substitution berücksichtigt werden. Eine Umbenennung  $\tilde{u}(\tau) := u(t_f \cdot \tau)$  darf nur für feste  $t_f$  verwendet werden und z.B. nicht bei der Ableitung nach  $t_f$ . Wir verdeutlichen dies an einem einzigen Summand:

$$D_{t_f} \left\{ \int_0^{t_f} u^2(t) dt \right\} = u^2(t_f), \quad (1.21)$$

$$D_{t_f} \left\{ t_f \int_0^1 \tilde{u}^2(\tau) d\tau \right\} = \int_0^1 \tilde{u}^2(\tau) d\tau \stackrel{\text{Resubst.}}{=} \frac{1}{t_f} \int_0^{t_f} u^2(t) dt, \quad (1.22)$$

$$D_{t_f} \left\{ t_f \int_0^1 u^2(t_f \cdot \tau) d\tau \right\} = \int_0^1 u^2(t_f \cdot \tau) d\tau + \underbrace{t_f \int_0^1 2u(t_f \cdot \tau) \cdot \tau d\tau}_{\stackrel{\text{p.I.}}{=} [u^2(t_f \cdot \tau) \cdot \tau]_{\tau=0}^1 - \int_0^1 u^2(t_f \cdot \tau) d\tau} = u^2(t_f). \quad (1.23)$$

Die zweite Variante, also Gleichung (1.22), liefert das falsche Ergebnis, da die Umbenennung die Abhängigkeit von  $t_f$  verschleiert. Die beiden anderen Varianten liefern das korrekte Ergebnis, welches nur Werte zur Endzeit  $t_f$  enthält. Mit der Hamilton-Funktion erhält man dagegen auch nach der Umbenennung die korrekte Formel:

$$0 = H(1) = t_f \cdot \left( 1 + \frac{\lambda}{2} u^2(1) + [x_2 p_1 + u p_2]_{\tau=1} \right). \quad (1.24)$$

### 1.2.2 Analyse des Multiplikators $\mu(t)$ der Zustandsbeschränkung

Die Regularität des Lagrange-Multiplikators  $\mu(t)$  der Zustandsgleichung wird nun nachgereicht. Bekanntlich kommt der Multiplikator aus dem Dualraum der Beschränkung. Da Zustandsbeschränkungen sinnvoll im Raum der stetigen Funktionen untersucht werden sollten ([96, Kap. 6.2]), haben wir mit  $\mu \in C([0, t_f])^* \cong \mathcal{M}([0, t_f])$  im allgemeinen ein reguläres Borel-Maß (siehe z.B. [5, Satz 4.22]). Dabei ist  $\mathcal{M}([0, t_f])$  der Raum aller regulären Borel-Maße auf  $[0, t_f]$ ; er kann bekanntlich identifiziert werden mit dem Dualraum der stetigen Funktionen  $C([0, t_f])^*$  durch

$$\langle \tilde{\mu}_{[0, t_f]}, y \rangle_{C([0, t_f])^*, C([0, t_f])} = \int_{[0, t_f]} y d\mu_{[0, t_f]} \quad \forall y \in C([0, t_f]).$$

Die Norm in  $\mathcal{M}([0, t_f])$  ist dabei die totale Variation (z.B. [89])

$$\|\mu\|_{\mathcal{M}([0, t_f])} = \sup \left\{ \int_{[0, t_f]} y d\mu : y \in C([0, t_f]) \text{ und } \|y\|_\infty \leq 1 \right\}.$$

Das Maß  $\mu(t)$  hat bei OC-ODE allerdings eine ganz spezielle Form – weshalb man in den Veröffentlichungen zu OC-ODE mit Zustandsbeschränkung selten auf Maße treffen wird. Betrachten wir hierzu ein Randstück  $[t_{\text{on}}, t_{\text{off}}]$ : Aufgrund der Regularitäten der Daten liefert ein Vorgehen analog zu Bergounioux und Kunisch [13] auf dem Randstück folgende Aussagen:

$$x_1 \equiv x_{\text{max}} \in H^4(\mathring{\mathcal{A}}) \implies u = \ddot{x}_1 = \ddot{x}_{\text{max}} \in H^2(\mathring{\mathcal{A}}) \quad (1.25)$$

$$\implies p_2 = -\lambda u \in H^2(\mathring{\mathcal{A}}) \implies \mu = \ddot{p}_2 - x_1 + x_d \in L^2(\mathring{\mathcal{A}}). \quad (1.26)$$

Das Maß ist also im Inneren des Randstückes regulär. Da weiterhin wegen der Komplementaritätsbedingung bekanntlich  $\mu|_{\mathcal{I}} = 0$  gilt, bleiben die Schaltzeiten  $S$  zu untersuchen. Diese sind aber nur endlich viele Punkte. Ein Maß  $\mu(t) \in \mathcal{M}([0, t_f])$  auf endlich vielen (isolierten) Punkten lässt sich explizit angeben (Casas [21] oder Hilfssatz 3.4.19 in Kap. 3). Wir notieren diese Idee noch einmal in anderen Worten:

Nach der Lebesgue-Zerlegung kann man jedes Maß  $\mu \in \mathcal{M}([0, t_f])$  eindeutig darstellen als  $\mu = \mu_a + \mu_s$  mit einem absolut-stetigen Anteil  $\mu_a$  und einem singulären Anteil  $\mu_s$ , welche auf disjunkten Mengen leben (siehe z.B. Alt [5]). Der Satz von Radon-Nikodym liefert zudem, dass  $\mu_a$  eine  $L^1$ -Dichte besitzt [5, Satz 4.11]. Im vorliegenden Fall ist also  $\mu_s$  auf die Menge der Schaltzeiten  $S$  konzentriert. Da dies endlich viele, isolierte Punkte (im Inneren) sind, liefert Hilfssatz 3.4.19 später die Tatsache, dass  $\mu_s$  in jedem dieser Punkte ein Dirac-Maß ist, mit einer Dirac-Dichte, d.h.  $\mu_s = \sum_{t_k \in S} \varrho_{t_k} \cdot \delta(t - t_k)$ .

Die ODE  $-\dot{p}_1 = \mu$  ist daher stückweise zu verstehen: Nach Superposition  $p_1 = p_1^I + p_1^{II}$  mit  $-\dot{p}_1^I = \mu_a \in L^2(0, t_f)$  und  $-\dot{p}_1^{II} = \mu_s$  finden wir  $p_1^I \in H^1(0, t_f) \hookrightarrow C([0, t_f])$  und der Treppenfunktion  $p_1^{II} = -\sum_{t_k \in S} \varrho_{t_k} \cdot \sigma(t - t_k)$  mit der Sprungfunktion (Heaviside-Funktion)

$$\sigma(t) := \begin{cases} 0, & t < 0, \\ 1, & t \geq 0. \end{cases}$$

Die Adjungierte  $p_1$  ist also eine stückweise stetige Funktion, welche an jeder Schaltzeit springen kann. Daher findet man in der OC-ODE Literatur i.d.R. direkt Sprungbedingungen anstelle des singulären Anteils  $\mu_s$ .

### 1.2.3 Notwendige Bedingungen des aufgeteilten Problems analog zu BDD

Die Idee des BDD-Ansatzes soll nun mit der formalen Lagrange-Technik kombiniert werden. Die in Frey, Bechmann, Rund und Pesch [39] entwickelte Technik, welche wir in Kap. 2 einführen und beweisen werden, wird hierzu einmal formal auf OC-ODE angewandt. Exemplarisch betrachten wir den Fall eines Randstückes  $\mathcal{A} = [t_{\text{on}}, t_{\text{off}}] \subset (0, t_f)$ .

Die Idee ist, die Identität  $x_1 \equiv x_{\text{max}}$  auf  $\mathcal{A}$  mittels eines Randwertproblems  $\ddot{x}_1 = \ddot{x}_{\text{max}}$ ,  $x_1(t_{\text{on}}^+) = x_{\text{max}}(t_{\text{on}})$ ,  $x_1(t_{\text{off}}^-) = x_{\text{max}}(t_{\text{off}})$  äquivalent umzuformen und die Zustandsgleichungen  $\ddot{x} = u$  einzusetzen. Die erhaltenen Bedingungen werden mit  $q \in L^2(\mathring{\mathcal{A}})$  bzw.  $\sigma_{\text{on}}, \sigma_{\text{off}} \in \mathbb{R}$  an die Lagrange-Funktion angekoppelt. Im Gegensatz zu dem BDD-Ansatz aus Abschnitt 1.1.3 wird die Konvention geändert: Wir stellen nicht zwei Anfangsbedingungen in  $t_{\text{on}}$ , sondern je eine Randbedingung in  $t_{\text{on}}$  und  $t_{\text{off}}$ . Schließlich soll die Methodik in Kap. 2 auf elliptische Randwertprobleme übertragen werden.

Die aufgeteilte Lagrange-Funktion analog zum BDD-Ansatz wird also definiert als

$$\begin{aligned}
\mathcal{L}_{\text{BDD}} &:= \int_{\mathcal{I}} 1 + \frac{\lambda}{2} u^2 - \dot{x}_1 p_1 + x_2 p_1 - \dot{x}_2 p_2 + u p_2 dt \\
&+ \int_{\dot{\mathcal{A}}} 1 + \frac{\lambda}{2} u^2 - \dot{x}_1 p_1 + x_2 p_1 - \dot{x}_2 p_2 + u p_2 dt \\
&+ \int_{\dot{\mathcal{A}}} (u - \ddot{x}_{\max}) q dt + (x_1(t_{\text{on}}^+) - x_{\max}(t_{\text{on}})) \sigma_{\text{on}} + (x_1(t_{\text{off}}^-) - x_{\max}(t_{\text{off}})) \sigma_{\text{off}} \\
&+ \sum_{s \in \{\text{on}, \text{off}\}} (x_2(t_s^+) - x_2(t_s^-)) r_{1,s} + (x_1(t_s^+) - x_1(t_s^-)) r_{2,s}.
\end{aligned} \tag{1.27a}$$

Man beachte, dass wir nicht einfach die Integrale in der Lagrange-Funktion aufspalten, sondern zusätzliche Nahtbedingungen mit Multiplikatoren  $r_{1,s}, r_{2,s} \in \mathbb{R}$  ankoppeln (siehe Kap. 2). Die Ableitungen werden mittels partieller Integration auf die adjungierten Zustände übertragen:

$$\begin{aligned}
\mathcal{L}_{\text{BDD}} &= \int_{\mathcal{I}} 1 + \frac{\lambda}{2} u^2 + x_1 \dot{p}_1 + x_2 p_1 + x_2 \dot{p}_2 + u p_2 dt \\
&+ \int_{\dot{\mathcal{A}}} 1 + \frac{\lambda}{2} u^2 + x_1 \dot{p}_1 + x_2 p_1 + x_2 \dot{p}_2 + u p_2 dt \\
&+ \int_{\dot{\mathcal{A}}} (u - \ddot{x}_{\max}) q dt + (x_1(t_{\text{on}}^+) - x_{\max}(t_{\text{on}})) \sigma_{\text{on}} + (x_1(t_{\text{off}}^-) - x_{\max}(t_{\text{off}})) \sigma_{\text{off}} \\
&+ \sum_{s \in \{\text{on}, \text{off}\}} (x_2(t_s^+) - x_2(t_s^-)) r_{1,s} + (x_1(t_s^+) - x_1(t_s^-)) r_{2,s} \\
&- [x_1 p_1 + x_2 p_2]_0^{t_{\text{f}}} - [x_1 p_1 + x_2 p_2]_{t_{\text{on}}^+}^{t_{\text{on}}^-} - [x_1 p_1 + x_2 p_2]_{t_{\text{off}}^+}^{t_{\text{off}}^-},
\end{aligned} \tag{1.27b}$$

Betrachten wir nun die Zustände und adjungierten Variablen auf  $\mathcal{I}$  und  $\mathcal{A}$  als unabhängig, so können wir durch Richtungsableitung die aufgeteilten notwendigen Bedingungen herleiten.<sup>2</sup> Sie lauten unter nachträglicher Elimination von  $r_{1,s}, r_{2,s}$ :

$$\begin{array}{lll}
\mathcal{I} : & \dot{\mathcal{A}} : & \partial \mathcal{A} : \\
\dot{x}_1 = x_2 & \dot{x}_1 = x_2 & x_1(t_s^-) = x_1(t_s^+), \tag{1.28a} \\
\dot{x}_2 = u & \dot{x}_2 = u & x_2(t_s^-) = x_2(t_s^+), \tag{1.28b} \\
-\dot{p}_1 = 0 & -\dot{p}_1 = 0 & p_1(t_s^-) = p_1(t_s^+) + \sigma_s, \tag{1.28c} \\
-\dot{p}_2 = p_1 & -\dot{p}_2 = p_1 & p_2(t_s^-) = p_2(t_s^+), \tag{1.28d} \\
p_2 + \lambda u = 0 & p_2 + \lambda u + q = 0, & \tag{1.28e} \\
& u = \ddot{x}_{\max} & x_1(t_s) = x_{\max}(t_s), \tag{1.28f}
\end{array}$$

wobei mit  $t_s, s \in \{\text{on}, \text{off}\}$  die Schaltzeiten bezeichnet werden. Dazu erhält man unverändert die Anfangs- und Endbedingungen sowie die Bedingung der freien Endzeit. Die Komplementaritätsbedingung haben wir bereits eingearbeitet. Zusätzlich ist  $x_1 < x_{\max}$  auf  $\mathcal{I}$  zu den notwendigen Bedingungen hinzuzunehmen.

<sup>2</sup> Grundlagen zur formalen Lagrange-Technik findet der Leser z.B. in [96, Kap. 2.10 und 4.7]. Die Idee der Lagrange-Technik am aufgeteilten System stammt aus unserer Arbeit [39]. Wir besprechen sie detailliert in Kap. 2.

In Kapitel 2 werden wir diese aufgeteilte Lagrange-Technik ausführlicher behandeln. Fürs Erste konzentrieren wir uns auf ihr Ergebnis: Wir erkennen alle bekannten notwendigen Bedingungen außer den Vorzeichenbedingungen und  $u(t_s^+) = u(t_s^-)$  bzw. die gleichbedeutende Bedingung  $q(t_s) = 0$ . Letztere fehlt, da  $\frac{d\mathcal{L}}{dt_s} = 0$  noch nicht ausgewertet wurde. Unter Verwendung der Leibnizregel folgt exemplarisch für  $t_s = t_{\text{on}}$  direkt für (1.27a), aber auch für (1.27b) nach zusätzlichem Umklammern:

$$\begin{aligned}
 \frac{d}{dt_s} \mathcal{L}_{\text{BDD}} &= \left[ \frac{\lambda}{2} u^2 + \underbrace{(x_2 - \dot{x}_1)}_{=0 \text{ stetig}} p_1 + \underbrace{(u - \dot{x}_2)}_{=0 (*)} p_2 \right]_{t_s^+}^{t_s^-} \\
 &+ \underbrace{(u(t_s^+) - \ddot{x}(t_s))}_{\stackrel{(*)}{=} 0} q(t_s^+) + \left( \underbrace{\dot{x}_1(t_s^+) - \dot{x}_{\text{max}}(t_s)}_{\stackrel{(**)}{=} 0} \right) \sigma_s \\
 &+ \left( \underbrace{\dot{x}_2(t_s^+) - \dot{x}_2(t_s^-)}_{\stackrel{(*)}{=} u(t_s^+) \quad \stackrel{(*)}{=} u(t_s^-)} \right) \underbrace{r_{1,s}}_{=-p_2(t_s^-) \stackrel{(*)}{=} \lambda u(t_s^-)} + \underbrace{(\dot{x}_1(t_s^+) - \dot{x}_1(t_s^-))}_{=0 \text{ stetig}} r_{2,s} \quad (1.29) \\
 &= \frac{\lambda}{2} (u^2(t_s^-) - u^2(t_s^+) + 2u(t_s^-)u(t_s^+) - 2u^2(t_s^-)) = -\frac{\lambda}{2} (u(t_s^-) - u(t_s^+))^2.
 \end{aligned}$$

Bei (\*) wird jeweils die höhere Regularität  $u \in H^2(\dot{\mathcal{A}})$  sowie  $u \in H^2(\mathcal{I})$  ausgenutzt. (\*\*) verwendet  $x_1 \equiv x_{\text{max}}$  auf  $[t_{\text{on}}, t_{\text{off}}]$ , was aus den restlichen notwendigen Bedingungen gefolgert werden kann. Das Verschwinden der Richtungsableitung liefert also die Stetigkeit von  $u$  an jeder der Schaltzeiten  $t_s \in S$  – ähnlich zu Gleichung (1.15).

Ein Vergleich mit der Ableitung  $D_{t_t} \mathcal{L}$  aus (1.20) zeigt, dass die Ableitung nach den Schaltzeiten schwieriger auszuführen ist: Nur bei passenden Nahtbedingungen liefert sie das korrekte Ergebnis. Wir leiten derartige Nahtbedingungen in Kap. 2 für ein elliptisches Problem her. In Kap. 3.6.1 werden Nahtbedingungen bei einem zeitabhängigen ODE-PDE Problem angesprochen.

Wie wir sehen, ist diese Ableitung nach den Schaltzeiten  $t_s$  bereits aufwändig. Bei OC-PDE entspricht der Schaltzeit  $t_s$  – als Rand einer verteilten aktiven Menge – nunmehr eine Randkurve in  $\mathbb{R}^2$  oder sogar eine geschlossene gekrümmte Fläche in  $\mathbb{R}^3$ . Die Leibnizregel wird gewissermaßen zur Shape-Ableitung, die Berechnung wird daher sehr aufwändig (siehe Frey, Bechmann, Rund und Pesch [39], Frey [37]).

Die aus dem BDD-Ansatz erhaltenen notwendigen Bedingungen enthalten keine Vorzeichenbedingungen an die Multiplikatoren mehr, da die Zustandsbeschränkung als Gleichung behandelt wurde und nicht als Ungleichung. Wir erhalten schwächere notwendige Bedingungen. Fordert man die Äquivalenz der notwendigen Bedingungen, so erhält man aus dem Vergleich der Bedingungen (1.28c-1.28e) mit den bekannten Bedingungen (1.5b-

1.5e), (1.6a) entsprechend

$$q(t_s) = 0, \quad (1.30a)$$

$$\ddot{q} = \eta^{(0)} \geq 0, \quad (1.30b)$$

$$\sigma_{\text{on}} = \varrho_{\text{on}} - \underbrace{\dot{q}(t_{\text{on}}^+)}_{\geq 0}, \quad (1.30c)$$

$$\sigma_{\text{off}} = \varrho_{\text{off}} + \underbrace{\dot{q}(t_{\text{off}}^-)}_{\geq 0} \quad (1.30d)$$

mit den Folgen  $q \leq 0$  und  $\sigma_{\text{on}}, \sigma_{\text{off}} \geq 0$ .

Vergleichen wir nun die Ergebnisse mit denen aus der Literatur bzw. aus Abschnitt 1.1.3: Bis auf die unterschiedliche Konvention bei der Forderung der Punktbedingungen – und in der Folge die Verteilung der Sprungbedingungen – erhalten wir die meisten notwendigen Bedingungen, mit Ausnahme der mit dem Maximumprinzip erhaltenen stärkeren notwendigen Bedingung (1.11a). Vorteil der aufgeteilten Lagrange-Technik ist, dass wir sie auch bei OC-PDE verwenden können (da sie ja gerade von dort – aus [39] – stammt).

### 1.3 Numerische Umsetzung über ein Mehrpunkt-Randwertproblem

Es bieten sich drei verschiedene Methoden an, um zustandsbeschränkte OC-ODE ausgehend von einem Mehrpunkt-Randwertproblem zu lösen: Verwendet die Methode auf  $\mathcal{A}$  direkt die Zustandsbeschränkung, so sprechen wir in Analogie zu OC-ODE vom JLS-Ansatz, greift man dagegen auf eine Umformung der Zustandsbeschränkung zurück, welche die Steuerung explizit enthält, so bezeichnen wir dies als BDD-Ansatz. Eine dritte Methode bezeichnen wir als Eliminationsmethode. Sie nutzt die Zustandsbeschränkung auf der aktiven Menge zur Elimination von Zuständen und Steuerungen (siehe [15, S. 119], [61, S. 256] bzw. der dort zitierten Dissertation von Dreyfus). Der Vollständigkeit halber sprechen wir sie kurz an:

#### 1.3.1 Die Eliminationsmethode

Bei der Eliminationsmethode wird auf  $\mathcal{A}$  die Zustandsbeschränkung der Ordnung  $m$  sukzessive ausgenutzt, um  $m$  Zustände und eine Steuerung aus dem System zu entfernen [15, S. 119]. Man löst also stückweise auf der aktiven und der inaktiven Menge eine unterschiedliche Anzahl von ODE.

Führt man die Eliminationsmethode am KKT-System aus, so müssen konsistent auch  $m$  adjungierte Gleichungen eliminiert werden. Die Auswahl kann nichttrivial sein. Als Alternative kann man die Elimination direkt in der Problemstellung ausführen und erst anschließend das KKT-System aufstellen. In diesem Fall sind die dualen Variablen bereits kanonisch reduziert. In den Arbeiten Bulirsch und Chudej [18], Chudej und Günther [28] und Chudej [27] wird diese Methode aus dem Kontext der DAE als Wahl von Minimalkoordinaten bezeichnet. Man erzielt dort sehr gute Ergebnisse anhand komplex gekoppelter Anwendungsprobleme aus dem Bereich der Flugbahnoptimierung.

Die Eliminationsmethode hat den Vorteil, dass weniger Gleichungen gelöst werden müssen. Zusätzlich zeichnen sich bei Anfangs- oder Anfangsrandwertproblemen Vorteile aus der Sicht der differential-algebraischen Gleichungen (DAE) heraus ab: Die eliminierten Nebenbedingungen werden exakt eingehalten – zu einem „Drift-Off-Effekt“ kann es nicht kommen (siehe [28]).

### 1.3.2 Numerische Umsetzung der verschiedenen indirekten Verfahren

Numerisch umgesetzt wurden i.d.R. nur der BDD-Ansatz sowie die Eliminationsmethode [29]. Methoden, welche die notwendigen Bedingungen von JLS ohne Elimination von ODE lösen, gibt es praktisch keine – aus folgenden Gründen: Prinzipiell bieten sich hierfür zwei Möglichkeiten, welche beide nicht in Frage kamen.

Entweder löst man auf  $\mathcal{I}$ ,  $\mathcal{A}$  jeweils alle ODE zusammen mit den Gradientengleichungen und fügt auf  $\mathcal{A}$  zusätzlich die algebraische Gleichung  $x_1 = x_{\max}$  hinzu. Hierzu benötigt man allerdings einen DAE-Solver – und man handelt sich einen ungünstigen Differentiationsindex von i.d.R.  $2m + 1$  ein ([28]), wobei  $m \geq 1$  die Ordnung der Zustandsbeschränkung ist.

Oder man versucht den Multiplikator  $\eta^{(0)}$  zu eliminieren, um auf eine reine ODE zu kommen. Die dazu nötige algebraische Gleichung erhält man allerdings erst durch häufiges ( $2m$ -maliges) Ableiten und Einsetzen, siehe Kap. 2.4.3 im Kontext von OC-PDE. Dabei erbringt man allerdings bereits den Aufwand, welchen man für den BDD-Ansatz oder eine Eliminationsmethode benötigt – und zieht diese aufgrund weiterer Vorteile entsprechend vor.

Den BDD-Ansatz setzt man folgendermaßen um: Man löst alle ODE auf  $\mathcal{A}$  und  $\mathcal{I}$  – allerdings mit unterschiedlichen rechten Seiten. Auf  $\mathcal{A}$  eliminiert man die Steuerung  $u_j$  durch den BDD-Ansatz, sowie den Multiplikator  $\eta^{(m)}$  durch  $L_{\text{BDD},u_j} = 0$ . Übrig bleibt (unter kanonischer Elimination der anderen Steuerungen durch ihr Steuergesetz) ein Mehrpunkt-Randwertproblem – ein ODE-System für Zustände und adjungierte Zustände mit stückweise gegebenen rechten Seiten und Übergangsbedingungen. Man umgeht also die Problematik der DAE. Das Mehrpunkt-Randwertproblem löst man i.d.R. mit der Mehrzielmethode (siehe z.B. [19], [20], [76]).

Der BDD-Ansatz gewinnt an Bedeutung bei gekoppelten Systemen aus mehreren Zuständen und Steuerungen, da man hier die Bestimmungsformel der Steuerung und des Multiplikators nicht einfach ablesen kann wie im Prototyp-Beispiel.

### 1.3.3 Direkte Verfahren

Neben indirekten Verfahren der Optimalsteuerung wird diese Arbeit auch die direkten Verfahren aufgreifen. Abschnitt 3.5.1 gibt einen Überblick über die möglichen Varianten direkter Verfahren.

Die Idee einer direkten Methode ist es, das Problem ohne Kenntnis der adjungierten Gleichungen zu lösen. Dazu sind die Nebenbedingungen und das Zielfunktional als Erstes zu diskretisieren. Man erhält ein endlich-dimensionales Problem, ein nichtlineares Optimierungsproblem (NLP). Dieses Problem wird mit etablierten Methoden der nichtlinearen Optimierung gelöst. Eine gute Übersicht zur nichtlinearen Optimierung bereiten z.B. die Bücher Nocedal und Wright [74] sowie Geiger und Kanzow [42].

## 1.4 Zusammenfassung

Anhand eines Prototyp-Problems eines OC-ODE mit einer punktweisen Zustandsbeschränkung zweiter Ordnung und einer freien Endzeit werden zwei Varianten der notwendigen Bedingungen nach dem Maximumprinzip besprochen: Der JLS-Ansatz (nach Jacobson, Lele und Speyer), sowie der BDD-Ansatz (nach Bryson, Denham und Dreyfus).

Um den Anschluss an OC-PDE zu finden, werden beide Varianten von notwendigen Bedingungen erneut hergeleitet – diesmal mit der formalen Lagrange-Technik. Der Lagrange-Multiplikator der Zustandsgleichung ist hier zunächst ein reguläres Borel-Maß. Dieses Maß wird genauer untersucht. Unter einer Strukturannahme an die aktive Menge kann das Maß zerlegt werden in einen stetigen, regulären Anteil und einen singulären Anteil, eine Treppenfunktion. Der Zusammenhang zwischen den Sprungbedingungen und dem Borel-Maß wird dadurch klar.

Zur numerischen Lösung transformiert man die notwendigen Bedingungen in ein Mehrpunkt-Randwertproblem und löst dieses mit der Mehrzielmethode oder dem Kollokationsverfahren. Numerische Methoden bauen meist auf dem BDD-Ansatz auf, da das Mehrpunkt-Randwertproblem hier eine DAE von deutlich geringerem Index (i.d.R. Differentiationsindex 1) ist. Bei Standard-Problemen lassen sich sogar alle algebraischen Gleichungen beseitigen, so dass ein Mehrpunkt-Randwertproblem bei gewöhnlichen Differentialgleichungen entsteht. Das Mehrpunkt-Randwertproblem nach dem JLS-Ansatz weist dagegen einen derart hohen Index auf, dass es numerisch nicht sinnvoll gelöst werden kann.

## Kapitel 2

# BDD-Ansätze bei elliptischen OC-PDE mit Zustandsbeschränkungen

Um OC-PDE mit punktwisen Zustandsbeschränkungen für die theoretische Analyse und für numerische Methoden zugänglich zu machen, haben sich zwei Regularisierungsmethoden etabliert: Die Moreau-Yosida Regularisierung (siehe Hintermüller und Kunisch [54], [55]) sowie die Lavrentiev-Regularisierung (siehe Meyer, Rösch und Tröltzsch [72] oder auch [70], [84]). Während Erstere die Ungleichungen der Zustandsbeschränkungen herausnimmt und in einem veränderten Zielfunktional berücksichtigt, führt Letztere die Ungleichungen mit sich – jedoch werden sie durch Regularisierung zu gemischten Steuerungs-Zustandsbeschränkungen verändert. Beide Regularisierungstechniken führen einen Regularisierungsparameter ein, welcher iterativ angepasst werden muss und im Grenzübergang das Ausgangsproblem approximiert. Eine ausführliche Darstellung der beiden Regularisierungstechniken – sowie einen Vergleich anhand einiger Beispiele – findet man in Bechmann und Frey [9].

Nach der Regularisierung lassen sich etablierte Lösungsmethoden für gemischte Steuerungs-Zustandsbeschränkungen oder reine Steuerungsbeschränkungen anwenden wie z.B. Innere-Punkte-Methoden (z.B. [10], [84]) oder die primal-duale Aktive-Mengen-Strategie (PDAS, siehe Bergounioux, Ito und Kunisch [11] sowie Hintermüller, Ito und Kunisch [53]). Im Falle reiner Zustandsbeschränkungen selbst lässt sich die PDAS dagegen nicht im Funktionenraum formulieren und zeigt numerisch eine Gitterabhängigkeit, d.h. die Anzahl der Iterationen steigt mit der Verfeinerung der Gitter ([12], [55]). Innere-Punkte-Methoden lassen sich dagegen auch im Funktionenraum formulieren, so dass sie sich ohne Regularisierung auf zustandsbeschränkte OC-PDE anwenden lassen, siehe Schiela [92].

BDD-Ansätze gibt es derzeit im Bereich von OC-PDE dagegen noch nicht. Diese Lücke schließen wir mit unserer Arbeit Frey, Bechmann, Rund und Pesch [39], welche in den Abschnitten 2.1 bis 2.3 beschrieben und eingeordnet wird. Die wesentlichen Beweise findet der Leser in [39], sie werden hier nicht wiederholt.

Durch BDD-Ansätze wird die Zustandsbeschränkung auf der aktiven Menge äquivalent in eine Steuerungs-Zustandsbeschränkung oder ggf. eine reine Steuerungsbeschränkung transformiert. Wichtiger Unterschied zur Lavrentiev-Regularisierung ist, dass hier transformiert und nicht regularisiert wird.

Nach der Problemformulierung betrachten wir zunächst die vorhandenen notwendigen Bedingungen. In Abschnitt 2.2 leiten wir dann neue notwendige Bedingungen nach einem Analogon zum BDD-Ansatz bei OC-ODE aus Kap. 1 her. Die veränderte Betrachtungsweise, welche die Aufteilung in verschiedene Gebiete mit Übergangsbedingungen an dem Interface dazwischen mit sich bringt, führt auf eine neuartige Problemstellung in Abschnitt 2.3, in welcher sich die Problemstellungen der Topologie- und Shape-Optimierung als auch Optimalen Steuerung vereinen. Wir nennen diese Problemstellung ein Mengen-Optimalsteuerungsproblem. Abschnitt 2.4 gibt einen Ausblick auf Eliminationsmethoden und den Begriff der Ordnung von Zustandsbeschränkungen, um die verschiedenen Ansätze und Methoden aus Kapitel 1 auch bei OC-PDE einschätzen zu können.

## 2.1 Problemstellung und notwendige Bedingungen

Um neue notwendige Bedingungen analog zum BDD-Ansatz aus Kapitel 1 aufsetzen zu können, müssen wir zuerst die einzelnen Zusammenhangskomponenten der aktiven Menge kategorisieren. Bei OC-ODE gibt es hier im Regelfall entweder Berührungspunkte oder Randstücke, also 0- und 1-dimensionale Komponenten der aktiven Menge. Ein OC-PDAE, welches bewiesenermaßen mit diesen beiden Kategorien auskommt, behandeln wir ausführlich in Kapitel 3. Bei OC-PDE kann die aktive Menge im Allgemeinen aber auch Komponenten anderer Dimension besitzen. In diesem Kapitel entwickeln wir zunächst Methoden für  $n$ -dimensionale Komponenten anhand eines elliptischen Prototyp-Beispiels im Falle  $n = 2$ .<sup>1</sup> Wir entscheiden uns für folgendes linear-quadratisches elliptisches OC-PDE:

### Definition 2.1.1 (Prototyp eines elliptischen OC-PDE)

Sei  $\Omega \subset \mathbb{R}^2$  offen und beschränkt mit  $C^{1,1}$ -Rand  $\Gamma$ . Gegeben sei  $y_d \in H^1(\Omega)$ ,  $y_{\max} \in H^4(\Omega)$  und  $\lambda > 0$ . Es sei  $u \in L^2(\Omega)$ . Gesucht ist die Lösung  $(\bar{y}, \bar{u})$  des Optimalsteuerungsproblems

$$\min \quad J(y, u) = \frac{1}{2} \|y - y_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2} \|u\|_{L^2(\Omega)}^2, \quad (2.1a)$$

$$\text{s.t.} \quad -\Delta y + y = u \quad \text{f. ü. in } \Omega, \quad (2.1b)$$

$$\partial_\nu y = 0 \quad \text{f. ü. an } \Gamma, \quad (2.1c)$$

$$g(y) := y - y_{\max} \leq 0 \quad \text{f. ü. in } \Omega. \quad (2.1d)$$

Dabei ist  $\nu$  das äußere Einheitsnormalenvektorfeld von  $\Omega$ .

Zur weiteren Analyse benötigt man die Existenz einer Lösung  $(\bar{y}, \bar{u})$  dieses OC-PDE. Da es sich um ein Prototyp-Problem handelt, findet man jede Menge Resultate in der Literatur. So besitzt zum einen die Zustandsgleichung bei dieser passend gewählten Randregularität zu jeder Steuerung  $u \in L^2(\Omega)$  nach Standardsätzen bekanntlich genau eine Lösung  $y \in H^2(\Omega)$  [44, Theorem 2.2.2.5]. Darüberhinaus besitzt das Optimalsteuerungsproblem eine eindeutige Lösung, siehe z.B. Casas [22, Theorem 1]. Infolgedessen wird man die Lösung aus Optimalitätsbedingungen heraus bestimmen, z.B. denen aus [22] bzw. Alibert und Raymond [4]:

<sup>1</sup> In der vorliegenden Arbeit beschränken wir uns auf  $n = 2$ . Die Gültigkeit aller Aussagen für  $n = 3$  prüfen wir derzeit. Ab  $n \geq 4$  fehlt die stetige Einbettung  $H^2(\Omega) \hookrightarrow C(\bar{\Omega})$  nach [5, Satz 8.13], so dass die Theorie in dieser Form nicht anwendbar ist.

**Lemma 2.1.2 (Notwendige Bedingungen nach Casas [22, Theorem 2])**

Es existieren Lagrange-Multiplikatoren  $p \in W^{1,s}(\Omega)$  und  $\mu \in \mathcal{M}(\bar{\Omega})$  mit  $s \in [1, 2)$ , so dass zusätzlich zum Randwertproblem (2.1b), (2.1c) und der Zustandsbeschränkung (2.1d) folgende Optimalitätsbedingungen erfüllt sind:

$$\begin{aligned} -\Delta p + p &= y - y_d + \mu|_{\Omega} && \text{f. ü. in } \Omega, \\ \partial_{\nu} p &= \mu|_{\Gamma} && \text{f. ü. an } \Gamma, \\ p + \lambda u &= 0 && \text{f. ü. in } \Omega, \\ \int_{\bar{\Omega}} (y - y_{\max}) \, d\mu &= 0, \quad \mu \geq 0. \end{aligned}$$

Dabei ist die adjungierte PDE distributionell zu verstehen. Gemeint ist ihre schwache Formulierung, siehe [22], [23] oder [96, Kap. 7.2.3]. Wir gehen darauf in Kapitel 3 näher ein.

Aus den Optimalitätsbedingungen kann man bei näherem Betrachten weitere Eigenschaften herleiten, wenn man zwischen aktiver und inaktiver Menge unterscheidet. Die grundlegende Quelle zu einer derartigen strukturellen Analyse der Lagrange-Multiplikatoren bildet Bergounioux und Kunisch [13]: Durch Aufteilung in aktive und inaktive Menge, Herleitung der Übergangsbedingungen dazwischen und Anwendung von Regularitätstheorie können die Autoren die Maße sehr gut charakterisieren. Es zeigt sich, dass sich der singuläre Teil des Maßes auf den Rand der aktiven Menge konzentriert. Die Zerlegung in aktive und inaktive Menge bildet die Grundlage für alle weitergehenden Untersuchungen dieses Kapitels, daher definieren wir:

**Definition 2.1.3 (Aktive Menge)**

Die aktive Menge  $\mathcal{A}$  der Zustandsbeschränkung  $g(y) \leq 0$  ist gegeben durch

$$\mathcal{A} := \{x \in \Omega : g(\bar{y}(x)) = 0\}.$$

Weiterhin definieren wir die inaktive Menge  $\mathcal{I} := \Omega \setminus \mathcal{A}$  und das Interface  $\gamma := \partial\mathcal{A}$ .

Wie bei OC-ODE müssen wir die weitere Analyse mit Regularitätsannahmen an die aktive Menge absichern wie z.B., dass die aktive Menge nur aus endlich vielen Zusammenhangskomponenten besteht. Für dieses Kapitel schließen wir darüberhinaus niederdimensionale Anteile der aktiven Menge sowie die leere Menge aus. Wir landen bei den an dieser Stelle üblichen Annahmen an die aktive Menge (vgl. [13]).

**Annahme 2.1.4 (Regularität der aktiven Menge)**

$$\begin{aligned} \mathcal{A} &= \bigcup_{i=1}^l \mathcal{A}_i, \quad \bar{\mathcal{A}}_i = \mathcal{A}_i, \quad \mathcal{A} \cap \Gamma = \emptyset, \\ \mathcal{A}_i \cap \mathcal{A}_j &= \emptyset, \quad i \neq j, \quad i, j \in \{1, \dots, l\}, \\ \mathcal{A}_i &\text{ ist zusammenhängend und besitzt einen } C^{1,1}\text{-Rand } \forall i = 1, \dots, l, \\ \mathring{\mathcal{A}} &\neq \emptyset. \end{aligned}$$

Das Interface  $\gamma$  ist eine Niveaulinie der  $H^2$ -Funktion  $\bar{y} - y_{\max}$ . Leider kann man die allgemeine Frage nach der Regularität von Niveaulinien von  $H^k$ -Funktionen derzeit nicht

beantworten, weshalb wir die folgende Analyse mit der vorletzten Zeile der Annahme absichern mussten; sie ist nötig, damit z.B. der zusammengesetzte Zustand später wieder in  $H^2(\Omega)$  landet, also zulässig für das Ausgangsproblem ist.

Da wir algorithmisch die aktive Menge iterativ bestimmen werden, arbeiten wir in jedem Zwischenschritt mit einer Schätzung der aktiven Menge  $\mathcal{B}$  bzw. der inaktiven Menge  $\mathcal{J}$ . Diese Mengen sollen ebenfalls obige Regularitätsannahmen erfüllen. Daher definieren wir den Begriff der Konfiguration:

**Definition 2.1.5 (Konfiguration)**

Eine Aufteilung  $(\mathcal{B}, \mathcal{J})$  der Menge  $\Omega$  mit  $\mathcal{J} := \Omega \setminus \mathcal{B}$ , wobei  $\mathcal{B}$  die Annahme 2.1.4 erfüllt, nennen wir eine Konfiguration. Das Interface der Konfiguration bezeichnen wir mit  $\beta = \partial\mathcal{B}$ .

Betrachten wir zunächst eine beliebige Konfiguration  $(\mathcal{B}, \mathcal{J})$ . Wir teilen die Größe  $y$  gemäß der Konfiguration auf:

$$y_{\mathcal{J}} := y|_{\mathcal{J}}, \quad y_{\mathcal{B}} := y|_{\mathcal{B}}.$$

Analog definieren wir  $u_{\mathcal{J}}$  bzw.  $u_{\mathcal{B}}$ . Das äußere Einheitsnormalenvektorfeld von  $\mathcal{B}$  bzw.  $\mathcal{J}$  an ihrem Interface  $\beta$  bezeichnen wir entsprechend mit  $\nu_{\mathcal{B}}$  bzw.  $\nu_{\mathcal{J}}$ . Analog verwenden wir  $y_{\mathcal{A}}$ ,  $y_{\mathcal{I}}$  etc. bei der optimalen Konfiguration  $(\mathcal{A}, \mathcal{I})$ . Mit dieser Notation können wir die notwendigen Bedingungen aus [13] in unsere Notation bringen. Wir werden sie später oft als Vergleich heranziehen.

**Lemma 2.1.6 (Notw. Bedingungen analog zu Bergounioux und Kunisch [13])**

Unter Annahme 2.1.4 existieren Lagrange-Multiplikatoren  $p \in H^1(\Omega)$  und  $\mu \in \mathcal{M}(\bar{\Omega})$ , so dass die Lösung  $(\bar{y}_{\mathcal{A}}, \bar{y}_{\mathcal{I}}, \bar{u}_{\mathcal{A}}, \bar{u}_{\mathcal{I}})$  zusätzlich zu den Zustandsgleichungen (2.1b), (2.1c) und der Zustandsbeschränkung (2.1d) folgende Optimalitätsbedingungen erfüllt:

$$-\Delta p_{\mathcal{I}} + p_{\mathcal{I}} = \bar{y}_{\mathcal{I}} - y_{\text{d}}, \quad \lambda \bar{u}_{\mathcal{I}} + p_{\mathcal{I}} = 0 \quad \text{in } \mathcal{I}, \quad (2.2a)$$

$$-\Delta p_{\mathcal{A}} + p_{\mathcal{A}} = \bar{y}_{\mathcal{A}} - y_{\text{d}} + \mu_{\mathcal{A}}, \quad \lambda \bar{u}_{\mathcal{A}} + p_{\mathcal{A}} = 0 \quad \text{in } \mathring{\mathcal{A}}, \quad (2.2b)$$

$$p_{\mathcal{I}} = p_{\mathcal{A}}, \quad \partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} p_{\mathcal{A}} = \mu|_{\gamma} =: \mu_{\gamma} \quad \text{an } \gamma, \quad (2.2c)$$

$$\partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}} = 0 \quad \text{an } \Gamma, \quad (2.2d)$$

$$\mu_{\mathcal{I}} = 0, \quad \mu_{\mathcal{A}} \geq 0, \quad \mu_{\gamma} \geq 0. \quad (2.2e)$$

Wir lesen heraus, dass die linksseitige und die rechtsseitige Dirichlet-Spur des adjungierten Zustandes  $p$  am Interface übereinstimmen, während die Differenz der Neumann-Spuren von  $p$  dort um  $\mu|_{\gamma}$  springt. Weiterhin findet man dort unter den geforderten Daten eine höhere Regularität  $p_{\mathcal{I}}, u_{\mathcal{I}} \in H^2(\mathcal{I})$ ,  $p_{\mathcal{A}}, u_{\mathcal{A}} \in H^2(\mathring{\mathcal{A}})$ . Auch das Maß  $\mu$  zeigt eine höhere Regularität:  $\mu|_{\mathcal{A}} \in L^2(\mathring{\mathcal{A}})$ ,  $\mu|_{\gamma} \in H^{\frac{1}{2}}(\gamma)$  sowie  $\mu|_{\mathcal{I}} = 0$ .<sup>2</sup> Der Multiplikator ist also auf die aktive Menge konzentriert, sein singulärer Anteil (nach der Lebesgue-Zerlegung) ist sogar nur auf das Interface  $\gamma$  konzentriert. Auch wenn  $\mu_{\gamma}$  entlang des Interfaces eine reguläre Dichte besitzt, bleibt es auf  $\Omega$  gesehen jedoch ein Maß. Ferner sei erwähnt, dass sich aus den notwendigen Bedingungen  $u_{\mathcal{I}} = u_{\mathcal{A}} \in H^{\frac{3}{2}}(\gamma)$  folgern lässt (siehe z.B. [56]).

<sup>2</sup> Die Räume  $H^{\frac{1}{2}}(\gamma)$ ,  $H^{\frac{3}{2}}(\gamma)$  findet man z.B. in Grisvard [44], Adams und Fournier [2] im Zusammenhang von Spuren von  $H^k$ -Funktionen.

## 2.2 Neue notwendige Bedingungen nach dem BDD-Ansatz

In [13] werden die notwendigen Bedingungen des in  $\mathcal{A}$  bzw.  $\mathcal{I}$  aufgeteilten Problems durch Testfunktionenwahl (in Kombination mit Aufteilung von  $L^2$ -Integralen und der Anwendung der Greenschen Formel) aus den nicht aufgeteilten Optimalitätsbedingungen aus [22] aufgestellt. Da wir anhand von BDD-Ansätzen neue Optimalitätsbedingungen herleiten wollen, müssen wir einen anderen Weg schaffen. In [39] haben wir daher eine neue Methodik geschaffen, welche wir hier behandeln werden. Die Methodik lautet im Überblick:

1. Äquivalentes Aufteilen des OC-PDE auf die feste, optimale Konfiguration  $(\mathcal{A}, \mathcal{I})$ .
2. Anwendung eines BDD-Ansatzes.
3. Herleiten der notwendigen Bedingungen für das geteilte Problem an der optimalen Konfiguration.
4. Erweiterung auf nicht optimale Konfigurationen  $(\mathcal{B}, \mathcal{J})$ :  $\mathcal{B}$  ist jetzt eine weitere Optimierungsvariable.
5. Herleitung der notwendigen Bedingung am Interface mittels „Shape-Lagrange-Technik“.

Dieser Abschnitt behandelt die Schritte 1 bis 3. Die letzten beiden Schritte folgen in Abschnitt 2.3.

Zunächst lässt sich die Zustandsgleichung äquivalent auf eine beliebige Konfiguration aufteilen, wenn man passende Übergangsbedingungen fordert:

### Lemma 2.2.1 (Aufteilung der Zustandsgleichung)

Das Randwertproblem der Zustandsgleichung (2.1b), (2.1c) auf  $\Omega$  lässt sich äquivalent aufteilen auf die Konfiguration  $(\mathcal{B}, \mathcal{J})$  zu<sup>3</sup>

$$-\Delta y_{\mathcal{J}} + y_{\mathcal{J}} = u_{\mathcal{J}} \quad \text{in } \mathcal{I}, \quad -\Delta y_{\mathcal{B}} + y_{\mathcal{B}} = u_{\mathcal{B}} \quad \text{in } \mathcal{B}, \quad (2.3a)$$

$$\partial_{\nu} y_{\mathcal{J}}|_{\Gamma} = 0 \quad \text{an } \Gamma, \quad y_{\mathcal{J}} = y_{\mathcal{B}} \quad \text{an } \beta, \quad (2.3b)$$

$$\partial_{\nu_{\mathcal{J}}} y_{\mathcal{J}} + \partial_{\nu_{\mathcal{B}}} y_{\mathcal{B}} = 0 \quad \text{an } \beta. \quad (2.3c)$$

*Beweis:*

Siehe [39, Sect. 3]. □

Diese Aufteilung anhand einer beliebigen Konfiguration benötigen wir erst in Schritt 4 – also in Abschnitt 2.3. Bis dahin arbeiten wir fest an der optimalen Konfiguration  $(\mathcal{A}, \mathcal{I})$ .

Kommen wir nun zu Schritt 2, dem Anwenden des BDD-Ansatzes. In [39] übertragen wir die Idee von Bryson, Denham und Dreyfus von OC-ODE auf OC-PDE. Wir erläutern im Folgenden unser Vorgehen und die dahinterstehenden Ideen. Zunächst wird die

<sup>3</sup> Im Folgenden lassen wir „f. ü.“ der Übersichtlichkeit halber stets weg.

Zustandsbeschränkung auf der aktiven Menge  $\mathcal{A}$  durch ein passendes Randwertproblem ersetzt. Wir teilen dazu die Zustandsbeschränkung auf in  $y_{\mathcal{I}} < y_{\max}$  und  $y_{\mathcal{A}} = y_{\max}$ .<sup>4</sup> Man beachte, dass wir durch diese Transformation den Ungleichungscharakter auf  $\mathcal{A}$  vernachlässigen, da wir die Beschränkung dort als Gleichung betrachten. Eine Vorzeichenbedingung an den zugehörigen Multiplikator verlieren wir daher, unsere notwendigen Bedingungen werden etwas schwächer sein als die klassischen notwendigen Bedingungen (vgl. [50] im Falle OC-ODE). Wir ergänzen diese Vorzeichenbedingungen später durch Vergleich mit den klassischen notwendigen Bedingungen (vgl. [50]).

**Bemerkung 2.2.2** Die Zustandsbeschränkung wird behandelt als Gleichungsbeschränkung auf  $\mathcal{A}$  und strikte Ungleichungsbeschränkung  $y_{\mathcal{I}} < y_{\max}$ ; letztere ist bei der optimalen Konfiguration nicht aktiv und kann daher bekanntlich gestrichen werden ohne die Lösung des Problems zu beeinflussen. Wenn wir später die Konfiguration frei wählen (siehe Abschnitt 2.3), müssen wir diese Bedingung allerdings wieder berücksichtigen.

### 2.2.1 Der Bryson-Denham-Dreyfus-Ansatz bei OC-PDE

Zunächst möchten wir den BDD-Ansatz für OC-PDE zugänglich machen. Die Idee überträgt sich wie folgt: Wir ersetzen die Identität  $g(y) = y_{\mathcal{A}} - y_{\max} \equiv 0$  auf  $\mathcal{A}$  äquivalent durch ein Randwertproblem der passenden Ordnung, so dass die Steuerung explizit auftritt. Bei OC-ODE war dieses Vorgehen bereits nicht eindeutig bestimmt, da man Freiheiten in der Wahl der Anfangs- und Randbedingungen hatte (vgl. [15]). Bei OC-PDE hat man zusätzlich Freiheiten in der Wahl des Differentialoperators. Bei der vorliegenden linearen Zustandsgleichung legen wir uns darauf fest, stets den Differentialoperator der Zustandsgleichung zu verwenden. Dennoch erhalten wir bereits beim Prototyp-Problem verschiedene BDD-Ansätze je nach Randbedingung: Wir bezeichnen die zwei typischen BDD-Ansätze gehörend zu Dirichlet- und Neumann-Randbedingung als BDD-Ansatz erster und zweiter Art (abgekürzt: BDD1, BDD2). Die zugehörigen Randwertprobleme lauten

$$-\Delta g + g = 0, \quad -\Delta g + g = 0 \quad \text{in } \mathring{\mathcal{A}}, \quad (2.4)$$

$$g = 0, \quad \partial_{\nu_{\mathcal{A}}} g = 0 \quad \text{an } \gamma. \quad (2.5)$$

Als zweiter Schritt ist die Zustandsgleichung einzusetzen, so dass die Steuerung explizit auftritt. In beiden Fällen liefert dies  $0 = -\Delta g + g = u_{\mathcal{A}} + \Delta y_{\max} - y_{\max}$  sowie die Randbedingung des entsprechenden Falles. Standardsätze nach Grisvard [44] liefern dank der  $C^{1,1}$ -Regularität des Interfaces  $\gamma$  die Äquivalenz der Lösung des Randwertproblems zu der Ausgangsgleichung  $y_{\mathcal{A}} = y_{\max}$  auf  $\mathcal{A}$ .

Verwendet man diese Umformung auf Ebene des Optimalsteuerungsproblems, so können wir mit Theorem 4.1 aus [39] beweisen, dass die Lösung davon unverändert bleibt.

### 2.2.2 Die Lagrange-Technik am aufgeteilten System

In [39] entwickeln wir eine neue Methode, um die Optimalitätsbedingungen des aufgeteilten Systems herzuleiten: Die Lagrange-Technik am aufgeteilten System. Die Technik gestattet es uns, die nötigen Veränderungen des BDD-Ansatzes einzubauen. Dass die Lagrange-Technik zum Ziel führt, beweisen wir in [39, Theorem 5.1] über die Zowe-Kurcysz-Bedingung<sup>5</sup>.

<sup>4</sup> Bei den Daten  $y_{\max}$ ,  $y_d$  werden wir der Übersichtlichkeit halber die Einschränkungen  $|_{\mathcal{I}}$  etc. weglassen. Sie gehen aus dem Kontext hervor.

<sup>5</sup> Hintergründe zu den verschiedenen Constraint-Qualifications findet der Leser in [96, Kap. 6].

Wir zeigen dort die Existenz von Lagrange-Multiplikatoren<sup>6</sup>  $p_{\mathcal{I}} \in L^2(\mathcal{I}, \Delta)$  bzw.  $p_{\mathcal{A}} \in L^2(\mathring{\mathcal{A}}, \Delta)$  zu den aufgeteilten Zustandsgleichungen (2.3), sowie  $(q, \sigma) \in L^2(\mathring{\mathcal{A}}) \times H^{-\frac{3}{2}}(\gamma)$ .<sup>7</sup> Da wir die Zustandsbeschränkung hier als Gleichungsnebenbedingung für die Steuerung behandeln, ist ihr Lagrange-Multiplikator  $q$  von Anfang an in  $L^2(\mathring{\mathcal{A}})$ . Hinterher werden wir höhere Regularität zeigen z.B.  $q \in H^2(\mathring{\mathcal{A}})$ .

Die Nahtbedingungen an  $\gamma$  erhalten eigene Multiplikatoren  $r_1, r_2$ . Damit definieren wir die Lagrange-Funktion  $\mathcal{L}$  des an der optimalen Konfiguration  $(\mathcal{A}, \mathcal{I})$  aufgeteilten OC-PDE zu

$$\mathcal{L}(y_{\mathcal{I}}, y_{\mathcal{A}}, u_{\mathcal{I}}, u_{\mathcal{A}}, r_1, r_2, q, \sigma) := \mathcal{L}_1(y_{\mathcal{I}}, y_{\mathcal{A}}, u_{\mathcal{I}}, u_{\mathcal{A}}, r_1, r_2) + \mathcal{L}_{\text{BDD1}}(q, \sigma) \quad (2.6)$$

mit dem allgemeinen Anteil

$$\begin{aligned} \mathcal{L}_1 := & \int_{\mathcal{I}} (y_{\mathcal{I}} - y_{\text{d}})^2 dx + \int_{\mathring{\mathcal{A}}} (y_{\mathcal{A}} - y_{\text{d}})^2 dx + \frac{\lambda}{2} \int_{\mathcal{I}} u_{\mathcal{I}}^2 dx + \frac{\lambda}{2} \int_{\mathring{\mathcal{A}}} u_{\mathcal{A}}^2 dx \\ & - \int_{\mathcal{I}} (-\Delta y_{\mathcal{I}} + y_{\mathcal{I}} - u_{\mathcal{I}}) p_{\mathcal{I}} dx - \int_{\mathring{\mathcal{A}}} (-\Delta y_{\mathcal{A}} + y_{\mathcal{A}} - u_{\mathcal{A}}) p_{\mathcal{A}} dx - \langle p_{\mathcal{I}}, \partial_{\nu} y_{\mathcal{I}} \rangle_{H^{-\frac{1}{2}}(\Gamma), H^{\frac{1}{2}}(\Gamma)} \\ & + \langle r_1, y_{\mathcal{I}} - y_{\mathcal{A}} \rangle_{H^{-\frac{3}{2}}(\gamma), H^{\frac{3}{2}}(\gamma)} + \langle r_2, \partial_{\nu_{\mathcal{I}}} y_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} y_{\mathcal{A}} \rangle_{H^{-\frac{1}{2}}(\gamma), H^{\frac{1}{2}}(\gamma)}, \end{aligned} \quad (2.7)$$

sowie dem speziellen Anteil des BDD-Ansatzes erster Art

$$\mathcal{L}_{\text{BDD1}} := \int_{\mathring{\mathcal{A}}} (u_{\mathcal{A}} + \Delta y_{\text{max}} - y_{\text{max}}) q dx + \langle \sigma, y_{\mathcal{A}} - y_{\text{max}} \rangle_{H^{-\frac{3}{2}}(\gamma), H^{\frac{3}{2}}(\gamma)}. \quad (2.8)$$

Dabei ist  $\langle \cdot, \cdot \rangle_{V^*, V}$  die gebräuchliche Bezeichnung für das Dualitätsprodukt auf einem Banachraum  $V$  und  $H^{-\frac{3}{2}}(\gamma) = H^{\frac{3}{2}}(\gamma)^*$  bzw.  $H^{-\frac{1}{2}}(\gamma) = H^{\frac{1}{2}}(\gamma)^*$ . Hervorzuheben ist, dass die präsentierte Lagrange-Funktion nicht einfach durch Aufteilen der Integrale der üblichen Lagrange-Funktion hervorgeht.

Wir führen nun die aufgeteilte Lagrange-Technik für den BDD-Ansatz erster Art aus und tabellieren anschließend die Unterschiede des anderen BDD-Ansatzes. Dessen Lagrange-Funktionen enthält entsprechend den speziellen Anteil

$$\mathcal{L}_{\text{BDD2}} := \int_{\mathring{\mathcal{A}}} (u_{\mathcal{A}} + \Delta y_{\text{max}} - y_{\text{max}}) q dx + \langle \sigma, \partial_{\nu_{\mathcal{A}}} (y_{\mathcal{A}} - y_{\text{max}}) \rangle_{H^{-\frac{1}{2}}(\gamma), H^{\frac{1}{2}}(\gamma)}. \quad (2.9)$$

**Bemerkung 2.2.3** Wir behandeln  $y_{\mathcal{I}}, y_{\mathcal{A}}$  als unabhängige Größen mit entsprechend unabhängigen Multiplikatoren. Die Lagrange-Technik liefert uns die Zusammenhänge dieser Größen als Teil des Ergebnisses. Insbesondere sind  $p_{\mathcal{I}}$  und  $p_{\mathcal{A}}$  an dieser Stelle noch nicht als Einschränkungen einer globalen Größe  $p$  auf die Mengen  $\mathcal{I}$  bzw.  $\mathcal{A}$  zu verstehen.

<sup>6</sup> Man beachte, dass jeder Ansatz seine eigenen dualen Variablen besitzt. Der Übersichtlichkeit halber wird kein weiterer Index eingeführt.

<sup>7</sup> Die Eigenschaften der verwendeten sehr schwachen Räume  $L^2(C, \Delta) := \{\varphi \in L^2(C) \mid \Delta \varphi \in L^2(C)\}$  findet man in [8, Section 7.1-9].

Die Berechnung der notwendigen Bedingungen mit der formalen Lagrange-Technik aus [96] teilt sich in einige Schritte auf: Zunächst rekonstruieren sich die aufgeteilten Zustandsgleichungen mit Nahtbedingungen durch Verschwinden der Richtungsableitungen bzgl.  $p_{\mathcal{I}}, p_{\mathcal{A}}, r_1, r_2$ . Anschließend werden die Laplace-Operatoren mittels (verallgemeinerter) Greenscher Formel von den Zuständen auf die Multiplikatoren übertragen. Jetzt lassen sich die restlichen notwendigen Bedingungen über Berechnung der Richtungsableitungen und übliche Variationsargumente ablesen (siehe z.B. [96]). Zusätzlich zu den aufgeteilten Zustandsgleichungen aus (2.3) erhalten wir unter nachträglicher Elimination der Multiplikatoren  $r_1, r_2$  folgende, beide Ansätze betreffende Optimalitätsbedingungen

$$-\Delta p_{\mathcal{I}} + p_{\mathcal{I}} = y_{\mathcal{I}} - y_d, \quad p_{\mathcal{I}} + \lambda u_{\mathcal{I}} = 0 \quad \text{in } \mathcal{I}, \quad (2.10a)$$

$$\partial_{\nu} p_{\mathcal{I}} = 0 \quad \text{an } \Gamma, \quad (2.10b)$$

$$y_{\mathcal{A}} = y_{\max} \quad \text{an } \gamma, \quad (2.10c)$$

sowie speziell für den ersten BDD-Ansatz die Bedingungen

$$-\Delta p_{\mathcal{A}} + p_{\mathcal{A}} = y_{\mathcal{A}} - y_d, \quad p_{\mathcal{A}} + \lambda u_{\mathcal{A}} + q = 0 \quad \text{in } \mathring{\mathcal{A}}, \quad (2.11a)$$

$$p_{\mathcal{I}} = p_{\mathcal{A}}, \quad \partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} p_{\mathcal{A}} = \sigma \quad \text{an } \gamma, \quad (2.11b)$$

$$u_{\mathcal{A}} = -\Delta y_{\max} + y_{\max} \quad \text{in } \mathring{\mathcal{A}}. \quad (2.11c)$$

Den zweiten BDD-Ansatz behandeln wir ausführlich in [39, Sect. 6]. Die Resultate behandeln wir tabellarisch im nächsten Abschnitt im Überblick.

Wir erkennen die von OC-ODE bekannten Hauptmerkmale des BDD-Ansatzes (vgl. Abschnitt 1.3.2): Zum einen besitzt das Optimalitätssystem eine Darstellungsformel (2.11c) für die Steuerung auf der aktiven Menge, zum anderen liefert die Gradientengleichung in (2.11a) direkt den Lagrange-Multiplikator  $q$  des BDD-Ansatzes. Diese beiden algebraischen Gleichungen erlaubten es bei OC-ODE, das KKT-System ohne Umweg auf eine stückweise gegebene ODE zu reduzieren, was stark zum praktischen Durchbruch der BDD-Ansätze beitrug ([75], [26]).

#### Bemerkung 2.2.4 (Höhere Regularitäten der dualen Variablen)

Aus unserem allgemeineren Theorem 7.5 aus [39] können wir höhere Regularitäten  $q \in H^2(\mathring{\mathcal{A}})$ ,  $p_{\mathcal{I}} \in H^2(\mathcal{I})$ ,  $p_{\mathcal{A}} \in H^2(\mathcal{A})$  und  $\sigma \in H^{\frac{1}{2}}(\gamma)$  folgern.

### 2.2.3 Vergleich der notwendigen Bedingungen

Im Gegensatz zu den klassischen notwendigen Bedingungen zeigen beide BDD-Ansätze folgende Unterschiede auf  $\mathring{\mathcal{A}}$ :

$$\text{klassisch: } -\Delta p_{\mathcal{A}} + p_{\mathcal{A}} = y_{\mathcal{A}} - y_d + \mu_{\mathring{\mathcal{A}}}, \quad p_{\mathcal{A}} + \lambda u_{\mathcal{A}} = 0, \quad y_{\mathcal{A}} = y_{\max}.$$

$$\text{BDD: } -\Delta p_{\mathcal{A}} + p_{\mathcal{A}} = y_{\mathcal{A}} - y_d, \quad p_{\mathcal{A}} + q + \lambda u_{\mathcal{A}} = 0, \quad u_{\mathcal{A}} = -\Delta y_{\max} + y_{\max}.$$

Als Regularität finden wir  $q \in H^2(\mathring{\mathcal{A}})$  [39, Prop. 5.2] im Gegensatz zu  $\mu_{\mathring{\mathcal{A}}} \in L^2(\mathring{\mathcal{A}})$  aus Lemma 2.1.6. Dies steht im Einklang mit OC-ODE ([66], [50]): Der Multiplikator des BDD-Ansatzes zeigt eine höhere Regularität verglichen zur direkten Ankopplung der Zustandsbeschränkung.

Zusätzlich unterscheiden sich alle drei Varianten in den notwendigen Bedingungen am Interface  $\gamma$ , dargestellt in folgender Tabelle:<sup>8</sup>

klassisch:	$p_{\mathcal{I}} = p_{\mathcal{A}}$	$\partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} p_{\mathcal{A}} = \mu_{\gamma}$	$\mu_{\gamma} \in H^{\frac{1}{2}}(\gamma)$
BDD1:	$p_{\mathcal{I}} = p_{\mathcal{A}}$	$\partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} p_{\mathcal{A}} = \sigma$	$\sigma \in H^{\frac{1}{2}}(\gamma)$
BDD2:	$p_{\mathcal{I}} = p_{\mathcal{A}} + \sigma$	$\partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} p_{\mathcal{A}} = 0$	$\sigma \in H^{\frac{3}{2}}(\gamma)$

Wir sehen insbesondere, dass sich die Regularität von  $\sigma$  beim Ansatz BDD2 erhöht. Dies liegt daran, dass die Randbedingung (2.5) des Randwertproblems auf Ableitungen zurückgreift. Die aus Maurer [66] bekannte Regel greift auch hier: Je höher die Ableitungen der Zustandsbeschränkung im Ansatz sind, desto regulärer sind die zugehörigen Lagrange-Multiplikatoren.

Betrachten wir abschließend die beiden Multiplikatoren  $p_{\mathcal{I}}$ ,  $p_{\mathcal{A}}$  als Teile eines globalen Multiplikators  $p$ , so erkennen wir Folgendes: Je nach BDD-Ansatz verändert man den adjungierten Zustand auf der aktiven Menge, indem man mit  $\sigma$  Knicke oder Sprünge erzeugt. So springt  $p$  beispielsweise im Falle des BDD2-Ansatzes am Interface, zeigt aber eine stetig fortsetzbare Spur der Normalenableitung.

Bei unseren BDD-Ansätzen steckt die Optimalitätsbedingung  $u_{\mathcal{I}} = u_{\mathcal{A}}$  an  $\gamma$  zunächst noch nicht im System, da eine zusätzliche Unbekannte  $q|_{\gamma}$  auftritt. Die Argumentation, welche bei den klassischen notwendigen Bedingungen aus Lemma 2.1.6 auf  $u_{\mathcal{I}} = u_{\mathcal{A}}$  an  $\gamma$  führt, ist daher hier nicht anwendbar.

Dank der Äquivalenz der Problemstellungen aus [39, Theorem 4.1] können wir die erhaltenen notwendigen Bedingungen mit denen aus Lemma 2.1.6 vergleichen. Der Vergleich liefert eine zusätzliche Randbedingung am Interface, sowie die fehlenden Vorzeichenbedingungen: Wir erhalten  $-\Delta q + q = \mu_{\dot{\mathcal{A}}}$ , sowie je nach Ansatz:

Ansatz	zusätzliche RB an $\gamma$	Vergleichsbedingung	VZ-Bedingung
BDD1	$q = 0$	$\mu_{\gamma} = \sigma + \partial_{\nu_{\mathcal{A}}} q$	$q \geq 0, \sigma \geq 0$
BDD2	$q = \sigma$	$\mu_{\gamma} = \partial_{\nu_{\mathcal{A}}} q$	$q \geq 0, \sigma \geq 0$

Mit der zusätzlichen Randbedingung ließe sich  $u_{\mathcal{I}} = u_{\mathcal{A}}$  an  $\gamma$  wieder folgern. Wir werden die Gleichung  $u_{\mathcal{I}} = u_{\mathcal{A}}$  an  $\gamma$  im nächsten Abschnitt aber auch aus einem anderen Zugang erhalten.

### Bemerkung 2.2.5 (Elimination der Maß-Anteile $\sigma$ , $\mu_{\gamma}$ )

Beim Ansatz BDD2 können wir  $\sigma$  und  $\mu_{\gamma}$  durch passende Spuren von  $q$  eliminieren. Dadurch treten im KKT-System keine Maße mehr auf, sondern nur noch Dirichlet- bzw. Neumann-Spuren einer  $H^2(\dot{\mathcal{A}})$ -Funktion  $q$ .

**Bemerkung 2.2.6** Die BDD-Ansätze lassen sich entsprechend auch durch spezielle Substitutionen aus den klassischen aufgeteilten notwendigen Bedingungen aus [13] herleiten. Allerdings führen nicht alle Substitutionen auf BDD-Ansätze und nicht jede Substitution eines BDD-Ansatzes ist naheliegend. Daher empfehlen wir das Vorgehen aus Abschnitt 2.2 bzw. [39].

<sup>8</sup> Man beachte, dass jeder Ansatz seine eigenen dualen Größen besitzt, d.h.  $p_{\mathcal{A}}$ ,  $q$ ,  $\sigma$  sind hier zeilenweise verschieden.

Als nächstes verlassen wir die optimale Konfiguration. Wir erhalten darauf ein sogenanntes Mengen-Optimalsteuerungsproblem. Infolgedessen leiten wir in [39, Formel (7.36)] die Bedingung  $u_{\mathcal{I}} = u_{\mathcal{A}}$  an  $\gamma$  aus einer partiellen Shape-Ableitung der Lagrange-Funktion nach dem Interface her. Dies entspricht der Ableitung nach den Schaltzeiten aus Kap. 1.2.3, Formel (1.29).

## 2.3 Das Mengen-Optimalsteuerungsproblem

Wir möchten jetzt das Optimalsteuerungsproblem aus einem anderen Blickwinkel betrachten. Wir geben dabei nur einen kurzen Leitfaden zu dem Abschnitt 7 aus [39]. Den interessierten Leser verweise ich auf unsere zukünftigen gemeinsamen Arbeiten (z.B. [38]) – sowie auf die Dissertationen meiner beiden Kollegen Simon Bechmann und Michael Frey [37].

Bei Mehrpunkt-Randwertproblemen im Bereich von OC-ODE in Kap.1 werden die Schaltzeiten – also die Randpunkte der aktiven Menge – als unbekannte Größen innerhalb des KKT-Systems behandelt. Ableitungen der Lagrange-Funktion nach den Schaltzeiten liefern hier zugehörige notwendige Bedingungen z.B. Formel (1.29). Alternativ liefert das Pontryaginsche Maximumprinzip eine Stetigkeits oder Sprungbedingung der Hamilton-Funktion an der Schaltzeit – siehe Gleichungen (1.6b), (1.11l, 1.11m).

Fortgeführt auf OC-PDE betrachten wir daher die Menge  $\mathcal{B}$  – bzw. ihren Rand  $\beta$  – als weitere Unbekannte. Die Topologie und die Form dieser Menge sollen optimiert werden, so dass im Optimum die aktive Menge der Zustandsbeschränkung erhalten wird.

Eine Methode, welche die aktive Menge bzw. deren Rand als Optimierungsvariable behandelt und iterativ bestimmt, entwickeln Hintermüller und Ring in [56]. Sie transformieren die notwendigen Bedingungen des Ausgangsproblems – unter Elimination von Variablen auf der aktiven Menge – in ein Problem der Shape-Optimierung und lösen es mit Level-Set-Methoden.

Eine weitere Methode entwickeln wir derzeit. Sie wird in Frey, Bechmann und Rund [38], sowie in Frey [37] erscheinen.

Die Betrachtungsweise von  $\mathcal{B}$  als Optimierungsvariable führt in unserer Arbeit [39, Section 7] auf eine neue Klasse von Problemstellungen, welche OC-PDE und Shapeoptimierung umschließt: Zu bestimmen ist gleichzeitig eine optimale Steuerung und eine Menge, so dass ein Zielfunktional unter stückweise gegebenen PDE-Nebenbedingungen minimiert wird. Wir möchten diese Problemstellung als Mengen-Optimalsteuerungsproblem bezeichnen:

**Definition 2.3.1 (Mengen-Optimalsteuerungsproblem, MOP)**

Als Mengen-Optimalsteuerungsproblem (MOP) bezeichnen wir das Problem

$$\begin{aligned}
\min J(y_{\mathcal{J}}, y_{\mathcal{B}}, u_{\mathcal{J}}, u_{\mathcal{B}}, \mathcal{B}) &:= \frac{1}{2} \|y_{\mathcal{J}} - y_{\text{d}}\|_{L^2(\mathcal{J})}^2 + \frac{1}{2} \|y_{\mathcal{B}} - y_{\text{d}}\|_{L^2(\mathcal{B})}^2 \\
&\quad + \frac{\lambda}{2} \|u_{\mathcal{J}}\|_{L^2(\mathcal{J})}^2 + \frac{\lambda}{2} \|u_{\mathcal{B}}\|_{L^2(\mathcal{B})}^2 \\
\text{s.t.} \quad & -\Delta y_{\mathcal{J}} + y_{\mathcal{J}} = u_{\mathcal{J}} && \text{in } \mathcal{J}, \\
& -\Delta y_{\mathcal{B}} + y_{\mathcal{B}} = u_{\mathcal{B}}, \quad u_{\mathcal{B}} = -\Delta y_{\text{max}} + y_{\text{max}} && \text{in } \overset{\circ}{\mathcal{B}}, \\
& \partial_{\nu} y_{\mathcal{J}} = 0 && \text{an } \Gamma, \\
& y_{\mathcal{J}} = y_{\mathcal{B}}, \quad 0 = \partial_{\nu_{\mathcal{J}}} y_{\mathcal{J}} + \partial_{\nu_{\mathcal{B}}} y_{\mathcal{B}} && \text{an } \beta, \quad (2.12) \\
& y_{\mathcal{B}} = y_{\text{max}} && \text{an } \beta, \\
& u_{\mathcal{J}} \in L^2(\mathcal{J}), \quad u_{\mathcal{B}} \in L^2(\mathcal{B}), \\
& y_{\mathcal{I}} < y_{\text{max}} && \text{in } \mathcal{J}, \\
& (\mathcal{B}, \mathcal{J}) \text{ ist eine Konfiguration.}
\end{aligned}$$

Wir sehen, dass alle Nebenbedingungen der Problemstellung berücksichtigt sind. Daher erhalten wir den Äquivalenzsatz [39, Theorem 7.1]:

**Lemma 2.3.2 (Äquivalenz von MOP und dem Ausgangsproblem)**

Das MOP und das Ausgangsproblem (2.1) besitzen dieselbe eindeutige Lösung.

*Beweis:*

*Jeder zulässige Punkt des MOP ist offensichtlich auch zulässig für das Ausgangsproblem (2.1). Weiterhin ist die Lösung des Ausgangsproblems auch zulässig für MOP. Da das Zielfunktional in beiden Problemen identisch ist, liefert ein einfacher Beweis durch Widerspruch das Resultat.  $\square$*

Wir müssen für diesen Beweis das MOP nicht einmal näher untersuchen.

Eine zugehörige notwendige Bedingung erreichen wir in [39, Formel (7.36)] anschließend formal über Techniken aus der Shape-Optimierung. Die erhaltene Bedingung liefert die Bedingung  $\bar{u}_{\mathcal{I}} = \bar{u}_{\mathcal{A}}$  an  $\gamma$ . Die damit verbundene Technik bezeichnen wir als „Shape-Lagrange-Technik“, siehe [37]. Vorteilhaft ist, dass die Shape-Lagrange-Technik keine Shape-Adjungierten benötigt.

Ein theoretisch gesichertes Vorgehen wird in Frey [37] anhand der totalen Shape-Ableitung einer Reformulierung des Problems untersucht, welches wir in [39, Sect. 7] als Bilevel-Optimierungsproblem einführen.

Wir gelangen auf diese Weise zum gleichen Ergebnis wie bei OC-ODE: Die Ableitung der Lagrange-Funktion nach der Schaltzeit in Gleichung (1.29) lieferte dort die Stetigkeit der Steuerung  $u$ .

## 2.4 Die Eliminationsmethode und die Ordnung von Zustandsbeschränkungen

Abschließend sollen zwei in Kap. 1 eingeführte Begriffe diskutiert werden. Beginnen wir mit der Eliminationsmethode:

### 2.4.1 Eliminationsmethode

Bei der Eliminationsmethode nutzt man die Zustandsbeschränkung um sukzessive Zustände und adjungierte Zustände (inklusive ihren Gleichungen) zu eliminieren. Im Falle des Prototyp-Beispiels ist die Zustandsbeschränkung sogar von der speziellen Form, dass sie alle Gleichungen und Variablen auf  $\mathcal{A}$  festlegt. Man kann in diesem Fall das Optimalitätssystem komplett auf die inaktive Menge reduzieren, ein freies Randwertproblem entsteht. In [56] wurde dieser Weg verfolgt. Wir behandeln diesen Weg auch in unserer anschließenden Arbeit über numerische Methoden [38]. Das freie Randwertproblem des Prototyp-Beispiels haben wir in [39, Sect. 7] notiert.

Bei Systemen mit mehreren Zustandsgrößen ist allerdings nicht mehr mit einer möglichen Reduktion auf  $\mathcal{I}$  zu rechnen, da im Allgemeinen nicht alle Variablen auf  $\mathcal{A}$  festgelegt werden. Wieviele Variablen eine skalare Zustandsbeschränkung festlegt, gab bei OC-ODE der Begriff der Ordnung an, welcher nun zu diskutieren ist.

### 2.4.2 Ordnung von Zustandsbeschränkungen bei OC-PDE

Bei OC-ODE gibt es den Begriff der Ordnung von Zustandsbeschränkungen (Def. 1.1.6). Eine treffende Definition für OC-PDE bedarf jedoch weitergehender Untersuchungen. Daher wird der Begriff im Folgenden lediglich anhand eines Beispiels erörtert.

Ist die Zustandsbeschränkung von höherer Ordnung, so muss das Vorgehen des BDD-Ansatzes (Ableiten und Einsetzen einer Zustandsgleichung) wiederholt werden, bis eine Steuerung auftritt. Diese strukturierte Methode gewinnt bei Zustandsbeschränkungen höherer Ordnung an Bedeutung, da man hier die Bestimmungsformel der Steuerung und des Multiplikators nicht einfach ablesen kann wie im Prototyp-Beispiel.

Da die Ordnung einer Zustandsbeschränkung nur von den Zustandsgleichungen und der Zustandsbeschränkung abhängt, beschränkt sich folgendes Beispiel auch nur auf diese Teile des OC-PDE:

#### Beispiel 2.4.1 (Zustandsbeschränkung höherer Ordnung)

Betrachten wir das System von Zustandsgleichungen

$$\begin{array}{llll} -\Delta y = -z + u, & -\Delta z = y, & -\Delta a = y + v, & \text{in } \Omega \\ y = 0, & z = 0, & a = 0 & \text{an } \Gamma. \end{array}$$

Die Steuerungen seien  $u, v \in L^2(\Omega)$ . Damit das System eindeutig lösbar ist, wurde es wie folgt aufgebaut: Die beiden PDE von  $(y, z)$  ähneln einem KKT-System eines Tracking-Problems einer Poissongleichung in  $y$  mit  $y_d = 0$ ,  $\lambda = 1$ , Steuerung  $-z$ , adjungiertem Zustand  $z$  und Quellterm  $u \in L^2(\Omega)$ . Bekanntlich besitzt dieses System für jedes  $u \in L^2(\Omega)$  genau eine Lösung. Wir betrachten die beiden PDE allerdings als Zustandsgleichungen. Darüber sitzt ein Randwertproblem für  $a$ . Für jedes  $v \in L^2(\Omega)$  und für jede Lösung  $y$  des Subsystems besitzt dieses nach Standardsätzen eine eindeutige Lösung  $a$ .

Betrachten wir nun verschiedene Zustandsbeschränkungen und ihre Einwirkung auf das System.

(a)  $z \leq z_{\max}$ ,  $z_{\max} \in H^8(\Omega)$ :

Die Zustandsbeschränkung  $z \leq z_{\max}$  ist bzgl. obiger Zustandsgleichungen von höherer Ordnung. Der Laplace-Operator und die Zustandsgleichungen sind zweimal anzuwenden, bis eine Steuerung in der Gleichung auftritt:

$$\begin{aligned} z = z_{\max} & \stackrel{-\Delta}{\Rightarrow} -\Delta z = -\Delta z_{\max}, & -\Delta z = y & \Rightarrow y = -\Delta z_{\max}, \\ & \stackrel{-\Delta}{\Rightarrow} -\Delta y = \Delta \Delta z_{\max}, & -\Delta y = -z + u & \Rightarrow -z + u = \Delta \Delta z_{\max}. \end{aligned}$$

Zur äquivalenten Umformung der Zustandsbeschränkung auf  $\mathcal{A}$  nach dem BDD-Ansatz bräuchte man hier 2 Randbedingungen am Interface  $\gamma$ , z.B. das Paar  $z = z_{\max}$ ,  $y = -\Delta z_{\max}$ .

Nach [15, S. 119] liefert die Ordnung einer Zustandsbeschränkung bei OC-ODE die Information, wieviele Zustände die Zustandsbeschränkung auf  $\mathcal{A}$  festlegt. Würde man die Definition einer Ordnung einer Zustandsbeschränkung bei OC-PDE im Falle von derartigen elliptischen Systemen 2. Ordnung auf dieser Grundlage aufbauen, dann hat dieses Beispiel Ordnung 2: Die Zustandsbeschränkung legt auf der aktiven Menge eine Steuerung ( $u$ ) sowie zwei Zustände ( $z$ ,  $y$ ) fest. Der dritte Zustand  $a$  ist durch Lösung einer PDE zu bestimmen.  $a$  und  $v$  sind auf der aktiven Menge aus verbliebenen KKT-Bedingungen numerisch zu bestimmen.

(b)  $y \leq y_{\max}$ ,  $y_{\max} \in H^4(\Omega)$ :

Die Zustandsbeschränkung wirkt sich hier ähnlich dem Prototyp-Problem: Einmaliges Ableiten und Einsetzen liefert die Bestimmungsformel  $u = -\Delta y_{\max} + z$ . Nur ein Zustand wird auf  $\mathcal{A}$  festgelegt. Einen Unterschied gekoppelter Systeme erkennen wir dennoch: Wir können zwar  $y$  und  $u$  durch Einsetzmethode eliminieren, kennen aber  $u$  selbst erst, nachdem  $z$  aus Lösung einer PDE auf  $\mathcal{A}$  bestimmt wurde.

**Bemerkung 2.4.2** Komplizierter wird die Definition einer Ordnung, wenn die Steuerungen und Zustände auf verschiedenen Gebieten vorliegen, z.B. bei Problemen mit ODE-PDE-Kopplung. Das Zusammenwirken ist dann sehr schwer erkennbar. In Kapitel 3 wird eine Formel für die Steuerung auf der aktiven Menge anhand eines Prototyp-Beispiels eines ODE-PDE gekoppelten Optimalsteuerungsproblems hergeleitet. Für komplexe Anwendungsprobleme wie das Brennstoffzellenmodell in Kapitel 4 ist es unklar, ob sich Identitäten für Steuerungen überhaupt angeben lassen. Am ehesten ist dies jedoch möglich mit einer strukturierten Methode, wie die der Bestimmung der Ordnung einer Zustandsbeschränkung.

**Bemerkung 2.4.3 (Strukturaussagen zur aktiven Menge)**

Bei OC-ODE ist man in der Lage, Struktursätze über die aktive Menge zu beweisen. Je nach Parität der Ordnung der Zustandsbeschränkung können hier entweder nur Berührungspunkte auftreten oder Berührungspunkte und Randstücke, siehe [61, S. 274]. Derartige Sätze sind auch bei OC-PDE wünschenswert, allerdings existieren sie derzeit nicht. Mittels Maximumprinzipien und Symmetrieaussagen erreichen wir in Kapitel 3 eine Strukturaussage bei einem Beispiel eines zeitabhängigen OC-PDAE mit Zustandsbeschränkungen. Allgemeinere Aussagen bleiben jedoch offen.

### 2.4.3 Herleitung von Maß-Darstellungen

Eng verbunden mit dem Begriff der Ordnung ist das Herleiten von Maß-Darstellungen: Die Vorgehensweise der Bestimmung der Ordnung einer Zustandsbeschränkung liefert eine Bestimmungsformel für die Steuerung auf  $\mathcal{A}$ . Führt man – nach Anwenden der Gradientengleichung – die gleiche Methodik ein zweites Mal aus, so erhält man eine Darstellung des Multiplikators  $\mu_{\mathcal{A}}$ . Greifen wir dazu obiges Beispiel auf und erweitern es auf ein OC-PDE:

#### Beispiel 2.4.4 (Maß-Darstellung bei höherer Ordnung)

$$\min J = \frac{1}{2} \|a - a_d\|_{L^2(\Omega)}^2 + \frac{\lambda_1}{2} \|u\|_{L^2(\Omega)}^2 + \frac{\lambda_2}{2} \|v\|_{L^2(\Omega)}^2 \quad \text{s.t.}$$

$$\begin{aligned} -\Delta y &= -z + u, & -\Delta z &= y, & -\Delta a &= y + v & \text{in } \Omega, \\ y &= 0, & z &= 0, & a &= 0 & \text{an } \Gamma, \end{aligned}$$

mit  $\lambda_i > 0$ ,  $a_d \in L^2(\Omega)$  und der Zustandsbeschränkung  $z \leq z_{\max}$  auf  $\bar{\Omega}$ ,  $z_{\max} \in H^8(\Omega)$ . Unter der Annahme der Existenz der Multiplikatoren  $\tilde{y}, \tilde{z}, \tilde{a}$  zu den PDE<sup>9</sup> und  $\mu \in \mathcal{M}(\Omega)$  zur Zustandsbeschränkung sowie  $z|_{\Gamma} < z_{\max}$  erhält man formal die adjungierten Gleichungen

$$\begin{aligned} -\Delta \tilde{y} &= \tilde{z} + \tilde{a}, & -\Delta \tilde{z} &= -\tilde{y} + \mu, & -\Delta \tilde{a} &= a - a_d & \text{in } \Omega, \\ \tilde{y} &= 0, & \tilde{z} &= 0, & \tilde{a} &= 0 & \text{an } \Gamma \end{aligned}$$

sowie die Gradientengleichungen  $\tilde{y} + \lambda_1 u = 0$ ,  $\tilde{a} + \lambda_2 v = 0$ . Für eine Konfiguration  $(\mathcal{A}, \mathcal{I})$  lässt sich auf  $\mathcal{A}$  eine Maß-Darstellung aus dem BDD-Ansatz bzw. dem Ergebnis der Ordnungsbestimmung  $u = \Delta \Delta z_{\max} + z$  gewinnen. Auf  $\mathcal{A}$  gilt:

$$\begin{aligned} & \tilde{y} = -\lambda_1 (\Delta \Delta z_{\max} + z), & & -\Delta \\ \implies & -\Delta \tilde{y} = \lambda_1 (\Delta^3 z_{\max} + \Delta z), & & \text{PDE einsetzen} \\ \implies & \tilde{z} + \tilde{a} = \lambda_1 (\Delta^3 z_{\max} - y), & & -\Delta \\ \implies & -\Delta \tilde{z} - \Delta \tilde{a} = -\lambda_1 (\Delta^4 z_{\max} - \Delta y), & & \text{PDE einsetzen} \\ \implies & -\tilde{y} + \mu_{\mathcal{A}} + a - a_d = -\lambda_1 (\Delta^4 z_{\max} - z + u). \end{aligned}$$

Mit dieser expliziten Formel kann  $\mu_{\mathcal{A}} := \mu|_{\mathcal{A}}$  eliminiert werden.

<sup>9</sup> Die Notation  $\tilde{y}$  für einen adjungierten Zustand zur Zustandsgleichung von  $y$  gewöhnen wir uns in Kapitel 4 an, da das dortige gekoppelte System 67 verschiedene adjungierte Zustände enthält.

## 2.5 Zusammenfassung

Um BDD-Ansätze auf OC-PDE übertragen zu können, haben wir die Problemstellung aufgeteilt auf die aktive und die inaktive Menge. Auf der aktiven Menge ist die Zustandsbeschränkung eine Gleichung, welche sich als Randwertproblem umformulieren lässt. Einsetzen der Zustandsgleichung lässt die Steuerung auftreten, wir erhalten eine Gleichungsnebenbedingung an die Steuerung im Inneren der aktiven Menge.

Das veränderte Problem erfüllt eine CQ, weshalb wir mit der Lagrange-Technik die notwendigen Bedingungen aufstellen können. Die letzte Gleichungsbedingung erhalten wir aus der partiellen Shape-Ableitung der Lagrange-Funktion nach dem Interface.

Wir erhalten damit alle notwendigen Bedingungen – außer den Vorzeichenbedingungen der Multiplikatoren der Zustandsbeschränkung, da wir diese rein als Gleichung behandelt hatten. Die notwendigen Bedingungen nach dem BDD-Ansatz sind damit zunächst schwächer als die etablierten notwendigen Bedingungen. Wir können die Vorzeichenbedingungen über einen Vergleich ergänzen.

Auf dem Wege treffen wir auf eine neue Problemstellung, welche Optimale Steuerung und Shapeoptimierung verbindet. Erste Untersuchungen sprechen dafür, dass sich dieses Mengen-Optimalsteuerungsproblem mit einer Shape-Lagrange-Technik behandeln lässt. Dies bietet den großen Vorteil, dass man die notwendigen Bedingungen unabhängig voneinander mittels partiellen Ableitungen der Lagrange-Funktion herleiten kann. Den Methoden zur Lösung der neuen notwendigen Bedingungen widmet sich unsere nächste Veröffentlichung [38].



## Kapitel 3

# Das Rocketcar - ein parabolisches OC-PDAE Problem mit Zustandsbeschränkungen

Das Problem des Rocketcars haben wir gemeinsam in einer Reihe von Veröffentlichungen behandelt: Pesch, Rund, von Wahl und Wendl [80], [79] sowie Wendl, Pesch und Rund [101]. Dabei ist der Artikel [80] noch nicht eingereicht, da er zusammen mit dem 2. Teil [81] erscheinen soll. Der 2. Teil wird die Ergebnisse dieser Dissertation enthalten – insbesondere die Maß-Darstellung sowie die Sprungbedingungen aus Abschnitt 3.4.

Die Abschnitte 3.2 und 3.5.2 wiederholen für die weitere Arbeit relevante Ergebnisse der einbezogenen Arbeiten [80], [79], [101] an den passenden Stellen. Darüberhinaus gibt Abschnitt 3.2.1 eine Zusammenfassung der einzelnen Arbeiten.

Allerdings sind weite Teile des Kapitels neu, unveröffentlicht und von mir – ohne Mithilfe meiner Kollegen – erarbeitet: Dazu gehören insbesondere die Abschnitte 3.3, 3.4, 3.6 und 3.7. Entsprechend werden sich in diesem Kapitel nur vereinzelt Verweise auf unsere gemeinsamen Arbeiten finden. Daher wurde das Kapitel so konzipiert, dass der Leser es unabhängig von den zugehörigen Veröffentlichungen lesen kann.

### 3.1 Einführung zu parabolischen OC-PDE mit Zustandsbeschränkungen

Notwendige Bedingungen für zustandsbeschränkte OC-PDE bei parabolischen Gleichungen liefern beispielsweise Casas in [24] oder Raymond und Zidani in [86]. So liefert Casas [24] die notwendigen Bedingungen für Randsteuerungsprobleme bei semilinearen parabolischen PDE mit Zustandsbeschränkungen. Die adjungierte PDE hat hier reguläre Borel-Maße als Quellterme; der Autor liefert einen Existenz- und Eindeutigkeitssatz für die adjungierte PDE mit Maßen und zeigt die Regularität des adjungierten Zustandes.

Das in Kapitel 1.2.2 angesprochene Problem, dass die Lagrange-Multiplikatoren der Zustandsbeschränkung im Allgemeinen nur reguläre Borel-Maße sind, beherrscht auch im parabolischen Fall die theoretische Analyse und die Numerik. Ohne  $L^\infty$ -Boxbeschränkungen an die Steuerung sind die Zustände im Allgemeinen nicht einmal stetig (außer bei verteilten Steuerungsproblemen in 1D), siehe Neitzel und Tröltzsch [73]. Daher be-

vorzuzugt man die Anwendung von Regularisierungsmethoden, wodurch die Maße aus der Theorie verschwinden und etablierte numerische Methoden anwendbar werden.

Im folgenden Kapitel wird ein Prototyp-Problem eines ODE-PDE beschränkten Optimalsteuerungsproblems mit Zustandsbeschränkungen behandelt, welches von der Struktur her sehr nahe an parabolischen OC-PDE liegt. Zunächst werden das Problem und seine notwendigen Bedingungen aufgestellt. Unter einer Strukturannahme an die aktive Menge lassen sich wie in den Kapiteln 1 und 2 detailliertere notwendige Bedingungen und Regularitätsaussagen herleiten.

Insbesondere werden die Sprungbedingungen von OC-ODE auf das gemischte ODE-PDE Problem übertragen. Der hierfür entscheidende Schritt ist eine explizite Darstellung des Maßes  $\mu \in \mathcal{M}(\bar{Q})$  aus handhabbaren Teilen. Dabei ist  $Q := \Omega \times (0, t_f)$  der Orts-Zeit-Zylinder. Das Ergebnis ist interessant: Bei dem behandelten Prototyp-Problem ist man in der Lage, die Tatsache aus Kapitel 2 bzw. Bergounioux und Kunisch [13] erstmalig auf eine niederdimensionale aktive Menge in Form einer Strecke im Inneren von  $Q$  zu übertragen: Der Multiplikator ist glatt im relativen Inneren der aktiven Menge und zeigt Dirac-Maße an den beiden Randpunkten. Diese Erkenntnis lässt sich zur weiteren theoretischen Analyse nutzen. Nach einem kurzen Überblick zur Lösung mit direkten Verfahren, widmet sich Abschnitt 3.6 den indirekten Verfahren. Analog zu OC-ODE lässt sich unter einer Strukturannahme an die aktive Menge ein Mehrpunkt-Anfangsrandwertproblem aufstellen. Das Kapitel schließt mit numerischen Experimenten.

## 3.2 Das Modell und seine verschiedenen Formulierungen

Das Modell des Raketenautos ist entstanden in Pesch, Rund, von Wahl und Wendl [80]. Unser Ziel war es, einen Prototyp eines OC-PDAE Problems zu erschaffen, welches die Gebiete OC-ODE, OC-PDE und OC-PDAE verbindet. Dadurch sind wir in der Lage, die Theorie und die Verfahren der verschiedenen Gebiete zu vergleichen und zu übertragen.

Ausgegangen sind wir von einem Problem der Flugbahn-Optimierung hypersonischer Flugzeuge aus Chudej, Pesch, Wächter et al. [30]. Die hohe Komplexität dieses Anwendungsproblems musste reduziert werden. So entschieden wir uns, die Bewegung auf eine Schiene einzuschränken (1D) und das Atmosphärenmodell einzusparen. Die Aufheizung sollte jedoch berücksichtigt werden, um die gekoppelte ODE-PDE Struktur zu erhalten.

Aufgrund seines Hintergrundes bezeichnen wir das resultierende Modell als das „Rocketcar“: Gesucht ist die Position  $w(t)$  eines Autos, welches sich auf einer Schiene ohne Kurven nach den Gesetzen der Newtonschen Mechanik bewegt. Dies führt auf gewöhnliche Differentialgleichungen. Dabei bezeichnen wir mit  $w(t)$  die Position des zur Punktmasse geschrumpften Raketenautos, d.h. die Position seines Schwerpunkts. Das Raketenauto heizt sich aufgrund der hohen Geschwindigkeit durch Reibung auf. Die Aufheizung des ausgedehnten Raketenautos modellieren wir anhand der Temperaturverteilung  $T(x, t)$  gemäß einer Wärmeleitungsgleichung, wobei wir mit  $x \in \Omega := (0, l)$  die Position innerhalb des Raketenautos bezeichnen. Die verschiedenen Reibungsmodelle (Coulomb, Stokes, Newton) liefern die Form der Kopplung des ODE-PDE Problems. Alle Größen des Systems seien bereits entdimensioniert.

Das Raketenauto soll anhand der Beschleunigung  $u(t)$  ausgehend von seiner Anfangsposition  $w_0$  und seiner Anfangsgeschwindigkeit  $\dot{w}_0$  zeitoptimal in den Ursprung der Phasenebene ( $w = 0, \dot{w} = 0$ ) gesteuert werden. Den Orts-Zeit-Zylinder bezeichnen wir mit  $Q := \Omega \times (0, t_f)$ . Wir erhalten ein Problem mit freier Endzeit  $t_f$ . OC-PDE mit freier Endzeit sind Thema der aktuellen Forschung, siehe z.B. Kunisch und Wachsmuth [64] für eine zeitoptimale Steuerung der Wellengleichung.

Dabei erzeugt die Beschleunigung  $u(t)$  Kosten, welche wir mit der  $L^2$ -Norm mit Kostenfaktor  $\lambda$  messen werden. Wir definieren nun die beiden Varianten des Optimalsteuerungsproblems, welche sich in der Art der Aufheizung unterscheiden: In „Problem 1“ wird das Auto auf seiner gesamten Länge aufgeheizt, in „Problem 2“ nur an seiner Spitze.

**Definition 3.2.1 (Rocketcar – Problem 1, aus [80])**

Wir betrachten das Optimalsteuerungsproblem

$$\min_{t_f > 0, u \in U_{\text{ad}}} J(t_f, u) := t_f + \frac{\lambda}{2} \int_0^{t_f} u^2 dt, \quad (3.1a)$$

$$\text{s. t.} \quad T_t - \Delta T = g(w_2(t)) \quad \text{in } Q := \Omega \times (0, t_f), \quad (3.1b)$$

$$\partial_\nu T + T = T_U \quad \text{an } \Sigma := \Gamma \times (0, t_f), \quad (3.1c)$$

$$T(x, 0) = T_0 \quad \text{in } \Omega, \quad (3.1d)$$

$$\dot{w}_1 = w_2, \quad \dot{w}_2 = u \quad \text{auf } (0, t_f), \quad (3.1e)$$

$$w_1(0) = w_0, \quad w_2(0) = \dot{w}_0, \quad (3.1f)$$

$$w_1(t_f) = 0, \quad w_2(t_f) = 0, \quad (3.1g)$$

$$T \leq T_{\text{max}} \quad \text{in } \bar{Q}, \quad (3.1h)$$

$$u \in U_{\text{ad}} := \{u \in L^2(0, t_f) \mid |u(t)| \leq u_{\text{max}}\}. \quad (3.1i)$$

Dabei ist  $\lambda > 0$ ,  $T_U$  die Umgebungstemperatur,  $\Gamma := \partial\Omega$ ,  $w_1 := w$ ,  $w_2 := \dot{w}$  und die Anfangsbedingungen erfüllen  $w_0^2 + \dot{w}_0^2 > 0$ . Die Ungleichungsrestriktionen seien gegeben durch  $T_{\text{max}} \in \mathbb{R}$ ,  $T_{\text{max}} > 0$  sowie  $u_{\text{max}} \in L^\infty(0, t_f)$ ,  $u_{\text{max}} > 0$ . Der Einfachheit halber betrachten wir in dieser Arbeit nur den Fall  $T_0 = 0$ ,  $T_U = 0$ .  $\nu$  bezeichnet den äußeren Einheitsnormalenvektor von  $\Omega$ .

**Definition 3.2.2 (Rocketcar – Problem 2, aus [80])**

Im zweiten Modell wird das Raketenauto nur an seiner Spitze in  $x = l$  durch Reibung aufgeheizt: Die Definition 3.2.1 wird übernommen. Lediglich die Gleichungen (3.1b), (3.1c) ändern sich zu

$$T_t - \Delta T = 0 \quad \text{in } Q, \quad (3.2a)$$

$$\partial_\nu T + T = \begin{cases} T_U & \text{in } \{0\} \times (0, t_f), \\ T_U + h(w_2(t)) & \text{in } \{1\} \times (0, t_f). \end{cases} \quad (3.2b)$$

Die Funktionen  $g$  und  $h$  modellieren die Reibung; daher müssen sie folgende Bedingungen erfüllen, formuliert für  $g$  ( $h$  analog):

$$g \in C^\infty(\mathbb{R} \setminus \{0\}) \cap C(\mathbb{R}), \quad (3.3a)$$

$$g \geq 0, \quad (3.3b)$$

$$g(z) = 0 \Leftrightarrow z = 0, \quad (3.3c)$$

$$g'(z) > 0, \forall z \in (0, \infty), \quad (3.3d)$$

$$g(z) = g(-z), \forall z \in \mathbb{R}. \quad (3.3e)$$

Dies berücksichtigt, dass die Reibung stets als Temperaturquelle fungiert (es sei denn es liegt keine Bewegung vor), dass die Reibung monoton mit dem Betrag der Geschwindigkeit ansteigt und unabhängig von der Richtung der Geschwindigkeit ist. Mögliche Reibungsmodelle bieten die Funktionen  $z \mapsto |z|^n$  für  $n = 1, 2, 3$ . Diese Bedingungen gehen in die Beweise der Abschnitte 3.3, 3.4 ein.

Problem 1 erinnert von der Struktur her an ein verteiltes Steuerungsproblem der Wärmeleitungsgleichung, Problem 2 an ein Randsteuerungsproblem; daher werden wir zur Unterscheidung diese Begriffe heranziehen. Wir konzentrieren uns in dieser Arbeit auf Problem 1. Für Problem 2 verweisen wir auf unsere Arbeiten [80], [81]. Das Problem ohne Steuerungskosten ( $\lambda = 0$ ) betrachten wir in [80].

Man beachte die Kopplungsstruktur beider Rocketcar-Probleme: Die ODE-Variablen gehen in die PDE ein. Die Zustandsbeschränkung schließt den Kreis: Die PDE-Variable beeinflusst hierüber die Steuerung und damit die ODE, wodurch ein vollständig gekoppeltes System entsteht.

### Bemerkung 3.2.3 (Rein zeitabhängige Steuerungen)

Es sei angemerkt, dass wir stets nur zeitabhängige Steuerungen haben. Diese Konstellation tritt aus technischer Sicht häufig auf, dass nur endlich viele zeitabhängige Steuersignale eingestellt werden dürfen. Unter anderem gilt dies auch für das PDAE-Modell der Brennstoffzelle in Kap. 4.

Mit in  $(x, t)$  verteilten aktiven Mengen wie in Kap. 2 ist aus der Dimension der Freiheitsgrade heraus daher i.d.R. nicht zu rechnen. Im vorliegenden Beispiel werden wir dies in Satz 3.3.7 auch beweisen.

### Umformulierung auf feste Endzeit

Zur numerischen Lösung des Problems müssen wir in der Regel auf feste Endzeit transformieren. Dazu führen wir eine neue Zeit  $\tau = t/t_f \in [0, 1]$  ein und suchen  $\tilde{u}(\tau) := u(t_f \tau)$  etc., wobei wir die Kennzeichnung  $\tilde{\cdot}$  im Folgenden aus Übersichtsgründen weglassen. Im Problem sind entsprechend alle Differentiale und Integrale zu substituieren. Problem 1

ist daher äquivalent zu dem Problem

$$\begin{aligned}
& \min_{u \in U_{\text{ad}}} J(y, u) := t_f + \frac{\lambda t_f}{2} \int_0^1 u^2(\tau) \, d\tau, \\
\text{s.t. } & \frac{1}{t_f} T_\tau(x, \tau) - \Delta T(x, \tau) = g(w_2(\tau)) && \text{in } \Omega \times (0, 1), \\
& \partial_\nu T(x, \tau) + T(x, \tau) = T_U && \text{an } \Gamma \times (0, 1), \\
& T(x, 0) = T_0 && \text{in } \Omega, \\
& \frac{1}{t_f} \dot{w}_1(\tau) = w_2(\tau), \quad w_1(0) = w_0, \quad w_1(1) = 0, \\
& \frac{1}{t_f} \dot{w}_2(\tau) = u(\tau), \quad w_2(0) = \dot{w}_0, \quad w_2(1) = 0, \\
& T(x, \tau) \leq T_{\text{max}} && \text{in } \bar{\Omega} \times [0, 1], \\
& u \in U_{\text{ad}} := \{u \in L^2(0, 1) \mid |u(\tau)| \leq u_{\text{max}}\}.
\end{aligned} \tag{3.4}$$

Die Transformation sorgt dafür, dass alle Differentialgleichungen Produkte von  $t_f$  und Zuständen bzw. der Steuerung enthalten (wenn wir mit  $t_f$  durchmultiplizieren). Das transformierte Problem ist an diesen Stellen nichtlinear, Probleme mit freier Endzeit sind daher stets iterativ zu lösen.

Interpretieren wir  $t_f$  als Steuerungsparameter, so beeinflussen wir einen Koeffizienten der PDE; allgemeinere Probleme dieser Art spielen z.B. bei der Steuerung bzw. Identifikation von Diffusionskoeffizienten eine Rolle [7], [33].

Wir können  $t_f$  aber auch als Zustand interpretieren, indem wir eine triviale ODE  $\frac{d}{d\tau} t_f = 0$  oder eine triviale PDE  $\frac{\partial}{\partial \tau} t_f = 0$  aufnehmen. In diesem Fall erhalten wir nichtlineare Zustandsgleichungen.

### Verbindung von OC-ODE, OC-PDE und OC-PDAE

Das Rocketcar-Modell erlaubt es, die Theorie und Methoden verschiedener Gebiete in Verbindung zu bringen und Bekanntes auf andere Gebiete zu übertragen. Grund dafür ist, dass sich das gemischte zustandsbeschränkte ODE-PDE auf drei verschiedene Varianten behandeln lässt: Zum einen kann man es als reines OC-ODE Problem behandeln, indem man die Lösung der PDE als Reihe über einen Separationsansatz bestimmt (semi-analytische Lösung). Dieses Vorgehen haben wir in [81] und [79] verfolgt.

Zum anderen lässt sich entsprechend die ODE durch Integration eliminieren. Bei dieser Variante erhalten wir ein reines OC-PDE Problem (siehe [81] und [101]). Bei beiden Vorgehensweisen erhält man ein anforderndes Problem, da die Elimination zu ausgefallenen Zusatzbedingungen führt: Im ersten Fall entstehen Integralgleichungen. Im zweiten Fall erhält man isoperimetrische Bedingungen sowie die Gegebenheit, dass neben der Steuerung zugleich auch ihre Stammfunktion im Problem auftritt.

Diese Arbeit verfolgt eine dritte, neue Variante: Das gemischte Problem wird direkt als Optimalsteuerungsproblem eines PDAE-Systems<sup>1</sup> behandelt. Abschnitt 3.4 liefert die notwendigen Bedingungen erster Ordnung für das PDAE-Problem, sowie eine detaillierte Analyse des Multiplikators der Zustandsbeschränkung. Im Abschnitt 3.3 davor werden einige inhärente Eigenschaften der Problemstellung bewiesen.

### 3.2.1 Inhalt der einbezogenen Veröffentlichungen

Ein Teil des Inhaltes der einbezogenen Veröffentlichungen wurde bereits in den vorigen Abschnitten zusammengefasst. Da einige Teile jedoch nicht aufgegriffen werden, liefert dieser Abschnitt eine Zusammenfassung jedes einzelnen Artikels.

#### Inhalt von [80]

Obwohl unser Artikel [80] noch nicht eingereicht ist, ist das Problem des Rocketcars dort entstanden. Wir beschreiben die Idee und die Problemstellung des Rocketcars, welches als Prototyp-Problem die Gebiete OC-ODE, OC-PDE und OC-PDAE verbinden soll. Als Grundlage lösen wir zunächst das Problem ohne Zustandsbeschränkung analytisch – jeweils für  $\lambda = 0$  (Bang-Bang-Lösungen) und für  $\lambda > 0$ . Wir lösen die PDE semianalytisch mit einem Separationsansatz unter Einsatz von Fourierreihen.

In einer Reihe von Theoremen zeigen wir die Existenz und Eindeutigkeit der Lösung der Zustandsgleichungen, die stetige Abhängigkeit von den Daten sowie diverse Eigenschaften der beiden Problemstellungen: Die Temperaturverteilung von Problem 1 erweist sich als nichtnegative, symmetrische klassische Lösung der Wärmeleitungsgleichung mit gewissen zusätzlichen Regularitätseigenschaften. Das globale Maximum der Temperatur liegt stets in der Wagenmitte  $x = l/2$  vor. Auch Problem 2 besitzt stets genau eine Lösung mit gewissen Regularitätseigenschaften. Das globale Maximum liegt hier für alle Steuerungen stets an der Spitze in  $x = l$ . Abschließend analysieren wir das Konvergenzverhalten der semianalytischen Lösungsformeln.

Die Veröffentlichung erscheint zusammen mit Teil 2 [81], welcher zum einen Ergebnisse der vorliegenden Dissertation und zum anderen ausführliche numerische Studien aus direkten Verfahren enthalten wird.

#### Inhalt von [79]

In [79] behandeln wir das Rocketcar als reines OC-ODE mit Zustandsbeschränkungen: Über einen Separationsansatz und Fourierreihen eliminieren wir die PDE durch eine semianalytische Lösungsformel. Das resultierende OC-ODE weist daraufhin allerdings eine nicht-lokale Zustandsbeschränkung auf. Daher führen wir einen dritten ODE-Zustand als Hilfsvariable ein. Er erfüllt eine Volterrasche Integro-DGL mit Reihenanteil und nunmehr eine punktweise Zustandsbeschränkung, weshalb sich die bestehende Theorie verwenden lässt. Die notwendigen Bedingungen enthalten ebenfalls eine derartige Integro-DGL, welche entsprechend rückwärts in der Zeit läuft.

---

<sup>1</sup> Die Literatur ist sich hier uneinig. Wir werden jedoch das ODE-PDE Problem auch als PDAE-System bezeichnen.

Die Zustandsbeschränkung ist von 2. Ordnung und kann daher bekanntlich zu Berührungspunkten und Randstücken führen. In der Tat beobachten wir numerisch sowohl einen Berührungspunkt als auch ein Randstück. Ergebnisse erzeugen wir mit der in Abschnitt 3.5.2 besprochenen direkten Methode – ohne die Lösungsformel zu verwenden, da die Optimalitätsbedingungen aufgrund der beiden Integro-DGL und der Reihenformel schwer zugänglich sind. Wir nutzen die hergeleiteten Optimalitätsbedingungen jedoch zur Verifikation der Lösung: Wir zeigen, dass die Schätzungen der Lagrange-Multiplikatoren aus dem direkten Verfahren die Optimalitätsbedingungen sehr gut erfüllen.

### Inhalt von [101]

In [101] eliminieren wir die beiden ODE durch Integration und behandeln das Rocketcar als reines OC-PDE. Die Randbedingungen der ODE führen zu zwei isoperimetrischen Nebenbedingungen an die Steuerung. Zusätzlich geht das Integral der Steuerung nicht-linear in die PDE ein, wodurch die Variationsungleichung sehr kompliziert wird: Über Integralausdrücke koppelt sie alle Werte der optimalen Steuerung vorwärts und rückwärts in der Zeit miteinander.

Wir schließen mit numerischen Ergebnissen aus dem direkten Verfahren. Insbesondere rekonstruieren wir den adjungierten Zustand der PDE aus seinen diskreten Schätzungen. Man kann aus den Ergebnissen bereits einige der Eigenschaften erahnen, welche im Zuge der vorliegenden Dissertation in Abschnitt 3.4 bewiesen werden. So erkennt man zwei Fundamentallösungen an den beiden Schaltpunkten und eine Knickbedingung im relativen Inneren der aktiven Menge.

## 3.3 Theoretische Behandlung der Problemstellung

Dieser Abschnitt umfasst einen Existenz- und Eindeutigkeitssatz für die Zustandsgleichungen sowie Regularitätsaussagen über die Zustände des Problems. Um Eindeutigkeit für eine feste Steuerung zu erreichen, sind die Endbedingungen (3.1g) dabei aus der Betrachtung zu nehmen, da das Zustandssystem andernfalls überbestimmt ist. Die Endbedingungen werden durch spezielle Wahl von  $u \in U_{\text{ad}} \subset L^\infty(0, t_f)$  und  $t_f$  im Laufe der Optimierung erfüllt; ihrer Erfüllbarkeit widmet sich Lemma 3.3.2. Dazu sei angemerkt, dass wir aufgrund der Steuerungsbeschränkungen (3.1i) mit  $u_{\text{max}} \in L^\infty(0, t_f)$  entsprechend stets  $u \in L^\infty(0, t_f)$  haben. Diese Eigenschaft ist wesentlich, um später die Regularität der Nichtlinearität  $g$  zeigen zu können.

### Lemma 3.3.1 (Existenz- und Eindeutigkeitssatz der Zustandsgleichungen)

Die Zustandsgleichungen (3.1b)–(3.1f) besitzen für jede Steuerung  $u \in L^\infty(0, t_f)$  eine eindeutige Lösung

$$(w_1, w_2, T) \in W^{2,\infty}(0, t_f) \times W^{1,\infty}(0, t_f) \times (W(0, t_f) \cap C(\bar{Q})) .$$

*Beweis:*

Man betrachtet zuerst das Anfangswertproblem  $\dot{w}_2(t) = u(t)$ ,  $w_2(0) = \dot{w}_0$  (d.h. unter Weglassen der Endbedingung). Nach wohlbekanntem Sätzen der Theorie gewöhnlicher Differentialgleichungen besitzt das Anfangswertproblem für jede rechte Seite  $u \in$

$L^\infty(0, t_f)$  genau eine Lösung  $w_2 \in W^{1,\infty}(0, t_f) \hookrightarrow C([0, t_f])$ . Aufgrund der Regularität der Nichtlinearität (3.3a) folgt entsprechend  $g(w_2(t)) \in L^\infty(0, t_f)$ . Es liegt also ein Anfangs-Randwertproblem der linearen Wärmeleitungsgleichung mit einer verteilten Quelle  $f(x, t) := g(w_2(t)) \in L^\infty(Q)$  vor. Daher liefert [73, Theorem 1.3] die Existenz und Eindeutigkeit der Temperaturverteilung  $T(x, t) \in W(0, t_f) \cap C(\bar{Q})$ .<sup>2</sup> Weiterhin erhält man dort stetige Abhängigkeit von den Daten. Das Anfangswertproblem der zweiten ODE  $\dot{w}_1 = w_2$ ,  $w_1(0) = w_0$  löst man analog dem der ersten.  $\square$

Damit das Optimalsteuerungsproblem eine Lösung hat, muss die Endzeit eine Mindestgröße haben, da sonst die Endbedingungen nicht eingehalten werden können. Man kann eine Abschätzung explizit angeben.

**Lemma 3.3.2** Es gilt  $t_f \geq t_{f,0} > 0$  mit

$$t_{f,0} := \begin{cases} -\dot{w}_0 + 2\sqrt{\frac{1}{2}\dot{w}_0^2 - w_0}, & w_0 < -\text{sign}(\dot{w}_0) \cdot \frac{1}{2}\dot{w}_0^2, \\ \dot{w}_0 + 2\sqrt{\frac{1}{2}\dot{w}_0^2 + w_0}, & w_0 > -\text{sign}(\dot{w}_0) \cdot \frac{1}{2}\dot{w}_0^2. \end{cases}$$

*Beweis:*

$t_{f,0}$  ist die Lösung für  $\lambda = 0$  ohne Zustandsbeschränkung aus Pesch [77, Übungsaufgabe 6b]. In weniger Zeit kann man die Endbedingungen (den Ursprung der Phasenebene) mit maximaler Geschwindigkeit nicht erreichen, es gibt für kleinere Endzeiten keinen zulässigen Punkt.  $\square$

Die Endzeit  $t_{f,0}$  muss nicht zu zulässigen Punkten führen, da die Zustandsbeschränkung verletzt sein kann; jedoch kann es für Endzeiten  $t_f < t_{f,0}$  keinen zulässigen Punkt geben. Die Formel lässt sich später in der Numerik nutzen, um unzulässige Startschätzungen für  $t_f$  auszusortieren. In der Theorie setzen wir die Existenz einer Lösung des Optimalsteuerungsproblems voraus. Eine Beschränkung an die Endzeit ist nicht nötig.

Die Zustände zeigen eine höhere Regularität als Lemma 3.3.1 erwarten lässt. Um dies zu zeigen, müssen wir uns im differenzierbaren Bereich des Reibungsmodells  $g(z)$  befinden. Daher die

**Annahme 3.3.3** Die Geschwindigkeitsverteilung  $w_2(t)$  besitze in  $(0, t_f)$  nur endlich viele (isolierte) Nullstellen.

Ohne Zustandsbeschränkung haben wir in [80] gezeigt, dass die optimale Geschwindigkeitsverteilung  $\bar{w}_2(t)$  in  $(0, t_f)$  maximal eine Nullstelle hat; diese tritt als Umkehrpunkt auf, wenn die Anfangsbedingungen einen einmaligen Richtungswechsel erfordern, d.h. wenn das Raketenauto z.B. zu Beginn in die falsche Richtung fährt. Mit Zustandsbeschränkung beobachten wir numerisch das gleiche Verhalten (siehe [79], [81]); die Annahme ist daher sinnvoll. Mit der Annahme lässt sich höhere Regularität beweisen:

<sup>2</sup> Die Definition von  $W(0, t_f) := \{u \in L^2(0, t_f; V) \mid u' \in L^2(0, t_f; V^*)\}$  findet man ausführlich in [96], [87]. Im vorliegenden Fall ist  $V = H^1(\Omega)$ .

**Satz 3.3.4 (Höhere Regularität der Temperatur)**

Erfüllt die Geschwindigkeitsverteilung  $w_2(t)$  die Annahme 3.3.3, so besitzt die Temperaturverteilung die Darstellung

$$T(x, t) = T_1(t) + T_2(x, t)$$

mit den Regularitäten  $T_1(t) \in C^1([0, t_f])$  bzw.  $T_2(x, t) \in C^\infty(\Omega \times (0, t_f])$ .

*Beweis:*

Unter der Annahme 3.3.3 und der Regularität von  $g$  (3.3a) folgt entsprechend aus  $w_2 \in W^{1,\infty}(0, t_f)$  jetzt auch  $g(w_2(t)) \in W^{1,\infty}(0, t_f)$ . Weiterhin ist die Tatsache auszunutzen, dass der Quellterm der Wärmeleitungsgleichung konstant im Ort ist. Aufgrund der Linearität der Zustandsgleichung kann man die Superposition  $T = T_1 + T_2$  betrachten mit

$$\begin{aligned} T_{1,t} - \Delta T_1 &= g(w_2(t)), & T_{2,t} - \Delta T_2 &= 0 & \text{in } Q, \\ \partial_\nu T_1 &= 0, & \partial_\nu T_2 + T_2 &= -T_1 & \text{an } \Sigma, \\ T_1(x, 0) &= 0, & T_2(x, 0) &= 0 & \text{in } \Omega. \end{aligned}$$

Durch diese Konstruktion lässt sich  $T_1$  analytisch angeben: Es gilt  $T_1(x, t) = z(t)$  mit  $z(t) \in W^{2,\infty}(0, t_f)$  aus  $\dot{z}(t) = g(\bar{w}_2(t))$ ,  $z(0) = 0$ . Wegen Sobolew-Einbettungssätzen [5, Satz 8.9] gilt  $W^{2,\infty}(0, t_f) \hookrightarrow C^1([0, t_f])$ .

Kommen wir zu  $T_2$ : Für die homogene lineare Wärmeleitungsgleichung gilt bekanntlich sehr hohe Regularität im Inneren, also insbesondere  $T_2 \in C^\infty(\Omega \times (0, t_f])$  (z.B. Evans [35, Theorem 8, S. 59]).

□

Im Ort zeigt sich die Temperaturverteilung aufgrund der speziellen Quelle also als beliebig glatt. Die Superposition aus dem Satz ist wichtig, sie wird in späteren Beweisen oft aufgegriffen. Weitere Regularitätsresultate liefert uns von Wahl in [80, Theoreme 4.7 und 4.10]. Widmen wir uns nun dem Optimalsteuerungsproblem.

**Annahme 3.3.5** Wir nehmen an, dass das Optimalsteuerungsproblem (3.1) eine Lösung  $(\bar{w}_1, \bar{w}_2, \bar{T}, \bar{u}, \bar{t}_f)$  besitzt, bei welcher die Zustandsbeschränkung (3.1h) nicht überall strikt erfüllt ist.

**Definition 3.3.6 (Aktive Menge der Zustandsbeschränkung)**

Die aktive Menge der Zustandsbeschränkung (3.1h) bezeichnen wir mit

$$\tilde{\mathcal{A}} := \{(x, t) \in \bar{Q} \mid \bar{T}(x, t) = T_{\max}\}.$$

Die Symmetrie des Problems sorgt dafür, dass die aktive Menge eine spezielle Form hat. Wir notieren den

**Satz 3.3.7** Die aktive Menge  $\tilde{\mathcal{A}}$  der Zustandsbeschränkung (3.1h) in Problem 1 hat die Form

$$\tilde{\mathcal{A}} \subset \{l/2\} \times [t_1, t_2],$$

mit  $t_1 > 0$  und  $t_2 < t_f$ .

*Beweis:*

Nach Satz 3.3.4 und [80, Theorem 4.7] ist die Temperaturverteilung  $T$  sehr regulär. [80, Theorem 4.6] liefert daher für jede Steuerung  $u \in L^\infty(0, t_f)$

$$\arg \max_{x \in \Omega} T(x, t) = \frac{l}{2} \quad \forall t \in (0, t_f).$$

Weiterhin gilt nach dortigem Theorem 4.4 dass  $T(x, t) > T(x, 0) = 0$  auf  $[0, l] \times (0, t_f]$ . Wegen der Endbedingung  $w_2(t_f) = 0$  und der Tatsache, dass  $w_2 \in H^1(0, t_f) \hookrightarrow C([0, t_f])$ , folgt

$$\max_Q T(x, t) = \max_{t \in [t_1, t_2]} T\left(\frac{l}{2}, t\right),$$

für  $0 < t_1 < t_2 < t_f$ . Da  $T_{\max} \equiv \text{const} > 0$  ist, gilt  $\mathcal{A} = \{(x, t) \in Q \mid \bar{T}(x, t) = \max_Q \bar{T}(x, t)\} \subset \{l/2\} \times [t_1, t_2]$ .  $\square$

Die Zustandsbeschränkung kann also maximal auf einer finiten Linie im Inneren des Orts-Zeit-Zylinders aktiv werden. In  $(x, t)$  verteilte aktive Mengen treten nicht auf. Die aktive Menge lässt sich in Folge von Satz 3.3.7 allein durch die Zeit charakterisieren. Wir definieren daher:

**Definition 3.3.8 (Aktive Zeiten der Zustandsbeschränkung)**

Die aktiven Zeiten der Zustandsbeschränkung bezeichnen wir mit

$$\mathcal{A} := \left\{ t \in (0, t_f) \subset \mathbb{R} \mid \bar{T}\left(\frac{l}{2}, t\right) = T_{\max} \right\}.$$

Die Tatsache, dass die Zustandsbeschränkung weder am Anfang noch am Ende aktiv ist, wird häufig in den Beweisen benötigt. Daher definieren wir den ersten und letzten aktiven Zeitpunkt:

**Definition 3.3.9 (Erster und letzter aktiver Zeitpunkt)**

Wir bezeichnen den ersten aktiven Zeitpunkt als  $t_E$ , den letzten als  $t_L$ , d.h.

$$\mathcal{A} \subset [t_E, t_L] \quad \cap \quad \{t_E, t_L\} \subset \partial \mathcal{A}.$$

Da das Maximum der Temperatur für jede Steuerung  $u \in L^\infty(0, t_f)$  stets in der Mitte liegt, reicht es die Zustandsbeschränkung auf der mittleren Linie zu fordern:

**Folgerung 3.3.10** Die Zustandsbeschränkung bei Problem 1 lässt sich äquivalent umformulieren:

$$T \leq T_{\max} \text{ auf } Q \quad \Leftrightarrow \quad T\left(\frac{l}{2}, t\right) \leq T_{\max}\left(\frac{l}{2}, t\right) \text{ auf } [0, t_f].$$

Die Aussagen dieses Abschnittes setzen direkt auf der Problemstellung auf und gelten daher für alle drei erwähnten Herangehensweisen. Der nächste Abschnitt widmet sich nun speziell der PDAE-Formulierung des Rocketcars.

### 3.4 Theoretische Behandlung als OC-PDAE Problem

In [79] haben wir das Rocketcar als reines OC-ODE behandelt, in [101] dagegen als reines OC-PDE. Als dritte Alternative wird das OC-PDAE im Folgenden erstmalig direkt behandelt. Es werden notwendige Bedingungen hergeleitet und unter einer Strukturannahme an die aktive Menge werden die entscheidenden Sätze 3.4.9 und 3.4.13 erreicht, welche uns eine Maß-Darstellung sowie die Sprungbedingungen liefern. In der Folge lässt sich eine höhere Regularität der Lösung des OC-PDAE zeigen.

Notwendige Bedingungen für parabolische OC-PDE mit Zustandsbeschränkungen findet man z.B. in [24, 86]. Da wir aber ein ODE-PDE gekoppeltes Problem haben, sind die notwendigen Bedingungen selbst herzuleiten; wir wählen hierzu die formale Lagrange-Technik (siehe z.B. [96]). Mit ihr lassen sich sogar bei komplexen Anwendungsproblemen – wie dem Brennstoffzellenmodell aus Kapitel 4 – die notwendigen Bedingungen effizient aufstellen. Ihre Anwendbarkeit wird man dann i.d.R. nicht beweisen können. Wir stellen daher fortan folgende Annahme:

**Annahme 3.4.1** Das Rocketcar-Problem aus Definition 3.2.1 besitze eine Lösung. Die Lagrange-Multiplikatoren  $p_1(t), p_2(t)$  zu den ODE und  $q(x, t)$  existieren und besitzen die nötige Regularität. Der Lagrange-Multiplikator  $\mu_{\bar{Q}} \in \mathcal{M}(\bar{Q})$  zu der Zustandsbeschränkung existiere.

$\mu_{\bar{Q}}$  ist also zunächst erneut ein reguläres Borel-Maß auf  $\bar{Q}$ . Allerdings erlaubt das Problem eine weitergehende Analyse und Spezifikation von  $\mu_{\bar{Q}}$ , welche später durchgeführt wird.

Die Existenz mehrerer lokal optimaler Lösungen können wir nicht ausschließen, da es sich um ein nichtlineares, nicht-konvexes Problem handelt. Numerisch beobachten wir allerdings stets eine eindeutige Lösung. Jede Lösung, welche obige Annahme erfüllt, muss Lösung der im Folgenden hergeleiteten notwendigen Bedingungen sein.

#### 3.4.1 Notwendige Bedingungen

Die Lagrange-Funktion für das Rocketcar-Problem aus Definition 3.2.1 lautet unter Aufnahme der Endbedingungen mit  $\alpha, \beta \in \mathbb{R}$  (vgl. Kap. 1.2.1)

$$\begin{aligned} \mathcal{L} = & J(y, u) - \iint_{\bar{Q}} (T_t - \Delta T - g(w_2(t))) q \, dx \, dt - \int_{\Sigma} (\partial_\nu T + T - T_U) q \, ds \, dt \\ & - \int_0^{t_f} (\dot{w}_1 - w_2) p_1 \, dt - \int_0^{t_f} (\dot{w}_2 - u) p_2 \, dt + \int_{\bar{Q}} (T(x, t) - T_{\max}) \, d\mu_{\bar{Q}} \\ & + \alpha w_1(t_f) + \beta w_2(t_f). \end{aligned} \quad (3.5)$$

Für die Steuerungsbeschränkungen werden dagegen keine Lagrange-Multiplikatoren eingeführt. Sie werden stattdessen wie üblich mit der Forderung  $u \in U_{\text{ad}}$  mit der konvexen zulässigen Menge der Steuerungen  $U_{\text{ad}}$  berücksichtigt. Die Theorie der Optimierung in Banach-Räumen kann mit derartigen konvexen Bedingungen an die Freiheitsgrade bekanntlich sehr geschickt umgehen; die Herleitung der notwendigen Bedingungen verkürzt sich.

**Bemerkung 3.4.2** Aufgrund von Folgerung 3.3.10 könnte man alternativ  $\int_0^{t_f} T(\frac{l}{2}, t) - T_{\max} d\mu(t)$  an die Lagrange-Funktion ankoppeln und hätte dadurch zu Beginn bereits strukturelle Aussagen über den Lagrange-Multiplikator. Wir bleiben vorerst bei der allgemeinen Version und präzisieren die Struktur des Multiplikators im Anschluss an die Herleitung der notwendigen Bedingungen.

Die notwendigen Bedingungen erhält man durch Bilden der Richtungsableitungen dieser Lagrange-Funktion mit anschließender partieller Integration. Wir behandeln hier nur die ausgefallenste Gleichung, nämlich diejenige adjungierte Gleichung, welche durch Richtungsableitung nach  $w_2$  entsteht:

$$\begin{aligned} 0 &= D_{w_2} \mathcal{L} \varphi = \iint_Q g'(w_2(t)) q \varphi dx dt + \int_0^{t_f} p_1 \varphi dt - \int_0^{t_f} p_2 \dot{\varphi} dt \\ &= \int_0^{t_f} \left[ p_1 + \dot{p}_2 + g'(w_2(t)) \int_{\Omega} q(x, t) dx \right] \varphi(t) dt, \quad \forall \varphi \in W_0^{1, \infty}(0, t_f). \end{aligned} \quad (3.6)$$

Die Randterme der partiellen Integration verschwinden, da die Testfunktion  $\varphi(0) = \varphi(t_f) = 0$  erfüllen muss, um die Randbedingungen von  $w_2$  einzuhalten. Die erhaltene Gleichung zeigt bereits eine typische Eigenschaft von ODE-PDE gekoppelten Problemen: Funktionen einer und mehrerer Veränderlicher sind miteinander gekoppelt über konstante Quellen (hier in der Zustandsgleichung von  $T$ ) oder als Gegenstück über Parameterintegrale (in obiger adjungierter Gleichung). Im Kapitel 4 werden wir weitere Effekte bei den Optimalitätsbedingungen des komplexen gekoppelten PDAE-Systems einer Brennstoffzelle finden. Insgesamt lauten die adjungierten Gleichungen

$$-q_t - \Delta q = \mu_Q \quad \text{in } Q, \quad (3.7a)$$

$$\partial_\nu q + q = \mu_\Sigma \quad \text{an } \Sigma, \quad (3.7b)$$

$$q(x, t_f) = \mu_f \quad \text{in } \Omega, \quad (3.7c)$$

$$-\dot{p}_1 = 0 \quad \text{auf } (0, t_f), \quad (3.7d)$$

$$-\dot{p}_2 = p_1 + g'(w_2(t)) \int_{\Omega} q(x, t) dx \quad \text{auf } (0, t_f), \quad (3.7e)$$

wobei  $\mu_Q := \mu_{\bar{Q}}|_Q$ ,  $\mu_\Sigma := \mu_{\bar{Q}}|_\Sigma$ ,  $\mu_f := \mu_{\bar{Q}}|_{t=t_f}$  Restriktionen des Maßes auf die entsprechenden Mengen sind. Lemma 3.4.4 sichert demnächst, dass  $\mu_\Sigma = 0$  und  $\mu_f = 0$  gilt. Zusätzlich ergeben sich  $\alpha = p_1(t_f)$  sowie  $\beta = p_2(t_f)$ . Die PDE ist im distributionellen Sinne zu verstehen; gemeint ist also ihre schwache Form

$$\begin{aligned} \iint_Q q \varphi_t dx dt + \iint_Q \nabla q \cdot \nabla \varphi dx dt + \iint_\Sigma q \varphi ds dt = \iint_Q \varphi(x, t) d\mu_{\bar{Q}} \quad (3.8) \\ \forall \varphi \in Y_0 \cap C^1(\bar{Q}), \end{aligned}$$

mit  $Y_0 := \{y \in C(\bar{Q}) \cap L^2(0, t_f; H^1(\Omega)) \mid y(x, 0) = 0 \forall x \in \Omega\}$ .

Die adjungierte PDE stimmt mit Gleichung (6.1) aus Casas [24] sowie Definition 3.4 aus Raymond und Zidani [86] überein. Die adjungierte Gleichung wird eindeutig lösbar, wenn wir zusätzlich die Formel der partiellen Integration bzw. die Greensche Formel fordern (vgl. [96, S. 284], [4]); Theorem 6.3 aus [24] liefert:

**Lemma 3.4.3 (Existenz- und Eindeigkeitsatz der adjungierten PDE)**

Sei  $t_f > 0$  fest. Dann besitzt die adjungierte PDE (3.8) eine eindeutige Lösung  $q \in L^r(0, t_f; W^{1,m}(\Omega))$  für alle  $r, m \in [1, 2)$  mit  $2/r + 1/m > 2$ , die zusätzlich die Gleichung

$$\iint_{\bar{Q}} (\varphi_t - \Delta \varphi) q \, dx \, dt + \iint_{\Sigma} (\partial_\nu \varphi + \varphi) q \, ds \, dt = \iint_{\bar{Q}} \varphi(x, t) \, d\mu_{\bar{Q}}, \quad \forall \varphi \in Y_0^\infty, \quad (3.9)$$

erfüllt. Dabei ist  $Y_0^\infty := \{y \in Y_0 \mid y_t - \Delta y \in L^\infty(Q), \quad \partial_\nu y + y \in L^\infty(\Sigma)\}$ .

Wir werden später sehen, dass die Regularität von  $q$  fast überall wesentlich höher ist. Wichtig ist, dass man für die adjungierte ODE keinerlei Randbedingungen erhält: Da die Zustände bereits an beiden Rändern festgelegt sind, sind die adjungierten Zustände hier frei. Die Endbedingungen erzeugen also erneut ein KKT-System mit asymmetrischer Verteilung der Randbedingungen (vgl. Bem. 1.1.5).

An die Steuerung sind konvexe Mengenbeschränkungen  $u \in U_{\text{ad}}$  gestellt. Die Richtungsableitung nach  $u$  darf in keine der zulässigen Richtungen negativ sein. Dies liefert die Variationsungleichung

$$\int_0^{t_f} (p_2 + \lambda u)(u - \bar{u}) \, dt \geq 0 \quad \forall u \in U_{\text{ad}}. \quad (3.10)$$

Sie resultiert punktweise in der üblichen Projektionsbeziehung  $\bar{u} = P_{U_{\text{ad}}}(-\frac{1}{\lambda} p_2)$  mit  $P_{U_{\text{ad}}}(f) := \min(u_{\text{max}}, \max(-u_{\text{max}}, f))$ , vgl. [96].

Für die Ableitung nach der freien Endzeit gehen wir analog zu Formel (1.20) vor. Hier erhält man unter Vorgriff auf die als nächstes erbrachten Erkenntnisse  $\mu_f = 0$  (Lemma 3.4.4), sowie  $q \equiv 0$  auf  $\Omega \times (t_L, t_f]$  (Folg. 3.4.5) die Bedingung zur freien Endzeit

$$\begin{aligned} 0 &= D_{t_f} \mathcal{L} = 1 + \frac{\lambda}{2} u^2(t_f) + \dot{w}_1(t_f) \underbrace{(\alpha - p_1(t_f))}_{=0} + \dot{w}_2(t_f) \underbrace{(\beta - p_2(t_f))}_{=0} + \left[ \underbrace{w_2}_{=0} p_2 + u p_2 \right]_{t_f} \\ &= 1 + \frac{\lambda}{2} u^2(t_f) + u(t_f) p_2(t_f), \end{aligned} \quad (3.11)$$

welche mit der Gleichung Hamilton-Funktion  $H(t_f) = 0$  aus [80] und der Theorie von OC-ODE mit freien Endzeiten nach Abschnitt 1.2.1 übereinstimmt. Im Allgemeinen wird diese Bedingung bei OC-PDAE Ortsintegrale über Differentialausdrücke verteilter Größen enthalten, wie es später bei den Bedingungen an die Schaltzeiten in (3.30) der Fall ist. Beim Rocketcar treten diese in  $t_f$  nicht auf, da die Zustandsbeschränkung zur Endzeit nicht aktiv ist. Dadurch entfällt einseitig die Kopplung: Für  $t > t_L$  spielt die Temperaturverteilung keine Rolle für die Optimierung.

Daneben gilt die Komplementaritätsbedingung

$$\mu_{\bar{Q}} \geq 0 \text{ in } \bar{Q}, \quad \iint_{\bar{Q}} [T(x, t) - T_{\text{max}}] \, d\mu_{\bar{Q}} = 0. \quad (3.12)$$

Die Gleichungen (3.1b)-(3.1h), (3.7), (3.10), (3.11), (3.12) formen die notwendigen Bedingungen 1. Ordnung, die KKT-Bedingungen.

Wir nutzen nun die Komplementaritätsbedingung, um die erste Strukturaussage des Lagrange-Multiplikators  $\mu_{\bar{Q}}$  zu erhalten. Wie bei elliptischen OC-PDE ([13]) gilt für parabolische OC-PDE und auch für das gemischte Problem des Rocketcars bekanntlich die Tatsache, dass der Lagrange-Multiplikator der Zustandsbeschränkung auf die aktive Menge konzentriert ist:

**Lemma 3.4.4** Es gilt  $\mu_{\bar{Q}}|_{\tilde{\mathcal{I}}} = 0$  mit der inaktiven Menge  $\tilde{\mathcal{I}} := Q \setminus \tilde{\mathcal{A}}$ .

**Folgerung 3.4.5** ( $q \equiv 0$  auf  $(t_L, t_f]$ )

Da  $\mu_{\bar{Q}}$  die einzige Inhomogenität des rückwärts gerichteten, linearen Anfangsrandwertproblems von  $q$  ist, ist  $q \equiv 0$  auf  $\bar{\Omega} \times (t_L, t_f]$ .

**Lemma 3.4.6** Es gilt die Maß-Zerlegung

$$\mu_{\bar{Q}} = \mu(t) \cdot \delta\left(x - \frac{l}{2}\right),$$

mit  $\mu(t) \in \mathcal{M}([t_E, t_L])$  und dem Dirac-Maß  $\delta(x - \frac{l}{2}) \in \mathcal{M}(\Omega)$  konzentriert in  $x = \frac{l}{2}$ .

*Beweis:*

Wir definieren  $\mu(t) := \mu_{\bar{Q}}|_{\{l/2\} \times [t_E, t_L]}$ ,  $\mu(t) \in \mathcal{M}([t_E, t_L])$ . Dann betrachten wir für ein festes  $y \in C(\bar{Q})$  das Dualitätsprodukt

$$\begin{aligned} \langle \mu_{\bar{Q}}, y \rangle_{C^*(\bar{Q}), C(\bar{Q})} &= \iint_{\bar{Q}} y \, d\mu_{\bar{Q}} \stackrel{\text{Lemma 3.4.4}}{=} \iint_{\tilde{\mathcal{A}}} y \, d\mu_{\bar{Q}} \stackrel{\text{Satz 3.3.7}}{=} \int_{\mathcal{A}} y\left(\frac{l}{2}, t\right) \, d\mu(t) \\ &= \iint_{\bar{Q}} y(x, t) \cdot \delta\left(x - \frac{l}{2}\right) \, dx \, d\mu(t) \quad \forall y \in C(\bar{Q}). \end{aligned}$$

Dies folgt wegen der Komplementaritätsbedingung und der Ausblendeigenschaft des Dirac-Impulses  $\delta(x - x_0)$  (z.B. [69]).  $\square$

Das Linienmaß  $\mu(t)$  soll weiter aufgeschlüsselt werden, um Übergangs- und Sprungbedingungen herzuleiten und die Regularitätsaussagen zu verbessern. Dazu benötigt man eine Strukturannahme an die aktive Menge  $\mathcal{A}$ . Greifen wir auf die bei OC-ODE übliche Annahme zurück:

**Annahme 3.4.7** Es gelte Annahme 1.1.3, d.h.  $\mathcal{A}$  bestehe nur aus endlich vielen, disjunkten Berührungspunkten  $t_c^j$  und Randstücken  $[t_{\text{on}}^i, t_{\text{off}}^i]$ . Weiterhin gelte  $|\bar{u}| < u_{\text{max}}$  auf  $\mathcal{A}$ , d.h. die aktiven Mengen der Steuerungsbeschränkungen und der Zustandsbeschränkung seien disjunkt.

Unsere numerischen Ergebnisse anhand direkter Verfahren aus [79], [81] bestätigen, dass diese beiden Annahmen gerechtfertigt sind. Weiterhin ist der entartete Fall auszuschließen. Daher die

**Annahme 3.4.8** Zu aktiven Zeiten  $t \in \mathcal{A}$  gilt stets  $w_2 \neq 0$ .

Die Annahme ist gerechtfertigt, da wegen  $T_{\max} > 0$  die Zustandsbeschränkung bei kleinen Geschwindigkeiten nicht aktiv sein wird.

Das Linienmaß  $\mu(t)$  wird nun auf Randstücken und Berührungspunkten genauer untersucht. Die Lage des (relativ gesehen) singulären Anteils lässt sich dabei auf endlich viele isolierte Punkte fixieren. Man findet Dirac-Maße in den Berührungspunkten, sowie in den Auf- und Absprungpunkten der Randstücke. Im Inneren der Randstücke stellt sich der Multiplikator als sehr glatt heraus. Beginnen wir mit der Analyse eines Randstückes.

### 3.4.2 Analyse von Randstücken

Für verteilte aktive Mengen – also Randstücke bei OC-ODE (Kap. 1), sowie  $n$ -dimensionale aktive Mengen bei elliptischen OC-PDE (Kap. 2) – gilt: Die Zustandsbeschränkung liefert auf der aktiven Menge zu regulären Daten eine höhere Regularität der optimalen Steuerung. Beim Rocketcar können wir dies im folgenden Abschnitt erstmalig auch bei einem parabolischen OC-PDAE mit aktiven Mengen in Form von Lebesgue-Nullmengen folgern.

Betrachten wir zunächst den Stand der Forschung. Von elliptischen Optimalsteuerungsproblemen mit Zustandsbeschränkungen aus [13], [56] weiß man, dass aktive Mengen von positivem Maße zu einer reguläreren optimalen Steuerung führen; infolgedessen steigen auch die Regularitäten des adjungierten Zustandes und des Lagrange-Multiplikators der Zustandsbeschränkung im Inneren der aktiven Menge. Für aktive Linien im  $\mathbb{R}^2$  erhält man nach [13] zwar im relativen Inneren noch die Knickbedingung des adjungierten Zustandes  $\frac{\partial p^+}{\partial \nu^+} + \frac{\partial p^-}{\partial \nu^-} = \mu_{\dot{\mathcal{A}}}$  in  $C(\dot{\mathcal{A}})^*$ , allerdings fällt die erhöhte Regularität weg, da keine zusätzliche Information in Normalenrichtung der aktiven Linie vorhanden ist.

Beim Rocketcar kann nach Satz 3.3.7 keine aktive Menge von positivem Maße vorliegen. Wir haben den Fall einer innerlich geschlitzten Menge. Eine derartige Menge ist schwer zu behandeln, da sie z.B. keinen Lipschitz-Rand hat, da sich die Menge am Schlitz nicht mehr nur auf einer Seite des Randes befindet (siehe Grisvard [44, S. 5]). Dennoch lässt sich eine höhere Regularität einer optimalen Steuerung zu aktiven Zeiten der Zustandsbeschränkung beweisen.

**Satz 3.4.9 (Höhere Regularität der optimalen Steuerung auf  $\bar{R}$ )**

Sei  $(\bar{T}, \bar{w}_1, \bar{w}_2, \bar{u}, q, p_1, p_2, \mu)$  eine optimale Lösung des Rocketcar-Problems im Falle eines Randstückes  $\bar{R}$  mit  $R := (t_{\text{on}}, t_{\text{off}})$ , d.h.  $\bar{T}(\frac{l}{2}, t) \equiv T_{\text{max}}$  auf  $\bar{R}$ . Dann gilt  $\bar{u} \in C^\infty(\bar{R})$ .

*Beweis:*

Wir greifen die Superposition  $\bar{T} = T_1 + T_2$  aus dem Beweis von Satz 3.3.4 auf mit  $T_1(x, t) = z(t)$  mit  $\dot{z}(t) = g(\bar{w}_2(t))$ ,  $z(0) = 0$  und  $T_2(x, t) \in C^\infty(Q)$ . Es gilt für die reguläre Kurve  $\bar{\gamma}(t) := (x(t), t)$  mit  $x(t) \equiv l/2$ ,  $t \in \bar{R}$  entsprechend  $\bar{T}(\bar{\gamma}(t)) \equiv T_{\text{max}} \in C^\infty(\bar{R})$  und daher

$$0 \equiv \frac{d}{dt} \{ \bar{T}(\bar{\gamma}(t)) \} = T_{1,t}(\frac{l}{2}, t) + T_{2,t}(\frac{l}{2}, t) = g(\bar{w}_2(t)) + \underbrace{T_{2,t}(\frac{l}{2}, t)}_{\in C^\infty(\bar{R})}.$$

Der hintere Summand ist wegen  $T_2(x, t) \in C^\infty(Q)$  und  $(\bar{R} \times \{\frac{l}{2}\}) \subset\subset Q$  beliebig glatt. Als Konsequenz ist  $g(\bar{w}_2(t)) =: \xi(t) \in C^\infty(\bar{R})$ .

Diese hohe Regularität ist als nächstes auf  $\bar{w}_2(t)$  zu übertragen. Zunächst gilt wegen Annahme 3.4.8 zum einen  $z := \bar{w}_2(t) \neq 0$  und wegen (3.3b), (3.3c) auch  $y := g(\bar{w}_2(t)) > 0 \forall t \in \bar{R}$ . Sei O.E.  $z = \bar{w}_2(t) > 0 \forall t \in \bar{R}$ .<sup>3</sup> Dann ist wegen (3.3d)  $y = g(z)$  invertierbar mit stetig differenzierbarer Umkehrfunktion  $z = g^{-1}(y)$  und es gilt die Umkehrregel  $g^{-1'}(y) = 1/g'(g^{-1}(y))$ . Wir machen einen Induktionsschritt:  $g^{-1}(y)$  ist  $C^1(\mathbb{R}^+)$ . Wegen  $z = g^{-1}(y) > 0$  ist nach (3.3a) auch  $g'(g^{-1}(y)) \in C^1(\mathbb{R}^+)$  und  $g'(g^{-1}(y)) > 0$ . Nach der Umkehrregel ist daher  $g^{-1'}(y) \in C^1(\mathbb{R}^+)$  und folglich  $g^{-1}(y) \in C^2(\mathbb{R}^+)$ . Induktion liefert  $g^{-1}(y) \in C^\infty(\mathbb{R}^+)$ .

Als Konsequenz erhält man wegen  $\xi|_{\bar{R}} > 0$  nun  $\bar{w}_2(t) = g^{-1}(\xi(t)) \in C^\infty(\bar{R})$  als Verkettung von  $C^\infty$ -Funktionen. Folglich ist auch  $\bar{u}(t) = \dot{\bar{w}}_2(t) \in C^\infty(\bar{R})$ .  $\square$

**Folgerung 3.4.10** Da nach Annahme 3.4.7 die Steuerungsbeschränkung auf  $[t_{\text{on}}, t_{\text{off}}]$  nicht aktiv ist, folgt aus (3.10)  $p_2 \in C^\infty(\bar{R})$  sowie aus (3.7e) wegen  $g'(\bar{w}_2(t))|_{\bar{R}} \neq 0$  und  $g'(\bar{w}_2(t)) \in C^\infty(\bar{R})$ :

$$\int_{\Omega} q(x, t) dx \in C^\infty(\bar{R}).$$

**Folgerung 3.4.11 (Feedback-Formel)**

Aufgrund von  $T_2 \in C^\infty(Q)$  ergibt sich die Feedback-Formel

$$g(\bar{w}_2(t)) = -T_{2,t} \left( \frac{l}{2}, t \right) = -T_{2,xx} \left( \frac{l}{2}, t \right) = -\bar{T}_{xx} \left( \frac{l}{2}, t \right) \quad (3.13)$$

und weiter

$$g'(\bar{w}_2(t)) \cdot \bar{u}(t) = -T_{2,tt} \left( \frac{l}{2}, t \right) = -T_{2,xxxx} \left( \frac{l}{2}, t \right) = -\bar{T}_{xxxx} \left( \frac{l}{2}, t \right). \quad (3.14)$$

<sup>3</sup> Der Fall  $\bar{w}_2 < 0$  folgt analog mit Symmetrie (3.3e).

Es sind also zwei Zeitableitungen nötig, bis die Steuerung explizit auftritt. Bei OC-ODE würde man nach Definition 1.1.6 von einer Zustandsbeschränkung 2. Ordnung sprechen, welche wir in [79] im Rahmen der Formulierung als OC-ODE auch zeigen. Ebenso würde man nach einer Semidiskretisierung mit der Linienmethode ein OC-ODE mit einer Zustandsbeschränkung 2. Ordnung erhalten.

Um die hohe Regularität auf das Maß übertragen zu können, braucht man den

**Hilfssatz 3.4.12** Es gilt  $\int_{\Gamma} q(s, t) ds \in C^{\infty}([0, t_f])$ .

*Beweis:*

Das Randintegral enthält in 1D die beiden Anteile  $\int_{\Gamma} q(s, t) ds = q(0, t) + q(l, t)$ . Man zeigt  $q(0, t) \in C^{\infty}([0, t_f])$ , der andere Anteil  $q(l, t)$  folgt analog oder durch Ausnutzen der Symmetrie des Problems.

Auf dem Bereich  $S := (0, \frac{l}{4}) \times (0, t_f)$  erfüllt  $q$  die homogene PDE  $-q_t - \Delta q = 0$ . Aufgrund der hohen inneren Regularität gilt entsprechend  $q \in C^{\infty}(S)$  (z.B. Evans [35, S. 59]). Da die Randbedingung  $\partial_{\nu} q(0, t) + q(0, t) = 0 \in C^{\infty}([0, t_f])$  ist, überträgt sich die hohe Regularität auch auf dieses Randstück. Da diese Tatsache in der Literatur nicht auffindbar ist, folgt ihr Beweis:

Als Hilfsproblem betrachtet man das Anfangsrandwertproblem der adjungierten Gleichung (3.7a)-(3.7c) auf einem vergrößerten Ortsgebiet  $\Omega_1 := (-\delta, l + \delta)$  für ein  $\delta > 0$ . Dann ist dessen Lösung  $q_1 \in C^{\infty}((-\delta, l/4) \times [0, t_f])$  und wegen Folgerung 3.4.5  $q_1 \equiv 0$  auf  $\bar{\Omega}_1 \times (t_L, t_f]$ . Daher ist  $q_1(0, t)$ , aber auch  $q_{1,x}(0, t) \in C^{\infty}([0, t_f])$  und analog für  $x = l$ . Nun löst man das Ausgangsproblem mit der Superposition  $q = q_1 + q_2$ . Dann muss  $q_2$  folgendes Anfangsrandwertproblem lösen:

$$\begin{aligned} -q_{2,t} - \Delta q_2 &= 0 && \text{in } Q, \\ \partial_{\nu} q_2 + q_2 &= -\partial_{\nu} q_1 - q_1 \in C^{\infty}([0, t_f]) && \text{an } \Sigma, \\ q_2(x, t_f) &= 0 && \text{in } \Omega. \end{aligned}$$

Wegen  $q_2 = q_1 = 0$  auf  $\bar{\Omega} \times (t_L, t_f]$  sind die Verträglichkeitsbedingungen zwischen der Randbedingung und der Anfangsbedingung aus [35, Sect. 7.1] stets erfüllt. Damit ist  $q_2 \in C^{\infty}(\bar{Q})$ , indem man [35, S. 367] auf Robin-Randbedingungen anpasst. Daher ist  $q(0, t) \in C^{\infty}([0, t_f])$ .  $\square$

### Satz 3.4.13 (Maßzerlegung auf Randstücken)

Für Problem 1 gilt auf einem Randstück  $\bar{R}$  die Maß-Zerlegung

$$\mu|_{\bar{R}} = \varrho_{\text{on}} \cdot \delta(t - t_{\text{on}}) + \varrho_{\text{off}} \cdot \delta(t - t_{\text{off}}) + \sigma(t),$$

mit  $\varrho_{\text{on}}, \varrho_{\text{off}} \in \mathbb{R}_0^+$  und

$$\sigma(t) := - \int_{\Omega} q_t(x, t) dx + \int_{\Gamma} q(s, t) ds \in C^{\infty}(\bar{R}), \quad \sigma(t) \geq 0, \quad (3.15)$$

d.h. es gilt für alle  $V$  offen mit  $\bar{V} \subset (0, t_f)$  und  $V \cap \mathcal{A} = \bar{R}$ :

$$\int_V \psi(t) d\mu(t) = \varrho_{\text{on}} \psi(t_{\text{on}}) + \varrho_{\text{off}} \psi(t_{\text{off}}) + \int_{\bar{R}} \psi(t) \sigma(t) dt \quad \forall \psi(t) \in C(V).$$

*Beweis:*

Wähle ein solches  $V$  und definiere  $V_L := V \cap (0, t_{\text{on}})$  und  $V_R := V \cap (t_{\text{off}}, t_f)$ . Dies ist aufgrund Annahme 3.4.7 stets möglich. Wir testen nun zuerst die adjungierte PDE (3.9) mit  $\varphi(x, t) = \psi(t) \cdot \zeta(x)$  mit  $\psi(t) \in C_0^\infty(V_L)$  (sowie mit 0 fortgesetzt auf  $[0, t_f]$ ) und  $\zeta(x) \equiv 1, \forall x \in \bar{\Omega}$ . Wegen  $\varphi(x, t) \in C^\infty(\bar{Q})$  und  $\varphi(x, 0) = 0$  ist dies eine zulässige Testfunktion. Wegen  $\mu|_{V_L} = 0$  folgt aus (3.9):

$$\int_{V_L} \dot{\psi}(t) \int_{\Omega} q(x, t) dx dt + \int_{V_L} \psi(t) \int_{\Gamma} q(s, t) ds dt = 0 \quad \forall \psi(t) \in C_0^\infty(V_L).$$

Da  $\xi(t) := \int_{\Omega} q(x, t) dx$  nach Lemma 3.4.3 zunächst nur in  $L^1(V_L)$  ist, betrachtet man  $\mathcal{T}_\xi(\psi) := \int_{V_L} \psi(t) \xi(t) dt \forall \psi(t) \in C_0^\infty(V_L)$  als reguläre Distribution. Die distributionelle Ableitung  $T_\xi$  erfüllt

$$T_\xi(\psi) = -T_\xi(\dot{\psi}) = - \int_{V_L} \dot{\psi}(t) \xi(t) dt = \int_{V_L} \psi(t) \int_{\Gamma} q(s, t) ds dt \quad \forall \psi(t) \in C_0^\infty(V_L)$$

und wird daher von  $\dot{\xi}(t) = \int_{\Gamma} q(s, t) ds$  erzeugt. Nach Hilfssatz 3.4.12 ist  $\dot{\xi}(t) \in C^\infty(\bar{V}_L)$ . Folglich ist  $\sigma|_{V_L} = 0$  und  $\xi(t) = \int_{\Omega} q(x, t) dx \in C^\infty(\bar{V}_L)$ .

Analog geht man auf  $V_R$  vor und erhält  $\sigma|_{V_R} = 0$  sowie  $\int_{\Omega} q(x, t) dx \in C^\infty(\bar{V}_R)$ . Wählt man in gleicher Konstellation dagegen  $\psi(t) \in C_0^\infty(V)$ , so folgt

$$\int_V \dot{\psi}(t) \int_{\Omega} q(x, t) dx + \int_V \psi(t) \int_{\Gamma} q(s, t) ds dt = \int_{\bar{R}} \psi(t) d\mu(t) \quad \forall \psi(t) \in C_0^\infty(V).$$

Aufgrund Folgerung 3.4.10 gilt bereits  $\int_{\Omega} q(x, t) dx \in C^\infty(\bar{R})$ ; daher lassen sich die Integrale auf der linken Seite aufteilen und auf jedem Abschnitt partiell integrieren unter Anwendung der Leibniz-Regel für Parameterintegrale. Es folgt:

$$\begin{aligned} \int_{V_L} \psi(t) \sigma(t) dt + \int_{\bar{R}} \psi(t) \sigma(t) dt + \int_{V_R} \psi(t) \sigma(t) dt + \psi(t_{\text{on}}) \underbrace{\left[ \int_{\Omega} q(x, t) dx \right]_{t_{\text{on}}^-}^{t_{\text{on}}^+}}_{=: \varrho_{\text{on}}} \\ + \psi(t_{\text{off}}) \underbrace{\left[ \int_{\Omega} q(x, t) dx \right]_{t_{\text{off}}^+}^{t_{\text{off}}^-}}_{=: \varrho_{\text{off}}} = \int_{\bar{R}} \psi(t) d\mu(t), \quad \forall \psi(t) \in C_0^\infty(V). \end{aligned}$$

Wegen  $\sigma|_{V_L} = \sigma|_{V_R} = 0$  gilt entsprechend

$$\int_{\bar{R}} \psi(t) \sigma(t) dt + \varrho_{\text{on}} \psi(t_{\text{on}}) + \varrho_{\text{off}} \psi(t_{\text{off}}) = \int_{\bar{R}} \psi(t) d\mu(t), \quad \forall \psi(t) \in C_0^\infty(V).$$

Unter Verkleinerung von  $V$  folgt dies auch für alle stetigen Funktionen, da  $C_0^\infty(V) \subset C(\bar{V}_2)$  dicht ist für ein  $V_2$  offen mit  $\bar{R} \subset V_2$  und  $\bar{V}_2 \subset V$ .

Wegen Folgerung 3.4.10 und Hilfssatz 3.4.12 ist entsprechend  $\sigma(t) \in C^\infty(\bar{R})$ . Die Vorzeichenbedingungen folgen aus  $\mu(t) \geq 0$ .  $\square$

**Bemerkung 3.4.14 (Sprungbedingungen)**

Der Satz liefert als Beiprodukt die Sprungbedingungen

$$\left[ \int_{\Omega} q(x, t) dx \right]_{t_{\text{on}}^+}^{t_{\text{on}}^-} = \varrho_{\text{on}} \geq 0, \quad \left[ \int_{\Omega} q(x, t) dx \right]_{t_{\text{off}}^+}^{t_{\text{off}}^-} = \varrho_{\text{off}} \geq 0. \quad (3.16)$$

Allgemein kann man aus  $\bar{w}_2(t) \in W^{1,\infty}(0, t_f)$  und (3.3a) folgern, dass  $g'(\bar{w}_2(t)) \in W^{1,\infty}(Z)$  für jede Zusammenhangskomponente  $Z$  von  $\{t \in (0, t_f) \mid \bar{w}_2(t) \neq 0\}$  gilt. Wegen Annahme 3.4.8 und (3.7e) übertragen sich die Sprungbedingungen daher auf

$$\left[ \dot{p}_2 \right]_{t_{\text{on}}^-}^{t_{\text{on}}^+} = g'(\bar{w}_2(t_{\text{on}})) \cdot \varrho_{\text{on}}, \quad \left[ \dot{p}_2 \right]_{t_{\text{off}}^-}^{t_{\text{off}}^+} = g'(\bar{w}_2(t_{\text{off}})) \cdot \varrho_{\text{off}}.$$

Sie sind das Analogon zu Sprungbedingungen bei OC-ODE aus Kap. 1.

**Bemerkung 3.4.15** Die Darstellung des Maes lasst sich problemlos auf inhomogene Robin-Randbedingungen und Neumann-Randbedingungen bertragen. Dirichlet-Randbedingungen muss man dagegen separat angehen, da hier die gewahlte Testfunktion nicht die dann notige Dirichlet-Randbedingung einhalt.

Zusatzlich gilt die folgende Knickbedingung analog zu [13]:

**Lemma 3.4.16** Innerhalb des Randstckes gilt

$$\mu(t) = \sigma(t) = q_x \left( \frac{l^-}{2}, t \right) - q_x \left( \frac{l^+}{2}, t \right) = 2 \cdot q_x \left( \frac{l^-}{2}, t \right) \quad \forall t \in R. \quad (3.17)$$

*Beweis:*

Der Beweis lauft analog zum elliptischen Fall in [13, Theorem 5]. Man definiert  $D := (l/4, 3l/4) \times R$ . Das Randstck  $\bar{R}$  trennt  $D$  in die beiden offenen Mengen  $D_L := (l/4, l/2) \times R$  und  $D_R := (l/2, 3l/4) \times R$ . Testen der sehr schwachen Form der adjungierten Gleichung (3.9) mit  $\varphi(x, t) \in C_0^\infty(D)$  liefert fur alle  $\varphi \in C_0^\infty(D)$ :

$$\begin{aligned} & \int_R \varphi \left( \frac{l}{2}, t \right) d\mu(t) - \iint_D \varphi_t q dx dt = - \iint_D \Delta \varphi q dx dt \\ & = - \int_R \left( \int_{l/4}^{l/2} \Delta \varphi q dx + \int_{l/2}^{3l/4} \Delta \varphi q dx \right) dt \\ & = - \int_R \left( \int_{l/4}^{l/2} \Delta q \varphi dx + \int_{l/2}^{3l/4} \Delta q \varphi dx \right) dt + \int_R - [\varphi_x q]_{\frac{l}{2}^+}^{\frac{l}{2}^-} + [q_x \varphi]_{\frac{l}{2}^+}^{\frac{l}{2}^-} dt. \end{aligned}$$

Zuletzt wurde die verallgemeinerte Greensche Formel verwendet; die letzte Gleichung ist daher vorerst distributionell zu verstehen, da wir noch nicht wissen, dass  $\Delta q \in L^2(\Omega \times R)$  ist. Auerdem ist die einseitige Spur  $q_x$  in  $x = l/2$  noch nicht erklart. Wie die

Spurabbildungen konkret zu bilden sind, findet man in [24, Theorem 6.2]. Ein gleiches Vorgehen auf  $D_L$  liefert formalen

$$\int_R \varphi\left(\frac{l}{2}, t\right) d\mu(t) - \iint_{D_L} \varphi_t q \, dx \, dt = - \iint_{D_L} \Delta q \varphi \, dx,$$

da hier die Randterme verschwinden. Analoges Vorgehen auf  $D_R$  und Kombination dieser drei Gleichungen lässt die verteilten Terme verschwinden, was dem Verwenden der homogenen PDE entspricht. Nur die Randterme bleiben stehen. Unter Verwendung der passenden Spurabbildungen lauten diese für das  $m \in [1, 2)$  aus Lemma 3.4.3:

$$\begin{aligned} \int_R \varphi\left(\frac{l}{2}, t\right) d\mu(t) &= \int_R \left[ q\left(\frac{l^+}{2}, t\right) - q\left(\frac{l^-}{2}, t\right) \right] \varphi_x\left(\frac{l}{2}, t\right) dt \\ &+ \left\langle q_x\left(\frac{l^-}{2}, t\right) - q_x\left(\frac{l^+}{2}, t\right), \varphi\left(\frac{l}{2}, t\right) \right\rangle_{W^{1/m, m'}(R)^*, W^{1/m, m'}(R)}. \end{aligned}$$

Nun sieht man zum einen  $q(\frac{l^+}{2}, t) = q(\frac{l^-}{2}, t)$  f.ü. auf  $R$ , da  $q \in L^r(0, t_f; W^{1, m}(\Omega))$  in Ortsrichtung nicht springen kann. Zum anderen folgt

$$\mu(t) = q_x\left(\frac{l^-}{2}, t\right) - q_x\left(\frac{l^+}{2}, t\right)$$

in  $C(R)^*$ . Höhere Regularität liefert Satz 3.4.13 wegen  $\mu(t) = \sigma(t)$ . Weiterhin ist  $q(x, t)$  offensichtlich achsensymmetrisch zu  $x = l/2$ .  $\square$

**Folgerung 3.4.17** Die Kombination von Satz 3.4.13 und Lemma 3.4.16 liefert  $\mu|_R = 2q_x(l/2, t) \in H^k(R) \forall k \in \mathbb{N}$ . Daher erhalten wir stückweise eine klassische Lösung  $q \in C^\infty([t_{\text{on}} + \varepsilon, t_{\text{off}} - \varepsilon]; B) \forall 0 < \varepsilon < t_{\text{off}} - t_{\text{on}}$  mit  $B := \{v \in C(\bar{\Omega}) \mid v|_{[0, l/2]} \in C^\infty([0, l/2]), v|_{[l/2, l]} \in C^\infty([l/2, l])\}$ .

**Bemerkung 3.4.18** Die Ergebnisse aus Satz 3.4.13 lassen sich wie folgt einordnen: Man wendet die Lebesgue-Zerlegung (Rudin [89, Theorem 6.9]) auf das Maß  $\mu|_{\bar{R}}$  an, d.h. man zerlegt  $\mu|_{\bar{R}} =: \mu_{\bar{R}} = \mu_a + \mu_s$  in einen absolut-stetigen Anteil  $\mu_a$  und einen singulären Anteil  $\mu_s$ . Der reguläre Anteil  $\mu_a$  besitzt nach dem Satz von Radon-Nikodým (z.B. [89], [5]) eine  $L^1$ -Dichte; wir wissen aus Satz 3.4.13, dass  $\mu_s|_R = 0$  ist und  $\mu_a|_R = \sigma(t) \in C^\infty(\bar{R})$  sogar eine noch bessere Dichte besitzt. Damit ist der singuläre Anteil auf  $\partial[t_{\text{on}}, t_{\text{off}}] = \{t_{\text{on}}, t_{\text{off}}\}$  konzentriert. Reguläre Borel-Maße auf isolierten Punkten sind aber bekanntlich stets Dirac-Maße (Hilfssatz 3.4.19).

### 3.4.3 Analyse von Berührungspunkten

Neben Randstücken können laut unserer Strukturannahme auch Berührungspunkte auftreten. Aufgrund Annahme 3.4.7 sind solche Berührungspunkte isoliert und von Randstücken separiert. Das zugehörige Maß liegt daher auf einem isolierten Punkt vor und ist nach folgendem Hilfssatz ein Dirac-Maß (vgl. auch Casas [23, Remark 4]).

#### Hilfssatz 3.4.19 (Isolierte Punkte und Dirac-Maße)

Ein reguläres Borel-Maß  $\mu \in \mathcal{M}(\bar{Q})$ , welches auf einer einpunktigen Menge  $\{(x_0, t_0)\}$  konzentriert ist mit  $(x_0, t_0) \in Q$ , ist ein Dirac-Maß.

*Beweis:*

Der Sachverhalt des Hilfssatzes ist bekannt und verbreitet; der Beweis liegt dem Autor leider nicht vor, da er laut [22] in der Dissertation von Eduardo Casas [21] steht. Zur Vollständigkeit wird er daher erbracht: Sei  $(x_0, t_0) \in Q$  der Punkt, in welchem das Maß  $\mu \in \mathcal{M}(\bar{Q})$  konzentriert ist. Die Borel- $\sigma$ -Algebra auf  $\{(x_0, t_0)\}$  beinhaltet nur die leere Menge und die gesamte Menge (den Punkt). Da Maße der leeren Menge stets die 0 zuordnen müssen, ist ein Maß eindeutig festgelegt durch  $\mu(\{(x_0, t_0)\}) = c \in \mathbb{R}$ . Dieses Maß wird mit Null fortgesetzt zu einem regulären Borel-Maß auf  $\bar{Q}$ . Es reicht bekanntlich, ein Maß auf einem durchschnittstabilen Erzeugendensystem der Borel- $\sigma$ -Algebra auf  $\bar{Q}$  zu definieren, wobei sich die Gesamtheit aller offenen Mengen  $A$  anbietet. Man definiert also

$$\begin{aligned} (x_0, t_0) \in A &\Leftrightarrow \mu(A) = c, \\ (x_0, t_0) \notin A &\Leftrightarrow \mu(A) = 0. \end{aligned}$$

Das Maß ist also ein  $c$ -faches des Dirac-Maßes. Auf offenen Mengen besitzt das Maß folglich eine Dirac-Dichte

$$\mu(A) = \iint_A c \cdot \delta(x - x_0) \cdot \delta(t - t_0) dx dt.$$

□

Wir notieren also

$$\mu|_{t_c^j} = \varrho_c^j \cdot \delta(t - t_c^j), \quad \varrho_c^j \in \mathbb{R}, \quad (3.18)$$

d.h. jeder Berührungspunkt liefert einen Dirac-Impuls. Insgesamt liefert damit jeder Schaltpunkt einen Dirac-Impuls.

Unter Ausnutzung der Vorzeichenbedingungen aus (3.12) können wir das Maß vollständig charakterisieren:

$$\mu(t) = \sum_{j=1}^m \varrho_c^j \delta(t - t_c^j) + \sum_{i=1}^n \varrho_{\text{on}}^i \delta(t - t_{\text{on}}^i) + \varrho_{\text{off}}^i \delta(t - t_{\text{off}}^i) + \sigma^i(t), \quad (3.19)$$

$$\mu_{\bar{Q}} = \mu(t) \cdot \delta(x - \frac{l}{2}), \quad \varrho_*^* \in \mathbb{R}, \quad \varrho_*^* \geq 0, \quad \sigma^i(t) \in C^\infty([t_{\text{on}}^i, t_{\text{off}}^i]), \quad \sigma^i(t) \geq 0.$$

#### **Bemerkung 3.4.20 (Sprungbedingungen bei einem Berührungspunkt)**

Mit der Technik aus dem Beweis von Satz 3.4.13 erhält man die Sprungbedingung

$$\left[ \int_{\Omega} q(x, t) dx \right]_{t_c^{j,+}}^{t_c^{j,-}} = \varrho_c^j \geq 0, \quad \Leftrightarrow \quad \left[ \dot{p}_2 \right]_{t_c^{j,+}}^{t_c^{j,-}} = g'(\bar{w}_2(t_c^j)) \cdot \varrho_c^j. \quad (3.20)$$

#### **3.4.4 Höhere Regularität**

Da die adjungierte PDE linear ist, könnte man die einzelnen Fälle der Maßzerlegung superponieren und auf diese Weise die Regularität von  $q(x, t)$  charakterisieren. Ein Dirac-Impuls liefert z.B. eine Fundamentallösung (siehe z.B. [35], [88]). Unter Anwendung der inversen Zeittransformation (siehe Abschnitt 4.6.2) folgt direkt das

**Lemma 3.4.21 (Fundamentallösung)**

Sei  $0 < t_s < t_f$ . Die Lösung des rückwärtsgerichteten Anfangswertproblems

$$\begin{aligned} -r_t - \Delta r &= \delta(x - x_s) \cdot \delta(t - t_s) && \text{in } \mathbb{R} \times (0, t_f), \\ r(x, t_f) &= 0 && \forall x \in \mathbb{R}, \end{aligned}$$

ist gegeben durch  $r = \Phi(x - x_s, t_s - t)$ , wobei  $\Phi(x, t)$  die (mit Null fortgesetzte) Fundamentallösung der Wärmeleitungsgleichung ist, d.h.  $\Phi(x, t) \in C^\infty(\mathbb{R}^2 \setminus \{(0, 0)\})$  und

$$\Phi(x, t) = \begin{cases} \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{x^2}{4t}\right) & \text{für } t > 0, \\ \text{n.d.} & \text{für } (x, t) = (0, 0), \\ 0 & \text{sonst.} \end{cases} \quad (3.21)$$

Numerisch lässt sich dies in Abschnitt 3.6.2 dank dem Superpositionsprinzip ausnutzen, um die Fundamentallösungen analytisch zu berechnen und so den Fehler zu reduzieren. Konzentrieren wir uns jedoch auf die Regularitäten der restlichen Variablen. Dazu sind zunächst folgende Definitionen nötig:

**Definition 3.4.22** Wir definieren die verschiedenen signifikanten Abschnitte der Zeitskala:

$$\begin{aligned} \mathcal{N} &:= \{t \in (0, t_f) \mid \bar{w}_2(t) = 0\}, \\ \mathcal{S} &:= \{t \in (0, t_f) \mid t = t_{\text{on}}^i \vee t = t_{\text{off}}^i \vee t = t_c^j, i = 1, \dots, n, j = 1, \dots, m\}, \\ \mathcal{B} &:= \{t \in [0, t_f] \mid |u(t)| = u_{\text{max}}\}, \\ \mathcal{C} &:= (0, t_f) \setminus (\mathcal{B} \cup \mathcal{A}). \end{aligned}$$

Die Dirac-Maße sind also auf die Schaltzeiten  $\mathcal{S}$  konzentriert;  $\mathcal{C}$  ist die Schnittmenge der inaktiven Mengen der Steuerungs- und Zustandsbeschränkungen.

**Definition 3.4.23** Wir definieren die stückweise glatten Funktionen als

$$\begin{aligned} C_{p,I}^\infty([0, t_f]) &:= \{u \in L^\infty(0, t_f) \mid \exists 0 = t_0 < t_1 < \dots < t_n = t_f \quad \forall k = 0, \dots, n-1 : \\ &u|_{[t_k, t_{k+1}]} \in C^\infty([t_k, t_{k+1}]), \quad I := \{t_1, \dots, t_{n-1}\}\}. \end{aligned}$$

$I$  bezeichnet also die Menge der Knickstellen. Ist die Funktion zusätzlich stetig oder stetig differenzierbar, so schreiben wir

$$\begin{aligned} C_{c,I}^\infty([0, t_f]) &:= C_{p,I}^\infty([0, t_f]) \cap C([0, t_f]), \\ C_{d,I}^\infty([0, t_f]) &:= C_{p,I}^\infty([0, t_f]) \cap C^1([0, t_f]). \end{aligned}$$

Obwohl  $q(x, t)$  in der Umgebung von  $(l/2, t_s)$  unbeschränkt ist für alle  $t_s \in \mathcal{S}$ , zeigt das Ortsintegral von  $q$  stückweise eine sehr hohe Regularität: Das Integral ist stückweise glatt und zeigt Sprungstellen in  $\mathcal{S}$  gemäß den Sprungbedingungen.

**Lemma 3.4.24** Es gilt

$$\int_{\Omega} q(x, t) dx \in C_{p,S}^{\infty}([0, t_f]) \quad \forall \varepsilon > 0.$$

*Beweis:*

Zum einen liefert Hilfssatz 3.4.12  $\int_{\Gamma} q(s, t) ds \in C^{\infty}([0, t_f])$ . Zum anderen gilt die Identität  $\sigma(t) := -\int_{\Omega} q_t(x, t) dx + \int_{\Gamma} q(s, t) ds = 0$  nach der Konstruktion im Beweis von Satz 3.4.13 sogar auf  $\bar{Z}$  für jede Zusammenhangskomponente  $Z$  von  $\mathcal{I}$ . Kombiniert mit dem bekannten Zusammenhang auf Randstücken folgt das Resultat.  $\square$

**Folgerung 3.4.25** Insbesondere ist  $\int_{\Omega} q(x, t) dx \in L^{\infty}(0, t_f)$  und daher  $p_2 \in W^{1,\infty}(0, t_f)$ .

Die Lösungen von Optimalsteuerungsproblemen zeichnen sich oft durch eine höhere Regularität aus, als es die Zustandsgleichungen erwarten lassen. Aus den adjungierten Gleichungen lassen sich ggf. Rückschlüsse auf die Regularität der Zustände ziehen. Im vorliegenden Beispiel finden wir sogar stückweise eine sehr hohe Regularität:

**Lemma 3.4.26 (Höhere Regularität der Zustände auf  $\mathcal{C}$ )**

Auf jeder Zusammenhangskomponente  $Z$  von  $\mathcal{C} \setminus \mathcal{N}$  gilt:

$$\bar{w}_2, \bar{u}, p_2 \in C^{\infty}(\bar{Z}).$$

*Beweis:*

Generell gilt  $\bar{w}_2 \in W^{1,\infty}(Z)$ . Daher gilt diese Folgerungskette mit  $m := 1$ :

$$\begin{aligned} \bar{w}_2 \in W^{m,\infty}(Z) &\xrightarrow{\bar{w}_2 \neq 0} g'(\bar{w}_2(t)) \in W^{m,\infty}(Z) && \xrightarrow[\text{(3.7e)}]{\text{Lemma 3.4.24}} \\ \bar{p}_2 \in W^{m,\infty}(Z) &\implies p_2, \bar{u} \in W^{m+1,\infty}(Z). \end{aligned}$$

Wegen  $\dot{\bar{w}}_2 = u$  ist aber dann  $\bar{w}_2 \in W^{m+2,\infty}(Z)$ . Induktion über  $m$  liefert  $\bar{w}_2, \bar{u}, p_2 \in W^{k,\infty}(Z) \quad \forall k \in \mathbb{N}$ . Also liegen diese Größen insbesondere in  $H^k(Z) \quad \forall k \in \mathbb{N}$  und nach Anwendung von Sobolew-Einbettungssätzen [5, Satz 8.9] folgt entsprechend  $\bar{w}_2, \bar{u}, p_2 \in C^{k-1}(\bar{Z}) \quad \forall k \in \mathbb{N}$ .  $\square$

Wir wissen bereits, dass die gleiche Regularität innerhalb von Randstücken vorliegt. Damit ist  $\bar{u}$  stückweise glatt, außer auf der aktiven Menge  $\mathcal{B}$  der Steuerungsbeschränkung; dort gilt  $\bar{u} = \pm u_{\max}$ , d.h. die Daten legen die Regularität fest. Der praktisch relevante Fall einer Maximalbeschleunigung  $u_{\max} = \text{const}$  liefert daher stückweise glatte Lösungen:

**Lemma 3.4.27** Gilt sogar  $u_{\max} \in C^{\infty}([0, t_f])$  so folgt

$$\begin{aligned} \bar{u} &\in C_{c,S \cup \mathcal{N} \cup \partial \mathcal{B}}^{\infty}([0, t_f]), & \bar{w}_2 &\in C_{d,S \cup \mathcal{N} \cup \partial \mathcal{B}}^{\infty}([0, t_f]), \\ g(\bar{w}_2(t)) &\in C_{c,S \cup \mathcal{N} \cup \partial \mathcal{B}}^{\infty}([0, t_f]). \end{aligned}$$

*Beweis:*

Zunächst gilt nach Folgerung 3.4.25  $p_2 \in W^{1,\infty}(0, t_f) \hookrightarrow C([0, t_f])$  und daher wegen der Projektionsformel auch  $\bar{u} \in W^{1,\infty}(0, t_f) \hookrightarrow C([0, t_f])$ . Zusätzlich wissen wir bereits  $\bar{u} \in C^\infty(\bar{R})$  für jedes Randstück aus 3.4.9. Weiterhin gilt auf jeder Zusammenhangskomponente  $Z$  von  $\mathcal{B}$  entweder  $\bar{u} \equiv u_{\max}$  oder  $\bar{u} \equiv -u_{\max}$ , in jedem Falle  $\bar{u} \in C^\infty(\bar{Z})$ . Zusammen mit Lemma 3.4.26 folgt  $\bar{u} \in C_{c,S \cup \mathcal{N} \cup \partial \mathcal{B}}^\infty([0, t_f])$ . Darauf aufbauend verwendet man  $\dot{w}_2 = u$  und (3.3a). Allerdings kann  $g(\bar{w}_2(t))$  an  $\mathcal{N}$  knicken, so dass wir nur  $g(\bar{w}_2(t)) \in C_{c,S \cup \mathcal{N} \cup \partial \mathcal{B}}^\infty([0, t_f])$  notieren. Für  $\mathcal{N} = \emptyset$  ist entsprechend zusätzlich  $g(\bar{w}_2(t)) \in C_{d,S \cup \partial \mathcal{B}}^\infty([0, t_f])$ .  $\square$

Die Temperaturverteilung ist dann ebenfalls stückweise regulär, allerdings müsste man hier spezielle Räume abstrakter Funktionen einführen, was zu weit abführt.

Dies sind genau die Regularitäten, mit welchen die numerische Methode im nächsten Abschnitt zu kämpfen haben wird. Die optimale Steuerung knickt also mitunter an allen Randpunkten der aktiven Mengen der Steuerungs- und Zustandsbeschränkungen, sowie am Umkehrpunkt ( $\bar{w}_2 = 0$ ). Zu den Zeiten  $\mathcal{N} \cap \bar{\mathcal{B}}$  kann es ferner keine Knicke geben.

Zuletzt knüpfen wir an den Begriff der Ordnung einer Zustandsbeschränkung aus Kapitel 1 und 2 an. Ohne den Begriff der Ordnung der Zustandsbeschränkung zu definieren, erkennen wir bereits eine wichtige Tatsache:

### Bemerkung 3.4.28 (Zur Zustandsbeschränkung)

1. Wir dürfen nicht einfach beliebiges „Ableiten und Einsetzen“ anwenden. So gilt beispielsweise folgende Folgerungskette nicht:

$$\bar{T} \equiv T_{\max} \text{ auf } \mathcal{A} \xrightarrow{\text{Ableiten}} \bar{T}_t - \Delta \bar{T} \equiv T_{\max,t} - \Delta T_{\max} \xrightarrow[\substack{\text{Einsetzen} \\ T_{\max} = \text{const}}]}{g(\bar{w}_2(t))} \equiv 0.$$

Denn i.A. ist  $\Delta T(l/2, t) \neq \Delta T_{\max}(l/2, t) \forall t \in \bar{R}$ . Genauer gilt nach dem Beweis von Satz 3.4.9 auf  $\mathcal{A}$ :  $g(\bar{w}_2(t)) = -\frac{d}{dt} \{T_2(\frac{l}{2}, t)\}$ . Diese rechte Seite verschwindet aus Energieerhaltung nur im stationären Fall (siehe auch die numerischen Resultate [79]). Eine nach Satz 3.4.9 erlaubte Vorgehensweise ist

$$\bar{T} \equiv T_{\max} \text{ auf } \mathcal{A} \xrightarrow{\text{Ableiten}} \bar{T}_t \equiv T_{\max,t} \xrightarrow[\substack{\text{Superposition} \\ T_{\max} = \text{const}}]}{g(\bar{w}_2(t))} + \underbrace{T_{2,t}(l/2, t)}_{=\Delta T_2(l/2,t)=\Delta T(\frac{l}{2},t)} \equiv 0,$$

wobei wieder die Superposition aus dem Beweis von Satz 3.3.4 mit  $T_2 \in C^\infty(Q)$  verwendet wird. Man darf also auf niederdimensionalen Anteilen der aktiven Menge nicht beliebige Ableitungen der Zustandsbeschränkung bilden, sondern nur dem Problem angepasste. Identitäten auf regulären  $C^1$ -Kurven darf man eben nur total nach dem Kurvenparameter differenzieren.

2. Die Zustandsbeschränkung legt auf  $\mathcal{A}$  die Größen  $w_2$  und  $u$  fest. Die Größen  $w_1$  und  $T$  sind aus Lösung ihrer Zustandsgleichung zu bestimmen.

Wenden wir uns nun den Lösungsmethoden zu. Beginnen wir mit den direkten Verfahren.

## 3.5 Direkte Verfahren

Da in Kap. 3 und 4 aufgrund verschiedener Anforderungen auch unterschiedliche direkte Verfahren entwickelt werden, gibt der folgende Abschnitt zunächst einen Überblick über die möglichen Varianten direkter Verfahren.

### 3.5.1 Übersicht direkter Verfahren

Bei direkten Verfahren werden das Zielfunktional und die Nebenbedingungen diskretisiert und einem Optimierer (NLP-Solver) übergeben. Dabei ist eine wichtige Entscheidung zu treffen: Verwendet man die volle oder die reduzierte Formulierung?

Bei der vollen Formulierung sind die Steuerungen und die Zustände Optimierungsvariablen. Die Zustandsgleichungen werden voll diskretisiert und dem Optimierer als Nebenbedingungen übergeben. Bei dieser vollen Diskretisierung werden die Zustandsgleichungen in Zeit und Ort diskretisiert und in einem großen Gleichungssystem gesammelt. Dieses Gleichungssystem wird dem Optimierer als Nebenbedingung übergeben.

Bei der reduzierten Formulierung erhält der Optimierer nur die wirklichen Freiheitsgrade – die Steuerungen – als Optimierungsvariablen. Die Zielfunktion wird als Unterfunktion übergeben, welche intern die Zustandsgleichungen löst, das Zielfunktional auswertet und nur Letzteres zurückgibt. Die reduzierte Formulierung kann die Zustandsgleichungen innerhalb der Unterfunktion zum einen aus der vollen Diskretisierung lösen, d.h. das große Gleichungssystem wird gelöst. Bei Anfangswertproblemen oder Anfangsrandwertproblemen gibt es aber eine effizientere Alternative: Die schrittweise Integration in der Zeit. Nach einer Semidiskretisierung im Ort wird das System hier schrittweise in der Zeit gelöst – mit passenden Einzelschritt- oder Mehrschrittverfahren.

Die volle Formulierung hat den Vorteil, dass sich reine Zustandsbeschränkungen und gemischte Steuerungs-Zustandsbeschränkungen berücksichtigen lassen. Im Gegenzug arbeitet man mit mehr Optimierungsvariablen, so dass nur Probleme geringer Dimension behandelt werden können. Das Anwendungsproblem der Brennstoffzelle in Kap. 4 lösen wir – in Abwesenheit von Zustandsbeschränkungen – daher mit der reduzierten Formulierung und der schrittweisen Integration. Dies erweist sich als sehr effizient bzgl. des Rechen- und Speicheraufwands.

### 3.5.2 Direkte Verfahren für das Rocketcar-Problem

Da das Rocketcar-Problem jedoch eine geringe Dimension hat, erlaubt uns die volle Formulierung eine Berücksichtigung der Zustandsbeschränkungen. Wir diskretisieren in  $x$  und  $t$  und sammeln alle Gleichungen in einem großen Gleichungssystem. In diesem Fall lässt sich die Zustandsbeschränkung kanonisch als punktweise Ungleichungsnebenbedingung einbauen.

Als Programmierumgebung wählen wir die Modellierungssprache AMPL [1], da wir auf diese Weise von ersten und zweiten analytischen Ableitungen profitieren. Wir lösen das Problem also mit AMPL in Kombination mit einem NLP-Solver. Wir entscheiden uns hier für das gängige und kostenlos verfügbare IPOPT [100], eine renommierte open-source Software.<sup>4</sup>

---

<sup>4</sup> IPOPT gewann 2011 den J. H. Wilkinson Prize for Numerical Software.

Die Ergebnisse des direkten Verfahrens findet man in unseren Arbeiten [79], [101], sowie in Abschnitt 3.7. Je nach Daten  $(w_0, \dot{w}_0, u_{\max}, T_{\max})$  beobachten wir dabei genau zwei verschiedene Fälle von aktiven Mengen. Entweder tritt ein Randstück auf, oder ein Berührungspunkt gefolgt von einem Randstück. Mit diesem Vorwissen können nun indirekte Verfahren der Optimalsteuerung entwickelt werden:

### 3.6 Indirekte Verfahren

Ein indirektes Verfahren für ein instationäres PDAE-System mit Zustandsbeschränkungen aufzusetzen ist eine anfordernde Aufgabe. Beim Rocketcar wissen wir apriori bereits aus den Sätzen über die Struktur der aktiven Menge, dass ein Linienmaß auftreten wird. Dieses ist in der Methode treffend aufzulösen. Auch die freie Endzeit ist treffend zu behandeln.

In einem ersten Schritt ist der Orts-Zeit-Zylinder  $Q$  passend zu unterteilen, um das Linienmaß an den Rand der Teilgebiete zu legen. Im vorliegenden Problem bedeutet dies, dass man das KKT-System auf der rechten Hälfte  $Q_R := (0, \frac{l}{2}) \times (0, t_f)$  sowie auf der linken Hälfte  $Q_L := (\frac{l}{2}, l) \times (0, t_f)$  separat betrachtet und dazwischen Übergangsbedingungen definiert. Beim Rocketcar kann man zeitgleich die Symmetrie ausnutzen, so dass generell nur eine der beiden Hälften zu lösen ist: In [80] haben wir gezeigt, dass die Lösung der PDE des Problems 1 achsensymmetrisch zu  $x = l/2$  ist. Dies überträgt sich offensichtlich auf den adjungierten Zustand  $q$  bis auf die Tatsache, dass er in  $x = l/2$  knickt. Beachtet man dies separat, kann man das KKT-System auf den halben Orts-Zeit-Zylinder  $Q_R$  reduzieren:

$$T_t - \Delta T = g(w_2(t)), \quad -q_t - \Delta q = 0 \quad \text{in } Q_R, \quad (3.22a)$$

$$\partial_\nu T(0, t) + T(0, t) = 0, \quad \partial_\nu q(0, t) + q(0, t) = 0 \quad \text{auf } (0, t_f), \quad (3.22b)$$

$$\partial_\nu T(l/2, t) = 0, \quad \partial_\nu q(l/2, t) = \frac{1}{2} \mu(t) \quad \text{auf } (0, t_f), \quad (3.22c)$$

$$T(x, 0) = 0, \quad q(x, t_f) = 0 \quad \text{in } \Omega, \quad (3.22d)$$

$$\dot{w}_1 = w_2, \quad \dot{w}_2 = u \quad \text{auf } (0, t_f), \quad (3.22e)$$

$$w_1(0) = w_0, \quad w_1(t_f) = 0, \quad (3.22f)$$

$$w_2(0) = \dot{w}_0, \quad w_2(t_f) = 0, \quad (3.22g)$$

$$-\dot{p}_1 = 0, \quad -\dot{p}_2 = p_1 + 2g'(w_2(t)) \int_0^{l/2} q(x, t) dx \quad \text{auf } (0, t_f), \quad (3.22h)$$

$$0 = 1 + \frac{\lambda}{2} u^2(t_f) + u(t_f) p_2(t_f), \quad (3.22i)$$

$$T(x, t) \leq T_{\max} \quad \text{in } \bar{Q}_R, \quad (3.22j)$$

$$\bar{u} = P_{U_{\text{ad}}} \left( -\frac{1}{\lambda} p_2 \right), \quad (3.22k)$$

$$0 = \iint_{\bar{Q}_R} [T(x, t) - T_{\max}] d\mu_{\bar{Q}}, \quad \mu_{\bar{Q}} \geq 0. \quad (3.22l)$$

### 3.6.1 Transformation auf ein Mehrpunkt-Anfangsrandwertproblem

Ziel ist es, das Maß ähnlich zu den Methoden aus Kapitel 1 treffend aufzulösen. Man stellt daher eine Hypothese an die aktive Menge analog zu OC-ODE: Eine Schaltfolge wird vorgegeben, d.h. die Anzahl und Reihenfolge der Berührungspunkte und Randstücke wird festgelegt. Anschließend werden die KKT-Bedingungen in Analogie zu OC-ODE als Mehrpunkt-Anfangsrandwertproblem formuliert. Die Dirac-Maße werden durch Sprungbedingungen und ihre analytisch gegebenen Fundamentallösungen ersetzt (welche bei OC-ODE einfache Sprungfunktionen waren, siehe Abschnitt 1.2.2).

Dann versucht man das KKT-System stückweise in der Zeit zu lösen. Findet man eine Lösung, so war die Annahme richtig. Wir haben einen KKT-Punkt gefunden. Findet man keine Lösung, so verändert man die Hypothese und startet neu. Gute Startschätzungen gewinnt man bei OC-ODE entweder aufgrund von Homotopie-Methoden [20] (schrittweises Absenken der oberen Schranke der Zustandsbeschränkung) oder anhand von Ergebnissen aus direkten Verfahren. Beides führt beim Rocketcar je nach Parameterauswahl entweder auf die Schaltfolge Berührungspunkt – Randstück oder auf ein einzelnes Randstück.

Exemplarisch wird im Folgenden ein Mehrpunkt-Anfangsrandwertproblem für den Fall eines Randstückes hergeleitet werden, d.h.  $\mathcal{A} = \bar{R} = [t_{\text{on}}, t_{\text{off}}]$  – da dieser Fall in den Studien des direkten Verfahrens am häufigsten auftritt. Das zugehörige Linienmaß lautet  $\mu(t) = \varrho_{\text{on}} \delta(t - t_{\text{on}}) + \varrho_{\text{off}} \delta(t - t_{\text{off}}) + \sigma(t)$ . Die beiden Schaltzeiten  $t_{\text{on}}, t_{\text{off}} \in \mathbb{R}$  sowie die Sprunghöhen  $\varrho_{\text{on}}, \varrho_{\text{off}} \in \mathbb{R}$  und die Funktion  $\sigma(t) \in C^\infty([t_{\text{on}}, t_{\text{off}}])$  treten als weitere Unbekannte im aufgeteilten KKT-System auf. Die Gleichungen (3.22j), (3.22l) werden analog zu Kap. 1 ersetzt durch

$$T(l/2, t) = T_{\max} \quad \text{für } t \in \mathcal{A} = [t_{\text{on}}, t_{\text{off}}], \quad (3.23a)$$

$$T(l/2, t) < T_{\max} \quad \text{für } t \in (0, t_f) \setminus \mathcal{A}, \quad (3.23b)$$

$$\mu(t) = \varrho_{\text{on}} \delta(t - t_{\text{on}}) + \varrho_{\text{off}} \delta(t - t_{\text{off}}) + \sigma(t), \quad (3.23c)$$

$$\varrho_{\text{on}}, \varrho_{\text{off}} > 0, \quad \sigma(t) \geq 0, \quad (3.23d)$$

$$\sigma(t) = 0 \quad \text{für } t \in (0, t_f) \setminus \mathcal{A}. \quad (3.23e)$$

Die Bedingungen (3.23b), (3.23d) werden erst a posteriori geprüft zur Verifikation der Strukturannahme.

Es bietet sich an, das Mehrpunkt-Anfangsrandwertproblem kombiniert mit einer Transformation auf feste Schaltzeiten umzusetzen. Dadurch kann die aktive Menge wesentlich genauer aufgelöst werden, wodurch insgesamt bessere Ergebnisse erzielt werden können. Daher teilt man die Zeitskala zuerst in drei Bereiche  $[0, t_{\text{on}}]$ ,  $[t_{\text{on}}, t_{\text{off}}]$ ,  $[t_{\text{off}}, t_f]$  auf und belegt jede Größe an den Nahtstellen doppelt (linker und rechter Wert). Die drei Bereiche werden im Folgenden mit I, II, und III nummeriert. Stetigkeits- und Sprungbedingungen fügt man als zusätzliche Gleichungen in das System ein. Zusätzlich wird jeweils auf feste Zeiten transformiert. Die Zeittransformation von Seite 38 ist hierzu zu verallgemeinern auf eine stückweise lineare Transformation auf die Zeitintervalle  $[0, a]$ ,  $[a, b]$  bzw.  $[b, c]$

mit festen  $a, b, c \in \mathbb{R}$  gemäß

$$\text{I)} \quad t = \tau \cdot \frac{t_{\text{on}}}{a}, \quad dt = \frac{t_{\text{on}}}{a} d\tau, \quad \tau \in [0, a], \quad (3.24a)$$

$$\text{II)} \quad t = t_{\text{on}} + \frac{t_{\text{off}} - t_{\text{on}}}{b - a} \cdot (\tau - a), \quad dt = \frac{t_{\text{off}} - t_{\text{on}}}{b - a} d\tau, \quad \tau \in [a, b], \quad (3.24b)$$

$$\text{III)} \quad t = t_{\text{off}} + \frac{t_{\text{f}} - t_{\text{off}}}{c - b} \cdot (\tau - 2), \quad dt = \frac{t_{\text{f}} - t_{\text{off}}}{c - b} d\tau, \quad \tau \in [b, c]. \quad (3.24c)$$

Zusätzlich sind entsprechend die Nebenbedingungen  $0 < t_{\text{on}} < t_{\text{off}} < t_{\text{f}}$  zu berücksichtigen.

Es werden nun alle Übergangsbedingungen aufgelistet: Beim Rocketcar sind die Größen  $p_1, p_2, w_1, w_2, u$  stetig in den Schaltpunkten  $t_{\text{on}}, t_{\text{off}}$  nach Abschnitt 3.4.4. Zusätzlich gelten für  $t_s \in S = \{t_{\text{on}}, t_{\text{off}}\}$ :

$$T(x, t_s^-) = T(x, t_s^+) \quad \forall x \in \Omega, \quad (3.25)$$

$$q(x, t_s^-) = q(x, t_s^+) \quad \forall x \in \Omega \setminus \{l/2\}. \quad (3.26)$$

In den Zentren der Fundamentallösungen lässt sich bei  $q$  keine Stetigkeit notieren. Die Umsetzung der Stetigkeit von  $q$  wird in Abschnitt 3.6.2 separat diskutiert.

Zusätzliche Nahtbedingungen sind nötig, damit die zusammengesetzte Temperaturverteilung die nötige Regularität aus Satz 3.3.4 erfüllt: Die Temperaturverteilung kann in  $x = l/2$  in Zeitrichtung weder springen noch knicken für alle  $t \in [t_E, t_L]$ . Daher gelten

$$T_t(l/2, t_{\text{on}}^-) = 0, \quad (3.27)$$

$$T_t(l/2, t_{\text{off}}^+) = 0. \quad (3.28)$$

Das gesamte Mehrpunkt-Anfangsrandwertproblem wird nun analog zu [76] und zu Kap. 1 aufgeschrieben. Auf jedem Stück ist ein Anfangswertproblem zu lösen mit Anfangsbedingungen gemäß folgender Tabelle, wobei Zustände vorwärts und adjungierte Zustände rückwärts in der Zeit integriert werden:

DGL	AW I:	AW II:	AW III:
$\dot{w}_1 = w_2$	$w_1(0) = w_0$	$w_1(t_{\text{on}}^+) = w_1(t_{\text{on}}^-)$	$w_1(t_{\text{off}}^+) = w_1(t_{\text{off}}^-)$
$\dot{w}_2 = u$	$w_2(0) = \dot{w}_0$	$w_2(t_{\text{on}}^+) = w_2(t_{\text{on}}^-)$	$w_2(t_{\text{off}}^+) = w_2(t_{\text{off}}^-)$
ARWP von $T$	$T(x, 0) = 0$	$T(x, t_{\text{on}}^+) = T(x, t_{\text{on}}^-)$	$T(x, t_{\text{off}}^-) = T(x, t_{\text{off}}^+)$
$-\dot{p}_2 = \dots$	$p_2(t_{\text{on}}^-) = p_2(t_{\text{on}}^+)$	$p_2(t_{\text{off}}^-) = p_2(t_{\text{off}}^+)$	$p_2(t_{\text{f}}) = \text{frei}$
ARWP von $q$	stetig <sup>5</sup>	stetig <sup>5</sup>	$q(x, t_{\text{f}}) = 0$

Tabelle 3.1: Stückweise gegebenes Anfangsrandwertproblem.

$p_1$  ist konstant und stetig, es handelt sich also in allen 3 Intervallen um die gleiche Konstante. Dazu sind in den Intervallen I und III zusätzliche Anfangs- und Endbedingungen zu erfüllen gemäß

<sup>5</sup> Die Stetigkeit von  $q$  wird in Abschnitt 3.6.1 separat diskutiert.

I	III
	$w_1(t_f) = 0$
	$w_2(t_f) = 0$
	$0 = 1 + \lambda u^2(t_f)/2 + u(t_f) p_2(t_f)$
$T_t(l/2, t_{\text{on}}^-) = 0$	$T_t(l/2, t_{\text{off}}^+) = 0$
$H(t_{\text{on}}^-) = H(t_{\text{on}}^+)$	$H(t_{\text{off}}^-) = H(t_{\text{off}}^+)$

Weiterhin ist die Bedingung  $T(l/2, t) = T_{\text{max}} \forall t \in [t_{\text{on}}, t_{\text{off}}]$  zu erfüllen. Dies sind genauso viele zusätzliche Bedingungen, wie es Unbekannte gibt. Die Unbekannten sind  $p_1$ ,  $p_2(t_f)$ ,  $t_{\text{on}}$ ,  $t_{\text{off}}$ ,  $t_f$ ,  $q_{\text{on}}$ ,  $q_{\text{off}}$ , sowie  $\sigma(t)$  auf  $[t_{\text{on}}, t_{\text{off}}]$ . Die letzten beiden Bedingungen an eine passende Hamilton-Funktion sind schwierig aufzustellen, weshalb es einen eigenen Abschnitt erfordert.

Dieses Mehrpunkt-Anfangsrandwertproblem löst man mit geeigneten Methoden. A posteriori prüft man dann die Richtigkeit der Schaltfolge anhand der Bedingungen (3.23b), (3.23d).

### Stetigkeit der Hamilton-Funktion in den Schaltpunkten

Die Stetigkeit der Steuerung in den Schaltzeiten  $t_s \in S$  steckt bereits im System. Daher erhält man an den Schaltzeiten analog zu Gleichung (1.8) jeweils eine andere weitere notwendige Bedingung, welche wir auf drei verschiedene Weisen erhalten können: Zum einen aus dem Maximumprinzip als  $H(t_s^+) = H(t_s^-)$ . Geschickt erhält man diese formal aus der „abstrakten“ Hamilton-Funktionen

$$H := 1 + \frac{\lambda}{2} u^2 + p_2 u + p_1 w_2 - \int_{\Omega} \nabla T \cdot \nabla q + g(w_2) q \, dx - \int_{\Gamma} T q \, ds. \quad (3.29)$$

Unter Ausnutzung der bereits bekannten Stetigkeiten von  $p_1$ ,  $p_2$ ,  $w_2$ ,  $u$ ,  $T$ ,  $T_x$  folgt

$$\begin{aligned} 0 &= H(t_s^+) - H(t_s^-) = - \int_{\Omega} \nabla T(x, t_s) \cdot (\nabla q(x, t_s^+) - \nabla q(x, t_s^-)) \, dx \\ &\quad + \int_{\Omega} g(w_2(t_s)) (q(x, t_s^+) - q(x, t_s^-)) \, dx - \int_{\Gamma} T(x, t_s) (q(x, t_s^+) - q(x, t_s^-)) \, ds \\ &= \int_{\Omega} \left[ \Delta T(x, t_s) + g(w_2(t_s)) \right] (q(x, t_s^+) - q(x, t_s^-)) \, dx. \end{aligned} \quad (3.30)$$

Alternativ geht man nach Kap. 2 vor und arbeitet mit einer aufgeteilten Lagrange-Funktion unter der zusätzlichen Nahtbedingung der Stetigkeit von  $T(x, t)$  in  $t_s$ . Man erhält formal die gleiche Bedingung.

Als dritte Variante lässt sich die Bedingung nach einer Semidiskretisierung im Ort mit der Linienmethode sehr einfach beweisen: Mittels

$$\int_{\Omega} T_t \varphi \, dx + \int_{\Omega} \nabla T \cdot \nabla \varphi \, dx + \int_{\Gamma} T \varphi \, ds = \int_{\Omega} g(w_2(t)) \varphi \, dx \quad (3.31)$$

und einem Finite-Elemente-Galerkinverfahren mit  $T(x, t) = \sum_{i=1}^r \alpha_i(t) \varphi_i(x)$  lautet das ODE-System der Wärmeleitungsgleichung nach der Linienmethode

$$M\dot{\vec{\alpha}} + A\vec{\alpha} = g(w_2(t)) \cdot M\vec{1},$$

mit der Massenmatrix  $M := (\int_{\Omega} \varphi_i \varphi_j dx)_{ij} \in \mathbb{R}^{r \times r}$ , der erweiterten Steifigkeitsmatrix  $A := (\int \nabla \varphi_i \cdot \nabla \varphi_j dx + \int_{\Gamma} \varphi_i \varphi_j ds)_{ij} \in \mathbb{R}^{r \times r}$  und  $\vec{1} := (1, \dots, 1)^T \in \mathbb{R}^r$ . Zusammen mit den restlichen Gleichungen liegt nun ein OC-ODE vor. Je nach Behandlung der Massenmatrix definiert man die Hamilton-Funktion  $H_1$  oder  $H_2$  daher als

$$H_1 := 1 + \frac{\lambda}{2} u^2 + p_2 u + p_1 w_2 + \vec{q}_1^T [g(w_2) \vec{1} - M^{-1} A \vec{\alpha}], \quad (3.32)$$

$$H_2 := 1 + \frac{\lambda}{2} u^2 + p_2 u + p_1 w_2 + \vec{q}_2^T [g(w_2) M \vec{1} - A \vec{\alpha}]. \quad (3.33)$$

Das Maximum-Prinzip aus OC-ODE liefert die Stetigkeit der Hamilton-Funktion an der Schaltzeit  $t_s$  und daher unter Verwendung der bereits bekannten Stetigkeitsbedingungen:

$$0 = [\vec{q}_1(t_s^+) - \vec{q}_1(t_s^-)]^T [g(t_s) \vec{1} - M^{-1} A \vec{\alpha}(t_s)], \quad (3.34)$$

$$0 = [\vec{q}_2(t_s^+) - \vec{q}_2(t_s^-)]^T [g(t_s) M \vec{1} - A \vec{\alpha}(t_s)]. \quad (3.35)$$

Als Unterschied zwischen den beiden Formulierungen erkennt man die Skalierung  $\vec{q}_1 = M \vec{q}_2$ . Der zweite Fall ist identisch mit der Finite-Element-Diskretisierung von (3.30).

### 3.6.2 Lösung des Mehrpunkt-Anfangsrandwertproblems

Zur Lösung eines derartigen Mehrpunkt-Anfangsrandwertproblems bieten sich verschiedene Möglichkeiten an. Zum einen kann man es mittels der Linienmethode in ein nichtlineares Mehrpunkt-Randwertproblem bei ODE transformieren und mit den dort etablierten Methoden lösen, also mit der Mehrzielmethode oder einem Kollokationsverfahren. Zum anderen kann man die Mehrzielmethode auf Mehrpunkt-Anfangsrandwertprobleme verallgemeinern. Diese beiden Varianten lassen sich bekanntlich sehr gut parallelisieren.

Eine dritte Möglichkeit ist es, ein Newton-Verfahren auf eine volle Diskretisierung der KKT-Bedingungen anzuwenden. Die Differentialgleichungen sind dabei auf den einzelnen Intervallen in Zeit und Ort zu diskretisieren (mit finiten Differenzen oder finiten Elementen o.ä.) und alle Bedingungen inklusive der Übergangsbedingungen sind in einem großen nichtlinearen Gleichungssystem zu sammeln. Dieses löst man anschließend mit etablierten Verfahren wie z.B. `fsolve` unter `MATLAB`, also z.B. mit globalisierten Varianten des Newton-Verfahrens.

### Behandlung von $q(x, t)$ und $\mu(t)$ im Mehrpunkt-ARWP

Ziel ist es, die singulären Maßanteile von  $\mu(t)$  analytisch aufzulösen. Genutzt wird die Superposition  $q = q_{\text{on}} + q_{\text{off}} + q_{\sigma}$  mit den beiden Fundamentallösungen  $q_{\text{on}} = \varrho_{\text{on}} \cdot \Phi(x - 1/2, t_{\text{on}} - t)$ ,  $q_{\text{off}} = \varrho_{\text{off}} \cdot \Phi(x - 1/2, t_{\text{off}} - t)$  aus Lemma 3.4.21. Damit  $q$  die adjungierte PDE löst, muss  $q_{\sigma}$  entsprechend Lösung des folgenden ARWP sein:

$$-q_{\sigma,t} - \Delta q_{\sigma} = 0 \quad \text{in } Q, \quad (3.36)$$

$$\partial_{\nu} q_{\sigma} + q_{\sigma} = -\partial_{\nu} (q_{\text{on}} + q_{\text{off}}) - (q_{\text{on}} + q_{\text{off}}) \quad \text{auf } \{0\} \times (0, t_f), \quad (3.37)$$

$$\partial_{\nu} q_{\sigma}(l/2, t) = \sigma(t)/2 \quad \text{auf } (0, t_f), \quad (3.38)$$

$$q_{\sigma}(x, t_f) = 0 \quad \text{in } \Omega. \quad (3.39)$$

Nur dieses ARWP ist zu diskretisieren, da die beiden Fundamentallösungen analytisch gegeben sind. Man formiert sie erst a posteriori. Während der Lösung der KKT-Bedingungen benötigt man zum einen die Randwerte der Fundamentallösungen z.B.

$$\zeta(t) := -\partial_\nu q_{\text{off}} - q_{\text{off}} = \begin{cases} \frac{\varrho_{\text{off}}}{\sqrt{4\pi(t_{\text{off}}-t)}} \cdot \exp\left(-\frac{l^2}{16(t_{\text{off}}-t)}\right) \cdot \left[\frac{l}{4(t_{\text{off}}-t)} - 1\right] & \text{auf } (0, t_{\text{off}}), \\ 0 & \text{auf } [t_{\text{off}}, t_f]. \end{cases} \quad (3.40)$$

Diese Quelle ist zwar theoretisch  $\zeta(t) \in C^\infty([0, t_f])$ , numerisch führt sie dennoch zu Problemen aufgrund der hohen Steigungen, siehe Abb. 3.1.

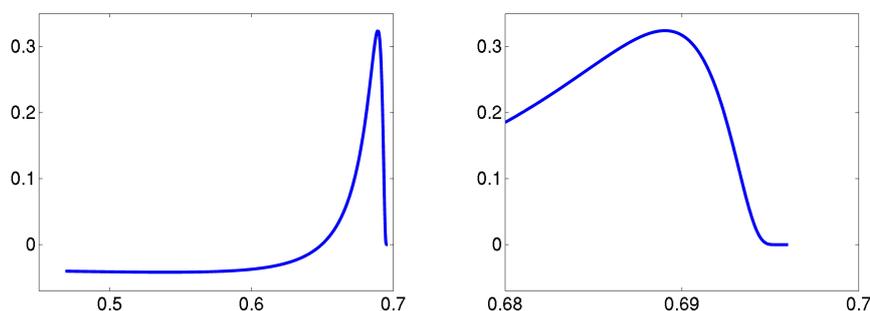


Abbildung 3.1: Glattheit des Quellterms  $\zeta(t)$  aus Gleichung (3.40) für die Parameter aus Abschnitt 3.7. Rechts: Zoom um  $t_{\text{off}}$ .

Zum anderen benötigt man die Ortsintegrale der Fundamentallösungen für die DGL von  $p_2$ , welche bekanntlich auf die Fehlerfunktion (erf) führen:

$$\int q_{\text{on}}(x, t) dx \stackrel{t < t_{\text{on}}}{=} \frac{1}{2} \varrho_{\text{on}} \operatorname{erf}\left(\frac{x - l/2}{2\sqrt{t_{\text{on}} - t}}\right) + \text{const}, \quad (3.41)$$

$$\int_{\Omega} q_{\text{on}}(x, t) dx = \begin{cases} \varrho_{\text{on}} \cdot \operatorname{erf}\left(\frac{l}{4\sqrt{t_{\text{on}} - t}}\right) & t < t_{\text{on}}, \\ \varrho_{\text{on}} & t = t_{\text{on}}^-, \\ 0 & t \geq t_{\text{on}}^+. \end{cases} \quad (3.42)$$

$q_{\text{off}}$  behandelt man analog. Auf diese Weise ist der singuläre Anteil des Maßes und der zugehörige unbeschränkte Anteil des adjungierten Zustandes  $q$  exakt aufgelöst. Folglich kann man sich größere Ortsgitter erlauben.

Wie wir in Kapitel 1 gesehen haben, ist das Mehrpunkt-ARWP aus Sicht der differentialalgebraischen Gleichungen zu untersuchen und gegebenenfalls abzuändern, was abschließend kurz angemerkt werden soll:

### Bemerkung 3.6.1 (DAE-Index des Mehrpunkt-ARWP)

Diskretisiert man das Mehrpunkt-ARWP mit der Linienmethode, so erhält man ein DAE, dessen Differentiationsindex zu untersuchen ist. Sollte er zu hoch sein, bieten sich die Gleichungen (3.13), (3.15) zur Indexreduktion an.

### 3.7 Numerische Experimente

Im Folgenden wird das Rocketcar-Problem mit direkten Verfahren gelöst. Die Ergebnisse werden verifiziert, indem die notwendigen Bedingungen anhand diskreter dualer Variablen aus IPOPT überprüft werden. Zur numerischen Umsetzung werden folgende Parameter gewählt:

$$\begin{aligned} w_0 &= -6, & \dot{w}_0 &= 0, & u_{\max} &= 1, \\ T_{\max} &= 1.5, & \lambda &= 0.1, & g(z) &= z^2. \end{aligned}$$

Der Ort wird äquidistant diskretisiert mit  $N$  Punkten, die Zeit  $\tau \in [0, 1]$  – nach der Transformation auf feste Endzeit – wird äquidistant in  $n_t$  Zeitpunkte unterteilt. Alle Zeitableitungen werden mit dem impliziten Euler-Verfahren diskretisiert, der Laplace-Operator wird mit dem üblichen symmetrischen Differenzenstern behandelt. Damit ergibt sich ein Verfahren der Ordnung 1 in der Zeit, sowie 2 im Ort.

Unter Verwendung des Crank-Nicolson-Verfahrens kann man bekanntlich bei gleichbleibendem Aufwand die Ordnung in der Zeit auf 2 steigern. Dass jedoch Vorsicht geboten ist bei der Verwendung des Crank-Nicolson-Verfahrens in Anwesenheit von Zustandsbeschränkungen zeigt Abschnitt 3.7.2.

Das direkte Verfahren liefert als Lösung diskrete Schätzungen der optimalen Steuerung und der optimalen Zustände. Die Ergebnisse zu den genannten Parametern sind exemplarisch dargestellt in Abb. 3.2. Die Werte der freien Endzeit findet man später in Tabelle 3.2. Weitere Ergebnisse findet man in den Arbeiten [79], [101].

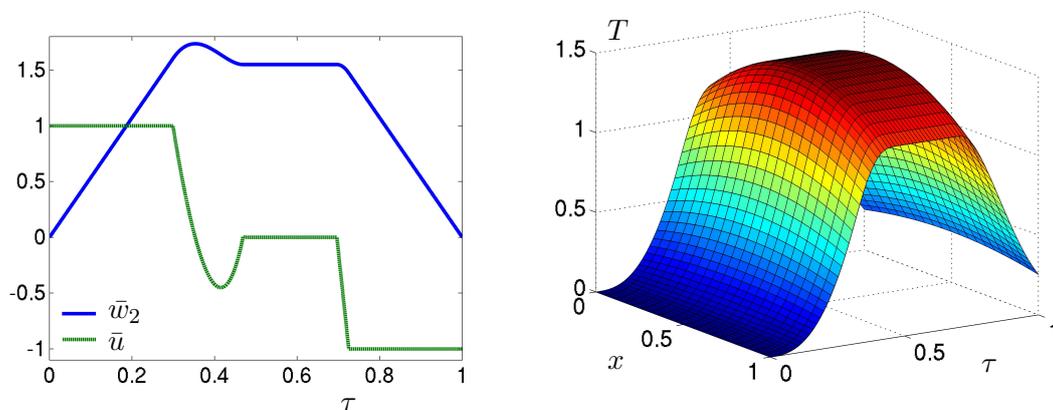


Abbildung 3.2: Ergebnisse des direkten Verfahrens: Optimale Steuerung  $\bar{u}(\tau)$  und optimale Zustände  $\bar{w}_2(\tau)$ ,  $\bar{T}(x, \tau)$  bei  $N = 21$  und  $n_t = 3000$ . Nur jeder 40. Zeitschritt ist dargestellt.

Bei den gewählten Parametern ist die Zustandsbeschränkung genau auf einem Randstück aktiv: Die Temperaturverteilung nimmt in der mittleren Linie  $x = l/2$  die Schranke  $T_{\max} = 1.5$  an. Die Geschwindigkeit  $\bar{w}_2(\tau)$  ist zu den aktiven Zeiten weit vom Stillstand entfernt. Die zur theoretischen Analyse benötigte Annahme 3.4.8 ist also zutreffend. Ebenso werden die vorhergesagten Regularitäten bestätigt: Die Steuerung ist stückweise glatt mit Knickstellen genau an den Auf- und Absprungpunkten der Steuerungs- und der Zustandsbeschränkung.

### 3.7.1 Verifikation der Ergebnisse direkter Verfahren

Direkte Verfahren zeichnen sich durch ihre Anwenderfreundlichkeit aus. Allerdings ist unklar, ob das Ergebnis wirklich optimal ist. Daher empfiehlt es sich, die notwendigen Bedingungen a posteriori abzu prüfen. In [79] führen wir dies für das Rocketcar als OC-ODE aus. Im Folgenden werden nun die notwendigen Bedingungen der Formulierung als OC-PDAE verifiziert.

Zur Verifikation dieser Ergebnisse sollen nun die notwendigen Bedingungen unter MATLAB überprüft werden. Grundlage dafür sind die diskreten Schätzungen der dualen Variablen, welche IPOPT zusätzlich ausgibt. Allerdings verwendet IPOPT zum einen eine entgegengesetzte Vorzeichenkonvention, zum anderen liegen die dualen Größen verschiedener Nebenbedingungen in unterschiedlichen Skalen vor. Dies ist bei der Rekonstruktion der dualen Größen zu beachten.

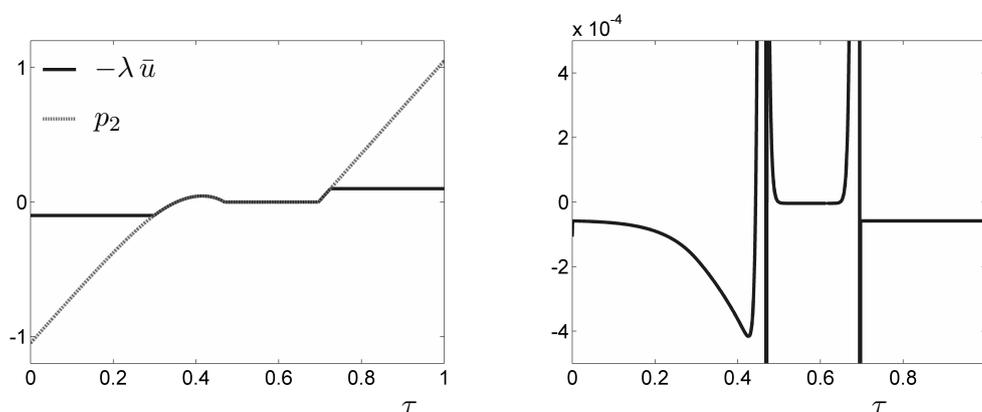


Abbildung 3.3: Links: Verifikation der Projektionsbeziehung: Vergleich von  $-\lambda \bar{u}(\tau)$  und  $p_2(\tau)$ . Rechts: Residuum der ODE von  $p_2$ .  $N = 21$ ,  $n_t = 3000$ .

Unter Beachtung dessen lassen sich die meisten Größen des KKT-Systems direkt angeben. Sehr genau erfüllt sind  $p_1 = \text{const}$ , sowie die Projektionsbeziehung  $\bar{u} = P_{U_{\text{ad}}}(-\frac{1}{\lambda} p_2)$ , siehe Abb. 3.3 (links). Das Residuum von (3.7e) findet man in Abb. 3.3 (rechts), wobei  $\dot{p}_2$  mit numerischer Differentiation und  $\int_{\Omega} q(x, \tau) dx$  mit der Trapezsumme berechnet wurde. Die Differentialgleichung ist also sehr genau erfüllt, obwohl IPOPT sie gar nicht kennt. Lediglich an den Schaltpunkten treten Fehler auf. Grund für das geringe Residuum sind die analytischen Ableitungen, welche AMPL automatisch zur Verfügung stellt.

Die Schätzung des adjungierten Zustandes  $q(x, \tau)$  aus IPOPT findet man in Abb. 3.4. Man erkennt die beiden Fundamentallösungen. Ein Zoom auf die aktiven Zeiten zeigt ein „Hausdach“: Der Zustand zeigt sich in Übereinstimmung mit der Theorie auf beiden Dachhälften beliebig glatt. Am Dachgiebel stellt sich aufgrund des Maßes der erwartete Knick ein.

Die Schaltzeiten und die Sprunghöhen  $\varrho_{\text{on}}$ ,  $\varrho_{\text{off}}$  sind aus den Knicken der Steuerung bzw. aus der Ableitung der Steuerung in Abb. 3.5 herauszulesen. Sie sind zusammen mit der Endzeit und den Endbedingungen der adjungierten ODE-Zustände in Tabelle 3.2 dargestellt – in Abhängigkeit der Diskretisierungsparameter. Man erkennt, dass die meisten Größen nahezu unabhängig von den Diskretisierungsparametern sind: Bereits bei groben Gittern stellen sich gute Ergebnisse ein. Lediglich die Sprunghöhe  $\varrho_{\text{on}}$  ist

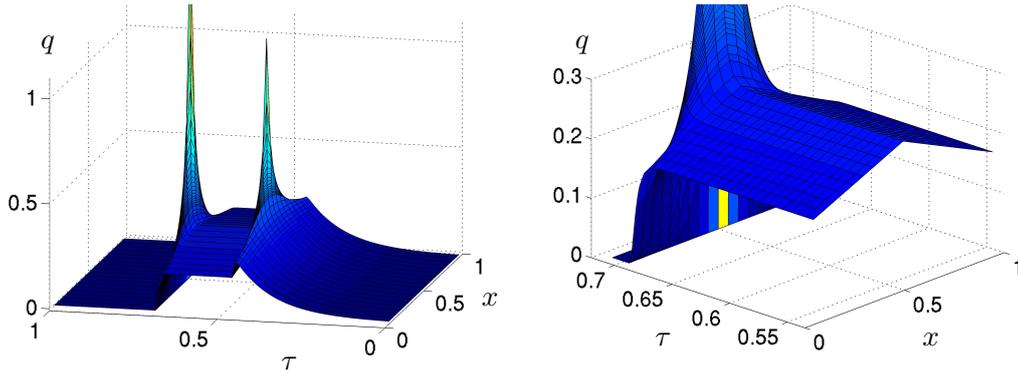


Abbildung 3.4: Diskrete Schätzung von  $q(x, \tau)$  aus dem direkten Verfahren. Rechts: Zoom auf den mittleren Anteil.  $N = 21$ ,  $n_t = 3000$ , nur jeder 40. Zeitschritt ist dargestellt.

sensitiv bzgl. der Zeitdiskretisierung. Die Ursache findet man in Abb. 3.5: Aufgrund der hohen Steigung der Kurve führen bereits kleine Diskretisierungsfehler in  $t_{\text{on}}$  zu großen Fehlern in der Sprunghöhe  $\varrho_{\text{on}}$ . Im Gegensatz dazu kann  $\varrho_{\text{off}}$  als Sprung zwischen zwei nahezu konstanten Funktionen sehr gut bestimmt werden.

$n_t$	$\varrho_{\text{on}}$	$\varrho_{\text{off}}$	$t_f$	$N$	$p_1$	$p_2(t_f)$	$t_{\text{on}}$	$t_{\text{off}}$	$t_f$
51	0.322	1.112	5.347	11	-0.6455	1.051	2.511	3.728	5.357
201	0.543	1.115	5.354	21	-0.6455	1.051	2.512	3.728	5.357
1001	0.592	1.116	5.356	41	-0.6455	1.051	2.512	3.728	5.357
3001	0.601	1.116	5.357						
5001	0.603	1.116	5.357						

Tabelle 3.2: Skalare Größen aus IPOPT bei verschiedenen Diskretisierungen. Links:  $N = 11$  Linien und verschiedenen Zeitdiskretisierungen. Rechts:  $n_t = 3000$  Zeitpunkte und verschiedene Ortsdiskretisierungen.

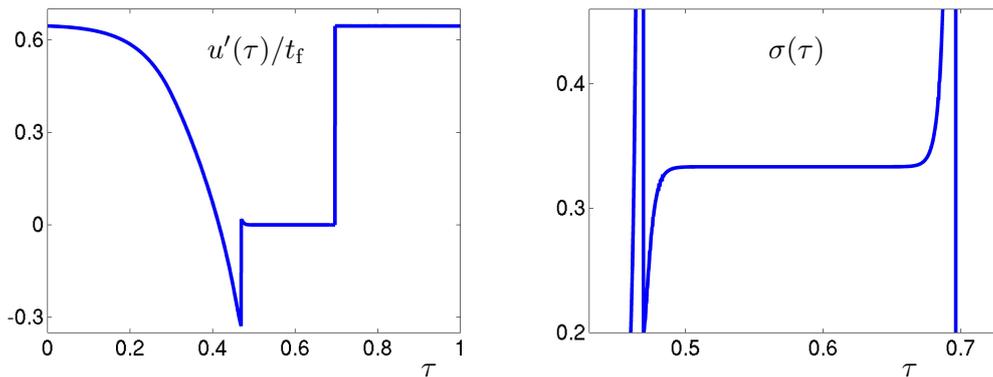


Abbildung 3.5: Links:  $u'(\tau)/t_f$ : Die Sprunghöhen  $\varrho_{\text{on}}$ ,  $\varrho_{\text{off}}$  lassen sich ablesen. Rechts:  $\sigma(\tau)$  aus (3.17).  $N = 11$ ,  $nt = 3000$ .

Das Linienmaß  $\sigma(t)$  wird nicht direkt ausgegeben, sondern ist aus dem Normalenableitungssprung in  $q_x(l/2, t)$  zu berechnen, siehe Abb. 3.5. Außer an den beiden Schaltzeiten erkennt man eine beliebig glatte Funktion in Übereinstimmung mit Satz 3.4.13. Die Vorzeichenbedingung ist erfüllt.

Betrachten wir die weiteren notwendigen Bedingungen bei  $N = 11$  und  $n_t = 3000$ . Die diskreten primalen und dualen Variablen aus IPOPT erfüllen die Transversalitätsbedingung sehr gut: Es ergibt sich  $H(t_f) = -5.5 \cdot 10^{-4}$ . Ebenso klein ist das Residuum der Feedback-Formel (3.13) auf der aktiven Menge, siehe Abb. 3.6. Zuletzt prüfen wir die Massdarstellung in Abb. 3.7. Links sieht man das Linienmaß  $\sigma(\tau)$  einmal aus der Massdarstellung (3.15) und einmal aus dem Normalenableitungssprung von  $q$  in (3.17). Das zweite Bild zeigt die Differenz. Im Inneren des Randstückes ist die Formel sehr gut erfüllt. An den Schaltpunkten kommt es zu Ungenauigkeiten, da die Fundamentallösungen bei nur  $N = 11$  Ortspunkten nicht genau genug aufgelöst werden.

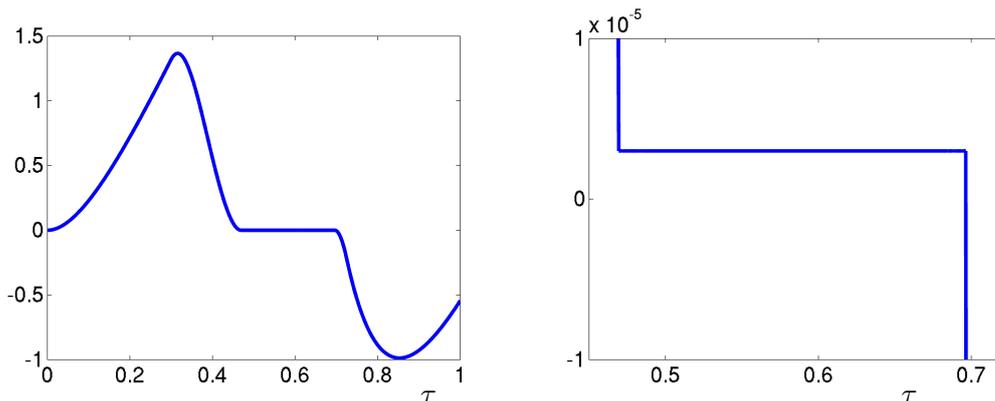


Abbildung 3.6: Residuum der Feedback-Formel (3.13). Rechts: Zoom auf die aktiven Zeiten, zu denen die Formel gelten soll.  $N = 11$ ,  $n_t = 3000$ .

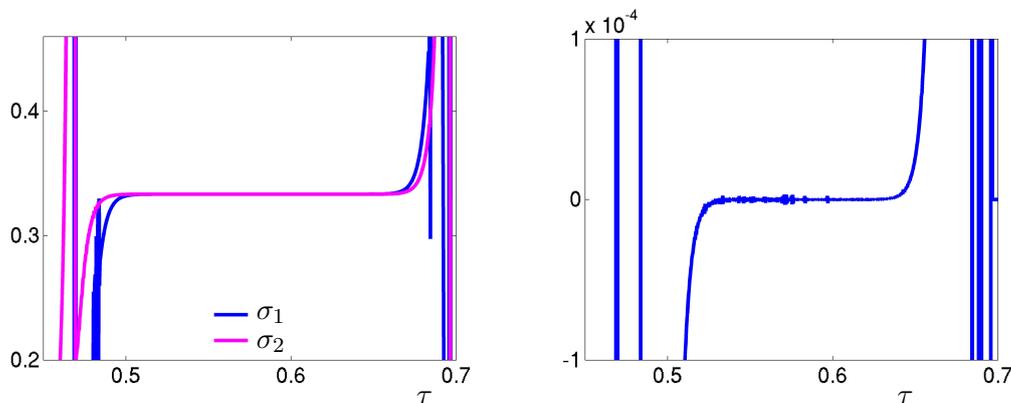


Abbildung 3.7: Links: Linienmaß  $\sigma(\tau)$  zum einen als  $\sigma_1$  aus der Massdarstellung (3.15), zum anderen als  $\sigma_2$  aus dem Normalenableitungssprung in  $q$  aus (3.17). Rechts: Differenz.  $N = 11$ ,  $n_t = 3000$ .

Das direkte Verfahren liefert gute Ergebnisse in kurzen Rechenzeiten. Vorallem die primären Variablen erweisen sich als robust gegenüber den Fehlern in den Schaltzeiten. Einige duale Variablen zeigen jedoch Ungenauigkeiten an den Schaltpunkten, da diese auf dem festen, äquidistanten Gitter nicht exakt aufgelöst werden können und da die Fundamentallösungen nicht exakt behandelt werden. Diese beiden Probleme lassen sich mit den vorgestellten indirekten Methoden aus Abschnitt 3.6 lösen, was jedoch den Rahmen dieser Arbeit sprengt.

### 3.7.2 Das Crank-Nicolson-Verfahren und Zustandsbeschränkungen

Unter gleichbleibendem Aufwand lässt sich die Ordnung des Diskretisierungsverfahrens in der Zeit bekanntlich erhöhen, indem man statt dem impliziten Euler-Verfahren auf das Crank-Nicolson-Verfahren umstellt. Letzteres weist jedoch gerade im Zusammenhang mit Zustandsbeschränkungen einen Nachteil auf, welcher zu berücksichtigen ist: Diskretisiert man innerhalb des direkten Verfahrens die Wärmeleitungsgleichung mit dem Crank-Nicolson-Verfahren, so zeigen einige duale Variablen regelmäßige Oszillationen. In Abb. 3.8 findet man  $\sigma(\tau)$  aus dem Normalenableitungssprung in  $q_x(l/2, \tau)$  bei beiden Verfahren. Der folgende Abschnitt liefert die Erklärung dieser Oszillationen.

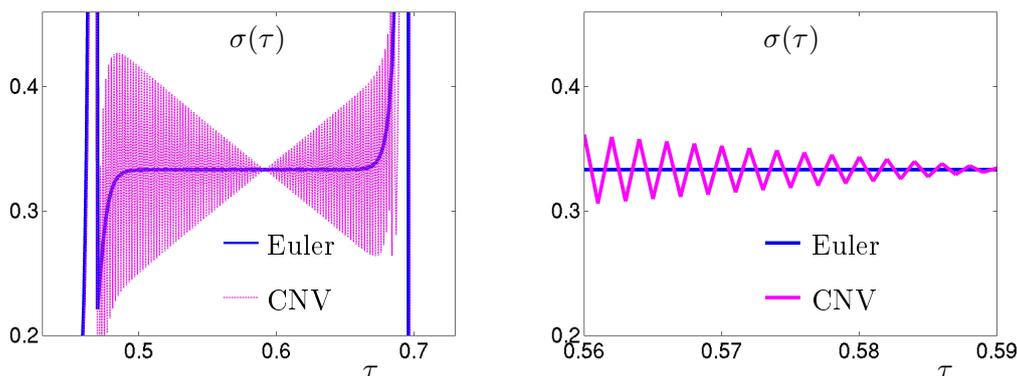


Abbildung 3.8: Rekonstruktion von  $\sigma(\tau)$  aus dem Normalenableitungssprung in  $q$ . Blau: Implizites Euler-Verfahren. Magenta: Crank-Nicolson-Verfahren. Rechts: Zoom.

Das Crank-Nicolson-Verfahren weist bei unstetigen oder nicht differenzierbaren Anfangsbedingungen bekanntlich Oszillationen auf, da es zwar A-stabil, aber nicht stark A-stabil ist ([62, p. 319]). Daher ist es zu dämpfen. Ein solches gedämpftes Crank-Nicolson-Verfahren erreicht man über die sogenannte „Rannacher-Glättung“, welche in der Finanzmathematik stark verbreitet ist ([91], [85]). Bei der Rannacher-Glättung werden die ersten beiden Schritte durch jeweils zwei Schritte des impliziten Euler-Verfahrens ersetzt, siehe auch [43].

Die adjungierte PDE besitzt aufgrund der Zustandsbeschränkung mehrere Zeiten, an welchen das Crank-Nicolson-Verfahren zu oszillieren beginnt. Beim Rocketcar sind dies genau die Schaltzeiten. Bei indirekten Verfahren sind daher die Zeitschritte vor den Schaltzeiten anzupassen.

Bei direkten Verfahren hat man allerdings keinen Zugriff auf die Diskretisierung der adjungierten Gleichungen. Verwendet man für die Zustandsgleichung das Crank-Nicolson-Verfahren, so haben wir in Abb. 3.8 gesehen, dass duale Variablen des direkten Verfahrens nicht gegen die kontinuierlichen Variablen konvergieren müssen. Im vorliegenden Beispiel hat dies nur wenige Auswirkungen auf die Lösung des direkten Verfahrens: Die relative Abweichung in der  $L^2$ -Norm von  $w_1$ ,  $w_2$ ,  $p_2$  bei den beiden Varianten ist jeweils kleiner als  $3 \cdot 10^{-4}$ . Entscheidend wird der Unterschied jedoch, wenn die Multiplikatoren weiterverwendet werden, z.B. als Startschätzung für ein indirektes Verfahren.

Für den allgemeinen Fall empfiehlt es sich daher, die Konvergenz der diskreten gegen die kontinuierlichen dualen Größen sicherzustellen. Für direkte Verfahren mit dem Crank-Nicolson-Verfahren unter Zustandsbeschränkungen ist daher bereits die Zustandsgleichung direkt vor den Schaltzeiten mit der Rannacher-Glättung zu versehen ([91], [102]).

Da man die Lage der Schaltpunkte nicht kennt, ist das Problem zunächst mit dem impliziten Euler-Verfahren zu lösen. Die daraus erhaltenen Näherungen der Schaltpunkte lassen sich zum einen zur Rannacher-Glättung nutzen, zum anderen zu angepassten pseudoadaptiven Gittern. Durch Letzteres werden die Schaltzeiten genauer aufgelöst, wodurch die Genauigkeit der Ergebnisse gesteigert wird.

Bei indirekten Verfahren über die Formulierung als Mehrpunkt-Anfangsrandwertproblem lässt sich die Rannacher-Glättung dagegen kanonisch einbauen, da die Lage der Schaltzeiten dank der Transformation 3.24 bekannt sind.

### 3.8 Zusammenfassung

Das Rocketcar ist ein OC-PDAE mit einer rein zeitabhängigen Steuerung und einer freien Endzeit. Die aktive Menge lebt auf einer Linie. Unter einer bei OC-ODE üblichen Strukturannahme an die aktive Menge kann diese nur aus Berührungspunkten und Randstücken bestehen. Beide Fälle erlauben eine detaillierte theoretische Analyse. Im relativen Inneren des Randstückes findet man sehr hohe Regularitäten. Der Lagrange-Multiplikator der Zustandsbeschränkung lässt sich in Folge dessen vollständig charakterisieren: Er ist stückweise glatt und zeigt nur an den Schaltpunkten Dirac-Impulse. Dort treten daher u.a. Fundamentallösungen der adjungierten Wärmeleitungsgleichung auf. Aufgrund der expliziten Maßdarstellung lassen sich Sprungbedingungen identifizieren, ähnlich zu den bekannten Sprungbedingungen aus OC-ODE.

Nach einer Strukturannahme an die aktive Menge lässt sich das KKT-System als Mehrpunkt-Anfangsrandwertproblem formulieren. Dank der Transformation auf feste Schaltzeiten liegen diese stets auf dem Gitter, weshalb die aktive Menge exakt aufgelöst werden kann. Zusätzlich erlaubt die Verwendung von Fundamentallösungen und von Sprungbedingungen die exakte Auflösung des singulären Anteils des Linienmaßes. Als nächster Schritt ist einer der vorgeschlagenen Mehrpunkt-ARWP-Solver umzusetzen.

Mit geringerem Programmieraufwand lässt sich das Rocketcar-Problem mit einer direkten Methode mittels AMPL und IPOPT lösen, welche sich als sehr effizient erweist. Dank analytischer erster und zweiter Ableitungen ergeben sich kurze Rechenzeiten und gute Ergebnisse. Das OC-PDAE lässt sich sogar unter Zustandsbeschränkungen lösen. Anhand der diskreten Schätzungen der dualen Variablen wird gezeigt, dass die notwendigen Bedingungen erfüllt sind. Lediglich das Linienmaß zeigt lokal um die Schaltpunkte auffällige Fehler. Bei Verwendung des Crank-Nicolson-Verfahrens ist dabei Vorsicht geboten: Es ist zu dämpfen, da andernfalls die diskreten Schätzungen der dualen Variablen zu Oszillationen neigen und daher nicht gegen die kontinuierlichen dualen Variablen konvergieren.

Das direkte Verfahren verwendet allerdings die volle Diskretisierung, weshalb die Anwendbarkeit auf große Anwendungsprobleme fraglich ist. Daher werden im nächsten Kapitel numerische Methoden für hochdimensionale Probleme entwickelt.



## Kapitel 4

# Optimale Steuerung von Schmelzkarbonat-Brennstoffzellen

Das folgende Kapitel widmet sich der Lösung eines komplexen Optimalsteuerungsproblems aus den Anwendungen mittels indirekten und direkten Methoden. Die Problemstellung entstand in einer interdisziplinären Arbeitsgemeinschaft aus Ingenieuren und Mathematikern im Rahmen des BMBF-Projektes „Optimierte Prozessführung von Brennstoffzellen-Systemen mit Methoden der Nichtlinearen Dynamik“ im Zeitraum 2002 - 2005, siehe Sundmacher, Kienle, Pesch et al. [94].

Nach einer Einführung in das Anwendungsgebiet beschäftigt sich Abschnitt 4.2 zunächst mit der Modellierung und einer passenden Diskretisierungs- und Lösungsmethode der Zustandsgleichungen, welche ein PDAE-System formen. In Abschnitt 4.3 wird das Optimalsteuerungsproblem definiert. Anschließend werden die notwendigen Bedingungen erster Ordnung mittels der formalen Lagrange-Technik hergeleitet und diskutiert. In der Folge werden indirekte Verfahren der Optimalsteuerung für das vorliegende OC-PDAE entwickelt. Am effizientesten erweist sich ein Quasi-Newton-Verfahren mit BFGS-Update und kubischer Liniensuche. Wir schließen mit numerischen Experimenten und einem Vergleich der Methoden.

Die Abschnitte 4.2, 4.3, 4.5 und Anhang A beschreiben Teile der einbezogenen Arbeit Rund und Chudej [90]. Zusätzlich findet man in [90] Simulationsergebnisse und Fehlerstudien zu dem in Abschnitt 4.2 beschriebenen Zustandssolver, welche in diesem Kapitel nicht wiederholt werden.

Die Abschnitte 4.4, 4.6, 4.7.3 und Anhang B sind dagegen alleinig im Rahmen dieser Dissertation entstanden.

Im Rahmen dieses Kapitels hat der Autor unter MATLAB ein Programmpaket von ca. 5.000-10.000 Code-Zeilen entwickelt, auf welches in der Kürze der Arbeit nicht eingegangen werden kann. Lediglich die Auswahl der Algorithmen wird beschrieben.

### 4.1 Hintergründe zu Brennstoffzellen

Die erste Erforschung und Nutzung von Brennstoffzellen geht auf militärische Einsatzgebiete vor allem in U-Booten und auf die Raumfahrt zurück, Bereiche in denen Kosten oft eine untergeordnete Rolle spielen. Die hohen Erwartungen, welche man in die Energiebe-

reitstellung durch Brennstoffzellen legt, begründen sich mit dem umfassenden Potential, mit dieser Technologie weite Bereiche der Energiebereitstellung zugleich zu reformieren.

Brennstoffzellen können stationär und mobil eingesetzt werden. Bei dem mobilen Einsatz stellen neben der Effizienz auch die Größe und das Gewicht der Anlage sowie die Speicherung des Brennstoffes besondere Aspekte dar. Ein wichtiges stationäres Einsatzgebiet ist die dezentrale Kraft-Wärme-Kopplung (KWK). Sie verspricht Energieeinsparungen und Emissionsminderungen. So nutzt man in Blockheizkraftwerken Strom und Abwärme einer Brennstoffzellenanlage und erreicht damit Gesamt-Wirkungsgrade bis ca. 90% [94].

In der Knallgasreaktion fusioniert Wasserstoff und Sauerstoff explosionsartig unter freier werdender Energie zu Wasser. Eine vergleichbare Reaktion läuft in der Brennstoffzelle ab, jedoch nicht explosionsartig sondern langsam und kontrolliert. Die beiden Teilreaktionen, Oxidation des Brennstoffes unter Abgabe von Elektronen und Reduktion des Oxidationsmittels unter Aufnahme von Elektronen, finden räumlich getrennt statt. Der Elektronenstrom zwischen diesen beiden Systemen ist für äußere Lasten nutzbar: Die Brennstoffzelle stellt elektrische Energie bereit.

#### 4.1.1 Typen von Brennstoffzellen

Man unterscheidet verschiedene Arten von Brennstoffzellen, welche sich in ihrer Betriebsart und damit in ihrem Anwendungsbereich stark unterscheiden. Für den mobilen Einsatz gibt es die Niedertemperatur-Brennstoffzellen. Zu den Niedertemperatur-Brennstoffzellen gehören die alkalische Brennstoffzelle (Alkaline Fuel Cell, AFC) und zwei Typen der Polymer-Membran-Brennstoffzelle (Polymer Exchange Membrane Fuel Cell, PEMFC, Direct Methanol Fuel Cell, DMFC). Die historische Vorreiterrolle spielt die AFC, welche schon in Apollo-Missionen, in Space Shuttles und in U-Booten der deutschen Marine zum Einsatz kam.

Im mittleren Temperaturbereich von 200°C laufen die sogenannten Phosphorsäure-Brennstoffzellen (Phosphoric Acid Fuel Cell, PAFC), benannt nach ihrem Elektrolyt. Als Hochtemperatur-Brennstoffzellen bezeichnet man zum einen die Schmelzkarbonat-Brennstoffzelle (Molten Carbonate Fuel Cell, MCFC) mit einer Betriebstemperatur von ca. 650°C, zum anderen Festoxid-Brennstoffzellen (Solid Oxide Fuel Cell, SOFC), welche bei 800 – 1000°C arbeiten. Hochtemperatur-Brennstoffzellen zeichnen sich durch die höchsten Wirkungsgrade aus. Allerdings erfordern die hohen Temperaturen große Systemausdehnungen, weshalb sie sich besonders zum stationären Einsatz eignen.

#### 4.1.2 Die Schmelzkarbonat-Brennstoffzelle (MCFC)

Schmelzkarbonat-Brennstoffzellen setzt man in Kleinkraftwerken oder zur dezentralen Wärme- und Energieversorgung ein. Die MCFC profitiert von ihrem hohen Wirkungsgrad, ihren schadstoffarmen Emissionen und der Flexibilität in der Art des Brennstoffes. Allerdings reagieren Schmelzkarbonat-Brennstoffzellen-Anlagen nur langsam auf veränderte Anforderungen und eignen sich daher primär zur Deckung der Grundlasten. Beim Hoch- und Herunterfahren der Anlage ist äußerste Vorsicht geboten, da hier bei falscher Steuerung Materialkorrosionen durch die Bildung von „Hotspots“ drohen, welche zu großen finanziellen Schäden führen können [94].

Blockheizkraftwerke mit einer MCFC-Einheit werden z.B. von zwei Firmen hergestellt. Zum einen gibt es das „Hotmodule“ der Firma *MTU Onsite Energy Fuel Cell Systems* bzw. der übergeordneten *Tognum AG* mit Sitz in Friedrichshafen. Neuere Anlagen dieses Typs stehen laut Herstellerangaben z.B. seit 2009 bei der Brauerei Erdinger Weißbräu, seit 2009 am Universitätsklinikum Gießen, seit 2010 beim Automobilzulieferer Reich GmbH oder seit November 2010 in der Heizzentrale Grünau in Zürich-Altstetten (Elektrizitätswerk der Stadt Zürich). Allerdings ist die Serienproduktion der Anlage kürzlich gescheitert, da der asiatische Kooperationspartner am 28.12.2010 vom Projekt abgesprungen war [3]. Produziert und vermarktet werden ähnliche Anlagen auch vom amerikanischen, NASDAQ-gelisteten Konkurrenten *FuelCell Energy*.

Das Modell, welches wir im nächsten Abschnitt aufgreifen werden, berücksichtigt alle wesentlichen Effekte, welche innerhalb einer MCFC ablaufen. Zusätzlich wurden die Peripherie sowie die Führung der Gasströme des Hotmodule abgebildet. Das Modell wurde kalibriert und verifiziert an einem Hotmodule der *IPF Heizkraftwerksbetriebsgesellschaft mbH* Magdeburg [45].

Das Herz des Hotmodule bildet ein Stapel aus ca. 300 Schmelzkarbonat-Brennstoffzellen, welche jeweils eine Fläche von  $0.8\text{ m}^2$  aufweisen [94]. Das Hotmodule hat eine Leistung ab 250 kW aufwärts und erreicht dabei einen elektrischen Wirkungsgrad von ca. 50%; mit Nutzung der Abwärme gibt der Hersteller sogar erreichbare Gesamtwirkungsgrade von ca. 90% an [94].

## 4.2 Das Modell und seine numerische Lösung

Aus dem Stapel der Schmelzkarbonat-Brennstoffzellen wird eine mittlere Zelle herausgegriffen und modelliert. Die verschiedenen Schichten der Zelle werden aufgrund ihrer vergleichsweise geringen Dicke zweidimensional modelliert. Die einzelnen Schichten interagieren miteinander in Richtung der dritten Dimension. Das vollständige Modell findet sich in den Dissertationen von Heidebrecht [51] und Sternberg [93]. Es liegt in verschiedenen Komplexitäten vor, so dass eine Hierarchie von Modellen zur Verfügung steht.

Um das Modell für indirekte Methoden der Optimalsteuerung von PDAE-Systemen zugänglich zu machen und um die Hierarchie zu erweitern, reduzieren wir das Modell in Rund und Chudej [90] auf örtlich eindimensional. Da ein Kreuzstrom-Modell wie in [51] in 1D nicht in Frage kommt, entscheiden wir uns aufbauend auf Erkenntnissen aus [93] für das Gegenstrom-Modell, d.h. die beiden Gasströme laufen in entgegengesetzter Richtung. Wir beschreiben nun kurz das Modell. Mehr Einzelheiten zur Modellierung findet der Leser in den Arbeiten [51, 93].

Abb. 4.1 zeigt schematisch den Aufbau der MCFC. Als Ortskoordinate wählen wir  $z \in [0, 1]$ , und lassen das Anodengas von links nach rechts strömen; das Kathodengas strömt folglich von  $z = 1$  nach  $z = 0$ . Die Zeit wird mit  $t \in [0, t_f]$  bezeichnet, der Orts-Zeit-Zylinder mit  $Q := (0, 1) \times (0, t_f)$ .

Das Herz der Brennstoffzelle besteht aus einem Galvanischen Element. Die beiden Elektroden sind getrennt von einem speziellen Elektrolyt (Index „e“ bzw. „s“), welcher nur Carbonat-Ionen  $\text{CO}_3^{2-}$  passieren lässt. Strom wird produziert, wenn die Anode mit Wasserstoff und die Kathode mit Sauerstoff versorgt werden. Dazu dienen die angelegten Gaskanäle. Der Anodengaskanal (Index „a“) wird mit einem Gemisch aus  $\text{CH}_4$  und  $\text{H}_2\text{O}$

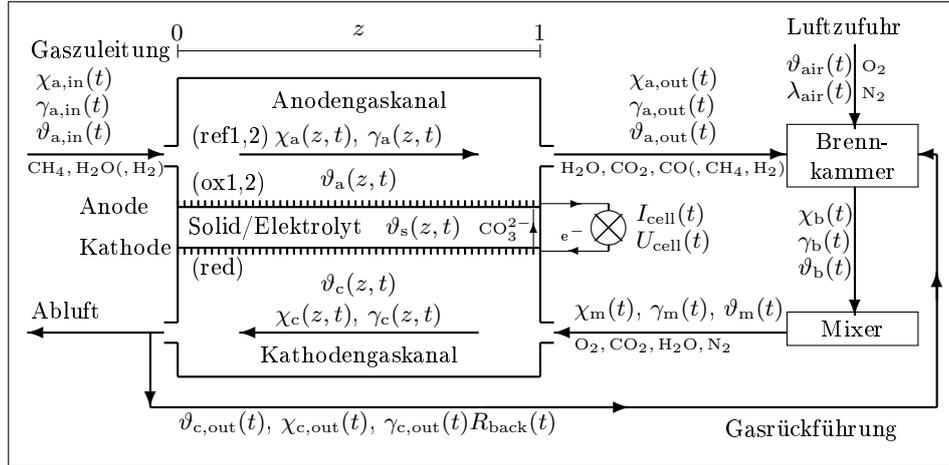
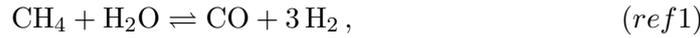


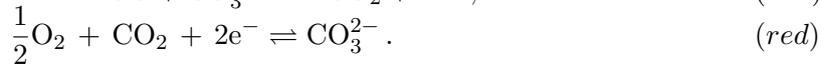
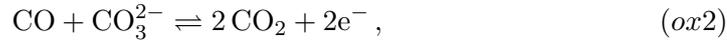
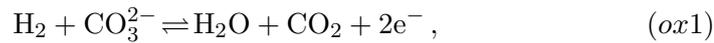
Abbildung 4.1: 1D Gegenstrom-Modell der MCFC (Grafik aus [90])

gespeist - plus möglicherweise ein Anteil  $H_2$  aus externer Reformierung. Aufgrund hoher Temperaturen findet innerhalb des Gaskanals interne Reformierung statt, d.h.  $H_2$  wird erzeugt durch die zwei Reduktionsreaktionen



Die Gesamtreaktion der Reformierung ist endotherm, verbraucht also Energie. Der Anodengas-Ausgang bei  $z = 1$  wird in eine Brennkammer (Index „b“) geführt, mit Luft gemischt und vollständig oxidiert, um anschließend über eine Mischkammer (Index „m“) in den Kathodengaskanal (Index „c“) zu münden. Am Kathodenausgang  $z = 0$  tritt warme, schadstofffreie Abluft aus, bestehend aus Luft,  $H_2O$  und  $CO_2$ . Um die Restwärme zu nutzen und die Prozesstemperatur zu stabilisieren, wird ein Teil dieser Abluft in die Brennkammer zurückgeleitet.

Der Strom wird aufgrund der beiden Oxidations-Reaktionen an der Anode sowie der Reduktionsreaktion an der Kathode produziert



Insgesamt treten im System sieben chemische Verbindungen auf. Sie werden zusammengefaßt in der Indexmenge  $\mathcal{I} := \{CH_4, H_2O, H_2, CO, CO_2, O_2, N_2\}$ .

Die auftretenden Zustände sind die Stoffstromdichte  $\gamma(z, t)$ , die Molenbrüche  $\chi_k(z, t)$  der einzelnen Verbindungen  $k \in \mathcal{I}$ , die Temperaturen  $\vartheta(z, t)$ , die Stromdichten  $i(z, t)$  und Ströme  $I(t)$ , die Potentiale  $\Phi(z, t)$  sowie die Zellspannung  $U_{cell}(t)$ .<sup>1</sup> Die Input-Variablen sind der Zellstrom  $I_{cell}(t)$ , welcher die äußere Last beschreibt, sowie die Konfiguration der in den Brennstoffzellenstapel einfließenden Gasströme (Eingänge von Anode und Brennkammer), welche wir im Abschnitt 4.3 als Steuerungen im Rahmen der Definition des Optimalsteuerungsproblems auswählen werden. Unter Beachtung der geänderten Vorzeichen der Ortsableitungen erhalten wir das PDAE-System in Abbildung 4.2.<sup>2</sup>

<sup>1</sup>  $(\gamma, \chi_k, \vartheta)$  der Brenn- und der Mischkammer hängen in der Modellierung nur von  $t$  ab.

<sup>2</sup> Die Variablen  $(z, t)$  werden der Übersichtlichkeit halber nur aufgeführt, wo es nicht aus dem Kontext hervorgeht.

<p>Gaskanäle und Solid:</p> $\frac{\partial \chi_{k,a}}{\partial t} + \gamma_a \vartheta_a \frac{\partial \chi_{k,a}}{\partial z} = f^{\chi_{k,a}}, \quad k \in \mathcal{I},$ $\frac{\partial \chi_{k,c}}{\partial t} - \gamma_c \vartheta_c \frac{\partial \chi_{k,c}}{\partial z} = f^{\chi_{k,c}}, \quad k \in \mathcal{I},$ $\frac{\partial \vartheta_a}{\partial t} + \gamma_a \vartheta_a \frac{\partial \vartheta_a}{\partial z} = f^{\vartheta_a},$ $\frac{\partial \vartheta_c}{\partial t} - \gamma_c \vartheta_c \frac{\partial \vartheta_c}{\partial z} = f^{\vartheta_c},$ $\frac{\partial(\gamma_a \vartheta_a)}{\partial z}(z, t) = f^{\gamma_a \vartheta_a},$ $-\frac{\partial(\gamma_c \vartheta_c)}{\partial z}(z, t) = f^{\gamma_c \vartheta_c},$ $c_{p,s} \frac{\partial \vartheta_s}{\partial t} - \frac{1}{\text{Pe}} \frac{\partial^2 \vartheta_s}{\partial z^2} = f^{\vartheta_s},$	<p>Randbedingungen:</p> $\chi_{k,a}(z=0, t) = \chi_{k,a,\text{in}}(t),$ $\chi_{k,c}(z=1, t) = \chi_{k,m}(t),$ $\vartheta_a(z=0, t) = \vartheta_{a,\text{in}}(t),$ $\vartheta_c(z=1, t) = \vartheta_m(t),$ $\gamma_a(z=0, t) = \gamma_{a,\text{in}}(t),$ $\gamma_c(z=1, t) = \gamma_m(t),$ $\frac{\partial \vartheta_s}{\partial z}(z=0, t) = \frac{\partial \vartheta_s}{\partial z}(z=1, t) = 0,$
<p>Brenn- und Mischkammer:</p> $\chi_{k,b}(t) = f^{\chi_{k,b}}, \quad k \in \mathcal{I},$ $\vartheta_b(t) = f^{\vartheta_b},$ $\gamma_b(t) = f^{\gamma_b},$ $\gamma_{\text{air}}(t) = f^{\gamma_{\text{air}}},$ $\frac{d\chi_{k,m}}{dt}(t) = f^{\chi_{k,m}}, \quad k \in \mathcal{I},$ $\frac{d\vartheta_m}{dt}(t) = f^{\vartheta_m},$ $\gamma_m(t) = f^{\gamma_m},$	<p>Anfangsbedingungen:</p> $\chi_{k,a}(z, t=0) = \chi_{k,a,0}(z),$ $\chi_{k,c}(z, t=0) = \chi_{k,c,0}(z),$ $\vartheta_a(z, t=0) = \vartheta_{a,0}(z),$ $\vartheta_c(z, t=0) = \vartheta_{c,0}(z),$ $\vartheta_s(z, t=0) = \vartheta_{s,0}(z),$ $\chi_{k,m}(t=0) = \chi_{k,m,0},$ $\vartheta_m(t=0) = \vartheta_{m,0},$
<p>Stromdichten:</p> $i_a(z, t) = f^{i_a},$ $i_c(z, t) = f^{i_c},$ $i_e(z, t) = f^{i_e},$ $i(z, t) = f^i,$	<p>Stromstärken:</p> $I_a(t) = \int_0^1 i_a \, dz,$ $I_c(t) = \int_0^1 i_c \, dz,$ $I_e(t) = \int_0^1 i_e \, dz,$
<p>Potentiale:</p> $\frac{dU_{\text{cell}}}{dt}(t) = f^{U_{\text{cell}}},$ $\frac{\partial \Phi_a^L}{\partial t}(z, t) = f^{\Phi_a^L},$ $\frac{\partial \Phi_c^L}{\partial t}(z, t) = f^{\Phi_c^L},$	<p>Anfangsbedingungen:</p> $U_{\text{cell}}(t=0) = U_{\text{cell},0},$ $\Phi_a^L(z, t=0) = \Phi_{a,0}^L(z),$ $\Phi_c^L(z, t=0) = \Phi_{c,0}^L(z).$

Abbildung 4.2: PDAE-System der Brennstoffzelle (1D-Gegenstrom-Modell)

Dabei sind alle Quellterme mittels Funktionen  $f^q$  abgekürzt, wobei der Index  $q$  für den entsprechenden Zustand steht. Die Quellterme sind längliche Funktionen von den Zuständen. Das vollständige Modell findet sich in [90] und im Anhang A.

In den Gaskanälen treten nichtlineare Advektions-Reaktionsgleichungen auf, also nichtlineare hyperbolische partielle Differentialgleichungen 1. Ordnung. Jeweils eine dieser Gleichungen hat eine idealisierte Zeitkonstante und degeneriert daher: Bei den Stoffstromdichten verschwindet der Summand der Zeitableitung. Bei Anwendung der Linienmethode wird dies im nächsten Abschnitt algebraische Gleichungen erzeugen. Weiterhin arbeiten wir mit einer Wärmeleitungsgleichung für die Solidtemperatur, sowie diversen gewöhnlichen Differentialgleichungen, algebraischen Gleichungen, im Ort degenerierten partiellen Differentialgleichungen (Potentiale  $\Phi_a^L, \Phi_c^L$ ) sowie Integralausdrücken. Insgesamt bildet dies ein vollständig gekoppeltes PDAE-Modell.

#### 4.2.1 Semidiskretisierung im Ort mit Differenzenverfahren

In [90] zeigen wir, dass die Zustände sehr glatt in Ortsrichtung sind, jedoch in Zeitrichtung große Gradienten und unterschiedliche Zeitkonstanten aufweisen. Anhand Fehlerstudien zeigen wir, dass eine Adaptivität in der Zeit dringend erforderlich ist: So übertrifft ein adaptives Zeitgitter mit 57 Zeitschritten bereits ein äquidistantes Zeitgitter mit 2000 Zeitschritten um Längen. Für den Ort genügen äquidistante Gitter; adaptive Ortsgitter erzielten in ersten Untersuchungen keine wesentlich besseren Ergebnisse. Daher entscheiden wir uns für eine Semidiskretisierung mit der Linienmethode. Dies erlaubt die Verwendung von etablierten, adaptiven DAE-Solvern wie z.B. `ode15s` unter MATLAB. Wir wählen also ein äquidistantes Ortsgitter  $z^{(i)} = (i - 1) \cdot h, i = 1, \dots, N$ , mit Schrittweite  $h = 1/(N - 1)$ . Die Funktionen  $f(z_i, t)$  werden approximiert durch ihre Werte auf den Linien  $f^{(i)}(t)$ .

Dennoch ist die Ortsdiskretisierung mit Bedacht zu wählen, da es sich um hyperbolische PDE handelt. Zur Sicherstellung der Stabilität sind unter den Differenzenverfahren die Upwind-Verfahren zu wählen, bei denen ein einseitiger Differenzenquotient entgegen der Flussrichtung verwendet wird [49, S.775]. Für die Anode ist daher ein Rückwärts-Differenzenquotient, für die Kathode ein Vorwärts-Differenzenquotient zu verwenden. Bei 2D- oder 3D-Problemen wird man dann i.d.R. auf Finite-Volumen-Methoden zurückgreifen (siehe z.B. [65]).

Bei den Erhaltungsgleichungen müssen die Upwind-Verfahren direkt auf die zugrundeliegende Erhaltungsform angewandt werden, damit sich das Erhaltungs-Prinzip auf die diskreten Gleichungen überträgt. Daher sind die Transportgleichungen wie in [93, Anhang C] zu diskretisieren. Die Koeffizienten werden dabei wie üblich am aktuellen Punkt ausgewertet, außer die Stoffstromdichten  $\gamma$ , welche am zurückliegenden Ortspunkt auszuwerten sind. Betroffen sind die PDE für  $\chi_{k,a}, \chi_{k,c}, \theta_a, \theta_c$ . Angewandt auf das Gegenstrom-Modell in 1D ergibt sich folgende Diskretisierung:

$$\frac{1}{\vartheta_a^{(i)}} \frac{\partial \chi_{k,a}^{(i)}}{\partial \tau} = -\gamma_a^{(i-1)} \frac{\chi_{k,a}^{(i)} - \chi_{k,a}^{(i-1)}}{h} + \sum_{j=\text{ox1,2,ref1,2}} (\nu_{k,j} - \chi_{k,a}^{(i)} \bar{\nu}_j) Da_j r_j^{(i)}, \quad (4.1a)$$

$$\frac{1}{\vartheta_c^{(i)}} \frac{\partial \chi_{k,c}^{(i)}}{\partial \tau} = \gamma_c^{(i+1)} \frac{\chi_{k,c}^{(i+1)} - \chi_{k,c}^{(i)}}{h} + (\nu_{k,\text{red}} - \chi_{k,c}^{(i)} \bar{\nu}_{\text{red}}) Da_{\text{red}} r_{\text{red}}^{(i)}. \quad (4.1b)$$

<p>Gaskanäle und Solid:</p> $\frac{\partial \chi_{k,a}^{(i)}}{\partial t} + \gamma_a^{(i-1)} \vartheta_a^{(i)} \frac{\chi_{k,a}^{(i)} - \chi_{k,a}^{(i-1)}}{h} = f^{\chi_{k,a}}(z^{(i)}, t), \quad k \in \mathcal{I},$ $\frac{\partial \chi_{k,c}^{(i)}}{\partial t} - \gamma_c^{(i+1)} \vartheta_c^{(i)} \frac{\chi_{k,c}^{(i+1)} - \chi_{k,c}^{(i)}}{h} = f^{\chi_{k,c}}(z^{(i)}, t), \quad k \in \mathcal{I},$ $\frac{\partial \vartheta_a^{(i)}}{\partial t} + \gamma_a^{(i-1)} \vartheta_a^{(i)} \frac{\vartheta_a^{(i)} - \vartheta_a^{(i-1)}}{h} = f^{\vartheta_a}(z^{(i)}, t),$ $\frac{\partial \vartheta_c^{(i)}}{\partial t} - \gamma_c^{(i+1)} \vartheta_c^{(i)} \frac{\vartheta_c^{(i+1)} - \vartheta_c^{(i)}}{h} = f^{\vartheta_c}(z^{(i)}, t),$ $\frac{\gamma_a^{(i)} \vartheta_a^{(i)} - \gamma_a^{(i-1)} \vartheta_a^{(i-1)}}{h} = f^{\gamma_a \vartheta_a}(z^{(i)}, t),$ $-\frac{\gamma_c^{(i+1)} \vartheta_c^{(i+1)} - \gamma_c^{(i)} \vartheta_c^{(i)}}{h} = f^{\gamma_c \vartheta_c}(z^{(i)}, t),$ $c_{p,s} \frac{\partial \vartheta_s^{(i)}}{\partial t} - \frac{\vartheta_s^{(i+1)} - 2\vartheta_s^{(i)} + \vartheta_s^{(i-1)}}{Pe \cdot h^2} = f^{\vartheta_s}(z^{(i)}, t),$	<p>Randbedingungen:</p> $\chi_{k,a}^{(1)}(t) = \chi_{k,a,\text{in}}(t),$ $\chi_{k,c}^{(N)}(t) = \chi_{k,m}(t),$ $\vartheta_a^{(1)}(t) = \vartheta_{a,\text{in}}(t),$ $\vartheta_c^{(N)}(t) = \vartheta_m(t),$ $\gamma_a^{(1)}(t) = \gamma_{a,\text{in}}(t),$ $\gamma_c^{(N)}(t) = \gamma_m(t),$ $\frac{\vartheta_s^{(2)} - \vartheta_s^{(0)}}{2h} = 0 = \frac{\vartheta_s^{(N+1)} - \vartheta_s^{(N-1)}}{2h},$
<p>Brenn- und Mischkammer:</p> $\chi_{k,b}(t) = f^{\chi_{k,b}}, \quad k \in \mathcal{I},$ $\vartheta_b(t) = f^{\vartheta_b},$ $\gamma_b(t) = f^{\gamma_b},$ $\gamma_{\text{air}}(t) = f^{\gamma_{\text{air}}},$ $\frac{d\chi_{k,m}}{dt}(t) = f^{\chi_{k,m}}, \quad k \in \mathcal{I},$ $\frac{d\vartheta_m}{dt}(t) = f^{\vartheta_m},$ $\gamma_m(t) = f^{\gamma_m},$	<p>Anfangsbedingungen:</p> $\chi_{k,a}^{(i)}(t=0) = \chi_{k,a,0}(z^{(i)}),$ $\chi_{k,c}^{(i)}(t=0) = \chi_{k,c,0}(z^{(i)}),$ $\vartheta_a^{(i)}(t=0) = \vartheta_{a,0}(z^{(i)}),$ $\vartheta_c^{(i)}(t=0) = \vartheta_{c,0}(z^{(i)}),$ $\vartheta_s^{(i)}(t=0) = \vartheta_{s,0}(z^{(i)}),$ $\chi_{k,m}(t=0) = \chi_{k,m,0},$ $\vartheta_m(t=0) = \vartheta_{m,0},$
<p>Stromdichten:</p> $i_a^{(i)}(t) = f^{i_a}(z^{(i)}, t),$ $i_c^{(i)}(t) = f^{i_c}(z^{(i)}, t),$ $i_e^{(i)}(t) = f^{i_e}(z^{(i)}, t),$ $i(t)^{(i)} = f^i(z^{(i)}, t),$	<p>Stromstärken: Quadraturformeln</p> $I_a(t) = \sum_{n=1}^N \alpha_n \cdot i_a^{(n)},$ $I_c(t) = \sum_{n=1}^N \alpha_n \cdot i_c^{(n)},$ $I_e(t) = \sum_{n=1}^N \alpha_n \cdot i_e^{(n)},$
<p>Potentiale:</p> $\frac{dU_{\text{cell}}}{dt}(t) = f^{U_{\text{cell}}},$ $\frac{\partial \Phi_a^L}{\partial t}(t) = f^{\Phi_a^L}(z^{(i)}, t),$ $\frac{\partial \Phi_c^L}{\partial t}(t) = f^{\Phi_c^L}(z^{(i)}, t),$	<p>Anfangsbedingungen:</p> $U_{\text{cell}}(t=0) = U_{\text{cell},0},$ $\Phi_a^L(t=0) = \Phi_{a,0}^L(z^{(i)}),$ $\Phi_c^L(t=0) = \Phi_{c,0}^L(z^{(i)}).$

Abbildung 4.3: Semidiskretisierung des PDAE-Systems der Brennstoffzelle

In der Anode ist also der Rückwärts-Differenzenquotient zu verwenden, in der Kathode dagegen der Vorwärts-Differenzenquotient. Für die entarteten PDE für  $\gamma_a \vartheta_a$  bzw.  $\gamma_c \vartheta_c$  sind die Upwind-Verfahren direkt am Produkt auszuführen. Die gesamte Semidiskretisierung ist dargestellt in Abb. 4.3, wobei jeweils  $i = 1, \dots, N$  gilt.<sup>3</sup>

Die homogene Neumann-Randbedingung an  $\vartheta_s$  wird mit höherer Ordnung aufgelöst vgl. [49, S.635]:

$$\begin{aligned} \frac{\partial^2 \vartheta_s}{\partial z^2}(z=0) &= \frac{2\vartheta_s^2 - 2\vartheta_s^1}{h^2}, \\ \frac{\partial^2 \vartheta_s}{\partial z^2}(z=1) &= \frac{2\vartheta_s^{N-1} - 2\vartheta_s^N}{h^2}. \end{aligned} \quad (4.2)$$

Die Ortsintegrale der Stromstärken werden mit einer Quadraturformel diskretisiert, z.B. der Trapezsumme oder der Simpsonsumme mit entsprechenden Gewichten  $\alpha_n$ . Durch maximale Elimination von Gleichungen (identisch verschwindende Molenbrüche, lineare explizite algebraische Gleichungen, Integrale nach Quadraturformeln) erhalten wir ein System aus 14 PDE und 5 ODE (siehe [90]). Die Linien der Dirichlet-Bedingungen werden eliminiert, d.h. die erste Linie der Anoden-Variablen, sowie die letzte Linie der Kathoden-Variablen werden jeweils aus dem System eliminiert. Die Solid-Temperatur sowie die Potentiale sind dagegen auf allen Linien zu berechnen. Damit ergibt sich bei den  $N$  Linien eine Gesamtzahl von  $11(N-1) + 3N + 5 = 14N - 6$  gewöhnlichen Differentialgleichungen.

Für die algebraischen Gleichungen und die entarteten PDE für  $(\gamma_a \vartheta_a)$  und  $(\gamma_c \vartheta_c)$  gibt es zwei Varianten: Da sie alle linear und gestaffelt sind, kann man sie zum einen als Ausdrücke eingeben und reduzieren<sup>4</sup>, so dass ein kleineres, stark besetztes, aber reines ODE-System resultiert. Zum anderen kann man die Gleichungen mitführen und ein etwas größeres, aber dünn besetztes DAE-System vom Index 1 erzeugen, welches mit entsprechendem DAE-Solver gelöst wird. Die Jacobi-Matrizen der rechten Seiten findet man in Abbildung 4.4.<sup>5</sup>

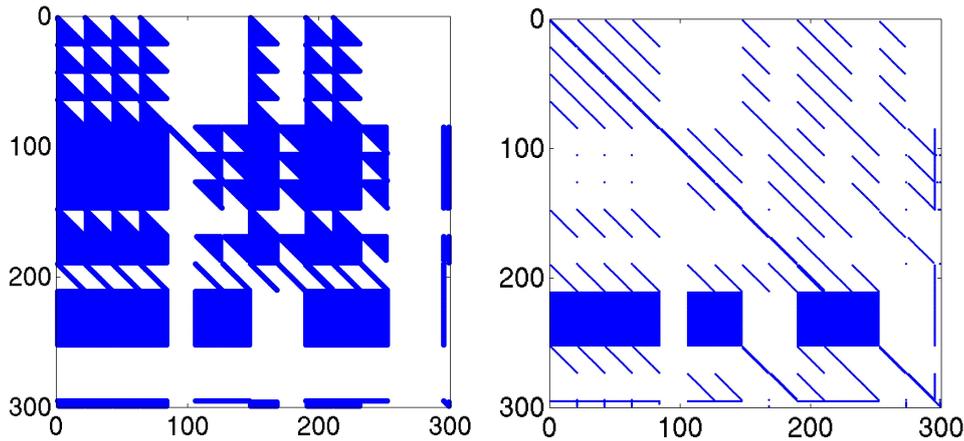


Abbildung 4.4: Besetztheitsstruktur der Jacobi-Matrix bei der ODE- und der DAE-Formulierung (spy-plot aus MATLAB).

<sup>3</sup>  $\vartheta_s^{(0)}$  und  $\vartheta_s^{(N+1)}$  sind Hilfswerte für die Neumann-Randbedingungen.

<sup>4</sup> Für  $\gamma_c$  nur im Falle  $R_{\text{back}} = 0$ .

<sup>5</sup> Abbildung vom Autor erstellt und in [90] veröffentlicht.

Wie wir sehen, kaufen wir uns bei der ODE-Formulierung stark besetzte Systemmatrizen ein, bei einer nur geringen Reduktion der Dimension (Reduktion um die weißen Zeilen und Spalten). Da DAE-Systeme vom Index 1 bekanntlich sehr effizient gelöst werden können (siehe Hairer et al. [47, 48]), sollte man die DAE-Formulierung vorziehen. In [93] wurde eine Mischung dieser beiden Varianten programmiert. Die algebraischen Gleichungen für  $\gamma_a$  wurden eliminiert. Die Partialdrücke in den Poren und  $\gamma_c$  wurden dagegen als algebraische Gleichungen ins DAE-System mit aufgenommen.

Die dicht besetzten Zeilen ließen sich sogar weiterhin auf 3 Zeilen reduzieren, indem man die algebraischen Größen  $I_a$ ,  $I_c$ ,  $I_e$  zu den Zuständen hinzunimmt. Numerisch haben wir diese Verbesserung derzeit noch nicht umgesetzt.

### 4.2.2 Zeitdiskretisierung

Für lineare Transportgleichungen ist bekannt, dass die Zeitschrittweite  $\tau$  bei expliziten Zeit-Integratoren durch eine CFL Bedingung  $\tau \leq C \cdot h$  an die Ortsschrittweite  $h$  gebunden ist. Bei parabolischen Problemen haben wir die striktere Bedingung  $\tau \leq Ch^2$ . Es müssen daher u.U. sehr viele kleine Zeitschritte gewählt werden, damit das Verfahren konvergiert. Implizite Differenzenverfahren in der Zeit erlauben dagegen größere Zeitschritte.

Das vorliegende Problem ist ein steifes DAE-System mit stark unterschiedlichen Zeitskalen ([90, 93]). Das DAE-System sollte daher mit impliziten und adaptiven Verfahren gelöst werden. Explizite Verfahren laufen in der Tat nur unter verschwindend kleinen Schrittweiten stabil ( $10^8 - 10^{10}$  Zeitschritte bei geringen Lastwechseln). Auch der Einsatz verschiedener semi-impliziter Verfahren (siehe z.B. Herzog und Kunisch [52, Remark 3.1]) führte nicht zum Ziel, da das DAE-System vermutlich zu viele nichtlineare Kopplungsterme enthält und das zugrunde liegende PDAE-System nicht zuletzt auch quasilineare PDE enthält, bei denen dann jeder Term nichtlinear ist bis auf die Zeitableitung.

Bei impliziten Einschnitt- oder Mehrschrittverfahren muss in jedem Zeitschritt ein nichtlineares Gleichungssystem gelöst werden. Bei kleinen Schrittweiten konvergiert die naheliegende Fixpunktiteration. Allgemeiner ist die Gleichung mit dem Newton-Verfahren oder seinen Abwandlungen iterativ zu lösen. Sehr gute Ergebnisse erzielt die DAE-Routine von MATLAB, `ode15s`, ein adaptives Mehrschrittverfahren variabler Ordnung zwischen 1 und 5 für steife ODE- und DAE-Systeme. Sie verwendet das vereinfachte Newton-Verfahren, wobei eine berechnete Jacobi-Matrix – wenn möglich – für mehrere Zeitschritte verwendet wird. Dadurch reduziert sich die Anzahl der benötigten Funktionsauswertungen drastisch, die Laufzeit sinkt. Die Jacobi-Matrizen stellen wir entweder mit numerischer Differentiation (`odenumjac`) oder mit automatischer Differentiation mittels dem Tool ADiMat [14] zur Verfügung. Beachten sollte man, dass die voreingestellte numerische Integration `odenumjac`, `numjac` in MATLAB durch Angabe des Besetztheitsgrades (Sparsity Pattern) bei der MCFC um ca. 36% beschleunigt werden kann (unabhängig von  $N$ ).<sup>6</sup>

---

<sup>6</sup> getestet für  $N \in \{21, 51, 151, 304\}$

### 4.2.3 Anfangswerte

Das DAE-System muss mit zulässigen Anfangswerten gestartet werden. Konstante Start-schätzungen erfüllen die Gleichungen nicht, der Gültigkeitsbereich des Modells wird dann gerade bei feinen Diskretisierungen nicht erreicht, das Verfahren bricht ab. Daher sind zulässige Anfangswerte zu generieren und in Tabellen abzuspeichern.

Für eine niedrige Ortsdiskretisierung liefert `ode15s` bei konstanten Anfangswerten korrekte Ergebnisse. So sind wir in der Lage, durch Simulation über einen langen Zeitraum stationäre Werte zu berechnen, welche in der Folge als stationäre Anfangswerte hergezo-gen werden.

Für eine feinere Ortsdiskretisierung konstruieren wir die passenden Anfangswerte itera-tiv: Für eine niedrige Anzahl an Linien liegt eine Startschätzung vor; wir interpolieren diese auf ein feineres Gitter und errechnen auf diesem erneut die stationäre Lösung. Ei-ne Iteration liefert entsprechend zuverlässige Startschätzungen für eine beliebig genaue Ortsdiskretisierung. Gestestet wurde das Verfahren erfolgreich, um von 21 Linien auf 641 Linien zu verfeinern. Damit liegen sehr gute Startschätzungen vor, mit denen auch der DAE-Solver arbeiten kann. Auf ein numerisches Hochfahren wie in [93] kann daher verzichtet werden.

### 4.2.4 Simulationsergebnisse

Mit der vorgestellten Semidiskretisierung und dem adaptiven Mehrschrittverfahren aus `ode15s` sind wir in der Lage, die Zustandsgleichungen effizient und genau zu lösen – wir haben also einen „Zustandssolver“. Ergebnisse der Simulationen haben wir in [90] aus-führlich dargestellt. In mehreren Studien haben wir dort ferner die Diskretisierungsfehler analysiert und das Übergangsverhalten des Systems bei einem akuten Lastwechsel cha-rakterisiert. Aufbauend auf dem Zustandssolver sind wir nun in der Lage, Algorithmen der Optimalen Steuerung aufzusetzen. Im nächsten Abschnitt identifizieren wir zunächst das Optimalsteuerungsproblem.

## 4.3 Das Optimalsteuerungsproblem der MCFC

Das Hauptziel der simulationsbasierten Optimierung einer MCFC-Anlage besteht dar-in, die Inputs (Gasströme) so zu steuern, dass die Effizienz der Anlage gesteigert wird, damit die gewünschte Last schnell und exakt geliefert wird und vor allem die Lebens-dauer der Komponenten erhöht wird: Zu niedrige oder zu hohe Temperaturen bringen das System aus dem Gleichgewicht und können zu dauerhaften Schäden führen. Zu hohe Temperaturunterschiede führen zu Spannungen in den festen Materialien, Spannungen führen schließlich zu Rissen, die MCFC wird irreparabel beschädigt.

Die Gasströme selbst können dabei nicht beliebig geändert werden, sondern nur inner-halb technisch bedingter Schranken. Aus diesen technischen Anforderungen erkennen wir ein zugrunde liegendes Problem der Optimalen Steuerung bei PDAE-Systemen unter Steuerungs- und Zustandsbeschränkungen, welches wir im folgenden Abschnitt definieren werden.

### 4.3.1 Auswahl der Steuerungen

Formulieren wir zunächst das Optimalsteuerungsproblem. Dazu sind zunächst die Steuerungen auszuwählen. Es gibt sieben technisch relevante Steuerungen, welche die Konfigurationen der Gaseinlässe festlegen:  $\gamma_{a,\text{in}}(t)$ ,  $\vartheta_{a,\text{in}}(t)$ ,  $\chi_{\text{CH}_4,a,\text{in}}(t)$ ,  $\chi_{\text{H}_2,a,\text{in}}(t)$ ,  $\lambda_{\text{air}}(t)$ ,  $\vartheta_{\text{air}}(t)$ ,  $R_{\text{back}}(t)$ . Die anderen Molenbrüche am Anodeneingang werden mit 0 eingespeist, bis auf das eliminierte  $\chi_{\text{H}_2\text{O},a,\text{in}}(t) = 1 - \chi_{\text{CH}_4,a,\text{in}}(t) - \chi_{\text{H}_2,a,\text{in}}(t)$ .

Allerdings muss man bei indirekten Methoden der Optimalsteuerung von mathematischer Seite her aufpassen. Steuerungen sollten hier idealerweise linear in die Zustandsgleichungen eingehen. Beim vorliegenden Brennstoffzellenmodell gehen die Steuerungen jedoch nichtlinear ein: Die meisten Steuerungen am Anoden-Eingang gehen ein zweites Mal an der Brennkammer ein, dort jedoch nichtlinear. Die Luftzufuhr ist in dieser Hinsicht ungünstig modelliert:

$$\gamma_{\text{air}}(t) = f^{\gamma_{\text{air}}} = \gamma_{a,\text{in}}(t) \frac{\lambda_{\text{air}}(t)}{\chi_{\text{O}_2,\text{air}}} \left( -\chi_{\text{O}_2,a,\text{in}}(t) - \sum_{k \in \mathcal{I}} \nu_{\text{O}_2, Ck} \chi_{k,a,\text{in}}(t) \right). \quad (4.3)$$

Wie wir sehen, geht zwar jede einzelne Steuerung linear ein, jedoch gibt es Produkte aus drei Steuerungen. Nach Ableiten liefert dies entsprechend mehrere quadratische Gleichungen in den Optimalitätsbedingungen. Gelegentlich lassen sich derartige Probleme mit einer Substitution beheben [96]. Dies gelingt uns auch im Falle der Brennstoffzelle, wobei wir das Ziel von linearen Steuerungen nicht ganz erreichen, sondern uns mit einer Situation dazwischen zufrieden geben müssen.

Wir definieren das Produkt  $\gamma_{a,\text{in}} \cdot \lambda_{\text{air}}$  als neue Steuerung. Damit handelt es sich in obiger Gleichung nur noch um Produkte zweier Steuerungen. Wir besprechen die Auswirkungen dieser Nichtlinearitäten in Abschnitt 4.4.4.

Als Steuerungen wählen wir also  $\vec{u}(t) = (u_1(t), \dots, u_7(t))$  mit<sup>7</sup>

$$\begin{aligned} u_1(t) &:= \gamma_{a,\text{in}}(t) \cdot \vartheta_{a,\text{in}}(t), & u_2(t) &:= \vartheta_{a,\text{in}}(t), & u_3(t) &:= \chi_{\text{CH}_4,a,\text{in}}(t), \\ u_4(t) &:= \chi_{\text{H}_2,a,\text{in}}(t), & u_5(t) &:= \gamma_{a,\text{in}}(t) \cdot \lambda_{\text{air}}(t), & u_6(t) &:= \vartheta_{\text{air}}(t), \\ u_7(t) &:= R_{\text{back}}(t). \end{aligned} \quad (4.4)$$

Daneben existiert eine weitere Input-Variable, die Stromstärke der Zelle  $I_{\text{cell}}$ . Sie definiert die gewünschte Leistung, welche die MCFC liefern soll, da die zugehörige Spannung der Zelle  $U_{\text{cell}}$  aus den Gleichungen folgt. Wir modellieren mit dieser Variable den vorliegenden Lastfall. Beispielsweise betrachten wir sprunghafte Lastwechsel oder periodische Lastwechsel.

Ein weiteres Ziel der Optimalen Steuerung ist es herauszufinden, mit welcher Steuerung sich der größte Effekt erzielen lässt. Denn die Verwirklichung einiger der Steuerungen verlangt zusätzlichen Aufwand (man denke an die Anschaffungs- und Betriebskosten der Steuereinrichtung). Die Zusammensetzung der Gasströme ist beispielsweise nicht so einfach zu ändern wie die Menge des eingespeisten Gases. Daher werden die numerischen Experimente in Abschnitt 4.7 auch mit Auswahlen dieser sieben Steuerungen arbeiten.

<sup>7</sup> Das zweite Produkt in  $u_1$  motivieren wir im nächsten Abschnitt.

### 4.3.2 Auswahl des Zielfunktional

Aus technischer Hinsicht liegt eine Pareto-Optimierungsaufgabe vor, also eine Aufgabe mit mehreren Zielen, welche abzuwägen sind. Wir formulieren jetzt mögliche Vertreter der einzelnen Ziele. Zum Beispiel kann man die Solidtemperatur oder ihren Gradienten an eine gewünschte, materialschonende Temperaturverteilung annähern, um die Lebensdauer zu optimieren:

$$Z_1 := \frac{1}{2} \int_0^{t_f} \int_0^1 (\vartheta_s - \vartheta_{\text{soll}})^2 dz dt, \quad (4.5)$$

$$Z_2 := \frac{1}{2} \int_0^{t_f} \int_0^1 \left( \frac{\partial \vartheta_s}{\partial z} - \frac{\partial \vartheta_{\text{soll}}}{\partial z} \right)^2 dz dt. \quad (4.6)$$

Ein anderes Ziel ist das schnelle und exakte Erreichen der geänderten Last

$$Z_3 := \frac{1}{2} \int_0^{t_f} (U_{\text{cell}} - U_{\text{cell, soll}})^2 dt. \quad (4.7)$$

Zusätzlich zum Erreichen des gewünschten Ziels muss man berücksichtigen, dass der Einsatz von Steuerungen nicht kostenfrei ist, sondern zu gewissen Kosten führt: Z.B. zu einem höheren Verbrauch von Brennstoff durch die Zelle oder durch nötiges Vorheizen einzelner Gasströme. Modellieren wir diese Kosten mit jeweiligen Kostenfaktoren  $\lambda_k$ ,  $k = 1, \dots, 7$  so lautet das zu minimierende Zielfunktional

$$\min_{\vec{u}(t)} J = Z_j + \frac{1}{2} \sum_{k=1}^7 \lambda_k \int_0^{t_f} u_k^2(t) dt, \quad (4.8)$$

wobei wir uns für eines der obigen Ziele entscheiden (oder eine Linearkombination der Ziele ansetzen).

### 4.3.3 Nebenbedingungen

Als Nebenbedingungen sollen die Zustandsgleichungen der MCFC aus Abb. 4.2 bzw. Anhang A erfüllt sein und zusätzlich Steuerungsbeschränkungen

$$u_{u,i}(t) \leq u_i(t) \leq u_{o,i}(t), \quad i = 1, \dots, 7, \quad (4.9)$$

mit  $u_{u,i}(t), u_{o,i}(t) \in L^\infty(0, t_f)$  und  $u_{u,i}(t) < u_{o,i}(t)$  für fast alle  $t$ . Alternativ schreiben wir dafür

$$u_i(t) \in U_{\text{ad},i} := \{u \in L^2(0, t_f) \mid u_{u,i}(t) \leq u(t) \leq u_{o,i}(t) \text{ f. ü.}\}. \quad (4.10)$$

bzw.  $\vec{u} \in U_{\text{ad}}$  mit  $U_{\text{ad}} := U_{\text{ad},1} \times \dots \times U_{\text{ad},7}$ . Die Problemstellung wird gelegentlich erweitert um punktweise Zustandsbeschränkungen an die Solidtemperatur

$$\vartheta_u(z, t) \leq \vartheta_s(z, t) \leq \vartheta_o(z, t) \quad (4.11)$$

oder an ihren Gradienten

$$\left\| \frac{\partial \vartheta_s}{\partial z}(z, t) \right\| \leq g(t). \quad (4.12)$$

Weiterhin müssen alle Molenbrüche stets in  $[0, 1]$  liegen, was weitere Zustandsbeschränkungen liefern würde. Diese sind jedoch bereits durch Abstimmung der Gleichungen stets erfüllt und entfallen daher (siehe [90]).

Insgesamt erhalten wir ein Dirichlet-Randsteuerungsproblem. Prototypen von Dirichlet-Randsteuerungsproblemen wurden bereits von einigen Autoren untersucht, beispielsweise geben Vexler und Kunisch [99] eine umfassende Darstellung. Derartige Probleme gelten als sehr anspruchsvoll, da die Theorie mit geringeren Regularitäten kämpfen muss und die Numerik umständlicher ist und geringere Konvergenzordnung besitzt. Allerdings handelt es sich in der Literatur um PDE zweiter Ordnung mit Dirichlet-Randsteuerung. Bei der MCFC werden dagegen die Dirichlet-Randbedingungen von PDE erster Ordnung gesteuert, so dass wir die Resultate nicht direkt übertragen können.

#### 4.3.4 Existenz von Lösungen

Damit ein Optimalsteuerungsproblem sinnvoll gestellt ist, braucht man die Existenz einer Lösung der Zustandsgleichungen für jede beliebige Steuerung  $\vec{u}(t)$ , welche die Box-Beschränkungen (4.10) erfüllt. Im Allgemeinen besitzen nichtlineare PDAE-Systeme keine eindeutige Lösung mehr. Die Lösungsstruktur sowie die Existenzbeweise werden sehr anfordernd. Wir setzen daher die Existenz und Eindeutigkeit des Modells im Betriebsbereich voraus.

**Annahme 4.3.1** Die Zustandsgleichungen besitzen für jede gemäß (4.10) zulässige Steuerung genau eine hinreichend reguläre lokale Lösung.

Zur Gültigkeit der Annahme muss der Anwender angepasste Steuerungsbeschränkungen wählen. In bisherigen numerischen Studien konnten wir keinen zweiten Lösungssast im Betriebsbereich der MCFC entdecken. Die Ergebnisse sind stets physikalisch sinnvoll. Das 2D-Modell wurde darüber hinaus sogar verifiziert in Gundermann, Heidebrecht und Sundmacher [45]. Weder Hysterese-Effekte noch das Auftreten von Schocks wurden bisher beobachtet. Derartige Fälle schließen wir daher für die Analyse aus.

Die Existenz einer optimalen Steuerung müsste man ebenfalls sehr aufwändig beweisen - falls es beim vorliegenden Anwendungsproblem überhaupt beweisbar ist. Wir gehen auch hier von der Existenz einer optimalen Steuerung aus:

**Annahme 4.3.2** Es existiert eine optimale Steuerung  $\vec{u}$  und zugehörige optimale Zustände des OC-PDAE aus Abb. 4.2 bzw. Anhang A. <sup>8</sup>

Numerisch verwenden wir Abstiegsverfahren mit Abbruchkriterien, so dass diese auf jeden Fall terminieren.

---

<sup>8</sup> Die optimalen Zustände erhalten der Übersichtlichkeit halber in diesem Kapitel keine separate Kennzeichnung. Sie geht aus dem Kontext hervor.

## 4.4 Notwendige Bedingungen

Bei der Behandlung größerer Anwendungsprobleme aus dem Bereich der Optimalen Steuerung mit indirekten Methoden empfiehlt sich folgendes Vorgehen: Zunächst sollte man die formale Lagrange-Technik anwenden, um Kenntnis über die (vermutlichen) Optimalitätsbedingungen zu erlangen. Zu beweisen, dass die so erhaltenen Optimalitätsbedingungen korrekt sind, wird dem Anwender dagegen nur in Spezialfällen möglich sein.

Eine CQ wie in den drei vorhergehenden Kapiteln werden wir hier nicht zeigen können: Bei reinen Gleichungsrestriktionen  $G(u) = 0$  – wie z.B. das PDAE-System der Zustandsgleichungen der MCFC ohne Zustandsbeschränkungen – ist die Slater-Bedingung nicht anwendbar. Die Zowe-Kurcyusz-Bedingung erfordert hier, dass der Operator  $G'(u)$  surjektiv ist. Diese einfach klingende Bedingung ist in der Praxis allerdings bereits schwer erfüllbar. Denn die Funktionenräume müssen passend gewählt werden, so dass alle Nichtlinearitäten noch differenzierbar aber gleichzeitig alle Multiplikatoren noch regulär genug sind (vgl. [96, S. 265]). Zudem fehlen dann i.d.R. die linearisierten Zustandsgleichungen sowie Sätze, welche die Existenz einer Lösung sicherstellen. Es bleibt oft nichts anderes übrig als die Existenz der Lagrange-Multiplikatoren anzunehmen.

**Annahme 4.4.1** Wir betrachten das Optimalsteuerungsproblem der MCFC aus Abschnitt 4.3 unter Weglassen der Zustandsbeschränkungen (4.11), (4.12). Dann sollen die Lagrange-Multiplikatoren zu den Gleichungsnebenbedingungen (also zum PDAE-System der MCFC) existieren und die nötige Regularität besitzen.

### 4.4.1 Herleitung der adjungierten Gleichungen der MCFC

In diesem Abschnitt leiten wir die adjungierten Gleichungen des 1D-Modell der MCFC ohne Zustandsbeschränkungen mit eliminierten Porendrücken mittels der formalen Lagrange-Technik her. Der Leser benötigt grundlegendes Vorwissen aus [96], z.B. das Berechnen von Richtungsableitungen von Funktionalen.

Die Herleitung des adjungierten PDAE-Systems ist ein größeres Unterfangen. Der Übersichtlichkeit halber wurden daher technische Details in den Anhang ausgegliedert, wie z.B. die 24 partiellen Ableitungen der Reaktionsraten oder der länglichen rechten Seiten des PDAE-Systems. Hauptaugenmerk liegt dagegen auf den Differentialoperatoren und den Integralausdrücken, sowie den Randbedingungen, welche zusammen den Charakter des adjungierten PDAE-Systems festlegen werden.

Je nach Ausführung der Zustandsgleichungen erhält man anders beschaffene adjungierte PDAE-Systeme. Wir werden im Folgenden einen Zugang wählen, bei welchem das adjungierte PDAE-System nahe an den primären Gleichungen und dabei relativ dünn besetzt ist. Grundlage dafür sind die Substitutionen  $v_a := \gamma_a \vartheta_a$  sowie  $v_c := \gamma_c \vartheta_c$ , welche bereits im Theorie-Teil von [93] Anwendung finden.<sup>9</sup> In den rechten Seiten der Brennkammer im Anhang ist an späterer Stelle entsprechend jeweils für  $\gamma_a$  der Bruch  $\frac{v_a}{\vartheta_a}$  zu setzen und für  $\gamma_c$  analog – und dies ist beim Berechnen der partiellen Ableitungen auch zu berücksichtigen.<sup>10</sup>

<sup>9</sup> Entsprechend verwenden wir  $f^{v_a} := f^{\gamma_a \vartheta_a}$ ,  $f^{v_c} := f^{\gamma_c \vartheta_c}$  sowie  $v_{a,in} := \gamma_{a,in} \vartheta_{a,in}$

<sup>10</sup> Man beachte, dass  $\vartheta_{a|c} \neq 0$ , da  $\vartheta_{a|c} = 0$  der absoluten Temperatur von 0 K entspricht.

Bei komplexen Anwendungsproblemen sollte man für die PDE und ihre Randbedingung jeweils eigene Lagrange-Multiplikatoren ansetzen. Die formale Lagrange-Technik liefert den Zusammenhang dieser Größen anschließend als Teil der Optimalitätsbedingungen. Bei Optimalsteuerungsproblemen bei skalaren Dirichlet- oder Neumann-Problemen<sup>11</sup> stellen sich beide Multiplikatoren als auf dem Rand identisch heraus. Bei Dirichlet-Randsteuerungsproblemen 2. Ordnung hängen sie über die Normalen-Spur miteinander zusammen, so dass die Gradientengleichung die Normalen-Spur des adjungierten Zustandes enthält, vgl. [99]. Im vorliegenden gekoppelten System werden wir eigene Zusammenhänge in Gl. (4.17) finden.

Unter der Annahme der Existenz der 67 Lagrange-Multiplikatoren  $\tilde{\chi}_{k,a}(z, t)$ ,  $\tilde{\vartheta}_a(z, t)$ , etc. der Gleichungen bzw.  $\hat{\chi}_{k,a}(t)$ ,  $\hat{\vartheta}_a(t)$ , etc. der Randbedingungen lässt sich nun eine Lagrange-Funktion definieren per<sup>12</sup>

---

<sup>11</sup> unter Verwendung der Conormalen-Ableitung

<sup>12</sup> Die Variablen  $(z, t)$  bzw.  $(t)$  sind der Übersichtlichkeit halber weggelassen, es sei denn ein fester Wert wurde eingesetzt.

$$\mathcal{L}(\dots) := J$$

$$\begin{aligned}
& - \sum_{k \in \mathcal{I}} \iint_Q \left( \frac{\partial \chi_{k,a}}{\partial t} + v_a \frac{\partial \chi_{k,a}}{\partial z} - f^{\chi_{k,a}} \right) \tilde{\chi}_{k,a} dz dt - \int_0^{t_f} \sum_{k \in \mathcal{I}} \left( \chi_{k,a}(0, t) - \chi_{k,a,\text{in}} \right) \hat{\chi}_{k,a} dt \\
& - \sum_{k \in \mathcal{I}} \iint_Q \left( \frac{\partial \chi_{k,c}}{\partial t} - v_c \frac{\partial \chi_{k,c}}{\partial z} - f^{\chi_{k,c}} \right) \tilde{\chi}_{k,c} dz dt - \int_0^{t_f} \sum_{k \in \mathcal{I}} \left( \chi_{k,c}(1, t) - \chi_{k,m} \right) \hat{\chi}_{k,c} dt \\
& - \iint_Q \left( \frac{\partial \vartheta_a}{\partial t} + v_a \frac{\partial \vartheta_a}{\partial z} - f^{\vartheta_a} \right) \tilde{\vartheta}_a dz dt - \int_0^{t_f} \left( \vartheta_a(0, t) - \vartheta_{a,\text{in}} \right) \hat{\vartheta}_a dt \\
& - \iint_Q \left( \frac{\partial \vartheta_c}{\partial t} - v_c \frac{\partial \vartheta_c}{\partial z} - f^{\vartheta_c} \right) \tilde{\vartheta}_c dz dt - \int_0^{t_f} \left( \vartheta_c(1, t) - \vartheta_m \right) \hat{\vartheta}_c dt \\
& - \iint_Q \left( \frac{\partial v_a}{\partial z} - f^{v_a} \right) \tilde{v}_a dz dt - \int_0^{t_f} \left( v_a(0, t) - v_{a,\text{in}} \right) \hat{v}_a dt \\
& - \iint_Q \left( - \frac{\partial v_c}{\partial z} - f^{v_c} \right) \tilde{v}_c dz dt - \int_0^{t_f} \left( v_c(1, t) - \gamma_m \vartheta_m \right) \hat{v}_c dt \tag{4.13} \\
& - \iint_Q \left( c_{p,s} \frac{\partial \vartheta_s}{\partial t} - \frac{1}{\text{Pe}} \frac{\partial^2 \vartheta_s}{\partial z^2} - f^{\vartheta_s} \right) \tilde{\vartheta}_s dz dt - \frac{1}{\text{Pe}} \int_0^{t_f} \left[ \frac{\partial \vartheta_s}{\partial z} \hat{\vartheta}_s \right]_0^1 dt \\
& - \sum_{k \in \mathcal{I}} \int_0^{t_f} \left( \chi_{k,b} - f^{\chi_{k,b}} \right) \tilde{\chi}_{k,b} dt - \int_0^{t_f} \left( \vartheta_b - f^{\vartheta_b} \right) \tilde{\vartheta}_b dt - \int_0^{t_f} \left( \gamma_b - f^{\gamma_b} \right) \tilde{\gamma}_b dt \\
& - \int_0^{t_f} \left( \gamma_{\text{air}} - f^{\gamma_{\text{air}}} \right) \tilde{\gamma}_{\text{air}} dt - \sum_{k \in \mathcal{I}} \int_0^{t_f} \left( \frac{d\chi_{k,m}}{dt} - f^{\chi_{k,m}} \right) \tilde{\chi}_{k,m} dt - \int_0^{t_f} \left( \gamma_m - f^{\gamma_m} \right) \tilde{\gamma}_m dt \\
& - \int_0^{t_f} \left( \frac{d\vartheta_m}{dt} - f^{\vartheta_m} \right) \tilde{\vartheta}_m dt - \iint_Q \left( i - f^i \right) \tilde{i} dt dz - \sum_{k \in \{a,c,e\}} \iint_Q \left( i_k - f^{i_k} \right) \tilde{i}_k dt dz \\
& - \sum_{k \in \{a,c,e\}} \int_0^{t_f} \left( I_k - \int_0^1 i_k dz \right) \tilde{I}_k dt - \int_0^{t_f} \left( \frac{dU_{\text{cell}}}{dt} - f^{u_{\text{cell}}} \right) \tilde{U}_{\text{cell}} dt \\
& - \iint_Q \left( \frac{\partial \Phi_a^L}{\partial t} - f^{\Phi_a^L} \right) \tilde{\Phi}_a^L dt dz - \iint_Q \left( \frac{\partial \Phi_c^L}{\partial t} - f^{\Phi_c^L} \right) \tilde{\Phi}_c^L dt dz .
\end{aligned}$$

Die Anfangsbedingungen werden wie üblich separat berücksichtigt bzw. in den Lösungsraum definiert. Beim Ableiten muss die Testfunktion entsprechend eine homogene Anfangsbedingung erfüllen. Soweit es effizient ist, werden die Ableitungen mittels partieller Integration bzw. Greenscher Formel bereits auf die dualen Variablen abgewälzt; die quasilinearen Anteile z.B.  $v_a \frac{\partial \chi_{k,a}}{\partial z}$  werden dagegen erst nach dem Differenzieren behandelt,

da nicht jede Variable die partielle Integration erfordert (z.B.  $v_a$  nicht). Mit partieller Integration erhalten wir beispielsweise

$$\begin{aligned} \iint_Q \frac{\partial \vartheta_a}{\partial t} \tilde{\vartheta}_a \, dt \, dz &= \int_0^1 \left[ \vartheta_a \tilde{\vartheta}_a \right]_0^{t_f} \, dz - \iint_Q \frac{\partial \tilde{\vartheta}_a}{\partial t} \vartheta_a \, dt \, dz, \\ \iint_Q \frac{\partial v_a}{\partial z} \tilde{v}_a \, dt \, dz &= \int_0^{t_f} [v_a \tilde{v}_a]_0^1 \, dt - \iint_Q \frac{\partial \tilde{v}_a}{\partial z} v_a \, dt \, dz. \end{aligned}$$

Durch Berechnung der Richtungsableitungen der Lagrange-Funktion lassen sich nun die Optimalitätsbedingungen aufstellen. Der kürzeren Schreibweise wegen werden fortan die Quellterme in Summen zusammengefasst mit Indexmengen benannt nach Zelle  $Z := \{\vartheta_a, \vartheta_c, \chi_{k,a}, \chi_{k,c}, \vartheta_s, v_a, v_c, i_a, i_c\}$ , Brennkammer  $B := \{\chi_{k,b}, \vartheta_b, \gamma_b, \gamma_{air}\}$ , Mischkammer  $M := \{\chi_{k,m}, \vartheta_m, \gamma_m\}$  und restlichen elektrischen Variablen  $E := \{\Phi_a^L, \Phi_c^L, i, i_e, U_{cell}\}$ <sup>13</sup>. Beispielhaft muss die Richtungsableitung nach  $\vartheta_a$  für alle Richtungen  $\varphi(z, t) \in \{\psi \in H \mid \psi(z, 0) = 0\}$  verschwinden:

$$\begin{aligned} 0 &= D_{\vartheta_a} \mathcal{L} \varphi = - \iint_Q - \frac{\partial \tilde{\vartheta}_a}{\partial t} \varphi + v_a \tilde{\vartheta}_a \frac{\partial \varphi}{\partial z} \, dz \, dt + \sum_{q \in Z} \iint_Q (D_{\vartheta_a} f^q \varphi) \tilde{q} \, dz \, dt \\ &\quad - \int_0^1 [\varphi \tilde{\vartheta}_a]_0^{t_f} \, dz - \int_0^{t_f} \varphi(0, t) \hat{\vartheta}_a \, dt + \sum_{q \in B} \int_0^{t_f} (D_{\vartheta_a} f^q \varphi(1, t)) \tilde{q} \, dt \quad (4.14) \\ &\stackrel{\text{p.I.}}{=} - \iint_Q \left( - \frac{\partial \tilde{\vartheta}_a}{\partial t} - \frac{\partial (v_a \tilde{\vartheta}_a)}{\partial z} \right) \varphi \, dz \, dt - \int_0^{t_f} [v_a \tilde{\vartheta}_a \varphi]_0^1 \, dt + \sum_{q \in Z} \iint_Q (D_{\vartheta_a} f^q \varphi) \tilde{q} \, dz \, dt \\ &\quad - \int_0^1 [\varphi \tilde{\vartheta}_a]_0^{t_f} \, dz - \int_0^{t_f} \varphi(0, t) \hat{\vartheta}_a \, dt + \sum_{q \in B} \int_0^{t_f} (D_{\vartheta_a} f^q \varphi(1, t)) \tilde{q} \, dt. \end{aligned}$$

Um daraus die adjungierten Gleichungen herauslesen zu können, bedient man sich der freien Wahl der Ableitungsrichtung  $\varphi \in \{\psi \in H \mid \psi(z, 0) = 0\}$ . Dabei ist  $H$  der Banach-Raum, in welchem der Zustand  $\vartheta_a$  liegt. Ohne Existenzsatz einer Lösung können wir  $H$  nicht explizit angeben. Wir nehmen jedoch an, dass  $C_0^\infty(Q) \subset H$  ist und dass obige Integrale auf  $Q$  entweder Skalarprodukte auf  $V := L^2(Q)$  sind oder Dualitätsprodukte auf  $V := L^p(Q)$  für  $1 < p < \infty$ . Da  $C_0^\infty(Q) \subset V$  jeweils dicht liegt, liefert die Wahl  $\varphi \in C_0^\infty(Q) \subset H$  die lokale Form der adjungierten PDE

$$\begin{aligned} - \frac{\partial \tilde{\vartheta}_a}{\partial t} - \frac{\partial (v_a \tilde{\vartheta}_a)}{\partial z} &= \sum_{q \in Z} \frac{\partial f^q}{\partial \vartheta_a} \tilde{q} \\ \Leftrightarrow - \frac{\partial \tilde{\vartheta}_a}{\partial t} - v_a \frac{\partial \tilde{\vartheta}_a}{\partial z} &= \frac{\partial v_a}{\partial z} \tilde{\vartheta}_a + \sum_{q \in Z} \frac{\partial f^q}{\partial \vartheta_a} \tilde{q}. \end{aligned}$$

Man beachte, dass die Ableitungen der Quellterme der Übersichtlichkeit halber erst im nächsten Abschnitt behandelt werden. Zunächst fällt auf, dass im Gegensatz zu semi-linearen PDAE-Systemen die hier vorliegenden quasilinearen Terme dazu führen, dass

<sup>13</sup>  $I_a, I_c, I_e$  werden direkt behandelt.

partielle Ableitungen von Zuständen in den Quelltermen der adjungierten Gleichungen auftreten.<sup>14</sup>

Auch die restlichen Terme lassen sich nach Integrationsbereichen separieren, vorausgesetzt die Spurabbildungen sind surjektiv in die entsprechenden  $L^2$ -Räume auf den verschiedenen Randteilen des Orts-Zeit-Zylinders:

$$\begin{aligned} v_a \tilde{\vartheta}_a &= \sum_{q \in B} \frac{\partial f^q}{\partial \vartheta_a} \tilde{q} && \text{auf } \{1\} \times (0, t_f), \\ \tilde{\vartheta}_a(z, t_f) &= 0 && \forall z \in \Omega, \\ \hat{\vartheta}_a &= v_a \tilde{\vartheta}_a && \text{auf } \{0\} \times (0, t_f). \end{aligned}$$

Aus der dritten Gleichung erkennen wir, wie wir die Randbedingung hätten verändern müssen, um sie mit dem gleichen Multiplikator ankoppeln zu können.

Mit dieser Methodik stellen wir alle adjungierten partiellen Differentialgleichungen auf:<sup>15</sup>

$$-\frac{\partial \tilde{\vartheta}_a}{\partial t} - v_a \frac{\partial \tilde{\vartheta}_a}{\partial z} = \frac{\partial v_a}{\partial z} \tilde{\vartheta}_a + \sum_{q \in Z} \frac{\partial f^q}{\partial \vartheta_a} \tilde{q}, \quad (4.15a)$$

$$-\frac{\partial \tilde{\vartheta}_c}{\partial t} + v_c \frac{\partial \tilde{\vartheta}_c}{\partial t} = -\frac{\partial v_c}{\partial z} \tilde{\vartheta}_c + \sum_{q \in Z} \frac{\partial f^q}{\partial \vartheta_c} \tilde{q}, \quad (4.15b)$$

$$-\frac{\tilde{\chi}_{k,a}}{\partial t} - v_a \frac{\partial \tilde{\chi}_{k,a}}{\partial z} = \frac{\partial v_a}{\partial z} \tilde{\chi}_{k,a} + \sum_{q \in Z} \frac{\partial f^q}{\partial \chi_{k,a}} \tilde{q}, \quad (4.15c)$$

$$-\frac{\tilde{\chi}_{k,c}}{\partial t} + v_c \frac{\partial \tilde{\chi}_{k,c}}{\partial z} = -\frac{\partial v_c}{\partial z} \tilde{\chi}_{k,c} + \sum_{q \in Z} \frac{\partial f^q}{\partial \chi_{k,a}} \tilde{q}, \quad (4.15d)$$

$$-c_{p,s} \frac{\partial \tilde{\vartheta}_s}{\partial t} - \frac{1}{\text{Pe}} \frac{\partial^2 \tilde{\vartheta}_s}{\partial z^2} = \sum_{q \in Z} \frac{\partial f^q}{\partial \vartheta_s} \tilde{q}, \quad (4.15e)$$

$$-\frac{\partial \tilde{v}_a}{\partial z} = -\sum_{k \in \mathcal{I}} \frac{\partial \chi_{k,a}}{\partial z} \tilde{\chi}_{k,a} - \frac{\partial \vartheta_a}{\partial z} \tilde{\vartheta}_a, \quad (4.15f)$$

$$\frac{\partial \tilde{v}_c}{\partial z} = \sum_{k \in \mathcal{I}} \frac{\partial \chi_{k,c}}{\partial z} \tilde{\chi}_{k,c} + \frac{\partial \vartheta_c}{\partial z} \tilde{\vartheta}_c. \quad (4.15g)$$

Wir erkennen bereits an den linken Seiten, dass der Typus der einzelnen partiellen Differentialgleichung unverändert bleibt, jedoch alle ersten Ableitungen ihr Vorzeichen ändern. Für die hyperbolischen Transportgleichungen bedeutet dies eine Umkehr des Flusses. Zusätzlich laufen die Transport- und Wärmeleitungsgleichungen nun ab der Endzeit rückwärts in der Zeit.

Die erhaltenen Randbedingungen auf  $\{1\} \times (0, t_f)$  ergeben sich zu

$$v_a \tilde{\vartheta}_a = \sum_{q \in B} \frac{\partial f^q}{\partial \vartheta_a(1, t)} \tilde{q}, \quad v_a \tilde{\chi}_{k,a} = \sum_{q \in B} \frac{\partial f^q}{\partial \chi_{k,a}(1, t)} \tilde{q}, \quad \tilde{v}_a = \sum_{q \in B} \frac{\partial f^q}{\partial v_a(1, t)} \tilde{q}, \quad (4.16a)$$

<sup>14</sup> Man vergleiche die adjungierten Gleichungen der Navier-Stokes-Gleichungen (siehe z.B. [32]).

<sup>15</sup> Den Beitrag des Zielfunktionals  $J$  fügen wir nachträglich hinzu, wenn wir das Zielfunktional spezifizieren (siehe Abschnitt 4.4.3).

sowie auf  $\{0\} \times (0, t_f)$  zu

$$v_c \tilde{\vartheta}_c = \sum_{q \in B} \frac{\partial f^q}{\partial \vartheta_c(0, t)} \tilde{q}, \quad v_c \tilde{\chi}_{k,c} = \sum_{q \in B} \frac{\partial f^q}{\partial \chi_{k,c}(0, t)} \tilde{q}, \quad \tilde{v}_c = \sum_{q \in B} \frac{\partial f^q}{\partial v_c(0, t)} \tilde{q}. \quad (4.16b)$$

Dazu gilt für die adjungierte Solidtemperatur auf  $\{0, 1\} \times (0, t_f)$  die Symmetriebedingung

$$\frac{\partial \tilde{\vartheta}_s}{\partial z} = 0. \quad (4.16c)$$

Weiterhin finden wir den Zusammenhang der adjungierten Variablen auf den Ortsrändern:

$$\hat{\vartheta}_a = v_a \tilde{\vartheta}_a, \quad \hat{\chi}_{k,a} = v_a \tilde{\chi}_{k,a}, \quad \hat{v}_a = \tilde{v}_a \quad \text{auf } \{0\} \times (0, t_f), \quad (4.17a)$$

$$\hat{\vartheta}_c = v_c \tilde{\vartheta}_c, \quad \hat{\chi}_{k,c} = v_c \tilde{\chi}_{k,c}, \quad \hat{v}_c = \tilde{v}_c \quad \text{auf } \{1\} \times (0, t_f), \quad (4.17b)$$

$$\hat{\vartheta}_s = \tilde{\vartheta}_s \quad \text{auf } \{0, 1\} \times (0, t_f). \quad (4.17c)$$

Diese Gleichungen werden wir sukzessive dazu nutzen, um die Variablen auf den linken Seiten aus den Optimalitätsbedingungen zu entfernen.

Die Endbedingungen sind jeweils homogen, sie werden hier nicht separat aufgeführt. Inhomogen werden sie in dem Moment, in welchem das Zielfunktional auch Tracking-Terme von Endzuständen enthält<sup>16</sup>.

Das DAE-System der Brenn- und der Mischkammer erzeugt ein ähnliches DAE-System von adjungierten Gleichungen:

$$\tilde{\chi}_{k,b} = \frac{\partial f^{\chi_{k,m}}}{\partial \chi_{k,b}} \tilde{\chi}_{k,m}, \quad (4.18a)$$

$$\tilde{\vartheta}_b = \sum_{q \in \{\vartheta_m, \gamma_m\}} \frac{\partial f^q}{\partial \vartheta_b} \tilde{q}, \quad (4.18b)$$

$$\tilde{\gamma}_b = \sum_{q \in B \cup M} \frac{\partial f^q}{\partial \gamma_b} \tilde{q}, \quad (4.18c)$$

$$\tilde{\gamma}_{\text{air}} = \sum_{q \in B} \frac{\partial f^q}{\partial \gamma_{\text{air}}} \tilde{q}, \quad (4.18d)$$

$$-\frac{d\tilde{\chi}_{k,m}}{dt} = \hat{\chi}_{k,c} + \frac{\partial f^{\chi_{k,m}}}{\partial \chi_{k,m}} = v_c(1, t) \cdot \tilde{\chi}_{k,c}(1, t) + \frac{\partial f^{\chi_{k,m}}}{\partial \chi_{k,m}} \tilde{\chi}_{k,m}, \quad (4.18e)$$

$$-\frac{d\tilde{\vartheta}_m}{dt} = \hat{\vartheta}_c + \sum_{q \in M} \frac{\partial f^q}{\partial \vartheta_m} \tilde{q} = v_c(1, t) \cdot \tilde{\vartheta}_c(1, t) + \sum_{q \in M} \frac{\partial f^q}{\partial \vartheta_m} \tilde{q}, \quad (4.18f)$$

$$\tilde{\gamma}_m = \vartheta_m \hat{v}_c = \vartheta_m \tilde{v}_c(1, t). \quad (4.18g)$$

Die Endbedingungen sind erneut homogen, also  $\tilde{\chi}_{k,m}(t_f) = \tilde{\vartheta}_m(t_f) = 0$ . Ausgefallener sind dagegen die adjungierten Gleichungen von  $U_{\text{cell}}(t)$  und  $I_k(t), k \in \{a, c, e\}$ , da hier Integralterme involviert sind. Als Richtungsableitung erhalten wir z.B. für  $U_{\text{cell}}$ :

$$\begin{aligned} 0 &= D_{U_{\text{cell}}} \mathcal{L} \varphi(t) = \int_0^{t_f} \frac{d\tilde{U}_{\text{cell}}}{dt} \varphi(t) dt + \iint_Q \sum_{q \in Z \cup E} \frac{\partial f^q}{\partial U_{\text{cell}}} (z, t) \tilde{q}(z, t) dz dt \\ &= \int_0^{t_f} \left( \frac{d\tilde{U}_{\text{cell}}}{dt} + \int_0^1 \sum_{q \in Z \cup E} \frac{\partial f^q}{\partial U_{\text{cell}}} (z, t) \tilde{q}(z, t) dz \right) \varphi(t) dt. \end{aligned} \quad (4.19)$$

<sup>16</sup> Also z.B. den Summanden  $\frac{1}{2} \int_0^1 (\vartheta_s(z, t_f) - \vartheta_{\text{soil},2}(z))^2 dz$ .

Testen mit  $\varphi \in C_0^\infty((0, t_f))$  liefert eine gewöhnliche Differentialgleichung mit Integraltermen von im Ort verteilten Funktionen, wie sie meist bei gemischten ODE-PDE Steuerungsproblemen auftritt - siehe auch Kapitel 3.4.1. Dieser Effekt tritt ebenso bei den anderen Variablen einer Veränderlichen  $I_k(t)$ ,  $k \in \{a, c, e\}$  auf. Insgesamt lauten die adjungierten Gleichungen der elektrischen Variablen:

$$\tilde{I}_k(t) = \sum_{q \in E} \int_0^1 \frac{\partial f^q}{\partial I_k} \tilde{q} dz, \quad k \in \{a, c, e\}, \quad (4.20a)$$

$$-\frac{d\tilde{U}_{\text{cell}}}{dt}(t) = \sum_{q \in Z \cup E} \int_0^1 \frac{\partial f^q}{\partial U_{\text{cell}}} \tilde{q} dz, \quad (4.20b)$$

$$-\frac{\partial \tilde{\Phi}_{a|c}^L}{\partial t}(z, t) = \sum_{q \in Z \cup E} \frac{\partial f^q}{\partial \Phi_{a|c}^L} \tilde{q}, \quad (4.20c)$$

$$\tilde{i}(z, t) = \sum_{q \in E} \frac{\partial f^q}{\partial i} \tilde{q}, \quad (4.20d)$$

$$\tilde{i}_k(z, t) = \tilde{I}_k(t) + \sum_{q \in E \cup \{\vartheta_s\}} \frac{\partial f^q}{\partial i_k} \tilde{q}, \quad k \in \{a, c, e\}. \quad (4.20e)$$

Bemerkenswert ist weiterhin die letzte Gleichung: Aus der Integralgleichung  $I_k(t) = \int_0^1 i_k(z, t) dz$  wird hier eine örtlich konstante Quelle  $\tilde{I}_k(t)$  in den algebraischen Gleichungen der adjungierten Stromstärken  $\tilde{i}_k$ ,  $k \in \{a, c, e\}$ .

#### 4.4.2 Herleitung der rechten Seiten

Bisher wurden die Differentialoperatoren sowie die Struktur der Übergangsbedingungen hergeleitet. Der eigentliche Aufwand des Aufstellens der adjungierten Gleichungen der MCFC versteckt sich jedoch in den rechten Seiten, welchen sich dieser Abschnitt widmet. Es muss hier eine große (bei ungünstigem Vorgehen dreistellige) Anzahl von partiellen Ableitungen von Funktionen mehrerer Veränderlicher berechnet werden. Diese Aufgabe sollte man bei komplexen Anwendungsproblemen mit analytischem bzw. automatischem Differenzieren erledigen lassen (z.B. mit MAPLE oder einem AD-Tool, siehe [www.autodiff.org](http://www.autodiff.org)).

Um den Überblick zu behalten, sollte man sich die Struktur der Zustandsgleichungen zunutze machen und Hilfsvariablen definieren. Dies erleichtert zum einen die Herleitung und sorgt für eine übersichtlichere Darstellung der Ergebnisse, zum anderen begünstigt dies nicht zuletzt die Programmierung. Betrachten wir zunächst die adjungierte Gleichung der Solidtemperatur

$$-c_{p,s} \frac{\partial \tilde{\vartheta}_s}{\partial t} - \frac{1}{\text{Pe}} \frac{\partial^2 \tilde{\vartheta}_s}{\partial z^2} = \sum_{q \in Z} \frac{\partial f^q}{\partial \vartheta_s} \tilde{q}. \quad (4.21)$$

Die Solidtemperatur tritt in den meisten Hilfsgrößen auf (Reaktionsraten  $r_j$ , Wärmeaustausch  $q_j$ ) und geht daher in die rechten Seiten aller PDE sowie in  $f^{i_a}$ ,  $f^{i_c}$  ein. Wir teilen das Auftreten von  $\vartheta_s$  in verschiedene Kategorien ein; nach direktem Auftreten und nach

Auftreten über Reaktionsraten  $r_j$ , welche entsprechend die Kettenregel erfordern. Wir führen nun alle Ableitungen aus, außer das Nachdifferenzieren der Reaktionsraten:

$$\begin{aligned}
& -c_{p,s} \frac{\partial \tilde{\vartheta}_s}{\partial t} - \frac{1}{\text{Pe}} \frac{\partial^2 \tilde{\vartheta}_s}{\partial z^2} = \tilde{\vartheta}_a \frac{\vartheta_a}{c_p} \left( St_{as} + c_p \sum_{j=\text{ox}1,2} \bar{\nu}_j^+ Da_j r_j \right) + \tilde{\vartheta}_c \left( \vartheta_c \bar{\nu}_{\text{red}}^- Da_{\text{red}} r_{\text{red}} \right. \\
& \quad \left. + \frac{\vartheta_c St_{cs}}{c_p} \right) + \tilde{\vartheta}_a \left( \frac{St_{as}}{c_p} + \sum_{j=\text{ox}1,2} \bar{\nu}_j^+ Da_j r_j \right) + \tilde{\vartheta}_c \left( \bar{\nu}_{\text{red}}^- Da_{\text{red}} r_{\text{red}} + \frac{St_{cs}}{c_p} \right) \\
& \quad + \tilde{\vartheta}_s \left( -St_{as} - St_{cs} + c_p \bar{\nu}_{\text{red}}^+ Da_{\text{red}} r_{\text{red}} + c_p \sum_{j=\text{ox}1,2} \bar{\nu}_j^- Da_j r_j \right) \\
& \quad + \sum_{j=\text{ox}1,2} \frac{\partial r_j}{\partial \vartheta_s} \left\{ \sum_{k \in \mathcal{I}} \tilde{\chi}_{k,a} \vartheta_a (\nu_{k,j} - \chi_{k,a} \bar{\nu}_j) Da_j + \tilde{\vartheta}_a \vartheta_a \bar{\nu}_j^+ Da_j (\vartheta_s - \vartheta_a) \right. \\
& \quad \quad \left. + \tilde{\vartheta}_a \left( \bar{\nu}_j^+ Da_j (\vartheta_s - \vartheta_a) + \vartheta_a \bar{\nu}_j Da_j \right) + \tilde{i}_a F n_j Da_j \right. \\
& \quad \quad \left. + \tilde{\vartheta}_s \left( -[\Delta_R h_j + n_j \Phi_a^L] Da_j - (\vartheta_a - \vartheta_s) \bar{\nu}_j^- Da_j c_p \right) \right\} \\
& \quad + \frac{\partial r_{\text{red}}}{\partial \vartheta_s} \left\{ \sum_{k \in \mathcal{I}} \tilde{\chi}_{k,c} \vartheta_c (\nu_{k,\text{red}} - \chi_{k,c} \bar{\nu}_{\text{red}}) Da_{\text{red}} + \tilde{\vartheta}_c \vartheta_c \bar{\nu}_{\text{red}}^- Da_{\text{red}} (\vartheta_s - \vartheta_c) \right. \\
& \quad \quad \left. + \tilde{\vartheta}_c \left( \vartheta_c \bar{\nu}_{\text{red}}^- Da_{\text{red}} + \bar{\nu}_{\text{red}}^- Da_{\text{red}} (\vartheta_s - \vartheta_c) \right) - \tilde{i}_c F n_{\text{red}} Da_{\text{red}} \right. \\
& \quad \quad \left. + \tilde{\vartheta}_s \left( -[\Delta_R h_{\text{red}} - n_{\text{red}} (U_{\text{cell}} - \Phi_c^L)] Da_{\text{red}} - (\vartheta_c - \vartheta_s) \bar{\nu}_{\text{red}}^+ Da_{\text{red}} c_p \right) \right\}. \tag{4.22}
\end{aligned}$$

Die ersten drei Zeilen enthalten die Terme aus direktem Auftreten. An der adjungierten Variable erkennen wir jeweils, aus welcher rechten Seite der Summand durch Ableiten nach  $\vartheta_s$  hervorgegangen ist. Wir erkennen bereits die Tatsache, dass die rechten Seiten an Länge noch einmal zunehmen (Ableiten von Produkten von Polynomen und Exponentialfunktionen in den Reaktionsraten  $r_j$ ). Um Übersicht zu gewinnen, definieren wir daher passende Hilfsvariablen. Uns fällt auf, dass alle Zustände, welche in einer Reaktionsrate auftreten, den gleichen Ausdruck  $\{\dots\}$  als Vorfaktor haben werden und sich nur im Nachdifferenzieren der Reaktionsrate unterscheiden werden. Dadurch motiviert definieren wir je Reaktionsrate eine Abkürzungen per  $\tilde{s}_j := \sum_{q \in \mathcal{Z}} \frac{\partial f^q}{\partial r_j}$ . Diese Abkürzungen findet man ausgerechnet in Anhang B.1; dort findet man auch die rechten Seiten der anderen adjungierten PDE.

Die rechten Seiten der Stromdichten und Potentiale stehen in Anhang B.2, gefolgt von den partiellen Ableitungen der Reaktionsraten in B.3. Die adjungierte Brenn- und Mischkammer steht in Anhang B.4. Die Übergangsbedingungen (4.16) erzeugen sehr längliche rechten Seiten, welche in Anhang B.5 nachlesbar sind.

#### 4.4.3 Beitrag des Zielfunktionals zu den adjungierten Gleichungen

Das Zielfunktional liefert die einzige Inhomogenität des linearen PDAE-Systems der adjungierten Gleichungen. Je nach Art des Zielfunktionals erscheinen folgende Inhomogenitäten auf der rechten Seite der entsprechenden PDE:

Zielfunktional	Gleichung für	Summand auf der rechten Seite
$Z_1$	$\tilde{\vartheta}_s$	$\vartheta_s - \vartheta_{\text{soll}}$
$Z_2$	$\tilde{\vartheta}_s$	$-(\Delta \vartheta_s - \Delta \vartheta_{\text{soll}})$
$Z_3$	$\tilde{U}_{\text{cell}}$	$U_{\text{cell}} - U_{\text{cell,soll}}$

$Z_2$  bedarf einer weiteren Erklärung: Wegen

$$D_{\vartheta_s} Z_2 \varphi = \iint_Q \left( \frac{\partial \vartheta_s}{\partial z} - \frac{\partial \vartheta_{\text{soll}}}{\partial z} \right) \cdot \frac{\partial \varphi}{\partial z} dz dt \quad (4.23)$$

tritt nach partieller Integration  $-\Delta \vartheta_s + \Delta \vartheta_{\text{soll}}$  als Quelle in der PDE von  $\tilde{\vartheta}_s$  auf; zusätzlich entstehen Randterme, welche entsprechend in den Randbedingungen von  $\tilde{\vartheta}_s$  landen; diese lauten dann

$$-\frac{1}{\text{Pe}} \frac{\partial \tilde{\vartheta}_s}{\partial z} = \frac{\partial \vartheta_s}{\partial z} - \frac{\partial \vartheta_{\text{soll}}}{\partial z} \quad \text{auf } \{0\} \times (0, t_f), \quad (4.24)$$

$$\frac{1}{\text{Pe}} \frac{\partial \tilde{\vartheta}_s}{\partial z} = \frac{\partial \vartheta_s}{\partial z} - \frac{\partial \vartheta_{\text{soll}}}{\partial z} \quad \text{auf } \{1\} \times (0, t_f). \quad (4.25)$$

Diese Ergebnisse stehen in Übereinstimmung zu denen aus Clever und Lang [31]. Dort wird in höheren Ortsdimensionen die Norm des Gradienten auf 0 geführt, um Temperaturspannungen bei der Glaskühlung zu vermeiden.

#### 4.4.4 Gleichungen für die Steuerungen

Da wir insgesamt ein Minimum der Lagrange-Funktion bzgl. den primären Variablen suchen, müssen wir auch die Richtungsableitungen nach den Steuerungen berechnen. Ohne Steuerungsbeschränkungen müssen diese alle verschwinden. Unter Beachtung von Steuerungsbeschränkungen müssen die Richtungsableitungen in alle zulässigen Richtungen nichtnegativ sein. Wie üblich bei Optimierungsproblemen in Banach-Räumen erfüllt die optimale Steuerung  $\vec{u}$  daher Variationsungleichungen der Form

$$\int_0^{t_f} \vec{f}'(\vec{u}) (\vec{u} - \vec{u}) dt \geq 0 \quad \forall \vec{u} \in U_{\text{ad}}. \quad (4.26)$$

Dabei ist  $\vec{f}'(\vec{u})$  der Gradient des auf die Steuerungen reduzierten Zielfunktional  $f(\vec{u})$  [96]. Wir können ihn als Gradient der Lagrange-Funktion bzgl. den Steuerungen  $\vec{f}'(\vec{u}) = \nabla_{\vec{u}} \mathcal{L}$  berechnen, d.h. wir berechnen die Richtungsableitungen der Lagrange-Funktion bzgl. der einzelnen Steuerungen und identifizieren dann daraus den Gradienten mit dem Satz von Riesz.

Wiederholen wir kurz den Teil  $\mathcal{L}_c$  der Lagrange-Funktion, welcher von den 7 Steuerungen abhängt, wobei wir die  $\nu_{\text{O}_2, Ck}$  aus Anhang A.5 einsetzen:

$$\begin{aligned} \mathcal{L}_c := & J + \int_0^{t_f} \vartheta_{a,\text{in}} \hat{\vartheta}_a + \left\{ \sum_{k \in \mathcal{I}} \chi_{k,a,\text{in}} \hat{\chi}_{k,a} \right\} + v_{a,\text{in}} \hat{v}_a + \xi R_{\text{back}} \\ & + \frac{\gamma_{\text{air}}}{\gamma_b} \vartheta_{\text{air}} \tilde{\vartheta}_b + \underbrace{\gamma_{a,\text{in}} \frac{\lambda_{\text{air}}}{\chi_{\text{O}_2,\text{air}}} \left( 2 \chi_{\text{CH}_4,a,\text{in}} + \frac{1}{2} \chi_{\text{H}_2,a,\text{in}} \right)}_{=f^{\gamma_{\text{air}}}} \tilde{\gamma}_{\text{air}} dt, \end{aligned}$$

mit der Abkürzung

$$\begin{aligned} \xi(t) := & \left\{ \sum_{k \in \mathcal{I}} \frac{\gamma_{c,\text{out}}(t)}{\gamma_b(t)} \chi_{k,c,\text{out}}(t) \tilde{\chi}_{k,b}(t) \right\} + \frac{\gamma_{c,\text{out}}(t)}{\gamma_b(t)} (\vartheta_{c,\text{out}}(t) - 1) \tilde{\vartheta}_b(t) \\ & + \gamma_{c,\text{out}}(t) \tilde{\gamma}_b(t). \end{aligned} \quad (4.27)$$

Da wir später nicht nur alle 7 Steuerungen auf einmal berechnen wollen, sondern auch Auswahlen davon optimal steuern möchten, müssen wir den Gradienten aus diesem Ausdruck jeweils durch Richtungsableitung bestimmen. Wir führen es hier für das Beispiel aller 7 Steuerungen mit den Substitutionen aus Abschnitt 4.3.1 auf. Nach Bilden der Richtungsableitung<sup>17</sup> setzen wir die Gleichungen (4.17a)-(4.17c) ein. Der Gradient ergibt sich darauf zu

$$\vec{f}'(\vec{u}(t)) = \begin{pmatrix} \lambda_1 u_1(t) + \tilde{v}_a(0, t) \\ \lambda_2 u_2(t) + u_1(t) \tilde{\vartheta}_a(0, t) \\ \lambda_3 u_3(t) + u_1(t) \tilde{\chi}_{\text{CH}_4, \text{a}}(0, t) + 2 u_5(t) \tilde{\gamma}_{\text{air}}(t) / \chi_{\text{O}_2, \text{air}} \\ \lambda_4 u_4(t) + u_1(t) \tilde{\chi}_{\text{H}_2, \text{a}}(0, t) + \frac{1}{2} u_5(t) \tilde{\gamma}_{\text{air}}(t) / \chi_{\text{O}_2, \text{air}} \\ \lambda_5 u_5(t) + \left(2u_3(t) + \frac{1}{2}u_4(t)\right) \tilde{\gamma}_{\text{air}}(t) / \chi_{\text{O}_2, \text{air}} \\ \lambda_6 u_6(t) + \frac{\gamma_{\text{air}}(t)}{\gamma_{\text{b}}(t)} \tilde{\vartheta}_{\text{b}}(t) \\ \lambda_7 u_7(t) + \xi(t) \end{pmatrix}. \quad (4.28)$$

Aufgrund der Substitution aus (4.4) ist der Gradient linear in den Steuerungen. Auf diesem Gradienten werden wir in Abschnitt 4.6 die Gradientenverfahren und Quasi-Newton-Verfahren aufsetzen.

Alle sieben Variationsungleichungen lassen sich analog zu [96, Lemma 2.26] punktweise formulieren. Dann erfüllt die optimale Steuerung  $(\bar{u}_1, \dots, \bar{u}_7)$  für fast alle  $t \in (0, t_f)$  folgende Variationsungleichungen in  $\mathbb{R}$ :

$$(\lambda_1 \bar{u}_1 + \tilde{v}_a(0, t)) (u - \bar{u}_1) \geq 0 \quad \forall u \in U_{\text{ad},1}, \quad (4.29\text{a})$$

$$(\lambda_2 \bar{u}_2 + \bar{u}_1 \tilde{\vartheta}_a(0, t)) (u - \bar{u}_2) \geq 0 \quad \forall u \in U_{\text{ad},2}, \quad (4.29\text{b})$$

$$(\lambda_3 \bar{u}_3 + \bar{u}_1 \tilde{\chi}_{\text{CH}_4, \text{a}}(0, t) + 2 \bar{u}_5 \tilde{\gamma}_{\text{air}} / \chi_{\text{O}_2, \text{air}}) (u - \bar{u}_3) \geq 0 \quad \forall u \in U_{\text{ad},3}, \quad (4.29\text{c})$$

$$(\lambda_4 \bar{u}_4 + \bar{u}_1 \tilde{\chi}_{\text{H}_2, \text{a}}(0, t) + \frac{1}{2} \bar{u}_5 \tilde{\gamma}_{\text{air}} / \chi_{\text{O}_2, \text{air}}) (u - \bar{u}_4) \geq 0 \quad \forall u \in U_{\text{ad},4}, \quad (4.29\text{d})$$

$$(\lambda_5 \bar{u}_5 + \left(2\bar{u}_3 + \frac{1}{2}\bar{u}_4\right) \tilde{\gamma}_{\text{air}} / \chi_{\text{O}_2, \text{air}}) (u - \bar{u}_5) \geq 0 \quad \forall u \in U_{\text{ad},5}, \quad (4.29\text{e})$$

$$(\lambda_6 \bar{u}_6 + \frac{\gamma_{\text{air}}}{\gamma_{\text{b}}} \tilde{\vartheta}_{\text{b}}) (u - \bar{u}_6) \geq 0 \quad \forall u \in U_{\text{ad},6}, \quad (4.29\text{f})$$

$$(\lambda_7 \bar{u}_7 + \xi) (u - \bar{u}_7) \geq 0 \quad \forall u \in U_{\text{ad},7}. \quad (4.29\text{g})$$

Im Gegensatz zu Prototyp-Problemen treten beim vorliegenden Optimalsteuerungsproblem auch Zustände sowie Produkte von adjungierten Variablen und Steuerungen auf. Zusammen mit den Zustandsgleichungen, den adjungierten Gleichungen, den Anfangs- und Randbedingungen vervollständigen diese Variationsungleichungen das Optimalitätssystem.

Die skalaren Variationsungleichungen für  $\bar{u}_1, \bar{u}_6, \bar{u}_7$  lassen sich wie im Lehrbuch auflösen zu expliziten Projektionsformeln<sup>18</sup>

$$\bar{u}_1 = P_{U_{\text{ad},1}} \left\{ -\frac{1}{\lambda_1} \tilde{v}_a(0, t) \right\}, \quad \bar{u}_6 = P_{U_{\text{ad},6}} \left\{ -\frac{1}{\lambda_6} \frac{\gamma_{\text{air}}}{\gamma_{\text{b}}} \tilde{\vartheta}_{\text{b}} \right\}, \quad \bar{u}_7 = P_{U_{\text{ad},7}} \left\{ -\frac{1}{\lambda_7} \xi \right\}. \quad (4.30\text{a})$$

<sup>17</sup>  $\chi_{\text{H}_2\text{O}, \text{a}, \text{in}}$  und  $\chi_{\text{H}_2\text{O}, \text{a}}$  werden eliminiert, siehe Abschnitt 4.4.5.

<sup>18</sup> Die Projektion ist wie in Kap. 3 definiert, also  $P_{U_{\text{ad},k}} \{x\} = \min(u_{\text{o},k}, \max(u_{\text{u},k}, x))$ .

Einseitig gekoppelte bzw. gestaffelte Variationsungleichungen lassen sich unter Beachtung der Reihenfolge ebenso klassisch lösen, allerdings treten hier Verkettungen von Projektionen auf:

$$\bar{u}_2 = P_{U_{\text{ad},1}} \left\{ -\frac{1}{\lambda_2} \bar{u}_1 \tilde{\vartheta}_a(0, t) \right\}. \quad (4.30b)$$

Dies ist so zu verstehen, dass zuerst das  $\bar{u}_1$  aus seiner Projektionsformel berechnet wird und anschließend bei bekanntem  $\bar{u}_1$  die Projektionsformel für  $\bar{u}_2$  ausgeführt wird. Als Konsequenz kann  $\bar{u}_2(t)$  daher theoretisch auch bei regulären Problemen im Inneren von  $U_{\text{ad},2}$  Knicke aufweisen.

Probleme bereiten dagegen beidseitig gekoppelte Systeme von Variationsungleichungen wie das System für  $\bar{u}_3$ ,  $\bar{u}_4$  und  $\bar{u}_5$  zusammen, da hier ein gestaffeltes Auflösen nicht ohne Fallunterscheidungen möglich ist. Obwohl die ersten Faktoren der Variationsungleichungen linear in allen Steuerungen sind, kann ein derartiges reelles, „lineares“ Variationsungleichungsproblem bereits eine komplizierte Lösungsstruktur aufweisen, wie das folgende Beispiel zeigt:

**Beispiel 4.4.2** Wir betrachten ein System aus zwei linearen<sup>19</sup>, reellen Variationsungleichungen

$$\begin{aligned} (u + v - 2p)(\varphi - u) &\geq 0, & \forall \varphi \text{ mit } 0 \leq \varphi \leq 2, \\ (u - v)(\psi - v) &\geq 0, & \forall \psi \text{ mit } 1 \leq \psi \leq 3. \end{aligned}$$

Im unbeschränkten Fall würden wir ein lineares Gleichungssystem mit der eindeutigen Lösung  $u = v = p$  erhalten. Diese Lösung gilt auch im beschränkten Fall, falls beide Schaltfunktionen  $g := u + v - 2p$ ,  $h := u - v$  verschwinden, was für  $1 \leq p \leq 2$  möglich ist. Allerdings treten mit den Beschränkungen gleichzeitig viele andere Lösungen auf, welche sich stellenweise dreifach überschneiden. Mittels Fallunterscheidung berechnet man die Lösungsstruktur zu

<u>Fälle:</u>	<u>Lösung:</u>	<u>Parameterbereich:</u>
$g = 0, \quad h = 0:$	$u = v = p$	$p \in \mathbb{R} \cap [1, 2] = [1, 2],$
$h > 0:$	$v = 1, u = 2p - 1$	$p \in (1, \infty) \cap \left[ \frac{1}{2}, \frac{3}{2} \right] = \left( 1, \frac{3}{2} \right],$
$h < 0:$	$v = 3, u = 2p - 3$	$p \in (-\infty, 3) \cap \left[ \frac{3}{2}, \frac{5}{2} \right] = \left[ \frac{3}{2}, \frac{5}{2} \right],$
$g > 0, \quad h < 0:$	$u = 0, v = 3$	$p < \frac{3}{2},$
$g < 0, \quad h > 0:$	$u = 2, v = 1$	$p > \frac{3}{2},$
$h < 0:$	$u = 2, v = 3$	$p > \frac{5}{2},$
$h = 0:$	$v = u = 2$	$p > 2.$

Die ersten drei Fälle, bei denen die Lösungen vom Parameter  $p$  abhängen, werden dabei zuletzt mit den Box-Beschränkungen in Einklang gebracht.

<sup>19</sup> Mit linear bezeichnen wir hier Variationsungleichungen, deren zugrundeliegende Gradienten linear in den Optimierungsvariablen sind.

Glücklicherweise haben wir es bei Optimalsteuerungsproblemen nicht mit derart allgemeinen Systemen von Variationsungleichungen zu tun: Die einzelnen Variationsungleichungen stammen von Richtungsableitungen einer skalaren Funktion, der Lagrange-Funktion. Entsprechend können wir die Systeme wieder in ein Optimierungsproblem umschreiben: Die drei gekoppelten Variationsungleichungen (4.29c)–(4.29e) sind notwendige Bedingungen erster Ordnung des Minimierungsproblems

$$\begin{aligned} \min_{(u_3, u_4, u_5) \in U_{\text{ad},3} \times U_{\text{ad},4} \times U_{\text{ad},5}} \varrho(u_3, u_4, u_5) := & \frac{1}{2} (\lambda_3 u_3^2 + \lambda_4 u_4^2 + \lambda_5 u_5^2) + u_1 (u_3 \hat{\chi}_{\text{CH}_4, \text{a}} \\ & + u_4 \hat{\chi}_{\text{H}_2\text{O}, \text{a}} + (1 - u_3 - u_4) \hat{\chi}_{\text{H}_2, \text{a}}) + \tilde{\gamma}_{\text{air}} \frac{u_5}{\chi_{\text{O}_2, \text{air}}} \left( 2u_3 + \frac{1}{2} u_4 \right). \end{aligned} \quad (4.31)$$

Aufgelistet sind hier entsprechend alle Teile der Lagrange-Funktion, welche von  $u_3$ ,  $u_4$  oder  $u_5$  abhängen. Dass  $(\bar{u}_3, \bar{u}_4, \bar{u}_5)$  Lösung dieses Minimierungsproblems sind, müsste man allerdings z.B. aus dem Maximumprinzip herleiten.

Für bekannte Zustände und adjungierte Zustände bedeutet dieses Minimierungsproblem also je Zeitschritt ein quadratisches Programm (QP) in 3 reellen Variablen. Die Hesse-Matrix von  $\varrho$  lautet mit  $\mu := \frac{\tilde{\gamma}_{\text{air}}}{2\chi_{\text{O}_2, \text{air}}}$

$$\text{Hess}_\varrho = \begin{pmatrix} \lambda_3 & 0 & 4\mu \\ 0 & \lambda_4 & \mu \\ 4\mu & \mu & \lambda_5 \end{pmatrix}. \quad (4.32)$$

Im Spezialfall  $\lambda_3 = \lambda_4 = \lambda_5 =: \alpha$  hat sie die Eigenwerte  $\alpha > 0$ ,  $\alpha + \sqrt{17\mu^2} > 0$  und  $\alpha - \sqrt{17\mu^2}$ . Das Optimierungsproblem ist in diesem Fall genau dann konvex, wenn  $|\tilde{\gamma}_{\text{air}}| \in \left[0, \frac{2}{\sqrt{17}} \alpha \chi_{\text{O}_2, \text{air}}\right]$ .  $\varrho$  muss daher nicht zwingend konvex sein. Im nicht-konvexen Fall liegt die Lösung stets auf dem Rand, d.h. die Box-Beschränkungen müssen dann stets aktiv sein. Denn ein Minimum im Inneren ist aufgrund des negativen Eigenwertes der konstanten Hesse-Matrix ausgeschlossen. Da sich  $\varrho$  mit der Zeit ändert, ist es denkbar, dass das Minimum von einem auf einen anderen Rand springt; die optimalen Steuerungen werden unstetig. Bei OC-ODE wurde ein derartiges Sprungverhalten z.B. bei der Steuerung eines Wiedereintrittsmanövers eines Spaceshuttles in [63] beobachtet. Ein Sprung in das Innere der Box-Beschränkungen ist dagegen nicht möglich.

**Bemerkung 4.4.3** Wählt man in der Modellierung dagegen die Steuerungen derart aus, dass sie auch in der Gleichung von  $\gamma_{\text{air}}$  linear auftreten, so können alle Variationsungleichungen als Projektionsformeln geschrieben werden. Dies ist z.B. der Fall, wenn wir nur  $(u_1, u_6, u_7)$  oder nur  $(u_1, u_2, u_5, u_6, u_7)$  steuern.

Bei den numerischen Verfahren in Abschnitt 4.6 brauchen wir nur den Gradienten  $\vec{f}'(\vec{u}(t))$ . Die Auflösbarkeit der Variationsungleichungen spielt nur bei der Fixpunktiteration eine Rolle. Die Konstruktion von Testbeispielen in Abschnitt 4.6.3 funktioniert in der präsentierten Form dagegen nur, wenn wir Projektionsformeln haben.

#### 4.4.5 Elimination von Molenbrüchen

In [90] haben wir die PDE für  $\chi_{N_2,c}$  sowie  $\chi_{H_2O,a}$  eliminiert durch Ausnutzen der algebraischen Gleichungen

$$\sum_k \chi_{k,a}(z, t) = 1, \quad \sum_k \chi_{k,c}(z, t) = 1 \quad \forall (z, t) \in Q. \quad (4.33)$$

Da diese beiden Zustände damit vollständig aus dem Modell entfernt werden, erhalten wir für sie auch keine adjungierten Gleichungen (man vergleiche die Eliminationsmethode bei OC-ODE aus Abschnitt 1.3.1). Um ein indirektes Verfahren programmieren zu können, müssen wir wissen, wie sich diese Elimination auf die restlichen adjungierten Gleichungen auswirkt.

Wir stellen fest, dass  $\chi_{N_2,c}$  von Grund auf in keiner einzigen Gleichung auftritt; daher beeinflusst dessen Elimination die adjungierten Gleichungen nicht. Auch  $\chi_{H_2O,a}$  tritt selbst nicht explizit in den Gleichungen auf, aber dafür in den Reaktionsraten. Jede andere Spezies tritt daher zum einen explizit auf, zum anderen mit negativem Vorzeichen an der Stelle von  $\chi_{H_2O,a}$ . Wir müssen also z.B.  $\frac{\partial r_{\text{refl}}}{\partial \chi_{CH_4,a}} - \frac{\partial r_{\text{refl}}}{\partial \chi_{H_2O,a}}$  anstatt  $\frac{\partial r_{\text{refl}}}{\partial \chi_{CH_4,a}}$  verwenden und jeweils  $\chi_{H_2O,a}$  durch obige Summe eliminieren. Am Gradienten muss dagegen nichts verändert werden.

#### 4.4.6 Erweiterung auf 2D

Die Verallgemeinerung des Optimalitätssystems auf das 2D-Modell ist problemlos möglich. Dabei ändern sich bei den adjungierten PDE lediglich die differentiellen Anteile, die länglichen Quellterme bleiben unverändert. Zusätzlich sind die Ortsintegrale der Mittelungsbedingungen [93, (3.29)-(3.31), (3.49)-(3.51)] am Eingang der Brennkammer zu berücksichtigen. Es handelt sich hierbei um eine weitere ODE-PDE Kopplung. Diese Mittelungsbedingungen führen analog zu den bisherigen Erkenntnissen zu örtlich konstanten Quelltermen in den Randbedingungen der adjungierten PDE der Gaskanäle ähnlich dem Effekt in Gleichung (4.20e).

#### 4.4.7 Zusammenfassung der notwendigen Bedingungen

Wir haben gesehen, dass die adjungierten Gleichungen jeweils vom gleichen Typus (parabolisch, hyperbolisch, gewöhnlich, algebraisch) sind. Allerdings kehren sich die Vorrugsrichtungen um: Sowohl die Ausbreitungsrichtung bei hyperbolischen PDE, als auch die Zeitpropagation bei parabolischen PDE, Transportgleichungen und ODE kehren sich um. Ebenso wirken die Kopplungen der verschiedenen Variablen über die Randbedingungen in die entgegengesetzte Richtung, so dass sich der gesamte Informationsfluss umkehrt. Dies ist dargestellt in Abbildung 4.5. Wir erhalten ein vollständig gekoppeltes Optimalitätssystem.

Die Brennkammer spielt die zentrale Rolle in dem Diagramm. Während sie bei der MCFC viele Eingänge besitzt, wirkt sie in der adjungierten MCFC als Verteiler. Die blauen Boxen stehen bei der MCFC für die Inputs. Bei der adjungierten MCFC sind es dagegen die Outputs. Die Variationsungleichungen generieren aus den Outputs der adjungierten MCFC entsprechend die Inputs der MCFC (rote Pfeile).



## 4.5 Direkte Verfahren der Optimalsteuerung

Direkte Verfahren bieten einen naheliegenden Ansatz, derartige Optimalsteuerungsprobleme zu lösen. Erste Ergebnisse des 2D MCFC Modells mittels direkter Verfahren findet man in [93]. Ausgefeiltere direkte Verfahren entwickeln wir für das 1D Modell in [90].

Wir arbeiten dort mit der reduzierten Formulierung (siehe Abschnitt 3.5.1), so dass die Anzahl der Freiheitsgrade des Optimierers deutlich reduziert ist. Gerade wenn die Ableitungen mit numerischer Differentiation erzeugt werden, ist dies essentiell. Darüber hinaus werden die Zustandsgleichungen nicht über eine volle Diskretisierung – wie beim Rocketcar in Abschnitt 3.5.2 – gelöst, sondern mit einer bzgl. Rechenzeit und Speicherbedarf effizienteren Alternative, der schrittweisen Integration: Mit einer Semidiskretisierung im Ort, gefolgt von einem Mehrschrittverfahren. Wir verwenden `ode15s` unter `MATLAB`, welches ein adaptiver impliziter DAE-Solver variabler Ordnung (zwischen 1 und 5) ist. Als Optimierer wird `fmincon` verwendet, also der etablierte NLP-Solver von `MATLAB`.

Wir erkennen dort einige Schwierigkeiten des MCFC-Modells, welche den Einsatz direkter Verfahren betreffen. Gerade die verschiedenen Zeitskalen der Zustände machen Probleme: Die sehr schnellen Variablen reagieren schlagartig auf eine sprunghaft geänderte Steuerung mit impulsartigen Überschwüngen. Je öfter die Steuerung springt, desto instabiler wird das System. Auf der anderen Seite brauchen wir sehr kleine Zeitschritte, um die schnellen Variablen auflösen zu können und um das steife DAE-System aus der Linienmethode hinreichend genau zu lösen. Stückweise konstante Steuerungen auf diesem feinen Zeitgitter der Zustände bereiten entsprechend Schwierigkeiten. Der Konvergenzbereich des direkten Verfahrens ist dann zu sehr eingeschränkt.

Daher wählen wir ein deutlich gröberes Zeitgitter  $0 = t_1^u < \dots < t_j^u < \dots < t_{N_u}^u = t_f$  für die Steuerungen. Zusätzlich entscheiden wir uns für stetige Steuerungen  $u_i$  aus linearen Splines durch die Stützpunkte  $(t_j^u, d_j^i)$ ,  $j = 1, \dots, N_u$ ,  $i = 1, \dots, 7$ .

Einen höheren Konvergenzbereich und eine niedrigere Laufzeit des Optimierers kann man erreichen, indem man zunächst Methoden der instantanen Steuerung anwendet. Derartige Ideen findet man z.B. in Hinze und Kunisch [58], Tröltzsch und Wachsmuth [97], sowie in Eppler und Tröltzsch [34]. Sie liefern – bei den hier untersuchten Zielfunktionalen – in kurzer Zeit sehr gute Steuerungen, welche dann als Startschätzung hergenommen werden können.

Damit konnten wir mit Einsatz von `fmincon` unter `MATLAB` in [90] gute Ergebnisse erzielen. Allerdings kann man sich dennoch nicht viele Steuerpunkte erlauben, da der Aufwand eines Optimierungsschrittes stark mit der Anzahl Steuerpunkte anwächst – aufgrund der numerischen Differentiation. Wir sehen das in Abschnitt 4.7 anhand einer extrem feinen Auflösung der Steuerungen.

In den nächsten beiden Abschnitten wird das Optimalsteuerungsproblem der MCFC erstmalig mit adjungierten-basierten Methoden gelöst. Dazu werden zunächst Methoden ausgewählt und entwickelt. In Abschnitt 4.7 folgen die numerischen Experimente.

## 4.6 Indirekte Verfahren der Optimalsteuerung

In Abschnitt 4.4 wurden notwendige Bedingungen 1. Ordnung (Optimalitätsbedingungen) hergeleitet zur optimalen Steuerung des Brennstoffzellenmodells. Darauf aufbauend werden wir in diesem Abschnitt indirekte Methoden entwickeln, d.h. iterative Methoden zur Lösung der Optimalitätsbedingungen. Zunächst erörtern wir, welche der etablierten Verfahren auf komplexe Anwendungsprobleme von der Form der MCFC anwendbar sind.

### 4.6.1 Auswahl möglicher Verfahren

Analog zur Backward-Sweep-Methode entscheiden wir uns dafür, nur gestaffelte Verfahren zu betrachten, d.h. Verfahren bei denen die Zustandsgleichungen und die adjungierten Gleichungen separat gelöst werden. Für Optimalsteuerungsprobleme bei zeitabhängigen PDAE-Systemen versprechen diese Effizienz in Bezug auf Speicher und Rechenzeit, da sowohl die Zustandsgleichungen als auch die adjungierten Gleichungen schrittweise in der Zeit gelöst werden können, so dass die Rechenzeit linear in der Anzahl der Zeitschritte skaliert. Im Gegensatz zu einer vollen Diskretisierung können wir uns dadurch sehr viele Zeitschritte erlauben und folglich die unterschiedlichen Zeitskalen sehr gut auflösen. Aber auch der nichtlineare Gleichungslöser profitiert hiervon: So besitzt er dank dem vorigen Zeitschritt bereits gute Startschätzungen und kann oft sogar die Jacobi-Matrix aus dem letzten Zeitschritt übernehmen (siehe Dokumentation von `ode15s`).

Ein anderer Einsatzzweck gestaffelter Verfahren sind Umgebungen, bei denen der Zustandssolver und der adjungierte Solver als Black Box vorliegen. Nachteil der vorgestellten gestaffelten Verfahren ist es, dass nicht die Konvergenzgeschwindigkeit z.B. eines Newton- bzw. SQP-Verfahrens erreicht wird.

Es kommen vier gestaffelte Verfahren in Frage, welche wir in diesem Abschnitt besprechen werden. Erste gestaffelte Verfahren bietet die Klasse der Gradientenverfahren. Für unser vorliegendes Problem mit Steuerungsbeschränkungen bieten sich das bedingte Gradientenverfahren und das projizierte Gradientenverfahren an (siehe [96], [52]). Gradientenverfahren sind i.d.R. global konvergent mit linearer Konvergenzrate.

Der Gradient lässt sich aber auch für Quasi-Newton-Verfahren verwenden, wie das BFGS-Verfahren (siehe z.B. [74]), welches ebenfalls gestaffelte Verfahren sind. Bei Quasi-Newton-Verfahren wird aus den Gradienten der bisherigen Schritte nach und nach eine bessere Approximation der Hesse-Matrix erstellt. Derartige Verfahren erreichen superlineare Konvergenz. Bei NLP sind sie stark verbreitet, zumal man im Ggs. zu einem Newton-Verfahren keine zweiten Ableitungen benötigt und zusätzliche Eigenschaften wie globale Konvergenz einbauen kann. Gradienten- und Quasi-Newton-Verfahren sind die Methoden, welche wir auf das MCFC-Problem anwenden können.

Ausgefeiltere Methoden wie die Primal-Duale Aktive-Mengen-Strategie (PDAS), das halbglatte Newton-Verfahren oder SQP-Verfahren sind dagegen nicht anwendbar. Sie benötigen in den Zwischenschritten die Lösung von gekoppelten Systemen aus (ggf. linearisierten) Zustandsgleichungen und adjungierten Gleichungen. Um derartige Methoden im vorliegenden Fall anzuwenden, benötigen wir eine volle Diskretisierung. Für das 1D-Modell wäre eine volle Diskretisierung bei heutigem Rechnerstand möglich. Für komplexe Anwendungsprobleme in 2D oder 3D kommt sie dagegen nicht in Frage.

Für Lagrange-Newton bzw. SQP-Verfahren benötigt man zudem das linearisierte KKT-System, welches zwei weitere PDAE-Systeme umschließt: Die linearisierten Zustandsgleichungen sowie die (in allen Variablen) linearisierten adjungierten Gleichungen. Für letztere braucht man die zweiten partiellen Ableitungen der rechten Seiten. Der Aufwand deren Herleitung ist daher erneut in der Größenordnung der Herleitung der adjungierten Gleichungen. Die Komplexität und dadurch bedingte Fehleranfälligkeit der Gleichungen dürften diesen Weg in der Praxis verbauen. Einzigen Lichtblick bietet hier der konsequente Einsatz von AD-Tools. Verfahren mit zweiten Ableitungen werden im Rahmen dieses Kapitels nicht untersucht.

Wir sprechen nun die einzelnen gestaffelten Verfahren an, beginnend mit den beiden Gradientenverfahren. Da sich im Ergebnis das Quasi-Newton-Verfahren als überlegen erwiesen hat, kommen wir relativ schnell darauf zu sprechen. Als viertes gestaffeltes Verfahren fügen wir zuletzt die Fixpunktiteration hinzu.

### Bedingtes Gradientenverfahren

Das bedingte Gradientenverfahren aus [96] lässt sich direkt zur Steuerung des MCFC-Modells nutzen. Der Übersicht halber bezeichnen wir die Variablen abstrakt als  $y_k$  (alle Zustände zur Iteration  $k$ ),  $u_k$  (alle Steuerungen) und  $p_k$  (alle adjungierten Zustände). Damit lautet das bedingte Gradientenverfahren:

#### Algorithmus 4.6.1 (Bedingtes Gradientenverfahren)

1. Sei  $u_1$  gegeben. Setze  $k = 1$ .
2. Berechne  $y_k$  aus den Zustandsgleichungen zu  $u_k$ .
3. Berechne  $p_k$  aus den adjungierten Gleichungen zu  $(y_k, u_k)$ .
4. Berechne den Gradienten des reduzierten Zielfunktional  
 $g_k = (g_k^i)_{i=1,\dots,7} = f'(u_k)$  unter Verwendung von  $(y_k, u_k, p_k)$ .
5. Formiere die Suchrichtung  $v_k$  durch

$$v_k^i = \begin{cases} u_{u,i}, & \text{für } g_k^i > 0, \\ \frac{1}{2}(u_{u,i} + u_{o,i}), & \text{für } g_k^i = 0, \\ u_{o,i}, & \text{für } g_k^i < 0. \end{cases}$$

6. Berechne eine Schrittweite  $s_k \in (0, 1]$  durch Schrittweitensuche.
7. Setze  $u_{k+1} := u_k + s_k \cdot v_k$ ,  $k = k + 1$  und gehe zu Schritt 2.

Natürlich ist noch ein entsprechendes Abbruchkriterium einzubauen (z.B. Norm des Gradienten klein etc.). Der mittlere Fall der Suchrichtung ist bekanntlich numerisch auszuschließen.

Das besondere am bedingten Gradientenverfahren ist, dass  $u_k + s \cdot v_k \in U_{\text{ad}}$  automatisch für alle  $s \in [0, 1]$  erfüllt ist – als Konvexkombination zweier zulässiger Steuerungen. Daher ist die Liniensuche gleichbedeutend mit einem Minimierungsproblem einer Funktion einer Veränderlichen  $s$ , die allerdings nicht analytisch gegeben ist. Im Gegensatz zu linear-quadratischen Problemen können wir die Schrittweitsuche daher nicht exakt ausführen, sondern greifen auf eine Strategie der Liniensuche aus der nichtlinearen Optimierung zurück, z.B. eine Schrittweitenstrategie nach Armijo (siehe z.B. [41], [74]).

### Gradienten-Projektionsverfahren

Eine schnellere Konvergenz erhält man i.d.R. mit dem Gradienten-Projektionsverfahren [96]. Obiger Algorithmus bleibt gültig, lediglich die Suchrichtung wird auf den Antigradienten geändert:

$$v_k := -g_k. \quad (4.34)$$

Da die vorgeschlagenen Steuerungen  $u_k + s_k \cdot v_k$  nun nicht mehr automatisch zulässig sind, projiziert man sie noch auf den zulässigen Bereich. Die Schrittweitenstrategie lautet also:

#### Algorithmus 4.6.2 (Schrittweitenstrategie des Gradienten-Projektionsverf.)

1. Gegeben sei eine Startschrittweite  $s_0$ . Setze  $m = 0$ .
2. Setze  $w_m := P_{U_{\text{ad}}} \{u_k + s_m \cdot v_k\}$  und löse die Zustandsgleichungen zu dieser Steuerung.
3. Wenn  $f(w_m)$  hinreichend kleiner als  $f(u_k)$ , akzeptiere die Schrittweite  $s_m$ .
4. Andernfalls berechne  $s_{m+1}$  z.B. nach Armijo, setze  $m = m + 1$  und gehe zu Schritt 2.

Obwohl wir in unserem Fall nicht alle Variationsungleichungen als Projektionsformeln schreiben können, liefert die Projektion an sich dennoch stets eine zulässige Steuerung.

### Quasi-Newton-Verfahren

Um ein Quasi-Newton-Verfahren an dieser Stelle einzusetzen, sollte man sich daher zunächst den Stand der Dinge vor Augen rufen: Wir haben einen adjungierten Solver und können aus dessen Ergebnissen den Gradienten des reduzierten Zielfunktional gemäß Abschnitt 4.4.4 formen. Wir können also sowohl den Wert als auch den Gradienten des reduzierten Zielfunktional angeben. Diese Bausteine reichen bereits aus, um Quasi-Newton-Verfahren anzuwenden zu können.

Da es sehr effiziente Quasi-Newton-Verfahren auf dem Markt gibt, möchten wir diese nutzen, um z.B. von effizienter Liniensuche zu profitieren. Wir gehen also den Schritt von indirekten Verfahren hin zum Aufruf einer NLP-Routine. Wir gehen dabei fast identisch zur direkten Methode vor: Wir rufen `fmincon` mit dem reduzierten Zielfunktional

auf. Der einzige Unterschied ist, dass wir diesmal den Gradienten mit übergeben, d.h. wir lösen zusätzlich innerhalb des Zielfunktional nach den Zustandsgleichungen noch die adjungierten Gleichungen und formen den (diskretisierten) Gradienten. Wir rufen `fmincon` auf zur Minimierung des reduzierten Zielfunktional. Die Unterfunktion des reduzierten Zielfunktional führt zu einem überreichten Steuervektor folgende Schritte in Reihenfolge aus:

1. Zustandsgleichungen lösen
2. Zielfunktional per Quadraturformel berechnen
3. Adjungierte Gleichungen lösen
4. Gradient formieren
5. Zielfunktionswert und Gradient zurückgeben

Abbildung 4.6: Ablauf einer Funktionsauswertung der Unterfunktion des Quasi-Newton-Verfahrens.

Beim direkten Verfahren fielen die Schritte 3, 4 weg und lediglich der Zielfunktionswert wurde zurückgegeben. `fmincon` produzierte die nötigen Ableitungen dann mittels numerischer Differentiation.

Die Aktive-Mengen-Strategie und das Innere-Punkte-Verfahren von `fmincon` verwenden dann eine Quasi-Newton-Methode mit BFGS-Update und kubischer Liniensuche. Eine höhere Genauigkeit erzielt man, wenn man nach Konvergenz des Quasi-Newton-Verfahrens ein paar Schritte einer Fixpunktiteration anschließt. Daher beschreiben wir zuletzt die Methode der Fixpunktiteration.

### Fixpunktiteration

Nicht vergessen sollte man die naheliegendste Methode, die Fixpunktiteration. Falls die Auswahl der Steuerungen es erlaubt, die Variationsungleichungen als Projektionsformeln oder geschachtelte Projektionsformeln (oder als noch allgemeinere Fixpunktformen) zu formulieren, dann lautet die Fixpunktiteration zur Lösung des KKT-Systems wie folgt:

#### Algorithmus 4.6.3 (Fixpunktiteration)

1. Sei  $u_k$  gegeben.
2. Berechne  $y_k$  aus den Zustandsgleichungen zu  $u_k$ .
3. Berechne  $p_k$  aus den adjungierten Gleichungen zu  $(y_k, u_k)$ .
4. Berechne  $u_{k+1}$  aus den Projektionsformeln zu  $(y_k, u_k, p_k)$ .
5. Iteriere Schritte 2. – 4. bis konvergent.

Allerdings konvergiert die Fixpunktiteration bekanntlich nur, falls die Kostenfaktoren  $\lambda_i$  in gewisser Weise groß genug sind, siehe [60, Sect. 3.5]. Da sich die Kontraktionsbedingung beim vorliegenden Problem schwer angeben lässt, beschränken wir uns auf numerische Studien.

Die Fixpunktiteration ließe sich sehr effizient mit der variationellen Diskretisierung aus Hinze [57] kombinieren. Dort ist vorgeschlagen, die Steuerung nicht zu diskretisieren, sondern ihre Diskretisierung anhand der Projektionsformel aus den Diskretisierungen der anderen Variablen zu erhalten. Innerhalb einer gestaffelten Methode lässt sich dies besonders effizient einbauen: Man löst beide PDAE-Systeme adaptiv und formt die neue Steuerung, wobei das zugehörige Gitter aus der Projektion und den Diskretisierungen von  $y$ ,  $p$  folgt. Einen solchen Projektionsschritt sollte man stets am Ende seiner Optimierung ausführen, um von der gesteigerten Fehlerordnung zu profitieren (vgl. [71]).

### 4.6.2 Adjungierter Solver und Wahl des Zeitgitters

Um ein gestaffeltes indirektes Verfahren aufsetzen zu können, braucht man zunächst einen adjungierten Solver. D.h. man muss in der Lage sein, die adjungierten Gleichungen für gegebene Steuerungen und Zustände zu simulieren. Da man an dieser Stelle bereits einen angepassten Zustandssolver geschrieben habt, kann man sein Vorwissen nutzen: Die adjungierten Gleichungen werden daher analog zu den Zustandsgleichungen diskretisiert, also mit der Linienmethode unter Anwendung von Upwind-Verfahren wie denen aus Kap. 4.2.1.

Nach Abschnitt 4.4.7 sind die Flussrichtungen in den adjungierten Gleichungen genau entgegengesetzt zu denen der Zustandsgleichungen. Die Techniken aus Abschnitt 4.2.1 sind entsprechend angepasst anzuwenden. Der hauptsächliche Programmieraufwand liegt an der (fehlerfreien) Bereitstellung der länglichen rechten Seiten und der Umsetzung der komplexen Kopplungsstruktur aus Abb. 4.5.

Anschließend möchten wir die adjungierten Gleichungen mit `ode15s` lösen. Da `ode15s` leider keine rückwärts gerichteten Gleichungen verarbeiten kann, müssen wir die inverse Zeittransformation  $\tau = t_f - t$ ,  $\tau \in [0, t_f]$  anwenden und die Vektoren konsistent umsortieren. Als Ergebnis erhalten wir einen in der Zeit adaptiven adjungierten Solver.

In [90] haben wir das Einschwingverhalten der Zustände eingehend untersucht. Es zeigt sich, dass zu Beginn und direkt nach einem Lastwechsel sehr viele Zeitschritte nötig sind, zu Ende hin dann immer weniger. Die adjungierten Zustände brauchen dagegen ein grundlegend anderes Gitter: Der adaptive DAE-Solver `ode15s` wählt bei der adjungierten MCFC zusätzlich viele Zeitschritte um die Startzeit  $t = t_f$  herum. Dies liegt daran, dass das adjungierte PDAE-System ebenso verschiedene Zeitskalen wie die Zustandsgleichungen zeigt. Allerdings führt hier die Endbedingung zu Schwingungen, welche im Laufe der (rückwärts laufenden) Zeit sich dann ausgleichen. Das unterschiedliche Einschwingen von Zuständen und adjungierten Zuständen sieht man exemplarisch in Abbildung 4.7. Während Zustände aufgrund der Anfangsbedingung und dem Lastwechsel (hier bei  $t = 300$ ) ins Schwingen kommen, sorgt bei den adjungierten Zuständen die Endbedingung für Schwingungen.

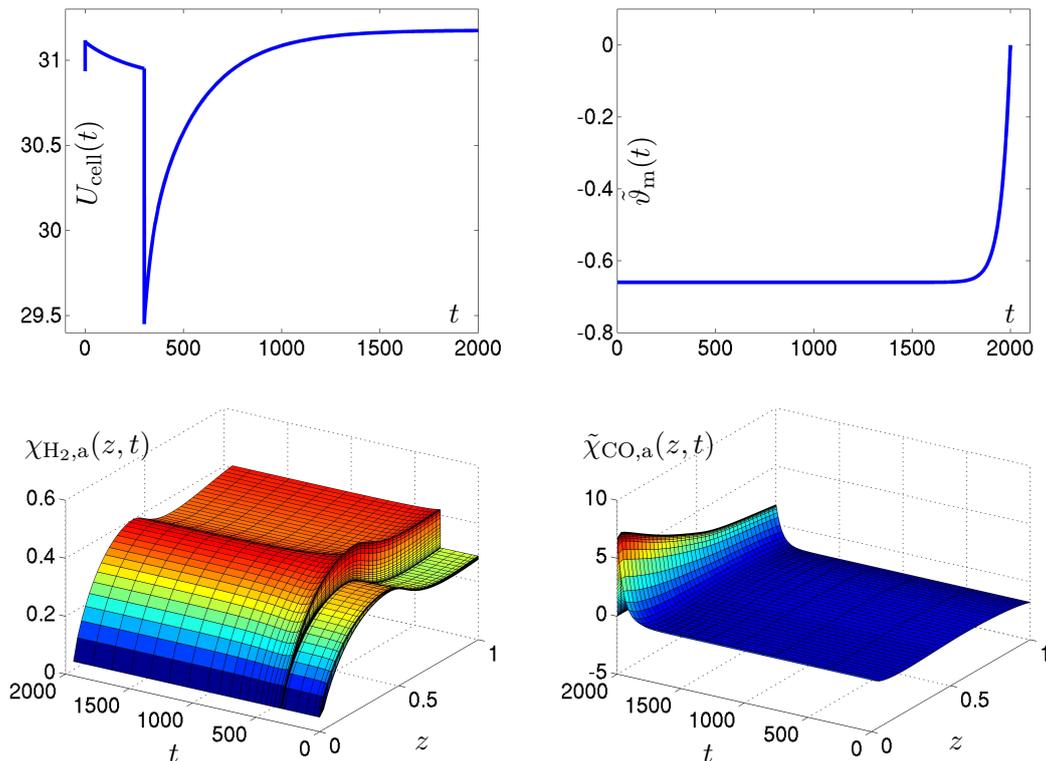


Abbildung 4.7: Links: Typisches Einschwingverhalten von Zuständen ( $U_{\text{cell}}(t)$ ,  $\chi_{\text{H}_2,\text{a}}(z, t)$ ) bei konstanten Steuerungen und einem Lastwechsel bei  $t = 300$ . Rechts: Typisches Einschwingverhalten von adjungierten Zuständen ( $\vartheta_m(t)$ ,  $\tilde{\chi}_{\text{CO},\text{a}}(z, t)$ ) bei konstanten Steuerungen, sowie zeitlich konstanten Zuständen und Sollzuständen.

Die adaptiven Gitter von Zuständen und adjungierten Zuständen sind folglich auch unter variablen Steuerungen in Bereichen kontrovers, siehe Abb. 4.8. Da die Steuerung von Zuständen und adjungierten Zuständen beeinflusst wird, empfiehlt es sich die Erkenntnisse zu kombinieren: Es wird die Vereinigung der beiden adaptiven Gitter als festes Gitter für einen Optimierungsdurchlauf verwendet. Die Steuerung lebt auf diesem Gitter. Zustände und adjungierte Zustände werden dennoch adaptiv gelöst, allerdings muss – innerhalb der gewählten Implementierung – das adaptive Gitter das feste Gitter zumindest enthalten. Wird eine Variable auf einem feineren Gitter abgerufen, so wird stets linear interpoliert.

#### Bemerkung 4.6.4 (Diskreter und kontinuierlicher adjungierter Zustand)

Da die Zustands- und adjungierten Gleichungen hier separat adaptiv gelöst werden, stimmt der erhaltene adjungierte Zustand i.A. nicht exakt mit dem diskreten adjungierten Zustand überein, vgl. Ulbrich [98], [103]. In Ziems und Ulbrich [103, Remark 3.7] wird gezeigt, dass sich dieses Problem mittels Gitterverfeinerung anhand Fehlerschätzern der kontinuierlichen Größen in den Griff bekommen lässt. Dabei werden die Gitter der Zustände und der adjungierten Zustände zugleich verfeinert.

Die Konvergenz des diskretisierten gegen den kontinuierlichen adjungierten Zustand ist in jedem Falle sicherzustellen, damit das korrekte Ergebnis erhalten wird, siehe auch Hager [46] im Falle von OC-ODE.

Bei den numerischen Ergebnissen in Abschnitt 4.7 wird daher eine sehr feine Zeitdiskretisierung gewählt, was möglich ist, da der Aufwand dank der schrittweisen Integration

nur linear in den Zeitschritten anwächst. Die Zeitgitter der Zustände und der adjungierten Zustände werden zugleich entsprechend genau verfeinert. Daher ist von einer Konvergenz gegen die kontinuierlichen adjungierten Zustände auszugehen. Die Ergebnisse werden dies später bestätigen.

Bereits erste Tests mit dem adjungierten Solver liefern ein interessantes Ergebnis: Wir erkennen zunächst numerisch und dann durch genaueren Blick auf die adjungierten Gleichungen, dass der adjungierte Wassergehalt in Kathode, Brennkammer und Mischkammer identisch verschwindet:

$$\tilde{\chi}_{\text{H}_2\text{O},c} \equiv 0, \quad \tilde{\chi}_{\text{H}_2\text{O},m} \equiv 0, \quad \tilde{\chi}_{\text{H}_2\text{O},b} \equiv 0. \quad (4.35)$$

Bei näherem Hinsehen erkennt man die Ursache:  $\chi_{\text{H}_2\text{O},c}$  ist an der Reduktionsreaktion nicht beteiligt. Der identisch verschwindende adjungierte Zustand verrät uns, dass die zugeordnete Zustandsgleichung sozusagen obsolet ist; sie wirkt sich nicht auf die Optimierung aus. Es gibt eine algebraische Relation, mit welcher die Zustandsvariable  $\chi_{\text{H}_2\text{O},c}$  aus dem System eliminiert werden kann. Und in der Tat könnten wir das Modell an dieser Stelle vereinfachen, indem wir die Tatsache nutzen, dass die Molenbrüche von  $\text{N}_2$  und  $\text{H}_2\text{O}$  in der Kathode, in Brenn- und Mischkammer stets in einem festen (zeitabhängigen) Verhältnis stehen:

$$\frac{\chi_{\text{H}_2\text{O},c}(z,t)}{\chi_{\text{N}_2,c}(z,t)} = \frac{\chi_{\text{H}_2\text{O},m}(t)}{\chi_{\text{N}_2,m}(t)} = \frac{\chi_{\text{H}_2\text{O},b}(t)}{\chi_{\text{N}_2,b}(t)} = c(t) \quad \forall z \in [0, 1]. \quad (4.36)$$

Aufbauend auf einem adjungierten Solver können wir nun gestaffelte indirekte Verfahren aufsetzen. Um die Effizienz der Optimierungsverfahren untersuchen zu können, erproben wir sie im nächsten Abschnitt zunächst an Testaufgaben und im Anschluss am gestellten Anwendungsproblem. Doch wie konstruiert man Testaufgaben bei derartig komplexen PDAE-Modellen?

### 4.6.3 Konstruktion von Testaufgaben

Zur fehlerfreien Programmierung benötigt man i.d.R. Testaufgaben, also Aufgabenstellungen bei denen man das Ergebnis bereits kennt. Für akademische Beispiele lassen sich diese oft analytisch erzeugen, siehe [96]. Hilfreiche Testbeispiele findet man z.B. für elliptische Randsteuerungsprobleme mit Steuerungsbeschränkungen in [25], für linear-elliptische, verteilte Steuerungsprobleme mit Zustandsbeschränkungen in [70] sowie [72]. Auch Parabolische Testbeispiele lassen sich so konstruieren, siehe z.B. [6], [83].

Bei Anwendungsproblemen kann man dagegen aufgrund der Komplexität keine analytischen Testbeispiele mehr erwarten. Daher erzeugen wir uns im Folgenden strukturiert numerische Testbeispiele. Beim vorliegenden KKT-System lassen sich numerische Testbeispiele jedoch erst nach Einführung einer weiteren Modellierungsgröße konstruieren. Um möglichst wenig zu verändern, führen wir daher eine Sollsteuerung  $\vec{u}_d = (u_k^d)_{k=1,\dots,7}$  ein, d.h. wir ändern die Steuerungskosten in (4.8) zu

$$\frac{1}{2} \sum_{k=1}^7 \lambda_k \int_0^{t_f} (u_k(t) - u_k^d(t))^2 dt. \quad (4.37)$$

Entsprechend ändert sich das KKT-System genau im jeweils ersten Summanden der einzelnen Komponenten des Gradienten: Aus  $\lambda_k u_k(t)$  wird entsprechend  $\lambda_k (u_k(t) - u_k^d(t))$ .

Dank diesem weiteren Freiheitsgrad können wir sehr schnell Testbeispiele generieren mit folgender Methodik:

1. Wähle die optimale Steuerung  $\vec{u}$  beliebig (aber zulässig)
2. Berechne die optimalen Zustände aus den Zustandsgleichungen
3. Wähle den Sollzustand  $U_{\text{cell,soll}}$ <sup>20</sup> beliebig und löse die adjungierten Gleichungen
4. Wähle ein  $\vec{u}_d$ , so dass die Variationsungleichungen erfüllt sind

Die Wahl von  $\vec{u}_d$  in Schritt 4 ist auf inaktiven Bereichen vorgegeben, auf aktiven Bereichen hat man dagegen Wahlfreiheiten. Man beachte, dass sich mit den erzeugten Testbeispielen Fehler im Zustandssolver oder im adjungierten Solver i.A. nicht erkennen lassen. Die Beispiele dienen hauptsächlich der Konvergenzanalyse der verschiedenen Optimierungsverfahren.

Erproben wir nun die programmierten Verfahren an derartigen Testbeispielen.

## 4.7 Numerische Experimente

In diesem Abschnitt betrachten wir die Ergebnisse indirekter Verfahren anhand einiger Beispiele. Im Zielfunktional  $J$  aus (4.8) wählen wir generell  $Z_3$  aus. Als Parameter seien  $t_f = 2000$  sowie ein Lastwechsel gemäß

$$I_{\text{cell}} = \begin{cases} 0.4, & t < 0.1, \\ 0.5, & t \geq 0.1, \end{cases}$$

gewählt. Im Ort wird mit  $N = 41$  Linien diskretisiert. Die Zeitgitter werden adaptiv von `ode15s` erstellt gemäß Abschnitt 4.6.2 aus der Simulation zur Startschätzung der Steuerungen mit folgenden Toleranzen `rtol`, `atol`<sup>21</sup>:

	rtol	atol	Zeitschritte in		
			Bsp. 1	Bsp. 2	Bsp. 4
Zustände	3	6	244	272	343
Adjungierte Zustände	3	4	280	262	426
Zusammen <sup>22</sup>			522	532	767

Tabelle 4.1: Toleranzen und Anzahl Zeitschritte der adaptiven Zeitdiskretisierung.

Die erhaltenen Gitter des 1. Testbeispiels sind dargestellt in Abb. 4.8. Die Zustände werden zu kleinen Zeiten sehr fein aufgelöst aufgrund der Änderungen der schnellen Variablen nach der Anfangsbedingung und nach dem Lastwechsel. Zu den beiden Zeiten  $t = 400$  und  $t = 1200$  knickt  $u_d$  und  $U_{\text{cell,soll}}$  (siehe Abb. 4.9), weshalb auch hier feiner aufgelöst wird.

<sup>20</sup> Oder  $\vartheta_{\text{soll}}$  – je nach Zielfunktional.

<sup>21</sup> Siehe Dokumentation von `ode15s`.

<sup>22</sup> Anfangs- und Endzeit zählt nur einmal.

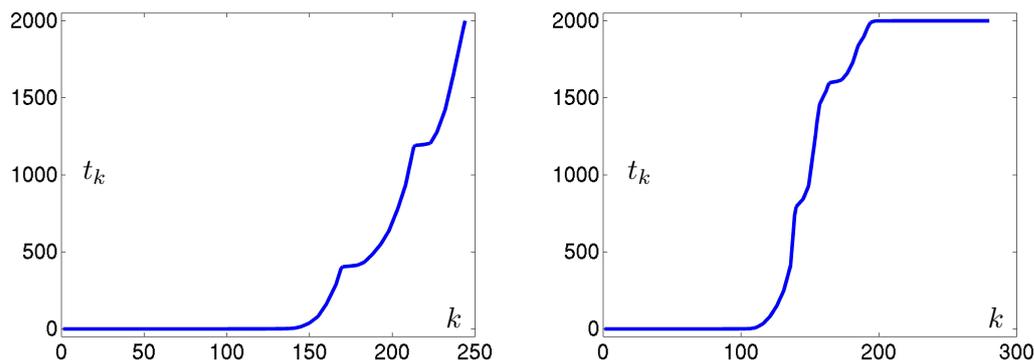


Abbildung 4.8: Adaptives Zeitgitter  $0 = t_1 < \dots < t_k < \dots < t_f$  über Index  $k$  des 1. Testbeispiels. Zustände links, adjungierte Zustände rechts.

#### 4.7.1 Lösung von Testbeispielen

Wir lösen zuerst zwei Testbeispiele und anschließend zwei konkrete Fragestellungen des Anwendungsproblems. Als Quasi-Newton-Verfahren das BFGS-Verfahren von `fmincon` ausgewählt, welches man über den Aufruf des Innere-Punkte-Verfahrens ohne Gleichungsnebenbedingungen erhält. Box-Beschränkungen an die Steuerung lassen sich hier günstig berücksichtigen.

##### 1. Testbeispiel: Steuerung von $v_{a,\text{in}}$

Im ersten Beispiel wählen wir nur eine Steuerung  $\vec{u}(t) = u_1(t) = v_{a,\text{in}}(t)$  mit Kostenfaktor  $\lambda_1 = 0.5$ . Alle nicht verwendeten Steuerungen eines Beispiels werden generell aufgefüllt mit konstanten Standardwerten aus folgender Tabelle:

$u_1$	$u_2$	$u_3$	$u_4$	$u_5$	$u_6$	$u_7$
$v_{a,\text{in}}$	$\vartheta_{a,\text{in}}$	$\chi_{\text{CH}_4,a,\text{in}}$	$\chi_{\text{H}_2,a,\text{in}}$	$\gamma_{a,\text{in}}(t) \cdot \lambda_{\text{air}}(t)$	$\vartheta_{\text{air}}$	$R_{\text{back}}$
4.4	3.0	0.28	0.05	3.3	1.5	0.3

Tabelle 4.2: Standardwerte der Steuerungen.

Anstelle von  $u_5$  wird hier eine konstante Luftzahl  $\lambda_{\text{air}} = 2.2$  verwendet. Als Steuerungsbeschränkung wählen wir  $3.2 \leq v_{a,\text{in}}(t) \leq 5.6$  und starten mit der Startschätzung  $v_{a,\text{in}}(t) \equiv 4.4$ .

Trifft man eine Auswahl von Steuerungen, so verändert sich von allen notwendigen Bedingungen lediglich der Gradient. Mit nur einer Steuerung  $v_{a,\text{in}}$  lautet er:<sup>23</sup>

$$f'(\vec{u}(t)) = \lambda_1(u_1(t) - u_1^d(t)) + \tilde{v}_a(0, t) + \frac{\lambda_{\text{air}}}{\vartheta_{a,\text{in}} \chi_{\text{O}_2,\text{air}}} \left( 2 \chi_{\text{CH}_4,a,\text{in}} + 0.5 \chi_{\text{H}_2,a,\text{in}} \right) \tilde{\gamma}_{\text{air}}(t)$$

<sup>23</sup> Letzter Summand kommt aus der Gleichung von  $\gamma_{\text{air}}$ , da  $\gamma_{a,\text{in}} = v_{a,\text{in}}/\vartheta_{a,\text{in}}$  bei Vorgabe der restlichen Inputs zu einem Auftreten von  $v_{a,\text{in}}$  führt.

Um die verschiedenen Optimierungsverfahren zu testen, generieren wir uns nach vorigem Abschnitt ein Testbeispiel. Wir entscheiden uns für

$$\bar{u}_1(t) = \begin{cases} 3.2, & t \leq 400, \\ 3.2 + 2.4 \cdot (t - 400)/800, & 400 < t < 1200, \\ 5.6, & 1200 \leq t \leq 2000 \end{cases}$$

und lösen dazu die Zustandsgleichungen. Dies generiert die optimalen Zustände, welche der Optimierer zu Ende hin liefern soll. Die optimale Zellspannung nennen wir  $U_{\text{cell,opt}}(t)$ . Mit  $U_{\text{cell,soll}}(t) := U_{\text{cell,opt}}(t) + 1$  können wir die adjungierten Gleichungen dazu lösen, mit der Wahl  $u_d(t) := \bar{u}_1(t) + \tilde{v}_a(0, t)/\lambda_1$  ist schließlich auch die Variationsungleichung erfüllt. Die erhaltene Sollzellspannung  $U_{\text{cell,soll}}(t)$  findet man in Abb. 4.9. Damit haben wir ein OC-PDAE definiert, dessen Lösung wir kennen. Untersucht wird nun, wie schnell und wie genau die einzelnen Optimierungsverfahren diese Lösung finden.

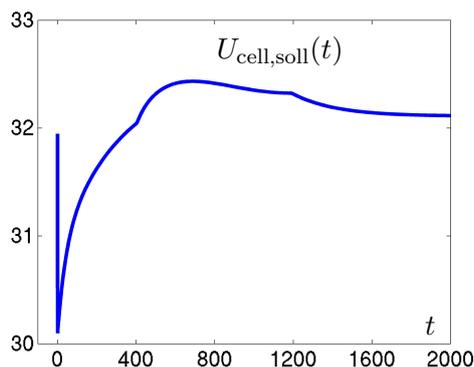


Abbildung 4.9: 1. Testbeispiel: Sollzustand  $U_{\text{cell,soll}}(t)$ .

Wir starten das indirekte Quasi-Newton-Verfahren. `fmincon` konvergiert sehr schnell, siehe Abbildung 4.10. Zu Beginn schafft es einen guten Abstieg, zu Ende hin wird es etwas langsamer.

Alternativ wechseln wir zu Ende des Verfahrens hin (nach 5 Schritten) auf eine Nachiteration mittels der Fixpunktiteration. Im Ggs. zu dem Quasi-Newton-Verfahren müssen wir uns hier auf kein festes Steuerungsgitter festlegen: Die Diskretisierung der Steuerung lässt sich günstig in jedem Schritt an die adaptiven Gitter der Zustände und der adjungierten Zustände anpassen, so dass die Genauigkeit gesteigert wird. Ein Wechsel auf die Fixpunktiteration kombiniert mit der treffenderen Diskretisierung der Steuerung erspart hier einige Funktionsauswertungen. Beim vorliegenden Beispiel konvergiert die Fixpunktiteration, auch wenn man sie von Beginn an startet.

Die schrittweise Konvergenz der Steuerung zur bekannten optimalen Steuerung sieht man in Abbildung 4.11.

Iter	F-count	f(x)	First-order optimality	Norm of step
0	1	7.88e+03	5.0e-01	
1	2	7.31e+03	2.4e-01	1.1e+01
2	3	6.98e+03	8.0e-02	8.8e+00
3	4	6.86e+03	2.4e-02	3.2e+00
4	5	6.79e+03	6.4e-03	1.9e+00
5	6	6.76e+03	2.3e-03	9.5e-01
6	7	6.75e+03	1.2e-03	3.3e-01
7	8	6.74e+03	1.3e-03	3.9e-01
8	9	6.74e+03	6.2e-04	1.4e-01
9	10	6.74e+03	8.7e-04	2.3e-02
10	12	6.74e+03	5.2e-04	3.3e-03
11	29	6.73e+03	2.2e-04	6.3e-06

Wechsel nach 5 Schritten auf die Fixpunktiteration:

5	6	6.76e+03	2.3e-03	9.5e-01
6	7	6.73e+03	FP	
7	8	6.73e+03	FP	

Abbildung 4.10: 1. Testbeispiel: Ausgabe des Quasi-Newton-Verfahrens. Der optimale Wert liegt bei  $6.73e+03$ . Iter: Iterationszähler, F-count: Anzahl Funktionsauswertungen der Unterfunktion aus Abb. 4.6, f(x): Wert des Zielfunktional. Unten: Wechsel nach 5 Schritten auf die Fixpunktiteration.

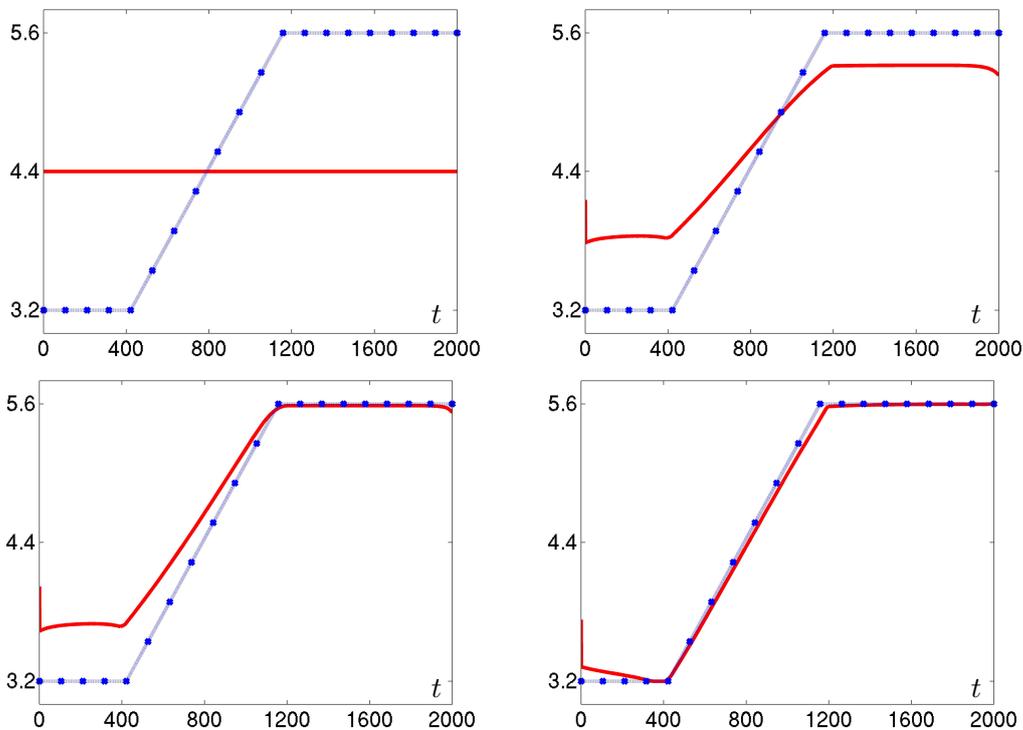


Abbildung 4.11: 1. Testbeispiel: Steuerung  $v_{a,in}(t)$  und optimale Steuerung (blau, gepunktet) in Schritten 0, 2, 5 des Quasi-Newton-Verfahrens, gefolgt vom 1. Schritt der Fixpunktiteration.

## 2. Testbeispiel: Steuerung von $\vartheta_{\text{air}}$

Im zweiten Beispiel wählen wir als Steuerung  $\vartheta_{\text{air}}(t)$  mit Steuerungsbeschränkung  $1 \leq \vartheta_{\text{air}}(t) \leq 2$ , Kostenfaktor  $\lambda_6 = 0.5$  und Startschätzung  $\vartheta_{\text{air}}(t) = 1.5$ . Alle anderen Steuerungen werden erneut konstant nach obiger Tabelle vorgegeben. Der Gradient lautet

$$f'(\vec{u}(t)) = \lambda_6(u_6(t) - u_6^d(t)) + \frac{\gamma_{\text{air}}(t)}{\gamma_b(t)} \tilde{\vartheta}_b(t).$$

Das Testbeispiel wird erzeugt mit  $\bar{u}(t) = P_{[1,2]} \{1.5 + 0.8 \cdot \sin(t \cdot \pi/500)\}$ . Weiterhin ist  $U_{\text{cell,soll}}(t) = U_{\text{cell,opt}}(t) + 10$  und  $u_6^d(t) = \bar{u}(t) + \frac{\gamma_{\text{air}}(t)}{\lambda_6 \gamma_b(t)} \tilde{\vartheta}_b(t)$ .

Nun werden die einzelnen Optimierungsverfahren an diesem Beispiel getestet. Die Laufzeiten einer Funktionsauswertung und eines Schrittes des jeweiligen Optimierungsverfahrens findet man in folgender Tabelle:

	Laufzeit je Schritt in sec	
Funktionsauswertung	10	(mit Gradient)
Indirektes Quasi-Newton-Verfahren	20	(bei nur 1 Funktionsauswertung)
Fixpunktiteration	25	
Direktes Quasi-Newton-Verfahren	6171	

Tabelle 4.3: 2. Testbeispiel: Laufzeiten der einzelnen Optimierungsverfahren bei 41 Linien und 532 Zeitschritten.

In Abb. 4.12 findet man die Ergebnisse verschiedener Verfahren anhand der Ausgabe von `fmincon`. Zunächst wird das indirekte Quasi-Newton-Verfahren aus Abschnitt 4.6 gestartet. Nach 5 Schritten werden zwei Fälle verglichen: Zuerst sieht man den Wechsel auf das direkte Verfahren (also das exakt gleiche Quasi-Newton-Verfahren, nur mit numerischer Differentiation statt Übergabe des Gradienten). Es schafft nur einen minimalen Abstieg bei extremen Laufzeiten: Aufgrund der hohen Diskretisierung der Steuerung (532 Steuerpunkte) braucht das direkte Verfahren mit numerischer Differentiation eine entsprechend hohe Anzahl an Funktionsauswertungen je Optimierungsschritt. Da jede Funktionsauswertung die Lösung des nichtlinearen PDAE-Systems der Zustandsgleichungen einschließt, erklärt sich die extreme Laufzeit des direkten Verfahrens: Es braucht 15 Stunden für die abgebildeten 8 Schritte, verglichen mit 3 Minuten, welche das indirekte Verfahren für die ersten 8 Schritte benötigt.

Als zweiter Fall wird nach den 5 Schritten des Quasi-Newton-Verfahrens auf die Fixpunktiteration gewechselt, welche schnell zum Ziel kommt. Die Anzahl der Funktionsauswertungen spiegelt die Laufzeit wieder: Je Funktionsauswertung sind beim indirekten Verfahren im Schnitt 18 sec, beim direkten 10 sec Rechenzeit nötig. Der restliche Aufwand eines Optimierungsschrittes ist vernachlässigbar, siehe Tab. 4.3.

Verwendet man das direkte Verfahren von Beginn an, so stagniert es nach einem ersten guten Abstieg, siehe Abb. 4.13. Die Rechenzeit ist enorm, da die Diskretisierung der Steuerung einfach zu fein ist. Man sollte für direkte Verfahren bei komplexen Anwendungsproblemen entweder angepasste Diskretisierungen verwenden (siehe Abschnitt 4.5) oder ausgefeiltere Methoden verwenden (siehe Abschnitt 4.7.3).

<p>1.) Indir. Quasi-Newton-Verfahren:</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Iter</th> <th style="text-align: left;">F-count</th> <th style="text-align: left;">f(x)</th> </tr> </thead> <tbody> <tr><td>0</td><td>1</td><td>1.5377e+04</td></tr> <tr><td>1</td><td>2</td><td>1.4659e+04</td></tr> <tr><td>2</td><td>3</td><td>1.4575e+04</td></tr> <tr><td>3</td><td>4</td><td>1.4413e+04</td></tr> <tr><td>4</td><td>5</td><td>1.4301e+04</td></tr> <tr style="border-bottom: 1px solid black;"><td>5</td><td>6</td><td>1.4284e+04</td></tr> <tr><td>6</td><td>7</td><td>1.4279e+04</td></tr> <tr><td>7</td><td>8</td><td>1.4279e+04</td></tr> <tr><td>8</td><td>9</td><td>1.4279e+04</td></tr> <tr><td>9</td><td>10</td><td>1.4277e+04</td></tr> <tr><td>10</td><td>12</td><td>1.4277e+04</td></tr> <tr><td>11</td><td>13</td><td>1.4276e+04</td></tr> <tr><td>12</td><td>16</td><td>1.4276e+04</td></tr> <tr><td>13</td><td>21</td><td>1.4276e+04</td></tr> <tr><td>14</td><td>26</td><td>1.4276e+04</td></tr> <tr><td>15</td><td>27</td><td>1.4276e+04</td></tr> </tbody> </table>	Iter	F-count	f(x)	0	1	1.5377e+04	1	2	1.4659e+04	2	3	1.4575e+04	3	4	1.4413e+04	4	5	1.4301e+04	5	6	1.4284e+04	6	7	1.4279e+04	7	8	1.4279e+04	8	9	1.4279e+04	9	10	1.4277e+04	10	12	1.4277e+04	11	13	1.4276e+04	12	16	1.4276e+04	13	21	1.4276e+04	14	26	1.4276e+04	15	27	1.4276e+04	<p>2.) Wechsel nach 5 Schritten auf a) direktes Verfahren:</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Iter</th> <th style="text-align: left;">F-count</th> <th style="text-align: left;">f(x)</th> </tr> </thead> <tbody> <tr><td>0</td><td>533</td><td>1.4284e+04</td></tr> <tr><td>1</td><td>1072</td><td>1.4284e+04</td></tr> <tr><td>2</td><td>1612</td><td>1.4284e+04</td></tr> <tr><td>3</td><td>2156</td><td>1.4284e+04</td></tr> <tr><td>4</td><td>2697</td><td>1.4284e+04</td></tr> <tr><td>5</td><td>3251</td><td>1.4284e+04</td></tr> <tr><td>6</td><td>3784</td><td>1.4284e+04</td></tr> <tr><td>7</td><td>4317</td><td>1.4284e+04</td></tr> <tr><td>8</td><td>4855</td><td>1.4283e+04</td></tr> </tbody> </table> <p>b) Fixpunktiteration:</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Iter</th> <th style="text-align: left;">F-count</th> <th style="text-align: left;">f(x)</th> </tr> </thead> <tbody> <tr><td>0</td><td>0</td><td>1.4284e+04</td></tr> <tr><td>1</td><td>1</td><td>1.4278e+04</td></tr> <tr><td>2</td><td>2</td><td>1.4276e+04</td></tr> </tbody> </table>	Iter	F-count	f(x)	0	533	1.4284e+04	1	1072	1.4284e+04	2	1612	1.4284e+04	3	2156	1.4284e+04	4	2697	1.4284e+04	5	3251	1.4284e+04	6	3784	1.4284e+04	7	4317	1.4284e+04	8	4855	1.4283e+04	Iter	F-count	f(x)	0	0	1.4284e+04	1	1	1.4278e+04	2	2	1.4276e+04
Iter	F-count	f(x)																																																																																												
0	1	1.5377e+04																																																																																												
1	2	1.4659e+04																																																																																												
2	3	1.4575e+04																																																																																												
3	4	1.4413e+04																																																																																												
4	5	1.4301e+04																																																																																												
5	6	1.4284e+04																																																																																												
6	7	1.4279e+04																																																																																												
7	8	1.4279e+04																																																																																												
8	9	1.4279e+04																																																																																												
9	10	1.4277e+04																																																																																												
10	12	1.4277e+04																																																																																												
11	13	1.4276e+04																																																																																												
12	16	1.4276e+04																																																																																												
13	21	1.4276e+04																																																																																												
14	26	1.4276e+04																																																																																												
15	27	1.4276e+04																																																																																												
Iter	F-count	f(x)																																																																																												
0	533	1.4284e+04																																																																																												
1	1072	1.4284e+04																																																																																												
2	1612	1.4284e+04																																																																																												
3	2156	1.4284e+04																																																																																												
4	2697	1.4284e+04																																																																																												
5	3251	1.4284e+04																																																																																												
6	3784	1.4284e+04																																																																																												
7	4317	1.4284e+04																																																																																												
8	4855	1.4283e+04																																																																																												
Iter	F-count	f(x)																																																																																												
0	0	1.4284e+04																																																																																												
1	1	1.4278e+04																																																																																												
2	2	1.4276e+04																																																																																												

Abbildung 4.12: 2. Testbeispiel: Durchlauf des indirekten Quasi-Newton-Verfahrens (links). Nach 5 Schritten wird alternativ auf zwei andere Verfahren gewechselt (rechts). **Iter**: Iterationszähler, **F-count**: Anzahl Funktionsauswertungen der Unterfunktion aus Abb. 4.6, **f(x)**: Wert des Zielfunktional. Der optimale Wert liegt bei  $1.4276e+04$ .

Iter	F-count	f(x)
0	533	1.5377e+04
2	1599	1.4609e+04
4	2682	1.4608e+04
6	3771	1.4607e+04

Abbildung 4.13: 2. Testbeispiel: Durchlauf des direkten Verfahrens (jeder zweite Schritt). Der optimale Wert liegt bei  $1.4276e+04$ .

Das Quasi-Newton-Verfahren erreicht sowohl mit, als auch ohne den Wechsel auf die Fixpunktiteration das theoretische Optimum. Beide Varianten liefern die gleiche Kurve  $U_{\text{cell,opt}}(t)$  in Abb. 4.14: Ausgehend von einer weit entfernten Startschätzung erreichen die Optimierer die gewünschte Zellspannung  $U_{\text{cell,soll}}(t)$  sehr gut. Eine weitere Annäherung steigert das Zielfunktional aufgrund zu stark anwachsender Steuerungskosten.

Zwei der optimalen Zustände findet man in Abb. 4.15. Zu Beginn erkennt man bei  $\chi_{a,\text{CH}_4}$  den starken Einfluss des Lastwechsels. Danach kommt es wie erwartet zu periodischen Schwankungen.  $\vartheta_a$  ist robuster gegenüber dem Lastwechsel. In  $z = 0$  zeigt es ausnahmsweise starke Gradienten in Ortsrichtung, was genauer aufgelöst ist in [90, Fig. 6].

In Grafik 4.16 sind zwei ausgewählte adjungierte Zustände abgebildet. Man erkennt um die Anfangsbedingung  $t_f = 2000$  große Gradienten in Zeitrichtung. Hier wählt der adaptive Solver daher viele, kurze Zeitschritte. Damit zeigen adjungierte Zustände ein vergleichbares Übergangsverhalten wie die Zustände, deren Verhalten ausführlich in [90] beschrieben ist: Bei nicht-stationären Anfangsbedingungen – sowie beim Lastwechsel –

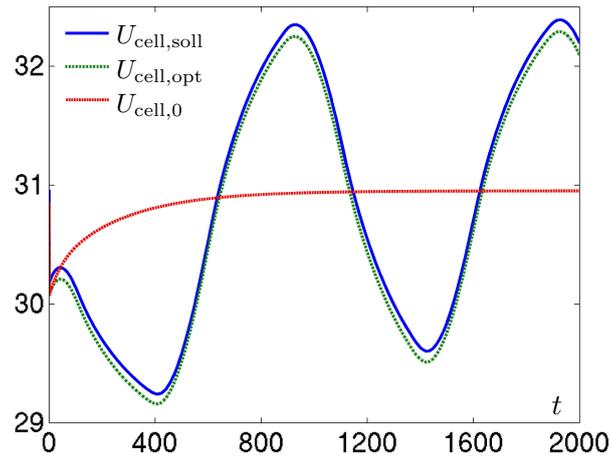


Abbildung 4.14: 2. Testbeispiel: Sollzustand  $U_{\text{cell,soll}}(t)$ , optimaler Zustand  $U_{\text{cell,opt}}(t)$  und Startschätzung  $U_{\text{cell,0}}(t)$ .

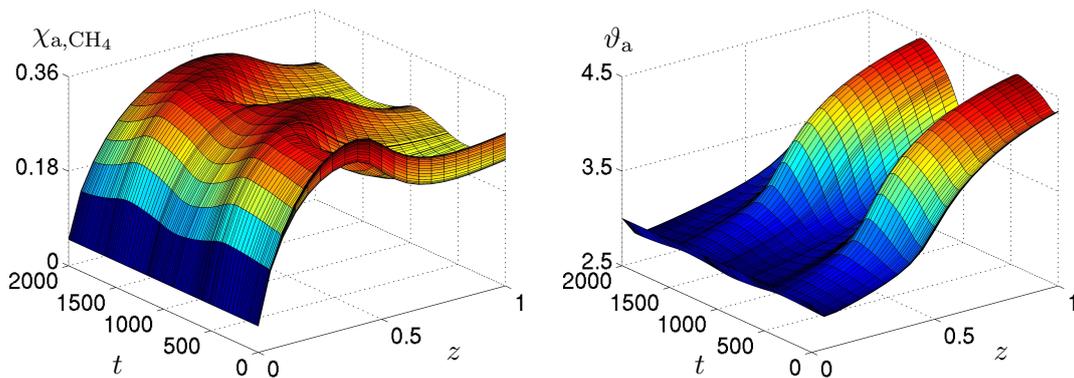


Abbildung 4.15: 2. Testbeispiel: Optimale Zustände  $\chi_{\text{a,CH}_4}(z, t)$  und  $\vartheta_{\text{a}}(z, t)$ . (Nur jeder dritte Zeitschritt ist zur Zeichnung verwendet worden.)

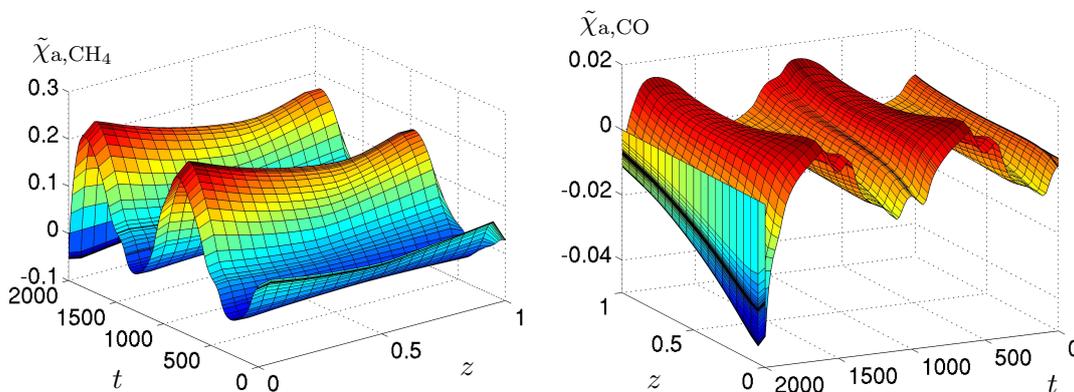


Abbildung 4.16: 2. Testbeispiel: Adjungierte Zustände  $\tilde{\chi}_{\text{a,CH}_4}$  und  $\tilde{\chi}_{\text{a,CO}}$  nach Schritt 5 des Quasi-Newton-Verfahrens. (Nur jeder dritte Zeitschritt ist zur Zeichnung verwendet worden.)

führen die unterschiedlichen Zeitskalen des Systems zu großen zeitlichen Änderungen oder gar zu Überschwingern. Der prinzipielle Unterschied liegt jedoch in der Tatsache, dass die adjungierten Gleichungen rückwärts in der Zeit gerichtet sind. Daher sind optimal gewählte Gitter von Zuständen und adjungierten Zuständen fundamental verschieden. Aus diesem Grund hatten wir hier zwei individuell angepasste Gitter vereinigt, um alle Variablen des KKT-Systems gut aufzulösen.

Die Konvergenz der Steuerung zur optimalen Steuerung findet man in Abb. 4.17.

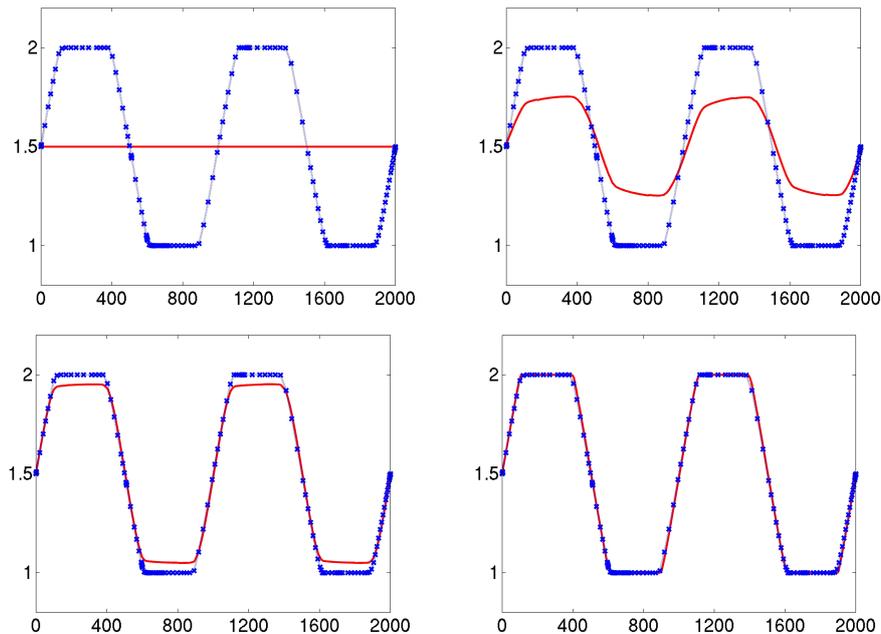


Abbildung 4.17: 2. Testbeispiel: Steuerung  $v_{\text{air}}(t)$  und optimale Steuerung (blau, gepunktet) in Schritten 0, 2, 5 des indirekten Quasi-Newton-Verfahrens, gefolgt vom 1. Schritt der Fixpunktiteration.

### 4.7.2 Lösung des Anwendungsproblems

Nachdem wir die Optimierungsverfahren und deren Konvergenzeigenschaften an zwei Testbeispielen untersucht haben, können wir uns an das Anwendungsproblem wagen.

### 3. Beispiel: Tracking unter verschiedenen Steuerungskosten

Wir greifen voriges Testbeispiel auf und ändern lediglich  $u_d = 0$  sowie die Steuerungskosten. Zunächst wählen wir einen sehr geringen Kostenfaktor  $\lambda_6 = 5 \cdot 10^{-5}$  und untersuchen als zweite Variante  $\lambda_6 = 5 \cdot 10^{-2}$ . Das indirekte Quasi-Newton-Verfahren konvergiert jeweils sehr schnell, siehe exemplarisch den ersten Fall in Abb. 4.18. Die Sollzellspannung wird dank der geringen Steuerungskosten sehr gut erreicht, siehe Abb. 4.19. Obwohl dieses Diagramm sehr an das entsprechende Diagramm des 2. Testbeispiels erinnert, sollte man sich vor Augen halten, dass es diesmal nur aufgrund der geringen Steuerungskosten entsteht. Im 2. Testbeispiel konnte ebenso aufwändig gesteuert werden, dort allerdings aufgrund der günstig gelegenen Sollsteuerung  $u_d$ .

Iter	F-count	f(x)	First-order optimality	Norm of step
0	1	1.023003e+03	1.5e-01	
1	2	6.792458e+02	7.5e-02	1.5e+00
2	3	1.338197e+02	9.4e-02	4.4e+00
3	4	6.127869e+01	6.4e-02	1.4e+00
4	5	1.888273e+01	4.5e-02	1.3e+00
5	6	6.433656e+00	4.8e-02	9.7e-01
6	8	6.133361e+00	3.3e-02	3.9e-01
7	20	6.133453e+00	7.5e-03	2.7e-02
8	30	6.131543e+00	7.4e-03	8.5e-04
9	54	6.131526e+00	7.5e-03	1.1e-08
10	57	6.131521e+00	7.5e-03	1.3e-08

Abbildung 4.18: 3. Beispiel: Indirektes Quasi-Newton-Verfahren zu  $\lambda_6 = 5 \cdot 10^{-5}$ .

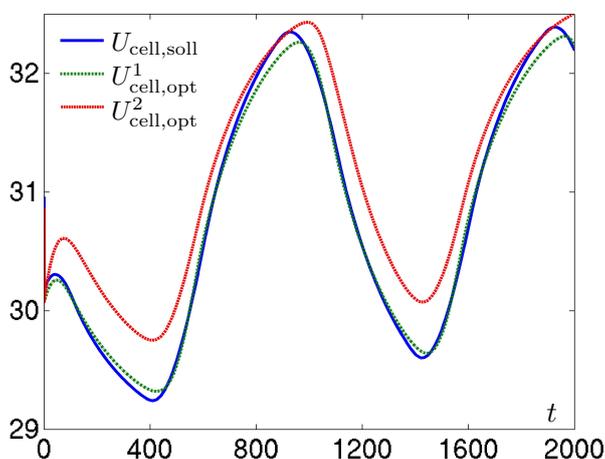


Abbildung 4.19: 3. Beispiel: Sollzustand  $U_{\text{cell,soll}}(t)$  und optimaler Zustand  $U_{\text{cell,opt}}^1(t)$  bei  $\lambda_6 = 5 \cdot 10^{-5}$  bzw.  $U_{\text{cell,opt}}^2(t)$  bei  $\lambda_6 = 5 \cdot 10^{-2}$  nach jeweils 6 Schritten des indirekten Quasi-Newton-Verfahrens.

Erhöht man den Kostenfaktor auf  $\lambda_6 = 5 \cdot 10^{-2}$ , so konvergiert das indirekte Quasi-Newton-Verfahren ähnlich schnell, bei einem entsprechend anderen Ergebnis: Aufgrund der höheren Steuerungskosten lässt sich die Sollzellspannung in Abb. 4.19 nicht so genau erreichen.

#### 4. Beispiel: Der schnelle Lastwechsel

Als letztes Experiment soll der schnelle Lastwechsel gelöst werden. Dazu wird der Sollzustand definiert nach [90] zu

$$U_{\text{cell,soll}} := \begin{cases} U_{\text{cell},0.4}, & t < 0.1, \\ U_{\text{cell},0.5}, & t \geq 0.1. \end{cases} \quad (4.38)$$

Dabei ist die rechte Seite jeweils die stationäre Zellspannung, welche sich nach langer Zeit zu  $I_{\text{cell}} = 0.4$  bzw.  $I_{\text{cell}} = 0.5$  und Standardwerten der Steuerung aus Tabelle 4.2

einstellt. Erneut werden obige zwei Steuerungskosten im Vergleich untersucht. Die restlichen Parameter übernehmen wir aus dem 2. Testbeispiel.

Das Quasi-Newton-Verfahren konvergiert hier langsamer. Nach 50 Schritten liefert es die Zellspannungen in Abb. 4.20. Wir sehen, dass bei minimalen Steuerungskosten der Lastwechsel trotz Steuerungsbeschränkungen rasant und exakt abgeschlossen werden kann. Dabei wird der Brennstoffzellenstapel allein anhand der Temperatur des Luftstroms gesteuert, welcher in die Brennkammer eingespeist wird.

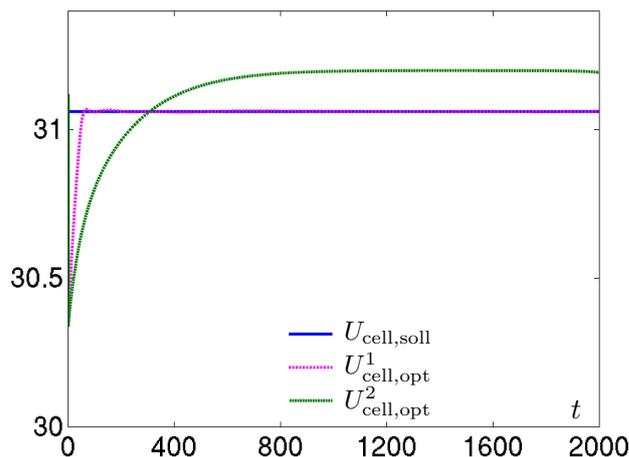


Abbildung 4.20: 4. Beispiel: Sollzustand  $U_{\text{cell,soll}}(t)$  und optimaler Zustand  $U_{\text{cell,opt}}^1(t)$  bei  $\lambda_6 = 5 \cdot 10^{-5}$  bzw.  $U_{\text{cell,opt}}^2(t)$  bei  $\lambda_6 = 5 \cdot 10^{-2}$  nach 50 Schritten des indirekten Quasi-Newton-Verfahrens.

Bei derart niedrigen Steuerungskosten wie  $\lambda_6 = 5 \cdot 10^{-5}$  divergiert die Fixpunktiteration. Eine kleine Norm des Gradienten führt aufgrund der Division durch  $\lambda_6$  offensichtlich bereits zu einer großen Differenz der nächsten Steuerung zur optimalen Steuerung.

Iter	F-count	f(x)	First-order optimality	Norm of step
0	1	3.137e+01	5.2e-02	
5	6	7.001e+00	1.2e-02	3.6e-01
10	13	6.343e+00	1.1e-02	1.2e-01
15	20	5.861e+00	5.8e-03	1.0e-01
20	27	5.687e+00	2.3e-03	2.9e-02
25	34	5.653e+00	1.2e-03	1.8e-02
30	42	5.488e+00	2.6e-03	1.4e-01
35	47	5.428e+00	1.2e-03	7.3e-03
40	56	5.383e+00	1.5e-03	2.4e-02
45	77	5.374e+00	3.9e-04	4.1e-03
50	129	5.372e+00	3.2e-04	1.8e-10

Abbildung 4.21: 4. Beispiel: Ausgabe des indirekten Quasi-Newton-Verfahrens zu  $\lambda_6 = 5 \cdot 10^{-5}$ .

### 4.7.3 Vergleich und Weiterentwicklung durch AD

Im folgenden werden die verschiedenen Verfahren der Optimalen Steuerung verglichen und Weiterentwicklungen werden vorgeschlagen.

#### Vergleich von direkten und indirekten Verfahren

Die Kombination eines gestaffelten indirekten Verfahrens mit der auf die Steuerung reduzierten Formulierung und einer schrittweisen Integration in der Zeit zeigt sich als sehr effizient. Innerhalb von wenigen Minuten lässt sich das komplexe Anwendungsproblem mit hoher Genauigkeit optimal steuern. Die indirekten Methoden konvergieren schnell und zuverlässig. Sie erlauben eine wesentlich feinere Auflösung der Steuerungen, da die Anzahl der Funktionsauswertungen je Optimierungsschritt im Vergleich zu den direkten Verfahren nicht von der Anzahl der Steuerpunkte abhängt. Der Gradient ist beim indirekten Verfahren glatter, so dass weniger numerische Artefakte zu beobachten sind. Allerdings stellt die Herleitung der notwendigen Bedingungen bei komplexen gekoppelten Systemen den Anwender vor eine große Herausforderung.

Direkte Verfahren mit numerischer Differentiation können in Genauigkeit und Rechenaufwand nicht mit indirekten Verfahren konkurrieren. Der Aufwand steigt stark mit der Anzahl der Freiheitsgrade. Sehr feine Steuergitter, verteilte Steuerungen oder die volle Formulierung (das Verwenden von Zuständen und Steuerungen als Optimierungsvariablen) erzeugen viele Freiheitsgrade und bereiten entsprechend Probleme. Man sollte die direkten Verfahren für solche Probleme kombinieren mit Methoden der instantanen Steuerung, der modellprädiktiven Regelung, der Modellreduktion oder der automatischen Differentiation.

Vorteil der direkten Verfahren ist ihre Anwenderfreundlichkeit: Der Anwender hat die Zustandsgleichungen und das Zielfunktional passend zu diskretisieren. Die aufwändige Herleitung der Optimalitätsbedingungen und deren Lösung entfällt dagegen.

#### Weiterentwicklung durch AD-Tools

Ziel ist es, die Vorteile beider Verfahren zu kombinieren. Ein Schlüssel hierfür ist die Automatische Differentiation<sup>24</sup> (AD), welche sich innerhalb direkter und indirekter Verfahren nutzen lässt. So steigert AD bei direkten Verfahren die Genauigkeit und reduziert den Rechenaufwand: Analytische Ableitungen sind bekanntlich nicht von der Auslöschung betroffen und erreichen daher eine hohe Genauigkeit. Zudem steigt der Rechenaufwand nicht in der Form, da AD-Tools mittlerweile das Ziel des linearen gestiegenen Aufwandes erreichen, d.h. der Aufwand den Gradienten zu berechnen soll nur das  $c$ -fache des Aufwandes der Funktionsauswertung sein – mit einem kleinen, vom Problem unabhängigen  $c$ .

Besonders effizient lässt sich AD bei der Verwendung fester Gitter einbauen. Das Rocketcar in Kap. 3.5.2 zeigt, dass man mit direkten Verfahren und AD dann beeindruckende Ergebnisse bei geringem Implementationsaufwand erzielen kann. Dank der Verwendung eines festen Gitters und einer vollen Diskretisierung lassen sich sogar Zustandsbeschränkungen berücksichtigen.

---

<sup>24</sup> Literatur und Tools findet man unter [www.autodiff.org](http://www.autodiff.org).

Das Integrieren von AD-Tools in direkte Methoden, welche auf der reduzierten Formulierung basieren, benötigt dagegen eine genauere Betrachtung durch den Anwender: AD-Tools lassen sich hier meist nicht auf das komplette Programmpaket anwenden, sondern effizient nur auf ausgewählte Subsysteme. Konstrukte wie Gittergeneratoren oder Adaptivitätsroutinen stellen z.B. besondere Herausforderungen dar [17]. So lässt sich insbesondere das reduzierte Zielfunktional der MCFC aufgrund der adaptiven Zeitdiskretisierung als Ganzes nicht mehr automatisch ableiten. Andere Konstrukte wie indirekte Löser linearer Gleichungssysteme sollte man nicht differenzieren, sondern sich sogenannter hierarchischer Ansätze bedienen (siehe z.B. [82, S. 1053] und dort genannte Referenzen). Dabei stellt man ein abgeleitetes Gleichungssystem auf und löst dieses bei Aufruf des Gradienten mit dem vorhandenen indirekten Löser. Einige AD-Tools setzen dies automatisch um.

Die passende Anwendung von AD auf einen großen Code im Bereich der Shapeoptimierung in der Aerodynamik findet man bei Gauger, Walther et al. [40].

Für die reduzierte Formulierung bei adaptiven Gittern sollte man auch die Kombination von indirekten Methoden mit AD in Erwägung ziehen. Für das MCFC-Problem bedeutet dies, dass man die Differentialoperatoren der adjungierten Gleichungen herleitet, den Hauptaufwand der rechten Seiten aber per AD berechnen lässt. AD trägt hier zu Effizienz und geringerer Fehleranfälligkeit. Zu Problemen mit der Differentiation der Adaptivitätsroutine kann es hier nicht kommen, da die adaptive Diskretisierung erst nach Anwenden des AD-Tools vorgenommen wird.

## 4.8 Zusammenfassung

Die Modellierung von Brennstoffzellen mit Differentialgleichungen liefert meist komplexe PDAE-Modelle. Der optimale Betrieb von Brennstoffzellenanlagen stellt daher ein Musterbeispiel eines komplexen OC-PDAE dar.

Ein konkretes PDAE-Modell einer Hochtemperatur-Brennstoffzelle vom Typ MCFC vereint DAE, nichtlineare hyperbolische Advektions-Reaktionsgleichungen, parabolische Diffusions-Reaktionsgleichungen und integrale Kopplungen aufgrund elektrischer Felder. Zusätzlich besitzt das Problem multiple Zeitskalen. Aufgrund der detaillierten Modellierung der chemischen Reaktionen weist das PDAE-System eine starke Kopplung über lange, nichtlineare Quellterme auf. Aus dem technischen Hintergrund heraus wird ein OC-PDAE mit 7 rein zeitabhängigen Steuerungen und Steuerungsbeschränkungen formuliert.

Aufgrund der hohen Komplexität des PDAE-Modells lassen sich die notwendigen Bedingungen nur mit der formalen Lagrange-Technik aufstellen. Die vielseitigen und nichtlinearen Kopplungen erzeugen jedoch eine Vielzahl von Summanden in den adjungierten Gleichungen. Zusammen mit der hohen Zahl von 67 verschiedenen dualen Variablen stellt aber auch die formale Lagrange-Technik bereits eine sehr aufwändige Angelegenheit dar. Bei derart komplexen Problemen sollte man daher als Alternative den Einsatz von Automatischer Differentiation in Erwägung ziehen.

Die notwendigen Bedingungen bestehen aus dem PDAE-Modell der Brennstoffzelle, einem ebenso großen PDAE-Modell von adjungierten Gleichungen, sowie 7 Variationsungleichungen für die Steuerungen, welche verzwickelt sind, da die Steuerungen nichtlinear in die Zustandsgleichungen eingehen.

Zur Lösung wird ein großes Software-Paket unter MATLAB entwickelt. Die Zustandsgleichungen werden mit der Linienmethode unter speziell auf die hyperbolischen Erhaltungsgleichungen angepasster finiter Differenzen diskretisiert. Anschließend werden etablierte adaptive und implizite DAE-Methoden eingesetzt, um die verschiedenen Zeitskalen des Modells treffend aufzulösen. Die Zustandsgleichungen können damit sehr schnell und sehr genau gelöst werden.

Da die notwendigen Bedingungen aufgrund der Komplexität nicht mit einer vollen Diskretisierung gelöst werden können, werden gestaffelte Verfahren entwickelt, bei welchen die beiden PDAE-Systeme zwar häufig, aber stets unabhängig voneinander gelöst werden müssen. Die adjungierten Gleichungen werden dabei analog zu den Zustandsgleichungen mit der Linienmethode und einer adaptiven, impliziten DAE-Methode gelöst. Als Optimierungsverfahren setzen sich etablierte Quasi-Newton-Verfahren mit BFGS-Update durch.

Die Optimierungsverfahren werden verglichen anhand numerischer Testbeispiele. Mehrere Studien belegen die Effizienz der entwickelten Methoden. Die indirekten Quasi-Newton-Verfahren zeigen dabei einen großen Konvergenzbereich. Dank gestaffelter Verfahren und schrittweiser Integration ist jede Funktions- und Gradientenauswertung sehr schnell berechnet. Da das etablierte Quasi-Newton-Verfahren zudem nur wenige Funktionsauswertungen und nur wenige Schritte zur Konvergenz benötigt, liegen die Ergebnisse auch bei feinen Diskretisierungen in wenigen Minuten vor. Darauf aufbauend lässt sich zuletzt das Anwendungsproblem in geringer Laufzeit und hoher Genauigkeit lösen.

# Kapitel 5

## Schluss

Diese Arbeit liefert Beiträge zur Optimalen Steuerung partiell-differential algebraischer Gleichungen. Insbesondere werden Zustandsbeschränkungen bei der Optimalen Steuerung gewöhnlicher und partieller Differentialgleichungen sowie gekoppelter Systeme untersucht. Die verschiedenen Konzepte dieser Gebiete werden verglichen, übertragen und eingeordnet.

Zentrale Ergebnisse sind die Übertragung der notwendigen Bedingungen nach Bryson, Denham und Dreyfus auf elliptische Optimalsteuerungsprobleme mit punktwisen Zustandsbeschränkungen, die Übertragung von Sprungbedingungen und Maßdarstellungen auf ein ODE-PDE beschränktes Optimalsteuerungsproblem mit Zustandsbeschränkungen bei niederdimensionalen aktiven Mengen, sowie die Entwicklung effizienter numerischer Methoden für komplexe Anwendungsprobleme.

Die Beiträge dieser Arbeit gliedern sich in vier Kapitel, deren Aspekte jeweils zusammengefasst werden:

Zunächst werden die Grundlagen aus OC-ODE wiederholt. Die notwendigen Bedingungen von Optimalsteuerungsproblemen mit gewöhnlichen Differentialgleichungen unter (punktwisen) Zustandsbeschränkungen werden im Falle verteilter aktiver Mengen zu differential algebraischen Gleichungen (DAE), da die Zustandsbeschränkung auf der aktiven Menge eine Gleichung darstellt. Zur numerischen Lösung transformiert man die notwendigen Bedingungen in ein Mehrpunkt-Randwertproblem und löst dieses mit der Mehrzielmethode oder dem Kollokationsverfahren. Es gibt zwei etablierte Systeme notwendiger Bedingungen, die direkte Ankopplung nach Jacobson, Lele und Speyer (JLS-Ansatz) sowie die indirekte Ankopplung nach Bryson, Denham und Dreyfus (BDD-Ansatz). Numerische Methoden bauen meist auf dem BDD-Ansatz auf, da das Mehrpunkt-Randwertproblem hier eine DAE von deutlich geringerem Index (i.d.R. Differentiationsindex 1) ist. Bei Standard-Problemen lassen sich beim BDD-Ansatz sogar alle algebraischen Gleichungen beseitigen, so dass ein Mehrpunkt-Randwertproblem bei gewöhnlichen Differentialgleichungen entsteht. Das Mehrpunkt-Randwertproblem nach dem JLS-Ansatz weist dagegen einen derart hohen Index auf, dass es numerisch nicht sinnvoll gelöst werden kann.

Über die formale Lagrange-Technik und eine Analyse des Lagrange-Multiplikators der Zustandsbeschränkung lassen sich unter einer Strukturannahme an die aktive Menge die Zusammenhänge zwischen dem Borel-Maß und den Sprungbedingungen herausarbeiten.

Optimalsteuerungsprobleme partieller Differentialgleichungen unter Zustandsbeschränkungen und verteilten aktiven Mengen führen analog zu Systemen partiell-differential

algebraischer Gleichungen. Im Hinblick auf ihre Vorteile bei gewöhnlichen Differentialgleichungen werden BDD-Ansätze erstmalig auf elliptische partielle Differentialgleichungen übertragen. Die Zustandsbeschränkung wird hierzu auf der aktiven Menge äquivalent in ein Randwertproblem transformiert. Bei der konsequenten Aufteilung der gesamten Problemstellung auf die aktive und die inaktive Menge erhält man zusätzliche Nebenbedingungen an die Zustände. Unter Sicherstellung der Zowe-Kurcyusz-Bedingung lassen sich die neuen notwendigen Bedingungen mit einer aufgeteilten Lagrange-Technik herleiten. Die beteiligten Lagrange-Multiplikatoren weisen höhere Regularitäten auf verglichen mit den klassischen notwendigen Bedingungen.

Als Verallgemeinerung des Mehrpunkt-Randwertproblems erhält man aus den neuen notwendigen Bedingungen ein Randwertproblem auf verschiedenen Gebieten mit Übergangsbedingungen. Das Interface zwischen den verschiedenen Gebieten stellt eine Optimierungsvariable dar, welche den Schaltzeiten des Mehrpunkt-Randwertproblems im Falle gewöhnlicher Differentialgleichungen entspricht. Eine notwendige Bedingung am Interface wird daher mit Techniken der Shapeoptimierung hergeleitet.

Zeitabhängige Anwendungsprobleme stellen uns häufig vor Optimalsteuerungsprobleme mit gekoppelten Systemen aus gewöhnlichen und partiellen Differentialgleichungen unter Zustandsbeschränkungen. Da sich im Ort verteilte Steuerungen selten in die Realität umsetzen lassen, enthalten viele Modelle nur endlich viele zeitveränderliche Steuersignale, welche optimal zu gestalten sind. Als Beispiel denke man an die Steuerung von Positionen, Geschwindigkeiten, Beschleunigungen, Amplituden, Stromstärken, Spannungen oder von Winkeln. Hier sind i.d.R. Methoden für niederdimensionale aktive Mengen zu entwickeln. Exemplarisch untersuchen wir ein derartiges Prototyp-Problem eines ODE-PDE beschränkten Optimalsteuerungsproblems mit punktweisen Zustandsbeschränkungen und freier Endzeit, das sogenannte Rocketcar-Problem.

In einer Reihe theoretischer Untersuchungen werden Sprungbedingungen aus der Optimalen Steuerung gewöhnlicher Differentialgleichungen auf das ODE-PDE beschränkte Optimalsteuerungsproblem übertragen. Der Lagrange-Multiplikator der Zustandsbeschränkung kann erstmals für niederdimensionale aktive Mengen vollständig angegeben werden. Dies ermöglicht im Prototyp-Beispiel weitgehende Regularitätsaussagen und eine Formulierung der notwendigen Bedingungen als Mehrpunkt-Anfangsrandwertproblem. Direkte und indirekte numerische Methoden werden entwickelt und diskutiert. Ergebnisse aus direkten Verfahren belegen die theoretischen Resultate.

Komplexe Anwendungsprobleme stellen für die theoretische und numerische Analyse große Herausforderungen dar. Bei der Untersuchung eines PDAE-Modells einer Brennstoffzellen-Anlage trifft man auf ein System aus ODE, parabolischen und hyperbolischen PDE, DAE sowie integralen und nicht-lokalen Kopplungstermen und multiplen Zeitskalen. Das zugehörige Optimalsteuerungsproblem sucht nach 7 rein zeitabhängigen Steuerungen unter technisch bedingten Box-Beschränkungen.

Zur Lösung dieses Problems werden direkte und indirekte Methoden entwickelt. Das Aufstellen der notwendigen Bedingungen erweist sich dank der Vielzahl von 67 verschiedenen Lagrange-Multiplikatoren und der verzweigten Kopplungsstruktur als äußerst diffizil. Die erhaltenen notwendigen Bedingungen lassen sich jedoch zu sehr effizienten Optimierungsmethoden nutzen. Dank gestaffelter Verfahren und der schrittweisen Integration der Teilsysteme erreicht man sehr geringe Laufzeiten je Optimierungsschritt. Die schrittweise Integration erlaubt zudem eine sehr feine, adaptive Zeitdiskretisierung, wodurch die verschiedenen Zeitskalen auf einem langen Zeithorizont sehr genau aufgelöst werden können. Die darauf aufgesetzten adjungierten-basierten Quasi-Newton-Verfahrens erlauben daher in numerischen Studien eine schnelle und genaue Steuerung des Brennstoffzellen-Modells.

# Anhang A

## Zustandsgleichungen der MCFC

### A.1 Gaskanäle und Solid

$$\begin{aligned}
\frac{\partial \chi_{k,a}}{\partial t} + \gamma_a \vartheta_a \frac{\partial \chi_{k,a}}{\partial z} &= f^{\chi_{k,a}} := \vartheta_a \sum_{j=\text{ox}1,2,\text{ref}1,2} (\nu_{k,j} - \chi_{k,a} \bar{\nu}_j) Da_j r_j, \quad k \in \mathcal{I}, \\
\frac{\partial \chi_{k,c}}{\partial t} - \gamma_c \vartheta_c \frac{\partial \chi_{k,c}}{\partial z} &= f^{\chi_{k,c}} := \vartheta_c (\nu_{k,\text{red}} - \chi_{k,c} \bar{\nu}_{\text{red}}) Da_{\text{red}} r_{\text{red}}, \quad k \in \mathcal{I}, \\
\frac{\partial \vartheta_a}{\partial t} + \gamma_a \vartheta_a \frac{\partial \vartheta_a}{\partial z} &= f^{\vartheta_a} := \frac{\vartheta_a}{c_p} \left( \sum_{j=\text{ref}1,2} -\Delta_R h_j Da_j r_j + \sum_{j=\text{ox}1,2} \bar{\nu}_j^+ Da_j r_j (\vartheta_s - \vartheta_a) c_p + q_{\text{as}} \right), \\
\frac{\partial \vartheta_c}{\partial t} - \gamma_c \vartheta_c \frac{\partial \vartheta_c}{\partial z} &= f^{\vartheta_c} := \underbrace{\vartheta_c \bar{\nu}_{\text{red}}^- Da_{\text{red}} r_{\text{red}} (\vartheta_s - \vartheta_c)}_{=0} + \frac{\vartheta_c q_{\text{cs}}}{c_p}, \\
\frac{\partial(\gamma_a \vartheta_a)}{\partial z} &= f^{\gamma_a \vartheta_a} := \sum_{j=\text{ref}1,2} -\Delta_R h_j Da_j r_j / c_p + \sum_{j=\text{ox}1,2} \bar{\nu}_j^+ Da_j r_j (\vartheta_s - \vartheta_a), \\
&\quad + \frac{q_{\text{as}}}{c_p} + \vartheta_a \sum_{j=\text{ox}1,2,\text{ref}1,2} \bar{\nu}_j Da_j r_j, \\
-\frac{\partial(\gamma_c \vartheta_c)}{\partial z} &= f^{\gamma_c \vartheta_c} := \vartheta_c \bar{\nu}_{\text{red}}^- Da_{\text{red}} r_{\text{red}} + \underbrace{\bar{\nu}_{\text{red}}^- Da_{\text{red}} r_{\text{red}} (\vartheta_s - \vartheta_c)}_{=0} + \frac{q_{\text{cs}}}{c_p}, \\
c_{p,s} \frac{\partial \vartheta_s}{\partial t} &= f^{\vartheta_s} := \frac{1}{\text{Pe}} \frac{\partial^2 \vartheta_s}{\partial z^2} + q_{\text{solid}} - q_{\text{as}} - q_{\text{cs}} \\
&\quad + \left( -(\vartheta_c - \vartheta_s) \bar{\nu}_{\text{red}}^+ Da_{\text{red}} r_{\text{red}} - \sum_{j=\text{ox}1,2} (\vartheta_a - \vartheta_s) \bar{\nu}_j^- Da_j r_j \right) c_p.
\end{aligned}$$

### A.2 Stromdichten und Potentiale

$$\begin{aligned}
i_a &= f^{i_a} := F \sum_{j=\text{ox}1,\text{ox}2} n_j Da_j r_j, \\
i_c &= f^{i_c} := -F n_{\text{red}} Da_{\text{red}} r_{\text{red}}, \\
i_e &= f^{i_e} := (\Phi_a^L - \Phi_c^L) \kappa_e, \\
I_a(t) &= \int_0^1 i_a dz, \quad I_c(t) = \int_0^1 i_c dz, \quad I_e(t) = \int_0^1 i_e dz, \\
i &= f^i := (c_a^{-1} + c_e^{-1} + c_c^{-1})^{-1} \left( \frac{i_a - I_a}{c_a} + \frac{i_e - I_e}{c_e} + \frac{i_c - I_c}{c_c} \right) + I_{\text{cell}}, \\
\frac{dU_{\text{cell}}}{dt} &= f^{U_{\text{cell}}} := \frac{I_a - I_{\text{cell}}}{c_a} + \frac{I_e - I_{\text{cell}}}{c_e} + \frac{I_c - I_{\text{cell}}}{c_c}, \\
\frac{\partial \Phi_a^L}{\partial t} &= f^{\Phi_a^L} := \frac{i_a - i}{c_a}, \quad \frac{\partial \Phi_c^L}{\partial t} = f^{\Phi_c^L} := \frac{i_a - i}{c_a} + \frac{i_e - i}{c_e}.
\end{aligned}$$

### A.3 Reaktionsraten und Ausdrücke

$$\begin{aligned}
q_{as} &= St_{as} (\vartheta_s - \vartheta_a), \\
q_{cs} &= St_{cs} (\vartheta_s - \vartheta_c), \\
q_{solid} &= (\Phi_a^L - \Phi_c^L) i_c / F + \sum_{j=ox1,2} [-\Delta_R h_j - n_j \Phi_a^L] Da_j r_j \\
&\quad + [-\Delta_R h_{red} + n_{red} (U_{cell} - \Phi_c^L)] Da_{red} r_{red}, \\
r_{ref1} &= \exp \left[ Arr_{ref1} \left( \frac{1}{\vartheta_{ref1}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \left( \chi_{CH_4,a} \chi_{H_2O,a} - \frac{\chi_{H_2,a}^3 \chi_{CO,a}}{K_{ref1}} \right), \\
r_{ref2} &= \exp \left[ Arr_{ref2} \left( \frac{1}{\vartheta_{ref2}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \left( \chi_{CO,a} \chi_{H_2O,a} - \frac{\chi_{H_2,a} \chi_{CO_2,a}}{K_{ref2}} \right), \\
r_{ox1} &= \exp \left[ Arr_{ox1} \left( \frac{1}{\vartheta_{ox1}^0} - \frac{1}{\vartheta_s} \right) \right] \left\{ \varphi_{H_2,a} \exp \left[ \frac{\alpha_{ox1}^+ n_{ox1} (-\Phi_a^L - \Delta\Phi_{ox1}^0)}{\vartheta_s} \right] \right. \\
&\quad \left. - \varphi_{H_2O,a} \varphi_{CO_2,a} \exp \left[ \frac{-(1 - \alpha_{ox1}^+) n_{ox1} (-\Phi_a^L - \Delta\Phi_{ox1}^0)}{\vartheta_s} \right] \right\}, \\
r_{ox2} &= \exp \left[ Arr_{ox2} \left( \frac{1}{\vartheta_{ox2}^0} - \frac{1}{\vartheta_s} \right) \right] \left\{ \varphi_{CO,a} \exp \left[ \frac{\alpha_{ox2}^+ n_{ox2} (-\Phi_a^L - \Delta\Phi_{ox2}^0)}{\vartheta_s} \right] \right. \\
&\quad \left. - \varphi_{CO_2,a}^2 \exp \left[ \frac{-(1 - \alpha_{ox2}^+) n_{ox2} (-\Phi_a^L - \Delta\Phi_{ox2}^0)}{\vartheta_s} \right] \right\}, \\
r_{red} &= \exp \left[ Arr_{red} \left( \frac{1}{\vartheta_{red}^0} - \frac{1}{\vartheta_s} \right) \right] \left\{ \varphi_{CO_2,c}^{-2} \exp \left[ \frac{\alpha_{red}^+ (U_{cell} - \Phi_c^L - \Delta\Phi_{red}^0)}{\vartheta_s} \right] \right. \\
&\quad \left. - \varphi_{O_2,c}^{0.75} \varphi_{CO_2,c}^{-0.5} \exp \left[ \frac{-\alpha_{red}^- (U_{cell} - \Phi_c^L - \Delta\Phi_{red}^0)}{\vartheta_s} \right] \right\}, \\
\Delta\Phi_{ox1}^0(\vartheta_s) &= 28.26 - 19.84 \vartheta_s, \\
\Delta\Phi_{ox2}^0(\vartheta_s) &= 20.98 - 17.86 \vartheta_s, \\
\Delta\Phi_{red}^0(\vartheta_s) &= 78.00 - 23.06 \vartheta_s, \\
K_{ref1}(\vartheta_a) &= \exp(30.19 - 90.41/\vartheta_a), \\
K_{ref2}(\vartheta_a) &= \exp(-3.97 + 14.57/\vartheta_a), \\
\varphi_{k,a} &= \chi_{k,a} + \frac{1}{D_{k,as}} \sum_{j=ox1,2} \nu_{k,j} Da_j r_j, \quad k \in \mathcal{I}, \\
\varphi_{k,c} &= \chi_{k,c} + \frac{1}{D_{k,cs}} \sum_{j=red} \nu_{k,j} Da_j r_j, \quad k \in \mathcal{I}.
\end{aligned}$$

### A.4 Brenn- und Mischkammer

Übergang aus den Gaskanälen:

$$\begin{aligned}
\chi_{k,a,out}(t) &= \chi_{k,a}(z=1, t), & \vartheta_{a,out}(t) &= \vartheta_a(z=1, t), & \gamma_{a,out}(t) &= \gamma_a(z=1, t), \\
\chi_{k,c,out}(t) &= \chi_{k,c}(z=0, t), & \vartheta_{c,out}(t) &= \vartheta_c(z=0, t), & \gamma_{c,out}(t) &= \gamma_c(z=0, t).
\end{aligned}$$

Brennkammer:

$$\begin{aligned}
\chi_{k,b}(t) &= f^{\chi_{k,b}} := \frac{\gamma_{a,\text{out}}(t)}{\gamma_b(t)} \left( \chi_{k,a,\text{out}}(t) + \underbrace{\sum_{j \in \mathcal{I}} \nu_{k,Cj} \chi_{j,a,\text{out}}(t)}_{=0} \right) \\
&\quad + \frac{\gamma_{c,\text{out}}(t) R_{\text{back}}(t)}{\gamma_b(t)} \left( \chi_{k,c,\text{out}}(t) + \underbrace{\sum_{j \in \mathcal{I}} \nu_{k,Cj} \chi_{j,c,\text{out}}(t)}_{=0} \right) \\
&\quad + \frac{\gamma_{\text{air}}(t)}{\gamma_b(t)} \left( \chi_{k,\text{air}} + \underbrace{\sum_{j \in \mathcal{I}} \nu_{k,Cj} \chi_{j,\text{air}}}_{=0} \right), \quad k \in \mathcal{I}, \\
\vartheta_b(t) &= f^{\vartheta_b} := \frac{\gamma_{a,\text{out}}(t)}{\gamma_b(t) c_p} \left( c_p (\vartheta_{a,\text{out}}(t) - 1) - \sum_{k \in \mathcal{I}} \chi_{k,a,\text{out}}(t) \Delta_C h_k \right) \\
&\quad + \frac{\gamma_{c,\text{out}}(t) R_{\text{back}}(t)}{\gamma_b(t) c_p} \left( c_p (\vartheta_{c,\text{out}}(t) - 1) - \underbrace{\sum_{k \in \mathcal{I}} \chi_{k,c,\text{out}}(t) \Delta_C h_k}_{=0} \right) \\
&\quad + \frac{\gamma_{\text{air}}(t)}{\gamma_b(t) c_p} \left( c_p (\vartheta_{\text{air}}(t) - 1) - \underbrace{\sum_{k \in \mathcal{I}} \chi_{k,\text{air}} \Delta_C h_k}_{=0} \right), \\
\gamma_b(t) &= f^{\gamma_b} := \gamma_{a,\text{out}}(t) \left( 1 + \sum_{k \in \mathcal{I}} \bar{\nu}_{C,k} \chi_{k,a,\text{out}}(t) \right) \\
&\quad + \gamma_{c,\text{out}}(t) R_{\text{back}}(t) \left( 1 + \underbrace{\sum_{k \in \mathcal{I}} \bar{\nu}_{C,k} \chi_{k,c,\text{out}}(t)}_{=0} \right) \\
&\quad + \gamma_{\text{air}}(t) \left( 1 + \underbrace{\sum_{k \in \mathcal{I}} \bar{\nu}_{C,k} \chi_{k,\text{air}}}_{=0} \right), \\
\gamma_{\text{air}}(t) &= f^{\gamma_{\text{air}}} := \gamma_{a,\text{in}}(t) \frac{\lambda_{\text{air}}(t)}{\chi_{\text{O}_2,\text{air}}} \left( -\chi_{\text{O}_2,a,\text{in}}(t) - \sum_{k \in \mathcal{I}} \nu_{\text{O}_2,Ck} \chi_{k,a,\text{in}}(t) \right) \\
&= \gamma_{a,\text{in}}(t) \frac{\lambda_{\text{air}}(t)}{\chi_{\text{O}_2,\text{air}}} \left( 2 \chi_{\text{CH}_4,a,\text{in}}(t) + \frac{1}{2} \chi_{\text{H}_2,a,\text{in}}(t) \right).
\end{aligned}$$

Mischkammer:

$$\begin{aligned}
\frac{d\chi_{k,m}}{dt} &= f^{\chi_{k,m}} := \gamma_b(t) (\chi_{k,b}(t) - \chi_{k,m}(t)) \frac{\vartheta_m(t)}{V_m}, \quad k \in \mathcal{I}, \\
\frac{d\vartheta_m}{dt} &= f^{\vartheta_m} := \gamma_b(t) (\vartheta_b(t) - \vartheta_m(t)) \frac{\vartheta_m(t)}{V_m} \\
&\quad - \frac{St_m (\vartheta_m - \vartheta_u)}{c_p} \frac{\vartheta_m(t)}{V_m}, \\
\gamma_m(t) &= f^{\gamma_m} := \gamma_b(t) \frac{\vartheta_b(t)}{\vartheta_m(t)} - \frac{St_m (\vartheta_m - \vartheta_u)}{c_p \vartheta_m(t)}.
\end{aligned}$$

## A.5 Konstanten

$c_p$	4.5	$F$	3.5/8	$\chi_{\text{O}_2,\text{air}}$	0.21
$c_{p,s}$	10000	$\kappa_e$	1	$\chi_{\text{N}_2,\text{air}}$	0.79
$St_{\text{as}}$	80.0	$c_a$	0.00001	$\vartheta_u$	1
$St_{\text{cs}}$	120.0	$c_e$	0.00001	Pe	2.3
$St_m$	1.0	$c_c$	0.00001	$V_m$	5

$k$	$\Delta_C h_k$	$\nu_{k,\text{ref1}}$	$\nu_{k,\text{ref2}}$	$\nu_{k,\text{ox1}}$	$\nu_{k,\text{ox2}}$	$\nu_{k,\text{red}}$
CH <sub>4</sub>	-323.85	-1	0	0	0	0
H <sub>2</sub> O	0	-1	-1	1	0	0
H <sub>2</sub>	-97.62	3	1	-1	0	0
CO	-114.22	1	-1	0	-1	0
CO <sub>2</sub>	0	0	1	1	2	1
O <sub>2</sub>	0	0	0	0	0	0.5
N <sub>2</sub>	0	0	0	0	0	0

$k$	$\nu_{k,C \text{ CH}_4}$	$\nu_{k,C \text{ H}_2\text{O}}$	$\nu_{k,C \text{ H}_2}$	$\nu_{k,C \text{ CO}}$	$\nu_{k,C \text{ CO}_2}$	$\nu_{k,C \text{ O}_2}$	$\nu_{k,C \text{ N}_2}$
CH <sub>4</sub>	-1	0	0	0	0	0	0
H <sub>2</sub> O	2	0	1	0	0	0	0
H <sub>2</sub>	0	0	-1	0	0	0	0
CO	0	0	0	-1	0	0	0
CO <sub>2</sub>	1	0	0	1	0	0	0
O <sub>2</sub>	-2	0	-0.5	-0.5	0	0	0
N <sub>2</sub>	0	0	0	0	0	0	0

$$\bar{\nu}_j = \sum_{k \in \mathcal{I}} \nu_{k,j}, \quad \bar{\nu}_j^+ = \sum_{k \in \mathcal{I}, \nu_{k,j} > 0} \nu_{k,j}, \quad \bar{\nu}_j^- = \sum_{k \in \mathcal{I}, \nu_{k,j} < 0} \nu_{k,j}, \quad \bar{\nu}_{C,j} = \sum_{k \in \mathcal{I}} \nu_{k,C \text{ } j},$$

$$D_{k,\text{as}} = D_{k,\text{cs}} = 100, \quad k \in \mathcal{I}.$$

$j$	$Da_j$	$Arr_j$	$\vartheta_j^0$	$\Delta_R h_j$	$\alpha_j^+$	$\alpha_j^-$	$n_j$
ref1	25.0	84.4	2.93	90.5			
ref2	100.0	6.2	2.93	-14.5			
ox1	5.0	21.6	2.93	56.0	0.5		2
ox2	5.0	21.6	2.93	42.0	0.5		2
red	0.3	31.2	2.93	156.0	2.5	0.5	2

## A.6 Anfangs- und Randwerte

Anfangswerte: (Für grobe Diskretisierungen)

k	$\chi_{k,a,0}$	$\chi_{k,c,0}$	$\chi_{k,m,0}$	$\chi_{k,a,in}$
CH <sub>4</sub>	2/7	0	0	
H <sub>2</sub> O	5/7	0.2	0.2	
H <sub>2</sub>	0.0	0.0	0.0	
CO	0.0	0.0	0.0	0
CO <sub>2</sub>	0.0	0.1	0.1	0
O <sub>2</sub>	0.0	0.1	0.1	0
N <sub>2</sub>	0.0	0.6	0.6	0

$\vartheta_{a,0}$	3.0	$\gamma_{a,0}$	1.0	$\Phi_{a,0}^S$	32.3	$i_{a,0}$	0.3
$\vartheta_{c,0}$	3.0	$\gamma_{c,0}$	6.0	$\Phi_{c,0}^S$	32.3	$i_{c,0}$	0.3
$\vartheta_{m,0}$	3.0	$\gamma_{m,0}$	0.0	$\Phi_{c,0}^L$	32.3	$i_{e,0}$	0.3
$\vartheta_{s,0}$	3.1						



## Anhang B

# Adjungierte Gleichungen der MCFC

### B.1 Gaskanäle und Solid

$$\begin{aligned}
-\frac{\partial \tilde{\vartheta}_a}{\partial t} - v_a \frac{\partial \tilde{\vartheta}_a}{\partial z} &= \frac{\partial v_a}{\partial z} \tilde{\vartheta}_a + \left[ \sum_{k \in \mathcal{I}} \tilde{\chi}_{k,a} \sum_{j=\text{ox}1,2,\text{ref}1,2} (\nu_{k,j} - \chi_{k,a} \bar{\nu}_j) Da_j r_j \right] \\
&+ \tilde{\vartheta}_a \left( \frac{f^{\vartheta_a}}{\vartheta_a} - \frac{\vartheta_a}{c_p} St_{\text{as}} - \vartheta_a \sum_{j=\text{ox}1,2} \bar{\nu}_j^+ Da_j r_j \right) + \tilde{\vartheta}_s \left( St_{\text{as}} - c_p \sum_{j=\text{ox}1,2} \bar{\nu}_j^- Da_j r_j \right) \\
&+ \tilde{v}_a \left( -\frac{St_{\text{as}}}{c_p} - \sum_{j=\text{ox}1,2} \bar{\nu}_j^+ Da_j r_j + \sum_{j=\text{ox}1,2,\text{ref}1,2} \bar{\nu}_j Da_j r_j \right) + \sum_{j=\text{ref}1,2} \tilde{s}_j \frac{\partial r_j}{\partial \vartheta_a}, \\
-\frac{\partial \tilde{\vartheta}_c}{\partial t} + v_c \frac{\partial \tilde{\vartheta}_c}{\partial t} &= -\frac{\partial v_c}{\partial z} \tilde{\vartheta}_c + \tilde{\vartheta}_c \left( \frac{f^{\vartheta_c}}{\vartheta_c} - \underbrace{\vartheta_c \bar{\nu}_{\text{red}}^- Da_{\text{red}} r_{\text{red}}}_{=0} - \frac{\vartheta_c}{c_p} St_{\text{cs}} \right) \\
&+ \tilde{\vartheta}_s \left( St_{\text{cs}} - c_p \bar{\nu}_{\text{red}}^+ Da_{\text{red}} r_{\text{red}} \right) + \left[ \sum_{k \in \mathcal{I}} \frac{f^{\chi_{k,c}}}{\vartheta_c} \tilde{\chi}_{k,c} \right] \\
&+ \tilde{v}_c \left( \bar{\nu}_{\text{red}} Da_{\text{red}} r_{\text{red}} - \underbrace{\bar{\nu}_{\text{red}}^- Da_{\text{red}} r_{\text{red}}}_{=0} - \frac{St_{\text{cs}}}{c_p} \right), \\
-\frac{\partial \tilde{\chi}_{k,a}}{\partial t} - v_a \frac{\partial \tilde{\chi}_{k,a}}{\partial z} &= \frac{\partial v_a}{\partial z} \tilde{\chi}_{k,a} - \tilde{\chi}_{k,a} \vartheta_a \sum_{j=\text{ox}1,2,\text{ref}1,2} \bar{\nu}_j Da_j r_j + \sum_{j=\text{ox}1,2,\text{ref}1,2} \tilde{s}_j \frac{\partial r_j}{\partial \chi_{k,a}}, \\
-\frac{\partial \tilde{\chi}_{k,c}}{\partial t} + v_c \frac{\partial \tilde{\chi}_{k,c}}{\partial z} &= -\frac{\partial v_c}{\partial z} \tilde{\chi}_{k,c} - \tilde{\chi}_{k,c} \vartheta_c \bar{\nu}_{\text{red}} Da_{\text{red}} r_{\text{red}} + \tilde{s}_{\text{red}} \frac{\partial r_{\text{red}}}{\partial \chi_{k,c}}, \\
-c_{p,s} \frac{\partial \tilde{\vartheta}_s}{\partial t} - \frac{1}{\text{Pe}} \frac{\partial^2 \tilde{\vartheta}_s}{\partial z^2} &= \tilde{\vartheta}_a \frac{\vartheta_a}{c_p} \left( St_{\text{as}} + c_p \sum_{j=\text{ox}1,2} \bar{\nu}_j^+ Da_j r_j \right) + \tilde{\vartheta}_c \left( \underbrace{\vartheta_c \bar{\nu}_{\text{red}}^- Da_{\text{red}} r_{\text{red}}}_{=0} + \frac{\vartheta_c St_{\text{cs}}}{c_p} \right) \\
&+ \tilde{v}_a \left( \frac{St_{\text{as}}}{c_p} + \sum_{j=\text{ox}1,2} \bar{\nu}_j^+ Da_j r_j \right) + \tilde{v}_c \left( \underbrace{\bar{\nu}_{\text{red}}^- Da_{\text{red}} r_{\text{red}}}_{=0} + \frac{St_{\text{cs}}}{c_p} \right) \\
&+ \tilde{\vartheta}_s \left( -St_{\text{as}} - St_{\text{cs}} + c_p \bar{\nu}_{\text{red}}^+ Da_{\text{red}} r_{\text{red}} + c_p \sum_{j=\text{ox}1,2} \bar{\nu}_j^- Da_j r_j \right) + \sum_{j=\text{ox}1,2,\text{red}} \tilde{s}_j \frac{\partial r_j}{\partial \vartheta_s}, \\
-\frac{\partial \tilde{v}_a}{\partial z} &= -\sum_{k \in \mathcal{I}} \frac{\partial \chi_{k,a}}{\partial z} \tilde{\chi}_{k,a} - \frac{\partial \vartheta_a}{\partial z} \tilde{\vartheta}_a, \\
\frac{\partial \tilde{v}_c}{\partial z} &= \sum_{k \in \mathcal{I}} \frac{\partial \chi_{k,c}}{\partial z} \tilde{\chi}_{k,c} + \frac{\partial \vartheta_c}{\partial z} \tilde{\vartheta}_c.
\end{aligned}$$

Hilfsgrößen:

$$\begin{aligned}
\tilde{s}_j &:= \sum_{k \in \mathcal{I}} \tilde{\chi}_{k,a} \vartheta_a (\nu_{k,j} - \chi_{k,a} \bar{\nu}_j) Da_j - \tilde{\vartheta}_a \frac{\vartheta_a}{c_p} \Delta_R h_j Da_j \\
&\quad + \tilde{v}_a \left( -\Delta_R h_j Da_j / c_p + \vartheta_a \bar{\nu}_j Da_j \right), \quad j = \text{ref1, ref2}, \\
\tilde{s}_j &:= \sum_{k \in \mathcal{I}} \tilde{\chi}_{k,a} \vartheta_a (\nu_{k,j} - \chi_{k,a} \bar{\nu}_j) Da_j + \tilde{\vartheta}_a \vartheta_a \bar{\nu}_j^+ Da_j (\vartheta_s - \vartheta_a) \\
&\quad + \tilde{v}_a \left( \bar{\nu}_j^+ Da_j (\vartheta_s - \vartheta_a) + \vartheta_a \bar{\nu}_j Da_j \right) + \tilde{i}_a F n_j Da_j \\
&\quad + \tilde{\vartheta}_s \left( [-\Delta_R h_j - n_j \Phi_a^L] Da_j - (\vartheta_a - \vartheta_s) \bar{\nu}_j^- Da_j c_p \right), \quad j = \text{ox1, ox2}, \\
\tilde{s}_{\text{red}} &:= \sum_{k \in \mathcal{I}} \tilde{\chi}_{k,c} \vartheta_c (\nu_{k,\text{red}} - \chi_{k,c} \bar{\nu}_{\text{red}}) Da_{\text{red}} + \underbrace{\tilde{\vartheta}_c \vartheta_c \bar{\nu}_{\text{red}}^- Da_{\text{red}} (\vartheta_s - \vartheta_c)}_{=0} \\
&\quad + \tilde{v}_c \left( \vartheta_c \bar{\nu}_{\text{red}} Da_{\text{red}} + \underbrace{\bar{\nu}_{\text{red}}^- Da_{\text{red}} (\vartheta_s - \vartheta_c)}_{=0} \right) - \tilde{i}_c F n_{\text{red}} Da_{\text{red}} \\
&\quad + \tilde{\vartheta}_s \left( [-\Delta_R h_{\text{red}} + n_{\text{red}} (U_{\text{cell}} - \Phi_c^L)] Da_{\text{red}} - (\vartheta_c - \vartheta_s) \bar{\nu}_{\text{red}}^+ Da_{\text{red}} c_p \right).
\end{aligned}$$

## B.2 Stromdichten und Potentiale

$$\begin{aligned}
-\frac{\partial \tilde{\Phi}_a^L}{\partial t}(z, t) &= \tilde{\vartheta}_s \left( \frac{i_e}{F} - \sum_{j=\text{ox1,2}} n_j Da_j r_j \right) + \tilde{i}_e \kappa_e + \sum_{j=\text{ox1,2}} \tilde{s}_j \frac{\partial r_j}{\partial \Phi_a^L}, \\
-\frac{\partial \tilde{\Phi}_c^L}{\partial t}(z, t) &= -\tilde{\vartheta}_s \left( \frac{i_e}{F} + n_{\text{red}} Da_{\text{red}} r_{\text{red}} \right) - \tilde{i}_e \kappa_e + \tilde{s}_{\text{red}} \frac{\partial r_{\text{red}}}{\partial \Phi_c^L}, \\
\tilde{i}(z, t) &= -\frac{1}{c_a} \tilde{\Phi}_a^L - \tilde{\Phi}_c^L \left( \frac{1}{c_a} + \frac{1}{c_e} \right), \\
\tilde{i}_a(z, t) &= \tilde{I}_a(t) + \tilde{i} (c_a^{-1} + c_e^{-1} + c_c^{-1})^{-1} \frac{1}{c_a} + \tilde{\Phi}_a^L \frac{1}{c_a} + \tilde{\Phi}_c^L \frac{1}{c_a}, \\
\tilde{i}_c(z, t) &= \tilde{I}_c(t) + \tilde{i} (c_a^{-1} + c_e^{-1} + c_c^{-1})^{-1} \frac{1}{c_c}, \\
\tilde{i}_e(z, t) &= \tilde{I}_e(t) + \tilde{\vartheta}_s (\Phi_a^L - \Phi_c^L) / F + \tilde{i} (c_a^{-1} + c_e^{-1} + c_c^{-1})^{-1} \frac{1}{c_e} + \tilde{\Phi}_c^L \frac{1}{c_e}, \\
\tilde{I}_k(t) &= \frac{1}{c_k} \tilde{U}_{\text{cell}}(t) - (c_a^{-1} + c_e^{-1} + c_c^{-1})^{-1} \frac{1}{c_k} \int_0^1 \tilde{i}(z, t) dz, \quad k \in \{a, c, e\}, \\
-\frac{d\tilde{U}_{\text{cell}}}{dt}(t) &= \int_0^1 \tilde{s}_{\text{red}}(z, t) \frac{\partial r_{\text{red}}}{\partial U_{\text{cell}}}(z, t) + \tilde{\vartheta}_s(z, t) n_{\text{red}} Da_{\text{red}} r_{\text{red}}(z, t) dz.
\end{aligned}$$

### B.3 Partielle Ableitungen der Reaktionsraten

$$\begin{aligned}
\frac{\partial r_{\text{ref1}}}{\partial \vartheta_a} &= \frac{Arr_{\text{ref1}}}{\vartheta_a^2} r_{\text{ref1}} + \exp \left[ Arr_{\text{ref1}} \left( \frac{1}{\vartheta_{\text{ref1}}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \left( 90.41 \frac{\chi_{\text{H}_2,\text{a}}^3 \chi_{\text{CO},\text{a}}}{K_{\text{ref1}} \vartheta_a^2} \right), \\
\frac{\partial r_{\text{ref1}}}{\partial \chi_{\text{CH}_4,\text{a}}} &= \exp \left[ Arr_{\text{ref1}} \left( \frac{1}{\vartheta_{\text{ref1}}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \chi_{\text{H}_2\text{O},\text{a}}, \\
\frac{\partial r_{\text{ref1}}}{\partial \chi_{\text{H}_2\text{O},\text{a}}} &= \exp \left[ Arr_{\text{ref1}} \left( \frac{1}{\vartheta_{\text{ref1}}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \chi_{\text{CH}_4,\text{a}}, \\
\frac{\partial r_{\text{ref1}}}{\partial \chi_{\text{H}_2,\text{a}}} &= \exp \left[ Arr_{\text{ref1}} \left( \frac{1}{\vartheta_{\text{ref1}}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \left( -3 \frac{\chi_{\text{H}_2,\text{a}}^2 \chi_{\text{CO},\text{a}}}{K_{\text{ref1}}} \right), \\
\frac{\partial r_{\text{ref1}}}{\partial \chi_{\text{CO},\text{a}}} &= \exp \left[ Arr_{\text{ref1}} \left( \frac{1}{\vartheta_{\text{ref1}}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \left( -\frac{\chi_{\text{H}_2,\text{a}}^3}{K_{\text{ref1}}} \right), \\
\frac{\partial r_{\text{ref2}}}{\partial \vartheta_a} &= \frac{Arr_{\text{ref2}}}{\vartheta_a^2} r_{\text{ref2}} + \exp \left[ Arr_{\text{ref2}} \left( \frac{1}{\vartheta_{\text{ref2}}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \left( -14.57 \frac{\chi_{\text{H}_2,\text{a}} \chi_{\text{CO}_2,\text{a}}}{K_{\text{ref2}} \vartheta_a^2} \right), \\
\frac{\partial r_{\text{ref2}}}{\partial \chi_{\text{CO},\text{a}}} &= \exp \left[ Arr_{\text{ref2}} \left( \frac{1}{\vartheta_{\text{ref2}}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \chi_{\text{H}_2\text{O},\text{a}}, \\
\frac{\partial r_{\text{ref2}}}{\partial \chi_{\text{H}_2\text{O},\text{a}}} &= \exp \left[ Arr_{\text{ref2}} \left( \frac{1}{\vartheta_{\text{ref2}}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \chi_{\text{CO},\text{a}}, \\
\frac{\partial r_{\text{ref2}}}{\partial \chi_{\text{H}_2,\text{a}}} &= \exp \left[ Arr_{\text{ref2}} \left( \frac{1}{\vartheta_{\text{ref2}}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \left( -\frac{\chi_{\text{CO}_2,\text{a}}}{K_{\text{ref2}}} \right), \\
\frac{\partial r_{\text{ref2}}}{\partial \chi_{\text{CO}_2,\text{a}}} &= \exp \left[ Arr_{\text{ref2}} \left( \frac{1}{\vartheta_{\text{ref2}}^0} - \frac{1}{\vartheta_a} \right) \right] \cdot \left( -\frac{\chi_{\text{H}_2,\text{a}}}{K_{\text{ref2}}} \right), \\
\frac{\partial r_{\text{ox1}}}{\partial \vartheta_s} &= \frac{Arr_{\text{ox1}}}{\vartheta_s^2} r_{\text{ox1}} + \exp \left[ Arr_{\text{ox1}} \left( \frac{1}{\vartheta_{\text{ox1}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \frac{n_{\text{ox1}} (\Phi_a^L + 28.26)}{\vartheta_s^2} \\
&\quad \cdot \left( \varphi_{\text{H}_2,\text{a}} \exp \left[ \frac{\alpha_{\text{ox1}}^+ n_{\text{ox1}} (-\Phi_a^L - \Delta\Phi_{\text{ox1}}^0)}{\vartheta_s} \right] \cdot \alpha_{\text{ox1}}^+ \right. \\
&\quad \left. + \varphi_{\text{H}_2\text{O},\text{a}} \varphi_{\text{CO}_2,\text{a}} \exp \left[ \frac{-(1 - \alpha_{\text{ox1}}^+) n_{\text{ox1}} (-\Phi_a^L - \Delta\Phi_{\text{ox1}}^0)}{\vartheta_s} \right] \cdot (1 - \alpha_{\text{ox1}}^+) \right), \\
\frac{\partial r_{\text{ox1}}}{\partial \varphi_{\text{H}_2,\text{a}}} &= \exp \left[ Arr_{\text{ox1}} \left( \frac{1}{\vartheta_{\text{ox1}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \exp \left[ \frac{\alpha_{\text{ox1}}^+ n_{\text{ox1}} (-\Phi_a^L - \Delta\Phi_{\text{ox1}}^0)}{\vartheta_s} \right], \\
\frac{\partial r_{\text{ox1}}}{\partial \varphi_{\text{H}_2\text{O},\text{a}}} &= \exp \left[ Arr_{\text{ox1}} \left( \frac{1}{\vartheta_{\text{ox1}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \left( -\varphi_{\text{CO}_2,\text{a}} \exp \left[ \frac{-(1 - \alpha_{\text{ox1}}^+) n_{\text{ox1}} (-\Phi_a^L - \Delta\Phi_{\text{ox1}}^0)}{\vartheta_s} \right] \right), \\
\frac{\partial r_{\text{ox1}}}{\partial \varphi_{\text{CO}_2,\text{a}}} &= \exp \left[ Arr_{\text{ox1}} \left( \frac{1}{\vartheta_{\text{ox1}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \left( -\varphi_{\text{H}_2\text{O},\text{a}} \exp \left[ \frac{-(1 - \alpha_{\text{ox1}}^+) n_{\text{ox1}} (-\Phi_a^L - \Delta\Phi_{\text{ox1}}^0)}{\vartheta_s} \right] \right), \\
\frac{\partial r_{\text{ox1}}}{\partial \Phi_a^L} &= \exp \left[ Arr_{\text{ox1}} \left( \frac{1}{\vartheta_{\text{ox1}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \frac{n_{\text{ox1}}}{\vartheta_s} \left( -\varphi_{\text{H}_2,\text{a}} \exp \left[ \frac{\alpha_{\text{ox1}}^+ n_{\text{ox1}} (-\Phi_a^L - \Delta\Phi_{\text{ox1}}^0)}{\vartheta_s} \right] \cdot \alpha_{\text{ox1}}^+ \right. \\
&\quad \left. - \varphi_{\text{H}_2\text{O},\text{a}} \varphi_{\text{CO}_2,\text{a}} \exp \left[ \frac{-(1 - \alpha_{\text{ox1}}^+) n_{\text{ox1}} (-\Phi_a^L - \Delta\Phi_{\text{ox1}}^0)}{\vartheta_s} \right] \cdot (1 - \alpha_{\text{ox1}}^+) \right),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial r_{\text{ox2}}}{\partial \vartheta_s} &= \frac{Arr_{\text{ox2}}}{\vartheta_s^2} r_{\text{ox2}} + \exp \left[ Arr_{\text{ox2}} \left( \frac{1}{\vartheta_{\text{ox2}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \frac{n_{\text{ox2}} (\Phi_a^L + 20.98)}{\vartheta_s^2} \\
&\quad \cdot \left( \varphi_{\text{CO}_2, \text{a}} \exp \left[ \frac{\alpha_{\text{ox2}}^+ n_{\text{ox2}} (-\Phi_a^L - \Delta\Phi_{\text{ox2}}^0)}{\vartheta_s} \right] \cdot \alpha_{\text{ox2}}^+ \right. \\
&\quad \left. + \varphi_{\text{CO}_2, \text{a}}^2 \exp \left[ \frac{-(1 - \alpha_{\text{ox2}}^+) n_{\text{ox2}} (-\Phi_a^L - \Delta\Phi_{\text{ox2}}^0)}{\vartheta_s} \right] \cdot (1 - \alpha_{\text{ox2}}^+) \right), \\
\frac{\partial r_{\text{ox2}}}{\partial \varphi_{\text{CO}_2, \text{a}}} &= \exp \left[ Arr_{\text{ox2}} \left( \frac{1}{\vartheta_{\text{ox2}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \exp \left[ \frac{\alpha_{\text{ox2}}^+ n_{\text{ox2}} (-\Phi_a^L - \Delta\Phi_{\text{ox2}}^0)}{\vartheta_s} \right], \\
\frac{\partial r_{\text{ox2}}}{\partial \varphi_{\text{CO}_2, \text{a}}} &= -2 \varphi_{\text{CO}_2, \text{a}} \cdot \exp \left[ Arr_{\text{ox2}} \left( \frac{1}{\vartheta_{\text{ox2}}^0} - \frac{1}{\vartheta_s} \right) \right] \exp \left[ \frac{-(1 - \alpha_{\text{ox2}}^+) n_{\text{ox2}} (-\Phi_a^L - \Delta\Phi_{\text{ox2}}^0)}{\vartheta_s} \right], \\
\frac{\partial r_{\text{ox2}}}{\partial \Phi_a^L} &= \exp \left[ Arr_{\text{ox2}} \left( \frac{1}{\vartheta_{\text{ox2}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \frac{n_{\text{ox2}}}{\vartheta_s} \left( -\varphi_{\text{CO}_2, \text{a}} \exp \left[ \frac{\alpha_{\text{ox2}}^+ n_{\text{ox2}} (-\Phi_a^L - \Delta\Phi_{\text{ox2}}^0)}{\vartheta_s} \right] \cdot \alpha_{\text{ox2}}^+ \right. \\
&\quad \left. - \varphi_{\text{CO}_2, \text{a}}^2 \exp \left[ \frac{-(1 - \alpha_{\text{ox2}}^+) n_{\text{ox2}} (-\Phi_a^L - \Delta\Phi_{\text{ox2}}^0)}{\vartheta_s} \right] \cdot (1 - \alpha_{\text{ox2}}^+) \right), \\
\frac{\partial r_{\text{red}}}{\partial \vartheta_s} &= \frac{Arr_{\text{red}}}{\vartheta_s^2} r_{\text{red}} + \exp \left[ Arr_{\text{red}} \left( \frac{1}{\vartheta_{\text{red}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \frac{U_{\text{cell}} - \Phi_c^L - 78.00}{\vartheta_s^2} \\
&\quad \cdot \left( -\varphi_{\text{CO}_2, \text{c}}^{-2} \exp \left[ \frac{\alpha_{\text{red}}^+ (U_{\text{cell}} - \Phi_c^L - \Delta\Phi_{\text{red}}^0)}{\vartheta_s} \right] \cdot \alpha_{\text{red}}^+ \right. \\
&\quad \left. - \varphi_{\text{O}_2, \text{c}}^{0.75} \varphi_{\text{CO}_2, \text{c}}^{-0.5} \exp \left[ \frac{-\alpha_{\text{red}}^- (U_{\text{cell}} - \Phi_c^L - \Delta\Phi_{\text{red}}^0)}{\vartheta_s} \right] \cdot \alpha_{\text{red}}^- \right), \\
\frac{\partial r_{\text{red}}}{\partial \varphi_{\text{CO}_2, \text{c}}} &= \exp \left[ Arr_{\text{red}} \left( \frac{1}{\vartheta_{\text{red}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \left\{ -2 \varphi_{\text{CO}_2, \text{c}}^{-3} \exp \left[ \frac{\alpha_{\text{red}}^+ (U_{\text{cell}} - \Phi_c^L - \Delta\Phi_{\text{red}}^0)}{\vartheta_s} \right] \right. \\
&\quad \left. + 0.5 \varphi_{\text{O}_2, \text{c}}^{0.75} \varphi_{\text{CO}_2, \text{c}}^{-1.5} \exp \left[ \frac{-\alpha_{\text{red}}^- (U_{\text{cell}} - \Phi_c^L - \Delta\Phi_{\text{red}}^0)}{\vartheta_s} \right] \right\}, \\
\frac{\partial r_{\text{red}}}{\partial \varphi_{\text{O}_2, \text{c}}} &= -0.75 \varphi_{\text{O}_2, \text{c}}^{-0.25} \varphi_{\text{CO}_2, \text{c}}^{-0.5} \exp \left[ Arr_{\text{red}} \left( \frac{1}{\vartheta_{\text{red}}^0} - \frac{1}{\vartheta_s} \right) \right] \exp \left[ \frac{-\alpha_{\text{red}}^- (U_{\text{cell}} - \Phi_c^L - \Delta\Phi_{\text{red}}^0)}{\vartheta_s} \right], \\
\frac{\partial r_{\text{red}}}{\partial \Phi_c^L} &= \exp \left[ Arr_{\text{red}} \left( \frac{1}{\vartheta_{\text{red}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \left( \varphi_{\text{CO}_2, \text{c}}^{-2} \exp \left[ \frac{\alpha_{\text{red}}^+ (U_{\text{cell}} - \Phi_c^L - \Delta\Phi_{\text{red}}^0)}{\vartheta_s} \right] \cdot \frac{-\alpha_{\text{red}}^+}{\vartheta_s} \right. \\
&\quad \left. - \varphi_{\text{O}_2, \text{c}}^{0.75} \varphi_{\text{CO}_2, \text{c}}^{-0.5} \exp \left[ \frac{-\alpha_{\text{red}}^- (U_{\text{cell}} - \Phi_c^L - \Delta\Phi_{\text{red}}^0)}{\vartheta_s} \right] \cdot \frac{\alpha_{\text{red}}^-}{\vartheta_s} \right), \\
\frac{\partial r_{\text{red}}}{\partial U_{\text{cell}}} &= \exp \left[ Arr_{\text{red}} \left( \frac{1}{\vartheta_{\text{red}}^0} - \frac{1}{\vartheta_s} \right) \right] \cdot \left( \varphi_{\text{CO}_2, \text{c}}^{-2} \exp \left[ \frac{\alpha_{\text{red}}^+ (U_{\text{cell}} - \Phi_c^L - \Delta\Phi_{\text{red}}^0)}{\vartheta_s} \right] \cdot \frac{\alpha_{\text{red}}^+}{\vartheta_s} \right. \\
&\quad \left. + \varphi_{\text{O}_2, \text{c}}^{0.75} \varphi_{\text{CO}_2, \text{c}}^{-0.5} \exp \left[ \frac{-\alpha_{\text{red}}^- (U_{\text{cell}} - \Phi_c^L - \Delta\Phi_{\text{red}}^0)}{\vartheta_s} \right] \cdot \frac{\alpha_{\text{red}}^-}{\vartheta_s} \right).
\end{aligned}$$

## B.4 Brenn- und Mischkammer

$$\begin{aligned}
\tilde{\gamma}_b &= \sum_{k \in \mathcal{I}} \frac{f^{\chi_{k,m}}}{\gamma_b} \tilde{\chi}_{k,m} + (\vartheta_b - \vartheta_m) \frac{\vartheta_m}{V_m} \tilde{\vartheta}_m + \frac{\vartheta_b}{\vartheta_m} \tilde{\gamma}_m - \frac{1}{\gamma_b} \left( f^{\vartheta_b} \tilde{\vartheta}_b + \sum_{k \in \mathcal{I}} f^{\chi_{k,b}} \tilde{\chi}_{k,b} \right), \\
\tilde{\chi}_{k,b} &= \frac{\gamma_b \vartheta_m}{V_m} \tilde{\chi}_{k,m}, & \tilde{\vartheta}_b &= \frac{\gamma_b \vartheta_m}{V_m} \tilde{\vartheta}_m + \frac{\gamma_b}{\vartheta_m} \tilde{\gamma}_m, \\
\tilde{\gamma}_{\text{air}} &= \tilde{\chi}_{\text{O}_2,b} \frac{\chi_{\text{O}_2,\text{air}}}{\gamma_b} + \tilde{\vartheta}_b \frac{\vartheta_{\text{air}} - 1}{\gamma_b} + \tilde{\gamma}_b, & \tilde{\gamma}_m &= \vartheta_m \tilde{v}_c, \\
-\frac{d\tilde{\chi}_{k,m}}{dt} &= [v_c \tilde{\chi}_{k,c}|_{(1,t)} - \frac{\gamma_b \vartheta_m}{V_m} \tilde{\chi}_{k,m}], \\
-\frac{d\tilde{\vartheta}_m}{dt} &= [v_c \tilde{\vartheta}_c|_{(1,t)} + \frac{1}{\vartheta_m} \left\{ \sum_{k \in \mathcal{I}} f^{\chi_{k,m}} \chi_{k,m} \right\} + \left( \frac{1}{\vartheta_m} f^{\vartheta_m} - \frac{\vartheta_m}{V_m} (\gamma_b + \frac{St_m}{c_p}) \right) \tilde{\vartheta}_m \\
&\quad - \frac{1}{\vartheta_m^2} \left( \gamma_b \vartheta_b + \frac{St_m \vartheta_u}{c_p} \right) \tilde{\gamma}_m.
\end{aligned}$$

## B.5 Übergangs- und Randbedingungen

$$\begin{aligned}
v_{a,\text{out}} \tilde{\vartheta}_a(1,t) &= -\frac{v_{a,\text{out}}}{\vartheta_{a,\text{out}}^2 \gamma_b} \left\{ \sum_{k \in \mathcal{I}} \left( \chi_{k,a,\text{out}}(t) + \sum_{j \in \mathcal{I}} \nu_{k,Cj} \chi_{j,a,\text{out}}(t) \right) \tilde{\chi}_{k,b} \right\} \\
&\quad - \frac{v_{a,\text{out}}}{\vartheta_{a,\text{out}}^2 \gamma_b c_p} \left( -c_p - \sum_{k \in \mathcal{I}} \chi_{k,a,\text{out}}(t) \Delta_C h_k \right) \tilde{\vartheta}_b \\
&\quad - \frac{v_{a,\text{out}}}{\vartheta_{a,\text{out}}^2} \left( 1 + \sum_{k \in \mathcal{I}} \bar{\nu}_{C,k} \chi_{k,a,\text{out}}(t) \right) \tilde{\gamma}_b, \\
v_{a,\text{out}} \tilde{\chi}_{k,a}(1,t) &= \frac{v_{a,\text{out}}}{\vartheta_{a,\text{out}} \gamma_b} \left( \tilde{\chi}_{k,b} + \sum_{j \in \mathcal{I}} \nu_{j,Ck} \tilde{\chi}_{j,b} \right) - \frac{v_{a,\text{out}}}{\vartheta_{a,\text{out}} \gamma_b c_p} \Delta_C h_k \tilde{\vartheta}_b + \frac{v_{a,\text{out}}}{\vartheta_{a,\text{out}}} \bar{\nu}_{C,k} \tilde{\gamma}_b, \\
\tilde{v}_a(1,t) &= \frac{1}{\vartheta_{a,\text{out}}(t) \gamma_b(t)} \left\{ \sum_{k \in \mathcal{I}} \left( \chi_{k,a,\text{out}}(t) + \sum_{j \in \mathcal{I}} \nu_{k,Cj} \chi_{j,a,\text{out}}(t) \right) \tilde{\chi}_{k,b} \right\} \\
&\quad + \frac{1}{\vartheta_{a,\text{out}} \gamma_b(t) c_p} \left( c_p (\vartheta_{a,\text{out}}(t) - 1) - \sum_{k \in \mathcal{I}} \chi_{k,a,\text{out}}(t) \Delta_C h_k \right) \tilde{\vartheta}_b \\
&\quad + \frac{1}{\vartheta_{a,\text{out}}} \left( 1 + \sum_{k \in \mathcal{I}} \bar{\nu}_{C,k} \chi_{k,a,\text{out}}(t) \right) \tilde{\gamma}_b, \\
v_{c,\text{out}} \tilde{\vartheta}_c(0,t) &= -\frac{v_{c,\text{out}} R_{\text{back}}}{\vartheta_{c,\text{out}}^2 \gamma_b} \left\{ \sum_{k \in \mathcal{I}} \chi_{k,c,\text{out}}(t) \tilde{\chi}_{k,b} \right\} + \frac{v_{c,\text{out}} R_{\text{back}}(t)}{\vartheta_{c,\text{out}}^2 \gamma_b(t)} \tilde{\vartheta}_b - \frac{v_{c,\text{out}} R_{\text{back}}}{\vartheta_{c,\text{out}}^2} \tilde{\gamma}_b, \\
v_{c,\text{out}} \tilde{\chi}_{k,c}(0,t) &= \frac{v_{c,\text{out}} R_{\text{back}}}{\vartheta_{c,\text{out}} \gamma_b} \tilde{\chi}_{k,b}, \\
\tilde{v}_c(0,t) &= \frac{R_{\text{back}}}{\vartheta_{c,\text{out}} \gamma_b} \left\{ \sum_{k \in \mathcal{I}} \chi_{k,c,\text{out}}(t) \tilde{\chi}_{k,b} \right\} + \frac{R_{\text{back}}}{\vartheta_{c,\text{out}} \gamma_b} (\vartheta_{c,\text{out}}(t) - 1) \tilde{\vartheta}_b + \frac{R_{\text{back}}}{\vartheta_{c,\text{out}}} \tilde{\gamma}_b, \\
\frac{\partial \tilde{\vartheta}_s}{\partial z}(0,t) &= 0 = \frac{\partial \tilde{\vartheta}_s}{\partial z}(1,t).
\end{aligned}$$



# Symbolverzeichnis

$\mathcal{A}$	aktive Menge
$\mathring{\mathcal{A}}, \bar{\mathcal{A}}, \partial\mathcal{A}$	Inneres, Abschluss und Rand der Menge $\mathcal{A}$
$C^k(\Omega)$	Menge der $k$ -mal stetig-differenzierbaren Funktionen auf $\Omega$
$C(\Omega)$	Menge der stetigen Funktionen auf $\Omega$
$C_0^\infty(\Omega)$	Menge der $\infty$ oft differenzierbaren Funktionen mit kompaktem Träger
$C([0, t_f]; L^2(\Omega))$	Raum der abstrakten, stetigen Funktionen mit Werten in $L^2(\Omega)$
$\nabla_{\vec{x}}$	Gradient bzgl. $\vec{x}$
$\delta(x - x_0)$	Dirac-Dichte zentriert in $x_0$
$\Delta$	Laplace-Operator
$\delta_\nu$	äußere Normalenableitung
$H$	Hamilton-Funktion
$H^k(\Omega)$	Sobolew-Raum, $H^k(\Omega) = W^{k,2}(\Omega)$
$\text{Hess}_f$	Hesse-Matrix von $f$
$\mathcal{L}$	Lagrange-Funktion
$L^p(\Omega)$	Lebesgue-Raum
$\mathcal{M}(\bar{\Omega})$	Raum aller regulären Borel-Maße auf $\bar{\Omega}$
$\nu$	äußeres Einheitsnormalenvektorfeld
$\mathbb{R}$	reelle Zahlen
$\text{sign}(x)$	Signum-Funktion, Vorzeichen von $x$
$T_t, T_{xx}$	partielle Ableitungen von $T(x, t)$ nach $t$ bzw. zweimal nach $x$
$V^*$	Dualraum von $V$
$W^{k,p}(\Omega)$	Sobolew-Raum
$y _{\mathcal{A}}$	Einschränkung der Funktion $y(x)$ auf die Menge $\mathcal{A}$
$\emptyset$	leere Menge
$\hookrightarrow$	stetige Einbettung
$[f(x)]_a^b$	$= f(b) - f(a), \quad [f(x)]_c := f(c)$
$\ f(x)\ _{L^2(\Omega)}$	$L^2$ -Norm der Funktion $f(x)$ auf $\Omega$
$\langle \cdot, \cdot \rangle_{V^*, V}$	Dualitätsprodukt auf einem Banachraum $V$



# Abkürzungsverzeichnis

AD	automatische Differentiation
AE	algebraische Gleichungen
arg min	Stelle, an der eine Funktion ihr Minimum annimmt
AWP	Anfangswertproblem
ARWP	Anfangs-Randwertproblem
BDD	Bryson, Denham, Dreyfus
BFGS-Verfahren	Verfahren von Broyden-Fletcher-Goldfarb-Shanno
CFL-Bedingung	Courant-Friedrichs-Lewy-Bedingung
const	konstant
DAE	differential-algebraische Gleichungen
FDTO	First-Discretize-Then-Optimize
FOTD	First-Optimize-Then-Discretize
f.ü.	fast überall
IP-Verfahren	Innere-Punkte Verfahren
JLS	Jacobson, Lele, Speyer
KKT-Bedingung	Karush-Kuhn-Tucker-Bedingung
KWK	Kraft-Wärme-Kopplung
LP	lineares Programm, lineares Optimierungsproblem
MCFC	Schmelzkarbonat-Brennstoffzelle (Molten Carbonate Fuel Cell)
max, min	maximiere, Maximum
min	minimiere, Minimum
NLP	nichtlineares Optimierungsproblem

OC-ODE	Optimalsteuerungsproblem mit gewöhnlichen Differentialgleichungen
OC-PDAE	OC-PDE mit zusätzlichen ODE oder AE
OC-PDE	Optimalsteuerungsproblem mit partiellen Differentialgleichungen
ODE	gewöhnliche Differentialgleichung (Ordinary Differential Equation)
PDAE	partiell-differential algebraische Gleichungen
PDE	partielle Differentialgleichung (Partial Differential Equation)
p.I.	partielle Integration
RB	Randbedingung
RWP	Randwertproblem
SQP	sequentielle quadratische Programmierung
s.t.	unter der Bedingung (subject to)
VZ	Vorzeichen

# Literaturverzeichnis

- [1] The AMPL Company (2011). <http://www.ampl.com/>.
- [2] Robert A. Adams und John J. F. Fournier. *Sobolev Spaces*. Elsevier, 2008.
- [3] Tognum AG. Ad-hoc-Meldung der Tognum AG vom 29.12.2010. <http://investoren.tognum.de/cgi-bin/show.ssp?companyName=tognum&language=German&id=3101&newsID=1016074>.
- [4] Jean Jaques Alibert und Jean Pierre Raymond. Boundary control of semilinear elliptic equations with discontinuous leading coefficients and unbounded controls. *Numerical Functional Analysis and Optimization*, 3&4:235–250, 1997.
- [5] Hans Wilhelm Alt. *Lineare Funktionalanalysis — Eine anwendungsorientierte Einführung*. Springer Berlin Heidelberg, 2006.
- [6] Nadir Arada, Jean Pierre Raymond und Fredi Tröltzsch. On an augmented lagrangian SQP method for a class of optimal control problems in Banach spaces. *Computational Optimization and Applications*, 22:369–398, 2002. 10.1023/A:1019763022415.
- [7] Ute Aßmann und Arnd Rösch. Identification of an unknown parameter function in the main part of an elliptic partial differential equation. *Preprint, Universität Duisburg-Essen*, 2011.
- [8] Jean Pierre Aubin. *Approximation of Elliptic Boundary-Value Problems*, volume XXVI of *Pure and Applied Mathematics*. Wiley-Interscience, New York, 1972.
- [9] Simon Bechmann und Michael Frey. *Regularisierungsmethoden für Optimalsteuerungsprobleme*, volume 80 of *Bayreuther Mathematische Schriften*. Mathematisches Institut der Universität Bayreuth, 2008.
- [10] Maïtine Bergounioux, Mounir Haddou, Michael Hintermüller und Karl Kunisch. A comparison of a Moreau-Yosida based active set strategy and interior point methods for constrained optimal control problems. *SIAM Journal on Optimization*, 11:495–521, 2000.
- [11] Maïtine Bergounioux, Kazufumi Ito und Karl Kunisch. Primal-dual strategy for constrained optimal control problems. *SIAM Journal on Control and Optimization*, 37:1176–1194, 1999.
- [12] Maïtine Bergounioux und Karl Kunisch. Primal-dual strategy for state-constrained optimal control problems. *Computational Optimization and Applications*, 22:193–224, 2002.

- [13] Maitine Bergounioux und Karl Kunisch. On the structure of Lagrange multipliers for state-constrained optimal control problems. *Systems & Control Letters*, 48:169–176, 2003.
- [14] Christian H. Bischof, H. Martin Bücker, Bruno Lang, Arno Rasch und Andre Vehreschild. Combining source transformation and operator overloading techniques to compute derivatives for MATLAB programs. In *Proceedings of the Second IEEE International Workshop on Source Code Analysis and Manipulation (SCAM 2002)*, pages 65–72, Los Alamitos, CA, USA, 2002. IEEE Computer Society.
- [15] Arthur Earl Bryson und Yu-Chi Ho. *Applied optimal control: optimization, estimation, and control*. Hemisphere Publishing Corporation, 1975.
- [16] Arthur Earl. Bryson, Jr., W. F. Denham und S. E. Dreyfus. Optimal programming problems with inequality constraints I: Necessary conditions for extremal solutions. *AIAA Journal*, 1(11):2544–2550, 1963.
- [17] Martin Bücker. Private Kommunikation (2011). RWTH Aachen.
- [18] Roland Bulirsch und Kurt Chudej. Effiziente und stabile Lösung eines reichweitenmaximalen Fluges mit Staudruckbeschränkung für ein Überschallflugzeug. *DGLR Jahrbuch*, 2:1141–1147, 1997.
- [19] Roland Zdeněk Bulirsch, Francesco Montrone und Hans Josef Pesch. Abort landing in the presence of windshear as a minimax optimal control problem. Part 1: Necessary conditions. *Journal of Optimization Theory and Applications*, 70:1–23, 1991.
- [20] Roland Zdeněk Bulirsch, Francesco Montrone und Hans Josef Pesch. Abort landing in the presence of windshear as a minimax optimal control problem. Part 2: Multiple shooting and homotopy. *Journal of Optimization Theory and Applications*, 70:223–254, 1991.
- [21] Eduardo Casas. *Análisis numérico de algunos problemas de optimización estructural*. PhD thesis, Univ. Santiago de Compostela (Spain), 1982.
- [22] Eduardo Casas. Control of an elliptic problem with pointwise state constraints. *SIAM Journal on Control and Optimization*, 4:1309–1322, 1986.
- [23] Eduardo Casas. Boundary control of semilinear elliptic equations with pointwise state constraints. *SIAM Journal on Control and Optimization*, 31:993–1006, 1993.
- [24] Eduardo Casas. Pontryagin’s principle for state-constrained boundary control problems of semilinear parabolic equations. *SIAM Journal on Control and Optimization*, 35(4):1297–1327, 1997.
- [25] Eduardo Casas, Mariano Mateos und Fredi Tröltzsch. Error estimates for the numerical approximation of boundary semilinear elliptic control problems. *Computational Optimization and Applications*, 31:193–219, 2005. 10.1007/s10589-005-2180-2.
- [26] Kurt Chudej. Private Kommunikation (2011). Universität Bayreuth.

- 
- [27] Kurt Chudej. *Effiziente Lösung zustandsbeschränkter Optimalsteuerungsaufgaben*. Habilitationsschrift, Fakultät für Mathematik und Physik, Universität Bayreuth, 2001.
- [28] Kurt Chudej und Michael Günther. A global state space approach for the numerical solution of state-constrained trajectory optimization problems. *Journal of Optimization Theory and Applications*, 103(1):75–93, 1999.
- [29] Kurt Chudej und Hans Josef Pesch. Private Kommunikation (2011). Universität Bayreuth.
- [30] Kurt Chudej, Hans Josef Pesch, Markus Wächter, Gottfried Sachs und Florent Le Bras. Instationary heat-constrained trajectory optimization of a hypersonic space vehicle by ODE-PDE-constrained optimal control. In Giuseppe Buttazzo, Aldo Frediani und Panos M. Pardalos, editors, *Variational Analysis and Aerospace Engineering*, volume 33 of *Optimization and its Applications*, pages 127–144. Springer New York, 2009.
- [31] Debora Clever und Jens Lang. Optimal control of radiative heat transfer in glass cooling with restrictions on the temperature gradient. *Optimal Control Applications and Methods*, 2011.
- [32] Juan Carlos De Los Reyes und Karl Kunisch. A semi-smooth Newton method for control constrained boundary optimal control of the Navier-Stokes equations. *Nonlinear Analysis*, 62:1289–1316, 2005.
- [33] Klaus Deckelnick und Michael Hinze. Identification of matrix parameters in elliptic PDEs. *Hamburger Beiträge zur Angewandten Mathematik*, 2011-05, 2011.
- [34] Karsten Eppler und Fredi Tröltzsch. Optimal control problems for the nonlinear heat equation. In Martin Grötschel, Sven O. Krumke und Jörg Rambau, editors, *Online Optimization of Large Scale Systems*, pages 173–184. Springer, Berlin, Heidelberg, 2001.
- [35] Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, 2002.
- [36] Gustav Feichtinger und Richard F. Hartl. *Optimale Kontrolle ökonomischer Prozesse*. Walter de Gruyter, Berlin, 1986.
- [37] Michael Frey. *Shape calculus applied to state-constrained elliptic optimal control problems (working title)*. PhD thesis, Universität Bayreuth, to appear 2012.
- [38] Michael Frey, Simon Bechmann und Armin Rund. Numerical solution of new necessary conditions for distributed optimal control problems of linear elliptic equations with state constraints (working title). In preparation, 2011.
- [39] Michael Frey, Simon Bechmann, Armin Rund und Hans Josef Pesch. New necessary conditions for distributed optimal control problems of linear elliptic equations with state constraints. In preparation, 2011.

- [40] Nicolas Gauger, Andrea Walther, Carsten Moldenhauer und Markus Widhalm. Automatic differentiation of an entire design chain for aerodynamic shape optimization. In Cameron Tropea, Suad Jakirlic, Hans-Joachim Heinemann, Rolf Henke und Heinz Hönliger, editors, *New Results in Numerical and Experimental Fluid Mechanics VI*, volume 96 of *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, pages 454–461. Springer Berlin, Heidelberg, 2008.
- [41] Carl Geiger und Christian Kanzow. *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer, Berlin, 1999.
- [42] Carl Geiger und Christian Kanzow. *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer, Berlin, 2002.
- [43] Christian Goll, Rolf Rannacher und Winnifried Wollner. Goal-oriented mesh adaptation in the space-time finite element approximation of the Black-Scholes equation. Preprint, 2011.
- [44] Pierre Grisvard. *Elliptic Problems in Nonsmooth Domains*. Pitman Advanced Publishing Program, Boston London Melbourne, 1985.
- [45] Matthias Gundermann, Peter Heidebrecht und Kai Sundmacher. Validation of a mathematical model using an industrial MCFC plant. *Journal of Fuel Cell Science and Technology*, 3:303–307, 2006.
- [46] William W. Hager. Runge-kutta methods in optimal control and the transformed adjoint system. *Numerische Mathematik*, 87(2):247–282, 2000.
- [47] Ernst Hairer, Christian Lubich und Michel Roche. *The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods*, volume 1409 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1989.
- [48] Ernst Hairer und Gerhard Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, volume 14 of *Springer Series in Computational Mathematics*. Springer, Berlin, 1996.
- [49] Martin Hanke-Bourgeois. *Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens*. B. G. Teubner, Stuttgart, 2009.
- [50] Richard F. Hartl, Suresh P. Sethi und Raymond G. Vickson. A survey of the maximum principles for optimal control problems with state constraints. *SIAM Review*, 37(2):181–218, 1995.
- [51] Peter Heidebrecht. *Modelling, Analysis and Optimisation of a Molten Carbonate Fuel Cell with Direct Internal Reforming (DIR-MCFC)*. PhD thesis, Universität Magdeburg, 2004.
- [52] Roland Herzog und Karl Kunisch. Algorithms for PDE-constrained optimization. *GAMM-Mitteilungen*, 33(2):163–176, 2010.
- [53] Michael Hintermüller, Kazufumi Ito und Karl Kunisch. The primal-dual active set method as a semi-smooth Newton method. *SIAM Journal on Optimization*, 13(3):865–888, 2003.

- [54] Michael Hintermüller und Karl Kunisch. PDE-constrained optimization subject to pointwise constraints on the control, the state, and its derivative. *SIAM Journal on Optimization*, 20(3):1133–1156, 2009.
- [55] Michael Hintermüller und Karl Kunisch. Stationary optimal control problems with pointwise state constraints. In *Numerical PDE Constrained Optimization*, volume 72 of *Lecture Notes in Computational Science and Engineering*. Springer, Berlin Heidelberg, 2009.
- [56] Michael Hintermüller und Wolfgang Ring. A level set approach for the solution of a state constrained optimal control problem. *Numerische Mathematik*, 98:135–166, 2004.
- [57] Michael Hinze. A variational discretization concept in control constrained optimization: The linear-quadratic case. *Journal of Computational Optimization and Application*, 30(1):45–63, 2005.
- [58] Michael Hinze und Karl Kunisch. On suboptimal control strategies for the Navier-Stokes equations. *ESAIM: Proceedings*, 4:181–198, 1998.
- [59] Michael Hinze, René Pinnau, Michael Ulbrich und Stefan Ulbrich. *Optimization with PDE constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer, Dordrecht, NL, 2009.
- [60] Michael Hinze und Morten Vierling. Variational discretization and semi-smooth Newton methods; implementation, convergence and globalization in PDE constrained optimization with control constraints. *Hamburger Beiträge zur Angewandten Mathematik*, 15, 2009.
- [61] David H. Jacobson, M. M. Lele und Jason L. Speyer. New necessary conditions of optimality for control problems with state-variable inequality constraints. *Journal of Mathematical Analysis and Applications*, 35(2):255–284, 1971.
- [62] Peter Knabner und Lutz Angermann. *Numerical Methods for Elliptic and Parabolic Partial Differential Equations*, volume 44 of *Texts in Applied Mathematics*. Springer New York, 2003.
- [63] Hermann Kreim, Bernd Kugelmann, Hans Josef Pesch und Michael H. Breitner. Minimizing the maximum heating of a re-entering space shuttle: An optimal control problem with multiple control constraints. *Optimal Control Applications and Methods*, 17(1):45–69, 1996.
- [64] Karl Kunisch und Daniel Wachsmuth. Time optimal control of the wave equation, its regularization and numerical realization. *Submitted*.
- [65] Randall J. Leveque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002.
- [66] Helmut Maurer. *Optimale Steuerprozesse mit Zustandsbeschränkungen*. Habilitationsschrift, Universität Würzburg, 1976.
- [67] Helmut Maurer. Differential stability in optimal control problems. *Applied Mathematics & Optimization*, 5:283–295, 1979. 10.1007/BF01442559.

- [68] Helmut Maurer. On the minimum principle for optimal control problems with state constraints. *Schriftenreihe des Rechenzentrums der Universität Münster*, 41, 1979.
- [69] Kurt Meyberg und Peter Vachenauer. *Höhere Mathematik 2*. Springer, Berlin, 2006.
- [70] Christian Meyer, Uwe Prüfert und Fredi Tröltzsch. On two numerical methods for state-constrained elliptic control problems. *Optimization Methods and Software*, 22(6):871–899, 2007.
- [71] Christian Meyer und Arnd Rösch. Superconvergence properties of optimal control problems. *SIAM Journal on Control and Optimization*, 43(3):970–985, 2004.
- [72] Christian Meyer, Arnd Rösch und Fredi Tröltzsch. Optimal control problems of PDEs with regularized pointwise state constraints. *Computational Optimization and Applications*, 33(2-3):209–228, 2006.
- [73] Ira Neitzel und Fredi Tröltzsch. On regularization methods for the numerical solution of parabolic control problems with pointwise state constraints. *ESAIM: Control, Optimisation and Calculus of Variations*, 15(2):426–453, 2009.
- [74] Jorge Nocedal und Stephen J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.
- [75] Hans Josef Pesch. Private Kommunikation (2011). Universität Bayreuth.
- [76] Hans Josef Pesch. Offline and online computation of optimal trajectories in the aerospace field. In Attilio Miele, Angelo und Salvetti, editor, *Applied Mathematics in Aerospace Science and Engineering*, volume 44 of *Mathematical Concepts and Methods in Science and Engineering*, pages 165–220. Plenum Press, New York, 1994.
- [77] Hans Josef Pesch. *Schlüsseltechnologie Mathematik*. Stuttgart: B. G. Teubner, 2002.
- [78] Hans Josef Pesch. Vorlesungsskript „Variationsrechnung und optimale Steuerung“, Universität Bayreuth, 2003.
- [79] Hans Josef Pesch, Armin Rund, Wolf von Wahl und Stefan Wendl. On some new phenomena in state-constrained optimal control if ODE as well as PDE are involved. *Control and Cybernetics*, 39:647–660, 2010.
- [80] Hans Josef Pesch, Armin Rund, Wolf von Wahl und Stefan Wendl. On a prototype class of ODE-PDE state-constrained optimal control problem. Part 1: Analysis of the state-unconstrained problems. Preprint, Universität Bayreuth, 2011.
- [81] Hans Josef Pesch, Armin Rund und Stefan Wendl. On a prototype class of ODE-PDE state-constrained optimal control problem. Part 2: The state-constrained problems. In preparation.
- [82] Markus Probst, Michael Lulfesmann, Mike Nicolai, H. Martin Bückner, Marek Behr und Christian H. Bischof. Sensitivity of optimal shapes of artificial grafts with respect to flow parameters. *Computer Methods in Applied Mechanics and Engineering*, 199(17-20):997 – 1005, 2010.

- [83] Uwe Prüfert und Fredi Tröltzsch. An interior point method for a parabolic optimal control problem with regularized pointwise state constraints. *Zeitschrift für Angewandte Mathematik und Mechanik*, 87(8–9):564–589, 2007.
- [84] Uwe Prüfert, Fredi Tröltzsch und Martin Weiser. The convergence of an interior point method for an elliptic control problem with mixed control-state constraints. *Computational Optimization and Applications*, 39(2):183–218, 2008.
- [85] Rolf Rannacher. Finite element solution of diffusion problems with irregular data. *Numerische Mathematik*, 43:309–327, 1984.
- [86] Jean Pierre Raymond und Housnaa Zidani. Pontryagin’s principle for state-constrained control problems governed by parabolic equations with unbounded controls. *SIAM Journal on Control and Optimization*, 36(6):1853–1879, 1998.
- [87] Jean Pierre Raymond und Housnaa Zidani. Hamiltonian Pontryagin’s principles for control problems governed by semilinear parabolic equations. *Applied Mathematics & Optimization*, 39:143–177, 1999.
- [88] Michael Renardy und Robert C. Rogers. *An Introduction to Partial Differential Equations*, volume 13 of *Texts in Applied Mathematics*. Springer, New York, 2001.
- [89] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill Book Company, New York, 1974.
- [90] Armin Rund und Kurt Chudej. Optimal Control for a Simplified 1D Fuel Cell Model. *Mathematical and Computer Modelling of Dynamical Systems*, accepted for publication, 2011.
- [91] Ekkehard Sachs. Private Kommunikation (2011). Universität Trier.
- [92] Anton Schiela. Barrier methods for optimal control problems with state constraints. *SIAM Journal on Optimization*, 20(2):1002–1031, 2009.
- [93] Kati Sternberg. *Simulation, Optimale Steuerung und numerische Sensitivitätsanalyse einer Schmelzkarbonat-Brennstoffzelle mithilfe eines partiell differential-algebraischen Gleichungssystems*. PhD thesis, Fakultät für Mathematik, Physik und Informatik, Universität Bayreuth, 2007.
- [94] Kai Sundmacher, Achim Kienle, Hans Josef Pesch, Joachim F. Berndt und Gerhard Huppmann, editors. *Molten Carbonate Fuel Cells: Modeling, Analysis, Simulation, and Control*. WILEY-VCH Verlag, Weinheim, 2007.
- [95] Fredi Tröltzsch. *Optimality Conditions for Parabolic Control Problems and Applications*, volume 62 of *Teubner Texte zur Mathematik*. Teubner, Leipzig, 1984.
- [96] Fredi Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen*. Vieweg+Teubner, Wiesbaden, 2009.
- [97] Fredi Tröltzsch und Daniel Wachsmuth. On convergence of a receding horizon method for parabolic boundary control. *Optimization Methods and Software*, 19(2):201–216, 2004.

- [98] Stefan Ulbrich. Generalized SQP methods with parareal-time-domain decomposition for time-dependent PDE-constrained optimization. In Lorenz T. Biegler, Omar Ghattas, Matthias Heinkenschloss, David Keyes und Bart van Bloemen Waanders, editors, *Real-Time PDE-Constrained Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2007.
- [99] Boris Vexler und Karl Kunisch. Constrained Dirichlet boundary control in  $L^2$  for a class of evolution equations. *SIAM J. Control and Optimization*, 46:1726–1753, 2007.
- [100] Andreas Wächter und Carl Laird. Software package IPOPT. <https://projects.coin-or.org/Ipopt>.
- [101] Stefan Wendl, Hans Josef Pesch und Armin Rund. On a state-constrained PDE optimal control problem arising from ODE-PDE optimal control. In Moritz Diehl, Francois Glineur, Elias Jarlebring und Wim Michiels, editors, *Recent Advances in Optimization and its Applications in Engineering*, pages 429–438. Springer, Berlin, 2010.
- [102] Winnifried Wollner. Private Kommunikation (2011). Universität Hamburg.
- [103] J. Carsten Ziems und Stefan Ulbrich. Adaptive multilevel inexact SQP methods for PDE-constrained optimization. *SIAM J. on Optimization*, 21(1):1–40, 2011.

# Anhang C

## Publikationen

### Liste der einbezogenen Veröffentlichungen dieser Arbeit:

- M. Frey, S. Bechmann, A. Rund, H.J. Pesch:** „New Necessary Conditions for Distributed Optimal Control Problems of Linear Elliptic Equations with State Constraints“. Preprint, Universität Bayreuth, 2011.
- S. Wendl, H.J. Pesch, A. Rund:** „On a State-Constrained PDE Optimal Control Problem arising from ODE-PDE Optimal Control“. In: M. Diehl, F. Glineur, W. Michiels (Eds.): *Recent Advances in Optimization and its Applications in Engineering*. Springer, Heidelberg, pp. 429-438, 2010.
- H.J. Pesch, A. Rund, W. von Wahl, S. Wendl:** „On Some New Phenomena in State-constrained Optimal Control if ODE as well as PDE are Involved“. *Control and Cybernetics*, 39, 3, pp. 647-660, 2010.
- H.J. Pesch, A. Rund, W. von Wahl, S. Wendl:** „On a Prototype Class of ODE-PDE State-constrained Optimal Control Problems. Part 1: Analysis of the State-unconstrained Problems“. Preprint, Universität Bayreuth, 2011.
- A. Rund, K. Chudej:** „Optimal Control for a simplified 1D Fuel Cell Model“. *Mathematical and Computer Modelling of Dynamical Systems*, published online 2011, in print 2012.

### Weitere Veröffentlichungen, die nicht Teil dieser Arbeit sind:

#### Beiträge in Zeitschriften

- K. Sternberg, K. Chudej, H.J. Pesch, A. Rund:** „Parametric Sensitivity Analysis of Fast Load Changes of a Dynamic MCFC Model“. *Journal of Fuel Cell Science and Technology*, 5(2), 021002 (6 pages), 2008.
- M. Scholle, A. Rund, N. Aksel:** „Drag Reduction and Improvement of Material Transport in Creeping Films“. *Archieve of Applied Mechanics* 75, 2-3, pp. 93-112, Springer, 2006.
- A. Rund:** „Fast Parallel Matrix Multiplication: The Strassen-Winograd Algorithm“. *Beiträge zum Wissenschaftlichen Rechnen: Ergebnisse des Gaststudentenprogramms 2004 des John von Neumann-Instituts für Computing*, Rüdiger Esser (Ed.), pp. 53-64, Jülich, 2004.

**Beiträge in Proceedings**

**S. Bechmann, M. Frey, H.J. Pesch, A. Rund:** „On some new Adjoining Approaches for State-Constrained Elliptic Optimal Control Problems“. AIP Conference Proceedings 1168/1, pp. 1322-1323, 2009.

**A. Rund, K. Sternberg, H.J. Pesch, K. Chudej:** „Optimal Control of a Large PDAE Molten Carbonate Fuel Cell Model“. Proceedings in Applied Mathematics and Mechanics 7, 2008.

**Beiträge in Bearbeitung**

**H.J. Pesch, A. Rund, S. Wendl:** „On a Prototype Class of ODE-PDE State-constrained Optimal Control Problems. Part 2: The State-constrained Problems“. In preparation.

**M. Frey, S. Bechmann, A. Rund:** „Numerical Solution of new Necessary Conditions for Distributed Optimal Control Problems of Linear Elliptic Equations with State Constraints“. In preparation.

**S. Kurz, A. Rund:** „Optimal Population Allocations for the Continuous Schelling Segregation Model“. In preparation.

**C.1**

**New Necessary Conditions for Distributed  
Optimal Control Problems of Linear Elliptic  
Equations with State Constraints**

Michael Frey, Simon Bechmann, Armin Rund und Hans Josef Pesch

Preprint, Universität Bayreuth, 2011



# New Necessary Conditions for Distributed Optimal Control Problems of Linear Elliptic Equations with State Constraints\*

Michael Frey<sup>†</sup>    Simon Bechmann<sup>†</sup>    Armin Rund<sup>†</sup>    Hans Josef Pesch<sup>†‡</sup>

## Abstract

This paper deals with the solution of linear-quadratic elliptic optimal control problems with pointwise state constraints. Based on two different reformulations of the state constrained optimal control problem and a hypothesis on the structure of the active set new necessary conditions are obtained with higher regularity of the multipliers associated with the state constraint including new jump and sign conditions. Measures are no longer an issue. These two approaches mimic the well-known Bryson-Denham-Dreyfus indirect adjoining method which is the preferred ansatz in solving state constrained optimal control problems with ordinary differential equations numerically.

Mathematically the reformulations lead to a new kind of set optimal control problem, where the active set of the state constraint, resp. the interface between the inactive and the active set are to be determined as part of the solution. For this purpose, optimal control theory and shape optimization have to be combined. Existence, uniqueness and regularity results constitute a major part of the investigations, in order to guarantee the solvability of different intermediate steps. In the end, the new necessary conditions can be formulated as a free boundary value problem.

Numerical results would be beyond the scope of this paper and will be published subsequently.

**Key words.** Elliptic optimal control problem, state constraint, necessary conditions, set optimal control problem, shape optimization, bilevel optimization, free boundary value problem.

**AMS subject classifications.** 49J20, 49K20, 49N10, 49Q10, 35J57

## 1 Introduction

The consideration of pointwise state inequality constraints is an essential tool for the modeling of realistic processes which are to be optimized according to a given performance index: optimization is usually the second step after their simulation. Such processes can be governed, for example, by ordinary or partial differential equations, in short ODE or PDE. This leads to optimal control problems with constraints in form of ODE, denoted henceforth by OC-ODE, resp. optimal control problems with PDE, denoted by OC-PDE.

---

\*The first author is supported by the German National Science Foundation under Grant No. PE 365/12-1.

<sup>†</sup>University of Bayreuth, Chair of Mathematics in Engineering Sciences, 95440 Bayreuth, Germany

<sup>‡</sup>Corresponding author: [hans-josef.pesch@uni-bayreuth.de](mailto:hans-josef.pesch@uni-bayreuth.de).

In the ODE case, state constraints can provide for the observance, for example, of dynamic pressure constraints in flight path optimization, of temperature constraints in chemical processes, or of collision avoidance in robotics. In the PDE case, temperature constraints are an issue as well if heat conduction is involved or turbulence must be restricted in flow control. References [9, 10] and [28] give examples of multiply constrained optimal control problems in flight path optimization, resp. microeconomics, and [18, 17] of cooling processes in steel production, or [15] of flow control problems governed by the Navier Stokes equation. In [14], an extremely complicated, large scale optimal control model for a coupled system of partial integro-differential-algebraic equations of dimension 28 is described in detail to simulate fast load changes for a certain type of fuel cells. Here a constraint on the temperature or even better on the gradient of the temperature would be utmost desirable to reduce the risk of material corrosion, but the model's complexity prevents so far any progress in this direction.

While state constrained OC-ODE are rather well understood, state constrained OC-PDE are a highly topical research area. Excellent surveys for OC-ODE, resp. OC-PDE can be found in Hartl, Sethi, Vickson [20], resp. in Hintermüller, Kunisch [22].

The major steps in investigating state constraints in optimal control problems of ODE trace back to the early papers of Bryson, Denham, Dreyfus [8] and Jacobson, Lele, Speyer [25]. Both approaches head for so-called indirect methods, today preferably denoted by *first optimize, then discretize*. Bryson, Denham, and Dreyfus replaced the state constraint by interior boundary conditions and simultaneously treated a certain derivative of the state constraint like a mixed control-state constraint (BDD-ansatz). Jacobson, Lele, and Speyer pursued the so-called direct adjoining approach (of the constraining function to the Hamiltonian), which corresponds to the common approach in OC-PDE; see Tröltzsch [33]. In 1976, Maurer succeeded in amalgamating those two approaches in his habilitation [26]; later published in [27]. His investigations revealed that the multipliers associated with the different total time derivatives of the constraining function up to a certain order, by which the Hamiltonian is augmented, become the regular, the higher the order of the derivative is. Hereby, the so-called order of the state constraint plays a crucial role and is responsible for the topology of the active set, i. e. the time instances where the state constraint is fulfilled with the equality sign.

In OC-ODE the Bryson-Denham-Dreyfus approach turned out to be superior in view of its numerical realization, cf. [9, 10, 20].

In OC-PDE the difficulties arise from the fact that the Lagrange multipliers associated with pointwise state constraints exist only as measures, in general; cf. Casas [11] or Tröltzsch [32]. This influences strongly both the analysis of optimality conditions and the numerics, where poor regularity often has an impact on the convergence behaviour of numerical methods, cf. Bergounioux et al. [4, 5], since classical (non-regularized) active set methods require an almost everywhere evaluation of the multiplier in order to estimate the active set.

Typical analytical techniques in obtaining first order optimality conditions are either based on Lagrange multiplier rules, see again [11], [32] (for elliptic as well as for parabolic problems), or on maximum principles, see Casas [12], Alibert and Raymond [1] (for elliptic problems) and Casas [13], Raymond and Zidani [30] (for parabolic problems). Recent textbooks are due to Tröltzsch [33] with the emphasis on the analysis and basic numerical ideas, and Hinze, Pinnau, M. Ulbrich, S. Ulbrich [24] with the emphasis on the algorithmic and numerical treatment on the infinite dimensional level. In addition, the survey paper of Herzog and Kunisch [21] on numerical methods for OC-PDE shall be mentioned here, too. A comprehensive survey on state constrained elliptic OC-PDE

including current regularization techniques and some numerical comparison tests can be found in [3].

The intention of the present paper is to build a bridge between OC-ODE and OC-PDE. Especially, we aim at the application of the Bryson-Denham-Dreyfus approach in PDE constrained optimal control in order to improve the regularity associated with the multipliers of state constraints. At first we state the model problem and summarize some well-known results in section 2. Based on an adapted splitting in sect. 3, we reformulate the state constraint analog to Bryson, Denham, Dreyfus and proof new necessary conditions in section 4 and 5. A second variant of the BDD-type ansatz in section 6 yields more regular multipliers. The approaches lead to a new kind of optimization problem in section 7, which we call set optimal control problem, where topology and shape optimization as well as optimal control merge. Our approach has some similarity with the approach of Hintermüller and Ring [23], who turn the OC-PDE into a shape optimization problem.

## 2 Problem formulation and background

We consider the following linear-quadratic optimal control problem with pointwise state constraints and distributed control as a driving example

$$\text{minimize} \quad J(y, u) := \frac{1}{2} \|y - y_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2} \|u - u_d\|_{L^2(\Omega)}^2 \quad (2.1a)$$

$$\text{subject to} \quad -\Delta y + y = u, \quad \text{a. e. in } \Omega, \quad (2.1b)$$

$$\partial_\nu y = 0, \quad \text{a. e. on } \Gamma, \quad (2.1c)$$

$$G(y) := y - y_{\max} \leq 0, \quad \text{a. e. in } \Omega, \quad (2.1d)$$

where  $\Omega \subset \mathbb{R}^2$  is a bounded domain with boundary  $\Gamma$  of class  $C^{1,1}$  and where  $\nu$  denotes the outward unit normal. Further, we require  $y_d \in H^1(\Omega)$ ,  $y_{\max} \in H^4(\Omega)$ ,  $u_d \in H^2(\Omega)$ ,  $\lambda > 0$ , and  $u \in L^2(\Omega)$ . During the course of the paper we will see where and why some of the data of the problem need to have higher regularity. Note that this usually should not be an issue in practical applications except the restricting  $C^{1,1}$ -smoothness of the boundary  $\Gamma$ .

Due to [19, Thm. 2.2.2.5], we know that the elliptic boundary-value problem possesses a unique weak solution in  $H^2(\Omega)$  for all controls  $u \in L^2(\Omega)$ . In addition, the optimal control problem admits a unique solution  $(\bar{y}, \bar{u})$  due to [11].

We introduce the active, resp. inactive set associated with the state constraint (2.1d)

$$\mathcal{A} := \{x \in \bar{\Omega} : G(\bar{y}) = 0\}, \quad (2.2)$$

$$\mathcal{I} := \Omega \setminus \mathcal{A}, \quad (2.3)$$

$$\gamma := \partial\mathcal{A}, \quad (2.4)$$

and require for the remaining part of this paper that (cf. Bergounioux and Kunisch [6]):

**Assumptions 2.1** *The active set  $\mathcal{A}$  fulfills*

$$\mathcal{A} = \bigcup_{i=1}^l \mathcal{A}_i, \quad \bar{\bar{\mathcal{A}}}_i = \bar{\mathcal{A}}_i, \quad \mathcal{A} \cap \Gamma = \emptyset,$$

$$\mathcal{A}_i \cap \mathcal{A}_j = \emptyset, \quad i \neq j, \quad i, j \in \{1, \dots, l\},$$

*$\mathcal{A}_i$  is connected with a  $C^{1,1}$ -boundary for each  $i$ ,*

where  $\overset{\circ}{\mathcal{B}}$  denotes the interior of a set  $\mathcal{B} \subset \mathbb{R}^2$  and  $\bar{\mathcal{B}}$  its closure.

These assumptions exclude active sets of zero measure, except the empty set, and avoid degeneracy phenomena such as appendices of non-zero measure attached to active components (compartments)  $\mathcal{A}_i$ . The authors are aware that active sets of zero measure may play a role, (cf. [11], [6]), but have to leave this question open here. In addition, the  $C^{1,1}$ -regularity of the boundary of the active set is not guaranteed.

In the following, we present some well-known first order necessary conditions, which reveal some insight to the structure of the multipliers for state constraints. Since  $y_{\max} \in H^4(\Omega) \hookrightarrow C^0(\bar{\Omega})$ , we have an explicit Slater point

$$(\hat{y}, \hat{u}) := (\min_{x \in \bar{\Omega}}(y_{\max}(x)) - 1, \min_{x \in \bar{\Omega}}(y_{\max}(x)) - 1).$$

Note that  $y_{\max} \in H^2(\Omega)$  would be sufficient here, but the  $H^4$ -regularity has to be required later on. Because of the existence of a Slater point,  $(\bar{y}, \bar{u})$  fulfills the following first order necessary conditions (cf. the quite similar result [6, Thm. 2]):

**Theorem 2.2** *Under Assumptions 2.1 let the pair  $(\bar{y}, \bar{u})$  be optimal for problem (2.1). Then there exist a real regular Borel measure  $\mu \in \mathcal{M}(\bar{\Omega})$  and an associated adjoint state  $\wp \in H^1(\Omega)$ , such that the optimality system (Karush-Kuhn-Tucker system) is given by (2.1b), (2.1c), (2.1d) and the following equations (2.5),*

$$-\Delta \wp_{\mathcal{I}} + \wp_{\mathcal{I}} = \bar{y}_{\mathcal{I}} - y_d, \quad a. e. \text{ in } \mathcal{I}, \quad (2.5a)$$

$$\partial_{\nu} \wp_{\mathcal{I}} = 0, \quad a. e. \text{ on } \Gamma, \quad (2.5b)$$

$$-\Delta \wp_{\mathcal{A}} + \wp_{\mathcal{A}} = \bar{y}_{\mathcal{A}} - y_d + \mu_{\mathring{\mathcal{A}}}, \quad a. e. \text{ in } \mathring{\mathcal{A}}, \quad (2.5c)$$

$$\wp_{\mathcal{I}} = \wp_{\mathcal{A}}, \quad a. e. \text{ on } \gamma, \quad (2.5d)$$

$$\partial_{\nu_{\mathcal{I}}} \wp_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} \wp_{\mathcal{A}} = \mu_{\gamma}, \quad a. e. \text{ on } \gamma, \quad (2.5e)$$

$$\lambda(\bar{u}_{\mathcal{I}} - u_d) + \wp_{\mathcal{I}} = 0, \quad a. e. \text{ in } \mathcal{I}, \quad (2.5f)$$

$$\lambda(\bar{u}_{\mathcal{A}} - u_d) + \wp_{\mathcal{A}} = 0, \quad a. e. \text{ in } \mathring{\mathcal{A}}, \quad (2.5g)$$

where  $\nu_{\mathcal{I}}, \nu_{\mathcal{A}}$  are the outward unit normals of  $\mathcal{I}$  and  $\mathcal{A}$ , respectively and where the subscripts  $(\cdot)_{\mathcal{I}}$  and  $(\cdot)_{\mathcal{A}}$  denote the restrictions on  $\mathcal{I}, \mathcal{A}$ . In addition, there holds  $(\wp_{\mathcal{I}}, \wp_{\mathcal{A}}) \in H^2(\mathcal{I}) \times H^2(\mathring{\mathcal{A}})$  and

$$\mu = \mu_{\mathcal{I}} + \mu_{\mathring{\mathcal{A}}} + \mu_{\gamma}, \quad \mu_{\mathcal{I}} = 0, \quad (2.5h)$$

$$\mu_{\mathring{\mathcal{A}}} = \lambda(-\Delta^2 y_{\max} + 2\Delta y_{\max} - y_{\max} - \Delta u_d + u_d) - y_{\max} + y_d \geq 0 \text{ in } L^2(\mathring{\mathcal{A}}), \quad (2.5i)$$

$$\mu_{\gamma} = \partial_{\nu_{\mathcal{I}}} \wp_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} \wp_{\mathcal{A}} \geq 0 \text{ in } H^{\frac{1}{2}}(\gamma). \quad (2.5j)$$

**Remark 2.3** Note that the higher regularity of the data  $u_d$  and  $y_{\max}$  comes into play here.

Compared to OC-ODE, the previous theorem corresponds to the Jacobson-Lele-Speyer approach. We point out that although  $\mu_{\gamma}$  and  $\mu_{\mathring{\mathcal{A}}}$  are regular on the sets  $\gamma$  and  $\mathring{\mathcal{A}}$ , the composition  $\mu$  is still a regular Borel measure on  $\Omega$ .

### 3 Reformulation adapted to the splitting of the constraints

A basic idea in this paper is to split the constraints adapted to inactive and active set. As a start, we split the boundary value problem (2.1b), (2.1c) equivalently in (cf. [7, Prop. 4.1], [2, Cor. 7.1-5])

$$\begin{aligned}
-\Delta y_{\mathcal{I}} + y_{\mathcal{I}} &= u_{\mathcal{I}}, & \text{a. e. in } \mathcal{I}, \\
-\Delta y_{\mathcal{A}} + y_{\mathcal{A}} &= u_{\mathcal{A}}, & \text{a. e. in } \mathring{\mathcal{A}}, \\
\partial_{\nu} y_{\mathcal{I}} &= 0, & \text{a. e. on } \Gamma, \\
y_{\mathcal{I}} &= y_{\mathcal{A}}, & \text{a. e. on } \gamma, \\
\partial_{\nu_{\mathcal{I}}} y_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} y_{\mathcal{A}} &= 0, & \text{a. e. on } \gamma,
\end{aligned}$$

with  $y_{\mathcal{I}} \in H^1(\mathcal{I}, \Delta) := \{y \in H^1(\mathcal{I}) \mid \Delta y \in L^2(\mathcal{I})\}$ ,  $y_{\mathcal{A}} \in H^1(\mathring{\mathcal{A}}, \Delta)$ . For properties of the space  $H^1(\mathring{\mathcal{A}}, \Delta)$  cf. [2], especially Sec. 6.1-3 and 7.1-9. Since the boundaries  $\gamma$  and  $\Gamma$  are sufficiently regular, we even get  $y_{\mathcal{I}} \in H^2(\mathcal{I})$ ,  $y_{\mathcal{A}} \in H^2(\mathring{\mathcal{A}})$  (cf. [19, Thm. 2.2.2.5]).

Additionally we replace the state constraint (2.1d) by an equation in the active set and a strict inequality in the inactive set. These observations lead to the following split model problem with interface conditions for given  $\mathcal{A}$ ,  $\mathcal{I}$ , and  $\gamma$ :

$$\begin{aligned}
\text{minimize } \mathcal{J}(y_{\mathcal{I}}, y_{\mathcal{A}}, u_{\mathcal{I}}, u_{\mathcal{A}}) &:= \frac{1}{2} \|y_{\mathcal{I}} - y_d\|_{L^2(\mathcal{I})}^2 + \frac{1}{2} \|y_{\mathcal{A}} - y_d\|_{L^2(\mathring{\mathcal{A}})}^2 \\
&\quad + \frac{\lambda}{2} \|u_{\mathcal{I}} - u_d\|_{L^2(\mathcal{I})}^2 + \frac{\lambda}{2} \|u_{\mathcal{A}} - u_d\|_{L^2(\mathring{\mathcal{A}})}^2
\end{aligned} \tag{3.1a}$$

$$\text{subject to } \quad -\Delta y_{\mathcal{I}} + y_{\mathcal{I}} = u_{\mathcal{I}}, \quad \text{a. e. in } \mathcal{I}, \tag{3.1b}$$

$$-\Delta y_{\mathcal{A}} + y_{\mathcal{A}} = u_{\mathcal{A}}, \quad \text{a. e. in } \mathring{\mathcal{A}}, \tag{3.1c}$$

$$\partial_{\nu} y_{\mathcal{I}} = 0, \quad \text{a. e. on } \Gamma, \tag{3.1d}$$

$$y_{\mathcal{I}} = y_{\mathcal{A}}, \quad \text{a. e. on } \gamma, \tag{3.1e}$$

$$\partial_{\nu_{\mathcal{I}}} y_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} y_{\mathcal{A}} = 0, \quad \text{a. e. on } \gamma, \tag{3.1f}$$

$$y_{\mathcal{A}} = y_{\max}, \quad \text{a. e. in } \mathring{\mathcal{A}}, \tag{3.1g}$$

$$y_{\mathcal{A}} = y_{\max}, \quad \text{a. e. on } \gamma, \tag{3.1h}$$

$$u_{\mathcal{A}} \in L^2(\mathring{\mathcal{A}}), \quad u_{\mathcal{I}} \in L^2(\mathcal{I}), \tag{3.1i}$$

$$y_{\mathcal{A}} \in H^2(\mathring{\mathcal{A}}), \quad y_{\mathcal{I}} \in H^2(\mathcal{I}), \tag{3.1j}$$

$$y_{\mathcal{I}} < y_{\max}, \quad \text{a. e. in } \mathcal{I}, \tag{3.1k}$$

$$\mathcal{A}, \mathcal{I}, \text{ and } \gamma \text{ are defined by (2.2)–(2.4)}. \tag{3.1l}$$

*Remark 3.1* Although the results in Theorem 2.2 suggest such a splitted reformulation of the model problem, it seems that it has not yet been introduced.

In the following, we propose a reformulation of the state constraint that lifts the regularity of the multiplier component  $\mu_{\mathcal{A}}$  on  $\mathring{\mathcal{A}}$  from  $L^2(\mathring{\mathcal{A}})$  to  $H^2(\mathring{\mathcal{A}})$ ; this allows for an appropriate discretization of the multiplier with standard finite elements. In a second variant we will also get rid of the  $\gamma$ -component of the multiplier. This gives new necessary conditions without any measures but piecewise  $H^2$ -multipliers.

#### 4 Reformulation of the optimal control problem: 1st variant

Now, we are going to apply the Bryson-Denham-Dreyfus approach to the above state constrained optimal control problem. The basic idea is to replace the state constraint in the active set  $y_A = y_{\max}$  on  $\mathring{A} \cup \gamma$ , cf. (3.1g), (3.1h), by an equivalent boundary value problem and afterwards to rewrite the result by means of the state equation.

We have  $G(\bar{y}_A) \equiv 0$ . Due to the regularity of  $\mathcal{A}$  this condition can be formulated equivalently as a boundary value problem by means of the elliptic differential operator of the state equation

$$-\Delta G(\bar{y}_A) + G(\bar{y}_A) = -\Delta \bar{y}_A + \bar{y}_A + \Delta y_{\max} - y_{\max} = 0, \quad \text{a. e. in } \mathring{A}, \quad (4.1a)$$

$$G(\bar{y}_A) = \bar{y}_A - y_{\max} = 0, \quad \text{a. e. on } \gamma. \quad (4.1b)$$

Using the state equation, one obtains

$$\Delta y_{\max} - y_{\max} + \bar{u}_A = 0, \quad \text{a. e. in } \mathring{A},$$

$$\bar{y}_A - y_{\max} = 0, \quad \text{a. e. on } \gamma.$$

Since the differential operator, which is used to achieve an explicit dependence on the control  $u$ , is of second order, the procedure of transferring the state constraint to an equivalent equality constraint on the subdomain  $\mathring{A}$  of  $\Omega$  is analog to that for a boundary arc of a second-order state constraint in OC-ODE; cf. [8, 20, 27].

This consideration gives rise to the following problem with interface conditions:

$$\begin{aligned} \text{minimize } \mathcal{J}(y_{\mathcal{I}}, y_A, u_{\mathcal{I}}, u_A) := & \frac{1}{2} \|y_{\mathcal{I}} - y_d\|_{L^2(\mathcal{I})}^2 + \frac{1}{2} \|y_A - y_d\|_{L^2(\mathring{A})}^2 \\ & + \frac{\lambda}{2} \|u_{\mathcal{I}} - u_d\|_{L^2(\mathcal{I})}^2 + \frac{\lambda}{2} \|u_A - u_d\|_{L^2(\mathring{A})}^2 \end{aligned} \quad (4.2a)$$

$$\text{subject to } \quad -\Delta y_{\mathcal{I}} + y_{\mathcal{I}} = u_{\mathcal{I}}, \quad \text{a. e. in } \mathcal{I}, \quad (4.2b)$$

$$-\Delta y_A + y_A = u_A, \quad \text{a. e. in } \mathring{A}, \quad (4.2c)$$

$$\partial_\nu y_{\mathcal{I}} = 0, \quad \text{a. e. on } \Gamma, \quad (4.2d)$$

$$y_{\mathcal{I}} = y_A, \quad \text{a. e. on } \gamma, \quad (4.2e)$$

$$\partial_{\nu_{\mathcal{I}}} y_{\mathcal{I}} + \partial_{\nu_A} y_A = 0, \quad \text{a. e. on } \gamma, \quad (4.2f)$$

$$u_A = -\Delta y_{\max} + y_{\max}, \quad \text{a. e. in } \mathring{A}, \quad (4.2g)$$

$$y_A = y_{\max}, \quad \text{a. e. on } \gamma, \quad (4.2h)$$

$$u_A \in L^2(\mathring{A}), \quad u_{\mathcal{I}} \in L^2(\mathcal{I}), \quad (4.2i)$$

$$y_A \in H^2(\mathring{A}), \quad y_{\mathcal{I}} \in H^2(\mathcal{I}), \quad (4.2j)$$

$$y_{\mathcal{I}} < y_{\max}, \quad \text{a. e. in } \mathcal{I}, \quad (4.2k)$$

$$\mathcal{A}, \mathcal{I}, \text{ and } \gamma \text{ are defined by (2.2)–(2.4)}. \quad (4.2l)$$

**Theorem 4.1** *Let Assumptions 2.1 be fulfilled. Then the original model problem (2.1) and the reformulation (4.2) possess the same unique optimal solution.*

*Proof.* The assertion follows directly from the more general proof of Theorem 7.1.

□

**Remark 4.2** By definition of  $\mathcal{I}$ , the inequality  $y_{\mathcal{I}} < y_{\max}$  can be regarded as an inactive inequality constraint  $y_{\mathcal{I}} \leq y_{\max}$ . Consequently, it could be omitted without influencing the optimal solution. However, as will be seen later on, it is beneficial to drag it along.

## 5 First order necessary conditions

This section is devoted to the derivation of first order necessary conditions for the reformulated model problem (4.2).

**Theorem 5.1** *Let  $(\bar{y}_T, \bar{y}_A, \bar{u}_T, \bar{u}_A)$  be the optimal solution to problem (4.2). Then there exist multipliers  $q \in L^2(\mathring{A})$ ,  $\sigma \in H^{-\frac{3}{2}}(\gamma)$ , and functions  $p_T \in L^2(\mathcal{I}, \Delta) := \{\varphi \in L^2(\mathcal{I}) \mid \Delta\varphi \in L^2(\mathcal{I})\}$ ,  $p_A \in L^2(\mathring{A}, \Delta)$ , such that*

$$-\Delta p_T + p_T = \bar{y}_T - y_d, \quad \text{a. e. in } \mathcal{I}, \quad (5.1a)$$

$$\partial_\nu p_T = 0, \quad \text{a. e. on } \Gamma, \quad (5.1b)$$

$$-\Delta p_A + p_A = \bar{y}_A - y_d, \quad \text{a. e. in } \mathring{A}, \quad (5.1c)$$

$$p_T = p_A, \quad \text{a. e. on } \gamma, \quad (5.1d)$$

$$\partial_{\nu_T} p_T = \sigma - \partial_{\nu_A} p_A, \quad \text{a. e. on } \gamma, \quad (5.1e)$$

$$\lambda(\bar{u}_T - u_d) + p_T = 0, \quad \text{a. e. in } \mathcal{I}, \quad (5.1f)$$

$$\lambda(\bar{u}_A - u_d) + p_A + q = 0, \quad \text{a. e. in } \mathring{A}. \quad (5.1g)$$

Note that due to the low regularity of  $\sigma$  the equations defining  $p_T$  and  $p_A$  only make sense for  $p_T \in L^2(\mathcal{I}, \Delta)$  and  $p_A \in L^2(\mathring{A}, \Delta)$ , and for  $\Gamma \in C^{1,1}$ . For properties of the space  $L^2(\mathring{A}, \Delta)$  cf. [2, Sec. 7.1-9]. We will prove higher regularity in Prop. 5.2.

*Proof.* Consider the linear, continuous (cf. [19, Thm. 2.3.3.2]) control-to-state operator  $S = (S_T, S_A) : L^2(\mathcal{I}) \times L^2(\mathring{A}) \rightarrow H^2(\mathcal{I}) \times H^2(\mathring{A})$ , which maps  $(u_T, u_A)$  to the solution  $(y_T, y_A)$  of (4.2b)–(4.2f), and the well-known trace operators

$$\tau_T : H^2(\mathcal{I}) \rightarrow H^{\frac{3}{2}}(\gamma),$$

$$\tau_A : H^2(\mathring{A}) \rightarrow H^{\frac{3}{2}}(\gamma).$$

Referring to Remark 4.2, we can omit the inequality (4.2k) and reduce the problem to

$$\begin{aligned} \text{minimize} \quad & f(u_T, u_A) := \mathcal{J}(S_T(u_T, u_A), S_A(u_T, u_A), u_T, u_A) \\ \text{subject to} \quad & \mathcal{G}(u_T, u_A) := \begin{pmatrix} \Delta y_{\max} - y_{\max} + u_A \\ \tau_A S_A(u_T, u_A) - \tau_A y_{\max} \end{pmatrix} = 0, \end{aligned} \quad (5.2)$$

where  $\mathcal{G} : L^2(\mathcal{I}) \times L^2(\mathring{A}) \rightarrow L^2(\mathring{A}) \times H^{\frac{3}{2}}(\gamma)$ . This reduced problem fits in the usual framework of nonlinear optimization in Banach spaces. Next, we will show that the Zowe-Kurcyusz constraint qualification is valid for the operator  $\mathcal{G}$  in  $(\bar{u}_T, \bar{u}_A)$ . Thus, for each arbitrary  $z_1 \in L^2(\mathring{A})$  and  $z_2 \in H^{\frac{3}{2}}(\gamma)$  we need to find  $(h_T, h_A) \in L^2(\mathcal{I}) \times L^2(\mathring{A})$  such that

$$\mathcal{G}'(h_T, h_A) = \begin{pmatrix} h_A \\ \tau_A S_A(h_T, h_A) \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

(Note that  $\mathcal{G}$  is continuously Fréchet differentiable, since  $S_A$  is continuous and affine.) As a start, this defines  $h_A := z_1$ . Now let  $(v_T^1, v_A^1) := S(0, h_A)$ . Consequently  $v_A^1 \in H^2(\mathring{A})$  and  $\tau_A v_A^1 \in H^{\frac{3}{2}}(\gamma)$ . Next, let  $v_A^2 \in H^2(\mathring{A})$  solve

$$\begin{aligned} -\Delta v_A^2 + v_A^2 &= 0, & \text{a. e. in } \mathring{A}, \\ \tau_A v_A^2 &= z_2 - \tau_A v_A^1, & \text{a. e. on } \gamma. \end{aligned}$$

Due to [19, Thm. 1.5.1.2] there exists  $v_{\mathcal{I}}^2 \in H^2(\mathcal{I})$  which suffices

$$\begin{aligned} \partial_\nu v_{\mathcal{I}}^2 &= 0, & \text{a. e. on } \Gamma, \\ \partial_{\nu_{\mathcal{I}}} v_{\mathcal{I}}^2 &= -\partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}}^2, & \text{a. e. on } \gamma, \\ \tau_{\mathcal{I}} v_{\mathcal{I}}^2 &= \tau_{\mathcal{A}} v_{\mathcal{A}}^2, & \text{a. e. on } \gamma. \end{aligned}$$

Defining  $h_{\mathcal{I}} := -\Delta v_{\mathcal{I}}^2 + v_{\mathcal{I}}^2 \in L^2(\mathcal{I})$ ,  $v_{\mathcal{I}} := v_{\mathcal{I}}^1 + v_{\mathcal{I}}^2$ , and  $v_{\mathcal{A}} := v_{\mathcal{A}}^1 + v_{\mathcal{A}}^2$ , it follows

$$\begin{aligned} -\Delta v_{\mathcal{I}} + v_{\mathcal{I}} &= -\Delta v_{\mathcal{I}}^1 + v_{\mathcal{I}}^1 - \Delta v_{\mathcal{I}}^2 + v_{\mathcal{I}}^2 = 0 + h_{\mathcal{I}}, & \text{a. e. in } \mathcal{I}, \\ -\Delta v_{\mathcal{A}} + v_{\mathcal{A}} &= -\Delta v_{\mathcal{A}}^1 + v_{\mathcal{A}}^1 - \Delta v_{\mathcal{A}}^2 + v_{\mathcal{A}}^2 = z_1 + 0, & \text{a. e. in } \mathring{\mathcal{A}}, \\ \partial_\nu v_{\mathcal{I}} &= \partial_\nu v_{\mathcal{I}}^1 + \partial_\nu v_{\mathcal{I}}^2 = 0 + 0, & \text{a. e. on } \Gamma, \\ \tau_{\mathcal{I}} v_{\mathcal{I}} - \tau_{\mathcal{A}} v_{\mathcal{A}} &= (\tau_{\mathcal{I}} v_{\mathcal{I}}^1 - \tau_{\mathcal{A}} v_{\mathcal{A}}^1) + (\tau_{\mathcal{I}} v_{\mathcal{I}}^2 - \tau_{\mathcal{A}} v_{\mathcal{A}}^2) = 0 + 0, & \text{a. e. on } \gamma, \\ \partial_{\nu_{\mathcal{I}}} v_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}} &= (\partial_{\nu_{\mathcal{I}}} v_{\mathcal{I}}^1 + \partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}}^1) + (\partial_{\nu_{\mathcal{I}}} v_{\mathcal{I}}^2 + \partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}}^2) = 0 + 0, & \text{a. e. on } \gamma, \end{aligned}$$

or in other words  $S(h_{\mathcal{I}}, z_1) = (v_{\mathcal{I}}, v_{\mathcal{A}})$ . Furthermore, there holds

$$\mathcal{G}'(h_{\mathcal{I}}, h_{\mathcal{A}}) = \begin{pmatrix} h_{\mathcal{A}} \\ \tau_{\mathcal{A}} S_{\mathcal{A}}(h_{\mathcal{I}}, h_{\mathcal{A}}) \end{pmatrix} = \begin{pmatrix} z_1 \\ \tau_{\mathcal{A}} v_{\mathcal{A}}^1 + (z_2 - \tau_{\mathcal{A}} v_{\mathcal{A}}^1) \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

Consequently, the Zowe-Kurcyusz constraint qualification is fulfilled and there exist Lagrange multipliers  $q \in L^2(\mathring{\mathcal{A}})$  and  $\sigma \in H^{\frac{3}{2}}(\gamma)^* = H^{-\frac{3}{2}}(\gamma)$ , so as to  $(\bar{u}_{\mathcal{I}}, \bar{u}_{\mathcal{A}}, q, \sigma)$  is a saddle point of the Lagrange function (cf. [33, Thm. 6.3])

$$\mathcal{L} : L^2(\mathcal{I}) \times L^2(\mathring{\mathcal{A}}) \times L^2(\mathring{\mathcal{A}}) \times H^{-\frac{3}{2}}(\gamma) \rightarrow \mathbb{R},$$

$$\mathcal{L}(u_{\mathcal{I}}, u_{\mathcal{A}}, q, \sigma) := f(u_{\mathcal{I}}, u_{\mathcal{A}}) + \int_{\mathring{\mathcal{A}}} q (u_{\mathcal{A}} + \Delta y_{\max} - y_{\max}) + \langle \langle \sigma, \tau_{\mathcal{A}} S_{\mathcal{A}}(u_{\mathcal{I}}, u_{\mathcal{A}}) - \tau_{\mathcal{A}} y_{\max} \rangle \rangle_{\gamma},$$

where  $\langle \langle \cdot, \cdot \rangle \rangle_M$  is the duality pairing between  $H^{-\frac{3}{2}}(M)$  and  $H^{\frac{3}{2}}(M)$  for  $M \in \{\gamma, \Gamma\}$ . Since the constraints (5.2) solely consist of equalities, it holds

$$\begin{aligned} 0 &= \partial_{u_{\mathcal{A}}} \mathcal{L}(\bar{u}_{\mathcal{I}}, \bar{u}_{\mathcal{A}}, q, \sigma) h \\ &= \int_{\mathcal{I}} (S_{\mathcal{I}}(\bar{u}_{\mathcal{I}}, \bar{u}_{\mathcal{A}}) - y_d) S_{\mathcal{I}}(0, h) + \int_{\mathring{\mathcal{A}}} (S_{\mathcal{A}}(\bar{u}_{\mathcal{I}}, \bar{u}_{\mathcal{A}}) - y_d) S_{\mathcal{A}}(0, h) \\ &\quad + \int_{\mathring{\mathcal{A}}} \lambda (\bar{u}_{\mathcal{A}} - u_d) h + \int_{\mathring{\mathcal{A}}} q h + \langle \langle \sigma, \tau_{\mathcal{A}} S_{\mathcal{A}}(0, h) \rangle \rangle_{\gamma}, \quad \forall h \in L^2(\mathring{\mathcal{A}}). \end{aligned}$$

Now let  $(p_{\mathcal{I}}, p_{\mathcal{A}}) \in L^2(\mathcal{I}, \Delta) \times L^2(\mathring{\mathcal{A}}, \Delta)$  be introduced as the solution to (5.1a)–(5.1e). Note that such functions are well defined, see, e. g., [2, Sec. 7.1-9] or [19, Sec. 1.5.3].



*Proof.* Because primal variables match, we find, by comparison of (5.1f) and (2.5f), and respectively of (5.1g) and (2.5g),

$$\begin{aligned} p_{\mathcal{I}} &= \wp_{\mathcal{I}} \in H^2(\mathcal{I}), \\ p_{\mathcal{A}} + q &= \wp_{\mathcal{A}} \in H^2(\mathring{\mathcal{A}}). \end{aligned} \tag{5.4}$$

By inserting (5.4) into (2.5c) and by comparing with (5.1c), one obtains

$$-\Delta q + q = \mu_{\mathring{\mathcal{A}}}, \text{ a. e. in } \mathring{\mathcal{A}}.$$

Similarly we get

$$\partial_{\nu_{\mathcal{A}}} q + \sigma = \mu_{\gamma}, \text{ a. e. on } \gamma$$

by equations (2.5e) and (5.1e). Since  $\wp$  and  $p$  are continuous across the interface  $\gamma$  due to (2.5d) and (5.1d), we have  $q = 0$  on  $\gamma$ . With  $\mu_{\mathring{\mathcal{A}}} \in L^2(\mathring{\mathcal{A}})$ , regularity theory for the boundary value problem (5.3c), (5.3d) yields  $q \in H^2(\mathring{\mathcal{A}})$  and, due to (5.4), one obtains  $p_{\mathcal{A}} \in H^2(\mathring{\mathcal{A}})$ . In addition, due to (5.1e) and  $H^2$ -regularity of  $p_{\mathcal{I}}$  and  $p_{\mathcal{A}}$ , there holds  $\sigma \in H^{\frac{1}{2}}(\gamma)$ . With  $\tau_{\mathcal{A}} q = 0$ ,  $\mu_{\mathring{\mathcal{A}}} \geq 0$ , and weak maximum principle we find  $q \geq 0$  and, in particular,  $\partial_{\nu_{\mathcal{A}}} q \leq 0$ ; hence  $0 \leq \mu_{\gamma} = \sigma + \partial_{\nu_{\mathcal{A}}} q \leq \sigma$ .  $\square$

**Remarks 5.3** Proposition 5.2 shows the direct connection between the traditional multipliers and the new ones.

- It should be emphasized that the multipliers  $q$  and  $\sigma$  inherit sign conditions from their counterparts  $\mu_{\mathring{\mathcal{A}}}$  and  $\mu_{\gamma}$ , although they belong to equality constraints.
- The improvement of regularity of  $q$  vs.  $\mu_{\mathring{\mathcal{A}}}$  is linked to the derivative of the equation  $G(\bar{y}_{\mathcal{A}}) \equiv 0$  in  $\mathring{\mathcal{A}}$ : The equation was differentiated twice (cf. (4.1a)) and the regularity improves from  $L^2(\mathring{\mathcal{A}})$  to  $H^2(\mathring{\mathcal{A}})$ . Analogously,  $\sigma$  does not gain any improvement in regularity compared to  $\mu_{\gamma}$ , which is due to the fact that the boundary condition (4.1b) does not contain differential operations. However, as we will see in the next section, improvements can be achieved by applying Neumann boundary conditions.
- The boundary condition  $q = 0$  on  $\gamma$  is not contained in the statement of Theorem 5.1 and is only obtained by comparison with the traditional optimality conditions. This gives an indication that the restriction of problem (4.2) to the a priori unknown active set  $\mathcal{A}$  inhibits the derivation of a condition of this type. In section 7, we will present an informal way how to obtain this condition.
- The lifting of the regularities can be obtained independently of the above comparison, in particular with Theorem 7.5 in section 7.
- Results, which are comparable to the first two items, are known from OC-ODE; compare Maurer [27, (5.9), (5.10)].

## 6 Reformulation of the optimal control problem: 2nd variant

Instead of the Dirichlet boundary value problem (4.1) we may also choose a Neumann boundary value problem as equivalent reformulation of the state constraint on the active set. In contrast to the previous BDD-ansatz, the boundary condition contains a differential operator. In particular, this leads to higher regularity of the corresponding multiplier. Starting with

$$\begin{aligned} -\Delta G(\bar{y}_A) + G(\bar{y}_A) &= -\Delta \bar{y}_A + \bar{y}_A + \Delta y_{\max} - y_{\max} = 0, & \text{a. e. in } \mathring{A}, \\ \partial_{\nu_A} G(\bar{y}_A) &= \partial_{\nu_A} \bar{y}_A - \partial_{\nu_A} y_{\max} = 0, & \text{a. e. on } \gamma, \end{aligned}$$

and using the state equation one obtains

$$\begin{aligned} \Delta y_{\max} - y_{\max} + \bar{u}_A &= 0, & \text{a. e. in } \mathring{A}, \\ \partial_{\nu_A} \bar{y}_A - \partial_{\nu_A} y_{\max} &= 0, & \text{a. e. on } \gamma. \end{aligned}$$

Consequently, as in section 4, we consider a reformulation of the optimal control problem (3.1), where we replace (3.1g), (3.1h) with

$$\Delta y_{\max} - y_{\max} + u_A = 0, \quad \text{a. e. in } \mathring{A}, \quad (6.1a)$$

$$\partial_{\nu_A} y_A - \partial_{\nu_A} y_{\max} = 0, \quad \text{a. e. on } \gamma. \quad (6.1b)$$

The existence and uniqueness result of Theorem 7.1 is valid in this situation, too. Thus, we can state new necessary conditions for the given reformulation.

**Theorem 6.1** *Let  $(\bar{y}_I, \bar{y}_A, \bar{u}_I, \bar{u}_A)$  be the optimal solution to problem (3.1), where (3.1g), (3.1h) is replaced with (6.1). Then there exist multipliers<sup>1</sup>  $q \in L^2(\mathring{A})$ ,  $\sigma \in H^{-\frac{1}{2}}(\gamma)$ , and functions  $p_I \in L^2(\mathcal{I}, \Delta)$ ,  $p_A \in L^2(\mathring{A}, \Delta)$ , such that*

$$-\Delta p_I + p_I = \bar{y}_I - y_d, \quad \text{a. e. in } \mathcal{I}, \quad (6.2a)$$

$$\partial_{\nu} p_I = 0, \quad \text{a. e. on } \Gamma, \quad (6.2b)$$

$$-\Delta p_A + p_A = \bar{y}_A - y_d, \quad \text{a. e. in } \mathring{A}, \quad (6.2c)$$

$$p_I = p_A + \sigma, \quad \text{a. e. on } \gamma, \quad (6.2d)$$

$$\partial_{\nu_I} p_I = -\partial_{\nu_A} p_A, \quad \text{a. e. on } \gamma, \quad (6.2e)$$

$$\lambda(\bar{u}_I - u_d) + p_I = 0, \quad \text{a. e. in } \mathcal{I}, \quad (6.2f)$$

$$\lambda(\bar{u}_A - u_d) + p_A + q = 0, \quad \text{a. e. in } \mathring{A}. \quad (6.2g)$$

*Proof.* The statement of Remark 4.2 is valid here, too. Thus, we omit the state constraint (3.1k) in the inactive set in the following. Using the continuous control-to-state operator  $S$  from the proof of the analog Theorem 5.1 and the continuous Neumann trace operator  $\partial_{\nu_A}: H^2(\mathring{A}) \rightarrow H^{\frac{1}{2}}(\gamma)$ , the considered optimal control problem becomes

$$\begin{aligned} \text{minimize} \quad f(u_I, u_A) &:= \frac{1}{2} \|S_I(u_I, u_A) - y_d\|_{L^2(\mathcal{I})}^2 + \frac{1}{2} \|S_A(u_I, u_A) - y_d\|_{L^2(\mathring{A})}^2 \\ &\quad + \frac{\lambda}{2} \|u_I - u_d\|_{L^2(\mathcal{I})}^2 + \frac{\lambda}{2} \|u_A - u_d\|_{L^2(\mathring{A})}^2 \\ \text{subject to} \quad \mathcal{G}(u_I, u_A) &:= \begin{pmatrix} \Delta y_{\max} - y_{\max} + u_A \\ \partial_{\nu_A} S_A(u_I, u_A) - \partial_{\nu_A} y_{\max} \end{pmatrix} = 0, \end{aligned} \quad (6.3)$$

<sup>1</sup>Although the multipliers differ from those of section 5, we omit an additional subscript for the sake of simplicity.

where  $\mathcal{G} : L^2(\mathcal{I}) \times L^2(\mathring{\mathcal{A}}) \rightarrow L^2(\mathring{\mathcal{A}}) \times H^{\frac{1}{2}}(\gamma)$ . The remaining proof follows the lines of the proof of Theorem 5.1, but is given for convenience. In order to validate the Zowe-Kurcyusz constraint qualification for the operator  $\mathcal{G}$ , we need to find  $(h_{\mathcal{I}}, h_{\mathcal{A}}) \in L^2(\mathcal{I}) \times L^2(\mathring{\mathcal{A}})$  such that

$$\mathcal{G}'(h_{\mathcal{I}}, h_{\mathcal{A}}) = \begin{pmatrix} h_{\mathcal{A}} \\ \partial_{\nu_{\mathcal{A}}} S_{\mathcal{A}}(h_{\mathcal{I}}, h_{\mathcal{A}}) \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},$$

where  $z_1 \in L^2(\mathring{\mathcal{A}})$ ,  $z_2 \in H^{\frac{1}{2}}(\gamma)$  are given arbitrarily. As a start, this defines  $h_{\mathcal{A}} := z_1$ . Now let  $(v_{\mathcal{I}}^1, v_{\mathcal{A}}^1) := S(0, h_{\mathcal{A}})$ . Consequently  $v_{\mathcal{A}}^1 \in H^2(\mathring{\mathcal{A}})$  and  $\partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}}^1 \in H^{\frac{1}{2}}(\gamma)$ . Next, let  $v_{\mathcal{A}}^2 \in H^2(\mathring{\mathcal{A}})$  solve

$$\begin{aligned} -\Delta v_{\mathcal{A}}^2 + v_{\mathcal{A}}^2 &= 0, & \text{a. e. in } \mathring{\mathcal{A}}, \\ \partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}}^2 &= z_2 - \partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}}^1, & \text{a. e. on } \gamma. \end{aligned}$$

Due to [19, Thm. 1.5.1.2] there exists  $v_{\mathcal{I}}^2 \in H^2(\mathcal{I})$  which suffices

$$\begin{aligned} \partial_{\nu} v_{\mathcal{I}}^2 &= 0, & \text{a. e. on } \Gamma, \\ \partial_{\nu_{\mathcal{I}}} v_{\mathcal{I}}^2 &= -\partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}}^2, & \text{a. e. on } \gamma, \\ \tau_{\mathcal{I}} v_{\mathcal{I}}^2 &= \tau_{\mathcal{A}} v_{\mathcal{A}}^2, & \text{a. e. on } \gamma. \end{aligned}$$

Defining  $h_{\mathcal{I}} := -\Delta v_{\mathcal{I}}^2 + v_{\mathcal{I}}^2 \in L^2(\mathcal{I})$ ,  $v_{\mathcal{I}} := v_{\mathcal{I}}^1 + v_{\mathcal{I}}^2$ , and  $v_{\mathcal{A}} := v_{\mathcal{A}}^1 + v_{\mathcal{A}}^2$ , it follows

$$\begin{aligned} -\Delta v_{\mathcal{I}} + v_{\mathcal{I}} &= -\Delta v_{\mathcal{I}}^1 + v_{\mathcal{I}}^1 - \Delta v_{\mathcal{I}}^2 + v_{\mathcal{I}}^2 = 0 + h_{\mathcal{I}}, & \text{a. e. in } \mathcal{I}, \\ -\Delta v_{\mathcal{A}} + v_{\mathcal{A}} &= -\Delta v_{\mathcal{A}}^1 + v_{\mathcal{A}}^1 - \Delta v_{\mathcal{A}}^2 + v_{\mathcal{A}}^2 = z_1 + 0, & \text{a. e. in } \mathring{\mathcal{A}}, \\ \partial_{\nu} v_{\mathcal{I}} &= \partial_{\nu} v_{\mathcal{I}}^1 + \partial_{\nu} v_{\mathcal{I}}^2 = 0 + 0, & \text{a. e. on } \Gamma, \\ \tau_{\mathcal{I}} v_{\mathcal{I}} - \tau_{\mathcal{A}} v_{\mathcal{A}} &= (\tau_{\mathcal{I}} v_{\mathcal{I}}^1 - \tau_{\mathcal{A}} v_{\mathcal{A}}^1) + (\tau_{\mathcal{I}} v_{\mathcal{I}}^2 - \tau_{\mathcal{A}} v_{\mathcal{A}}^2) = 0 + 0, & \text{a. e. on } \gamma, \\ \partial_{\nu_{\mathcal{I}}} v_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}} &= (\partial_{\nu_{\mathcal{I}}} v_{\mathcal{I}}^1 + \partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}}^1) + (\partial_{\nu_{\mathcal{I}}} v_{\mathcal{I}}^2 + \partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}}^2) = 0 + 0, & \text{a. e. on } \gamma, \end{aligned}$$

or in other words  $S(h_{\mathcal{I}}, z_1) = (v_{\mathcal{I}}, v_{\mathcal{A}})$ . Furthermore, there holds

$$\mathcal{G}'(h_{\mathcal{I}}, h_{\mathcal{A}}) = \begin{pmatrix} h_{\mathcal{A}} \\ \partial_{\nu_{\mathcal{A}}} S_{\mathcal{A}}(h_{\mathcal{I}}, h_{\mathcal{A}}) \end{pmatrix} = \begin{pmatrix} z_1 \\ \partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}}^1 + (z_2 - \partial_{\nu_{\mathcal{A}}} v_{\mathcal{A}}^1) \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

Consequently, the Zowe-Kurcyusz constraint qualification is fulfilled and there exist Lagrange multipliers  $q \in L^2(\mathring{\mathcal{A}})$  and  $\sigma \in H^{\frac{1}{2}}(\gamma)^* = H^{-\frac{1}{2}}(\gamma)$ , so as to  $(\bar{u}_{\mathcal{I}}, \bar{u}_{\mathcal{A}}, p, \sigma)$  is a saddle point of the Lagrange function (cf. [33, Thm. 6.3])

$$\mathcal{L} : L^2(\mathcal{I}) \times L^2(\mathring{\mathcal{A}}) \times L^2(\mathring{\mathcal{A}}) \times H^{-\frac{1}{2}}(\gamma) \rightarrow \mathbb{R},$$

$$\mathcal{L}(u_{\mathcal{I}}, u_{\mathcal{A}}, q, \sigma) := f(u_{\mathcal{I}}, u_{\mathcal{A}}) + \int_{\mathring{\mathcal{A}}} q(u_{\mathcal{A}} + \Delta y_{\max} - y_{\max}) + \left\langle \sigma, \partial_{\nu_{\mathcal{A}}} S_{\mathcal{A}}(u_{\mathcal{I}}, u_{\mathcal{A}}) - \partial_{\nu_{\mathcal{A}}} y_{\max} \right\rangle_{\gamma}.$$

Here again, we use the notation  $\langle \cdot, \cdot \rangle_M$ , resp.  $\langle\langle \cdot, \cdot \rangle\rangle_M$  for the duality pairings between  $H^{-\frac{1}{2}}(M)$  and  $H^{\frac{1}{2}}(M)$ , resp.  $H^{-\frac{3}{2}}(M)$  and  $H^{\frac{3}{2}}(M)$  for  $M \in \{\gamma, \Gamma\}$ . Since the constraints (6.3) solely consist of equalities, there holds

$$\begin{aligned} 0 &= \partial_{u_{\mathcal{A}}} \mathcal{L}(\bar{u}_{\mathcal{I}}, \bar{u}_{\mathcal{A}}, q, \sigma)h \\ &= \int_{\mathcal{I}} (S_{\mathcal{I}}(\bar{u}_{\mathcal{I}}, \bar{u}_{\mathcal{A}}) - y_d) S_{\mathcal{I}}(0, h) + \int_{\mathring{\mathcal{A}}} (S_{\mathcal{A}}(\bar{u}_{\mathcal{I}}, \bar{u}_{\mathcal{A}}) - y_d) S_{\mathcal{A}}(0, h) \\ &\quad + \int_{\mathring{\mathcal{A}}} \lambda(\bar{u}_{\mathcal{A}} - u_d)h + \int_{\mathring{\mathcal{A}}} qh + \left\langle \sigma, \partial_{\nu_{\mathcal{A}}} S_{\mathcal{A}}(0, h) \right\rangle_{\gamma}, \quad \forall h \in L^2(\mathring{\mathcal{A}}). \end{aligned}$$

Now let  $(p_{\mathcal{I}}, p_{\mathcal{A}}) \in L^2(\mathcal{I}, \Delta) \times L^2(\mathring{\mathcal{A}}, \Delta)$  be introduced as the solution to (6.2a)–(6.2e). Note that such functions are well defined, see, e. g., [2, Sec. 7.1-9] or [19, Sec. 1.5.3]. Using the suitable Green's formula (cf. [19, p. 57]) we can proceed

$$\begin{aligned}
0 &= \int_{\mathcal{I}} (-\Delta p_{\mathcal{I}} + p_{\mathcal{I}}) S_{\mathcal{I}}(0, h) + \int_{\mathring{\mathcal{A}}} (-\Delta p_{\mathcal{A}} + p_{\mathcal{A}}) S_{\mathcal{A}}(0, h) \\
&\quad + \int_{\mathring{\mathcal{A}}} \lambda (\bar{u}_{\mathcal{A}} - u_d) h + \int_{\mathring{\mathcal{A}}} q h + \langle \sigma, \partial_{\nu_{\mathcal{A}}} S_{\mathcal{A}}(0, h) \rangle_{\gamma} \\
&= \int_{\mathcal{I}} p_{\mathcal{I}} \underbrace{(-\Delta S_{\mathcal{I}}(0, h) + S_{\mathcal{I}}(0, h))}_{=0} + \int_{\mathring{\mathcal{A}}} p_{\mathcal{A}} \underbrace{(-\Delta S_{\mathcal{A}}(0, h) + S_{\mathcal{A}}(0, h))}_{=h} \\
&\quad + \int_{\mathring{\mathcal{A}}} \lambda (\bar{u}_{\mathcal{A}} - u_d) h + \int_{\mathring{\mathcal{A}}} q h + \langle \sigma, \partial_{\nu_{\mathcal{A}}} S_{\mathcal{A}}(0, h) \rangle_{\gamma} \\
&\quad - \langle \langle \underbrace{\partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}}}_{=-\partial_{\nu_{\mathcal{A}}} p_{\mathcal{A}}}, \underbrace{\tau_{\mathcal{I}} S_{\mathcal{I}}(0, h)}_{=\tau_{\mathcal{A}} S_{\mathcal{A}}(0, h)} \rangle \rangle_{\gamma} + \langle \underbrace{\tau_{\mathcal{I}} p_{\mathcal{I}}}_{=\tau_{\mathcal{A}} p_{\mathcal{A}} + \sigma}, \underbrace{\partial_{\nu_{\mathcal{I}}} S_{\mathcal{I}}(0, h)}_{=-\partial_{\nu_{\mathcal{A}}} S_{\mathcal{A}}(0, h)} \rangle_{\gamma} \\
&\quad - \langle \langle \partial_{\nu_{\mathcal{A}}} p_{\mathcal{A}}, \tau_{\mathcal{A}} S_{\mathcal{A}}(0, h) \rangle \rangle_{\gamma} + \langle \tau_{\mathcal{A}} p_{\mathcal{A}}, \partial_{\nu_{\mathcal{A}}} S_{\mathcal{A}}(0, h) \rangle_{\gamma} \\
&\quad - \langle \langle \underbrace{\partial_{\nu} p_{\mathcal{I}}}_{=0}, \tau S_{\mathcal{I}}(0, h) \rangle \rangle_{\Gamma} + \langle \tau p_{\mathcal{I}}, \underbrace{\partial_{\nu} S_{\mathcal{I}}(0, h)}_{=0} \rangle_{\Gamma} \\
&= \int_{\mathring{\mathcal{A}}} p_{\mathcal{A}} h + q h + \lambda (\bar{u}_{\mathcal{A}} - u_d) h, \quad \forall h \in L^2(\mathring{\mathcal{A}}).
\end{aligned}$$

Finally, we obtain  $\lambda (\bar{u}_{\mathcal{A}} - u_d) + p_{\mathcal{A}} + q = 0$  in  $L^2(\mathring{\mathcal{A}})$  and, analogously,  $\lambda (\bar{u}_{\mathcal{I}} - u_d) + p_{\mathcal{I}} = 0$  in  $L^2(\mathcal{I})$ , which completes the derivation of the claimed conditions.  $\square$

Next, we investigate the connection between the multipliers of Theorem 6.1 and those of the traditional optimality system of Theorem 2.2. As additional result we get higher regularity for several variables and some sign conditions.

**Proposition 6.2** *Let  $(\bar{y}_{\mathcal{I}}, \bar{y}_{\mathcal{A}}, \bar{u}_{\mathcal{I}}, \bar{u}_{\mathcal{A}})$  be the optimal solution to problem (3.1), where (3.1g), (3.1h) is replaced with (6.1), let  $p_{\mathcal{I}}, p_{\mathcal{A}}, q$ , and  $\sigma$  be defined by (6.2), and let  $(\wp_{\mathcal{I}}, \wp_{\mathcal{A}}) \in H^2(\mathcal{I}) \times H^2(\mathring{\mathcal{A}})$ ,  $(\mu_{\mathring{\mathcal{A}}}, \mu_{\gamma}) \in L^2(\mathring{\mathcal{A}}) \times H^{\frac{1}{2}}(\gamma)$  be the multipliers of the common optimality system (2.5). Then there holds*

$$p_{\mathcal{I}} = \wp_{\mathcal{I}}, \quad \text{a. e. in } \mathcal{I}, \quad (6.4a)$$

$$p_{\mathcal{A}} + q = \wp_{\mathcal{A}}, \quad \text{a. e. in } \mathring{\mathcal{A}}, \quad (6.4b)$$

$$-\Delta q + q = \mu_{\mathring{\mathcal{A}}}, \quad \text{a. e. in } \mathring{\mathcal{A}}, \quad (6.4c)$$

$$q = \sigma, \quad \text{a. e. on } \gamma, \quad (6.4d)$$

$$\partial_{\nu_{\mathcal{A}}} q = \mu_{\gamma}, \quad \text{a. e. on } \gamma, \quad (6.4e)$$

$$q \geq 0, \quad \text{a. e. in } \mathring{\mathcal{A}}, \quad (6.4f)$$

$$\sigma \geq 0, \quad \text{a. e. on } \gamma, \quad (6.4g)$$

and the regularities improve to  $p_{\mathcal{I}} \in H^2(\mathcal{I})$ ,  $p_{\mathcal{A}} \in H^2(\mathring{\mathcal{A}})$ ,  $q \in H^2(\mathring{\mathcal{A}})$ , and  $\sigma \in H^{\frac{3}{2}}(\gamma)$ .

*Proof.* Because primal variables match, we find, by comparison of (6.2f) and (2.5f), and respectively of (6.2g) and (2.5g),

$$\begin{aligned} p_{\mathcal{I}} &= \wp_{\mathcal{I}} \in H^2(\mathcal{I}), \\ p_{\mathcal{A}} + q &= \wp_{\mathcal{A}} \in H^2(\mathring{\mathcal{A}}). \end{aligned} \tag{6.5}$$

By inserting (6.5) into (2.5c) and by comparing with (6.2c), one obtains

$$-\Delta q + q = \mu_{\mathring{\mathcal{A}}}, \text{ a. e. in } \mathring{\mathcal{A}}.$$

Similarly we get

$$\partial_{\nu_{\mathcal{A}}} q = \mu_{\gamma}, \text{ a. e. on } \gamma$$

by equations (2.5e) and (6.2e). Since  $\mu_{\mathring{\mathcal{A}}} \in L^2(\mathring{\mathcal{A}})$  and  $\mu_{\gamma} \in H^{\frac{1}{2}}(\gamma)$ , regularity theory for the boundary value problem (6.4c), (6.4d) yields  $q \in H^2(\mathring{\mathcal{A}})$  and, due to (6.5), one obtains  $p_{\mathcal{A}} \in H^2(\mathring{\mathcal{A}})$ . Since  $\wp$  is continuous across the interface  $\gamma$  due to (2.5d), and since the jump condition (6.2d) holds, we furthermore have  $\sigma = \tau_{\mathcal{A}} q \in H^{\frac{3}{2}}(\gamma)$ . With  $\mu_{\gamma} \geq 0$  and  $\mu_{\mathring{\mathcal{A}}} \geq 0$  we find  $q \geq 0$  by weak maximum principle. Applying the Sobolev embedding theorem  $q \in H^2(\mathring{\mathcal{A}}) \hookrightarrow C(\mathcal{A})$  one finally gets  $\sigma \geq 0$ .  $\square$

**Remarks 6.3** A comparison of the two discussed Bryson-Denham-Dreyfus type approaches reveals the following results:

- The choice of the boundary condition in the ansatz leads to different jump conditions for the adjoint state. In particular, a Dirichlet boundary condition generates a kink (i. e. a jump in the derivative), whereas a Neumann boundary condition yields a jump. Consequently, one might use this degree of freedom with respect to further analysis or numerical treatment.
- In contrast to the ansatz of section 4, the regularity of  $\sigma$  improves vs. those of  $\mu_{\gamma}$  from  $H^{\frac{1}{2}}(\gamma)$  to  $H^{\frac{3}{2}}(\gamma)$ . This is due to the fact, that we utilize the first order differential operator  $\partial_{\nu_{\mathcal{A}}}$  in the reformulation (6.1b) of the state constraint.
- The second Bryson-Denham-Dreyfus type ansatz yields necessary conditions in which the original multipliers  $\mu_{\mathring{\mathcal{A}}}$  and  $\mu_{\gamma}$  are replaced with the  $H^2$ -regular  $q$  and its Dirichlet trace  $\sigma$ . Thus, in contrast to the first Bryson-Denham-Dreyfus type ansatz, the measure nature of the necessary conditions has disappeared.
- Higher regularity of the multipliers can be obtained via application of Theorem 7.5 in this ansatz, too. Thus, the comparison with the classical necessary conditions is not mandatory.

## 7 Set Optimal Control Problem

So far the derivation of new necessary conditions was based on an equivalent reformulation of the original model problem, which implicitly used the knowledge of the optimal active set  $\mathcal{A}$ . Of course  $\mathcal{A}$  is not known a priori, and, consequently, it is worth generalizing the problem formulations of sections 4 and 6 to a set optimal control problem, where the right splitting of  $\Omega$  in active and inactive sets has to be found as well. Exemplarily we use the ansatz of section 4.

$$\begin{aligned} \text{minimize } \mathfrak{J}(y_{\mathcal{J}}, y_{\mathcal{B}}, u_{\mathcal{J}}, u_{\mathcal{B}}, \mathcal{B}) &:= \frac{1}{2} \|y_{\mathcal{J}} - y_d\|_{L^2(\mathcal{J})}^2 + \frac{1}{2} \|y_{\mathcal{B}} - y_d\|_{L^2(\mathring{\mathcal{B}})}^2 \\ &+ \frac{\lambda}{2} \|u_{\mathcal{J}} - u_d\|_{L^2(\mathcal{J})}^2 + \frac{\lambda}{2} \|u_{\mathcal{B}} - u_d\|_{L^2(\mathring{\mathcal{B}})}^2 \end{aligned} \quad (7.1a)$$

$$\text{subject to } \quad -\Delta y_{\mathcal{J}} + y_{\mathcal{J}} = u_{\mathcal{J}}, \quad \text{a. e. in } \mathcal{J}, \quad (7.1b)$$

$$\quad -\Delta y_{\mathcal{B}} + y_{\mathcal{B}} = u_{\mathcal{B}}, \quad \text{a. e. in } \mathring{\mathcal{B}}, \quad (7.1c)$$

$$\quad \partial_{\nu} y_{\mathcal{J}} = 0, \quad \text{a. e. on } \Gamma, \quad (7.1d)$$

$$\quad y_{\mathcal{J}} = y_{\mathcal{B}}, \quad \text{a. e. on } \beta, \quad (7.1e)$$

$$\quad \partial_{\nu_{\mathcal{J}}} y_{\mathcal{J}} + \partial_{\nu_{\mathcal{B}}} y_{\mathcal{B}} = 0, \quad \text{a. e. on } \beta, \quad (7.1f)$$

$$\quad u_{\mathcal{B}} = -\Delta y_{\max} + y_{\max}, \quad \text{a. e. in } \mathring{\mathcal{B}}, \quad (7.1g)$$

$$\quad y_{\mathcal{B}} = y_{\max}, \quad \text{a. e. on } \beta, \quad (7.1h)$$

$$\quad u_{\mathcal{B}} \in L^2(\mathring{\mathcal{B}}), \quad u_{\mathcal{J}} \in L^2(\mathcal{J}), \quad (7.1i)$$

$$\quad y_{\mathcal{B}} \in H^2(\mathring{\mathcal{B}}), \quad y_{\mathcal{J}} \in H^2(\mathcal{J}), \quad (7.1j)$$

$$\quad y_{\mathcal{J}} < y_{\max}, \quad \text{a. e. in } \mathcal{J}, \quad (7.1k)$$

$$\quad \Omega = \mathcal{B} \dot{\cup} \mathcal{J}, \quad \beta := \partial \mathcal{B}, \quad \mathcal{B} \text{ fulfills Assumptions 2.1.} \quad (7.1l)$$

**Theorem 7.1** *Let Assumptions 2.1 be fulfilled. Then the original model problem (2.1) and the reformulation (7.1) possess the same unique optimal solution.*

*Proof.* We show that the optimal solution to the original model problem is optimal for the reformulation, as well. Let  $(\bar{y}, \bar{u})$  be the optimal solution to (2.1), and let  $\mathcal{A}$  be defined as in (2.2). Defining  $\mathcal{B} := \mathcal{A}$ ,  $\mathcal{J} := \Omega \setminus \mathcal{A}$ ,  $\bar{y}_{\mathcal{J}} := \bar{y}|_{\mathcal{J}}$ ,  $\bar{y}_{\mathcal{B}} := \bar{y}|_{\mathcal{B}}$ ,  $\bar{u}_{\mathcal{J}} := \bar{u}|_{\mathcal{J}}$ , and  $\bar{u}_{\mathcal{B}} := \bar{u}|_{\mathcal{B}}$ , it is obvious that we obtain a feasible point for (7.1). Now suppose this point is not optimal, in other words there exists  $(\hat{y}_{\mathcal{J}}, \hat{y}_{\mathcal{B}}, \hat{u}_{\mathcal{J}}, \hat{u}_{\mathcal{B}}, \hat{\mathcal{B}})$  with  $\mathfrak{J}(\hat{y}_{\mathcal{J}}, \hat{y}_{\mathcal{B}}, \hat{u}_{\mathcal{J}}, \hat{u}_{\mathcal{B}}, \hat{\mathcal{B}}) < \mathfrak{J}(\bar{y}_{\mathcal{J}}, \bar{y}_{\mathcal{B}}, \bar{u}_{\mathcal{J}}, \bar{u}_{\mathcal{B}}, \mathcal{B})$ . Regarding the equivalence between the boundary value problem (2.1b), (2.1c) and its splitted version, it can easily be verified that  $(\hat{y}, \hat{u})$  defined by

$$\hat{y} = \begin{cases} \hat{y}_{\mathcal{J}}, & \text{in } \hat{\mathcal{J}}, \\ \hat{y}_{\mathcal{B}}, & \text{in } \hat{\mathcal{B}}, \end{cases} \quad \text{and} \quad \hat{u} = \begin{cases} \hat{u}_{\mathcal{J}}, & \text{in } \hat{\mathcal{J}}, \\ \hat{u}_{\mathcal{B}}, & \text{in } \hat{\mathcal{B}}, \end{cases}$$

is a feasible point for the original problem (2.1). This yields

$$J(\hat{y}, \hat{u}) = \mathfrak{J}(\hat{y}_{\mathcal{J}}, \hat{y}_{\mathcal{B}}, \hat{u}_{\mathcal{J}}, \hat{u}_{\mathcal{B}}, \hat{\mathcal{B}}) < \mathfrak{J}(\bar{y}_{\mathcal{J}}, \bar{y}_{\mathcal{B}}, \bar{u}_{\mathcal{J}}, \bar{u}_{\mathcal{B}}, \mathcal{B}) = J(\bar{y}, \bar{u}),$$

in contradiction to optimality of  $(\bar{y}, \bar{u})$ . The uniqueness of the solution of the reformulation is shown with the same arguments as above together with the uniqueness of the solution of the original model problem.  $\square$

*Remark 7.2* As far as we know, problems of this type have not been studied in literature up to now. We call it a set optimal control problem for obvious reasons. The set of all admissible sets  $\mathcal{B}$  (cf. (7.11)) does not possess a vector space structure. Consequently, problem (7.1) does not fit in the usual framework of nonlinear optimization in Banach spaces. It is a hybrid problem which possesses Banach space variables and a set variable, whose topology and shape has to be determined. In this context Assumptions 2.1 rather defines the set of admissible sets (comparable to  $u_{\mathcal{J}} \in L^2(\mathcal{J})$ ), than being an explicit restriction.

Next, we analyze the set optimal control problem in more detail. Obviously, problem (7.1) is equivalent to the bilevel optimization problem

$$\begin{aligned} \min_{\mathcal{B}} \quad & \min_{y_{\mathcal{J}}, y_{\mathcal{B}}, u_{\mathcal{J}}, u_{\mathcal{B}}} \quad \mathfrak{J}(y_{\mathcal{J}}, y_{\mathcal{B}}, u_{\mathcal{J}}, u_{\mathcal{B}}; \mathcal{B}). \quad (7.2) \\ \text{s.t. (7.1k), (7.1l)} \quad & \text{s.t. (7.1b)–(7.1j)} \end{aligned}$$

A crucial point in this consideration is where to place inequality (7.1k). If it was treated as a constraint in the optimization with respect to the function space variables (in the following denoted as inner optimization), this subproblem might possess no optimum for a given splitting of  $\Omega$ . Hence, we use it as a constraint in the optimization with respect to the set variable. If the inequality was omitted, the optimal solution would be the same as the one of the original model problem without state constraint.

*Remark 7.3* The bilevel optimization problem (7.2) occurs here in a natural way, while Hintermüller, Ring [23] construct a similar problem by penalizing optimality conditions redundant for the inner optimization problem in the outer optimization problem.

Now we see, that the inner optimization yields a unique solution for any suitable splitting of  $\Omega$ .

**Theorem 7.4** *Let  $\Omega = \mathcal{B} \dot{\cup} \mathcal{J}$  be a fixed splitting compatible to Assumptions 2.1. Then the inner optimization problem (7.1a)–(7.1j) has a unique solution.*

*Proof.* Let  $\Omega = \mathcal{B} \dot{\cup} \mathcal{J}$  be the given splitting. Due to constraints (7.1g), (7.1h), and (7.1c), the variables  $u_{\mathcal{B}}$  and  $y_{\mathcal{B}}$  are fixed on the set  $\mathcal{B}$ . Hence, the subproblem can equivalently be reduced to the set  $\mathcal{J}$ :

$$\begin{aligned} \text{minimize} \quad & f(u_{\mathcal{J}}) := \frac{1}{2} \|S(u_{\mathcal{J}}) - y_d\|_{L^2(\mathcal{J})}^2 + \frac{\lambda}{2} \|u_{\mathcal{J}} - u_d\|_{L^2(\mathcal{J})}^2 \\ \text{subject to} \quad & u_{\mathcal{J}} \in U := \{u_{\mathcal{J}} \in L^2(\mathcal{J}) \mid \partial_{\nu_{\mathcal{J}}} S(u_{\mathcal{J}}) = -\partial_{\nu_{\mathcal{B}}} y_{\max} \text{ on } \beta\}, \end{aligned}$$

where  $S : L^2(\mathcal{J}) \rightarrow H^2(\mathcal{J})$  maps  $u_{\mathcal{J}}$  to the solution of the boundary value problem (7.1b), (7.1d), and  $y_{\mathcal{J}} = y_{\max}$  on  $\beta$ . Since  $S$  and  $\partial_{\nu_{\mathcal{J}}}$  are continuous (cf. [19, Thm. 2.3.3.2, 1.5.1.2]), and since  $U$  is a closed and convex subset of  $L^2(\mathcal{J})$ , the reduced problem has a unique optimal solution (cf. [33, Thm. 2.16]).  $\square$

A consequence of Theorem 7.4 is that the existence result of Theorem 5.1 is valid for any feasible splitting and its corresponding optimal solution. Therefore, the inner optimization for a given splitting is equivalent to solving the first order necessary conditions of Theorem 5.1. Since Proposition 5.2 allows the lifting of the regularity of the multipliers only at the optimal configuration by comparing our necessary conditions with the classical ones, we now need an analogous result for a given arbitrary splitting.

**Theorem 7.5** *Let  $\Omega = \mathcal{B} \dot{\cup} \mathcal{J}$  be a fixed splitting compatible to Assumptions 2.1. Then the boundary value problem*

$$\begin{aligned} -\Delta y_{\mathcal{J}} + y_{\mathcal{J}} &= -\frac{1}{\lambda}p + u_d, & \text{a. e. in } \mathcal{J}, \\ -\Delta p_{\mathcal{J}} + p_{\mathcal{J}} &= y_{\mathcal{J}} - y_d, & \text{a. e. in } \mathcal{J}, \\ \partial_{\nu} y_{\mathcal{J}} &= 0, & \text{a. e. on } \Gamma, \\ \partial_{\nu} p_{\mathcal{J}} &= 0, & \text{a. e. on } \Gamma, \\ y_{\mathcal{J}} &= y_{\max}, & \text{a. e. on } \beta, \\ \partial_{\nu_{\mathcal{J}}} y_{\mathcal{J}} &= \partial_{\nu_{\mathcal{J}}} y_{\max}, & \text{a. e. on } \beta, \end{aligned}$$

possesses a unique solution  $(y_{\mathcal{J}}, p_{\mathcal{J}}) \in H^2(\mathcal{J}) \times H^2(\mathcal{J})$ .

*Proof.* The proof is divided into three steps. First we transform the system into a fourth order boundary value problem with homogeneous boundary values. Therefore, we start with  $\tilde{y} \in H^4(\mathcal{J})$  fulfilling  $\tilde{y} = y_{\max}$ ,  $\partial_{\nu_{\mathcal{J}}} \tilde{y} = \partial_{\nu_{\mathcal{J}}} y_{\max}$  on  $\beta$ , and  $\partial_{\nu} \tilde{y} = 0$  on  $\Gamma$ , which exists due to  $y_{\max} \in H^4(\Omega)$  and  $\text{dist}(\Gamma, \beta) \geq \delta > 0$  for a fixed  $\delta$  according to Assumptions 2.1. Let  $\tilde{p} \in H^2(\mathcal{J})$  fulfill  $-\Delta \tilde{p} + \tilde{p} = \tilde{y} - y_d$  in  $\mathcal{J}$ ,  $\partial_{\nu} \tilde{p} = 0$  on  $\Gamma$ . Then  $\hat{y} := y_{\mathcal{J}} - \tilde{y}$ ,  $\hat{p} := p_{\mathcal{J}} - \tilde{p}$  solve

$$-\Delta \hat{y} + \hat{y} = -\frac{1}{\lambda} \hat{p} + F, \quad \text{a. e. in } \mathcal{J}, \quad (7.3a)$$

$$-\Delta \hat{p} + \hat{p} = \hat{y}, \quad \text{a. e. in } \mathcal{J}, \quad (7.3b)$$

$$\partial_{\nu} \hat{y} = 0, \quad \text{a. e. on } \Gamma, \quad (7.3c)$$

$$\partial_{\nu} \hat{p} = 0, \quad \text{a. e. on } \Gamma, \quad (7.3d)$$

$$\hat{y} = 0, \quad \text{a. e. on } \beta, \quad (7.3e)$$

$$\partial_{\nu_{\mathcal{J}}} \hat{y} = 0, \quad \text{a. e. on } \beta, \quad (7.3f)$$

with  $F := u_d + \Delta \tilde{y} - \tilde{y} - \frac{1}{\lambda} \tilde{p} \in H^2(\mathcal{J})$ . By substituting (7.3b) in (7.3a) we obtain

$$\Delta^2 \hat{p} - 2\Delta \hat{p} + \left(1 + \frac{1}{\lambda}\right) \hat{p} = F, \quad \text{a. e. in } \mathcal{J}, \quad (7.4a)$$

$$\partial_{\nu} \hat{p} = 0, \quad \text{a. e. on } \Gamma, \quad (7.4b)$$

$$\partial_{\nu} (-\Delta \hat{p} + \hat{p}) = 0, \quad \text{a. e. on } \Gamma, \quad (7.4c)$$

$$-\Delta \hat{p} + \hat{p} = 0, \quad \text{a. e. on } \beta, \quad (7.4d)$$

$$\partial_{\nu_{\mathcal{J}}} (-\Delta \hat{p} + \hat{p}) = 0, \quad \text{a. e. on } \beta. \quad (7.4e)$$

In a second step we prove existence and uniqueness for a solution  $\hat{p} \in H^2(\mathcal{J})$  of this fourth order boundary value problem. Defining  $H := \{v \in H^2(\mathcal{J}) \mid \partial_{\nu} v = 0 \text{ on } \Gamma\}$  and the bilinear form

$$B : H \times H \rightarrow \mathbb{R}, (v, w) \mapsto (\Delta v, \Delta w) + 2(\nabla v, \nabla w) + \left(1 + \frac{1}{\lambda}\right)(v, w) - \int_{\beta} (w \partial_{\nu_{\mathcal{J}}} v + v \partial_{\nu_{\mathcal{J}}} w),$$

where  $(\cdot, \cdot)$  denotes the scalar product in  $L^2(\mathcal{J})$ , resp.  $L^2(\mathcal{J})^2$ . In fact, the bilinear form  $B$  is a scalar product on  $H$ , since it is symmetric and positive definite, which is seen by integration by parts:

$$B(v, v) = \int_{\mathcal{J}} (\Delta v - v)^2 + \frac{1}{\lambda} v^2.$$

Since  $F \in L^2(\mathcal{J})$ , the linear form  $(F, \cdot)$  is in  $(H, B(\cdot, \cdot))^*$ . With Riesz representation theorem we have existence and uniqueness of  $\check{p} \in H$  such that

$$B(\check{p}, w) = (F, w), \quad \forall w \in H. \quad (7.5)$$

Note that (7.5) is the weak formulation of the boundary value problem (7.4), which is easily verified by integration by parts and well-known variational arguments. The third step brings the proof full circle. Let  $\check{y} \in H^2(\mathcal{J})$  be the unique solution to (7.3a), (7.3c), (7.3e) with  $\hat{p}$  replaced by  $\check{p}$ . Since  $\check{p}$  fulfills (7.3d) by definition of  $H$ , the equations (7.3b) and (7.3f) remain to be shown. According to the second step, it holds for all  $w \in H$ :

$$\begin{aligned} & (F, w) \\ &= (\Delta\check{p}, \Delta w) + (\nabla\check{p}, \nabla w) - \int_{\beta} \check{p} \partial_{\nu_{\mathcal{J}}} w + \left(\frac{1}{\lambda}\check{p}, w\right) + (\nabla\check{p}, \nabla w) + (\check{p}, w) - \int_{\beta} w \partial_{\nu_{\mathcal{J}}}\check{p} \\ &= (\Delta\check{p}, \Delta w) - (\check{p}, \Delta w) + \int_{\Gamma} \underbrace{\check{p} \partial_{\nu} w}_{=0} + (\Delta\check{y} - \check{y} + F, w) - (\Delta\check{p}, w) + (\check{p}, w) + \int_{\Gamma} w \underbrace{\partial_{\nu}\check{p}}_{=0} \\ &= (\Delta\check{p} - \check{p} + \check{y}, \Delta w - w) + (F, w) + \int_{\Gamma} w \underbrace{\partial_{\nu}\check{y}}_{=0} + \int_{\beta} w \partial_{\nu_{\mathcal{J}}}\check{y} - \int_{\Gamma} \underbrace{\check{y} \partial_{\nu} w}_{=0} - \int_{\beta} \underbrace{\check{y}}_{=0} \partial_{\nu_{\mathcal{J}}} w. \end{aligned}$$

This implies

$$0 = (\Delta\check{p} - \check{p} + \check{y}, \Delta w - w) + \int_{\beta} w \partial_{\nu_{\mathcal{J}}}\check{y}, \quad \forall w \in H.$$

For an arbitrary  $s \in L^2(\mathcal{J})$  there exists a unique  $w \in H$  which solves  $-\Delta w + w = -s$  in  $\mathcal{J}$  and  $w = 0$  on  $\beta$ . Therefore, it holds (7.3b) and consequently (7.3f), too. By applying the inverse transformations of the first step we conclude the proof.  $\square$

**Remark 7.6** The regularity results of Theorem 7.5 also yield higher smoothness of the multipliers  $q$  and  $\sigma$  by (5.1e) and (5.1g).

We continue with the bilevel optimization problem (7.2) and reformulate it again: In a first step we formulate a minimization problem with the optimality system of the inner problem as constraints of the outer problem. This reformulation fits in the framework of PDE constrained topology and/or shape optimization. In a second step we replace the optimality system by a saddle point problem for the Lagrangian of the inner optimization and obtain

$$\min_{\substack{\mathcal{B} \\ \text{s.t. (7.1k), (7.1l)}}} \left( \min_{(y_{\mathcal{J}}, y_{\mathcal{B}}, u_{\mathcal{J}}, u_{\mathcal{B}})} \max_{(p_{\mathcal{J}}, p_{\mathcal{B}}, q, \sigma, \zeta, \xi)} \mathcal{L}(y_{\mathcal{J}}, y_{\mathcal{B}}, u_{\mathcal{J}}, u_{\mathcal{B}}; p_{\mathcal{J}}, p_{\mathcal{B}}, q, \sigma, \zeta, \xi; \mathcal{B}) \right)$$

with

$$\begin{aligned}
\mathcal{L}(y_{\mathcal{J}}, y_{\mathcal{B}}, u_{\mathcal{J}}, u_{\mathcal{B}}; p_{\mathcal{J}}, p_{\mathcal{B}}, q, \sigma, \zeta, \xi; \mathcal{B}) &:= \mathcal{J}(y_{\mathcal{J}}, y_{\mathcal{B}}, u_{\mathcal{J}}, u_{\mathcal{B}}) \\
&- \int_{\mathcal{J}} \nabla y_{\mathcal{J}} \nabla p_{\mathcal{J}} + (y_{\mathcal{J}} - u_{\mathcal{J}}) p_{\mathcal{J}} - \int_{\mathring{\mathcal{B}}} \nabla y_{\mathcal{B}} \nabla p_{\mathcal{B}} + (y_{\mathcal{B}} - u_{\mathcal{B}}) p_{\mathcal{B}} \\
&+ \int_{\beta} (y_{\mathcal{J}} - y_{\mathcal{B}}) \zeta + \int_{\beta} \partial_{\nu_{\mathcal{J}}} y_{\mathcal{J}} (\xi + p_{\mathcal{J}}) + \partial_{\nu_{\mathcal{B}}} y_{\mathcal{B}} (\xi + p_{\mathcal{B}}) \\
&+ \int_{\mathring{\mathcal{B}}} (\Delta y_{\max} - y_{\max} + u_{\mathcal{B}}) q + \int_{\beta} (y_{\mathcal{B}} - y_{\max}) \sigma.
\end{aligned}$$

This final reformulation is very close to the framework of Delfour and Zolésio in [16, Chp. 10, Thm. 5.1]. By their theorem one can obtain the shape derivative of the value of the minimax problem as partial shape derivative of the Lagrangian evaluated at the minimaximizer. An analogous result with respect to a topology derivative is not known to the best knowledge of the authors.

From now on we proceed on an informal way to derive our final necessary conditions. First we omit the inequality (7.1k) temporarily. The derivatives of the Lagrangian with respect to the function space variables yield the well-known optimality system of the inner optimization. In addition, we compute the partial shape derivative of  $\mathcal{L}$  at

$$(\bar{y}_{\mathcal{I}}, \bar{y}_{\mathcal{A}}, \bar{u}_{\mathcal{I}}, \bar{u}_{\mathcal{A}}; p_{\mathcal{I}}, p_{\mathcal{A}}, q, \sigma, \zeta = \partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}}, \xi = -p_{\mathcal{I}}|_{\gamma}; \mathcal{A})$$

and obtain, for all perturbation fields  $V \in \mathcal{V} := \{V \in C^{1,1}(\bar{\Omega}, \mathbb{R}^2) | V = 0 \text{ on } \Gamma\}$  according to [31]:

$$\begin{aligned}
\partial_{\mathcal{B}} \mathcal{L}(\dots)[V] &= \\
&\int_{\gamma} (V \cdot \nu_{\mathcal{I}}) \left[ \underbrace{\frac{1}{2}(\bar{y}_{\mathcal{I}} - y_d)^2 - \frac{1}{2}(\bar{y}_{\mathcal{A}} - y_d)^2}_{=0} + \underbrace{\frac{\lambda}{2}(\bar{u}_{\mathcal{I}} - u_d)^2}_{=\frac{1}{2\lambda} p_{\mathcal{I}}^2} - \underbrace{\frac{\lambda}{2}(\bar{u}_{\mathcal{A}} - u_d)^2}_{=\frac{1}{2\lambda} (p_{\mathcal{A}} + q)^2} \right] \\
&+ \int_{\gamma} (V \cdot \nu_{\mathcal{I}}) \left[ -\nabla \bar{y}_{\mathcal{I}} \nabla p_{\mathcal{I}} + \underbrace{(\bar{u}_{\mathcal{I}} - \bar{y}_{\mathcal{I}}) p_{\mathcal{I}}}_{=-\frac{1}{\lambda} p_{\mathcal{I}} + u_d} + \nabla \bar{y}_{\mathcal{A}} \nabla p_{\mathcal{A}} + \underbrace{(\bar{y}_{\mathcal{A}} - \bar{u}_{\mathcal{A}}) p_{\mathcal{A}}}_{=\frac{1}{\lambda} (p_{\mathcal{A}} + q) - u_d} \right] \\
&+ \int_{\gamma} (V \cdot \nu_{\mathcal{I}}) \left[ -\underbrace{(\Delta y_{\max} - y_{\max} + \bar{u}_{\mathcal{A}}) q}_{=0} \right] \\
&+ \int_{\gamma} (V \cdot \nu_{\mathcal{I}}) \left[ \underbrace{(\partial_{\nu_{\mathcal{I}}} \bar{y}_{\mathcal{I}} + \partial_{\nu_{\mathcal{A}}} \bar{y}_{\mathcal{A}}) \partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}} + \kappa_{\mathcal{I}} (\bar{y}_{\mathcal{I}} - \bar{y}_{\mathcal{A}}) \partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}}}_{=0} \right. \\
&\quad + \partial_{\nu_{\mathcal{I}}}^2 \bar{y}_{\mathcal{I}} \underbrace{(-p_{\mathcal{I}} + p_{\mathcal{I}})}_{=0} + \partial_{\nu_{\mathcal{I}}} \bar{y}_{\mathcal{I}} \partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}} + \kappa_{\mathcal{I}} \partial_{\nu_{\mathcal{I}}} \bar{y}_{\mathcal{I}} \underbrace{(-p_{\mathcal{I}} + p_{\mathcal{I}})}_{=0} \\
&\quad - \partial_{\nu_{\mathcal{A}}}^2 \bar{y}_{\mathcal{A}} \underbrace{(-p_{\mathcal{I}} + p_{\mathcal{A}})}_{=0} - \partial_{\nu_{\mathcal{A}}} \bar{y}_{\mathcal{A}} \partial_{\nu_{\mathcal{A}}} p_{\mathcal{A}} - \kappa_{\mathcal{A}} \partial_{\nu_{\mathcal{A}}} \bar{y}_{\mathcal{A}} \underbrace{(-p_{\mathcal{I}} + p_{\mathcal{A}})}_{=0} \\
&\quad \left. - \partial_{\nu_{\mathcal{I}}}^2 \bar{y}_{\mathcal{A}} \underbrace{(\partial_{\nu_{\mathcal{A}}} \bar{y}_{\mathcal{A}} - \partial_{\nu_{\mathcal{A}}} y_{\max})}_{=0} \sigma + \kappa_{\mathcal{I}} \underbrace{(\bar{y}_{\mathcal{A}} - y_{\max})}_{=0} \sigma \right] \\
&= \int_{\gamma} (V \cdot \nu_{\mathcal{I}}) \left[ \underbrace{-\nabla_{\gamma} \bar{y}_{\mathcal{I}} \nabla_{\gamma} p_{\mathcal{I}} + \nabla_{\gamma} \bar{y}_{\mathcal{A}} \nabla_{\gamma} p_{\mathcal{A}}}_{=0} - \underbrace{\frac{1}{2\lambda} p_{\mathcal{I}}^2 + \frac{1}{2\lambda} p_{\mathcal{A}}^2}_{=0} - \frac{1}{2\lambda} q^2 \right],
\end{aligned}$$

where we have used (7.1b)–(7.1h), (5.1) and  $y_d \in H^1(\Omega)$ . Here  $\kappa_{\mathcal{I}} = -\kappa_{\mathcal{A}}$  denote the mean curvature of  $\gamma$ , and  $\nabla_{\gamma}$  the tangential gradient.

Equation (5.3d), which was obtained by comparison with the traditional necessary conditions, now guarantees that the partial shape derivative of  $\mathcal{L}$  vanishes at the optimal solution of the original model problem. Vice versa, this can be interpreted as justification for our informal way to derive first order necessary conditions for the set optimal control problem (7.1) by means of a Lagrange principle. The shape derivative exactly yields the missing condition of Theorem 5.1 (cf. Remarks 5.3):

$$0 = -\frac{1}{2\lambda} q^2 \stackrel{(5.1d)}{=} -\frac{1}{2\lambda} (p_{\mathcal{A}} + q - p_{\mathcal{I}})^2 \stackrel{(5.1f)}{\stackrel{(5.1g)}{=}} -\frac{\lambda}{2} (\bar{u}_{\mathcal{A}} - \bar{u}_{\mathcal{I}})^2 \Leftrightarrow \bar{u}_{\mathcal{A}} = \bar{u}_{\mathcal{I}}, \text{ on } \gamma.$$

This is the analog to the continuity of the Hamiltonian across junction points of boundary arcs for autonomous problems in OC-ODE, see [27, p. 22, (iii)], [20, (5.15)].

In summary the vanishing partial shape derivative of the Lagrangian  $\mathcal{L}$  is compatible with the inequality (7.1k).

*Remark 7.7* The conceptual ideas of this section are not mandatorily connected with the Bryson-Denham-Dreyfus reformulation of the state constraint. In stead of  $u_{\mathcal{B}} = -\Delta y_{\max} + y_{\max}$ , we could use  $y_{\mathcal{B}} = y_{\max}$  a.e. in  $\mathring{\mathcal{B}}$ , which leads to the traditional necessary conditions of Theorem 2.2.

We end the paper by assembling the new necessary conditions in form of free boundary problems:

for the BDD-ansatz of section 4:

$$\begin{array}{llll} -\Delta \bar{y}_{\mathcal{I}} + \bar{y}_{\mathcal{I}} = \bar{u}_{\mathcal{I}}, & \text{in } \mathcal{I}, & -\Delta p_{\mathcal{I}} + p_{\mathcal{I}} = \bar{y}_{\mathcal{I}} - y_d, & \text{in } \mathcal{I}, \\ -\Delta \bar{y}_{\mathcal{A}} + \bar{y}_{\mathcal{A}} = \bar{u}_{\mathcal{A}}, & \text{in } \mathring{\mathcal{A}}, & -\Delta p_{\mathcal{A}} + p_{\mathcal{A}} = \bar{y}_{\mathcal{A}} - y_d, & \text{in } \mathring{\mathcal{A}}, \\ \partial_{\nu} \bar{y}_{\mathcal{I}} = 0, & \text{on } \Gamma, & \partial_{\nu} p_{\mathcal{I}} = 0, & \text{on } \Gamma, \\ \bar{y}_{\mathcal{I}} = \bar{y}_{\mathcal{A}}, & \text{on } \gamma, & p_{\mathcal{I}} = p_{\mathcal{A}}, & \text{on } \gamma, \\ \partial_{\nu_{\mathcal{I}}} \bar{y}_{\mathcal{I}} = -\partial_{\nu_{\mathcal{A}}} \bar{y}_{\mathcal{A}}, & \text{on } \gamma, & \partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}} = \sigma - \partial_{\nu_{\mathcal{A}}} p_{\mathcal{A}}, & \text{on } \gamma, \\ -\Delta y_{\max} + y_{\max} = \bar{u}_{\mathcal{A}}, & \text{in } \mathring{\mathcal{A}}, & \lambda (\bar{u}_{\mathcal{I}} - u_d) = -p_{\mathcal{I}}, & \text{in } \mathcal{I}, \\ \bar{y}_{\mathcal{A}} = y_{\max}, & \text{on } \gamma, & \lambda (\bar{u}_{\mathcal{A}} - u_d) = -p_{\mathcal{A}} - q, & \text{in } \mathring{\mathcal{A}}, \\ \bar{y}_{\mathcal{I}} < y_{\max}, & \text{in } \mathcal{I}, & q = 0 & \text{on } \gamma, \\ & & \sigma \geq 0 & \text{on } \gamma, \\ & & q \geq 0 & \text{in } \mathring{\mathcal{A}}, \end{array}$$

and for the BDD-ansatz of section 6 (with different functions  $q, \sigma = q|_{\gamma}, p_{\mathcal{A}}$ ):

$$\begin{array}{llll} -\Delta \bar{y}_{\mathcal{I}} + \bar{y}_{\mathcal{I}} = \bar{u}_{\mathcal{I}}, & \text{in } \mathcal{I}, & -\Delta p_{\mathcal{I}} + p_{\mathcal{I}} = \bar{y}_{\mathcal{I}} - y_d, & \text{in } \mathcal{I}, \\ -\Delta \bar{y}_{\mathcal{A}} + \bar{y}_{\mathcal{A}} = \bar{u}_{\mathcal{A}}, & \text{in } \mathring{\mathcal{A}}, & -\Delta p_{\mathcal{A}} + p_{\mathcal{A}} = \bar{y}_{\mathcal{A}} - y_d, & \text{in } \mathring{\mathcal{A}}, \\ \partial_{\nu} \bar{y}_{\mathcal{I}} = 0, & \text{on } \Gamma, & \partial_{\nu} p_{\mathcal{I}} = 0, & \text{on } \Gamma, \\ \bar{y}_{\mathcal{I}} = \bar{y}_{\mathcal{A}}, & \text{on } \gamma, & p_{\mathcal{I}} = p_{\mathcal{A}} + q, & \text{on } \gamma, \\ \partial_{\nu_{\mathcal{I}}} \bar{y}_{\mathcal{I}} = -\partial_{\nu_{\mathcal{A}}} \bar{y}_{\mathcal{A}}, & \text{on } \gamma, & \partial_{\nu_{\mathcal{I}}} p_{\mathcal{I}} = -\partial_{\nu_{\mathcal{A}}} p_{\mathcal{A}}, & \text{on } \gamma, \\ -\Delta y_{\max} + y_{\max} = \bar{u}_{\mathcal{A}}, & \text{in } \mathring{\mathcal{A}}, & \lambda (\bar{u}_{\mathcal{I}} - u_d) = -p_{\mathcal{I}}, & \text{in } \mathcal{I}, \\ \partial_{\nu_{\mathcal{A}}} \bar{y}_{\mathcal{A}} = \partial_{\nu_{\mathcal{A}}} y_{\max}, & \text{on } \gamma, & \lambda (\bar{u}_{\mathcal{A}} - u_d) = -p_{\mathcal{A}} - q, & \text{in } \mathring{\mathcal{A}}, \\ \bar{y}_{\mathcal{I}} < y_{\max}, & \text{in } \mathcal{I}, & q \geq 0 & \text{on } \gamma. \end{array}$$

*Remark 7.8* These sets of necessary conditions build the basis for our new numerical method presented in a forthcoming paper. Note that, despite of the shape optimization problem (7.2), shape adjoints will not play a role as in [23]. Instead, we will determine the free interface by a so-called shape linearization technique.

## 8 Conclusions

This paper presents two main ideas: The first is introducing a splitting of the domain  $\Omega$  into two sets, which should have an interpretation as active, resp. inactive sets of the state constraint. This idea leads to a set optimal control problem whose analysis is non-standard. It combines the research fields of optimal control of PDEs and shape, resp. topology optimization and it reveals encouraging insight in the structure of PDE optimal control problems with state constraints.

The second idea is the transfer of the Bryson-Denham-Dreyfus ansatz for state constrained optimal control with ordinary differential equations to state constrained elliptic optimal control problems. Two variants have been investigated leading to new first order necessary conditions. The multipliers associated with the state constraint distinguish themselves from the well-known ones by higher regularity. In particular, the new multiplier in the second ansatz (section 6) is no longer a measure. This lifting of regularity corresponds to the results known from OC-ODE, when comparing the necessary conditions of the BDD approach and the approach of Jacobson, Lele and Speyer, which is known to have some drawbacks in practical applications.

Depending on the ansatz in the present paper, the adjoint variable exhibits a jump condition either in its value or in its normal derivative. Moreover, the new multipliers exhibit relations and further properties which are highly comparable to known results in OC-ODE.

The research is now open to further investigations. First of all the informal application of a Lagrange principle for treating the set optimal control problem requires a mathematically rigorous proof. In a next step, the new treatment of state constraints in PDE-constrained optimization should be generalizable to semilinear elliptic problems and more complicated state constraints. An other interesting topic is the treatment of state constrained parabolic optimal control problems.

Further open questions are concerned with the introduction of an order concept like in ordinary optimal control theory and its possible implications on the topology of active sets. Some first observations for a coupled control system of a parabolic and an ordinary differential equation have already been made; see [29, 34].

In a forthcoming paper, an adapted numerical scheme for solving the new first order optimality systems is presented, that exhibits some encouraging properties. In particular, the numerical treatment is based on the perspective of the set optimal control problem, cf. section 7.

## Acknowledgements

The authors are indebted to Christian Simader, University of Bayreuth, for his many valuable hints concerning PDE theory. Thanks are also due to Stephan Schmidt, University of Trier, for helpful discussions concerning shape optimization.

## References

- [1] J. J. Alibert and J. P. Raymond. Boundary control of semilinear elliptic equations with discontinuous leading coefficients and unbounded controls. *Numerical Functional Analysis and Optimization*, 3&4:235–250, 1997.
- [2] J. P. Aubin. *Approximation of Elliptic Boundary-Value Problems*, volume XXVI of *Pure and Applied Mathematics*. Wiley-Interscience, New York, 1972.
- [3] S. Bechmann and M. Frey. *Regularisierungsmethoden für Optimalsteuerungsprobleme*, volume 80 of *Bayreuther Mathematische Schriften*. Mathematisches Institut der Universität Bayreuth, 2008.
- [4] M. Bergounioux, M. Haddou, M. Hintermüller, and K. Kunisch. A comparison of a Moreau-Yosida based active set strategy and interior point methods for constrained optimal control problems. *SIAM Journal on Optimization*, 11:495–521, 2000.
- [5] M. Bergounioux and K. Kunisch. Primal-dual strategy for state-constrained optimal control problems. *Computational Optimization and Applications*, 22:193–224, 2002.
- [6] M. Bergounioux and K. Kunisch. On the structure of Lagrange multipliers for state-constrained optimal control problems. *Systems & Control Letters*, 48:169–176, 2003.
- [7] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*, volume 15 of *Springer Series in Computational Mathematics*. Springer, Berlin Heidelberg New York, 1991.
- [8] A. E. Bryson, Jr., W. F. Denham, and S. E. Dreyfus. Optimal programming problems with inequality constraints I: Necessary conditions for extremal solutions. *AIAA Journal*, 1(11):2544–2550, 1963.
- [9] R. Z. Bulirsch, F. Montrone, and H. J. Pesch. Abort landing in the presence of windshear as a minimax optimal control problem. Part 1: Necessary conditions. *Journal of Optimization Theory and Applications*, 70:1–23, 1991.
- [10] R. Z. Bulirsch, F. Montrone, and H. J. Pesch. Abort landing in the presence of windshear as a minimax optimal control problem. Part 2: Multiple shooting and homotopy. *Journal of Optimization Theory and Applications*, 70:223–254, 1991.
- [11] E. Casas. Control of an elliptic problem with pointwise state constraints. *SIAM Journal on Control and Optimization*, 4:1309–1322, 1986.
- [12] E. Casas. Boundary control of semilinear elliptic equations with pointwise state constraints. *SIAM Journal on Control and Optimization*, 31:993–1006, 1993.
- [13] E. Casas. Pontryagin’s principle for state-constrained boundary control problems of semilinear parabolic equations. *SIAM Journal on Control and Optimization*, 35(4):1297–1327, 1997.
- [14] K. Chudej, H. J. Pesch, and K. Sternberg. Optimal control of load changes for molten carbonate fuel cell systems: A challenge in PDE constrained optimization. *SIAM Journal on Applied Mathematics*, 70(2):621–639, 2009.

- [15] J. C. De Los Reyes and R. Griesse. State-constrained optimal control of the three-dimensional stationary Navier-Stokes equations. *Journal of Mathematical Analysis and Applications*, 343(1):257–272, 2008.
- [16] M. C. Delfour and J. P. Zolésio. *Shape and Geometries*, volume 22 of *Advances in Design and Control*. SIAM, Philadelphia, 2nd edition, 2011.
- [17] K. Eppler and F. Tröltzsch. Fast optimization methods in the selective cooling of steel. In M. Grötschel, S. O. Krumke, and J. Rambau, editors, *Online Optimization of Large Scale Systems*, pages 185–204. Springer, Berlin Heidelberg, 2001.
- [18] K. Eppler and F. Tröltzsch. Optimal control problems for the nonlinear heat equation. In M. Grötschel, S. O. Krumke, and J. Rambau, editors, *Online Optimization of Large Scale Systems*, pages 173–184. Springer, Berlin, Heidelberg, 2001.
- [19] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Pitman Advanced Publishing Program, Boston London Melbourne, 1985.
- [20] R. F. Hartl, S. P. Sethi, and R. G. Vickson. A survey of the maximum principles for optimal control problems with state constraints. *SIAM Review*, 37(2):181–218, 1995.
- [21] R. Herzog and K. Kunisch. Algorithms for PDE-constrained optimization. *GAMM-Mitteilungen*, 33(2):163–176, 2010.
- [22] M. Hintermüller and K. Kunisch. Stationary optimal control problems with pointwise state constraints. In *Numerical PDE Constrained Optimization*, volume 72 of *Lecture Notes in Computational Science and Engineering*. Springer, Berlin Heidelberg, 2009.
- [23] M. Hintermüller and W. Ring. A level set approach for the solution of a state constrained optimal control problem. *Numerische Mathematik*, 98:135–166, 2004.
- [24] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer, Dordrecht, NL, 2009.
- [25] D. H. Jacobson, M. M. Lele, and J. L. Speyer. New necessary conditions of optimality for control problems with state-variable inequality constraints. *Journal of Mathematical Analysis and Applications*, 35(2):255–284, 1971.
- [26] H. Maurer. *Optimale Steuerprozesse mit Zustandsbeschränkungen*. Habilitationsschrift, Universität Würzburg, 1976.
- [27] H. Maurer. On the minimum principle for optimal control problems with state constraints. *Schriftenreihe des Rechenzentrums der Universität Münster*, 41, 1979.
- [28] H. Maurer and H. J. Pesch. Direct optimization methods for solving a complex state-constrained optimal control problem in microeconomics. *Applied Mathematics and Computation*, 204(2):568–579, 2008.
- [29] H. J. Pesch, A. Rund, W. von Wahl, and S. Wendl. On some new phenomena in state-constrained optimal control if ODE as well as PDE are involved. *Control and Cybernetics*, 39:647–660, 2010.

- [30] J. P. Raymond and H. Zidani. Hamiltonian Pontryagin's principles for control problems governed by semilinear parabolic equations. *Applied Mathematics & Optimization*, 39:143–177, 1999.
- [31] J. Sokołowski and J. P. Zolésio. *Introduction to Shape Optimization*. Springer, 1992.
- [32] F. Tröltzsch. *Optimality Conditions for Parabolic Control Problems and Applications*, volume 62 of *Teubner Texte zur Mathematik*. Teubner, Leipzig, 1984.
- [33] F. Tröltzsch. *Optimal Control of Partial Differential Equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, 2010.
- [34] S. Wendl, H. J. Pesch, and A. Rund. On a state-constrained PDE optimal control problem arising from ODE-PDE optimal control. In M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels, editors, *Recent Advances in Optimization and its Applications in Engineering*, pages 429–438. Springer, Berlin, 2010.

## C.2

**On a Prototype Class of ODE-PDE  
State-constrained Optimal Control Problems.  
Part 1:  
Analysis of the State-unconstrained Problems**

Hans Josef Pesch, Armin Rund, Wolf von Wahl, Stefan Wendl

Preprint, Universität Bayreuth, 2011



# On a Prototype Class of ODE-PDE State-constrained Optimal Control Problems. Part 1: Analysis of the State-unconstrained Problems

Hans Josef Pesch\*    Armin Rund†    Wolf von Wahl‡    Stefan Wendl§

## Abstract

In this bipartite paper, we investigate a new class of optimal control problems with constraints in form of a coupled system of ordinary and partial differential equations. We call this kind of problems hypersonic rocket car problems, since they are inspired, on the one hand, by the well-known rocket car problem, and, on the other hand, by a recently investigated flight path trajectory optimization problem for a hypersonic aircraft.

The hypersonic rocket car problems can be considered as undressed abstract examples for a class of staggered state-constrained ODE-PDE-constrained optimal control problems with a similar coupling structure as the hypersonic aircraft problem. This simplification allows to obtain analytical solutions to a certain extent which is prohibited by the enormous complexity of the flight path optimization problem.

The analysis of structural questions concerning the existence of boundary arcs and touch points of state constraints is the aim of the second part of this paper and is novel in the context of PDE-constrained optimal control. We obtain results, which are, at a first glance, similar to state-constrained ODE optimal control problems and show their relation to the differentiation index of the related partial differential algebraic equation system along state-constrained subarcs. At a second glance, new phenomena are observed caused by the non-local character of the state constraint in the ODE context leading to additional constraints on the ODE states from the beginning of the process on.

**Key words.** Optimal control of partial differential equations, ODE-PDE-constrained optimization, state constraints, partial differential algebraic equations.

---

\*University of Bayreuth, Chair of Mathematics in Engineering Sciences, 95440 Bayreuth, Germany.  
Corresponding author: [hans-josef.pesch@uni-bayreuth.de](mailto:hans-josef.pesch@uni-bayreuth.de)

†University of Bayreuth, Mathematics in Engineering Sciences, 95440 Bayreuth, Germany

‡University of Bayreuth, Chair of Mathematics VI (Partial Differential Equations and Mathematical Physics), 95440 Bayreuth, Germany

§University of Bayreuth, Mathematics in Engineering Sciences, 95440 Bayreuth, Germany. This author was supported via the International Doctorate Program *Identification, Optimization and Control with Applications in Modern Technologies* within the Elite Network of Bavaria

## 1 Introduction

Realistic mathematical models of dynamical processes from scientific or engineering background may often have to consider different physical phenomena leading to coupled systems of equations that include partial and ordinary differential equations as well as algebraic equations. Frequently, their numerical solution is only the first step. The identification of system parameters and the control of such systems are tackled subsequently. Mathematically one obtains optimization problems where the constraints are given by the underlying dynamical process. Because of their complexity such optimization problems are not widely studied in literature, neither theoretically nor numerically.

One such example was an optimal control problem recently studied by Chudej et. al. [4] and M. Wächter [12]. It describes the flight of a hypersonic aircraft under the objective of minimum fuel consumption. The flight trajectory is described, as usual, by a system of ordinary differential equations (ODE). However, due to the hypersonic flight conditions a thermal protection system is indispensable and must therefore be taken into account in the model. This leads to a quasi-linear heat equation with non-linear boundary conditions, which is coupled with the ODE. A major goal of the optimization is limiting the peak temperature of the thermal protection system, inducing a pointwise state constraint, which couples the PDE with the ODE reversely.

However there is little to no hope to achieve any insight in the structure of this problem and its solution due its enormous complexity, which makes anything beyond numerical analysis nigh impossible. To obtain an unobstructed view, we decided to design a model problem as simple as possible, yet containing the key features of ODE-PDE optimal control.

This model problem we would like to call the *hypersonic rocket car* problem. Its ODE-part consists of the *rocket car on a rail track* problem, first studied by Bushaw [2] (see also [3]) in the pioneering days of ODE optimal control and is augmented by a heat equation, mimicing heating due to friction.

The subject of the paper is the theoretical and numerical analysis of this prototype of an ODE-PDE optimal control problem, with the major aim to answer structural questions about the existence of state-constrained subarcs or touch points. With view on complex real-life problems we focus on first-order necessary conditions only. Even these conditions are hardly manageable for problems of this kind not to mention second-order sufficiency conditions.

The paper may also lead to new contributions in the field of state-constrained optimal control problems with PDEs in particular with view to the connection of state-constrained optimal control problems to free boundary problems, where the boundary (interface) between the active and inactive sets is treated as optimization variable; cf. Hintermüller, Ring [5] and Bechmann et. al. [1]. This approach contains the spirit of determining the junction points in ODE state-constrained problems as optimization parameters of a multi-point boundary value problem; cf. [8].

The paper is divided into two parts. Part 1 will, as a prelude and for better comparison and illustration, shed some light on the state-unconstrained problem and provide the necessary theory for Part 2 [10], which deals with the analysis and the numerics of the full state-constrained problem.

## 2 The hypersonic rocket car problems

As mentioned before the *hypersonic rocket car* problem is based on the classical *rocket car on a rail track* problem, whose aim it is to drive a rocket car from a given starting position and velocity to the origin of the phase plane by controlling its acceleration. Additionally, the cars temperature  $T$  is calculated with its speed entering the heat equation. In case of Coulomb, Stokes or Newton friction the source term of the PDE has to be proportional to the vehicle's speed, its square or cube. This temperature finally is not allowed to exceed a certain threshold  $T_{\max}$ .

For the heating we consider two cases. In the first case, it is assumed to be induced over the entire length  $l$  of the car. This leads to an ODE-PDE-constrained optimization problem with an indirectly distributed control of the PDE via the ODE state variables by a semi-linear coupling term, subsequently referred to as Problem 1. In the second case, the PDE is controlled through a type of boundary control indirectly introduced by the ODE state variables, subsequently referred to as Problem 2. Here, the heat flux is assumed to be dependent also on the car's speed according to the aforementioned friction laws.

In view of the hypersonic aircraft problem, these two problems constitute extremely simplified versions of the optimal control of the instationary heat flux into the thermal protection system of the aircraft's body, resp., at its stagnation point.

In the following, the ODE state variable  $w$  denotes the one-dimensional position of the car depending on time  $t$  with the terminal time  $t_f$  unspecified. The PDE state variable  $T$  stands for the temperature and depends on time as well as the spatial coordinate  $x$  describing the position within the car. The control is denoted by  $u$  and stands for the acceleration of the car. The PDE is controlled only indirectly via the velocity  $\dot{w}$  of the car.

The hypersonic rocket car problems are now given as follows:

$$\min_{u \in U} \left\{ t_f + \frac{1}{2} \lambda \int_0^{t_f} u^2 dt \right\}, \quad \lambda \geq 0, \quad (2.1)$$

subject to

$$\ddot{w}(t) = u(t) \quad \text{in } (0, t_f), \quad (2.2a)$$

$$w(0) = w_0, \quad \dot{w}(0) = \dot{w}_0, \quad (2.2b)$$

$$w(t_f) = 0, \quad \dot{w}(t_f) = 0, \quad (2.2c)$$

$$U := \{u \in L^2(0, t_f) : |u(t)| \leq u_{\max} \text{ almost everywhere in } [0, t_f]\}, \quad (2.2d)$$

and either

1st Problem: *Distributed control of the PDE via an ODE state variable:*

$$\frac{\partial T}{\partial t}(x, t) - \frac{\partial^2 T}{\partial x^2}(x, t) = g(\dot{w}(t)) \quad \text{in } (0, l) \times (0, t_f), \quad (2.3a)$$

$$T(x, 0) = T_0 \quad \text{on } (0, l), \quad (2.3b)$$

$$-\frac{\partial T}{\partial x}(0, t) = -(T(0, t) - T_0),$$

$$\frac{\partial T}{\partial x}(l, t) = -(T(l, t) - T_0) \quad \text{on } (0, t_f), \quad (2.3c)$$

or

*2nd Problem: Boundary control of the PDE via an ODE state variable:*

$$\frac{\partial T}{\partial t}(x, t) - \frac{\partial^2 T}{\partial x^2}(x, t) = 0 \text{ in } (0, l) \times (0, t_f), \quad (2.4a)$$

$$T(x, 0) = T_0 \text{ on } (0, l), \quad (2.4b)$$

$$-\frac{\partial T}{\partial x}(0, t) = -(T(0, t) - T_0),$$

$$\frac{\partial T}{\partial x}(l, t) = -(T(l, t) - T_0) + h(\dot{w}(t)) \text{ on } (0, t_f) \quad (2.4c)$$

with  $T_0 \in \mathbb{R}$ ,  $T_0 \geq 0$ ,  $u_{\max} \in \mathbb{R}$ ,  $u_{\max} > 0$ , and non-negative continuously differentiable functions  $g$  and  $h \in C^\infty(\mathbb{R} \setminus \{0\}) \cap C(\mathbb{R})$  satisfying  $g(z) = 0$  and  $h(z) = 0$ , if and only if  $z = 0$ , and finally subject to a pointwise state constraint of type

$$T(x, t) \leq T_{\max} \text{ in } [0, l] \times [0, t_f], \quad T_{\max} = \text{const} > 0. \quad (2.5)$$

Physically reasonable choices for  $g$  and  $h$  are, for example,  $z \mapsto |z|^n$ ,  $n = 1, 2, 3$ , according to Coulomb, Stokes, and Newton friction. Negative values for  $z$ , i. e., when the rocket car drives backward, require some additional devices of the rocket car, in order to preserve the physical interpretability, such as a front and a rear engine in case of Problem 1 or the ability of instantaneous turns. (Take it with a grain of salt.)

### 3 The state-unconstrained hypersonic rocket car problems

For better illustration and to allow comparison with the results of the full-blown ODE-PDE problem lets first have a brief look at the pure ODE problem:

#### 3.1 Solution of the ODE-part. Inactive state constraint

Rewriting the second-order ODE as a system of two first-order ODEs in  $\mathbf{w} := (w_1, w_2) := (w, \dot{w})$ ,  $\dot{w}_1 = w_2$ ,  $\dot{w}_2 = u$ , and defining the Hamiltonian by

$$H(\mathbf{w}, \mathbf{p}, u) = 1 + \frac{1}{2} \lambda u^2 + p_1 w_2 + p_2 u,$$

the ODE adjoint equations are

$$\dot{p}_1 = -H_{w_1} = 0, \quad \dot{p}_2 = -H_{w_2} = -p_1.$$

The minimum principle yields

$$u(t) = P_{[-u_{\max}, u_{\max}]} \left( -\frac{1}{\lambda} p_2 \right),$$

where  $P_{[a,b]}(z) := \min \{b, \max \{a, z\}\}$  denotes the projection of  $\mathbb{R}$  onto the interval  $[a, b]$ .

Furthermore, the terminal time  $t_f$  is determined by

$$H|_{t_f} = 1 + \frac{1}{2} \lambda u(t_f)^2 + p_1(t_f) \underbrace{w_2(t_f)}_{=0} + p_2(t_f) u(t_f) = 0.$$

An elementary calculation shows that  $p_2(t) = -C(t - t_f) + p_2(t_f)$  with  $p_1(t) = C = \text{const}$ . Hence  $p_2$  is linear; cf. [9]. Due to this linearity, the optimal control law

induces at most two junction points, when taking box constraints on the control  $u$  into account. If  $\lambda$  is sufficiently small, the switching structure generally consists either of the three subarcs  $u(t) = u_{\max}$ ,  $u(t) = u^{\text{free}}(t) := -\frac{1}{\lambda} p_2(t)$ , and  $u(t) = -u_{\max}$  with the two junction points  $t_1$  and  $t_2$  or the other way around. Herewith, an analytical solution of the state-unconstrained problem can be obtained, however the necessary computations are rather longsome yet basic and are therefore omitted here. Their results are shown in the phase diagrams in Fig. 3.1 for  $\lambda = 0$  and in Fig. 3.2 for  $\lambda = 10^{-1}$ .

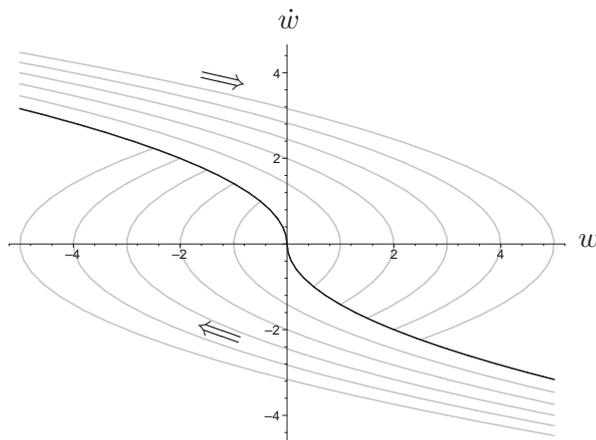


Figure 3.1: Optimal trajectories of the minimum-time problem ( $\lambda = 0$ ,  $|u| \leq u_{\max} = 1$ ) in the phase plane (grey). The black S-shaped curve is the switching curve, above which and on the left part of which the optimal (bang-bang) control is  $u = -1$  and below which and on the right part of which it is  $u = 1$ .

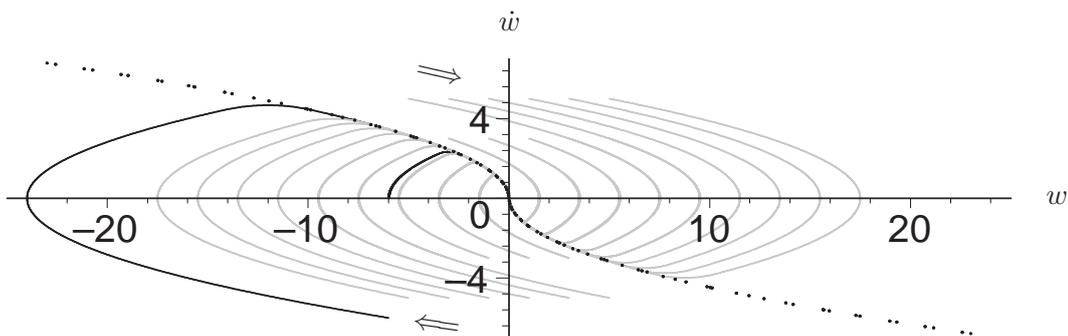


Figure 3.2: Optimal trajectories of the regularized minimum-time problem ( $\lambda = 10^{-1}$ ,  $|u| \leq u_{\max} = 1$ ) in the phase plane (grey). The dotted black is the switching curve of Fig. 3.1 and serves as envelope curve here. The two other black curves are the optimal solutions for the starting conditions  $w_0 = -6$  and  $\dot{w}_0 = 0$  resp.  $w_0 = -6$  and  $\dot{w}_0 = -6$ . Those will be picked up again later on. Of course, the optimal control is not bang-bang now, but has a sharp transition phase instead, from  $u = -1$  to  $u = 1$ , resp. vice versa in the vicinity of the switching curve.

### 3.2 Solution of the heat equation. Inactive state constraint

Furthermore it is also possible to derive a semi-analytical solution of the heat equation for both problems (of course still depending on the input  $\dot{w} = w_2$  from the ODE).

Without loss of generality, we may choose  $T_0 := 0$ . Additionally, in the second problem a transformation to homogeneous Robin conditions can be achieved by  $\hat{T} := T - \frac{1}{2} e^{x-l} h(\dot{w})$ . Then the second problem reads, again replacing  $\hat{T}$  by  $T$ , as

$$T_t(x, t) - T_{xx}(x, t) = \frac{1}{2} e^{x-l} \left( h(\dot{w}(t)) - \frac{d}{dt} h(\dot{w}(t)) \right) \text{ in } (0, l) \times (0, t_f), \quad (3.1a)$$

$$T(x, 0) = -\frac{1}{2} e^{x-l} h(\dot{w}(0)) \text{ on } (0, l), \quad (3.1b)$$

$$-T_x(0, t) = -T(0, t), \quad T_x(l, t) = -T(l, t) \text{ on } (0, t_f). \quad (3.1c)$$

Considering the homogeneous parts of the PDEs of the two problems, a separation of variables,  $T(x, t) = \xi(x) \tau(t)$ , leads to the eigenvalue problem

$$\xi'' + \mu \xi = 0 \quad \text{with} \quad 0 < \mu =: k^2$$

with the associated boundary conditions  $-\xi'(0) = -\xi(0)$  and  $\xi'(l) = -\xi(l)$ . In addition, one obtains a differential equation for  $\tau$ ,  $\dot{\tau} + \mu \tau = 0$ . For both problems, the analysis yields the same eigenfunctions

$$\varphi_n(x) = k_n \cos k_n x + \sin k_n x$$

and the same eigenvalues determined by either

$$\frac{2k}{k^2 - 1} = \tan k l \quad (\text{if } l \neq \frac{1}{2k} (2m - 1) \pi) \quad (3.2a)$$

or

$$\frac{k^2 - 1}{2k} = \cot k l \quad (\text{if } l \neq \frac{1}{2k} (2m + 1) \pi). \quad (3.2b)$$

They satisfy the asymptotic behaviour

$$\lim_{n \rightarrow \infty} \left( \mu_n - (n - 1)^2 \left( \frac{\pi}{l} \right)^2 \right) = 0.$$

Their dependence on  $l$  will furtheron be suppressed.

Because of the symmetry of the boundary conditions, these eigenfunctions are orthogonal, but not normalized with respect to the usual Hilbert space scalar product. To facilitate further calculations, we mainly use the normalized eigenfunctions from now on,

$$\phi_n(x) = \frac{1}{N_n} (k_n \cos k_n x + \sin k_n x)$$

$$\text{with } N_n^2 := \|\varphi_n\|_{L^2(0,l)}^2 = \int_0^l \varphi_n^2(x) dx = \frac{l}{2} k_n^2 + \frac{l}{2} + 1.$$

By Fourier's method one finally obtains the following solutions, which can be easily evaluated:

For the 1st problem, Eqs. (2.3):

$$\begin{aligned}
T(x, t) &= \sum_{n=1}^{\infty} \left[ \int_0^t g(\dot{w}(s)) e^{-k_n^2(t-s)} ds \right] \cdot \left( \int_0^l \phi_n(y) dy \right) \phi_n(x) \\
&= \sum_{n=1}^{\infty} \left[ \int_0^t g(\dot{w}(s)) e^{-k_n^2(t-s)} ds \right] \\
&\quad \cdot \frac{1}{N_n} \left[ \sin k_n l + \frac{1}{k_n} (1 - \cos k_n l) \right] \phi_n(x). \tag{3.3}
\end{aligned}$$

For the 2nd problem, Eqs. (3.1):

$$\begin{aligned}
T(x, t) &= -\frac{1}{2} e^{-l} \sum_{n=1}^{\infty} \left[ h(\dot{w}(0)) e^{-k_n^2 t} - \int_0^t (h(\dot{w}(s)) - \frac{d}{ds} h(\dot{w}(s))) e^{-k_n^2(t-s)} ds \right] \\
&\quad \cdot \left( \int_0^l e^y \phi_n(y) dy \right) \phi_n(x) + \frac{1}{2} h(\dot{w}(t)) e^{x-l} \tag{3.4a}
\end{aligned}$$

$$= \frac{1}{2} \sum_{n=1}^{\infty} \left[ (1 + k_n^2) \int_0^t h(\dot{w}(s)) e^{-k_n^2(t-s)} ds \right] \frac{1}{N_n} \sin k_n l \phi_n(x) \tag{3.4b}$$

The representation (3.4b) is obtained by partial integration and explicit computation of the  $y$ -integral. Despite of its compact form, the series representation (3.4b) converges slower compared to (3.4a). This is caused by the  $\mathcal{O}(n^2)$ -term  $(1 + k_n^2)$ . In Theorem 4.14, this will later be explained in more detail. Moreover, the jump discontinuities of  $\ddot{w} = u$  always cancel each other in (3.4b).

Herewith one can compute the temperature profiles for the trajectories of Fig. 3.2 for the Problems 1 and 2 (In all computations the length  $l$  of the car has been set to 1); see Figs. 3.3 and 3.4.

All graphical results shown here were obtained by symbolic computations using the mathematical and analytical software package MAPLE 10.

Please note that for Problem 1 the temperature is symmetric in space with respect to  $x = \frac{l}{2}$ , where the spatial maximum for each time  $t$  is situated. For Problem 2 there is unfortunately no such symmetry, but one can observe that the absolute maximum of the temperature is always at  $x = l$ . The latter implies that the coordinate  $x^*$  of the maximum of  $T$  for arbitrary  $t \in [0, t_f]$  satisfies  $x^* < l$ , if and only if the state constraint is not active. Those observations will be proven in the course of the next chapter and will come in very handy in Part 2 of the paper.

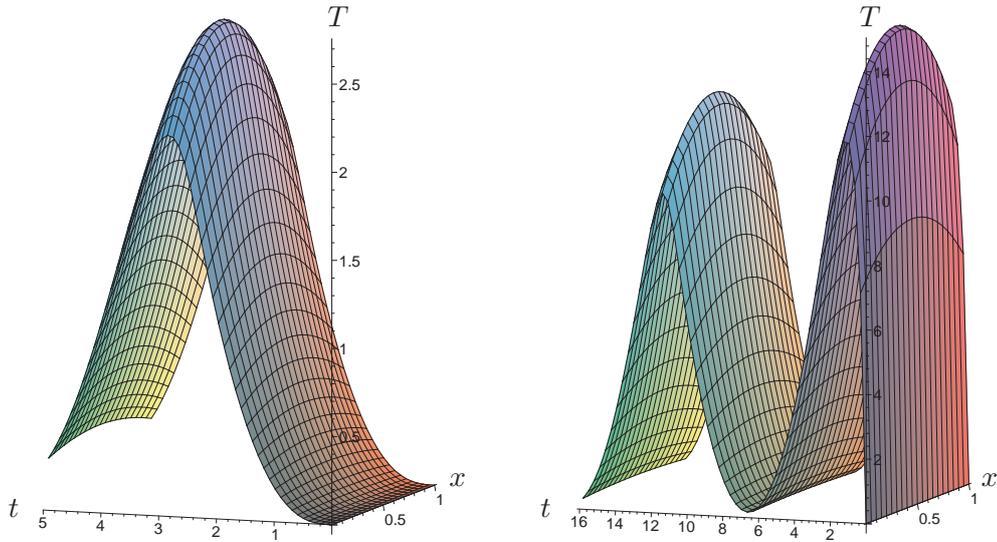


Figure 3.3: Temperature profiles for Problem 1 along state-unconstrained trajectories; cf. Fig. 3.2. Data:  $\lambda = 10^{-1}$ ,  $u_{\max} = 1$ ,  $w_0 = -6$ ,  $\dot{w}_0 = 0$  (left), resp.  $\dot{w}_0 = -6$  (right), and  $g(z) := z^2$ .

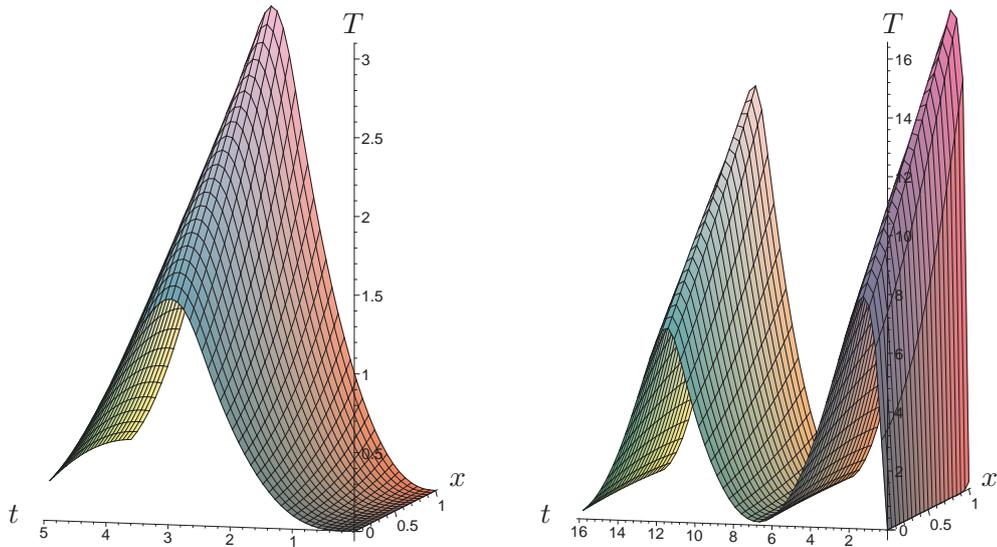


Figure 3.4: Temperature profiles for Problem 2 along state-unconstrained trajectories; cf. Fig. 3.2. Data:  $\lambda = 10^{-1}$ ,  $u_{\max} = 1$ ,  $w_0 = -6$ ,  $\dot{w}_0 = 0$  (left), resp.  $\dot{w}_0 = -6$  (right), and  $h(z) := z^2$ .

#### 4 Theoretical analysis of the coupled ODE-PDE system

In the following we summarize some properties, the solutions of the two different ODE-PDE systems have, independent of the choice of the control  $u \in U$ . Firstly, we present the main result:

**Theorem 4.1** *Let  $g$  and  $h \in C(\mathbb{R})$ . Let  $t_f > 0$  be fixed. Then Problems 1 and 2 given by Eqs. (2.2a), (2.2b), (2.3), resp. (2.2a), (2.2b), (2.4) possess each one and only one solution  $(w, T)$  in the space  $W_\infty^1(0, t_f) \times (W(0, t_f) \cap C(\bar{Q}))$  for all  $u \in U$ .*

*Moreover, the solutions depend continuously on the data, for example,*

$$\begin{aligned} & \|w\|_{W_\infty^1(0, t_f)} + \|T\|_{W(0, t_f)} + \|T\|_{C(\bar{Q})} \\ & \leq c_{\text{ODE}} (|w_0| + |\dot{w}_0| + \|u\|_{L^\infty(0, t_f)}) + c_{\text{PDE}} (\|g(\dot{w})\|_{L^\infty(0, t_f)} + T_0) \end{aligned}$$

*with constants  $c_{\text{ODE}}$ , resp.  $c_{\text{PDE}}$  independent of  $u$ , resp.  $g$  or  $h$ .*

As usual we define the Banach spaces  $W_p^m(0, t_f) := \{w \in L^p(0, t_f) : \frac{d^m}{dt^m} w \in L^p(0, t_f)\}$ . Moreover,  $W(0, t_f) := \{y \in L^2(0, t_f; V) \mid y' \in L^2(0, t_f; V^*)\}$  denotes the Hilbert space of square-integrable abstract functions with values in  $V := W_2^1(0, l)$  and (distributional) derivative  $y' \in L^2(0, t_f; V^*)$  with values in its dual  $V^*$ , see [11, p. 146].  $Q$  stands for the space-time cylinder  $(0, l) \times (0, t_f)$ .

*Proof.* According to standard theorems of existence, uniqueness and continuous dependence on data, the ODE (2.2a), (2.2b) possesses one and only one solution for all  $u \in U \subset L^\infty(0, t_f)$ , which depends continuously on the data:

$$\|w\|_{W_\infty^1(0, t_f)} \leq c_{\text{ODE}} (|w_0| + |\dot{w}_0| + \|u\|_{L^\infty(0, t_f)}) ,$$

since

$$\begin{aligned} \ddot{w}(t) = u(t) & \implies \dot{w}(t) = \dot{w}_0 + \int_0^t u(s) \, ds \implies |\dot{w}(t)| \leq |\dot{w}_0| + t_f \cdot \|u\|_{L^\infty(0, t_f)} \\ & \implies \|\dot{w}\|_{L^\infty(0, t_f)} \leq c (|\dot{w}_0| + \|u\|_{L^\infty(0, t_f)}) . \end{aligned}$$

Analogously we obtain  $\|w\|_{L^\infty(0, t_f)} \leq c (|w_0| + \|\dot{w}\|_{L^\infty(0, t_f)})$ .

Here and in the following,  $c$  always denotes a generic constant. Due to  $\dot{w} \in L^\infty(0, t_f)$  and  $g \in C(\mathbb{R})$ , we conclude  $g(\dot{w}(t)) \in L^\infty(0, t_f)$ .

Then, according to [11], Theorem 7.12, p. 378, both initial-boundary value problems possess unique solutions in the spaces stated. Moreover, the solutions depend continuously on all data of the problem, e. g. for Problem 1,

$$\|T(\cdot, t)\|_{W(0, t_f)} + \|T\|_{C(\bar{Q})} \leq c (\|g(\dot{w})\|_{L^\infty(0, t_f)} + T_0) .$$

Combining the two estimates completes the proof.  $\square$

**Remarks 4.2** *Theorem 4.1 gives rise to continuous, generally non-linear solution operators*

$$S: U \rightarrow W_\infty^1(0, t_f) \times W(0, t_f) \cap C(\bar{Q}), \quad u \mapsto (w, T) .$$

*Note that the nonlinearity of  $S$  is solely induced by the nonlinearities of  $g$ , resp.  $h$ . In the second part of this paper, we will show that  $S$  is even continuously Fréchet differentiable which will be used to prove the existence of a Lagrange multiplier associated with the state constraint (2.5).*

*The solutions are explicitly given by the formulae (3.3) and (3.4).*

*If the terminal time is unprescribed, we can linearly transform the time interval  $[0, t_f]$  into a fixed interval  $[0, 1]$  for a normalized time. This introduces a factor  $\frac{1}{t_f}$  in front of each time derivative of the ODE-PDE system. By a boundedness assumption on the optimal time  $t_f \leq t_f^{\max}$ , the proof applies accordingly.*

#### 4.1 Problem 1: Distributed control of the PDE by the ODE

Next we will show that solutions of Problem 1, Eqs. (2.3), are positive in  $[0, l] \times (0, t_f]$  and possess strong maxima w. r. t.  $x$  for all  $t > 0$  on the vertical line  $x = \frac{l}{2}$ . Moreover, we will investigate their regularity properties.

**Assumptions 4.3** Let  $T(x, t)$  be a solution of Problem 1, Eqs. (2.3), with  $T_0 = 0$  in  $[0, l] \times [0, t_f]$  and with the following properties,

(A1)  $T$  is continuous in  $[0, l] \times [0, t_f]$ ,  $\partial_x^i T$ ,  $i = 1, \dots, 3$ , as well as  $\partial_x^i T_t$ ,  $i = 0, 1$ , exist and are continuous in  $[0, l] \times (0, t_f]$ ;

(A2)  $T_x$  possesses a continuation in  $C^0([0, l] \times [0, t_f])$ , also denoted by  $T_x$ .

**Theorem 4.4** Under assumptions 4.3 there holds:

- a) If  $g(\dot{w}) > 0$  for  $[0, l] \times (0, t_f]$ , then  $T \geq 0$  in  $[0, l] \times [0, t_f]$ .
- b) If  $g(\dot{w}) > 0$  for  $[0, l] \times (0, t_f]$ , then  $T > 0$  in  $[0, l] \times (0, t_f]$ .

*Proof.* Part a). We assume that  $T$  has a negative minimum in  $(x_0, t_0) \in [0, l] \times [0, t_f]$ . Then  $t_0 > 0$ .

(i) Let  $(x_0, t_0)$  be an interior point of  $[0, l] \times (0, t_f]$ . Because of  $T_t(x_0, t_0) = T_x(x_0, t_0) = 0$ , Eq. (2.3a) yields  $T_{xx}(x_0, t_0) < 0$ , which contradicts to  $(x_0, t_0)$  being a minimizer. For,  $T(x, t_0) = T(t_0, x_0) + \frac{1}{2} T_{xx}(x_0, t_0) (x - x_0)^2 + \mathcal{O}((x - x_0)^3)$ .

(ii) Let  $x_0 = 0$ . Then Eq. (2.3c) yields  $T(x, t_0) = T(0, t_0) + T_x(0, t_0)x + \mathcal{O}(x^2) = T(0, t_0)(1 + x) + \mathcal{O}(x^2)$ , which also leads to a contradiction.

(iii) Let  $x_0 = l$ . Then Eq. (2.3c) yields  $T(x, t_0) = T(l, t_0) + T_x(l, t_0)(x - l) + \mathcal{O}((x - l)^2) = T(l, t_0)(1 + l - x) + \mathcal{O}((x - l)^2)$ , which leads to a contradiction, too.

(iv) Let  $t_0 = t_f$  and  $0 < x_0 < l$ . Then we have  $T_t(x_0, t_f) \leq 0$ . For,  $T_t(x_0, t_f) > 0$  would imply  $T(x_0, t) = T(x_0, t_f) + T_t(x_0, s)(t - t_f) < T(x_0, t_f)$  for  $t < s < t_f$  and  $t_f - t$  sufficiently small, thus a contradiction.

Therefore, by Eq. (2.3a) there must hold  $T_{xx}(x_0, t_f) < 0$ . The expansion  $T(x, t_f) = T(x_0, t_f) + \frac{1}{2} T_{xx}(x_0, t_f) (x - x_0)^2 + \mathcal{O}((x - x_0)^3)$  then leads again to a contradiction.

Hence,  $T$  cannot have a negative minimum.  $\square$

*Proof.* Part b). We assume that  $T$  has a minimum in  $(x_0, t_0) \in [0, l] \times (0, t_f]$ , which equals zero. As in the proof of Part a), (i), (iv), we can immediately exclude  $0 < x_0 < l$ .

Let  $x_0 = 0$ . Because of  $T(0, t_0) = 0$  and Eq. (2.3c), there holds  $T(x, t_0) = \frac{1}{2} T_{xx}(0, t_0) x^2 + \mathcal{O}(x^3)$ . Since Eq. (2.3a) yields  $T_{xx}(0, t_0) < 0$ , the last equation leads to a contradiction. We proceed analogously if  $x_0 = l$ .  $\square$

**Theorem 4.5** Under assumptions 4.3 there holds  $T(x, t) = T(l - x, t)$ .

*Proof.* Under the assumptions (A1), Problem 1 (2.3) possesses one and only one solution, since  $\mathcal{L}$  defined by

$$\mathcal{L}y := -y_{xx},$$

$$\mathcal{D}(\mathcal{L}) := \{y : y \in H^2(0, l), (-y_x + y)(0) = 0, (y_x + y)(l) = 0\}$$

is a positive definite self-adjoint operator in  $L^2(0, l)$ . Here,  $H^2(0, l)$  denotes the Hilbert space of all functions having square integrable weak derivatives up to order 2.

If  $T$  solves Problem 1, then  $\hat{T}(x, t) := T(l - x, t)$  solves it, too.  $\square$

**Theorem 4.6** Under assumptions 4.3 it holds:  $T$  takes its strong maximum in  $x = \frac{l}{2}$  for each  $t_0 \in (0, t_f)$ .  $T(x, t_0)$  increases strictly monotonic in  $[0, \frac{l}{2}]$  and decreases strictly monotonic in  $[\frac{l}{2}, l]$ .

*Proof.* Because of Theorem 4.5 there holds

$$T_x\left(\frac{l}{2}, t\right) = 0, \quad 0 \leq t \leq t_f.$$

Since there holds  $T_{xt} - T_{xxx} = 0$ ,  $T_x$  takes its maximum and minimum on the parabolic boundary of  $[0, \frac{l}{2}] \times [0, t_f]$ . Because of Theorem 4.4 a) and Eq. (2.3b), we obtain  $T_x(x, t) \geq 0$  in  $[0, \frac{l}{2}] \times [0, t_f]$ . Because of Theorem 4.4 b) and Eq. (2.3c),  $T_x(0, t_0) > 0$ . Hence,  $T$  at first increases strictly monotonic.

Let us assume that  $T_x(x, t_0)$  possesses its first zero in  $x_0$ ,  $0 < x_0 < \frac{l}{2}$ . There holds  $T_x$  is real analytic in  $(0, l) \times (0, t_f)$ , i. e.,

$$T_x(x, t_0) = T_{xx}(x_0, t_0)(x - x_0) + \sum_{k=2}^{\infty} \frac{1}{k!} \partial_x^k T_x(x_0, t_0) (x - x_0)^k \quad \text{in } |x - x_0| < \varepsilon.$$

Because of  $T_x(x, t_0) > 0$  in  $x_0 - \varepsilon < x < x_0$ , the first non-vanishing coefficient of the preceding expansion must have the form  $\frac{1}{(2l)!} \partial_x^{2l} T_x(x_0, t_0) > 0$ . Hence  $T(x, t_0)$  passes strictly monotonic increasing through  $(x_0, t_0)$ . Since only a finite number of zeros can lie in each interval  $(\eta, l - \eta)$ , the proposition follows immediately.  $\square$

**Theorem 4.7** Let  $T$  be the unique weak solution of

$$T_t + \mathcal{L}T = g(\dot{w}), \quad (4.1a)$$

$$T(0) = 0, \quad (4.1b)$$

i. e.,  $T \in C^0([0, t_f], L^2(0, l)) \cap L^2(0, t_f; \mathcal{D}(\mathcal{L}^{\frac{1}{2}}))$ , such that

$$-\int_0^{t_f} (T, \psi_t) ds + \int_0^{t_f} (\mathcal{L}^{\frac{1}{2}} T, \mathcal{L}^{\frac{1}{2}} \psi) ds = \int_0^{t_f} (g(\dot{w}), \psi) ds + (T(0) = 0, \psi(0)) \quad (4.2)$$

for all test functions  $\psi \in C^1([0, t_f], \mathcal{D}(\mathcal{L}^{\frac{1}{2}}))$  with  $\psi(t_f) = 0$ .

Then

$$T(t) = \int_0^t e^{-(t-s)\mathcal{L}} g(\dot{w}(s)) \, ds, \quad (4.3)$$

and  $T$  has the following additional regularity properties:

$$T(t) \in \mathcal{D}(\mathcal{L}), \quad 0 \leq t \leq t_f, \quad (4.4a)$$

$$\mathcal{L}T \in C^0([0, t_f], L^2(0, l)), \quad (4.4b)$$

$$T \in C^1([0, t_f], L^2(0, l)), \quad (4.4c)$$

$$T \in \bigcap_{\eta, 0 \leq \eta < \frac{1}{2}} C^0((0, t_f], C^{3+\eta}([0, l])), \quad (4.4d)$$

$$T_t \in \bigcap_{\eta, 0 \leq \eta < \frac{1}{2}} C^0((0, t_f], C^{1+\eta}([0, l])). \quad (4.4e)$$

In particular,  $T$  is a classical solution of (4.1) in  $[0, l] \times (0, t_f]$  and satisfies (4.1) in the strong sense, i. e.,  $\frac{d}{dt}T + \mathcal{L}T = g(\dot{w})$  in  $[0, t_f]$ .

**Remarks 4.8** As for the definition of a weak solution, we refer to Lions [7], p. 44. There it is shown that the weak solution uniquely exists for right hand sides in (4.1) having even weaker regularity properties than  $g(\dot{w})$ .

*Proof.* Since  $\ddot{w} = u \in L^\infty(0, t_f)$ , the term  $g(\dot{w})$  is Lipschitzian from  $[0, t_f]$  into  $\mathbb{R}$ . Thus

$$g(\dot{w}) \in C^{0,1}([0, t_f], L^2(0, l)),$$

and integration by parts in (4.3) yields

$$T(t) = \mathcal{L}^{-1}g(\dot{w}(t)) - \mathcal{L}^{-1}e^{-t\mathcal{L}}g(\dot{w}(0)) - \int_0^t \mathcal{L}^{-1}e^{-(t-s)\mathcal{L}} \frac{d}{dt}g(\dot{w})|_{t=s} \, ds. \quad (4.5)$$

This implies

$$\mathcal{L}T(t) = g(\dot{w}(t)) - e^{-t\mathcal{L}}g(\dot{w}(0)) - \int_0^t e^{-(t-s)\mathcal{L}} \frac{d}{dt}g(\dot{w})|_{t=s} \, ds,$$

$$\mathcal{L}T \in C^0([0, t_f], L^2(0, l)),$$

cf. [13], Sätze IV. 1.1 and IV. 1.2. In particular, the properties (4.4a, b, c) are satisfied and  $\frac{d}{dt}T + \mathcal{L}T = g(\dot{w})$  in  $[0, t_f]$ . Thus, the unique weak solution of (4.1) is given by (4.3). Since  $g(\dot{w})$  does not depend on  $x$ , we have

$$g(\dot{w}) \in C^0([0, t_f], C^2([0, l])).$$

Moreover,

$$\mathcal{L}^\kappa e^{-t\mathcal{L}}g(\dot{w}(0)) \in C^0([0, t_f], L^2(0, l)), \quad \kappa \in \mathbb{N},$$

thus

$$e^{-t\mathcal{L}}g(\dot{w}(0)) \in C^\infty([0, l] \times (0, t_f]) .$$

Since for any  $\varrho \in (0, 1]$  it holds

$$\mathcal{L}^{1-\varrho} \int_0^t e^{-(t-s)\mathcal{L}} \frac{d}{dt}g(\dot{w})|_{t=s} ds = \int_0^t \mathcal{L}^{1-\varrho} e^{-(t-s)\mathcal{L}} \frac{d}{dt}g(\dot{w})|_{t=s} ds$$

with an integrable and in  $0 \leq s < t$  continuous kernel, standard imbeddings furnish properties (4.4d, e).  $\square$

**Remarks 4.9** Property (4.4b) implies the continuity of  $T$  and  $T_x$  in  $[0, l] \times [0, t_f]$  as needed in Theorems 4.4 – 4.6; see Assumption 4.3.

Next we deal with  $T_{tt}$  and  $\partial_x^4 T$ . To that end the concept of maximal regularity for parabolic equations can be used to advantage. We refer to [13], Chapter V. The functional-analytic version of this concept, as explained in this reference, is most easily applied to the present problem. First, we remark that  $\frac{d}{dt}g(\dot{w}(t)) = g'(\dot{w}(t))\ddot{w}(t)$ . Thus the chain rule holds. This is seen by taking appropriate approximations of  $w$ .

**Theorem 4.10** For any  $\varepsilon$ ,  $0 < \varepsilon < t_f$ , and any  $r \geq 2$  we have

$$\begin{aligned} T_{tt} &\in L^r(\varepsilon, t_f; L^2(0, l)) , \\ \partial_x^4 T &\in L^r(\varepsilon, t_f; L^2(0, l)) . \end{aligned}$$

*Proof.* Let  $\varepsilon \in (0, t_f)$ . Let  $\zeta = \zeta_\varepsilon \in C^2([0, +\infty), \mathbb{R}^+)$  with  $\zeta(t) = 0$  in  $[0, \frac{\varepsilon}{2}]$ ,  $\zeta(t) = 1$  for  $t \geq \varepsilon$ . Let  $r \geq 2$ . Then

$$\begin{aligned} Z_t + \mathcal{L}Z &= \zeta_t g(\dot{w}) + \zeta g'(\dot{w})\ddot{w} - \zeta_{tt}T - \zeta_t T_t , \\ Z(0) &= 0 \end{aligned}$$

has a unique solution  $Z$  with

$$\begin{aligned} Z_t &\in L^r(0, t_f; L^2(0, l)) , \\ Z &\in L^r(0, t_f; \mathcal{D}(\mathcal{L})) ; \end{aligned}$$

cf. [13], Satz V.2.5. Setting

$$z = \int_0^t Z ds ,$$

we obtain

$$z_t + \mathcal{L}z = \zeta g(\dot{w}) - \zeta_t T ,$$

and, by uniqueness,  $z = \zeta T$ ,  $Z = (\zeta T)_t$ .  $\square$

## 4.2 Problem 2: Boundary control of the PDE by the ODE

We now deal with the second problem (2.4). As for the regularity question, the form (3.1) is more favourable for a first step. We start with these equations. In order to distinguish between  $T$  in (2.4) and  $T$  in (3.1) we set

$$\hat{T} = T - \frac{1}{2} e^{x-l} h(\dot{w}) \hat{=} T \quad \text{in Eqs. (3.1),}$$

$$T \quad \text{as in Eqs. (2.4).}$$

Let us mention the interpolation space

$$\mathcal{I}_{1-\frac{1}{r},r} := \left\{ y : \int_0^\infty \frac{1}{t^r} \| (e^{-t\mathcal{L}} - I) y \|^r dt < \infty \right\}$$

as introduced in [13], Definition V.2.1, for any  $r > 1$ . Then  $\mathcal{I}_{1-\frac{1}{r},r}$  is close to  $\mathcal{D}(\mathcal{L}^{1-\frac{1}{r}})$ . More precisely, for any  $\varepsilon > 0$  we have

$$\mathcal{D}(\mathcal{L}^{1-\frac{1}{r}+\varepsilon}) \subset \mathcal{I}_{1-\frac{1}{r},r} \subset \mathcal{D}(\mathcal{L}^{1-\frac{1}{r}-\varepsilon}), \quad (4.6)$$

$$\mathcal{I}_{1-\frac{1}{2},2} = \mathcal{D}(\mathcal{L}^{\frac{1}{2}})$$

with continuous imbeddings and equivalent norms, respectively; cf. [13], Chapter V., pp. 50, 59, Satz V.2.1. The appropriate first regularity theorem for (3.1) is now at hand.

**Theorem 4.11** *There is one and only one solution*

$$\hat{T} \in \bigcap_{r, r \geq 2} C^0([0, t_f], \mathcal{I}_{1-\frac{1}{r},r})$$

with

$$\hat{T}_t \in \bigcap_{r, r \geq 2} L^r(0, t_f; L^2(0, l)),$$

$$\hat{T} \in \bigcap_{r, r \geq 2} L^r(0, t_f; \mathcal{D}(\mathcal{L})),$$

$$\hat{T}_t(t) + \mathcal{L}\hat{T}(t) = \frac{1}{2} e^{x-l} \left( h(\dot{w}(t)) - \frac{d}{dt} h(\dot{w}(t)) \right) \quad \text{a. e. in } (0, t_f), \quad (4.7)$$

$$\hat{T}(x, 0) = -\frac{1}{2} e^{x-l} h(\dot{w}(0)) \quad \text{on } [0, l].$$

*Proof.*  $\hat{T}(x, 0)$  is in  $\mathcal{D}(\mathcal{L})$ . The right hand side in (3.1a) is in  $L^\infty(0, t_f; L^\infty(0, l))$  since  $h(\dot{w})$  has the distributional derivative  $h'(\dot{w}) \ddot{w} = h'(\dot{w}) u$ . Satz V.2.5 in [13] completes the proof.  $\square$

In the next step we again apply the maximum principle, now to Eqs. (2.4).

**Theorem 4.12** *T as well as  $T_x$  of problem (2.4) are continuous in  $[0, l] \times [0, t_f]$ . Moreover,  $T - T_0 \geq 0$  in  $[0, l] \times [0, t_f]$  and assumes its global maximum in  $[0, l] \times [0, t_f]$  on  $x = l$ .*

*Proof.* Equation (4.6), Theorem 4.11, and standard imbeddings show that  $T$  and  $T_x$  are continuous in  $[0, l] \times [0, t_f]$ . Indeed,  $T_x$  is even Hölder continuous in  $x$  and therefore in  $t$ ; cf. [6], pp. 78f. In particular,  $T$  is a classical solution of  $T_t - T_{xx} = 0$  in  $(0, l) \times (0, t_f)$  and therefore even real analytic there.

Assume now that  $T - T_0$  has a negative minimum in  $[0, l] \times [0, t_f]$ . Due to the maximum principle this is assumed on the parabolic boundary of  $[0, l] \times [0, t_f]$ . The only possibilities are  $x = 0$  and  $x = l$ . At a minimum point  $(0, t_0)$  on  $x = 0$ , we have, because of  $T(0, t_0) - T_0 < 0$ ,

$$\frac{\partial(T - T_0)}{\partial x}(0, t_0) = T(0, t_0) - T_0 < 0,$$

and, at a minimum point  $(l, t_0)$  on  $x = l$ , it holds, for the same reason,

$$\frac{\partial(T - T_0)}{\partial x}(l, t_0) = -(T(l, t_0) - T_0) + h(\dot{w}(t_0)) > 0.$$

However, both of these possibilities lead to a contradiction to  $T(0, t_0) - T_0$  and  $T(l, t_0) - T_0$  resp., being minimal.

Assume that  $T - T_0$  has a (global) positive maximum in  $(x_0, t_0)$ . It is assumed on  $\{0\} \times [0, t_f]$  or  $\{l\} \times [0, t_f]$ . For  $x_0 = 0$  we have

$$\frac{\partial(T - T_0)}{\partial x}(0, t_0) = T(0, t_0) - T_0 > 0$$

which is a contradiction. Hence, it must lie on  $\{l\} \times [0, t_f]$ . If the maximum of  $T - T_0$  vanishes, we have  $T \equiv T_0$ .  $\square$

**Remarks 4.13** *Note that Theorem 4.12 does not imply that  $l = \arg \max_{x \in [0, l]} T(x, t)$  for all  $t \in [0, t_f]$ .*

Finally, we investigate the convergence behaviour of the series (3.3) and (3.4), resp., representing the solutions of the two Problems (2.3) and (2.4) or (3.1), resp.

**Theorem 4.14** **a)** *The (time dependent) coefficients*

$$a_n(t) := \left[ \int_0^t g(\dot{w}(s)) e^{-k_n^2(t-s)} ds \right] \cdot \left( \int_0^l \phi_n(y) dy \right) \frac{k_n}{N_n},$$

$$b_n(t) := \left[ \int_0^t g(\dot{w}(s)) e^{-k_n^2(t-s)} ds \right] \cdot \left( \int_0^l \phi_n(y) dy \right) \frac{1}{N_n}$$

*of the series in  $\cos k_n x$  and  $\sin k_n x$  in the solution formula of Problem 1 are of order  $\mathcal{O}(n^{-4})$  uniformly in  $[0, t_f]$ . The coefficients associated with even modes vanish.*

*The series in  $\cos k_n x$  and  $\sin k_n x$  in the solution of Problem 1, Eq. (3.3), converge absolutely.*

b) *The (time dependent) coefficients*

$$\begin{aligned}\bar{a}_n(t) &:= \left[ (1 + k_n^2) \int_0^t h(\dot{w}(s)) e^{-k_n^2(t-s)} ds \right] \cdot \left( \int_0^l e^y \phi_n(y) dy \right) \frac{k_n}{N_n}, \\ \bar{b}_n(t) &:= \left[ (1 + k_n^2) \int_0^t h(\dot{w}(s)) e^{-k_n^2(t-s)} ds \right] \cdot \left( \int_0^l e^y \phi_n(y) dy \right) \frac{1}{N_n}\end{aligned}$$

of the series in  $\cos k_n x$  and  $\sin k_n x$  in the solution formula of Problem 2 are of order  $\mathcal{O}(n^{-4})$  uniformly in  $[\varepsilon, t_f]$ ,  $0 < \varepsilon \leq t_f$ .

The series in  $\cos k_n x$  and  $\sin k_n x$  in the solution of Problem 2, Eq. (3.4), converge absolutely.

*Proof.* Part a). We investigate each factor of the coefficients separately. For  $t > 0$  we have

$$0 < \tau_n(t) := \int_0^t g(\dot{w}(s)) e^{-k_n^2(t-s)} ds < c \int_0^t e^{-k_n^2(t-s)} ds < c \frac{1}{k_n^2} = \mathcal{O}(n^{-2}).$$

Furthermore we obtain

$$\frac{1}{N_n^2} = \frac{1}{\frac{l}{2} k_n^2 + \frac{l}{2} + 1} = \frac{2}{l} \frac{1}{k_n^2} \frac{1}{1 + \frac{l+2}{l} \frac{1}{k_n^2}} = \frac{2}{l} \frac{1}{k_n^2} (1 + \mathcal{O}(k_n^{-2})) = \mathcal{O}(n^{-2}),$$

if  $|\frac{l+2}{l} \frac{1}{k_n^2}| < 1$ , which can always be fulfilled for sufficiently large  $n$ , e. g. for  $n \geq 2$ , if  $l = 1$ .

Substituting (3.2b) in

$$\gamma_n := \int_0^l \phi_n(y) dy = \frac{1}{N_n} \left[ \sin k_n l + \frac{1}{k_n} (1 - \cos k_n l) \right],$$

one obtains

$$\gamma_n = \frac{1}{N_n} \left( \frac{2}{k_n} \cos k_n l + \frac{1}{k_n^2} \sin k_n l + \frac{1}{k_n} (1 - \cos k_n l) \right) = \mathcal{O}(k_n^{-2}) = \mathcal{O}(n^{-2}).$$

Because of  $a_n = \tau_n(t) \gamma_n \frac{1}{N_n} k_n$  and  $b_n = \tau_n(t) \gamma_n \frac{1}{N_n}$ , we at least have a decay of the order  $\mathcal{O}(n^{-4})$  of the coefficients of the cosine- and sine-terms of  $\varphi(x)$  uniformly in  $t$ . This also implies absolute convergence of the  $\cos k_n x$ - $\sin k_n x$  series.

Moreover, the coefficients associated with even modes vanish due to the symmetry of the eigenfunctions. This can be seen as follows. For all  $n \in \mathbb{N}$  and all  $x \in [-\frac{l}{2}, \frac{l}{2}]$ , there holds, using the addition formulas of sine and cosine,

$$\begin{aligned}\varphi_n\left(\frac{l}{2} + x\right) &= \begin{cases} +\varphi_n\left(\frac{l}{2} - x\right) \\ -\varphi_n\left(\frac{l}{2} - x\right) \end{cases} &\iff \tan k_n \frac{l}{2} = \begin{cases} \frac{1}{k_n} \\ -k_n \end{cases} \\ & &\iff \frac{2 \tan k_n \frac{l}{2}}{1 - \tan^2 k_n \frac{l}{2}} = \frac{2 k_n}{k_n^2 - 1} = \tan k_n l.\end{aligned}$$

The upper branch holds if and only if  $n$  is odd, the lower one if and only if  $n$  is even. Hence, we have, for all  $\kappa \in \mathbb{N}$ ,

$$\begin{aligned}\varphi_{2\kappa-1}\left(\frac{l}{2}+x\right) &= \varphi_{2\kappa-1}\left(\frac{l}{2}-x\right), \\ \varphi_{2\kappa}\left(\frac{l}{2}+x\right) &= -\varphi_{2\kappa}\left(\frac{l}{2}-x\right) \quad \text{for } -\frac{l}{2} \leq x \leq \frac{l}{2},\end{aligned}\tag{4.8}$$

i. e., the eigenfunctions  $\varphi_{2\kappa-1}$ , resp.  $\phi_{2\kappa-1}$  are even, the eigenfunctions  $\varphi_{2\kappa}$ , resp.  $\phi_{2\kappa}$  are odd w. r. t.  $x = \frac{l}{2}$ . Therefore we have

$$\int_0^l \phi_{2\kappa-1}(y) \, dy = 2 \int_0^{\frac{l}{2}} \phi_{2\kappa-1}(y) \, dy \quad \text{and} \quad \int_0^l \phi_{2\kappa}(y) \, dy = 0.$$

□

*Proof.* Part b). For Problem 2, we proceed analogously and use the representation (3.4a). We now define  $\bar{\tau}_n$  by

$$\bar{\tau}_n(t) := h(\dot{w}(0)) e^{-k_n^2 t} - \int_0^t \left( h(\dot{w}(s)) - \frac{d}{ds} h(\dot{w}(s)) \right) e^{-k_n^2 (t-s)} \, ds.\tag{4.9}$$

Similarly to the first part of the proof, one obtains

$$\bar{\tau}_n(t) = \mathcal{O}(e^{-n^2 t}) + \mathcal{O}(n^{-2}) \quad \text{for } t \in (0, t_f].$$

The rest of the proof proceeds analogously. By using (3.2b), one obtains

$$\bar{\gamma}_n := \int_0^l e^y \phi_n(y) \, dy = \frac{1}{N_n} e^l \sin k_n l = \frac{1}{N_n} e^l \cos k_n l \frac{2k_n}{k_n^2 - 1} = \mathcal{O}(n^{-2}),$$

which also sums up to an asymptotic decay of the coefficients of order  $\mathcal{O}(n^{-4})$  at least, but now in  $[\varepsilon, t_f]$ ,  $0 < \varepsilon \leq t_f$ . □

**Remarks 4.15** We now exploit Theorem 4.14. Since the  $\cos k_n x$ - $\sin k_n x$  series are no longer orthogonal, we need to estimate the series of the absolute values for  $T$  and its derivatives.

According to Theorem 4.7, the solution of Problem 1 is in the class  $C^0((0, t_f], C^{3+\varepsilon}([0, l]))$ ,  $0 \leq \varepsilon < \frac{1}{2}$ , together with  $T_t \in C^0((0, t_f], C^{1+\varepsilon}([0, l]))$ ,  $0 \leq \varepsilon < \frac{1}{2}$ . As for the  $\cos k_n x$ - $\sin k_n x$ -series we have absolute uniform convergence for the twice differentiated series  $T_{xx}$ .

In Problem 2 however, the  $\cos k_n x$ - $\sin k_n x$ -series improve on Theorem 4.11 for  $t > 0$ .

**Theorem 4.16** As for  $\hat{T}$  in Theorem 4.11 we have

$$\hat{T} \in \bigcap_{\varepsilon, 0 \leq \varepsilon < 1} C^0([\delta, t_f], C^{2+\varepsilon}([0, l])),$$

$$\hat{T}_t \in L^\infty(\delta, t_f; C^\varepsilon([0, l])), \quad 0 < \delta < t_f.$$

*Proof.* By Theorem 4.14 we have

$$\partial_x^2 \hat{T}(x, t) = - \sum_{n=1}^{\infty} k_n^2 (\bar{a}_n(t) \cos k_n x + \bar{b}_n(t) \sin k_n x)$$

with  $\bar{a}_n(t)$ ,  $\bar{b}_n(t)$  continuous in  $t$  and  $|\bar{a}_n(t)| \leq \frac{c}{k_n^4}$  as well as  $|\bar{b}_n(t)| \leq \frac{c}{k_n^4}$ . Now,

$$\frac{|\cos k_n x - \cos k_n x'|}{|x - x'|^\varepsilon} \leq c k_n^\varepsilon, \quad 0 \leq \varepsilon < 1, \quad x \neq x'.$$

A corresponding estimate holds for  $|\sin k_n x - \sin k_n x'|$ . This proves the first assertion. The second one follows from the equation for  $\hat{T}$ .  $\square$

## 5 Conclusions and Outlook

This paper is inspired by several optimal control problems of real-life applications with side conditions in form of staggered systems of unilaterally coupled equations of different type, such as ordinary and partial differential equations, where a pointwise state constraint on variables of the highest level regenerates the dependency to the lowest level. We have studied a class of abstract optimal control problems, called hypersonic rocket car problems, for a controlled second-order ordinary differential equation, a state variable of which controls a heat equation through semi-linear coupling terms, either via a source term or a boundary condition. In the first part of this bipartite paper, theoretical results are obtained concerning existence and uniqueness of solutions of the state-unconstrained problems as well as the existence of a continuous control-to-state operator. Moreover, the positivity of the optimal solutions, their regularity properties and the existence of global maxima are thoroughly investigated.

Part 2 of the paper is then devoted to the state-constrained versions of the hypersonic rocket car problems. In particular, we will tighten the state constraint step by step. According to Theorem 4.6, resp. Theorem 4.11, the maximum temperature is attained either along  $x = \frac{l}{2}$  (for Problem 1) or  $x = l$  (for Problem 2). By means of those double hump solutions [Figs. 3.3 (right), resp. 3.4 (right)] the set of active constraints will be analyzed. In particular, the concept of order of a state constraint known in ODE optimal control will be generalized to PDE optimal control. For this analysis, the regularity results of Theorems 4.10 and 4.16 turn out to be sufficient to perform the successive differentiation process for determining the state constraint's order. Finally, Theorem 4.14, in particular the speed of convergence of the series representations will be of importance how to obtain the numerical results of Part 2.

## References

- [1] FREY, M., BECHMANN, S., RUND, A., AND PESCH, H. J., *New Necessary Conditions for Distributed Optimal Control Problems of Linear Elliptic Equations with State Constraints*, in preparation.
- [2] BUSHAW, D. W., PhD Thesis, supervised by Solomon Lefschetz, Department of Mathematics, Princeton University, 1952, published as D. W. BUSHAW, *Differential Equations with a Discontinuous Forcing Term*. Experimental Towing Tank, Stevens Institute of Technology, Report No. 469, January 1953.

- 
- [3] BUSHAW, D. W., *Optimal Discontinuous Forcing Terms*, Contributions to the Theory of Nonlinear Oscillations IV, Annals of Mathematics Studies 41, Princeton, 29–52, 1958.
  - [4] CHUDEJ, K., PESCH, H. J., WÄCHTER, M., SACHS, G., AND LE BRAS, F., *Instationary Heat Constrained Trajectory Optimization of a Hypersonic Space Vehicle*, in: A. Frediani, G. Butazzo (Eds.), Variational Analysis and Aerospace Engineering. — Berlin: Springer, 2009.
  - [5] HINTERMÜLLER, M. AND RING, W., *A level set approach for the solution of a state-constrained optimal control problem*, Numerische Mathematik, Vol. 98, No. 1, 135–166, 2004.
  - [6] LADYZHENSKAYA, O. A., SOLONNIKOV, V. A., AND URAL'CEVA, N. N., *Linear and Quasilinear Equations of Parabolic Type*, American Mathematical Society, Providence, R. I., 1968.
  - [7] LIONS, J. L., *Equations Differentielles Operationelles et Problèmes aux Limites*, Springer, Berlin, Heidelberg, New York, 1961.
  - [8] PESCH, H. J., *A practical guide to the solution of real-life optimal control problems*, Control and Cybernetics, Vol. 23, 7–60, 1994.
  - [9] PESCH, H. J., *Schlüsseltechnologie Mathematik — Einblicke in aktuelle Anwendungen der Mathematik*, B. G. Teubner, Wiesbaden, Germany, 2002.
  - [10] PESCH, H. J., RUND, A., VON WAHL, W., AND WENDL, S., *On a Prototype Class of ODE-PDE State-constrained Optimal Control Problems. Part 2: The State-constrained Problems*, in preparation.
  - [11] TRÖLTZSCH, F., *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, American Mathematical Society Graduate Studies in Mathematics, Providence, Vol. 112, 2010.
  - [12] WÄCHTER, M., *Optimalflugbahnen im Hyperschall unter Berücksichtigung der instationären Aufheizung*, PhD Thesis, Technische Universität München, Faculty of Mechanical Engineering, Munich, Germany, 2004.
  - [13] WAHL, W. VON, *Funktionalanalysis III, Kapitel 4, Kapitel 5*, [http://www.diffgleichg.uni-bayreuth.de/de/team/von\\_Wahl\\_Wolf/script/](http://www.diffgleichg.uni-bayreuth.de/de/team/von_Wahl_Wolf/script/).



### C.3

## On Some New Phenomena in State-constrained Optimal Control if ODE as well as PDE are Involved

Hans Josef Pesch, Armin Rund, Wolf von Wahl und Stefan Wendl

Control and Cybernetics 39(3), pp. 647-660, 2010



# On some new phenomena in state-constrained optimal control if ODEs as well as PDEs are involved

Hans Josef Pesch\*    Armin Rund\*    Wolf von Wahl†    Stefan Wendl\*

## Abstract

In this paper we investigate a new class of optimal control problems with ODE as well as PDE constraints. We would like to call them "hypersonic rocket car problems", since they were inspired, on the one hand, by the well known rocket car problem from the early days of ODE optimal control, on the other hand by a recently investigated flight path trajectory optimization problem for a hypersonic aircraft.

The hypersonic rocket car problems mimic the latter's coupling structure, yet in an strongly simplified form. They can therefore be seen as prototypes of ODE-PDE control problems. Due to their relative simplicity they allow to a certain degree to obtain analytical solutions and insights into the structure of the adjoints, which would currently be unthinkable with complex real life problems.

Our main aim is to derive and verify necessary optimality conditions. Most of the obtained results bear a lot of similarities with state constrained ODE optimal control problems, yet we also observed some new phenomena.

**Key words.** Optimal control of partial differential equations, ODE-PDE-constrained optimization, state constraints, non-local state-constraints, integro-state constraints, optimal control problems for integro-differential equations, jump conditions.

---

\*Chair of Mathematics in Engineering Sciences, University of Bayreuth, D-95440 Bayreuth, Germany  
Corresponding author: [hans-josef.pesch@uni-bayreuth.de](mailto:hans-josef.pesch@uni-bayreuth.de)

†Chair of Mathematics VI (Partial Differential Equations and Mathematical Physics), University of Bayreuth, D-95440 Bayreuth, Germany



## C.4

# On a State-Constrained PDE Optimal Control Problem arising from ODE-PDE Optimal Control

Stefan Wendl, Hans Josef Pesch und Armin Rund

In: M. Diehl, F. Glineur, W. Michiels (Eds.): Recent Advances in Optimization  
and its Applications in Engineering

Springer, Heidelberg, pp. 429-438, 2010



# On a State-Constrained PDE Optimal Control Problem arising from ODE-PDE Optimal Control

S. Wendl\*      H. J. Pesch\*      A. Rund\*

## Abstract

The subject of this paper is an optimal control problem with ODE as well as PDE constraints. As it was inspired, on the one hand, by a recently investigated flight path optimization problem of a hypersonic aircraft and, on the other hand, by the so called "rocket car on a rail track"-problem from the pioneering days of ODE optimal control, we would like to call it "hypersonic rocket car problem".

While it features essentially the same ODE-PDE coupling structure as the aircraft problem, the rocket car problem's level of complexity is significantly reduced. Due to this fact it is possible not just to obtain easily interpretable results but also a certain degree of insight into the structure of the adjoints. Therefore, the rocket car problem can be seen as a prototype of an ODE-PDE optimal control problem. The main objective of this paper is the derivation of first order necessary optimality conditions.

**Key words.** Optimal control of partial differential equations, ODE-PDE-constrained optimization, state constraints

---

\*Chair of Mathematics in Engineering Sciences, University of Bayreuth, D-95440 Bayreuth, Germany  
Corresponding author: [hans-josef.pesch@uni-bayreuth.de](mailto:hans-josef.pesch@uni-bayreuth.de)



## C.5

# Optimal Control for a simplified 1D Fuel Cell Model

Armin Rund und Kurt Chudej

Mathematical and Computer Modelling of Dynamical Systems, 2011

DOI: [10.1080/13873954.2011.642389](https://doi.org/10.1080/13873954.2011.642389)



# Optimal Control for a Simplified 1D Fuel Cell Model

A. Rund\*      K. Chudej\*

## Abstract

Molten carbonate fuel cells are a promising technology for the operation of future stationary power plants. In order to enhance service life, a detailed knowledge of their dynamical behavior is essential. The possibility of fast and save load changes is important for daily operation of these power plants.

To predict the dynamical behaviour of fuel cells a hierachy of mathematical models has been developed in the past. Recently a systematic model reduction was applied to a 2D crossflow model. We present here the new 1D counterflow model and discuss a suitable discretization method. Accordingly we set up a method of optimal control following the first-discretize-then-optimize approach. Results are shown for simulation and optimal control in case of load changes.

**Key words.** Fuel cell; model reduction; numerical simulation; optimal control; PDAE model

**AMS subject classifications.** 35M33, 49M37, 65M20, 90C90

---

\*Chair of Mathematics in Engineering Sciences, University of Bayreuth, D-95440 Bayreuth, Germany  
Corresponding author: [armin.rund@uni-bayreuth.de](mailto:armin.rund@uni-bayreuth.de)