

Dissertation

Zeitabhängige optimale Steuerung der  
Wärmeleitungsgleichung mit  
Zustandsbeschränkungen  
mit Anwendung im Laserstrahlschmelzen

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

von  
Dipl.-Math. Simon Bechmann  
aus Nürnberg

1. Gutachter: Prof. Dr. Hans Josef Pesch
2. Gutachter: Prof. Dr. Anton Schiela

Tag der Einreichung: 19.12.2016

Tag des Kolloquiums: 23.06.2017



# Vorwort

## Zusammenfassung

In dieser Arbeit werden zustandsbeschränkte Optimalsteuerungsprobleme der Wärmeleitungsgleichung betrachtet. Die untersuchten Probleme weisen dabei eine spezielle Struktur auf: Die Steuerung in der rechten Seite der Wärmeleitungsgleichung und die Zustandsbeschränkung sind rein zeitabhängig. Diese besondere Struktur kann vielfach genutzt werden.

Im Theorieteil dieser Arbeit werden Regularitätsaussagen für die Lösungen der Optimalsteuerungsprobleme hergeleitet. Unter Verwendung der Technik des alternativen Optimalitätssystems kann die Stetigkeit der optimalen Steuerung für Probleme, in denen die zeitabhängige Zustandsbeschränkung aus einem ortsintegralen Term besteht, bewiesen werden, wenn die im Modell vorgegebenen Daten hinreichend regulär gewählt sind. Mit weiteren Annahmen an die Daten erhält man die absolute Stetigkeit der optimalen Steuerung auf Teilintervallen und eine höhere Regularität des Lagrange-Multiplikators zur Zustandsbeschränkung. In weiteren Untersuchungen wird die Technik des alternativen Optimalitätssystems angepasst und weiterentwickelt. So kann die Stetigkeit der optimalen Steuerung auch für Probleme nachgewiesen werden, bei denen die Daten in der Zustandsbeschränkung im Ort irregulär gewählt sind. Es können sogar allgemeinere Probleme mit rein zeitabhängigen Zustandsbeschränkungen in Operatorform betrachtet werden. Als Beispiel kann die Stetigkeit der optimalen Steuerung bei Problemen mit punktwiser Beschränkung des Zustands an einem fest vorgegebenen Ortspunkt gezeigt werden. Dies erlaubt es, eine interessante Schlussfolgerung für Optimalsteuerungsprobleme mit rein zeitabhängiger Steuerung und punktwiser Zustandsbeschränkung in Ort und Zeit zu ziehen.

Im Algorithmen- und Numerikteil dieser Arbeit wird ein neuer Optimierungsalgorithmus für die betrachteten Optimalsteuerungsprobleme entwickelt. Dieser Algorithmus stellt eine Erweiterung der primal-dualen aktive-Mengen-Strategie für diskretisierte zustandsbeschränkte Optimalsteuerungsprobleme nach Bergounioux und Kunisch dar. Zusätzlich fließen Ideen von Mehrfachschießverfahren, die bei der optimalen Steuerung gewöhnlicher Differentialgleichungen genutzt werden, ein. Die Leistungsfähigkeit des Algorithmus wird anhand mehrerer Modellprobleme getestet. Für reguläre Probleme zeigt sich eine Gitterunabhängigkeit des Verfahrens. Bei anspruchsvolleren Problemen mit noch stetiger optimaler Steuerung erhält man bei feinerer Diskretisierung eine moderate Erhöhung der Anzahl an Iterationen des Verfahrens. Insgesamt schlägt sich der Algorithmus typischerweise besser als die primal-duale aktive-Mengen-Strategie oder die aktive-Mengen-Strategie in Matlab. Allerdings sollte an dieser Stelle beachtet werden, dass das im Rahmen der Arbeit neu entwickelte Verfahren und die primal-duale aktive-Mengen-Strategie bei Modellproblemen mit Zustandsbeschränkungen der Ordnung grö-

ßer eins Konvergenzprobleme haben. Hinsichtlich der Resultate aus dem Theoriekapitel dieser Arbeit kann mithilfe der numerischen Ergebnisse gezeigt werden, dass die dort getroffenen Aussagen scharf zu sein scheinen, wenn das Gesamtintervall betrachtet wird. Andererseits sollten aber bei Problemen mit regulären Daten bessere Regularitätsaussagen auf Teilintervallen möglich sein.

Der letzte Teil der Arbeit widmet sich der optimalen Steuerung der Leistung des Lasers beim Laserstrahlschmelzen. Ein neu entwickeltes Modell beschreibt die Temperaturverteilung im Werkstück in Situationen, in denen der Laser einen Mäander am Rand eines Islands abfährt. Basierend auf diesem Modell können linear-quadratische Optimierungsprobleme zur Berechnung optimaler Steuerungsstrategien formuliert werden. Wegen zusätzlicher Beschränkungen kann der in dieser Arbeit neu entwickelte Algorithmus nicht zum Lösen der Probleme eingesetzt werden. Er wird deshalb ersetzt durch die `quadprog` Routine in Matlab. Die berechneten Strategien zur Anpassung der Leistung des Lasers werden auf den Maschinen zum 3D-Druck metallischer Bauteile getestet. Es stellt sich heraus, dass die Strategie, die die Leistung des Lasers nicht nur bei der Querfahrt im Mäander am Rande eines Islands, sondern auch auf einer kurzen Strecke zurück ins Innere des Bauteils absenkt, eine signifikante Verbesserung des Schmelzbildes mit sich bringt. Folglich kann auch eine bessere Oberflächenstruktur des Werkstücks erreicht werden.

## Abstract

In this thesis state constrained optimal control problems of the heat equation are considered. The investigated problems possess a special structure. The control in the right hand side of the heat equation and the state constraint are only time-dependent. This special structure can be used several times.

In the theoretical part of this work regularity results for the solutions of the optimal control problems are shown. Using the technique of alternative optimality systems, the continuity of the optimal control in situations in which the state constraint consists of a time-dependent spatially integrated term can be proven, if the data in the problems are chosen regularly enough. Under additional assumptions on the data the absolute continuity of the optimal control on subintervals and higher regularity of the Lagrange multiplier of the state constraint are obtained. In further investigations the technique of alternative optimality systems is adapted and enhanced. This allows to prove the continuity of the optimal control for problems with higher spatial irregularity of the data in the state constraint. Even problems with time-dependent state constraints formulated with a general linear continuous operator can be considered. As an example the continuity of the optimal control for problems with a pointwise state constraint at a fixed point in space can be shown. This has some interesting consequences for optimal control problems with time-dependent controls and pointwise state constraints in space and time.

In the algorithmic and numerics part of this thesis a new optimization algorithm for the considered problem class is developed. This algorithm is an extension of the primal-dual active set strategy for discretized state constrained optimal control problems by Bergounioux and Kunisch. Additionally, ideas of multiple shooting algorithms, which are used for the optimal control of ordinary differential equations, are incorporated. The performance of the algorithm is tested on several examples. Mesh independence for regular problems and moderate mesh dependence for more sophisticated problems with still continuous optimal control are obtained. Overall, the algorithm performs typically better than the primal-dual active set strategy or the active set strategy in Matlab. However, it should be noted that the newly developed algorithm and the primal-dual active set strategy exhibit convergence problems for problems with state constraints of order greater than one. With regard to the findings in the theoretical part of this work the numerical results show that the derived propositions seem to be precise, if the total interval is considered. On the other hand, better regularity results on subintervals for problems with regular data should be possible.

The last part of the thesis is devoted to the optimal control of the power of a laser during the laser melting process. A newly developed model describes the temperature distribution in the workpiece in situations in which the laser moves through the turning point in a meander at the boundary of an island. Based on this model linear quadratic optimization problems are formulated. With these problems optimal strategies for the adjustment of the power of the laser can be computed. Due to additional constraints, the newly developed algorithm of this thesis cannot be used. Therefore, the algorithm is replaced by the Matlab `quadprog` routine. The resulting strategies are tested on metal 3D printing machines. It turns out that the strategy that reduces the power of the laser not only during the transverse motion in the meander at the boundary of an island, but

also on a short way backwards to the interior brings about significant improvements in the melt structure. In consequence, better quality of the surface of the workpiece can be achieved.

## **Danksagung**

An dieser Stelle der Arbeit möchte ich die Gelegenheit nutzen, mich bei allen Personen zu bedanken, die maßgeblich zum Gelingen der Arbeit beigetragen haben.

Besonderer Dank gilt hierbei meinem Betreuer Prof. Dr. Hans Josef Pesch, der mich dazu inspirierte, zustandsbeschränkte Optimalsteuerungsprobleme partieller Differentialgleichungen unter dem Blickwinkel der Resultate der optimalen Steuerung gewöhnlicher Differentialgleichungen zu untersuchen. Er stand mir während der gemeinsamen Zeit am Lehrstuhl als auch in der Zeit außerhalb der Universität als wichtiger und ausdauernder Ansprechpartner zur Verfügung. Neben den fundierten wissenschaftlichen Kenntnissen und den daraus resultierenden befruchtenden Diskussionen schätze ich besonders seine Fähigkeit junge Personen anzuleiten und zu motivieren.

Des Weiteren möchte ich ganz besonders Prof. Dr. Kurt Chudej danken, der mir stets mit Rat und Tat zur Seite stand. Durch seinen Einsatz ermöglichte er es mir insbesondere auch außerhalb des Lehrstuhls an meiner Promotion zu arbeiten.

Ein wichtiger Bestandteil beim Anfertigen dieser Arbeit waren auch die tiefgehenden und inspirierenden Diskussionen mit Kollegen. Für die vielen Ratschläge geht ein großer Dank nochmals an Dr. Julia Fischer, Dr. Michael Frey, Dr. Nils Altmüller, Dr. Armin Rund, Dr. Stefan Wendl und Dr. Isabella Stock.

Ohne den großen und steten Rückhalt durch meine Familie, meine Eltern, meine Geschwister und meine Frau Nadja wäre es mir nicht möglich gewesen diese Arbeit anzufertigen. Ihr habt mir die Kraft zur Arbeit an der Promotion gegeben. Hierfür möchte ich Euch nochmals ganz herzlich danken.

Lauf, November 2017

Simon Bechmann





# Inhaltsverzeichnis

<b>Vorwort</b>	<b>III</b>
Zusammenfassung . . . . .	III
Abstract . . . . .	V
Danksagung . . . . .	VII
<b>1. Einführung</b>	<b>1</b>
<b>2. Theorie</b>	<b>5</b>
2.1. Zeitabhängige optimale Steuerung mit Zustandsbeschränkung in integraler Form . . . . .	5
2.1.1. Klassische Resultate . . . . .	8
2.1.2. Erhöhte Regularitäten mit alternativem adjungierten Zustand . . . . .	12
2.1.3. Zusammenhang mit der optimalen Steuerung gewöhnlicher Differentialgleichungen . . . . .	18
2.1.4. Erhöhte Regularitäten mit geschicktem Testen . . . . .	23
2.1.5. Erhöhte Regularitäten bei erhöhter Regularität des Zustands . . . . .	27
2.2. Zeitabhängige optimale Steuerung mit Zustandsbeschränkung in Operatorform . . . . .	32
<b>3. Algorithmen und numerische Versuche</b>	<b>41</b>
3.1. PDASSP Algorithmus . . . . .	41
3.1.1. Schaltpunktsuche und PDAS Strategie . . . . .	41
3.1.2. Zeitdiskretisiertes reduziertes Problem . . . . .	44
3.1.3. Grundidee des PDASSP Algorithmus . . . . .	48
3.1.4. Verfeinerter PDASSP Algorithmus . . . . .	57
3.2. Numerische Versuche . . . . .	64
3.2.1. Implementierung . . . . .	65
3.2.2. Modellprobleme und Resultate . . . . .	68
<b>4. Optimale Steuerung beim Laserstrahlschmelzen</b>	<b>85</b>
4.1. Modell . . . . .	86
4.2. Simulation des Modells . . . . .	89
4.3. Optimierung der Leistung des Lasers . . . . .	90
4.4. Umsetzung der Strategien auf der Anlage . . . . .	92
4.5. Strategien, Testbauteile und Auswertung . . . . .	93
4.5.1. Bauteile 1 . . . . .	93
4.5.2. Bauteile 2 . . . . .	97
<b>A. Anhang</b>	<b>103</b>

**Literatur**

**111**

# 1. Einführung

Das Themengebiet der optimalen Steuerung partieller Differentialgleichungen zeichnet sich nach Ansicht des Autors durch das Zusammenspiel verschiedener Teilgebiete der Mathematik aus. So werden bei der Optimierungstheorie im Banachraum und der Regularitätstheorie partieller Differentialgleichungen funktionalanalytische Methoden eingesetzt. Zudem kommen bei der konzeptionellen Entwicklung von Optimierungsalgorithmen Kenntnisse auf dem Gebiet der nichtlinearen Optimierung zum Tragen. In der Umsetzung der entwickelten Algorithmen und beim Lösen realer Anwendungsprobleme spielen schließlich die Numerik und das Scientific Computing eine entscheidende Rolle. In den Ausführungen dieser Arbeit soll all diesen Aspekten in der herausfordernden Situation der Optimalsteuerung parabolischer Differentialgleichungen mit Zustandsbeschränkungen Rechnung getragen werden. Die Arbeit gliedert sich deshalb in die folgenden drei Teile.

Im ersten Teil der Arbeit werden theoretische Betrachtungen zu einem zustandsbeschränkten Optimalsteuerungsproblem der Wärmeleitungsgleichung mit einem Tracking-Type-Zielfunktional angestellt. Das Problem ist motiviert durch ein Anwendungsproblem im Kontext des 3D-Drucks metallischer Bauteile und weist eine spezielle Struktur auf: Die Steuerung in der rechten Seite der Zustandsgleichung und die Zustandsbeschränkung sind nur von der Zeit abhängig. Ziel dieses Teils der Arbeit ist es, für Probleme dieser Struktur eine möglichst hohe Regularität der Lösungen des zugehörigen Optimalitätssystems unter minimalen Anforderungen an die Daten herzuleiten. Insbesondere soll die Stetigkeit der optimalen Steuerung bewiesen werden.

In der Literatur finden sich diverse Arbeiten mit Problemen ähnlicher Struktur. Optimalitätsbedingungen erster Ordnung für allgemeine zustandsbeschränkte parabolische Probleme und Regularitätsaussagen zu parabolischen Differentialgleichungen mit maßwertiger rechter Seite findet man in den Arbeiten von Casas [18], Barbu und Precupanu [4]. Hinreichende Bedingungen zweiter Ordnung für semilinear parabolische Probleme mit rein zeitabhängiger Steuerung aber punktwisen Zustandsbeschränkungen werden von de Los Reyes u. a. [20] behandelt. Mit der selben Zielsetzung betrachten Goldberg und Tröltzsch [27] Randsteuerungsprobleme bei vergleichbarer Struktur der Zustandsbeschränkung. Ein höheres Regularitätsresultat der optimalen Steuerung bei einer rein zeitabhängigen, integralen Zustandsbeschränkung konnte in Meidner, Rannacher und Vexler [45] als Nebenprodukt bei der Herleitung von Fehlerabschätzungen bewiesen werden. Maßgeblich beeinflusst wurden die Ausführungen in dieser Arbeit von den Darstellungen in den Monographien [11, 41], in denen mit dieser Arbeit vergleichbare Regularitätsresultate hergeleitet werden. Die Herangehensweise und die grundsätzliche Beweisidee, ein alternatives Optimalitätssystem zur Herleitung der erhöhten Regularität der optimalen Steuerung zu nutzen, ist überwiegend den Darstellungen in Bonnans und Jaisson [11] nachempfunden. Im Unterschied zu den Ausführungen dort weisen die im Rahmen dieser

Arbeit betrachteten Zustandsbeschränkungen aber eine allgemeinere Form auf. In Lasiecka [41] wird nachgewiesen, dass die optimale Steuerung von beschränkter Variation ist und deshalb der optimale Zustand eine höhere Regularität besitzen muss. Dies kann im weiteren Fortgang genutzt werden, um die Stetigkeit der optimalen Steuerung mittels einer Hamiltonfunktion und einer sogenannten „corner condition“, die durch Variation der Lagrangefunktion nach den Schaltpunkten erhalten wird, herzuleiten. Die Idee, die beschränkte Variation der optimalen Steuerung zu zeigen und hierdurch erhöhte Regularität des optimalen Zustands nachzuweisen, wird in dieser Arbeit in ähnlicher Form verfolgt. Den Ausarbeitungen in Bonnans und Jaisson [11] folgend wird aber nicht mit einer Hamiltonfunktion gearbeitet. Zusätzlich sei angemerkt, dass auch in Lasiecka [41] die Zustandsbeschränkungen in der Form dieser Arbeit nicht abgedeckt sind. Im Unterschied zu den beiden genannten Monographien wird in den nachfolgenden Ausführungen außerdem grundsätzlich mit Ortsgebieten niedrigerer Randregularität gearbeitet.

Konkret ist das Theoriekapitel wie folgt gegliedert: Im ersten Teil wird ein Modellproblem mit rein zeitabhängiger Zustandsbeschränkung in integraler Form betrachtet. Nach Einführung des zugehörigen reduzierten Problems werden Existenz- und Eindeutigkeitsaussagen und ein klassisches Optimalitätssystem formuliert. Unter ersten regulären Annahmen an die Daten kann hieraus eine alternative adjungierte Gleichung und ein alternatives Optimalitätssystem abgeleitet werden. Dies erlaubt es, höhere Regularität der Lösungen des Optimalitätssystems nachzuweisen. Die erhaltenen Resultate werden in einem nächsten Abschnitt mit den Resultaten der optimalen Steuerung gewöhnlicher Differentialgleichungen verglichen und in Beziehung gesetzt. In den weiteren Betrachtungen des Theoriekapitels geht es darum, vergleichbare Regularitätsresultate herzuleiten, wenn die Daten in der Zustandsbeschränkung so gewählt sind, dass ein alternativer adjungierter Zustand nicht mehr kanonisch eingeführt werden kann. Unter Ausnutzung der speziellen Struktur des Modellproblems wird die Technik des alternativen Optimalitätssystems weiterentwickelt. In einem zweiten Teil des Theoriekapitels wird eine Erweiterung auf Probleme mit rein zeitabhängiger Zustandsbeschränkung in Operatorform vorgenommen.

Im Algorithmenteil der Arbeit wird der sogenannte PDASSP Algorithmus zum Lösen der Optimalsteuerungsprobleme aus dem Theorieteil der Arbeit entwickelt und an speziell konstruierten Modellproblemen getestet. Der PDASSP Algorithmus stellt eine Erweiterung der primal-dualen aktive-Mengen-Strategie (PDAS) für zustandsbeschränkte diskretisierte Probleme nach Bergounioux und Kunisch [6] dar. Zusätzlich fließen in den Algorithmus Ideen einer Schaltpunktsuche, die bei der optimalen Steuerung gewöhnlicher Differentialgleichungen genutzt wird [16, 17, 49, 8, 10], ein. Hierbei wird die im Theorieteil der Arbeit bewiesene Stetigkeit der optimalen Steuerung maßgeblich genutzt. Das Ziel des neu entwickelten Verfahrens ist es, möglichst wenig Gitterabhängigkeit zu zeigen.

Der Algorithmen- und Numerikteil dieser Arbeit ist wie folgt aufgebaut. Nach einem kurzen Überblick über das Vorgehen bei einer Schaltpunktsuche und bei der primal-dualen aktive-Mengen-Strategie wird das diskretisierte reduzierte Problem zu den Problemen aus dem Theorieteil der Arbeit aufgestellt. Basierend auf einer detaillierten Betrachtung der aus numerischen Berechnungen erhaltenen Größen während der Iterationen

---

des PDAS Algorithmus wird die Grundidee für die Erweiterung des Algorithmus vorgestellt. Diese Idee wird verfeinert, um allgemeine Probleme behandeln zu können und die Konvergenz des neu entwickelten Algorithmus sicher zu stellen. Nach Hinweisen zur Implementation werden verschiedene Modellprobleme vorgestellt. Anhand dieser wird einerseits der PDASSP Algorithmus getestet, andererseits wird aber auch die Schärfe der Aussagen aus dem Theorieteil der Arbeit bewertet.

Der Anwendungsteil der Arbeit widmet sich dem Problem der optimalen Steuerung der Leistung des Lasers beim Laserstrahlschmelzen. Mit dem Laserstrahlschmelzen ist es möglich, komplexe metallische Bauteile generativ zu fertigen. Durch eine Anpassung der Leistung des Lasers in Umkehrpunkten beim mäanderförmigen Abfahren der sogenannten Islands soll erreicht werden, dass sich eine gleichmäßigere Schmelzspur und damit eine höhere Qualität des Bauteils ausbildet.

Simulationen der Temperaturverteilung im Werkstück und Auswertungen zum Verhalten der Schmelzspur an schräg zulaufenden Bauteilgrenzen werden in Ilin u. a. [37] vorgenommen. Nach Kenntnis des Autors findet sich in der Literatur aber noch kein Modell zur Simulation der Temperaturverteilung in einem Werkstück, wenn ein Laser die Umkehrpunkte eines Mäanders abfährt. Es wird deshalb zunächst ein solches Modell entwickelt. Basierend auf diesem Modell werden erste Simulationen des Prozesses vorgenommen. Anschließend wird ein linear-quadratisches zustandsbeschränktes reduziertes Optimalsteuerungsproblem formuliert. In einer Diskussion wird angeführt, weshalb der im Rahmen dieser Arbeit entwickelte PDASSP Algorithmus nicht zum Lösen der Probleme eingesetzt werden kann und stattdessen die Matlab Routine `quadprog` verwendet werden muss. Nach einer kurzen Darstellung, wie die berechneten Leistungssteuerungsstrategien auf 3D-Metalldruckern umgesetzt werden können, werden anhand speziell entwickelter Bauteile Tests für verschiedenste Verfahren durchgeführt. Die Ergebnisse werden mithilfe von Makro- und Mikroskopaufnahmen und mit Schlibbildern bewertet. Ein einfach umsetzbares Verfahren, das eine signifikante Verbesserung der Ausprägung der Schmelzspur bewirkt, wird hierbei identifiziert.



## 2. Theorie

### 2.1. Zeitabhängige optimale Steuerung mit Zustandsbeschränkung in integraler Form

Im Rahmen dieses Kapitels soll das nachfolgende zustandsbeschränkte Optimalsteuerungsproblem einer Wärmeleitungsgleichung betrachtet werden. Die konkrete Aufgabenstellung zeichnet sich dadurch aus, dass in der rechten Seite der Zustandsgleichung  $u(t)e(x, t)$  eine rein zeitabhängige Steuerung  $u(t)$  auftritt. Die Zustandsbeschränkung  $\int_{\Omega} y(x, t)\delta(x, t) dx - y_{\max}(t)$  ist ebenfalls nur von der Zeit  $t$  abhängig.

Die exakte Formulierung des Modellproblems lautet:

$$\min_{y \in W(0, T), u \in L^2(0, T)} J(y, u) := \frac{1}{2} \iint_Q (y(x, t) - y_d(x, t))^2 dx dt + \frac{\lambda}{2} \int_I u(t)^2 dt \quad (2.1a)$$

unter der Gleichungsnebenbedingung

$$y_t(x, t) - \Delta y(x, t) = u(t) \cdot e(x, t) \quad \text{f.ü. in } Q, \quad (2.1b)$$

$$\partial_\nu y(x, t) + y(x, t) = 0 \quad \text{f.ü. auf } \Sigma, \quad (2.1c)$$

$$y(x, 0) = y_0(x) \quad \text{f.ü. in } \Omega, \quad (2.1d)$$

und der Zustandsbeschränkung

$$\int_{\Omega} y(x, t)\delta(x, t) dx - y_{\max}(t) \leq 0 \quad \text{für alle } t \in I. \quad (2.1e)$$

Dabei seien  $\Omega \subset \mathbb{R}^d, d \geq 1$ , ein beschränktes Gebiet mit Rand  $\Gamma$  der Klasse  $C^{0,1}$  (Lipschitzgebiet),  $I := [0, T] \subset \mathbb{R}, 0 < T < \infty$ , das betrachtete Zeitintervall sowie  $Q := \Omega \times (0, T), \Sigma := \Gamma \times (0, T)$ .  $\lambda > 0$  sei ein reeller Parameter. Die weiteren Daten sollen den folgenden Bedingungen genügen: Verteilungsfunktion  $e \in L^\infty(0, T; L^2(\Omega))$ , Mittelungsfunktion  $\delta \in C([0, T]; L^2(\Omega))$ , Zielzustand  $y_d \in L^2(0, T; L^2(\Omega))$ , Anfangswert  $y_0 \in L^2(\Omega)$ , sowie Zustandsschranke  $y_{\max} \in C([0, T])$ , wobei  $\int_{\Omega} y_0 \delta(x, 0) dx < y_{\max}(0)$  gelte. Als zulässiger Raum für die Steuerungen sei

$$u \in L^2(0, T)$$

gefordert. In dem Modellproblem wird bewusst auf Steuerungsbeschränkungen verzichtet, obwohl diese in technischen Anwendungen häufig gegeben sind. Dies stellt aus theoretischer Sicht insofern einen schwierigeren Fall dar, als dass man aus den üblichen

Boxbeschränkungen  $u_a \leq u \leq u_b$  mit  $u_a, u_b \in L^\infty(0, T)$  für zulässige Steuerungen sofort  $u \in L^\infty(0, T)$  erhalten würde. In Folge ergäbe sich eine erhöhte Regularität des zugehörigen Zustands  $y$ . Die theoretischen Aussagen in dieser Arbeit sind aber auch ohne diese erhöhten Regularitäten möglich. Nichtsdestotrotz sind die verwendeten Beweistechniken und die Resultate dieser Arbeit auf den Fall von Steuerungsbeschränkungen übertragbar, wenn man beispielsweise  $u_a, u_b \in C([0, T])$  fordert und zusätzlich verlangt, dass sich die aktiven Mengen von Zustands- und Steuerungsbeschränkung nicht überschneiden.

An dieser Stelle sei außerdem darauf hingewiesen, dass im weiteren Fortgang der Arbeit grundsätzlich mit Banachraumwertigen Funktionen gearbeitet wird. Nur an Stellen, an denen die unterschiedlichen Abhängigkeiten der Funktionen von  $(x, t)$  oder  $t$  herausgestellt werden sollen, werden diese explizit aufgeführt. Wenn es der Übersichtlichkeit dient, wird komplett auf die Notation der Abhängigkeiten verzichtet.

Im Folgenden soll auf die Wohldefiniertheit der Zustandsgleichung (2.1b)–(2.1d) und der Zustandsbeschränkung (2.1e) eingegangen werden. Als Zustandsraum wird im Modellproblem der Raum  $W(0, T)$  genannt, der wie üblich definiert sei durch

$$W(0, T) := \{y \in L^2(0, T; H^1(\Omega)) \mid y' \in L^2(0, T; H^1(\Omega)^*)\},$$

wobei  $y'$  die distributionelle Zeitableitung von  $y$  bezeichne und die zugehörige Norm gegeben sei durch

$$\|y\|_{W(0, T)} := \left( \int_0^T \|y(t)\|_{H^1(\Omega)}^2 + \|y'(t)\|_{H^1(\Omega)^*}^2 dt \right)^{\frac{1}{2}}.$$

Um dies zu rechtfertigen soll eine wohlbekanntete Umformulierung der Zustandsgleichung vorgenommen werden [54]. Für  $\tilde{v}, w \in H^1(\Omega)$  wird durch die Bilinearform  $a: H^1(\Omega) \times H^1(\Omega) \rightarrow \mathbb{R}$

$$a(\tilde{v}, w) := \int_{\Omega} \nabla \tilde{v} \nabla w dx + \int_{\Gamma} \tilde{v} w ds(x)$$

in der schwachen Formulierung der Zustandsgleichung (2.1b)–(2.1d) mit

$$\langle A\tilde{v}, w \rangle_{H^1(\Omega)^*, H^1(\Omega)} = a(\tilde{v}, w) \quad \forall \tilde{v}, w \in H^1(\Omega)$$

ein linear stetiger Operator  $A: H^1(\Omega) \rightarrow H^1(\Omega)^*$  definiert. Wegen  $e \in L^\infty(0, T; L^2(\Omega))$  lässt sich außerdem eine Funktion  $E: [0, T] \rightarrow H^1(\Omega)^*$  mit  $\|E(t)\|_{H^1(\Omega)^*} \leq \|e(\cdot, t)\|_{L^2(\Omega)}$  definieren durch

$$\langle E(t), v \rangle_{H^1(\Omega)^*, H^1(\Omega)} = \int_{\Omega} e(x, t)v(x) dx \quad \forall v \in H^1(\Omega).$$

In Folge dessen gilt  $E \in L^\infty(0, T; H^1(\Omega)^*)$  und

$$\|uE\|_{L^2(0, T; H^1(\Omega)^*)}^2 \leq \int_0^T u(t)^2 \cdot \|e(\cdot, t)\|_{L^2(\Omega)}^2 dt \leq \|e\|_{L^\infty(0, T; L^2(\Omega))}^2 \|u\|_{L^2(0, T)}^2. \quad (2.2)$$



Man kann die Zustandsgleichung deshalb auch als Evolutionsproblem

$$y'(t) + Ay(t) = u(t) \cdot E(t) \quad \text{f.ü. in } [0, T], \quad (2.3a)$$

$$y(0) = y_0 \quad (2.3b)$$

mit rechter Seite in  $L^2(0, T; H^1(\Omega)^*)$  auffassen, so dass sich die folgenden Aussagen treffen lassen.

**Lemma 2.1.1.** *Es seien die Voraussetzungen des Modellproblems erfüllt, insbesondere gelte also  $e \in L^\infty(0, T; L^2(\Omega))$ . Dann besitzt die Zustandsgleichung (2.3) zu jedem  $u \in L^2(0, T)$  und jedem Anfangswert  $y_0 \in L^2(\Omega)$  genau eine Lösung  $y \in W(0, T)$ . Diese erfüllt mit einer Konstanten  $c > 0$*

$$\|y\|_{W(0, T)} \leq c (\|u\|_{L^2(0, T)} \cdot \|e\|_{L^\infty(0, T; L^2(\Omega))} + \|y_0\|_{L^2(\Omega)}). \quad (2.4)$$

Des Weiteren hat man für die Funktion  $g: [0, T] \rightarrow \mathbb{R}$ ,  $g(t) := \int_{\Omega} y(x, t) \delta(x, t) dx - y_{\max}(t)$  in der Zustandsbeschränkung:

$$g \in C([0, T]).$$

*Beweis.* Den Beweis dieses klassischen Existenz- und Eindeutigkeitsatzes findet man in Gajewski, Gröger und Zacharias [23], Emmrich [21] oder Wloka [55], wobei man die Beschränktheit und Koerzivität der Bilinearform  $a$  nutzt. Die Abschätzung (2.4) folgt aus der stetigen Abhängigkeit der Lösung von der rechten Seite und vom Anfangswert, wenn man die Ungleichung (2.2) berücksichtigt. Die Regularität der Funktion  $g$  ergibt sich unmittelbar aus der Einbettung  $W(0, T) \hookrightarrow C([0, T]; L^2(\Omega))$  und den Voraussetzungen  $\delta \in C([0, T]; L^2(\Omega))$ ,  $y_{\max} \in C([0, T])$ .  $\square$

**Bemerkung 2.1.1.** Im Kontext der obigen Ausarbeitungen ließen sich im Unterschied zu den Forderungen im Modellproblem ( $e \in L^\infty(0, T; L^2(\Omega))$ ) durchaus allgemeinere Funktionen  $E \in L^\infty(0, T; H^1(\Omega)^*)$  betrachten. Für die weiteren Ausarbeitungen ist dies allerdings nicht zweckmäßig, weshalb an dieser Stelle darauf verzichtet wurde.

Basierend auf den bisherigen Ergebnissen lässt sich das Modellproblem (2.1) umformulieren und ein reduziertes Problem einführen. Das Lemma 2.1.1 garantiert die Existenz eines linear stetigen Lösungsoperators  $S: L^2(0, T) \rightarrow W(0, T)$ , der jede Steuerung  $u$  bei fest vorgegebener Verteilungsfunktion  $e$  und Anfangswert  $y_0 = 0$  auf den zugehörigen Zustand  $y_u$  abbildet. Bezeichnet man mit  $y_{\text{init}}$  die Lösung der Zustandsgleichung zum Anfangswert  $y_0$  und homogener rechter Seite, so lässt sich ein affin linearer Steuerungs-Zustands-Operator  $\hat{S}$  definieren durch

$$\hat{S}: L^2(0, T) \rightarrow W(0, T), \quad u \mapsto \hat{S}(u) := y = y_u + y_{\text{init}} = Su + y_{\text{init}}.$$

Des Weiteren ist durch

$$\hat{G}: L^2(0, T) \rightarrow C([0, T]), \quad u \mapsto \hat{G}(u)(t) := \int_{\Omega} \hat{S}(u)(t) \delta(t) dx - y_{\max}$$

ein affin linearer Operator gegeben. Dieser besitzt wegen

$$\left\| \int_{\Omega} y(t)\delta(t) \, dx \right\|_{C([0,T])} \leq \|y\|_{C([0,T];L^2(\Omega))} \|\delta\|_{C([0,T];L^2(\Omega))} \quad \forall y, \delta \in C([0,T];L^2(\Omega))$$

und wegen der stetigen Einbettung  $W(0,T) \hookrightarrow C([0,T];L^2(\Omega))$  den linear stetigen Anteil

$$G: L^2(0,T) \rightarrow C([0,T]), u \mapsto \int_{\Omega} Su(t)\delta(t) \, dx.$$

Mit Hilfe der Definition des reduzierten Zielfunktional

$$\mathcal{J}: L^2(0,T) \rightarrow \mathbb{R}, \mathcal{J}(u) := J(\hat{S}(u), u) = \frac{1}{2} \left\| \hat{S}(u) - y_d \right\|_{L^2(0,T;L^2(\Omega))}^2 + \frac{\lambda}{2} \|u\|_{L^2(0,T)}^2$$

und des konvexen Kegels

$$K := \{g \in C([0,T]) \mid g(t) \leq 0 \text{ für alle } t \in [0,T]\}$$

lässt sich das Modellproblem (2.1) äquivalent als konvexe Optimierungsaufgabe im Banachraum formulieren.

$$\min_{u \in L^2(0,T)} \mathcal{J}(u) \quad \text{unter der Nebenbedingung} \quad \hat{G}(u) \in K. \quad (2.5)$$

### 2.1.1. Klassische Resultate

In diesem Abschnitt werden für das reduzierte Problem (2.5) klassische Resultate dargeboten, unter anderem der folgende Existenz- und Eindeigkeitssatz.

**Satz 2.1.1.** *Das Problem (2.5) besitzt unter den gegebenen Voraussetzungen des Modellproblems (2.1) eine eindeutige, optimale Lösung  $\bar{u} \in L^2(0,T)$ .*

*Beweis.* Ein Beweis der Aussage findet sich z.B. in Hinze u. a. [35]. Der Beweis der Existenzaussage beruht im Wesentlichen auf der schwachen Unterhalbstetigkeit des konvexen reduzierten Zielfunktional  $\mathcal{J}$  und der schwachen Folgenabgeschlossenheit der abgeschlossenen und konvexen Menge  $K$ . Die Eindeutigkeit der optimalen Lösung ergibt sich aufgrund der strikten Konvexität von  $\mathcal{J}$  für  $\lambda > 0$ .  $\square$

Um für das reduzierte konvexe Problem qualifizierte Optimalitätsbedingungen herzuleiten, fordert man typischerweise eine „constraint qualification“. Im Rahmen dieser Arbeit sei die folgende Slater-Bedingung angenommen.

**Annahme 2.1.1.** Es existiere ein Element  $\tilde{u} \in L^2(0,T)$ , so dass  $\hat{G}(\tilde{u}) \in \text{int } K$ , wobei  $\text{int } K$  das Innere des Kegels  $K$  bezeichne.

**Bemerkung 2.1.2.** Die Forderung  $\hat{G}(\tilde{u}) \in \text{int } K$  bedeutet nach der Definition von  $K \subset C([0, T])$  konkret die Existenz eines Zustands  $\tilde{y} = \hat{S}(\tilde{u})$  mit  $\int_{\Omega} \tilde{y}(t)\delta(t) dx < y_{\max}(t)$  für alle  $t \in [0, T]$ . Aufgrund der großen Wahlfreiheit für die Daten  $e$  und  $\delta$  ist dies im Allgemeinen hier nicht nachweisbar. Um jedoch überhaupt die Möglichkeit zu haben, eine entsprechende Steuerung  $\tilde{u}$  zu finden, muss für den Anfangswert  $y_0$  zwingend die Bedingung  $\int_{\Omega} y_0\delta(0) dx < y_{\max}(0)$  gefordert werden (vgl. Modellproblem (2.1)).

Aus der Optimierungstheorie im Banachraum ergeben sich mit den Ausführungen in [54, Kap. 6.1.1, 43] auf Grund der Konvexität der Operatoren  $\mathcal{J}$  und  $\hat{G}$  die folgenden Optimalitätsbedingungen.

**Satz 2.1.2.** Sei  $\bar{u}$  die optimale Lösung des Problems (2.5) und sei die Slater-Bedingung aus Annahme 2.1.1 erfüllt. Dann existiert ein  $\eta \in K^-$ , mit dem negativen dualen Kegel

$$K^- := \left\{ \eta \in C([0, T])^* : \langle \eta, g \rangle_{C([0, T])^*, C([0, T])} \leq 0 \quad \forall g \in K \right\},$$

so dass mit der Lagrangefunktion

$$\mathcal{L} : L^2(0, T) \times C([0, T])^* \rightarrow \mathbb{R}, (u, \eta) \mapsto \mathcal{J}(u) + \left\langle \eta, \hat{G}(u) \right\rangle_{C([0, T])^*, C([0, T])}$$

die Sattelpunkteigenschaft

$$D_u \mathcal{L}(\bar{u}, \eta)h = 0 \quad \forall h \in L^2(0, T)$$

und die komplementäre Schlupfbedingung

$$\eta \in K^- \quad \text{und} \quad \left\langle \eta, \hat{G}(\bar{u}) \right\rangle_{C([0, T])^*, C([0, T])} = 0$$

erfüllt sind. Dabei bezeichne  $D_u$  die partielle Frechet-Ableitung nach  $u$ .

Die Resultate des Satzes sollen nun näher spezifiziert werden. Dabei spielen die Funktionen beschränkter Variation eine entscheidende Rolle. Die exakten Definitionen sowie eine kompakte Übersicht über wichtige Eigenschaften dieser Funktionen v.a. im Kontext der hier behandelten Optimalsteuerungsprobleme findet sich in Gerds [24, Ch. 2]. Im Folgenden werden diese Eigenschaften wiederholt eingesetzt, wobei nur noch dann ein Verweis auf eine Literaturquelle vorgenommen wird, wenn es sich um ein Resultat handelt, dass sich nicht standardmäßig in der Literatur finden lässt. Es seien auf  $[0, T]$  definiert:

- $BV(0, T)$  der Raum aller reellwertigen *Funktionen beschränkter Variation*,
- $NBV_T(0, T)$  der Raum aller reellwertigen *normalisierten Funktionen beschränkter Variation*. Dieser besteht aus allen auf  $(0, T)$  rechtsseitig stetigen Funktionen  $\mu \in BV(0, T)$  mit  $\mu(T) = 0$ .

An dieser Stelle sei darauf hingewiesen, dass im Unterschied zu den Ausarbeitungen in Gerds [24], jedoch analog zu der Darstellung in Bonnans und Jaisson [11], die Normalisierung der Funktion  $\mu \in BV(0, T)$  so gewählt wurde, dass  $\mu(T) = 0$  und nicht  $\mu(0) = 0$  gilt. Da diese Bedingung nur der Festlegung einer additiven Konstanten dient, ist dies ohne weiteres möglich.

**Korollar 2.1.1.** Sei  $\bar{u}$  die optimale Lösung des Problems (2.5) und sei die Slater-Bedingung aus Annahme 2.1.1 erfüllt. Dann existiert ein  $\mu \in NBV_T(0, T)$ , so dass die folgenden Bedingungen erfüllt sind:

- Sattelpunkteigenschaft der Lagrangefunktion:  $\forall h \in L^2(0, T)$ :

$$0 = \iint_Q (\hat{S}(\bar{u}) - y_d) Sh \, dxdt + \lambda \int_I \bar{u}h \, dt + \int_I \left( \int_\Omega Sh \cdot \delta \, dx \right) d\mu \quad (2.6)$$

- komplementäre Schlupfbedingung:

$$\int_I g \, d\mu \leq 0 \quad \forall g \in K \quad \text{und} \quad 0 = \int_I \hat{G}(\bar{u}) \, d\mu. \quad (2.7)$$

*Beweis.* Der Raum  $NBV_T(0, T)$  versehen mit der passenden Norm (Totalvariation) kann mit dem Dualraum der stetigen Funktionen  $C([0, T])^*$  identifiziert werden [48, S. 266, 43, S. 113, 1]. Zu jedem  $\eta \in C([0, T])^*$  korrespondiert deshalb ein eindeutiges  $\mu \in NBV_T(0, T)$ , so dass die Optimalitätsbedingungen aus Satz 2.1.2 nach Gerdts [24, S. 131] äquivalent geschrieben werden können als:

$$\eta \in K^- \Leftrightarrow \int_0^T g(t) \, d\mu(t) \leq 0 \quad \forall g \in K$$

und

$$0 = \langle \eta, \hat{G}(\bar{u}) \rangle_{C([0, T])^*, C([0, T])} = \int_0^T \hat{G}(\bar{u})(t) \, d\mu(t)$$

sowie mit der Definition der Lagrangefunktion

$$\begin{aligned} 0 &= D_u \mathcal{L}(\bar{u}, \eta)h \\ &= \iint_Q (\hat{S}(\bar{u}) - y_d) Sh \, dxdt + \lambda \int_I \bar{u}h \, dt + \int_I \left( \int_\Omega Sh \cdot \delta \, dx \right) d\mu \quad \forall h \in L^2(0, T). \end{aligned}$$

□

**Bemerkung 2.1.3.** Die Funktion  $\mu \in NBV_T(0, T)$  sollte nicht als Lagrange-Multiplikator bezeichnet werden. Vielmehr handelt es sich bei  $\mu$  um die den Lagrange-Multiplikator  $d\mu$  erzeugende Funktion. Besitzt  $\mu$  nur einen absolut stetigen Anteil, so hat man die Darstellung  $\int_I \cdot d\mu(t) = \int_I \cdot \mu_t(t) \, dt$  mit  $\mu_t \in L^1(0, T)$ . Entsprechend assoziiert man in diesem Fall den Begriff Lagrange-Multiplikator mit der Funktion  $\mu_t$ , also der Zeitableitung von  $\mu$ .

Wie üblich führt man an dieser Stelle einen adjungierten Zustand ein, um Optimalitätsbedingungen in nichtvariationeller Form zu erhalten. In diesen tauchen banachraumwertige Funktionen beschränkter Variation  $BV(0, T; X)$ ,  $X$  Banachraum, auf. Einen Überblick über deren Eigenschaften und die Anwendung bei Optimalsteuerungsproblemen findet sich in Barbu und Precupanu [4, Ch. 1]. Eine wichtige Eigenschaft dieser Funktionen ist (analog zum reellwertigen Fall), dass sie für alle  $t \in (0, T)$  rechts- und linksseitige Grenzwerte besitzen. An den beiden Rändern existiert jeweils der entsprechende einseitige Grenzwert. Bis auf abzählbar viele Unstetigkeitsstellen sind die Funktionen stetig.

**Satz 2.1.3.** *Seien die Voraussetzungen des Modellproblems (2.1) und die Slater-Bedingung aus Annahme 2.1.1 erfüllt. Sei  $\bar{u} \in L^2(0, T)$  die eindeutige optimale Lösung des Problems (2.5). Dann existieren eine Funktion  $\mu \in NBV_T(0, T)$  sowie ein adjungierter Zustand  $p \in L^2(0, T; H^1(\Omega)) \cap L^\infty(0, T; L^2(\Omega)) \cap BV(0, T; H^1(\Omega)^*)$ , so dass die folgenden Bedingungen erfüllt sind:*

- *adjungierte Gleichung:*

$$\begin{aligned} \int_I \langle \varphi_t, p \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt + \int_I \langle Ap, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt \\ = \iint_Q (\hat{S}(\bar{u}) - y_d) \varphi dx dt + \iint_Q \delta \varphi dx d\mu \quad \forall \varphi \in W(0, T) \text{ mit } \varphi(0) = 0, \end{aligned} \quad (2.8)$$

- *Sattelpunkteigenschaft der Lagrangefunktion:*

$$\lambda \bar{u}(t) + \int_\Omega p e dx(t) = 0 \quad \text{f.ü. in } [0, T], \quad (2.9)$$

- *komplementäre Schlupfbedingung:  $\mu$  ist monoton steigend auf  $[0, T]$  und konstant auf jedem Intervall  $(t_1, t_2) \subset [0, T]$  mit  $t_1 < t_2$  und  $\hat{G}(\bar{u})(t) < 0$  für alle  $t \in (t_1, t_2)$ .*

*Beweis.* Ausgangspunkt für den Beweis ist Korollar 2.1.1, das insbesondere die Existenz der Funktion  $\mu \in NBV_T(0, T)$  mit den entsprechenden Eigenschaften garantiert. Die Schlupfbedingung in der veränderten Form lässt sich aus der Beziehung (2.7) wie in Gerdt's [24, S. 131 ff.] beschrieben herleiten. In Folge soll deshalb nur noch auf die adjungierte Gleichung und die Sattelpunkteigenschaft der Lagrangefunktion eingegangen werden. Nach Casas [18, Thm. 6.4] und Barbu und Precupanu [4, Cor. 3.1] existiert ein  $p \in L^2(0, T; H^1(\Omega)) \cap L^\infty(0, T; L^2(\Omega)) \cap BV(0, T; H^1(\Omega)^*)$ , so dass im distributionellen Sinne

$$\begin{aligned} -p' + Ap &= \hat{S}(\bar{u}) - y_d + \delta \mu', \\ \lim_{h \searrow 0} p(T - h) &= [\mu(T)] \cdot \delta(T) \end{aligned}$$

eingehalten ist. Hierbei sei  $[\mu(T)] = \mu(T) - \lim_{h \searrow 0} \mu(T - h)$  der Sprung der Funktion  $\mu \in NBV_T(0, T)$  zum Zeitpunkt  $T$ . Wegen  $\iint_Q \cdot \delta dx d\mu \in (C([0, T]; L^2(\Omega)))^*$  und  $Ap \in L^2(0, T; H^1(\Omega)^*)$  würde sich an dieser Stelle mit Barbu und Precupanu [4, Ch. 1.3] sogar darauf schließen lassen, dass  $p$  als Summe zweier Funktionen in  $BV(0, T; L^2(\Omega))$  und  $H^1(0, T; H^1(\Omega)^*)$  geschrieben werden kann. Dieses Resultat wird hier aber nicht benötigt, da später ohnehin eine explizite Darstellung für eine additive Zerlegung von  $p$  hergeleitet wird. Weiter erfüllt  $p$  gemäß Barbu und Precupanu [4, Cor. 3.2] für alle  $\varphi \in C^1([0, T]; H^1(\Omega))$  mit  $\varphi(0) = 0$ :

$$\begin{aligned} \int_I \langle \varphi_t, p \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt + \int_I \langle Ap, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt \\ = \iint_Q (\hat{S}(\bar{u}) - y_d) \varphi dx dt + \iint_Q \delta \varphi dx d\mu. \end{aligned}$$

Mit der Dichtigkeit von  $C^1([0, T]; H^1(\Omega))$  in  $W(0, T)$  [51, Ch. 7] ergibt sich die Gültigkeit dieser Beziehung auch für alle  $\varphi \in W(0, T)$  mit  $\varphi(0) = 0$  und somit die adjungierte Gleichung in der angeführten Form.

Funktionen  $Sh$  zu beliebigem  $h \in L^2(0, T)$  gehören dem Raum  $W(0, T)$  an und erfüllen nach der Definition des Operators  $S$  gerade die Bedingung  $(Sh)(0) = 0$ . Man kann deshalb die adjungierte Gleichung mit den Funktionen  $Sh$  testen und erhält:

$$\begin{aligned} \int_I \langle (Sh)_t, p \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt + \int_I \langle Ap, Sh \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt \\ = \iint_Q (\hat{S}(\bar{u}) - y_d) \cdot Sh dxdt + \iint_Q \delta \cdot Sh dx d\mu \quad \forall h \in L^2(0, T). \end{aligned}$$

Wegen

$$\int_I \langle Ap, Sh \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt = \int_I \langle A(Sh), p \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt$$

ergibt sich nach Testen der Zustandsgleichung (2.3) mit  $p \in L^2(0, T; H^1(\Omega))$

$$\int_0^T \left( \int_{\Omega} pe dx \right) h dt = \iint_Q (\hat{S}(\bar{u}) - y_d) \cdot Sh dxdt + \iint_Q \delta \cdot Sh dx d\mu.$$

Für alle  $h \in L^2(0, T)$  gilt deshalb mit der Beziehung (2.6)

$$\int_0^T \lambda \bar{u} h dt + \int_0^T \left( \int_{\Omega} pe dx \right) h dt = 0$$

und damit

$$\lambda \bar{u}(t) + \int_{\Omega} pe dx(t) = 0 \quad \text{f.ü. in } [0, T].$$

□

Wegen  $\bar{u} = -\frac{1}{\lambda} \int_{\Omega} pe dx$  in  $I$  ergibt sich aus  $p \in L^\infty(0, T; L^2(\Omega))$  und  $e \in L^\infty(0, T; L^2(\Omega))$  wegen Gleichung (2.9) erhöhte Regularität für die optimale Steuerung.

**Korollar 2.1.2.** *Die optimale Lösung  $\bar{u}$  des Problems (2.5) gehört dem Raum  $L^\infty(0, T)$  an, wenn die Slater-Bedingung aus Annahme 2.1.1 erfüllt ist.*

### 2.1.2. Erhöhte Regularitäten mit alternativem adjungierten Zustand

In diesem Kapitel soll der Frage nachgegangen werden unter welchen zusätzlichen Bedingungen an die Daten eine weitere Erhöhung der Regularität der optimalen Steuerung möglich ist. Das folgende Vorgehen zum Nachweis erhöhter Regularität ist dem Vorgehen in Bonnans und Jaisson [11] nachempfunden. Entscheidend wird dabei sein, dass die bisherige Adjungierte  $p$  additiv zerlegt werden kann in einen regulären Anteil und einen in der Zeit irreguläreren Anteil, in dem die Funktion  $\mu$  explizit auftritt. Um die Zerlegung durchzuführen, muss der Operator  $A$  auf  $\delta(t)$  fast überall in  $[0, T]$  angewandt werden können. Es sei deshalb die folgende zusätzliche Bedingung gefordert.

**Annahme 2.1.2.** Die Mittelungsfunktion  $\delta$  gehöre statt  $C([0, T]; L^2(\Omega))$  dem Raum  $W(0, T) \hookrightarrow C([0, T]; L^2(\Omega))$  an.

Des Weiteren werden diverse Eigenschaften der Funktionen beschränkter Variation eine entscheidende Rolle spielen, insbesondere die Formel der partiellen Integration für Stieltjes-Integrale. Als Referenz sei hier wieder auf Gerds [24] verwiesen.

Als erstes wichtiges Ergebnis erhält man ein zum klassischen Optimalitätssystem 2.1.3 äquivalentes Optimalitätssystem.

**Satz 2.1.4.** *Seien die Voraussetzungen des Modellproblems (2.1), die Slater-Bedingung aus Annahme 2.1.1 und die Zusatzannahme 2.1.2 erfüllt. Sei  $\bar{u} \in L^2(0, T)$  die eindeutige optimale Lösung des Problems (2.5). Dann existieren eine Funktion  $\mu \in NBV_T(0, T)$  sowie ein alternativer adjungierter Zustand  $q \in W(0, T)$ , so dass die folgenden Bedingungen erfüllt sind:*

- *alternative adjungierte Gleichung:*

$$\begin{aligned} -q'(t) + Aq(t) &= F(t) + \mu(t)(-\delta'(t) + A\delta(t)) \quad \text{f.ü. in } [0, T], \\ q(T) &= 0, \end{aligned}$$

mit  $F: [0, T] \rightarrow H^1(\Omega)^*$  definiert durch

$$\langle F(t), v \rangle_{H^1(\Omega)^*, H^1(\Omega)} = \int_{\Omega} (\hat{S}(\bar{u})(t) - y_d(t))v \, dx \quad \forall v \in H^1(\Omega),$$

- *Sattelpunkteigenschaft der Lagrangefunktion:*

$$\lambda \bar{u}(t) + \int_{\Omega} (q - \mu\delta)e \, dx(t) = 0 \quad \text{f.ü. in } [0, T],$$

- *komplementäre Schlupfbedingung:  $\mu$  ist monoton steigend auf  $[0, T]$  und konstant auf jedem Intervall  $(t_1, t_2) \subset [0, T]$  mit  $t_1 < t_2$  und  $\hat{G}(\bar{u})(t) < 0$  für alle  $t \in (t_1, t_2)$ .*

*Beweis.* Definiert man  $q := p + \mu\delta$ , so hat man für  $\delta \in W(0, T)$ ,  $p \in L^2(0, T; H^1(\Omega))$  und  $\mu \in NBV_T(0, T) \subseteq L^\infty(0, T)$  die Regularität  $q \in L^2(0, T; H^1(\Omega))$ . Eingesetzt in die adjungierte Gleichung für  $p$  ergibt sich äquivalent für alle  $\varphi \in W(0, T)$  mit  $\varphi(0) = 0$

$$\begin{aligned} &\int_I \langle \varphi_t, q - \mu\delta \rangle_{H^1(\Omega)^*, H^1(\Omega)} \, dt + \int_I \langle A(q - \mu\delta), \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} \, dt \\ &= \iint_Q (\hat{S}(\bar{u}) - y_d)\varphi \, dxdt + \iint_Q \delta\varphi \, dxd\mu \quad \Leftrightarrow \\ &\int_I \langle \varphi_t, q \rangle_{H^1(\Omega)^*, H^1(\Omega)} \, dt + \int_I \langle Aq, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} \, dt \\ &= \iint_Q (\hat{S}(\bar{u}) - y_d)\varphi \, dxdt + \iint_Q \delta\varphi \, dxd\mu \\ &\quad + \int_I \langle \varphi_t, \mu\delta \rangle_{H^1(\Omega)^*, H^1(\Omega)} \, dt + \int_I \mu \langle A\delta, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} \, dt. \end{aligned}$$

Da  $\int_{\Omega} \varphi \delta dx(t)$  für  $\varphi, \delta \in W(0, T)$  nach Gajewski, Gröger und Zacharias [23, S. 147] absolut stetig ist mit

$$\frac{d}{dt} \int_{\Omega} \varphi \delta dx(t) = \langle \varphi_t(t), \delta(t) \rangle_{H^1(\Omega)^*, H^1(\Omega)} + \langle \delta_t(t), \varphi(t) \rangle_{H^1(\Omega)^*, H^1(\Omega)} \quad \text{f.ü. in } [0, T]$$

und da  $\mu(T) = 0, \varphi(0) = 0$  gilt, kann man mit der Formel der partiellen Integration für Stieltjes-Integrale [24, S. 66, 48, 11] umformen

$$\int_I \langle \varphi_t, \mu \delta \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt = - \int_I \mu \langle \delta_t, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt - \int_I \left( \int_{\Omega} \varphi \delta dx \right) d\mu. \quad (2.11)$$

Somit sind die obigen Gleichungen für alle  $\varphi \in W(0, T)$  mit  $\varphi(0) = 0$  äquivalent zu

$$\begin{aligned} & \int_I \langle q, \varphi_t \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt + \int_I \langle Aq, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt \\ &= \iint_Q (\hat{S}(\bar{u}) - y_d) \varphi dx dt - \int_I \mu \langle \delta_t, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt \\ &+ \int_I \mu \langle A\delta, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt. \end{aligned}$$

Dies ist aber mit der Funktion  $F: [0, T] \rightarrow H^1(\Omega)^*$  definiert durch

$$\langle F(t), v \rangle_{H^1(\Omega)^*, H^1(\Omega)} = \int_{\Omega} (\hat{S}(\bar{u})(t) - y_d(t)) v dx \quad \forall v \in H^1(\Omega),$$

gerade die schwache Formulierung zum Evolutionsproblem

$$\begin{aligned} -q'(t) + Aq(t) &= F(t) + \mu(t)(-\delta'(t) + A\delta(t)) \quad \text{f.ü. in } [0, T], \\ q(T) &= 0. \end{aligned}$$

Die alternative Adjungierte  $q$  besitzt deshalb als Lösung der Evolutionsgleichung die höhere Regularität  $q \in W(0, T)$ . Die alternative Formulierung in der Sattelpunkteigenenschaft der Lagrangefunktion ergibt sich sofort durch Einsetzen von  $p = q - \mu\delta$ .  $\square$

Die optimale Steuerung  $\bar{u}$  besitzt also die Struktur

$$\lambda \bar{u} = - \int_{\Omega} qe dx + \underbrace{\mu}_{\in NBV_T(0, T)} \cdot \int_{\Omega} \delta e dx \quad \text{in } I. \quad (2.12)$$

Um bessere Regularitätsresultate als  $\bar{u} \in L^\infty(0, T)$  zu erhalten, ist es nötig, rechts- und linksseitige Limiten von  $\bar{u}$  zu beliebigen Zeitpunkten  $t \in [0, T]$  bilden zu können. Da dies für Funktionen beschränkter Variation gewährleistet ist, müssen Bedingungen für die Verteilungsfunktion  $e$  gefordert werden, so dass die zeitabhängigen Funktionen  $\int_{\Omega} qe dx$  und  $\int_{\Omega} \delta e dx$  die Bedingung ebenfalls erfüllen. Ein erstes Resultat ist bereits möglich, wenn man  $e \in C([0, T]; L^2(\Omega))$  verlangt.

**Korollar 2.1.3.** *Liegt die Verteilungsfunktion  $e$  im Raum  $C([0, T]; L^2(\Omega))$ , so sind die Funktionen  $\int_{\Omega} \delta e dx$  und  $\int_{\Omega} qe dx$  unter Annahme 2.1.2 stetig auf  $[0, T]$ . Ist  $\int_{\Omega} \delta e dx = 0$  in  $[0, T]$ , so hat man  $\bar{u} \in C([0, T])$ .*



*Beweis.* Die Stetigkeit der Funktionen  $\int_{\Omega} \delta e \, dx$  und  $\int_{\Omega} qe \, dx$  auf  $[0, T]$  folgt aus der Einbettung  $W(0, T) \hookrightarrow C([0, T]; L^2(\Omega))$  und der Forderung  $e \in C([0, T]; L^2(\Omega))$ . Für die weitere Aussage nutzt man die Bedingung (2.12).  $\square$

**Bemerkung 2.1.4.** Die Bedingung  $\int_{\Omega} \delta e \, dx(t) = 0$  stellt insofern keinen exotischen Sonderfall dar, als dass z.B. bei zueinander disjunkten Energieeintrags- und Messbereichen diese Bedingung bei entsprechender Modellierung von  $e$  und  $\delta$  gerade erfüllt ist.

Für weitreichendere Aussagen müssen stärkere Annahmen getroffen werden.

**Annahme 2.1.3.** Die Verteilungsfunktion  $e$  gehöre dem Raum  $W(0, T)$  an. Des Weiteren erfülle der Startwert  $y_0$  die Bedingung  $Ay_0 \in L^2(\Omega)$ .

Unter den Annahmen 2.1.2 und 2.1.3 sind die Funktionen  $\int_{\Omega} \delta e \, dx$  und  $\int_{\Omega} qe \, dx$  absolut stetig [23], so dass unmittelbar aus der Beziehung (2.12) wegen  $\int_{\Omega} qe \, dx(T) = 0$  folgt

**Lemma 2.1.2.**  $\bar{u} \in NBV_T(0, T)$ . Ist  $\int_{\Omega} \delta e \, dx = 0$  in  $[0, T]$ , so ist  $\bar{u}$  absolut stetig in  $[0, T]$ .

Dieses Ergebnis erlaubt es, die nach der Zeit differenzierte Zustandsgleichung zu betrachten und so, wie in Lasiecka [41], höhere Regularität für die optimalen Zustände  $\bar{y} := \hat{S}(\bar{u})$  nachzuweisen, wenn man zusätzlich eine erhöhte Regularität des Anfangswertes zur Verfügung hat. Die Grundlage hierfür bildet das folgende Lemma:

**Lemma 2.1.3.** Sei die Annahme 2.1.3 erfüllt. Dann besitzt der zu einer Steuerung  $u \in BV(0, T)$  gehörige Zustand  $y$  als Lösung des Evolutionsproblems (2.3) neben  $y \in W(0, T)$  die erhöhte Regularität  $y' \in L^2(0, T; H^1(\Omega)) \cap BV(0, T; H^1(\Omega)^*)$ . Des Weiteren löst  $Ay$  die Evolutionsgleichung

$$\begin{aligned} (Ay)'(t) + A(Ay)(t) &= u(t) \cdot Ae(t) \quad \text{f.ü. in } [0, T], \\ Ay(0) &= Ay_0, \end{aligned}$$

so dass  $Ay \in W(0, T)$  gilt.

*Beweis.* Die Aussage folgt unmittelbar aus dem Regularitätssatz A.0.1 im Anhang, indem man den Gelfand-Dreier  $H^1(\Omega) \subset L^2(\Omega) \subset H^1(\Omega)^*$  und die Beschränktheit und Koerzivität der Bilinearform  $a$ , die  $A$  erzeugt, nutzt.  $\square$

Für die weiteren Betrachtungen ist es zweckmäßig die Funktion in der Zustandsbeschränkung des reduzierten Problems ausgewertet am Optimum  $\bar{g}: [0, T] \rightarrow \mathbb{R}$ ,  $\bar{g}(t) := \hat{G}(\bar{u})(t)$  sowie die aktive und inaktive Menge

$$\mathcal{A} := \{t \in [0, T] \mid \bar{g}(t) = 0\}, \quad \mathcal{I} := \{t \in [0, T] \mid \bar{g}(t) < 0\} = [0, T] \setminus \mathcal{A}$$

einzuführen. Außerdem werden in der Folge banachraumwertige absolut stetige Funktionen aus  $AC^{k,p}(0, T; X)$ ,  $X$  Banachraum, genutzt. Für  $k \in \mathbb{N}$  und  $1 \leq p \leq \infty$  besteht der Raum  $AC^{k,p}(0, T; X)$  aus allen absolut stetigen Funktionen  $u: [0, T] \rightarrow X$ , deren Zeitableitungen für  $j = 1, 2, \dots, k-1$  absolut stetig sind und für  $j = 1, 2, \dots, k$  dem Raum  $L^p(0, T; X)$  angehören. Einen Überblick über deren Eigenschaften findet man in Barbu und Precupanu [4, Ch. 1]. Für den reellwertigen Fall wird kurz  $AC^{k,p}(0, T)$  geschrieben.

**Satz 2.1.5.** *Seien neben den Grundvoraussetzungen aus Satz 2.1.4 insbesondere die Bedingungen  $Ay_0 \in L^2(\Omega)$ ,  $e, \delta \in W(0, T)$ ,  $\delta_t \in C([0, T]; H^1(\Omega)^*)$  sowie  $y_{\max} \in C^1([0, T])$  erfüllt. Dann ist die optimale Lösung  $\bar{u}$  des Problems (2.5) stetig auf  $[0, T]$  und absolut stetig auf der inaktiven Menge  $\mathcal{I}$ . Des Weiteren gilt für die Funktion in der Zustandsbeschränkung:  $\bar{g} \in C^1([0, T])$ . Wird die Bedingung  $\int_{\Omega} \delta e \, dx(t) \neq 0$  für alle  $t$  in der aktiven Menge  $\mathcal{A}$  eingehalten, so besitzt  $\mu$  keinen diskreten Anteil:  $\mu \in NBV_T(0, T) \cap C([0, T])$ .*

*Beweis.* Für die gegebenen Daten sind  $\int_{\Omega} \delta e \, dx$  und  $\int_{\Omega} qe \, dx$  absolut stetige Funktionen auf  $[0, T]$ . Die komplementäre Schlupfbedingung in Satz 2.1.4 garantiert  $\mu$  konstant auf der inaktiven Menge. Insofern sind die Aussagen wegen Beziehung (2.12) nur noch für  $t$  aus der aktiven Menge  $\mathcal{A}$  zu zeigen. An dieser Stelle sei darauf hingewiesen, dass die aufgrund der Slater-Bedingung geforderte Bedingung  $\int_{\Omega} y_0 \delta(0) \, dx < y_{\max}(0)$  (siehe Modellproblem (2.1)) mit der Stetigkeit der Funktion  $\bar{g}$  die Existenz eines  $\varepsilon > 0$  impliziert, so dass  $\mathcal{A} \subseteq [\varepsilon, T]$ .

Sei  $t \in \mathcal{A}$ . Für die absolut stetige Funktion  $\bar{g} = \int_{\Omega} \bar{y} \delta \, dx - y_{\max}$  gilt unter Verwendung der Zustandsgleichung fast überall in  $[0, T]$ :

$$\begin{aligned} \frac{d}{dt} \bar{g}(t) &= \langle \bar{y}_t(t), \delta(t) \rangle_{H^1(\Omega)^*, H^1(\Omega)} + \langle \delta_t(t), \bar{y}(t) \rangle_{H^1(\Omega)^*, H^1(\Omega)} - \frac{d}{dt} y_{\max}(t) \\ &= \langle -A\bar{y}(t), \delta(t) \rangle_{H^1(\Omega)^*, H^1(\Omega)} + \bar{u}(t) (e(t), \delta(t))_{L^2(\Omega)} + \langle \delta_t(t), \bar{y}(t) \rangle_{H^1(\Omega)^*, H^1(\Omega)} \\ &\quad - \frac{d}{dt} y_{\max}(t). \end{aligned}$$

Wegen der erhöhten Regularität von  $\bar{y}$  und  $A\bar{y}$  nach dem obigen Lemma 2.1.3 hat man bei den gegebenen Daten

$$\frac{d}{dt} \bar{g} = \underbrace{(-A\bar{y}, \delta)_{L^2(\Omega)}}_{\in AC^{1,1}(0, T)} + \underbrace{\bar{u}}_{\in NBV_T(0, T)} \underbrace{(e, \delta)_{L^2(\Omega)}}_{\in AC^{1,1}(0, T)} + \underbrace{\langle \delta_t, \bar{y} \rangle_{H^1(\Omega)^*, H^1(\Omega)}}_{\in C([0, T])} - \underbrace{\frac{d}{dt} y_{\max}}_{\in C([0, T])}.$$

Für den weiteren Beweis benötigt man den Sprung einer Funktion  $b: [0, T] \rightarrow \mathbb{R}$  entsprechender Regularität zum Zeitpunkt  $t \in (0, T]$ . Dieser sei definiert durch:

$$[b(t)] := \lim_{h \searrow 0} b(t+h) - \lim_{h \searrow 0} b(t-h) \quad \text{für } t \in (0, T)$$

sowie  $[b(T)] := b(T) - \lim_{h \searrow 0} b(T-h)$ . Da insbesondere für Funktionen in  $NBV_T(0, T)$  ein solcher wohldefiniert ist und die Funktion  $\bar{g}$  auf  $\mathcal{A}$  ihr Maximum annimmt, ergibt sich für den Sprung der Funktion  $\frac{d}{dt} \bar{g}$  zu einem beliebigen Zeitpunkt  $t \in \mathcal{A}$

$$0 \geq \left[ \frac{d}{dt} \bar{g}(t) \right] = [\bar{u}(t)] \cdot (e(t), \delta(t))_{L^2(\Omega)}. \quad (2.13)$$

Berücksichtigt man, dass nach Satz 2.1.4 für  $t \in (0, T]$  wegen  $q, \delta, e \in W(0, T)$

$$\lambda[\bar{u}(t)] = [\mu(t)] \cdot (e(t), \delta(t))_{L^2(\Omega)}$$

und wegen der Monotonität von  $\mu$

$$[\mu(t)] \geq 0$$

gilt, so folgt für alle  $t \in \mathcal{A}$ :

$$\begin{aligned} 0 &\geq \left[ \frac{d}{dt} \bar{g}(t) \right] \cdot [\mu(t)] = [\bar{u}(t)] \cdot [\mu(t)] \cdot (e(t), \delta(t))_{L^2(\Omega)} \\ &= \lambda [\bar{u}(t)]^2 \geq 0. \end{aligned}$$

Unter Berücksichtigung der Aussage auf der inaktiven Menge ergibt sich  $[u(t)] = 0$  für alle  $t \in [0, T]$  und wegen Beziehung (2.13)  $\bar{g} \in C^1([0, T])$ . Es bleibt noch zu zeigen, dass  $\mu$  auf  $\mathcal{A}$  keinen diskreten Anteil besitzt. Wegen  $\lambda [\bar{u}(t)] = [\mu(t)] \cdot (e(t), \delta(t))_{L^2(\Omega)}$  gilt  $\mu \in C([0, T])$ , falls  $\int_{\Omega} \delta e \, dx(t) \neq 0$  für alle  $t \in \mathcal{A}$ .  $\square$

Abschließend soll noch gezeigt werden unter welchen Voraussetzungen sogar der Cantor-Anteil von  $\mu$  verschwindet.

**Satz 2.1.6.** *Sei  $\hat{G}(\bar{u})$  aktiv auf einem Intervall, d.h. es existieren  $t^{en}, t^{ex} \in \mathbb{R}$  mit  $0 < t^{en} < t^{ex} \leq T$  und  $\hat{G}(\bar{u}) = 0$  auf  $[t^{en}, t^{ex}]$ , und seien die stärkeren Annahmen des Kapitels erfüllt. Seien zusätzlich  $\delta_t \in AC^{1,2}(0, T; H^1(\Omega)^*)$ ,  $y_{\max} \in AC^{2,1}(0, T)$  und sei  $\int_{\Omega} \delta e \, dx(t) \neq 0$  für alle  $t \in [t^{en}, t^{ex}]$ . Dann sind die optimale Steuerung  $\bar{u}$  und die Funktion  $\mu$  absolut stetig auf  $[t^{en}, t^{ex}]$ .*

*Beweis.* Für alle  $t \in [t^{en}, t^{ex}]$  gilt:

$$\begin{aligned} 0 &= \frac{d}{dt} \hat{G}(\bar{u})(t) \\ &= \langle -A\bar{y}(t), \delta(t) \rangle_{H^1(\Omega)^*, H^1(\Omega)} + \bar{u}(t) (e(t), \delta(t))_{L^2(\Omega)} + \langle \delta_t(t), \bar{y}(t) \rangle_{H^1(\Omega)^*, H^1(\Omega)} \\ &\quad - \frac{d}{dt} y_{\max}(t) \end{aligned}$$

$$\int_{\Omega} \delta e \, dx(t) \neq 0 \Rightarrow \bar{u}(t) = \frac{\langle A\bar{y}(t), \delta(t) \rangle_{H^1(\Omega)^*, H^1(\Omega)} - \langle \delta_t(t), \bar{y}(t) \rangle_{H^1(\Omega)^*, H^1(\Omega)} + \frac{d}{dt} y_{\max}(t)}{(e(t), \delta(t))_{L^2(\Omega)}}.$$

Da alle Ausdrücke des Zählers und des Nenners wegen der geforderten Regularität der Daten absolut stetig sind, ergibt sich die absolute Stetigkeit von  $\bar{u}$  auf  $[t^{en}, t^{ex}]$ . Die absolute Stetigkeit der Funktion  $\mu$  folgt hieraus wegen

$$\mu(t) = \frac{\lambda \bar{u}(t) + (q(t), e(t))_{L^2(\Omega)}}{(\delta(t), e(t))_{L^2(\Omega)}}.$$

$\square$

### 2.1.3. Zusammenhang mit der optimalen Steuerung gewöhnlicher Differentialgleichungen

In diesem Kapitel sollen zum einen die Ergebnisse der bisherigen Kapitel in den Kontext der optimalen Steuerung gewöhnlicher Differentialgleichungen gestellt und zum anderen auf Besonderheiten bei der optimalen Steuerung partieller Differentialgleichungen eingegangen werden. Im Fokus steht dabei nicht, Resultate in der allgemeinsten Form zu präsentieren, sondern exemplarische Beispiele zu liefern. Ein für dieses Kapitel interessanter Überblick über Begrifflichkeiten findet sich in Gerdt's [24, Kap. 3.2.3].

**Definition 2.1.1.** Mit der Funktion  $g: [0, T] \rightarrow \mathbb{R}$ ,  $g(t) := \hat{G}(u)(t)$  in der Zustandsbeschränkung definiert man:

- Falls  $g(t) = 0$  für alle  $t \in [t^{\text{en}}, t^{\text{ex}}] \subseteq [0, T]$  mit  $t^{\text{en}} < t^{\text{ex}}$ , so heißt  $[t^{\text{en}}, t^{\text{ex}}]$  *Randstück*. Falls ein  $\delta > 0$  existiert, so dass  $g(t) < 0$  für  $t \in [t^{\text{en}} - \delta, t^{\text{en}}]$  bzw.  $t \in (t^{\text{ex}}, t^{\text{ex}} + \delta]$ , so bezeichnet man  $t^{\text{en}}$  bzw.  $t^{\text{ex}}$  als *Aufsprungpunkt* bzw. *Absprungpunkt*.
- Falls  $t^{\text{to}} \in [0, T]$  die Bedingung  $g(t^{\text{to}}) = 0$  und  $g(t) < 0$  für alle  $t \in [t^{\text{to}} - \delta, t^{\text{to}} + \delta]$  mit  $\delta > 0$  erfüllt, so nennt man diesen Zeitpunkt *Kontaktpunkt*. Ist zusätzlich  $[\frac{d}{dt}g(t^{\text{to}})] = 0$  so bezeichnet man diesen Zeitpunkt als *Berührungspunkt*.

Für dieses Kapitel sei die folgende Annahme getroffen:

**Annahme 2.1.4.** Es seien alle Annahmen des vorhergehenden Kapitels 2.1.2 erfüllt, außerdem gelte  $\int_{\Omega} e \delta dx(t) \neq 0$  für alle  $t \in I$ . Des Weiteren setze sich die aktive Menge  $\mathcal{A}$  der zur optimalen Steuerung  $\bar{u}$  gehörigen Funktion in der Zustandsbeschränkung  $\bar{g}$  aus  $n$  (disjunkten) Randstücken und  $m$  Kontaktpunkten zusammen ( $n, m \in \mathbb{N}_0$ ) und es gelte  $\mathcal{A} \subset (0, T)$ .

Als erstes Ergebnis lässt sich an dieser Stelle festhalten, dass am Optimum  $\bar{u}$  des Problems (2.5) jeder auftretende Kontaktpunkt ein Berührungspunkt sein muss, da nach Satz 2.1.5 die Zustandsschranke ausgewertet für die optimale Steuerung die erhöhte Regularität  $\bar{g} \in C^1([0, T])$  besitzt.

Es ist zweckmäßig die folgenden Mengen einzuführen (falls die entsprechenden Mengen nicht leer sind).

**Definition 2.1.2.** Zu einer aktiven Menge  $\mathcal{A}$  korrespondieren:

- Die Menge der Aufsprungpunkte  $\tau_{\text{en}} = \{t_1^{\text{en}} < t_2^{\text{en}} < \dots < t_n^{\text{en}}\}$ ,
- die Menge der Absprungpunkte  $\tau_{\text{ex}} = \{t_1^{\text{ex}} < t_2^{\text{ex}} < \dots < t_n^{\text{ex}}\}$ ,
- die Menge der Berührungspunkte  $\tau_{\text{to}} = \{t_1^{\text{to}} < t_2^{\text{to}} < \dots < t_m^{\text{to}}\}$ ,
- sowie die Menge der Randstücke  $\mathcal{A}_b = \bigcup_{i=1}^n [t_i^{\text{en}}, t_i^{\text{ex}}]$ .

Mit dieser Definition lässt sich die aktive Menge schreiben als  $\mathcal{A} = \tau_{\text{to}} \cup \mathcal{A}_b$ . Um eine weitere interessante Aussage treffen zu können, soll zunächst die *Ordnung der Zustandsbeschränkung* bestimmt werden. Unter Verwendung der Zustandsgleichung (2.3) ergibt

sich für  $g = \int_{\Omega} y \delta \, dx - y_{\max}$  in  $[0, T]$ :

$$\frac{d}{dt}g = \langle y_t, \delta \rangle_{H^1(\Omega)^*, H^1(\Omega)} + \langle \delta_t, y \rangle_{H^1(\Omega)^*, H^1(\Omega)} - \frac{d}{dt}y_{\max} \quad (2.14a)$$

$$= \langle -Ay, \delta \rangle_{H^1(\Omega)^*, H^1(\Omega)} + u(t) \langle e, \delta \rangle_{L^2(\Omega)} + \langle \delta_t, y \rangle_{H^1(\Omega)^*, H^1(\Omega)} - \frac{d}{dt}y_{\max}. \quad (2.14b)$$

Da die Steuerung in der ersten Ableitung explizit auftritt und  $\int_{\Omega} e \delta \, dx(t) \neq 0$  für alle  $t \in I$  nach Annahme 2.1.4 eingehalten ist, besitzt die Zustandsbeschränkung die Ordnung 1. Überträgt man die Ergebnisse aus der optimalen Steuerung von gewöhnlichen Differentialgleichungen zum Verhalten der Funktionen im Optimalitätssystem an Berühr-, Auf- und Absprungpunkten nach Jacobson, Lele und Speyer [39] und Maurer [44] (vgl. [9]), so dürften bei der Zustandsbeschränkung erster Ordnung nur nicht essenzielle Berührungspunkte auftreten, wenn die verschärfte Legendre-Clebsch-Bedingung eingehalten ist. Für jeden Berührungspunkt  $t^{\text{to}} \in (0, T)$  müsste also gelten  $[\mu(t^{\text{to}})] = 0$ . Entsprechendes garantiert Satz 2.1.5. An Auf- und Absprungpunkten eines Randstücks hat man gemäß Satz 2.1.5 außerdem die Stetigkeit der Steuerung  $\bar{u}$  und der Funktion  $\mu$ , was ebenfalls exakt der Theorie der optimalen Steuerung gewöhnlicher Differentialgleichungen entspricht.

In einem nächsten Schritt soll aus dem im Kapitel 2.1.2 eingeführten alternativen Optimalitätssystem ein weiteres Optimalitätssystem hergeleitet werden, welches bei der optimalen Steuerung gewöhnlicher Differentialgleichungen typischerweise für Schießverfahren genutzt wird [10]. Ein Überblick über verschiedene Formulierungen von Optimalitätssystemen ist in Hartl, Sethi und Vickson [31] aufgeführt. Eine mathematisch exakte Herleitung des folgenden Systems für einen allgemeinen Fall im Kontext der optimalen Steuerung von gewöhnlichen Differentialgleichungen findet sich in Bonnans und Hermant [12]. Für ein spezielles Problem bei partiellen Differentialgleichungen wird das resultierende Optimalitätssystem und die grundsätzliche Herangehensweise in Bonnans und Jaisson [11] dargestellt. Diese Herangehensweise soll an dieser Stelle mit einigen Zwischenschritten rekapituliert werden.

Zunächst wird die Zustandsbeschränkung  $\bar{g}(t) \leq 0$  für alle  $t \in [0, T]$  in die Bedingungen  $\bar{g}(t) < 0$  für alle  $t \in [0, T] \setminus \mathcal{A}$  und  $\bar{g}(t) = 0$  für alle  $t \in \mathcal{A} = \tau_{\text{to}} \cup \mathcal{A}_b$  umformuliert. Im Rahmen des *BDD-Ansatzes* [14, 15] wird die algebraische Gleichung  $\bar{g} = 0$  auf Randstücken  $\mathcal{A}_b$  ersetzt durch ein Anfangswertproblem mit der Ordnung der Zustandsbeschränkung, in dem Fall erster Ordnung also  $\bar{g}(t_i^{\text{en}}) = 0$ ,  $\frac{d}{dt}\bar{g} = 0$  in  $[t_i^{\text{en}}, t_i^{\text{ex}}]$  für  $i = 1, \dots, n$ . Die differentielle Bedingung wird anschließend wie bei der Herleitung der Ordnung der Zustandsbeschränkung (2.14) in eine algebraische Bedingung umgeschrieben, die explizit von der Steuerung abhängig ist:

$$0 = \frac{d}{dt}\bar{g} \quad (2.15a)$$

$$= \langle -A\bar{y}, \delta \rangle_{H^1(\Omega)^*, H^1(\Omega)} + \bar{u} \langle e, \delta \rangle_{L^2(\Omega)} + \langle \delta_t, \bar{y} \rangle_{H^1(\Omega)^*, H^1(\Omega)} - \frac{d}{dt}y_{\max} \quad \text{in } \mathcal{A}_b. \quad (2.15b)$$

Da alle auftretenden Berührungspunkte nicht essenziell sind, kann auf die Bedingung  $\bar{g}(t) = 0$  für alle  $t \in \tau_{\text{to}}$  bei der weiteren Betrachtung verzichtet werden. Diese Bedingung wird am Optimum ohnehin eingehalten. Es gilt sich nun noch der den Lagrange-Multiplikator  $d\mu$

erzeugenden Funktion  $\mu$  zuzuwenden. Bei den im Rahmen der Schießverfahren betrachteten Optimalitätssysteme soll die Funktion  $\mu$  außerhalb der aktiven Menge verschwinden, was sich folgerichtig auf die adjungierte Gleichung auswirkt. Die entsprechenden Beziehungen können aus dem bisherigen alternativen Optimalitätssystem in Satz 2.1.4 wie folgt hergeleitet werden.

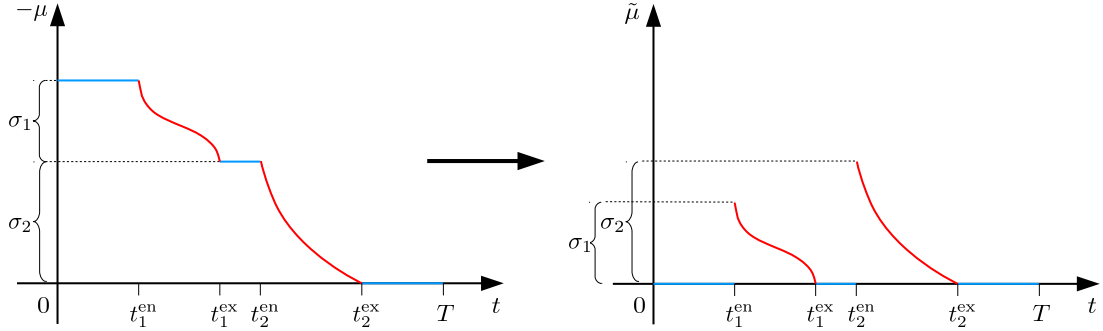


Abbildung 2.1.: Zusammenhang zwischen  $\mu$  und  $\tilde{\mu}$

Man zerlegt die monoton steigende, stetige Funktion  $\mu \in NBV_T(0, T)$  additiv in

$$-\mu = \tilde{\mu} + \sum_{i=1}^n \mathbf{1}_{[0, t_i^{\text{en}})} \cdot \sigma_i$$

mit  $\tilde{\mu} \in BV(0, T)$  und

$$\sigma_i := -\mu(t_i^{\text{en}}) - (-\mu(t_i^{\text{ex}})) = \mu(t_i^{\text{ex}}) - \mu(t_i^{\text{en}}) \quad (\geq 0)$$

sowie

$$\mathbf{1}_{[0, t_i^{\text{en}})}(t) = \begin{cases} 1 & \text{für } t \in [0, t_i^{\text{en}}), \\ 0 & \text{für } t \in [t_i^{\text{en}}, T], \end{cases}$$

für  $i = 1, \dots, n$ . Die Funktion  $\tilde{\mu}$  ist an den Aufsprungpunkten nicht mehr stetig, verschwindet aber wie gewünscht außerhalb  $\mathcal{A}_b$  und ist wegen der Monotonie von  $\mu$  auf  $\mathcal{A}_b$  monoton fallend. In den Absprungpunkten ist sie nach Konstruktion weiterhin stetig, so dass  $\lim_{h \rightarrow 0} \tilde{\mu}(t_i^{\text{ex}} - h) = 0$  für  $i = 1, \dots, n$  gelten muss. Insbesondere gilt auch  $\tilde{\mu} \geq 0$ . Die neu eingeführten Sprungparameter  $\sigma_i$  müssen wegen der Stetigkeit von  $\mu$  die Bedingung  $\sigma_i = \lim_{h \rightarrow 0} \tilde{\mu}(t_i^{\text{en}} + h)$  für  $i = 1, \dots, n$  erfüllen. Mit der Zerlegung von  $\mu$  lässt sich nun ein neuer alternativer adjungierter Zustand  $\tilde{q}$  einführen. Hierzu betrachte man die Definition der alternativen Adjungierten aus Satz 2.1.4:

$$q = p + \mu\delta = \underbrace{p - \tilde{\mu}\delta}_{=: \tilde{q}} - \sum_{i=1}^n \mathbf{1}_{[0, t_i^{\text{en}})} \sigma_i \delta.$$

Diese Funktion  $\tilde{q} \in L^2(0, T; H^1(\Omega))$  besitzt die distributionelle Zeitableitung ( $\varphi \in C_0^\infty(0, T)$ )

$$\tilde{q}'(\varphi) = \int_0^T q_t \varphi \, d\tau - \sum_{i=1}^n \delta(t_i^{\text{en}}) \sigma_i \varphi(t_i^{\text{en}}) + \sum_{i=1}^n \mathbf{1}_{[0, t_i^{\text{en}})} \sigma_i \int_0^T \delta_t \varphi \, d\tau$$

und erfüllt

$$A\tilde{q} = Aq + \sum_{i=1}^n \mathbf{1}_{[0, t_i^{\text{en}})} \sigma_i A\delta.$$

Aus der adjungierten Gleichung für  $q$  lassen sich deshalb mittels der Definition von  $\tilde{\mu}$  und  $\tilde{q}$  mit  $\langle F(t), v \rangle_{H^1(\Omega)^*, H^1(\Omega)} = \int_{\Omega} (\hat{S}(\bar{u})(t) - y_d(t))v \, dx$  für alle  $v \in H^1(\Omega)$  die Beziehungen

$$\begin{aligned} -\tilde{q}'(t) + A\tilde{q}(t) &= F(t) - \tilde{\mu}(t)(-\delta'(t) + A\delta(t)) \quad \text{f.ü. in } [0, T] \setminus (\tau_{\text{en}} \cup \tau_{\text{ex}}), \\ \tilde{q}(T) &= 0, \\ [\tilde{q}(t_i^{\text{en}})] &= -\sigma_i \delta(t_i^{\text{en}}) \quad i = 1, \dots, n, \\ [\tilde{q}(t_i^{\text{ex}})] &= 0 \quad i = 1, \dots, n \end{aligned}$$

herleiten, wobei analog zum reellwertigen Fall  $[q(t)]$  den Sprung der Funktion  $q$  zum Zeitpunkt  $t \in (0, T)$  bezeichne. Die resultierende Bedingung aus der Sattelpunkteigenschaft der Lagrangefunktion lautet mit der Definition von  $\tilde{q}$ :

$$\lambda \bar{u}(t) + \int_{\Omega} (\tilde{q} + \tilde{\mu}\delta)e \, dx(t) = 0 \quad \text{f.ü. in } [0, T].$$

Alles in allem erhält man das folgende Optimalitätssystem:

- Zustandsgleichung:

$$\bar{y}'(t) + A\bar{y}(t) = \bar{u}(t) \cdot E(t) \quad \text{f.ü. in } [0, T], \quad (2.16a)$$

$$\bar{y}(0) = y_0, \quad (2.16b)$$

mit  $\langle E(t), v \rangle_{H^1(\Omega)^*, H^1(\Omega)} = \int_{\Omega} e(x, t)v(x) \, dx$  für alle  $v \in H^1(\Omega)$ ,

- adjungierte Gleichung:

$$-\tilde{q}'(t) + A\tilde{q}(t) = F(t) - \tilde{\mu}(t)(-\delta'(t) + A\delta(t)) \quad \text{f.ü. in } [0, T] \setminus (\tau_{\text{en}} \cup \tau_{\text{ex}}), \quad (2.16c)$$

$$\tilde{q}(T) = 0, \quad (2.16d)$$

$$[\tilde{q}(t_i^{\text{en}})] = -\sigma_i \delta(t_i^{\text{en}}) \quad i = 1, \dots, n, \quad (2.16e)$$

$$[\tilde{q}(t_i^{\text{ex}})] = 0 \quad i = 1, \dots, n, \quad (2.16f)$$

mit  $\langle F(t), v \rangle_{H^1(\Omega)^*, H^1(\Omega)} = \int_{\Omega} (\bar{y}(t) - y_d(t))v \, dx$  für alle  $v \in H^1(\Omega)$ ,

- Sattelpunkteigenschaft der Lagrangefunktion:

$$\lambda \bar{u}(t) + \int_{\Omega} (\tilde{q} + \tilde{\mu}\delta)e \, dx(t) = 0 \quad \text{f.ü. in } [0, T], \quad (2.16g)$$

- Bedingung auf Randstücken:

$$\bar{g}(t_i^{\text{en}}) = 0, \quad \frac{d}{dt}\bar{g} = 0 \text{ in } [t_i^{\text{en}}, t_i^{\text{ex}}] \text{ in Variante (2.15),} \quad i = 1, \dots, n, \quad (2.16h)$$

- Bedingung außerhalb von Randstücken:

$$\tilde{\mu} = 0 \quad \text{in } I \setminus \mathcal{A}_b, \quad (2.16i)$$

- Bedingung an Auf- und Absprungpunkten:

$$\sigma_i = \lim_{h \searrow 0} \tilde{\mu}(t_i^{\text{en}} + h), \quad 0 = \lim_{h \searrow 0} \tilde{\mu}(t_i^{\text{ex}} - h), \quad i = 1, \dots, n, \quad (2.16j)$$

- Bedingung außerhalb von Randstücken:

$$\bar{g} < 0 \quad \text{f.ü. in } I \setminus \mathcal{A}_b, \quad (2.16k)$$

- Bedingung auf Randstücken:

$$\tilde{\mu} \text{ monoton fallend in } \mathcal{A}_b. \quad (2.16l)$$

Bei der Umsetzung eines Schießalgorithmus bereitet die adjungierte Gleichung jedoch große Probleme, da entgegen der Diffusionsrichtung integriert werden müsste. Im Rahmen dieser Arbeit wird deshalb in Teil 3 ein anderes Verfahren zur Lösung dieser Probleme verfolgt.

An dieser Stelle sei auch auf eine in einem gewissen Sinne umgekehrte Herangehensweise zur Herleitung der Optimalitätsbedingungen hingewiesen. Bei dieser Herangehensweise zerteilt man die Zustandsgleichung in aktive und inaktive Intervalle bereits vor der Herleitung der Optimalitätsbedingungen. Anschließend führt man auf der aktiven Menge, wie oben beschrieben, den BDD-Ansatz aus und versucht für das resultierende Optimierungsproblem die Korrektheit des Lagrangeformalismus zu verifizieren. Man hat dann von Anfang an Lagrange-Multiplikatoren zur Zustandsbeschränkung, die nur auf der aktiven Menge leben. Ein solcher Ansatz wurde im Kontext einer elliptischen partiellen Differentialgleichungen beispielsweise in Frey [22] verfolgt. Der hier gewählte Ansatz erscheint im Vergleich aber für ein parabolisches Problem deutlich zweckmäßiger.

Abschließend soll auf Probleme bzw. Besonderheiten eingegangen werden, die beim Übertragen der Beweistechnik des alternativen Optimalitätssystems von gewöhnlichen zu partiellen Differentialgleichungen zu beachten sind.

Bei parabolischen Gleichungen tritt neben dem Differentialoperator in der Zeit auch der Operator  $A$  auf. Diese Tatsache wirkt sich maßgeblich auf die zu erwartende Regularität des alternativen adjungierten Zustands  $q$  aus. In der Herleitung der Differentialgleichung für den alternativen adjungierten Zustand im Beweis des Satzes 2.1.4 wurde die höhere (Orts-)Regularität der Mittelungsfunktion  $\delta \in L^2(0, T; H^1(\Omega))$  genutzt. So konnte man die im Vergleich zur klassischen Adjungierten  $p$  höhere Regularität von  $q$  als Lösung einer regulären Differentialgleichung herleiten,  $q \in W(0, T)$ . Kann aber der Operator  $A: H^1(\Omega) \rightarrow H^1(\Omega)^*$  nicht mehr auf  $\delta(t)$  angewandt werden, so muss mit einem anderen Konzept eine Differentialgleichung von  $q$  in schwacher Formulierung hergeleitet werden.  $q$  wird als Lösung der resultierenden Differentialgleichung nur noch eine niedrigere Regularität besitzen. Der alternative adjungierte Zustand ist also nicht zwangsläufig



regulärer als der klassische. Nichtsdestotrotz ist es möglich, gewinnbringend mit einer derartigen Adjungierten zu arbeiten, wie in Kapitel 2.1.5 ausgeführt wird.

Eine weitere Besonderheit bei der Übertragung des Konzeptes des alternativen Optimalitätssystems auf die optimale Steuerung partieller Differentialgleichungen wird bei der Betrachtung der Zustandsbeschränkung offenkundig. In der Zustandsbeschränkung taucht der linear stetige Integraloperator  $W(0, T) \rightarrow C([0, T])$ ,  $y \mapsto \int_{\Omega} y \delta dx$  auf. Diese Darstellung in integraler Form erlaubt es in Satz 2.1.4, einen neuen verteilten adjungierten Zustand  $q := p + \mu \delta$  einzuführen. Entscheidend hierbei ist, dass der Term  $\int_I (\int_{\Omega} \varphi \delta dx) d\mu$  für Testfunktionen  $\varphi \in W(0, T)$  mit  $\varphi(0) = 0$  umformuliert werden kann (vgl. Gleichung (2.11)):

$$\int_I \left( \int_{\Omega} \varphi \delta dx \right) d\mu = - \int_I \langle \varphi_t, \mu \delta \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt - \int_I \mu \langle \delta_t, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt.$$

Würde aber in der Zustandsbeschränkung statt des Operators in integraler Form ein beliebiger linearer Operator  $B$ , der den Raum der Zustände auf  $C([0, T])$  abbildet, auftreten, so wäre bei entsprechender Regularität von  $B\varphi$  nur noch eine Umformung der Form

$$\int_I B\varphi d\mu = - \int_I (B\varphi)_t \mu dt$$

möglich. Es kann also im Unterschied zum Fall der optimalen Steuerung gewöhnlicher Differentialgleichungen bei Zustandsbeschränkungen der Form  $By - y_{\max} \leq 0$  nicht stets ein (verteilter) alternativer adjungierter Zustand eingeführt werden. Unter gewissen Strukturvoraussetzungen ist es in Kapitel 2.2 mit einer Modifikation der bisherigen Beweistechnik jedoch möglich, auch für ein Optimalsteuerungsproblem mit einer parabolischen Zustandsgleichung und nichtintegraler Darstellung der Zustandsbeschränkung weitreichende Aussagen zu treffen.

#### 2.1.4. Erhöhte Regularitäten mit geschicktem Testen

Es stellt sich die Frage, inwiefern eine erhöhte Regularität der optimalen Steuerung ohne Erhöhung der Ortsregularität der Mittelungsfunktion  $\delta$  erreicht werden kann. Betrachtet man die Ausarbeitungen in Kapitel 2.1.2, so erkennt man, dass ein wesentlicher Punkt in der dortigen Beweisführung die Einführung des alternativen adjungierten Zustands  $q$  unter Verwendung erhöhter Ortsregularität  $\delta \in L^2(0, T; H^1(\Omega))$  ist. Mit diesem konnte die Bedingung

$$\lambda \bar{u}(t) = - \int_{\Omega} p e dx(t) = - \int_{\Omega} q e dx(t) + \mu(t) \int_{\Omega} \delta e dx(t) \quad \text{f.ü. in } [0, T]$$

hergeleitet werden. Aufgrund der speziellen Struktur des betrachteten Optimalsteuerungsproblems ist es aber wegen der anschließenden Mittelung mit der Funktion  $e$  eigentlich nicht nötig, eine Zerlegung der Adjungierten  $p$  an sich vorzunehmen. Es genügt, die rein zeitabhängige Funktion  $\int_{\Omega} p e dx$  additiv zu zerlegen und dabei die Funktion  $\mu$  zu isolieren. Dies gelingt in den folgenden Ausarbeitungen durch geschicktes Testen der adjungierten Gleichung. Zur Durchführung muss die folgende Zusatzforderung gestellt werden.

**Annahme 2.1.5.** Die Datenfunktionen  $\delta$  und  $e$  erfüllen die höheren Regularitätsanforderungen:

- $\delta \in AC^{1,1}([0, T]; L^2(\Omega))$ ,
- $e \in W(0, T) \cap AC^{1,1}([0, T]; L^2(\Omega))$ .

**Satz 2.1.7.** Seien die Voraussetzungen des Modellproblems (2.1), die Slater-Bedingung aus Annahme 2.1.1 und die Annahme 2.1.5 erfüllt. Sei  $\bar{u} \in L^2(0, T)$  die optimale Lösung des Problems (2.5). Dann existieren eine Funktion  $\mu \in NBV_T(0, T)$  sowie eine Funktion  $q_e \in AC^{1,1}(0, T)$ , so dass die folgenden Bedingungen erfüllt sind:

- *Sattelpunkteigenschaft der Lagrangefunktion:*

$$\lambda \bar{u}(t) + q_e(t) - \mu(t) \int_{\Omega} \delta e \, dx(t) = 0 \quad \text{f.ü. in } [0, T], \quad (2.17)$$

- *komplementäre Schlupfbedingung:*  $\mu$  ist monoton steigend auf  $[0, T]$  und konstant auf jedem Intervall  $(t_1, t_2) \subset [0, T]$  mit  $t_1 < t_2$  und  $\hat{G}(\bar{u})(t) < 0$  für alle  $t \in (t_1, t_2)$ .

*Beweis.* Der Satz 2.1.3 garantiert die Existenz einer Adjungierten  $p \in L^2(0, T; H^1(\Omega)) \cap L^\infty(0, T; L^2(\Omega)) \cap BV(0, T; H^1(\Omega)^*)$  und einer Funktion  $\mu \in NBV_T(0, T)$ , so dass die adjungierte Gleichung

$$\begin{aligned} & \int_I \langle \varphi_t, p \rangle_{H^1(\Omega)^*, H^1(\Omega)} \, dt + \int_I \langle Ap, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} \, dt \\ &= \iint_Q (\hat{S}(\bar{u}) - y_d) \varphi \, dx dt + \iint_Q \delta \varphi \, dx d\mu \quad \forall \varphi \in W(0, T) \text{ mit } \varphi(0) = 0 \end{aligned}$$

erfüllt ist. Man betrachte nun die Funktion  $h \cdot e$  mit beliebigem  $h \in C_0^\infty(0, T)$ . Diese liegt für  $e \in W(0, T)$  in  $W(0, T)$  und erfüllt  $h(0) \cdot e(0) = 0$ . Sie ist also eine zulässige Testfunktion. Eingesetzt erhält man

$$\begin{aligned} & \int_0^T \langle e_t, p \rangle_{H^1(\Omega)^*, H^1(\Omega)} h \, dt + \int_0^T (e, p)_{L^2(\Omega)} h_t \, dt + \int_0^T \langle Ap, e \rangle_{H^1(\Omega)^*, H^1(\Omega)} h \, dt \\ &= \int_0^T (\hat{S}(\bar{u}) - y_d, e)_{L^2(\Omega)} h \, dt + \int_0^T (\delta, e)_{L^2(\Omega)} h \, d\mu. \end{aligned}$$

Aufgrund der gemachten Zusatzvoraussetzungen  $\delta, e \in AC^{1,1}([0, T]; L^2(\Omega))$  ist die zeitabhängige Funktion  $(\delta, e)_{L^2(\Omega)}$  absolut stetig. Somit lässt sich der letzte Summand mittels partieller Integration für Stieltjes-Integrale [24, S. 66, 48, 11] unter Verwendung von  $h \in C_0^\infty(0, T)$  umformen in:

$$\int_0^T (\delta, e)_{L^2(\Omega)} h \, d\mu = - \int_0^T \frac{d}{dt} \left( (\delta, e)_{L^2(\Omega)} \right) \mu h \, dt - \int_0^T (\delta, e)_{L^2(\Omega)} \mu \frac{d}{dt} h \, dt.$$

Eingesetzt und sortiert ergibt sich für alle  $h \in C_0^\infty(0, T)$

$$\int_0^T \left( (e, p)_{L^2(\Omega)} + (\delta, e)_{L^2(\Omega)} \mu \right) h_t dt \quad (2.18a)$$

$$= - \int_0^T \langle e_t, p \rangle_{H^1(\Omega)^*, H^1(\Omega)} h dt - \int_0^T \langle Ap, e \rangle_{H^1(\Omega)^*, H^1(\Omega)} h dt \quad (2.18b)$$

$$+ \int_0^T \left( \hat{S}(\bar{u}) - y_d, e \right)_{L^2(\Omega)} h dt - \int_0^T \frac{d}{dt} \left( (\delta, e)_{L^2(\Omega)} \right) \mu h dt. \quad (2.18c)$$

Setzt man  $q_e : [0, T] \rightarrow \mathbb{R}$  mit

$$q_e(t) := \int_{\Omega} p e dx(t) + \mu(t) \int_{\Omega} \delta e dx(t),$$

so hat man mit den gegebenen Zusatzvoraussetzungen  $q_e \in L^\infty(0, T)$ . Außerdem ist  $q_e$  wegen Gleichung (2.18) schwach differenzierbar mit Ableitung

$$\begin{aligned} (q_e)_t &= \underbrace{\langle e_t, p \rangle_{H^1(\Omega)^*, H^1(\Omega)}}_{\in L^1(0, T)} + \underbrace{\langle Ap, e \rangle_{H^1(\Omega)^*, H^1(\Omega)}}_{\in L^1(0, T)} - \underbrace{\left( \hat{S}(\bar{u}) - y_d, e \right)_{L^2(\Omega)}}_{\in L^2(0, T)} \\ &\quad + \underbrace{\frac{d}{dt} \left( (\delta, e)_{L^2(\Omega)} \right)}_{\in L^1(0, T)} \underbrace{\mu}_{\in L^\infty(0, T)} \end{aligned}$$

in  $L^1(0, T)$ . Nach Barbu und Precupanu [4, Ch. 1] kann der Raum  $W^{1,1}(0, T)$  aber mit dem Raum  $AC^{1,1}(0, T)$  identifiziert werden. Aus der Sattelpunkteigenschaft der Lagrangefunktion (vgl. Satz 2.1.3) und mit der Definition von  $q_e$  folgt schließlich fast überall in  $[0, T]$ :

$$0 = \lambda \bar{u}(t) + \int_{\Omega} p e dx(t) = \lambda \bar{u}(t) + q_e(t) - \mu(t) \int_{\Omega} \delta e dx(t).$$

Die Formulierung der komplementären Schlupfbedingung findet man bereits in Satz 2.1.3.  $\square$

Aus dem obigen Satz lassen sich mit analoger Beweistechnik zu Kapitel 2.1.2 diverse Schlüsse ziehen. Die Beweise sind deshalb in diesem Kapitel kurz gehalten. Wegen der Annahme 2.1.5 erhält man aus der Bedingung (2.17) mit der absoluten Stetigkeit von  $\int_{\Omega} \delta e dx$  und  $q_e$  unmittelbar:

**Korollar 2.1.4.** *Unter den Annahmen 2.1.1, 2.1.5 gehört die optimale Steuerung  $\bar{u}$  des Problems (2.5) dem Raum  $BV(0, T)$  an. Ist  $\int_{\Omega} \delta e dx = 0$  in  $[0, T]$ , so hat man  $\bar{u} \in AC^{1,1}(0, T)$ .*

Fordert man  $Ay_0 \in L^2(\Omega)$ , so folgt mit der erhöhten Regularität der optimalen Steuerung  $\bar{u} \in BV(0, T)$  und dem Gelfand-Dreier  $H^1(\Omega) \subset L^2(\Omega) \subset H^1(\Omega)^*$  wegen

$e \in W(0, T)$  gemäß Satz A.0.1 im Anhang die erhöhte Regularität des optimalen Zustands  $\bar{y} = \hat{S}(\bar{u})$ :

$$\bar{y} \in L^2(0, T; H^1(\Omega)) \text{ und } \bar{y}' \in L^2(0, T; H^1(\Omega)) \cap BV(0, T; H^1(\Omega)^*) \quad (2.19)$$

sowie  $A\bar{y} \in W(0, T)$  als Lösung einer Evolutionsgleichung. Dieses Resultat erlaubt es auch unter den Voraussetzungen dieses Kapitels die Stetigkeit der optimalen Steuerung nachzuweisen.

**Satz 2.1.8.** *Seien neben den Grundvoraussetzungen und den Annahmen 2.1.1, 2.1.5 die Bedingungen  $Ay_0 \in L^2(\Omega)$ ,  $y_{\max} \in C^1([0, T])$  und  $\delta_t \in C([0, T]; H^1(\Omega)^*)$  erfüllt. Dann ist die optimale Lösung  $\bar{u}$  des Problems (2.5) und die Funktion  $\frac{d}{dt}\bar{g} = \frac{d}{dt}(\hat{G}(\bar{u}))$  stetig auf  $[0, T]$ . Außerdem ist  $\bar{u}$  absolut stetig auf der inaktiven Menge  $\mathcal{I}$ . Wird die Bedingung  $\int_{\Omega} \delta e \, dx(t) \neq 0$  für alle  $t$  in der aktiven Menge  $\mathcal{A}$  eingehalten, so besitzt  $\mu \in BV(0, T)$  keinen diskreten Anteil,  $\mu \in NBV_T(0, T) \cap C([0, T])$ .*

*Beweis.* Aus der erhöhten Regularität von  $\bar{y}$  folgt insbesondere  $\bar{y} \in AC^{1,2}([0, T]; H^1(\Omega))$ , so dass mit der Bedingung  $\delta \in AC^{1,1}([0, T]; L^2(\Omega))$  die zeitabhängige Funktion  $\int_{\Omega} \bar{y}\delta \, dx$  absolut stetig ist. Mit  $y_{\max} \in C^1([0, T])$  ist somit auch  $\bar{g} = \hat{G}(\bar{u})$  absolut stetig und fast überall in  $[0, T]$  gilt mit der Zustandsgleichung:

$$\begin{aligned} \frac{d}{dt}\bar{g}(t) &= (\bar{y}_t(t), \delta(t))_{L^2(\Omega)} + (\delta_t(t), \bar{y}(t))_{L^2(\Omega)} - \frac{d}{dt}y_{\max}(t) \\ &= (-A\bar{y}(t) + \bar{u}(t)e(t), \delta(t))_{L^2(\Omega)} + (\delta_t(t), \bar{y}(t))_{L^2(\Omega)} - \frac{d}{dt}y_{\max}(t). \end{aligned}$$

Aufgrund der Ortsregularität von  $\delta_t \in L^1(0, T; L^2(\Omega))$  und  $\bar{y} \in AC^{1,2}([0, T]; H^1(\Omega))$  kann man für  $t \in [0, T]$  umformen:  $(\delta_t(t), \bar{y}(t))_{L^2(\Omega)} = \langle \delta_t(t), \bar{y}(t) \rangle_{H^1(\Omega)^*, H^1(\Omega)}$ . Wegen  $e, A\bar{y} \in W(0, T) \hookrightarrow C([0, T]; L^2(\Omega))$ ,  $\delta_t \in C([0, T]; H^1(\Omega)^*)$  und  $\bar{u} \in BV(0, T)$  gilt für den Sprung der Funktion  $\bar{g} = \hat{G}(\bar{u})$  zu einem Zeitpunkt  $t \in [0, T]$  schließlich:

$$\left[ \frac{d}{dt}\bar{g}(t) \right] = [\bar{u}(t)] \cdot (e(t), \delta(t))_{L^2(\Omega)}.$$

Der Satz 2.1.7 liefert die weiteren Zutaten, so dass man vollkommen analog zum Beweis von Satz 2.1.5 vorgehen kann, um die gewünschten Ergebnisse zu beweisen.  $\square$

**Bemerkung 2.1.5.** Im Unterschied zum vorherigen Kapitel ist es nicht mehr ohne Weiteres möglich nachzuweisen, dass auf aktiven Intervallen der Cantor-Anteil von  $\mu$  verschwindet und  $\bar{u}$  absolut stetig ist. Dies liegt an der Funktion  $(A\bar{y}, \delta)_{L^2(\Omega)}$ , die ohne weitere Zusatzforderungen nur in  $C([0, T])$  liegt. Für absolute Stetigkeit dieser Funktion unter Beibehaltung der Ortsregularität von  $\delta$  müsste man die Regularität von  $A\bar{y}$  erhöhen. Dies könnte man dadurch bewerkstelligen, dass man die Bedingungen an die Daten  $y_0, e$  so erhöht, dass entweder  $A\bar{y}$  als Lösung einer parabolischen Differentialgleichung mit regulärer rechter Seite erhöhte Regularität besitzt (maximale parabolische Regularität) oder dass für die Funktion  $A\bar{y}$  als Lösung der Gleichung

$$\begin{aligned} (A\bar{y})'(t) + A(A\bar{y})(t) &= \bar{u}(t) \cdot Ae(t) \quad \text{f.ü. in } [0, T], \\ A\bar{y}(0) &= Ay_0 \end{aligned}$$

wieder der Regularitätssatz A.0.1 anwendbar ist. Dieses Spiel ließe sich unter Anpassung der Daten natürlich beliebig fortsetzen. Man beachte dabei jedoch generell, dass man sich bei Forderungen wie  $Ae \in W(0, T)$  oder  $Ae \in L^2(0, T; L^2(\Omega))$  Randbedingungen für die Funktion  $e$  einhandelt.

### 2.1.5. Erhöhte Regularitäten bei erhöhter Regularität des Zustands

Eine weitere Möglichkeit, um irreguläre Mittelungsfunktionen  $\delta$  betrachten zu können, ergibt sich, wenn die Lösungen der Zustandsgleichung an sich höhere Regularität aufweisen. Es sei deshalb exemplarisch die folgende Annahme getroffen.

**Annahme 2.1.6.** Der Rand  $\Gamma$  von  $\Omega$  besitze die Regularität  $C^{1,1}$ . Außerdem gelte  $y_0 \in H^1(\Omega)$ .

Nach Lions [42, Ch. 1] weist dann jede Lösung  $y$  der Zustandsgleichung (2.1b) - (2.1d) für  $u \in L^2(0, T)$  und  $e \in L^\infty(0, T; L^2(\Omega))$  erhöhte Regularität auf:

$$y \in W_{H_\Gamma^2(\Omega), L^2(\Omega)}(0, T) := \{ \hat{y} \in L^2(0, T; H_\Gamma^2(\Omega)) \mid \hat{y}_t \in L^2(0, T; L^2(\Omega)) \},$$

mit  $\|y\|_{W_{H_\Gamma^2(\Omega), L^2(\Omega)}(0, T)} := \|y\|_{L^2(0, T; H_\Gamma^2(\Omega))} + \|y_t\|_{L^2(0, T; L^2(\Omega))}$ . Dabei bezeichne  $H_\Gamma^2(\Omega)$  den Teilraum von  $H^2(\Omega)$ , dessen Funktionen die Randbedingung  $\partial_\nu y + y = 0$  einhalten. In kanonischer Weise versehen mit der  $H^2(\Omega)$ -Norm ist der Raum  $H_\Gamma^2(\Omega)$  aufgrund des wohlbekannten Spursatzes in Grisvard [29] für  $C^{1,1}$ -Gebiete ein abgeschlossener Unterraum von  $H^2(\Omega)$  und somit selbst wieder ein separabler reflexiver Banachraum. Des Weiteren ist der Raum dicht eingebettet in  $L^2(\Omega)$ , da er insbesondere alle  $C_0^\infty(\Omega)$  Funktionen enthält.

Aus den vorangehenden Kapiteln ist ersichtlich, dass für den Nachweis der höheren Regularität der optimalen Steuerung, zeitabhängige Funktionen wie beispielsweise  $\int_\Omega y \delta \, dx$  absolut stetig sein müssen. Die hinzugewonnene höhere Regularität für den ersten Faktor ermöglicht dabei die Reduktion der Regularität des zweiten Faktors. Entsprechendes wird in dem nachfolgenden Lemma dargestellt und kann allgemein für die Situation eines Gelfand-Dreiers im Anhang formuliert werden. Der Beweis ist an den symmetrischen Fall, in dem beide Faktoren in  $W(0, T)$  liegen, angelehnt [23, S. 147 f.]. Er unterscheidet sich aber wesentlich darin, dass eine der beiden Funktionen aufgrund mangelnder Regularität nicht mehr in den Raum  $C([0, T]; H)$  ( $H$  Hilbertraum) eingebettet werden kann.

**Lemma 2.1.4.** Sei  $f \in W_{H_\Gamma^2(\Omega), L^2(\Omega)}(0, T)$  und  $g \in W_{L^2(\Omega), H_\Gamma^2(\Omega)^*}(0, T)$ , wobei

$$W_{L^2(\Omega), H_\Gamma^2(\Omega)^*}(0, T) := \{ \hat{g} \in L^2(0, T; L^2(\Omega)) \mid \hat{g}' \in L^2(0, T; H_\Gamma^2(\Omega)^*) \}$$

mit  $\|g\|_{W_{L^2(\Omega), H_\Gamma^2(\Omega)^*}(0, T)} := \|g\|_{L^2(0, T; L^2(\Omega))} + \|g'\|_{L^2(0, T; H_\Gamma^2(\Omega)^*)}$ . Dann lässt sich für die Funktion  $[0, T] \rightarrow \mathbb{R}, t \mapsto (f(t), g(t))_{L^2(\Omega)}$  die folgende Formel der partiellen Integration für  $s, t \in [0, T]$  herleiten:

$$(f(t), g(t))_{L^2(\Omega)} = (f(s), g(s))_{L^2(\Omega)} + \int_s^t (f'(\tau), g(\tau))_{L^2(\Omega)} + \langle g'(\tau), f(\tau) \rangle_{H_\Gamma^2(\Omega)^*, H_\Gamma^2(\Omega)} \, d\tau.$$

*Beweis.* Das Lemma folgt unmittelbar aus Lemma A.0.1 im Anhang, indem man mit dem Gelfand-Dreier  $H_{\Gamma}^2(\Omega) \subset L^2(\Omega) \subset H_{\Gamma}^2(\Omega)^*$  arbeitet.  $\square$

Mit der erhöhten Regularität des Zustands unter der Annahme 2.1.6 hat man gemäß des eben bewiesenen Lemmas bereits für  $\delta \in W_{L^2(\Omega), H_{\Gamma}^2(\Omega)^*}(0, T)$  die absolute Stetigkeit von  $[0, T] \rightarrow \mathbb{R}$ ,  $t \mapsto \int_{\Omega} y(t)\delta(t) dx$ . Es sei deshalb die Annahme getroffen:

**Annahme 2.1.7.**  $\delta \in C([0, T]; L^2(\Omega)) \cap W_{L^2(\Omega), H_{\Gamma}^2(\Omega)^*}(0, T)$ .

Es stellt sich dann jedoch die Frage, inwieweit für eine Mittelungsfunktion dieser Regularität wieder ein alternativer adjungierter Zustand eingeführt werden kann. Im Vergleich zu Kapitel 2.1.2 besteht das Problem, dass auf die Funktion  $\delta$  der Operator  $A$  nicht angewandt werden kann. Der Ausweg ist die Einführung eines alternativen adjungierten Zustands, der im Vergleich zur klassischen Adjungierten  $p$  aus Kapitel 2.1.1 eine niedrigere Regularität aufweist. Die Regularität ist aber nichtsdestotrotz ausreichend, um in Kombination mit dem Lemma 2.1.4 und der erhöhten Regularität der weiteren Größen weitreichende Aussagen zu treffen.

**Satz 2.1.9.** *Seien die Voraussetzungen des Modellproblems (2.1), die Slater-Bedingung aus Annahme 2.1.1 und die Zusatzannahmen 2.1.6, 2.1.7 erfüllt. Sei  $\bar{u} \in L^2(0, T)$  die optimale Lösung des Problems (2.5). Dann existieren eine Funktion  $\mu \in NBV_T(0, T)$  sowie ein alternativer adjungierter Zustand*

$$q \in \{ \hat{q} \in L^{\infty}(0, T; L^2(\Omega)) \mid \hat{q}' \in L^2(0, T; H_{\Gamma}^2(\Omega)^*) \},$$

so dass die folgenden Bedingungen erfüllt sind:

- *alternative adjungierte Gleichung:*

$$\begin{aligned} \iint_Q q (\varphi_t - \Delta \varphi) dx dt &= \iint_Q (\hat{S}(\bar{u}) - y_d) \varphi dx dt - \int_I \mu \langle \delta', \varphi \rangle_{H_{\Gamma}^2(\Omega)^*, H_{\Gamma}^2(\Omega)} dt \\ &\quad - \iint_Q \mu \delta \Delta \varphi dx dt \quad \forall \varphi \in W_{H_{\Gamma}^2(\Omega), L^2(\Omega)}(0, T) \text{ mit } \varphi(0) = 0, \end{aligned}$$

- *Sattelpunkteigenschaft der Lagrangefunktion:*

$$\lambda \bar{u}(t) + \int_{\Omega} (q - \mu \delta) e dx(t) = 0 \quad \text{f.ü. in } [0, T], \quad (2.20)$$

- *komplementäre Schlupfbedingung:  $\mu$  ist monoton steigend auf  $[0, T]$  und konstant auf jedem Intervall  $(t_1, t_2) \subset [0, T]$  mit  $t_1 < t_2$  und  $\hat{G}(\bar{u})(t) < 0$  für alle  $t \in (t_1, t_2)$ .*

*Beweis.* Ausgangspunkt des Beweises ist das Optimalitätssystem in Satz 2.1.3. In der Folge ist nachzuweisen, dass die klassische Adjungierte  $p$  zerlegt werden kann in  $p = q - \mu \delta$ , wobei der alternative adjungierte Zustand  $q$  obige adjungierte Gleichung erfüllt.

Sei hierzu  $\varphi \in W_{H^2_\Gamma(\Omega), L^2(\Omega)}(0, T)$  mit  $\varphi(0) = 0$  beliebig. Insbesondere gilt  $\varphi \in W(0, T)$ , so dass mit der klassischen adjungierten Gleichung (2.8)

$$\begin{aligned} & \int_I \langle \varphi_t, p \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt + \int_I \langle Ap, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt \\ &= \iint_Q (\hat{S}(\bar{u}) - y_d) \varphi dxdt + \iint_Q \delta \varphi dx d\mu \quad \forall \varphi \in W_{H^2_\Gamma(\Omega), L^2(\Omega)}(0, T) \text{ mit } \varphi(0) = 0. \end{aligned}$$

Mit partieller Integration kann man unter Verwendung der Randbedingungen  $\partial_\nu \varphi + \varphi = 0$  umformen:

$$\langle Ap, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} = \int_\Omega \nabla p \nabla \varphi dx + \int_\Gamma p \varphi ds(x) = - \int_\Omega p \Delta \varphi dx.$$

Es gilt also  $\forall \varphi \in W_{H^2_\Gamma(\Omega), L^2(\Omega)}(0, T)$  mit  $\varphi(0) = 0$ :

$$\iint_Q p(\varphi_t - \Delta \varphi) dxdt = \iint_Q (\hat{S}(\bar{u}) - y_d) \varphi dxdt + \iint_Q \delta \varphi dx d\mu. \quad (2.21)$$

Nach dem Lemma A.0.1 ist  $[0, T] \rightarrow \mathbb{R}, t \mapsto \int_\Omega \delta \varphi dx$  absolut stetig, so dass mit der Formel der partiellen Integration für Stieltjes-Integrale [24, S. 66, 48, 11]

$$\iint_Q \delta \varphi dx d\mu = - \int_I \left( \int_\Omega \delta \varphi dx \right)_t \mu dt = - \iint_Q \mu \delta \varphi_t dxdt - \int_I \mu \langle \delta', \varphi \rangle_{H^2_\Gamma(\Omega)^*, H^2_\Gamma(\Omega)} dt$$

folgt. Hierbei beachte man, dass die Anfangs- und Endterme wegen  $\varphi(0) = 0$  und  $\mu(T) = 0$  wegfallen.

Definiert man  $q := p + \mu \delta$ , so gehört  $q$  wegen  $p \in L^\infty(0, T; L^2(\Omega))$ ,  $\mu \in BV(0, T)$  und  $\delta \in C([0, T]; L^2(\Omega))$  insbesondere dem Raum  $L^\infty(0, T; L^2(\Omega))$  an. Eingesetzt in die Gleichung (2.21) ergibt sich mit der hergeleiteten Umformung für  $\iint_Q \delta \varphi dx d\mu$ :

$$\begin{aligned} \iint_Q q(\varphi_t - \Delta \varphi) dxdt &= \iint_Q (\hat{S}(\bar{u}) - y_d) \varphi dxdt + \iint_Q \mu \delta (\varphi_t - \Delta \varphi) dxdt + \iint_Q \delta \varphi dx d\mu \\ &= \iint_Q (\hat{S}(\bar{u}) - y_d) \varphi dxdt - \int_I \mu \langle \delta', \varphi \rangle_{H^2_\Gamma(\Omega)^*, H^2_\Gamma(\Omega)} dt \\ &\quad - \iint_Q \mu \delta \Delta \varphi dxdt \quad \forall \varphi \in W_{H^2_\Gamma(\Omega), L^2(\Omega)}(0, T) \text{ mit } \varphi(0) = 0. \end{aligned}$$

Um die Regularität der distributionellen Ableitung  $q'$  herzuleiten, geht man wie in Tröltzsch [54, S. 119 ff.] vor. Man testet die eben hergeleitete Gleichung für den alternativen adjungierten Zustand mit Funktionen  $h \cdot v$ , wobei  $h \in C_0^\infty(0, T)$  und  $v \in H^2_\Gamma(\Omega)$  beliebig seien, und erhält:

$$\int_I (qh_t, v)_{L^2(\Omega)} dt = \int_I (q, \Delta v)_{L^2(\Omega)} h dt + \int_I (\hat{S}(\bar{u}) - y_d, v)_{L^2(\Omega)} h dt \quad (2.22a)$$

$$- \int_I \mu \langle \delta', v \rangle_{H^2_\Gamma(\Omega)^*, H^2_\Gamma(\Omega)} h dt - \int_I \mu (\delta, \Delta v)_{L^2(\Omega)} h dt. \quad (2.22b)$$

Für die Integranden auf der rechten Seite gilt für fast alle  $t \in [0, T]$

$$\begin{aligned} \left| (q(t), \Delta v)_{L^2(\Omega)} \right| &\leq \|q(t)\|_{L^2(\Omega)} \|v\|_{H_{\Gamma}^2(\Omega)}, \\ \left| \left( \hat{S}(\bar{u})(t) - y_d(t), v \right)_{L^2(\Omega)} \right| &\leq \left\| \hat{S}(\bar{u})(t) - y_d(t) \right\|_{L^2(\Omega)} \|v\|_{H_{\Gamma}^2(\Omega)}, \\ \left| \mu(t) \langle \delta'(t), v \rangle_{H_{\Gamma}^2(\Omega)^*, H_{\Gamma}^2(\Omega)} \right| &\leq |\mu(t)| \left\| \delta'(t) \right\|_{H_{\Gamma}^2(\Omega)^*} \|v\|_{H_{\Gamma}^2(\Omega)}, \\ \left| \mu(t) (\delta(t), \Delta v)_{L^2(\Omega)} \right| &\leq |\mu(t)| \|\delta(t)\|_{L^2(\Omega)} \|v\|_{H_{\Gamma}^2(\Omega)}. \end{aligned}$$

Für festes  $t$  sind also

$$\begin{aligned} F_1(t) &: v \mapsto (q(t), \Delta v)_{L^2(\Omega)}, \\ F_2(t) &: v \mapsto \left( \hat{S}(\bar{u})(t) - y_d(t), v \right)_{L^2(\Omega)}, \\ F_3(t) &: v \mapsto \mu(t) \langle \delta'(t), v \rangle_{H_{\Gamma}^2(\Omega)^*, H_{\Gamma}^2(\Omega)}, \\ F_4(t) &: v \mapsto \mu(t) (\delta(t), \Delta v)_{L^2(\Omega)} \end{aligned}$$

linear stetige Funktionale auf  $H_{\Gamma}^2(\Omega)$  mit

$$\begin{aligned} \|F_1(t)\|_{H_{\Gamma}^2(\Omega)^*} &\leq \|q(t)\|_{L^2(\Omega)}, \\ \|F_2(t)\|_{H_{\Gamma}^2(\Omega)^*} &\leq \left\| \hat{S}(\bar{u})(t) - y_d(t) \right\|_{L^2(\Omega)}, \\ \|F_3(t)\|_{H_{\Gamma}^2(\Omega)^*} &\leq |\mu(t)| \left\| \delta'(t) \right\|_{H_{\Gamma}^2(\Omega)^*}, \\ \|F_4(t)\|_{H_{\Gamma}^2(\Omega)^*} &\leq |\mu(t)| \|\delta(t)\|_{L^2(\Omega)}. \end{aligned}$$

Da mit den gegebenen Regularitäten  $\|q\|_{L^2(\Omega)} \in L^\infty(0, T)$ ,  $\left\| \hat{S}(\bar{u}) - y_d \right\|_{L^2(\Omega)} \in L^2(0, T)$ ,  $|\mu| \|\delta'\|_{H_{\Gamma}^2(\Omega)^*} \in L^2(0, T)$  und  $|\mu| \|\delta\|_{L^2(\Omega)} \in L^\infty(0, T)$  gilt, hat man

$$F_i \in L^2(0, T; H_{\Gamma}^2(\Omega)^*) \quad \text{für } i = 1, \dots, 4.$$

Verwendet man die Definition der Funktionale in Gleichung (2.22), so ergibt sich für alle  $v \in H_{\Gamma}^2(\Omega)$

$$\left( \int_I q h_t \, dt, v \right)_{L^2(\Omega)} = \left\langle \int_I (F_1 + F_2 - F_3 - F_4) h \, dt, v \right\rangle_{H_{\Gamma}^2(\Omega)^*, H_{\Gamma}^2(\Omega)}$$

und deshalb

$$\int_I q h_t \, dt = \int_I (F_1 + F_2 - F_3 - F_4) h \, dt$$

in  $H_{\Gamma}^2(\Omega)^*$ . Die distributionelle Ableitung  $q' = -F_1 - F_2 + F_3 + F_4$  gehört also dem Raum  $L^2(0, T; H_{\Gamma}^2(\Omega)^*)$  an.



Wegen  $p = q - \mu\delta$  lässt sich schließlich die Sattelpunkteigenschaft der Lagrangefunktion ausdrücken als

$$\lambda \bar{u}(t) + \int_{\Omega} (q - \mu\delta)e \, dx(t) = 0 \quad \text{f.ü. in } [0, T].$$

□

Fordert man die erhöhte Regularität

**Annahme 2.1.8.**  $e \in W_{H^2_F(\Omega), L^2(\Omega)}(0, T)$ ,

so ist als unmittelbare Konsequenz auch die Funktion  $[0, T] \rightarrow \mathbb{R}$ ,  $t \mapsto \int_{\Omega} q(t)e(t) \, dx$  absolut stetig. Man beachte dabei jedoch, dass man sich an dieser Stelle Randbedingungen für die Datenfunktion  $e$  einhandelt.

Das Vorgehen zum Herleiten der weiteren Resultate ist in der Folge weitestgehend analog zu den Ausarbeitungen in Kapitel 2.1.2. Es müssen nur die angepassten Funktionsräume und Hilfssätze verwendet werden. Es wird deshalb, soweit keine Unterschiede in der Argumentation auftreten, auf eine Darstellung der Beweise verzichtet.

Die Sattelpunkteigenschaft der Lagrangefunktion (2.20) liefert:

**Korollar 2.1.5.** *Unter den Annahmen 2.1.1, 2.1.6, 2.1.7 und 2.1.8 gehört die optimale Steuerung  $\bar{u}$  dem Raum  $BV(0, T)$  an. Ist  $\int_{\Omega} \delta e \, dx = 0$  in  $[0, T]$ , so hat man  $\bar{u} \in AC^{1,1}(0, T)$ .*

Fordert man eine erhöhte Regularität des Anfangswertes, so lässt sich erhöhte Regularität für den optimalen Zustand  $\bar{y}$  nachweisen.

**Lemma 2.1.5.** *Sei die Annahme 2.1.6 erfüllt,  $Ay_0 \in H^1(\Omega)$ ,  $u \in BV(0, T)$  sowie  $e \in W_{H^2_F(\Omega), L^2(\Omega)}(0, T)$ . Dann besitzt die Lösung  $y$  der Zustandsgleichung (2.1b) - (2.1d) neben  $y \in W_{H^2_F(\Omega), L^2(\Omega)}(0, T)$  die zusätzliche Regularität*

$$y_t \in L^2(0, T; H^2_F(\Omega)) \cap BV(0, T; L^2(\Omega)),$$

sowie  $Ay \in W_{H^2_F(\Omega), L^2(\Omega)}(0, T)$ .

*Beweis.* Die Wahl der Daten garantiert, dass das Lemma 2.1.3 aus Kapitel 2.1.2 anwendbar ist. Es ist deshalb  $Ay$  als Lösung des Evolutionsproblems

$$\begin{aligned} (Ay)'(t) + A(Ay)(t) &= u(t) \cdot Ae(t) \quad \text{f.ü. in } [0, T], \\ Ay(0) &= Ay_0, \end{aligned}$$

gegeben. Für  $Ay_0 \in H^1(\Omega)$  und mit der erhöhten Randregularität aus Annahme 2.1.6 folgt gemäß Lions [42, Ch. 1]  $Ay \in W_{H^2_F(\Omega), L^2(\Omega)}(0, T)$ . Da Funktionen aus dem Raum  $W_{H^2_F(\Omega), L^2(\Omega)}(0, T)$  insbesondere dem Raum  $AC^{1,2}([0, T]; L^2(\Omega))$  angehören, folgt die erhöhte Regularität für  $y_t$  aus der Zustandsgleichung (2.3) mit der Tatsache, dass  $u \cdot e \in L^2(0, T; H^2_F(\Omega)) \cap BV(0, T; L^2(\Omega))$  für  $u \in BV(0, T)$  und  $e \in W_{H^2_F(\Omega), L^2(\Omega)}(0, T)$ . □

Der optimale Zustand  $\bar{y}$  besitzt wegen  $\bar{u} \in BV(0, T)$  diese erhöhte Regularität, so dass sich u.a. wegen  $\int_{\Omega} A\bar{y} \cdot \delta \, dx \in AC^{1,1}(0, T)$  beweisen lässt:

**Satz 2.1.10.** *Seien neben den Grundvoraussetzungen insbesondere die Annahmen 2.1.1, 2.1.6 sowie die Bedingungen  $Ay_0 \in H^1(\Omega)$ ,  $e \in W_{H^2(\Omega), L^2(\Omega)}(0, T)$ ,  $\delta \in C([0, T]; L^2(\Omega))$ ,  $\delta' \in C([0, T]; H^2_1(\Omega)^*)$  sowie  $y_{\max} \in C^1([0, T])$  erfüllt. Dann ist die optimale Lösung  $\bar{u}$  des Problems (2.5) stetig auf  $[0, T]$  und absolut stetig auf der inaktiven Menge  $\mathcal{I}$ . Des Weiteren gilt für die Funktion in der Zustandsbeschränkung:  $\bar{g} \in C^1([0, T])$ . Wird die Bedingung  $\int_{\Omega} \delta e \, dx(t) \neq 0$  für alle  $t$  in der aktiven Menge  $\mathcal{A}$  eingehalten, so besitzt  $\mu$  keinen diskreten Anteil:  $\mu \in NBV_T(0, T) \cap C([0, T])$ .*

Unter geeigneten Zusatzvoraussetzungen lässt sich wieder herleiten, dass der Cantor-Anteil von  $\mu$  verschwindet.

**Satz 2.1.11.** *Sei  $\bar{g}$  aktiv auf einem Intervall, d.h. es existieren  $t^{en}, t^{ex} \in \mathbb{R}$  mit  $0 < t^{en} < t^{ex} \leq T$  und  $\bar{g} = 0$  auf  $[t^{en}, t^{ex}]$ , und seien alle bisherigen Annahmen des Kapitels erfüllt. Sei zusätzlich  $\delta' \in AC^{1,2}([0, T]; H^2_1(\Omega)^*)$ ,  $y_{\max} \in AC^{2,1}(0, T)$  und sei  $\int_{\Omega} \delta e \, dx(t) \neq 0$  für alle  $t \in [t^{en}, t^{ex}]$ . Dann sind die optimale Steuerung  $\bar{u}$  und die Funktion  $\mu$  absolut stetig auf  $[t^{en}, t^{ex}]$ .*

## 2.2. Zeitabhängige optimale Steuerung mit Zustandsbeschränkung in Operatorform

In diesem Kapitel sollen die bisher entwickelten Techniken so kombiniert werden, dass Zustandsbeschränkungen in allgemeinerer Form untersucht werden können. Insbesondere sollen auch solche Zustandsbeschränkungen betrachtet werden, die keine integrale Form  $\int_{\Omega} y\delta \, dx$  mit einer Funktion  $\delta$  in einem Lebesgue-Raum aufweisen. Hierfür ist es zielführend, möglichst reguläre Zustände zu betrachten. Deshalb seien die folgenden Annahmen getroffen:

**Annahme 2.2.1.** Es seien die Grundannahmen des Modellproblems (2.1) erfüllt. Sei  $2 \leq p < \infty$ . Die Verteilungsfunktion  $e$  gehöre dem Raum  $e \in L^\infty(0, T; L^{\tilde{p}}(\Omega))$  mit  $2 \leq p \leq \tilde{p}$  an und es gelte für den Anfangswert  $y_0 = 0$ .

**Bemerkung 2.2.1.** Im Unterschied zu den vorherigen Kapiteln bezeichne  $p$  in diesem Kapitel einen Integrationsindex. Für den (klassischen) adjungierten Zustand wird die Darstellung  $\wp$  verwendet. Für den Anfangswert ist in diesem Kapitel vereinfachend die Annahme  $y_0 = 0$  getroffen. Wenn man mit den entsprechenden Interpolationsräumen arbeitet [3], ließen sich die nachfolgenden Ergebnisse aber auch auf den Fall inhomogener Anfangsdaten übertragen.

Es soll in der Folge mit dem Konzept der maximalen parabolischen Regularität gearbeitet werden. Man betrachte deshalb die Einschränkung des variationellen Operators  $A: H^1(\Omega) \rightarrow H^1(\Omega)^*$  aus Kapitel 2.1 auf den Raum  $L^p(\Omega)$ . Dieser Operator  $\mathcal{A}_p$  sei definiert durch

$$\begin{aligned} \mathcal{A}_p: L^p(\Omega) \supset D(\mathcal{A}_p) &\rightarrow L^p(\Omega) \\ v &\mapsto \mathcal{A}_p v := Av \end{aligned}$$

für  $v \in D(\mathcal{A}_p) := \{v \in H^1(\Omega) \mid Av \in L^p(\Omega)\}$ . Nach Griepentrog, Kaiser und Rehberg [28] ist der Operator  $\mathcal{A}_p$  abgeschlossen und dicht definiert. Versieht man  $D(\mathcal{A}_p)$  mit der Graphennorm

$$\|v\|_{D(\mathcal{A}_p)} := \|v\|_{L^p(\Omega)} + \|\mathcal{A}_p v\|_{L^p(\Omega)},$$

so besitzt  $\mathcal{A}_p$  außerdem für jedes  $p \in [2, \infty)$  maximale parabolische  $L^2(0, T; L^p(\Omega))$ -Regularität. D.h. zu jedem  $f \in L^2(0, T; L^p(\Omega))$  existiert genau ein  $y_f \in H^1(0, T; L^p(\Omega)) \cap L^2(0, T; D(\mathcal{A}_p))$ , so dass

$$\begin{aligned} \frac{\partial y_f}{\partial t}(t) + \mathcal{A}_p y_f(t) &= f(t) \quad \text{f.ü. in } [0, T], \\ y_f(0) &= 0. \end{aligned}$$

Des Weiteren ist der Operator

$$\frac{\partial}{\partial t} + \mathcal{A}_p: \{y_f \in L^2(0, T; D(\mathcal{A}_p)) \cap H^1(0, T; L^p(\Omega)) \mid y_f(0) = 0\} \rightarrow L^2(0, T; L^p(\Omega))$$

ein topologischer Isomorphismus [28, 20].

Die rechte Seite  $u \cdot e$  der Zustandsgleichung (2.1b)–(2.1d) im Modellproblem (2.1) gehört wegen der Annahme 2.2.1 dem Raum  $L^2(0, T; L^p(\Omega))$  an und erfüllt die Bedingung  $\|ue\|_{L^2(0, T; L^p(\Omega))} \leq \|u\|_{L^2(0, T)} \cdot \|e\|_{L^\infty(0, T; L^p(\Omega))}$ . Mit den bisherigen Betrachtungen gibt es deshalb zu jeder Steuerung  $u \in L^2(0, T)$  genau einen Zustand

$$y \in Y_p := AC^{1,2}([0, T]; L^p(\Omega)) \cap L^2(0, T; D(\mathcal{A}_p))$$

als Lösung der Zustandsgleichung in der Form

$$\frac{\partial y}{\partial t}(t) + \mathcal{A}_p y(t) = u(t) \cdot e(t) \quad \text{f.ü. in } [0, T], \tag{2.23}$$

$$y(0) = 0. \tag{2.24}$$

Außerdem lässt sich wie in Kapitel 2.1 ein linear stetiger Steuerungs-Zustands-Operator

$$S: L^2(0, T) \rightarrow Y_{p,0} := \{y \in Y_p \mid y(0) = 0\}, u \mapsto y$$

eingeführen.

Statt der bisher untersuchten Zustandsbeschränkung in Integralform  $\int_\Omega y \delta \, dx - y_{\max} \leq 0$  sollen nun allgemeinere Zustandsbeschränkungen untersucht werden. Hierzu definiert man:

**Definition 2.2.1.** Der Beobachtungsoperator  $B: Y_{p,0} \rightarrow C([0, T])$  sei ein linear stetiger Operator.

Mit diesem Operator sei die Menge der zulässigen Zustände durch die Bedingung

$$By - y_{\max} \leq 0$$

beschrieben. Wegen der Stetigkeit des Operators  $B$  auf  $Y_{p,0}$  kann man den linear stetigen Operator  $G: L^2(0, T) \rightarrow C([0, T])$ ,  $u \mapsto BSu$  einführen. Bei unverändertem Zielunktional lässt sich das resultierende Optimalsteuerungsproblem mit dem affin linearen Operator  $\hat{G}: L^2(0, T) \rightarrow C([0, T])$ ,  $u \mapsto Gu - y_{\max}$  wieder als reduziertes Problem im Banachraum schreiben (vgl. Problem (2.5)):

$$\min_{u \in L^2(0, T)} \mathcal{J}(u) \text{ unter der Nebenbedingung } \hat{G}(u) \in K = \{g \in C([0, T]) \mid g \leq 0\}. \quad (2.25)$$

Entsprechend gelten auch für diesen Fall die Existenz- und Eindeutigkeitsaussage von Satz 2.1.1 sowie unter Forderung einer Slater-Bedingung (Annahme 2.1.1) die Optimalitätsbedingungen in Satz 2.1.2. Mit der Charakterisierung von  $C([0, T])^*$  wie in Kapitel 2.1.1 ergibt sich unmittelbar das Korollar:

**Korollar 2.2.1.** *Sei  $\bar{u}$  die optimale Lösung des Problems (2.25) und sei die Slater-Bedingung aus Annahme 2.1.1 erfüllt. Dann existiert ein  $\mu \in NBV_T(0, T)$ , so dass die folgenden Bedingungen erfüllt sind:*

- *Sattelpunkteigenschaft der Lagrangefunktion:*

$$0 = \iint_Q (S\bar{u} - y_d)Sh \, dxdt + \lambda \int_I \bar{u}h \, dt + \int_I BSh \, d\mu \quad \forall h \in L^2(0, T), \quad (2.26)$$

- *komplementäre Schlupfbedingung:*

$$\int_I g \, d\mu \leq 0 \quad \forall g \in K \quad \text{und} \quad 0 = \int_I \hat{G}(\bar{u}) \, d\mu.$$

Unter Einführung eines adjungierten Zustands  $\wp$  erhält man das folgende Optimalitätssystem:

**Satz 2.2.1.** *Seien die Annahme 2.2.1 und die Slater-Bedingung aus Annahme 2.1.1 erfüllt. Sei  $B$  wie in Definition 2.2.1 definiert und sei  $\bar{u} \in L^2(0, T)$  die eindeutige optimale Lösung des Problems (2.25). Dann existieren eine Funktion  $\mu \in NBV_T(0, T)$  sowie ein adjungierter Zustand  $\wp \in L^2(0, T; L^{p'}(\Omega))$  mit  $\frac{1}{p} + \frac{1}{p'} = 1$ , so dass die folgenden Bedingungen erfüllt sind:*

- *adjungierte Gleichung:*

$$\iint_Q \wp(\varphi_t + A_p\varphi) \, dxdt \quad (2.27)$$

$$= \iint_Q (S\bar{u} - y_d)\varphi \, dxdt + \int_I (B\varphi) \, d\mu \quad \forall \varphi \in Y_{p,0}, \quad (2.28)$$

- *Sattelpunkteigenschaft der Lagrangefunktion:*

$$\lambda \bar{u}(t) + \int_{\Omega} \wp e \, dx(t) = 0 \quad \text{f.ü. in } [0, T],$$

- *komplementäre Schlupfbedingung:  $\mu$  ist monoton steigend auf  $[0, T]$  und konstant auf jedem Intervall  $(t_1, t_2) \subset [0, T]$  mit  $t_1 < t_2$  und  $\hat{G}(\bar{u})(t) < 0$  für alle  $t \in (t_1, t_2)$ .*

*Beweis.* Ausgangspunkt des Beweises ist Korollar 2.2.1, das insbesondere die Existenz eines  $\mu \in NBV_T(0, T)$  garantiert. In einem ersten Schritt soll die Existenz eines adjungierten Zustands  $\wp \in L^2(0, T; L^{p'}(\Omega))$ , der die adjungierte Gleichung erfüllt, gezeigt werden. Sei hierzu  $f \in L^2(0, T; L^p(\Omega))$  beliebig und sei  $\tilde{S}: L^2(0, T; L^p(\Omega)) \rightarrow Y_{p,0}$  der linear stetige Lösungsoperator, der jede rechte Seite  $f$  auf die zugehörige Lösung  $y_f$  von

$$\begin{aligned} \frac{\partial y_f}{\partial t}(t) + \mathcal{A}_p y_f(t) &= f(t) \quad \text{f.ü. in } [0, T], \\ y_f(0) &= 0 \end{aligned}$$

abbildet. Dann sind die Ausdrücke in der rechten Seite der Gleichung

$$\iint_Q \wp f \, dx dt = \iint_Q (S\bar{u} - y_d) \tilde{S} f \, dx dt + \int_I (B\tilde{S}f) \, d\mu \quad \forall f \in L^2(0, T; L^p(\Omega))$$

mit den getroffenen Voraussetzungen linear stetige Funktionale auf  $L^2(0, T; L^p(\Omega))$ , so dass unmittelbar die Existenz genau eines  $\wp \in L^2(0, T; L^{p'}(\Omega))$  mit  $\frac{1}{p} + \frac{1}{p'} = 1$  folgt. Da aber  $\frac{\partial}{\partial t} + \mathcal{A}_p$  einen topologischen Isomorphismus zwischen  $Y_{p,0}$  und  $L^2(0, T; L^p(\Omega))$  darstellt, existiert zu jedem  $\varphi \in Y_{p,0}$  genau ein  $f \in L^2(0, T; L^p(\Omega))$ , so dass  $\wp$  die Gleichung

$$\iint_Q \wp (\varphi_t + \mathcal{A}_p \varphi) \, dx dt = \iint_Q (S\bar{u} - y_d) \varphi \, dx dt + \int_I (B\varphi) \, d\mu \quad \forall \varphi \in Y_{p,0}$$

erfüllt.

Mit dem oben eingeführten Steuerungs-Zustands-Operator  $S: L^2(0, T) \rightarrow Y_{p,0}$  ist die Funktion  $Sh$  für beliebiges  $h \in L^2(0, T)$  eine zulässige Testfunktion in der adjungierten Gleichung. Man erhält unter Verwendung der Zustandsgleichung

$$\begin{aligned} \iint_Q \wp (eh) \, dx dt &= \iint_Q \wp ((Sh)_t + \mathcal{A}_p(Sh)) \, dx dt \\ &= \iint_Q (S\bar{u} - y_d)(Sh) \, dx dt + \int_I (B(Sh)) \, d\mu \quad \forall h \in L^2(0, T) \end{aligned}$$

und somit aus Gleichung (2.26)

$$\lambda \bar{u}(t) + \int_{\Omega} \wp e \, dx(t) = 0 \quad \text{f.ü. in } [0, T]. \quad (2.29)$$

Für die Formulierung der komplementären Schlupfbedingung in dieser Form sei wieder auf Gerdts [24, S. 131 ff.] verwiesen.  $\square$

Um wie in den vorhergehenden Kapiteln die Stetigkeit der optimalen Steuerung nachzuweisen, müssen einige Forderungen an den Operator  $B$  und die Funktion  $e$  gestellt werden. Diese sind einerseits einschränkend, andererseits aber noch hinreichend allgemein, als dass interessante Fälle abgedeckt werden können. Ein konkretes Beispiel findet sich am Ende des Kapitels.

**Annahme 2.2.2.** Die Verteilungsfunktion  $e \in L^\infty(0, T; L^{\bar{p}}(\Omega))$  und der Beobachtungsoperator  $B: Y_{p,0} \rightarrow C([0, T])$  erfüllen die folgenden Zusatzbedingungen:

- $e \in Y_p$ ,
- $\forall h \in C_0^\infty(0, T): (B(e \cdot h))(t) = (Be)(t) \cdot h(t) \quad \forall t \in [0, T]$ ,
- $Be \in AC^{1,1}(0, T)$ .

**Bemerkung 2.2.2.** Die erste Bedingung impliziert insbesondere, dass die Verteilungsfunktion die Randbedingungen  $\partial_\nu e + e = 0$  erfüllen muss. Dies stellt eine große Einschränkung dar. Wirkt  $e$  allerdings nur im Inneren von  $\Omega$ , so ist diese Bedingung bei passender Modellierung eingehalten. Die erste Bedingung wird in der Folge benötigt, um die in Kapitel 2.1.4 eingeführte Technik des geschickten Testens anwenden zu können. Die zweite und dritte Bedingung werden zur Herleitung der schwachen Differenzierbarkeit der zeitabhängigen Funktion  $\int_\Omega \wp e \, dx$  benötigt.

**Lemma 2.2.1.** *Seien neben den Voraussetzungen des Satzes 2.2.1 die Annahme 2.2.2 erfüllt. Dann existiert eine Funktion  $q_e \in AC^{1,1}(0, T)$ , so dass für die optimale Steuerung  $\bar{u}$  mit der Funktion  $\mu \in NBV_T(0, T)$  gilt:*

$$\lambda \bar{u}(t) + q_e(t) - \mu(t)(Be)(t) = 0 \quad \text{f.ü. in } [0, T]. \quad (2.30)$$

*Beweis.* Sei  $h \in C_0^\infty(0, T)$  beliebig. Dann liegt die Funktion  $h \cdot e$  für  $e \in Y_p$  in  $Y_{p,0}$  und ist somit eine zulässige Testfunktion in der adjungierten Gleichung (2.27). Eingesetzt erhält man unter Verwendung der Annahme 2.2.2:

$$\begin{aligned} & \int_I \left( \int_\Omega \wp e \, dx \right) h_t \, dt + \int_I \left( \int_\Omega \wp (e_t + \mathcal{A}_p e) \, dx \right) h \, dt \\ &= \int_I \left( \int_\Omega (S\bar{u} - y_d) e \, dx \right) h \, dt + \int_I \underbrace{(B(eh))}_{=(Be)h} \, d\mu \quad \forall h \in C_0^\infty(0, T). \end{aligned}$$

Da die zeitabhängige Funktion  $Be$  nach Voraussetzung absolut stetig ist, lässt sich der letzte Summand mittels partieller Integration für Stieltjes-Integrale [24, S. 66, 48, 11] umformen:

$$\int_0^T (Be)h \, d\mu = - \int_0^T (Be)_t \mu h \, dt - \int_0^T (Be)\mu h_t \, dt.$$

Eingesetzt und sortiert ergibt sich:

$$\int_0^T \left( \int_\Omega \wp e \, dx + (Be)\mu \right) h_t \, dt = \int_0^T \left( \int_\Omega -\wp (e_t + \mathcal{A}_p e) + (S\bar{u} - y_d) e \, dx \right) h \, dt \quad (2.31)$$

$$- \int_0^T (Be)_t \mu h \, dt, \quad \forall h \in C_0^\infty(0, T). \quad (2.32)$$

Definiert man  $q_e : [0, T] \rightarrow \mathbb{R}$  als

$$q_e(t) := \int_{\Omega} \wp(t)e(t) \, dx + (Be)(t)\mu(t),$$

so hat man mit den gegebenen Voraussetzungen  $q_e \in L^1(0, T)$ . Wegen (2.31) ist  $q_e$  schwach differenzierbar mit Ableitung

$$(q_e)_t = \underbrace{\int_{\Omega} \wp(e_t + \mathcal{A}_p e) \, dx}_{\in L^1(0, T)} - \underbrace{\int_{\Omega} (S\bar{u} - y_d)e \, dx}_{\in L^2(0, T)} + \underbrace{(Be)_t}_{\in L^1(0, T)} \underbrace{\mu}_{\in L^\infty(0, T)}$$

in  $L^1(0, T)$ . Die Identifikation von  $W^{1,1}(0, T)$  mit  $AC^{1,1}(0, T)$  gemäß Barbu und Precupanu [4] liefert  $q_e \in AC^{1,1}(0, T)$ . Nutzt man die Definition von  $q_e$ , so folgt mit Beziehung (2.29) fast überall in  $[0, T]$ :

$$0 = \lambda\bar{u}(t) + \int_{\Omega} \wp(t)e(t) \, dx(t) = \lambda\bar{u}(t) + q_e(t) - \mu(t)(Be)(t).$$

□

Aus der Bedingung (2.30) folgt mit der absoluten Stetigkeit von  $Be$  und  $q_e$  unmittelbar wie gehabt:

**Korollar 2.2.2.** *Seien die Voraussetzungen aus Lemma 2.2.1 erfüllt. Dann gehört die optimale Steuerung  $\bar{u}$  dem Raum  $BV(0, T)$  an. Ist  $Be = 0$  in  $[0, T]$ , so hat man  $\bar{u} \in AC^{1,1}(0, T)$ .*

In der Folge soll die Stetigkeit der optimalen Steuerung in  $[0, T]$  ohne die Bedingung  $Be = 0$  nachgewiesen werden. Hierfür ist eine zusätzliche Annahme an die Daten nötig:

**Annahme 2.2.3.** Die Daten des Problems seien so gewählt, dass für die optimale Steuerung  $\bar{u}$  die Funktion  $\bar{g} = \hat{G}(\bar{u}) \in C([0, T])$  eine derartige Regularität besitzt, dass der Sprung der Zeitableitung von  $\bar{g}$  für alle  $t \in (0, T]$  wohldefiniert ist. Für diesen gelte außerdem:

$$\left[ \frac{d}{dt} \bar{g}(t) \right] = \left[ \frac{d}{dt} B(S\bar{u})(t) \right] = [\bar{u}(t)] \cdot Be(t). \quad (2.33)$$

Ein Beispiel am Ende des Kapitels wird zeigen, dass diese einschränkenden Bedingungen durchaus noch interessante Fälle abdecken.

**Satz 2.2.2.** *Seien die Annahmen 2.2.1, 2.2.2 und 2.2.3 erfüllt. Dann ist die optimale Lösung  $\bar{u}$  des Problems (2.25) und die Funktion  $\frac{d}{dt} \bar{g} = \frac{d}{dt} (\hat{G}(\bar{u}))$  stetig auf  $[0, T]$ . Außerdem ist  $\bar{u}$  absolut stetig auf  $\mathcal{I}$ . Wird die Bedingung  $Be(t) \neq 0$  für alle  $t$  in der aktiven Menge  $\mathcal{A}$  eingehalten, so besitzt  $\mu \in BV(0, T)$  keinen diskreten Anteil:  $\mu \in NBV_T(0, T) \cap C([0, T])$ .*

*Beweis.* Der Beweis läuft analog zum Beweis des Satzes 2.1.5 in Kapitel 2.1.2 und sei an dieser Stelle nochmals in Kürze ausgeführt. Nach dem Lemma 2.2.1 hat man die absolute Stetigkeit von  $\bar{u}$  auf der inaktiven Menge wegen der komplementären Schlupfbedingung aus Satz 2.2.1 ( $\mu$  konstant auf  $\mathcal{I}$ ) und der absoluten Stetigkeit von  $Be$  gemäß Annahme 2.2.2. Außerdem gilt für  $t \in \mathcal{A}$  wegen Bedingung (2.30):

$$\lambda [\bar{u}(t)] = [\mu(t)](Be)(t). \quad (2.34)$$

Da die Funktion  $\bar{g}$  auf der aktiven Menge gerade maximal ist, muss deren Zeitableitung unter Verwendung der Annahme 2.2.3 die folgende Bedingung einhalten:

$$0 \geq \left[ \frac{d}{dt} \bar{g}(t) \right] = [\bar{u}(t)](Be)(t).$$

Die Monotonie von  $\mu$  garantiert  $[\mu(t)] \geq 0$  für alle  $t \in \mathcal{A}$ , so dass mit der Bedingung (2.34) für alle  $t \in \mathcal{A}$  folgt:

$$0 \geq \left[ \frac{d}{dt} \bar{g}(t) \right] \cdot [\mu(t)] = [\bar{u}(t)][\mu(t)](Be)(t) = \lambda [\bar{u}(t)]^2 \geq 0.$$

Hieraus folgt die Stetigkeit von  $\frac{d}{dt} \bar{g}$  unmittelbar mit der Annahme 2.2.3. Die Funktion  $\mu$  kann gemäß Gleichung (2.34) keinen diskreten Anteil besitzen, falls  $Be(t) \neq 0$  für alle  $t$  in  $\mathcal{A}$  gilt.  $\square$

In der Folge soll exemplarisch ein Operator angeführt und entsprechende Bedingungen an die Daten dargestellt werden, so dass die bisherigen Annahmen eingehalten sind.

**Ein Beispiel** In den vorangehenden Kapiteln wurde, um es im Kontext einer physikalischen Interpretation der Zustandsbeschränkung zu formulieren, der Zustand stets über einen mit der Funktion  $\delta$  gemittelten Bereich gemessen. Es stellt sich die Frage, ob die Aussagen zum Verhalten der optimalen Steuerung auch für ein Problem mit einer punktwisen Auswertung des Zustands möglich sind. Der Einfachheit halber soll dies für den Fall der Beobachtung des Zustands an einem (zeitlich) konstanten Ort  $x = b \in \Omega$  untersucht werden. Eine Erweiterung auf den Fall der Beobachtung entlang regulärer Kurven sollte aber möglich sein. Konkret betrachte man also den Operator

$$B: Y_{p,0} \rightarrow C([0, T]), y \mapsto y(b, \cdot),$$

von dem es in der Folge nachzuweisen gilt, dass er wohldefiniert ist und die geforderten Zusatzbedingungen einhält, wenn die weiteren Daten entsprechend gewählt sind.

Nach de Los Reyes u. a. [20, Cor. 1, Thm. 3, Cor. 2] ist für Lipschitzgebiete  $\Omega$  und für  $2 \leq p < \infty$ ,  $p > \frac{d}{2}$ ,  $p$  hinreichend groß der Raum  $Y_{p,0}$  stetig eingebettet in den Raum  $C^\alpha([0, T] \times \bar{\Omega})$  mit einem  $\alpha > 0$ . Es sei deshalb zunächst  $e \in L^\infty(0, T; L^\infty(\Omega))$  gewählt, so dass die punktwise Auswertung des Zustands für jede beliebige Steuerung  $u \in L^2(0, T)$  wohldefiniert ist und für den zugehörigen Zustand gilt:  $y(b, \cdot) \in C([0, T])$ . Wegen der stetigen Einbettung von  $Y_{p,0}$  in  $C^\alpha([0, T] \times \bar{\Omega})$  ist der gewählte Operator  $B$  offensichtlich stetig, so dass der Existenzsatz 2.1.1 angewandt werden kann. Unter der



Annahme einer Slater-Bedingung ergibt sich außerdem das Optimalitätssystem in Satz 2.1.2 und Korollar 2.2.1.

Um die Annahme 2.2.2 zu erfüllen, muss man von  $e$  weitere Bedingungen verlangen. Es sei deshalb statt  $e \in L^\infty(0, T; L^\infty(\Omega))$  die schärfere Bedingung

$$e \in AC^{1,2}([0, T]; C(\bar{\Omega})) \cap L^2(0, T; W_0^{2,\infty}(\Omega))$$

gefordert, so dass  $e$  dem Raum  $Y_p$  für jedes  $p$  mit  $2 \leq p < \infty$  angehört. Insbesondere ist dabei wegen  $\partial_\nu e = e = 0$  auf  $\Sigma$  die Randbedingung  $\partial_\nu e + e = 0$  eingehalten. Um die zusätzliche Bedingung  $Be \in AC^{1,1}(0, T)$  für die konkrete Wahl von  $B$  zu erfüllen, wird  $e \in AC^{1,2}([0, T]; C(\bar{\Omega}))$  statt beispielsweise  $e \in AC^{1,2}([0, T]; L^\infty(\Omega))$  verlangt. Eine kurze Rechnung für den Operator  $B$  wie oben definiert zeigt außerdem wie gewünscht:

$$\forall h \in C_0^\infty(0, T) : B(e \cdot h)(t) = e(b, t) \cdot h(t) = (Be)(t) \cdot h(t), \quad \forall t \in [0, T].$$

Es sei an dieser Stelle darauf hingewiesen, dass es mit den bisherigen Forderungen möglich ist, die Aussagen von Korollar 2.2.2 zu treffen. Konkret folgt also die absolute Stetigkeit der optimalen Steuerung, wenn  $e(b, t) = 0$  für alle  $t \in [0, T]$  eingehalten ist.

Um die Annahme 2.2.3 nachzuweisen, muss man überprüfen, dass für den zur optimalen Steuerung  $\bar{u}$  gehörigen optimalen Zustand  $\bar{y}$  für  $t \in (0, T]$  gilt:

$$\left[ \frac{d}{dt} (\bar{y}(b, \cdot) - y_{\max})(t) \right] = \left[ \left( \frac{\partial}{\partial t} \bar{y}(b, \cdot) \right) (t) \right] = [\bar{u}(t)] \cdot e(b, t). \quad (2.35)$$

Um diese Beziehungen herzuleiten, nutzt man wie in den vorherigen Kapiteln die Tatsache, dass der optimale Zustand höhere Regularität besitzt, da aus 2.2.2 bereits  $\bar{u} \in BV(0, T)$  bekannt ist. Ein entsprechendes Ergebnis liefert der Regularitätssatz A.0.2 im Anhang, der analog zum Kapitel 2.1 auch für den Fall des in diesem Kapitel betrachteten Zugangs mit dem Operator  $\mathcal{A}_p$  formuliert werden kann. Die Funktionen  $\mathcal{A}_p \bar{y}$  und  $e$  gehören dann jedem  $Y_p$  mit  $2 \leq p < \infty$  an, so dass man mit den zu Beginn des Beispiels erwähnten Einbettungssätzen aus de Los Reyes u. a. [20] insbesondere  $e, \mathcal{A}_p \bar{y} \in C([0, T] \times \bar{\Omega})$  hat. Aus der Zustandsgleichung (2.23) folgen mit  $\bar{u} \in BV(0, T)$  schließlich die Beziehungen (2.35), wenn die Bedingung  $y_{\max} \in C^1([0, T])$  eingehalten ist.

Alles in allem erhält man auch für den Fall einer punktweisen Zustandsbeschränkung der Form

$$y(b, \cdot) - y_{\max} \leq 0, \quad b \in \Omega,$$

unter Zusatzforderungen an die Daten die Stetigkeit der optimalen Steuerung.

Aus dieser Tatsache ergibt sich eine interessante Folgerung: Betrachtet man nämlich *nicht* den Fall einer punktweisen Zustandsbeschränkung in einem fixen Ortspunkt (oder wie zu Beginn des Beispiels erwähnt entlang einer entsprechend regulären Kurve in  $\Omega$ ) sondern eine punktweise Beschränkung der Form  $y(x, t) \leq y_{\max}(x, t)$  für alle  $(x, t) \in [0, T] \times \bar{\Omega}$ , so lässt sich auch für diesen Fall die Stetigkeit der optimalen Steuerung herleiten, wenn der sich einstellende optimale Zustand tatsächlich nur ausschließlich entlang eines fixen Ortspunktes bzw. entlang einer hinreichend regulären Kurve aktiv

wird. Dies scheint auf den ersten Blick einen Sonderfall darzustellen. Hat man allerdings im Hinterkopf, dass es zu jedem Zeitpunkt  $t \in [0, T]$  für die Steuerung nur einen skalaren Freiheitsgrad  $u(t) \in \mathbb{R}$  gibt, so wird dies bei entsprechenden Bedingungen an die Daten (auch bei realen Problemen) eher der Normalfall sein.

# 3. Algorithmen und numerische Versuche

## 3.1. PDASSP Algorithmus

### 3.1.1. Schaltpunktsuche und PDAS Strategie

In diesem Kapitel sollen die Grundideen zweier Algorithmen vorgestellt werden, die zur Lösung von Optimalsteuerungsproblemen der Form (2.5) in Kapitel 2.1 eingesetzt werden können: Eine Schaltpunktsuche und die primal-duale aktive-Mengen-Strategie (PDAS), wobei letztere zunächst formal zu verstehen ist (Näheres hierzu findet sich am Ende dieses Kapitels). Die beiden Algorithmen dienen als Grundlage für den in dieser Arbeit entwickelten PDASSP Algorithmus.

Sowohl der Schaltpunktsuche als auch der primal-dualen aktive-Mengen-Strategie liegt die Idee zugrunde, auf Basis eines Optimalitätssystems iterativ die Schätzungen für die aktive Menge zur Zustandsbeschränkung so anzupassen, dass letztendlich die aktive Menge am Optimum  $\mathcal{A}$  approximiert wird. Hierzu werden während der Iterationen die Ungleichungen in der Zustandsbeschränkung  $\hat{G}(\bar{u}) \leq 0$  und die Vorzeichenbedingung an den Lagrange-Multiplikator  $\mu' \geq 0$  relaxiert. Stattdessen fordert man in einem Iterationsschritt  $k$  der Algorithmen jeweils  $\hat{G}(u_k) = 0$  auf der aktuellen Schätzung der aktiven Menge  $\mathcal{A}_k$  und  $\mu' = 0$  auf der inaktiven Menge  $\mathcal{I}_k$ . Es entsteht ein gekoppeltes Gleichungssystem bestehend aus der Zustandsgleichung, der adjungierten Gleichung und der Sattelpunkteigenschaft der Lagrangefunktion. In den nachfolgenden Prinzipskizzen 3.1 und 3.2 sind mögliche Lösungen eines solchen Gleichungssystems dargestellt. Anhand dieser soll das Vorgehen zur Veränderung der Schätzungen der aktiven Menge für die beiden Algorithmen erklärt werden.

**Schaltpunktsuche** Im Rahmen einer Schaltpunktsuche versucht man, die Kontaktpunkte und die Auf- und Absprungpunkte (Schaltpunkte) so zu verschieben, dass die zusätzliche Sprungbedingung am Optimum eingehalten ist. In dieser Arbeit ist diese Bedingung bei hinreichend regulären Daten gerade durch die in Kapitel 2.1 bewiesene Stetigkeit der optimalen Steuerung gegeben. An einem Aufsprungpunkt  $t_1^{\text{en},k}$  muss also beispielsweise für den Sprung  $[\bar{u}(t_1^{\text{en},k})] = 0$  gelten. Für die Berechnung der Veränderung der Schaltpunkte setzt man ein Newtonverfahren ein. Sind die Schaltpunkte so variiert worden, dass die Stetigkeit der optimalen Steuerung eingehalten ist, prüft man a posteriori, ob für die gegebene Konfiguration auch die relaxierten Ungleichungsbeschränkungen eingehalten sind. Falls dies nicht der Fall ist, muss eine Anpassung der Schätzung der Topologie der aktiven Menge vorgenommen werden.

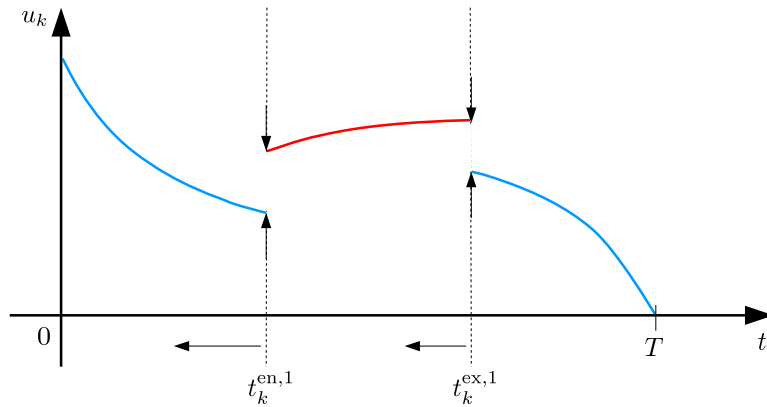


Abbildung 3.1.: Prinzip der Veränderung der aktiven Menge bei einer Schaltungssuche

Als Vorteile dieses Ansatzes sind aus Sicht des Autors die Möglichkeit der sehr exakten Bestimmung der Schaltpunkte und die lokal guten Konvergenzeigenschaften eines Newton-Verfahrens zu nennen. Als Nachteile handelt man sich, wie für ein Newton-Verfahren üblich, das Problem der Globalisierung ein, so dass unter Umständen für die Konvergenz des Verfahrens gute Startschätzungen benötigt werden. Zusätzlich muss eine Schätzung für die Topologie der aktiven Menge von außen neu vorgegeben werden, falls die Vorzeichenbedingungen für die gegebene Konfiguration nicht eingehalten sind. Die Berechnung der Sensitivitäten für das Newton-Verfahren ist außerdem mit zusätzlichem Rechenaufwand verbunden. Aus Implementationssicht muss eine detaillierte Behandlung der aktiven Menge vorgenommen werden. Kontaktpunkte und Auf- und Absprungpunkte von Randstücken müssen separat verwaltet werden.

Zur Umsetzung der Schaltungssuche werden bei der optimalen Steuerung gewöhnlicher Differentialgleichungen typischerweise Schießverfahren eingesetzt. Hierbei verwendet man die Formulierung des Optimalitätssystems aus Kapitel 2.1.3. Neben den Schaltpunkten treten dann die Sprungparameter an den Schaltpunkten und die Schätzung für den Anfangswert der adjungierten Gleichung als zu bestimmende Variablen auf. Wie bereits in Kapitel 2.1.3 erwähnt, bestünde bei der Anwendung von Schießverfahren für die Probleme dieser Arbeit das grundsätzliche Problem, dass die adjungierte Gleichung entgegen der Diffusionsrichtung integriert werden müsste. Theoretische Betrachtungen hinsichtlich der Wohldefiniertheit der Schießverfahren bei der optimalen Steuerung gewöhnlicher Differentialgleichungen werden in Bonnans und Hermant [10] angestellt. Einen Überblick über die numerische Umsetzung und den Einsatz der Verfahren bei realen Anwendungsproblemen findet man in Pesch [49] und [16, 17, 8].

**Primal-duale aktive-Mengen-Strategie** Bei der primal-dualen aktive-Mengen-Strategie, die erstmals für steuerungsbeschränkte Probleme in Bergounioux, Ito und Kunisch [5] eingeführt wurde, wird die Veränderung der Schätzung der aktiven Menge  $\mathcal{A}_k$  basierend auf punktwisen Informationen der primalen und dualen Variablen vorgenommen. Konkret ist die neue Schätzung für die aktive Menge im Schritt  $k + 1$  des Verfahrens

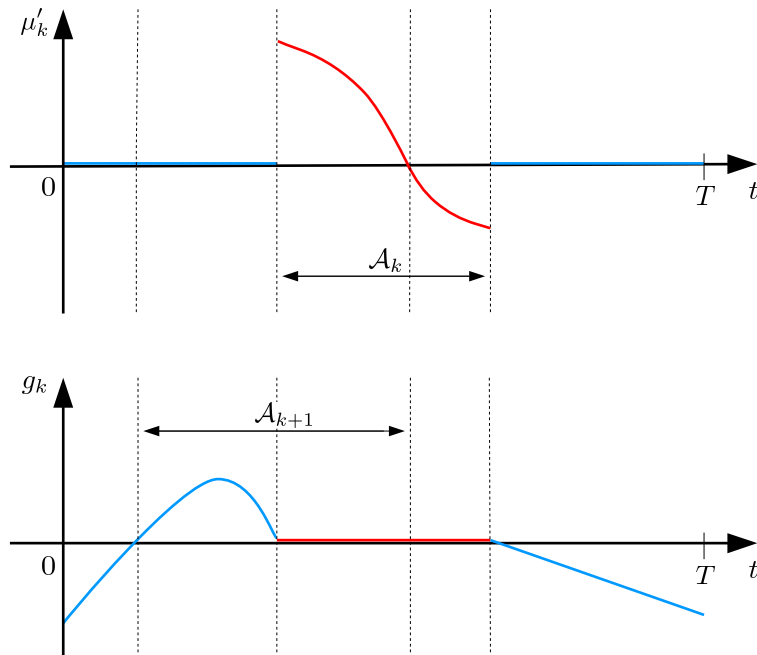


Abbildung 3.2.: Prinzip der Veränderung der aktiven Menge im PDAS Algorithmus

gegeben als

$$\mathcal{A}_{k+1} = \{t \in [0, T] \mid \mu'_k(t) + \hat{G}(u_k)(t) > 0\}. \quad (3.1)$$

Wegen der Bedingung  $\hat{G}(u_k) = 0$  auf  $\mathcal{A}_k$  wird die aktive Menge gerade um die Zeitpunkte vergrößert, in denen die Funktion  $g_k := \hat{G}(u_k)$  positiv ist. Gleichzeitig wird die Menge  $\mathcal{A}_k$  um diejenigen Zeitpunkte verringert, in denen  $\mu'_k$  nicht positiv ist. Der Algorithmus stoppt an der optimalen aktiven Menge, wenn keine Veränderungen mehr vorgenommen werden. Im Kontext von Problemen mit Zustandsbeschränkungen ist die Vorschrift (3.1) aber nur formal zu verstehen, da die distributionelle Ableitung  $\mu'$  keine punktweise Interpretation besitzen muss. Zwar konnte in Teil 2 unter Zusatzbedingungen an die Daten gezeigt werden, wann die den Lagrange-Multiplikator erzeugende Funktion  $\mu$  keine diskreten Anteile besitzt oder sogar absolut stetig ist, für die nicht optimale Konfiguration müssen diese Regularitätsaussagen jedoch nicht gelten. Ganz im Gegenteil finden sich bei den numerischen Betrachtungen in Kapitel 3.1.3 in der distributionellen Ableitung  $\mu'_k$  tatsächlich Maßanteile. Will man dieses Problem beheben, so ergeben sich zwei Möglichkeiten: Um den Algorithmus auch für zustandsbeschränkte Probleme im Funktionenraum formulieren zu können, führt man regularisierte Probleme ein, die (in einem gewissen Sinne) gegen das Ausgangsproblem konvergieren und den Vorteil bieten, dass alle beteiligten Funktionen punktweise interpretierbar sind [32, 34, 46, 47]. Alternativ betrachtet man von Anfang an eine diskretisierte Version des Problems (2.5), so dass die genannten Probleme nicht auftauchen [6]. Dieser Ansatz bildet die Grundlage für den im Rahmen dieser Arbeit entwickelten PDASSP Algorithmus. Es wird deshalb in Kapitel 3.1.2 ein diskretisiertes Modellproblem hergeleitet und in Kapitel 3.1.3 die

primal-duale aktive-Mengen-Strategie für diesen Fall exakt formuliert werden. Bei Untersuchungen des Konvergenzverhaltens dieses Algorithmus wird die Tatsache, dass  $\mu'_k$  Maßanteile besitzt, nichtsdestotrotz gewisse negative Konsequenzen nach sich ziehen.

Abschließend sollen noch kurz einige Vorteile der primal-dualen aktive-Mengen-Strategie genannt werden. Zum einen kann für den Fall eines zustandsbeschränkten diskretisierten Problems die globale Konvergenz des Verfahrens gezeigt werden [6]. Zum anderen lässt sich die primal-duale aktive-Mengen-Strategie sehr einfach implementieren, da nur mit der aktiven Menge an sich gearbeitet wird. Eine Unterscheidung nach Kontaktpunkt und Randstück und eine Bestimmung von Auf- und Absprungpunkten muss nicht vorgenommen werden. Insbesondere müssen keine Schätzungen für die Topologie der aktiven Menge vorgegeben werden. Die Topologie wird während der Iterationen des Algorithmus automatisch angepasst.

Für eine vollständige Darstellung sei an dieser Stelle auch auf Innere-Punkte-Verfahren hingewiesen, die ebenfalls zum Lösen der Probleme (2.5) eingesetzt werden könnten und im Funktionenraum formulierbar sind. Exemplarisch seien die Arbeiten von Schiela genannt [53, 52].

### 3.1.2. Zeitdiskretisiertes reduziertes Problem

Zur Herleitung eines in der Zeit semidiskretisierten Problems der Optimalsteuerungsaufgabe aus Kapitel 2.1 geht man von folgender Formulierung des Problems aus:

$$\min_{y \in W(0,T), u \in L^2(0,T)} J(y, u) := \frac{1}{2} \iint_Q (y - y_d)^2 dx dt + \frac{\lambda}{2} \int_I u^2 dt \quad (3.2a)$$

unter Berücksichtigung der Zustandsgleichung

$$\begin{aligned} \int_I \langle y_t, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt + \int_I \langle Ay, \varphi \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt \\ = \iint_Q u e \varphi dx dt \quad \forall \varphi \in L^2(0, T; H^1(\Omega)) \quad (3.2b) \\ y(0) = y_0 \end{aligned}$$

und der Zustandsbeschränkung

$$\int_{\Omega} y \delta dx \leq y_{\max} \quad \text{in } I. \quad (3.2c)$$

In den weiteren Ausführungen werden häufiger in der Zeit stückweise konstante Funktionen benötigt. Deshalb sei für einen Banachraum  $X$  und ein Intervall  $\Lambda \subset \mathbb{R}$  mit  $\mathcal{P}_0(\Lambda, X)$  der Raum aller auf dem Intervall  $\Lambda$  definierten Polynome vom Grad 0 mit Bild in  $X$  eingeführt. Außerdem seien folgende Annahmen getroffen:

**Annahme 3.1.1.**  $y_{\max}$  und  $\delta$  seien konstant in Zeitrichtung, also  $y_{\max}(t) = c$ ,  $c \in \mathbb{R}$ , und  $\delta(t) = c_f$ ,  $c_f \in L^2(\Omega)$ , für  $t \in [0, T]$ .

Die weiteren Schritte sind dem Vorgehen in Meidner, Rannacher und Vexler [45] nachempfunden: Zunächst nimmt man eine Semidiskretisierung der Zustandsgleichung in der Zeit mit „discontinuous Galerkin“ Methoden vor, wobei die Steuerung  $u$  jedoch noch als zeitkontinuierliche Funktion in  $L^2(0, T)$  betrachtet wird. Man unterteilt das Intervall  $I$  äquidistant in  $n$  Zeitintervalle der Länge  $\tau = \frac{T}{n}$ . Hierfür führt man die Zeitpunkte  $0 = t_0 < t_1 < \dots < t_{n-1} < t_n = T$ ,  $t_k = k \cdot \tau$  ein und setzt  $I_k = (t_{k-1}, t_k]$  für  $k = 1, \dots, n$ , so dass  $I = \{0\} \cup I_1 \cup I_2 \cup \dots \cup I_n$  ist. Als Ansatzfunktionen für die Zustände  $y$  und die Testfunktionen  $\varphi$  werden in der Folge in der Zeit stückweise konstante Funktionen mit Bild in  $H^1(\Omega)$  gewählt werden, also Funktionen aus dem Raum

$$W_\tau^0 := \{w^\tau \in L^2(0, T; H^1(\Omega)) \mid w^\tau|_{I_k} \in \mathcal{P}_0(I_k, H^1(\Omega)), k = 1, \dots, n\}.$$

Für eine Funktion  $w \in W_\tau^0$  definiert man für  $k = 1, \dots, n$

$$w^{\tau, k} := w^\tau(t_k),$$

so dass  $w^\tau(t) = w^{\tau, k}$  für alle  $t \in I_k$  gilt. Die in der Zeit semidiskretisierte Version der Zustandsgleichung (3.2b) ist gegeben durch

$$\begin{aligned} \int_I \langle Ay^\tau, \varphi^\tau \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt + \sum_{k=2}^n \int_\Omega (y^{\tau, k} - y^{\tau, k-1}) \varphi^{\tau, k} dx + \int_\Omega y^{\tau, 1} \varphi^{\tau, 1} dx & \quad (3.3) \\ = \iint_Q ue\varphi^\tau dxdt + \int_\Omega y_0 \varphi^{\tau, 1} dx & \quad \forall \varphi^\tau \in W_\tau^0. \end{aligned}$$

Diese Gleichung besitzt wie im kontinuierlichen Fall eine eindeutige Lösung  $y^\tau \in W_\tau^0$  [45]. Außerdem existiert ein affin linearer Operator  $\hat{S}^\tau: L^2(0, T) \rightarrow W_\tau^0$ ,  $u \mapsto y^\tau = S^\tau u + y_{\text{init}}^\tau$ . Hierbei ist  $S^\tau: L^2(0, T) \rightarrow W_\tau^0$  ein linear stetiger Operator, der jede Steuerung  $u \in L^2(0, T)$  auf die zugehörige Lösung  $y^\tau \in W_\tau^0$  der Gleichung (3.3) zum Anfangswert  $y_0 = 0$  abbildet. Die Funktion  $y_{\text{init}}^\tau \in W_\tau^0$  stellt die Lösung der Gleichung (3.3) zur Steuerung  $u = 0$  und Anfangswert  $y_0$  dar.

Nach der Betrachtung der zeitdiskretisierten Version der Zustandsgleichung soll nun eine diskretisierte Version der Zustandsbeschränkung hergeleitet werden. Da es sich bei Lösungen  $y^\tau$  der Zustandsgleichung (3.3) um in der Zeit stückweise konstante Funktionen handelt, kann die Zustandsbeschränkung (3.2c) unter Verwendung der Annahme 3.1.1 äquivalent formuliert werden als

$$\int_\Omega y^{\tau, k} \delta dx - y_{\max} \leq 0 \quad \text{für } k = 1, \dots, n.$$

Definiert man den affin linearen Operator

$$\hat{G}^\tau: L^2(0, T) \rightarrow \mathbb{R}^n, u \mapsto \left( \int_\Omega (S^\tau u)(t_k) \delta dx + \int_\Omega y_{\text{init}}^\tau(t_k) \delta dx - y_{\max} \right)_{k=1, \dots, n}$$

mit linear stetigem Anteil  $G^\tau: L^2(0, T) \rightarrow \mathbb{R}^n$ ,  $u \mapsto \left( \int_\Omega (S^\tau u)(t_k) \delta dx \right)_{k=1, \dots, n}$ , so kann man alles in allem das Optimalsteuerungsproblem formulieren als:

$$\min_{u^\tau \in L^2(0, T)} \mathcal{J}_\tau(u^\tau) := J(\hat{S}^\tau(u^\tau), u^\tau) \quad \text{unter der Nebenbedingung} \quad \hat{G}^\tau(u^\tau) \leq 0. \quad (3.4)$$

Nach Meidner, Rannacher und Vexler [45] ist bei hinreichend kleiner Schrittweite  $\tau$  auch für dieses Problem die Existenz einer eindeutigen optimalen Lösung  $\bar{u}^\tau$  gesichert. Des Weiteren lässt sich für dieses Problem analog zum Vorgehen in Kapitel 2.1.1 ein Optimalitätssystem herleiten. Da dieses in der Folge nicht im Detail benötigt wird, wird auf dessen Darstellung an dieser Stelle verzichtet. Stattdessen sei nur auf ein wichtiges Ergebnis hingewiesen, für das eine zusätzliche Annahme nötig ist:

**Annahme 3.1.2.**  $e \in E_\tau^0 := \{e \in L^\infty(0, T; L^2(\Omega)) \mid e|_{I_k} \in \mathcal{P}_0(I_k, L^2(\Omega)), k = 1, \dots, n\}$ .

Unter dieser Voraussetzung gehört die optimale Lösung  $\bar{u}^\tau$  des Problems (3.4) dem Raum

$$U_\tau^0 := \{u^\tau \in L^2(0, T) \mid u^\tau|_{I_k} \in \mathcal{P}_0(I_k, \mathbb{R}), k = 1, \dots, n\}$$

an [45].  $\bar{u}^\tau$  ist somit ebenfalls eine stückweise konstante Funktion. In Folge kann man sich bei der Optimierung ohne Einschränkung auf den Teilraum  $U_\tau^0 \subset L^2(0, T)$  zurückziehen und statt des Problems (3.4) die Optimierungsaufgabe

$$\min_{u^\tau \in U_\tau^0} \mathcal{J}_\tau(u^\tau) \quad \text{unter der Nebenbedingung} \quad \hat{G}^\tau(u^\tau) \leq 0 \quad (3.5)$$

betrachten.

Abschließend soll für dieses endlichdimensionale reduzierte Problem eine explizite Darstellung hergeleitet werden. Jedes  $u^\tau \in U_\tau^0$  lässt sich mittels Indikatorfunktionen  $\varepsilon_i \in U_\tau^0$  mit

$$\varepsilon_i(t) = \begin{cases} 1, & t \in I_i \\ 0, & t \in I \setminus I_i \end{cases}$$

schreiben als  $u^\tau = \sum_{i=1}^n u_i \cdot \varepsilon_i$  mit  $u_i \in \mathbb{R}$ . Setzt man  $y_i^\tau := S^\tau \varepsilon_i$ , so hat man

$$\hat{S}^\tau(u^\tau) = y^\tau = \sum_{i=1}^n u_i \cdot y_i^\tau + y_{\text{init}}^\tau. \quad (3.6)$$

Damit lässt sich die Bedingung  $\hat{G}^\tau(u^\tau) \leq 0$  schreiben als

$$\sum_{i=1}^n \left( \int_{\Omega} y_i^{\tau,k} \delta \, dx \right) \cdot u_i + \int_{\Omega} y_{\text{init}}^{\tau,k} \delta \, dx - y_{\text{max}} \leq 0 \quad \text{für } k = 1, \dots, n.$$

Definiert man die Einträge  $a_{k,i}$  der Matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  und  $b_k$  des Vektors  $\vec{b} \in \mathbb{R}^n$  durch

$$\mathbf{a}_{k,i} := \int_{\Omega} y_i^{\tau,k} \delta \, dx, \quad b_k := - \int_{\Omega} y_{\text{init}}^{\tau,k} \delta \, dx + y_{\text{max}},$$

so kann man die Bedingung  $\hat{G}^\tau(u^\tau) \leq 0$  äquivalent formulieren als

$$\mathbf{A} \cdot \vec{u} \leq \vec{b}.$$



Setzt man  $u^\tau = \sum_{i=1}^n u_i \cdot \varepsilon_i$  in das Zielfunktional  $\mathcal{J}_\tau$  ein und verwendet die Darstellung (3.6), erhält man

$$\begin{aligned} \mathcal{J}_\tau(u^\tau) &= \frac{1}{2} \iint_Q \left( \sum_{i=1}^n u_i \cdot y_i^\tau \right)^2 dxdt + \iint_Q \left( \sum_{i=1}^n u_i \cdot y_i^\tau \right) (y_{\text{init}}^\tau - y_d) dxdt \\ &\quad + \frac{1}{2} \iint_Q (y_{\text{init}}^\tau - y_d)^2 dxdt + \frac{\lambda}{2} \int_I \left( \sum_{i=1}^n u_i \cdot \varepsilon_i \right)^2 dt \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n u_i u_j \iint_Q y_i^\tau y_j^\tau dxdt + \sum_{i=1}^n u_i \cdot \iint_Q y_i^\tau (y_{\text{init}}^\tau - y_d) dxdt \\ &\quad + \frac{1}{2} \iint_Q (y_{\text{init}}^\tau - y_d)^2 dxdt + \frac{\lambda}{2} \sum_{i=1}^n \sum_{j=1}^n u_i u_j \int_I \varepsilon_i \varepsilon_j dt. \end{aligned}$$

Mit der symmetrischen Matrix  $\mathbf{H} \in \mathbb{R}^{n \times n}$  und dem Vektor  $\vec{f} \in \mathbb{R}^n$ , gegeben durch

$$\mathbf{h}_{i,j} := \iint_Q y_i^\tau y_j^\tau dxdt + \lambda \int_I \varepsilon_i \varepsilon_j dt, \quad f_i := \iint_Q y_i^\tau (y_{\text{init}}^\tau - y_d) dxdt,$$

ist das Optimierungsproblem (3.5) letztendlich bis auf die additive Konstante

$$\frac{1}{2} \iint_Q (y_{\text{init}}^\tau - y_d)^2 dxdt$$

äquivalent zum endlichdimensionalen reduzierten quadratischen Optimierungsproblem

$$\min_{\vec{u} \in \mathbb{R}^n} \frac{1}{2} \vec{u}^\top \mathbf{H} \vec{u} + \vec{f}^\top \vec{u} \quad \text{unter der Nebenbedingung} \quad \mathbf{A} \vec{u} \leq \vec{b}. \quad (3.7)$$

Existiert eine zulässige Steuerung, also ein  $\vec{u} \in \mathbb{R}^n$  so dass  $\mathbf{A} \vec{u} \leq \vec{b}$  gilt, so besitzt das quadratische Problem (3.7) für  $\lambda > 0$  eine eindeutige optimale Lösung  $\vec{u}$ . Diese ist charakterisiert durch die Existenz eines Lagrange-Multiplikators  $\vec{\eta} \in \mathbb{R}^n$ , mit dem das folgende Optimalitätssystem erfüllt ist [2]

$$\mathbf{A} \vec{u} \leq \vec{b}, \quad (3.8a)$$

$$\vec{\eta} \geq 0, \quad (3.8b)$$

$$\eta_j \cdot (\mathbf{A} \vec{u} - \vec{b})_j = 0 \quad \text{für } j = 1, \dots, n, \quad (3.8c)$$

$$\mathbf{H} \vec{u} + \vec{f} + \mathbf{A}^\top \vec{\eta} = 0. \quad (3.8d)$$

**Bemerkung 3.1.1.** Gemäß Meidner, Rannacher und Vexler [45] besteht für hinreichend reguläre Probleme zwischen der Approximation eines Lagrange-Multiplikators  $\mu_t \in L^1(0, T)$  aus Kapitel 2.1.1 mit stückweise konstanten Funktionen  $\mu_t^\tau \in U_\tau^0$  und dem Lagrange-Multiplikator  $\vec{\eta}$  in (3.8) gerade der Zusammenhang

$$\mu_t^\tau|_{I_j} = \frac{1}{\tau} \cdot \eta_j \quad \text{für } j = 1, \dots, n.$$

### 3.1.3. Grundidee des PDASSP Algorithmus

Der in dieser Arbeit entwickelte PDASSP Algorithmus stellt eine Erweiterung der primal-dualen aktive-Mengen-Strategie (PDAS) für diskretisierte zustandsbeschränkte Optimalsteuerungsprobleme dar. Der PDAS Algorithmus mit Untersuchungen zum Konvergenzverhalten ist für derartige Probleme in Bergounioux und Kunisch [6] dargestellt. An dieser Stelle sei aber auch auf die Arbeit von Hager und Ianculescu [30] hingewiesen.

Bevor die im Rahmen dieser Arbeit verwendete Umsetzung der primal-dualen aktive-Mengen-Strategie für das reduzierte Problem (3.7) dargestellt wird, seien noch einige Definitionen vorgenommen. Für  $\vec{v}, \vec{w} \in \mathbb{R}^n$  bezeichne  $\{i \mid \vec{v} \leq \vec{w}\}$  die Menge aller Indizes, für die die  $i$ -te Koordinate von  $\vec{v} - \vec{w}$  nicht positiv ist. Die aktive und inaktive Menge für das Problem (3.7) ist dann gegeben durch

$$\mathcal{A}^d := \{i \mid \mathbf{A}\vec{u} = \vec{b}\}, \quad \mathcal{I}^d := \{i \mid \mathbf{A}\vec{u} < \vec{b}\}.$$

Die Schätzungen dieser Mengen im  $k$ -ten Schritt des PDAS Algorithmus seien definiert als  $\mathcal{A}_k^d$  bzw.  $\mathcal{I}_k^d$ . Damit lautet der PDAS Algorithmus:

**Daten :** Matrizen und Vektoren  $\mathbf{H}, \vec{f}, \mathbf{A}, \vec{b}$  des Problems (3.7), maximale Iterationsanzahl  $k_{\max}$

**Ergebnis :** Lösung  $\vec{u}$  von (3.7), Multiplikator  $\vec{\eta}$ , Anzahl an Iterationen  $k$

- 1 Setze  $\mathcal{A}_1^d = \emptyset, \vec{\eta}_1 = \vec{0}$  und  $k = 1$ ;
- 2 Bestimme  $\vec{u}_1 \in \mathbb{R}^n$  als Lösung von  $\mathbf{H}\vec{u}_1 = -\vec{f}$ ;
- 3 **solange**  $k \leq k_{\max}$  **tue**
- 4     Setze  $k = k + 1$  ;
- 5     Bestimme neue aktive und inaktive Menge

$$\mathcal{A}_k^d = \{i \mid \vec{\eta}_k + \mathbf{A}\vec{u}_k - \vec{b} > \vec{0}\}, \quad \mathcal{I}_k^d = \{i \mid \vec{\eta}_k + \mathbf{A}\vec{u}_k - \vec{b} \leq \vec{0}\};$$

- 6     **wenn**  $\mathcal{A}_k^d = \mathcal{A}_{k-1}^d$  **dann**
- 7         Setze  $k = k - 1$ ;
- 8         STOP;
- 9     **Ende**
- 10     Bestimme  $\vec{u}_k \in \mathbb{R}^n, \vec{\eta}_k \in \mathbb{R}^n$  als Lösung von

$$\begin{aligned} \mathbf{H}\vec{u}_k + \mathbf{A}^\top \vec{\eta}_k &= -\vec{f}, \\ \mathbf{A}\vec{u}_k &= \vec{b} \quad \text{in } \mathcal{A}_k^d, \\ \vec{\eta}_k &= \vec{0} \quad \text{in } \mathcal{I}_k^d; \end{aligned}$$

- 11 **Ende**

**Algorithmus 1 : PDAS**

Im Stopkriterium wird der Iterationsindex um eine Iteration reduziert, da in dem weiteren Fortgang der Arbeit beim Vergleich von Algorithmen die Anzahl der gelösten Optimalitätssysteme gezählt werden wird. Um nachzuweisen, dass in Schritt 10 des

PDAS Algorithmus stets eine eindeutige Lösung existiert, verwendet man typischerweise die Tatsache, dass das formulierte Gleichungssystem im  $k$ -ten Schritt der PDAS gerade das Optimalitätssystem des gleichungsbeschränkten Problems

$$\min_{\vec{u} \in \mathbb{R}^n} \frac{1}{2} \vec{u}^\top \mathbf{H} \vec{u} + \vec{f}^\top \vec{u} \quad \text{unter der Nebenbedingung} \quad \mathbf{A} \vec{u} = \vec{b} \quad \text{auf } \mathcal{A}_k^d$$

darstellt. Die Existenz eines Lagrange-Multiplikators und die Lösbarkeit des Optimalitätssystems für dieses Hilfsproblem wäre garantiert, wenn man die Surjektivität der Abbildung  $\vec{u} \mapsto (\mathbf{A} \vec{u} - \vec{b})|_{\mathcal{A}_k^d}$  nachweisen könnte. Dies stellt im Rahmen der in dieser Arbeit frei wählbaren Datenfunktionen keine leicht nachweisbare Aussage dar und ist stark von der konkreten Wahl der Datenfunktionen abhängig. Deswegen sei die folgende Annahme getroffen, die insbesondere beinhaltet, dass jedes der Teilprobleme in Schritt 10 des PDAS Algorithmus eine eindeutige Lösung besitzt.

**Annahme 3.1.3.** Alle Daten seien so gewählt, dass der PDAS Algorithmus ausgehend von jeder beliebig vorgegebenen Indexmenge  $\mathcal{A}_k^d \subset \{1, 2, \dots, n\}$  in endlich vielen Schritten konvergiert.

Das Konvergenzverhalten des PDAS Algorithmus bei Problemen der Form (3.7) soll nun näher analysiert werden. In den folgenden Abbildungen werden die aus dem PDAS Algorithmus in der  $k$ -ten Iteration erhaltenen Lösungen aufgeführt. Hierzu werden statt des Vektors  $\vec{u}^k$  und des Multiplikators  $\vec{\eta}^k$  die korrespondierenden stückweise konstanten Funktionen  $u_k^\tau$  und  $\mu_{t,k}^\tau$  dargestellt (Kapitel 3.1.2). Aus Gründen der Übersichtlichkeit werden in den Darstellungen statt Stufenplots Linienplots gewählt. Die Datenpunkte an den Endpunkten jedes Zeitintervalls  $I_i$ , beispielsweise  $(t_i, u_k^\tau(t_i))$  für  $i = 1, \dots, n$ , werden per Geradenstück verbunden. Gehört der Endpunkt in der  $k$ -ten Iteration der PDAS zur aktuellen Schätzung der aktiven Menge, so wird das Geradenstück rot eingefärbt, andernfalls blau. Die Approximationen aus den vorausgegangenen  $k - 1$  Iterationen der PDAS werden einheitlich grau dargestellt. Die zu den Indexmengen  $\mathcal{A}^d$  und  $\mathcal{I}^d$  korrespondierenden Mengen an aktiven und inaktiven Zeitpunkten seien definiert durch

$$\mathcal{A}^{ti} := \{t_i \in \{t_1, \dots, t_n\} | i \in \mathcal{A}^d\}, \quad \mathcal{I}^{ti} := \{t_i \in \{t_1, \dots, t_n\} | i \in \mathcal{I}^d\}.$$

Analog setzt man für die Mengen in der  $k$ -ten Iteration der PDAS

$$\mathcal{A}_k^{ti} := \{t_i \in \{t_1, \dots, t_n\} | i \in \mathcal{A}_k^d\}, \quad \mathcal{I}_k^{ti} := \{t_i \in \{t_1, \dots, t_n\} | i \in \mathcal{I}_k^d\}.$$

Der PDAS Algorithmus wird zunächst auf das Problem 1 aus Kapitel 3.2.2 mit einer Zeitdiskretisierung von  $n = 100$  Zeitschritten angewandt. Es zeigt sich folgendes Verhalten, welches sich auch in anderen Testproblemen widerspiegelt. Nachdem im ersten Schritt der primal-dualen aktive-Mengen-Strategie die Lösung des unbeschränkten Problems berechnet wurde, liefert der Algorithmus im zweiten Schritt bereits eine gute Schätzung der aktiven Menge  $\mathcal{A}^{ti}$  am Optimum (Abbildung 3.3). Des Weiteren ist die Steuerung auf Teilintervallen sehr glatt. Nur beim Übergang von inaktiver zu aktiver Menge und umgekehrt treten Unstetigkeiten in Form großer Sprünge auf. An dieser Stelle sei nochmals daran erinnert, dass „kleine“ Sprünge in  $u_2^\tau$  zu den Zeitpunkten im

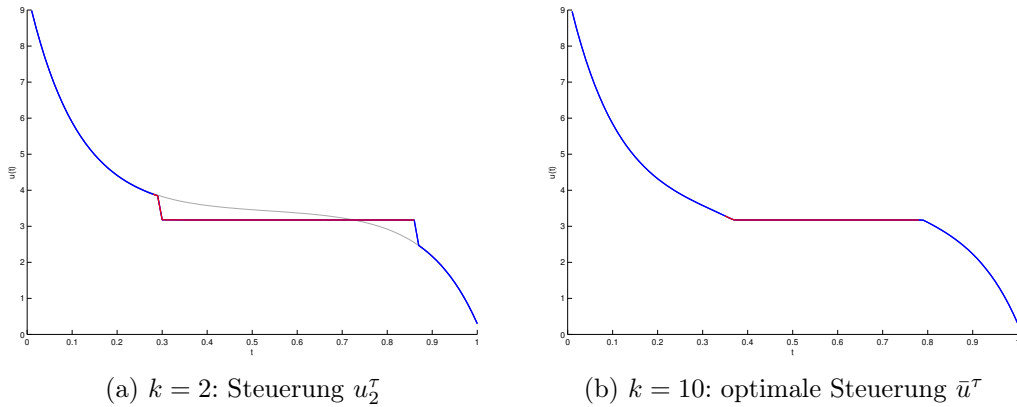


Abbildung 3.3.: Problem 1: Steuerung  $u_k^\tau$  nach  $k$  Iterationen der PDAS für  $n = 100$ ; blau/rot: Lösungen aus aktuellem Schritt  $k$  auf aktiver (rot) und inaktiver Menge (blau), grau: vorangegangene Lösungen

Zeitgitter in natürlicher Weise auftreten, da es sich um stückweise konstante Funktionen in  $U_\tau^0$  handelt. Bei genauer Betrachtung fällt auf, dass die Unstetigkeit der Steuerung nicht zum Zeitpunkt des Aufsprungs auf die aktive Menge ( $=: t_k^{\text{en}}$ ) sondern zwischen diesem und dem zweiten Punkt ( $=: t_k^{\text{en}+1}$ ) in der aktiven Menge auftritt. Beim Absprung von der aktiven Mengen ist ein Sprung zwischen dem letzten aktiven Punkt  $t_k^{\text{ex}}$  und dem ersten erneut inaktiven Zeitpunkt ( $=: t_k^{\text{ex}+1}$ ) vorhanden. Dies lässt sich bei Betrachtung des Multiplikators in Abbildung 3.4 wie folgt erklären. Im ersten und letzten Datenpunkt der aktiven Menge  $t_k^{\text{en}}$  und  $t_k^{\text{ex}}$  treten im Graphen von  $\mu_{t,2}^\tau$  hohe Spitzen auf. Diese hohen Spitzen verursachen über die Kopplungen von  $u^\tau$  und  $\mu_t^\tau$  im Optimalitätssystem jeweils einen Sprung der Steuerung zum darauffolgenden Zeitintervall.

Darüber hinaus indizieren die Spitzen in der Darstellung 3.4 von  $\mu_{t,2}^\tau$ , dass bei Betrachtung des Problems im nicht diskretisierten Fall die Stammfunktion  $\mu_2 \in NBV_T(0, T)$  für die Konfiguration in der zweiten Iteration nicht absolut stetig wäre. Vielmehr besäße  $\mu_2$  einen diskreten Anteil, so dass in  $d\mu_2$  Diracmaße vorhanden wären. Allerdings konzentrieren sich diese Unregelmäßigkeiten nur auf den Beginn und das Ende der aktiven Menge. Zwischen den beiden Spitzen verläuft  $\mu_{t,2}^\tau$  sehr regulär. Im Zusammenhang mit Schritt 5 im PDAS Algorithmus zum Update der aktiven Menge ist es interessant, das Vorzeichen der Funktion  $\mu_{t,2}^\tau$  zu betrachten. Aus der aktiven Menge werden im aktuellen Schritt gerade die Zeitpunkte entfernt, in denen  $\mu_{t,2}^\tau$  nicht positiv ist. Dies sind nur die beiden Zeitpunkte  $t_k^{\text{en}}$  und  $t_k^{\text{ex}}$ , in denen die Spitzen vorhanden sind. Der reguläre Anteil innerhalb der aktiven Menge ist positiv. Entsprechend wird die Indexmenge  $\mathcal{A}_2^d$  nur um zwei Indizes verkleinert (Tabelle 3.1).

Dieses Verhalten von  $\mu_{t,k}^\tau$  wiederholt sich über die weiteren Iterationen hinweg (Abbildung 3.5). Der reguläre Anteil von  $\mu_{t,k}^\tau$  variiert während der Iterationen des PDAS Algorithmus, und damit der Variation der Schätzung der aktiven Menge, in der Nähe der optimalen aktiven Menge nur sehr schwach und bleibt stets positiv. Jeweils im ersten aktiven Zeitpunkt  $t_k^{\text{en}}$  und letzten aktiven Zeitpunkt  $t_k^{\text{ex}}$  treten Spitzen in  $\mu_{t,k}^\tau$  auf, da dort das Verhalten eines Diracmaßes nachgeahmt wird. Nur in diesen Zeitpunkten

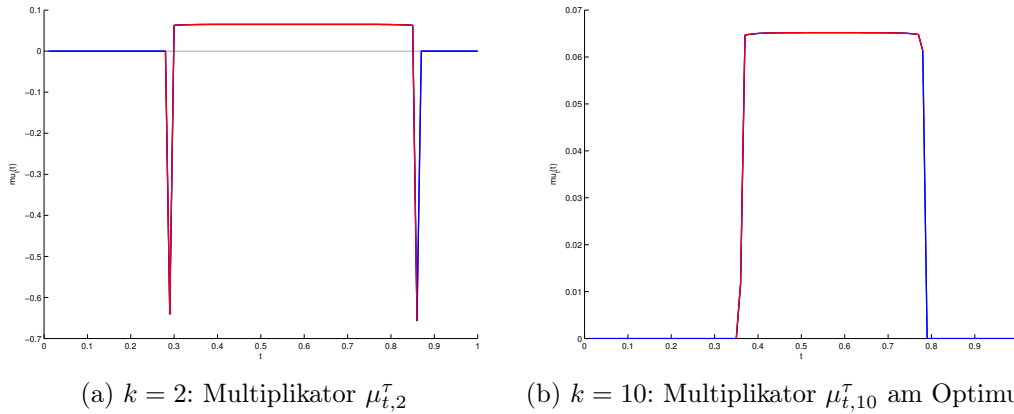


Abbildung 3.4.: Problem 1: Multiplikator  $\mu_{t,k}^\tau$  nach  $k$  Iterationen der PDAS für  $n = 100$ ; blau/rot: Lösungen aus aktuellem Schritt  $k$  auf aktiver (rot) und inaktiver Menge (blau), grau: vorangegangene Lösungen

ist  $\mu_{t,k}^\tau$  negativ, so dass bis zur Konvergenz des Verfahrens in Iteration 10 je Iteration maximal der erste und der letzte Index aus  $\mathcal{A}_k^d$  eliminiert werden (Tabelle 3.1).

PDAS Iteration $k$	1	2	3	4	5	6	7	8	9	10
Anzahl der Elemente $n_k$ in $\mathcal{A}_k^d$	0	58	56	54	52	50	48	46	44	43

Tabelle 3.1.: Entwicklung der aktiven Menge während der PDAS für  $n = 100$

Dies hat zur Folge, dass der PDAS Algorithmus bei einer Diskretisierung mit doppelter Anzahl an Zeitschritten ab der Iteration, ab der nur noch eine Verringerung um Anfangs- und Endindex auftritt, doppelt so viele Schritte benötigt (Tabelle 3.2). Der Algorithmus ist gitterabhängig, was auch in Bergounioux und Kunisch [6] und Bergounioux u. a. [7] festgestellt wurde.

Anzahl der Zeitschritte $n$	100	200	400	800
Anzahl an Iterationen	10	16	32	62

Tabelle 3.2.: Anzahl an Iterationen des PDAS Algorithmus in Abhängigkeit der Anzahl der Zeitschritte für Problem 1

An dieser Stelle sei darauf hingewiesen, dass bei einer Vergrößerung der aktiven Menge in Schritt 5 des Algorithmus die PDAS keine Gitterabhängigkeit zeigt. Eine Vergrößerung der aktiven Menge würde mit der Approximation der Funktion  $g$  in der Zustandsbeschränkung erfolgen. Diese Funktion ist auch für die nicht optimale Konfiguration hinreichend regulär. Insbesondere beinhaltet sie keine Diracmaß Anteile (Abbildung 3.6).

In dieser Arbeit soll ein Algorithmus entwickelt werden, der möglichst keine Gitterabhängigkeit zeigt. Die Idee hierbei ist, die in dem Teil 2 bewiesenen Resultate der Stetigkeit der optimalen Steuerung zu nutzen, und die primal-duale aktive-Mengen-Strategie

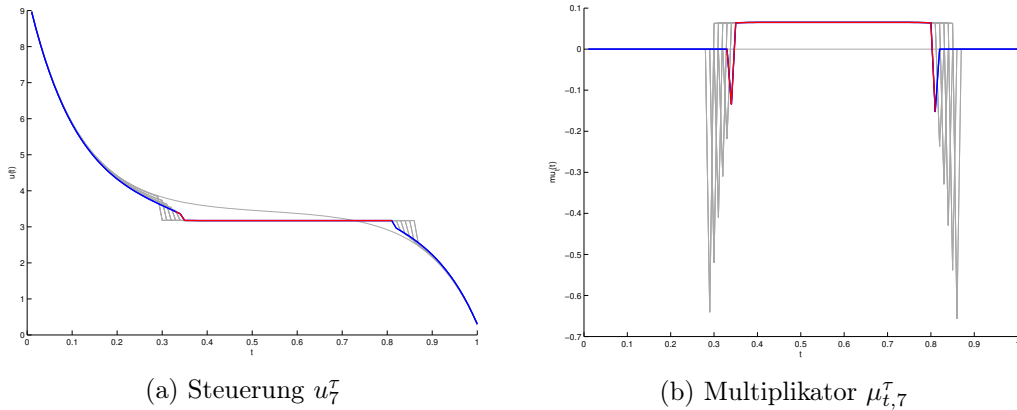


Abbildung 3.5.: Problem 1: Steuerung und Multiplikator nach  $k = 7$  Iterationen der PDAS für  $n = 100$ ; blau/rot: Lösungen aus aktuellem Schritt  $k$  auf aktiver (rot) und inaktiver Menge (blau), grau: vorangegangene Lösungen

mit einer Schaltungssuche zu verknüpfen. Ein derartiges Vorgehen wurde in einem gewissen Sinne erstmalig im Kontext der optimalen Steuerung von zustandsbeschränkten elliptischen Gleichungen in Frey [22] verwendet. Dort wurde ausgehend von einer Schätzung der aktiven Menge aus dem ersten Schritt des PDAS Algorithmus mit einem Shape-Newton-Verfahren versucht, die Form der aktiven Menge so zu variieren, dass die Stetigkeit der optimalen Steuerung erreicht wird. Im Rahmen dieses Vorgehens wurde für jede Iteration des Shape-Newton-Verfahrens ein neues Gitter erzeugt, um möglichst exakte Resultate zu erhalten. Ein Problem dieses Ansatzes ist, dass die aus dem PDAS Verfahren erhaltene Startschätzung nicht notwendigerweise im Konvergenzbereich des Shape-Newton-Verfahrens liegen muss. In dieser Arbeit soll ein anderer Ansatz verfolgt werden. Eine Schaltungssuche soll direkt in die Iterationen des PDAS Algorithmus für das diskretisierte Problem integriert werden, um bessere, nicht gitterabhängige, Vorschläge für die neue aktive (Index-)Menge zu erhalten. Die zugrundeliegende Zeitdiskretisierung soll während der Schaltungssuche nicht verändert und Konvergenz des Verfahrens über die PDAS Strategie sicher gestellt werden. In diesem Sinne stellt der nachfolgend entwickelte PDASSP Algorithmus eine Erweiterung der primal-dualen aktive-Mengen-Strategie dar.

Konzeptionell geht man in einem ersten Versuch für das obige Beispiel mit einer aktiven Menge wie folgt vor:

1. Führe die PDAS Strategie so lange aus, bis sich die Schätzung für die neue aktive Menge in Schritt 5 des Algorithmus am Aufsprungpunkt oder am Absprungpunkt der aktiven Menge nur noch um einen Zeitpunkt im Vergleich zur vorherigen Schätzung verändert.
2. Versuche die Verringerung der aktiven Menge an diesem Zeitpunkt durch einen Schritt einer Schaltungssuche zu verbessern. Führe einen Schritt in einem Newton-

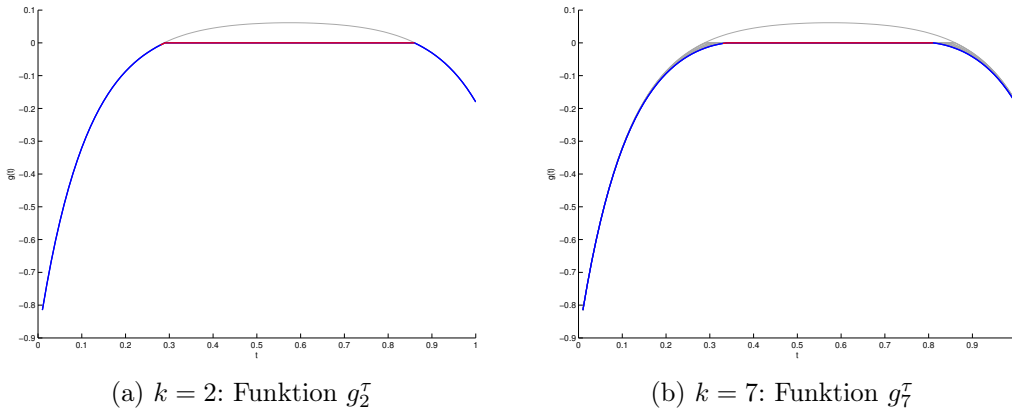


Abbildung 3.6.: Problem 1: Funktion  $g^\tau$  in der Zustandsbeschränkung nach  $k$  Iterationen der PDAS für  $n = 100$ ; blau/rot: Lösungen aus aktuellem Schritt  $k$  auf aktiver (rot) und inaktiver Menge (blau), grau: vorangegangene Lösungen

Verfahren aus um im  $k$ -ten Schritt

$$[u_k^\tau(t_k^{\text{en}})] := u_k^\tau(t_k^{\text{en}+1}) - u_k^\tau(t_k^{\text{en}}) \approx 0 \quad \text{bzw.} \quad [u_k^\tau(t_k^{\text{ex}})] := u_k^\tau(t_k^{\text{ex}+1}) - u_k^\tau(t_k^{\text{ex}}) \approx 0$$

zu erreichen.

3. Runde die aus dem Newton-Verfahren erhaltene neue Schätzung für den Anfangs- oder Endpunkt auf den nächsten im Zeitgitter vorhandenen Zeitpunkt und passe entsprechend die aktive Menge an.
4. Falls das Newton-Verfahren einen vernünftigen Vorschlag für die aktive Menge liefert, fahre mit dieser Schätzung im 1. Schritt dieses Konzepts fort. Andernfalls verwende den Vorschlag aus PDAS und springe zu Schritt 1..

Ein wichtiger Punkt in diesem Konzept ist, dass die Schaltpunktsuche nicht genutzt werden soll, um Auf- und Absprungpunkte mit einer hohen Genauigkeit zu detektieren. Dies wäre im Kontext einer Approximation der Steuerung mit stückweise konstanten Funktionen in  $U_\tau^0$  nicht zielführend. Vielmehr geht es darum, eine bessere Schätzung für das Update der aktiven (Index-)Menge mit vertretbarem Rechenaufwand zu liefern, insbesondere wenn man auch den Rundungsvorgang in Schritt 3. und das mögliche Umschalten auf den PDAS Vorschlag in Schritt 4. berücksichtigt.

In Schritt 2. stellt sich deshalb die Frage, wie die für ein Newton-Verfahren nötigen Ableitungen effizient berechnet werden können. Dabei gilt zu beachten, dass grundsätzlich durch die Kopplung über das Optimalitätssystem der Sprung der Steuerung in  $t^{\text{en}}$  auch abhängig ist von der Veränderung von  $t^{\text{ex}}$  und umgekehrt. Im numerischen Tests zeigte sich jedoch, dass dieser Einfluss sehr gering ist. So wird beispielsweise der Sprung in  $t^{\text{en}}$  überwiegend von der Variation von  $t^{\text{en}}$  selbst beeinflusst. In Folge wird in den weiteren Schritten nur noch diese Abhängigkeit berücksichtigt. Da es allerdings nach

wie vor zu aufwendig wäre, die nötigen Sensitivitäten über zugehörige Sensitivitätsgleichungen zu berechnen, soll ein anderer Weg verfolgt werden. Das Newton Verfahren in Schritt 2. ersetze man durch einen Schritt eines Sekantenverfahrens. Dies lässt sich mit einer Abwandlung im ersten Schritt des Konzepts effizient implementieren.

Man führt die Vorschläge aus der PDAS Strategie so lange aus, bis sich die aktive Menge am Auf- oder Absprungpunkt jeweils im Vergleich zur letzten Iteration um einen Index, und zur vorletzten Iteration um zwei Indizes verändert hat. In der  $k$ -ten Iteration überprüft man beispielsweise, ob  $t_k^{\text{en}} = t_{k-1}^{\text{en}+1}$  und  $t_{k-1}^{\text{en}} = t_{k-2}^{\text{en}+1}$ . Zusätzlich speichert man in jeder Iteration die Sprungwerte  $[u_k^\tau(t_k^{\text{en}})]$  ab. Die Veränderung  $\Delta t_k^{\text{en}}$  im zweiten Schritt des oberen Konzepts ist bei Verwendung eines Sekantenverfahrens dann gegeben durch

$$\Delta t_k^{\text{en}} = -\frac{\tau}{[u_{k-1}^\tau(t_{k-1}^{\text{en}})] - [u_{k-2}^\tau(t_{k-2}^{\text{en}})]} \cdot [u_{k-1}^\tau(t_{k-1}^{\text{en}})].$$

Übersetzt in Indizes  $\Delta i_k^{\text{en}}$ , um die der alte Aufsprungpunkt  $t_{k-1}^{\text{en}}$  auf den neuen Zeitpunkt  $t_{k,\text{neu}}^{\text{en}}$  im Zeitgitter verschoben werden muss, hat man mit einer Rundungsfunktion  $r$  auf die nächste ganze Zahl

$$\Delta i_k^{\text{en}} = r\left(\frac{\Delta t_k^{\text{en}}}{\tau}\right).$$

Das resultierende Vorgehen ist in der Abbildung 3.7 graphisch veranschaulicht. Um in

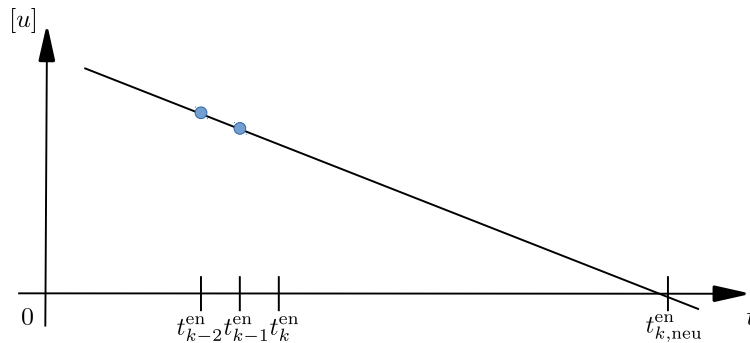


Abbildung 3.7.: Prinzipskizze zur Bestimmung eines neuen Aufsprungpunktes mit Sekantenverfahren

Schritt 4 zu bewerten, ob die erhaltene Veränderung tatsächlich sinnvoll ist, vergleicht man  $\Delta i_k^{\text{en}}$  mit der Anzahl an Elementen in der aktiven Menge  $n_k^{\text{PDAS}} := |\mathcal{A}_k^d|$  aus dem Vorschlag der PDAS Strategie. Da Veränderungen von Auf- und Absprungpunkt simultan stattfinden können, wird im Rahmen dieser Arbeit die Bedingung gefordert, dass mit der Gaußklammer  $\lfloor \cdot \rfloor$

$$\Delta i_k^{\text{en}} \leq \left\lfloor \frac{n_k^{\text{PDAS}}}{2} \right\rfloor$$

gelten muss, damit durch eine simultane Veränderung von Auf- und Absprungpunkt die Topologie der aktiven Menge nicht verändert wird. Um sicherzustellen, dass mindestens



der Vorschlag aus der PDAS Strategie ausgeführt wird, verlangt man zusätzlich

$$\Delta i_k^{\text{en}} \geq 1.$$

Zusammen:

$$\Delta i_k^{\text{en}} = \max\left\{1, \min\left\{\Delta i_k^{\text{en}}, \left\lfloor \frac{n_k^{\text{PDAS}}}{2} \right\rfloor\right\}\right\}.$$

Im Kontext dieser Beschränkungen an  $\Delta i_k^{\text{en}}$  ist es im Rahmen eines Algorithmus aus Effizienzgründen nur dann sinnvoll einen Schritt eines Sekantenverfahrens auszuführen, wenn für die Schätzung der neuen aktiven Menge aus Schritt 5 des PDAS die Bedingung  $n_k^{\text{PDAS}} \geq 4$  erfüllt ist.

Als erste Basisversion eines aus der primal-dualen aktive-Mengen-Strategie (PDAS) und einer Schaltpunktsuche (SP) kombinierten PDASSP Algorithmus erhält man für den Fall einer aktiven Menge wie in Problem 1 aus Kapitel 3.2.2 den Algorithmus 2. Dabei seien  $i_k^{\text{en}}$  und  $i_k^{\text{ex}}$  analog zu  $t_k^{\text{en}}$  und  $t_k^{\text{ex}}$  als das kleinste bzw. größtes Element in  $\mathcal{A}_k^d$  bezeichnet.  $\vec{u}_k(i)$  sei die  $i$ -te Koordinate von  $\vec{u}_k$  und für  $i = 1, \dots, n-1$  sei  $[\vec{u}_k(i)]$  definiert als  $[\vec{u}_k(i)] := \vec{u}_k(i+1) - \vec{u}_k(i)$ .

Die Ergebnisse des PDASSPbasic Algorithmus 2 für Problem 1 bei einer Zeitdiskretisierung mit  $n = 100$  Zeitschritten sind in Abbildung 3.8 und Tabelle 3.3 dargestellt.

PDASSPbasic Iteration $k$	1	2	3	4	5	6
Anzahl der Elemente $n_k$ in $\mathcal{A}_k^d$	0	58	56	46	44	43

Tabelle 3.3.: Entwicklung der aktiven Menge während des PDASSPbasic für Problem 1 für  $n = 100$

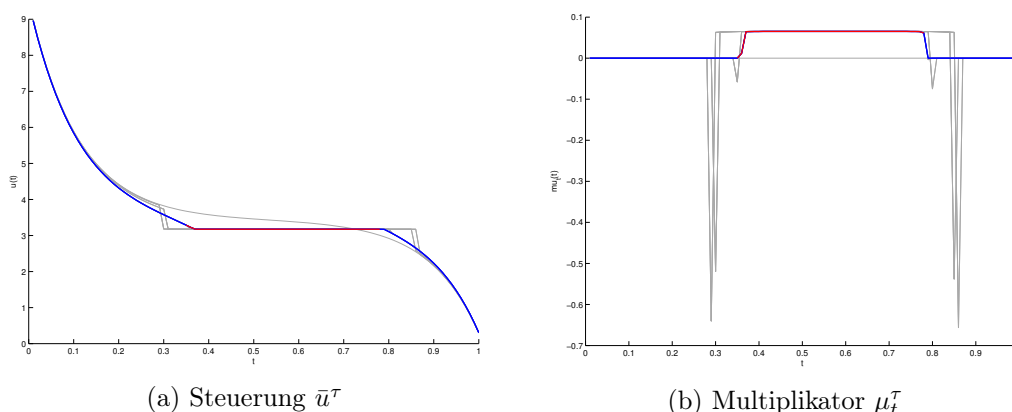


Abbildung 3.8.: Problem 1: Steuerung  $\bar{u}^\tau$  und Multiplikator  $\mu_t^\tau$  nach 6 Iterationen des PDASSPbasic für  $n = 100$ ; blau/rot: Lösungen aus aktuellem Schritt  $k$  auf aktiver (rot) und inaktiver Menge (blau), grau: vorangegangene Lösungen

**Daten :** Matrizen und Vektoren  $\mathbf{H}$ ,  $\vec{f}$ ,  $\mathbf{A}$ ,  $\vec{b}$  des Problems (3.7), maximale Iterationsanzahl  $k_{\max}$

**Ergebnis :** Lösung  $\vec{u}$  von (3.7), Multiplikator  $\vec{\eta}$ , Iterationsanzahl  $k$

1 Setze  $\mathcal{A}_1^d = \emptyset$ ,  $\vec{\eta}_1 = \vec{0}$  und  $k = 1$ ;

2 Bestimme  $\vec{u}_1 \in \mathbb{R}^n$  als Lösung von  $\mathbf{H}\vec{u}_1 = -\vec{f}$ ;

3 **solange**  $k \leq k_{\max}$  **tue**

4     Setze  $k = k + 1$  ;

5     Bestimme neue aktive und inaktive Menge

$$\mathcal{A}_k^d = \{i \mid \vec{\eta}_k + \mathbf{A}\vec{u}_k - \vec{b} > \vec{0}\}, \quad \mathcal{I}_k^d = \{i \mid \vec{\eta}_k + \mathbf{A}\vec{u}_k - \vec{b} \leq \vec{0}\};$$

6     **wenn**  $\mathcal{A}_k^d = \mathcal{A}_{k-1}^d$  **dann**

7         Setze  $k = k - 1$  ;

8         STOP;

9     **Ende**

10    **wenn**  $k \geq 4$  **dann**

11        Bestimme  $n_k^{\text{PDAS}} = |\mathcal{A}_k^d|$ ;

12        **wenn**  $n_k^{\text{PDAS}} \geq 4$  **dann**

13            Bestimme  $i_k^{\text{en}}$  und  $i_k^{\text{ex}}$ ;

14            **wenn**  $i_k^{\text{en}} = i_{k-1}^{\text{en}} + 1$  **und**  $i_{k-1}^{\text{en}} = i_{k-2}^{\text{en}} + 1$  **dann**

15                Setze  $\Delta t_k^{\text{en}} = -\frac{\tau}{[\vec{u}_{k-1}(i_{k-1}^{\text{en}})] - [\vec{u}_{k-2}(i_{k-2}^{\text{en}})]} \cdot [\vec{u}_{k-1}(i_{k-1}^{\text{en}})]$ ;

16                Berechne  $\Delta i_k^{\text{en}} = r \left( \frac{\Delta t_k^{\text{en}}}{\tau} \right)$ ;

17                Korrigiere  $\Delta i_k^{\text{en}} = \max\{1, \min\{\Delta i_k^{\text{en}}, \lfloor \frac{n_k^{\text{PDAS}}}{2} \rfloor\}\}$ ;

18                Passe  $\mathcal{A}_k^d$  an, so dass  $i_{k,\text{neu}}^{\text{en}} = i_{k-1}^{\text{en}} + \Delta i_k^{\text{en}}$ ;

19            **Ende**

20            **wenn**  $i_k^{\text{ex}} = i_{k-1}^{\text{ex}} - 1$  **und**  $i_{k-1}^{\text{ex}} = i_{k-2}^{\text{ex}} - 1$  **dann**

21                Setze  $\Delta t_k^{\text{ex}} = -\frac{-\tau}{[\vec{u}_{k-1}(i_{k-1}^{\text{ex}})] - [\vec{u}_{k-2}(i_{k-2}^{\text{ex}})]} \cdot [\vec{u}_{k-1}(i_{k-1}^{\text{ex}})]$ ;

22                Berechne  $\Delta i_k^{\text{ex}} = r \left( \frac{\Delta t_k^{\text{ex}}}{\tau} \right)$ ;

23                Korrigiere  $\Delta i_k^{\text{ex}} = \min\{-1, \max\{\Delta i_k^{\text{ex}}, -\lfloor \frac{n_k^{\text{PDAS}}}{2} \rfloor\}\}$ ;

24                Passe  $\mathcal{A}_k^d$  an, so dass  $i_{k,\text{neu}}^{\text{ex}} = i_{k-1}^{\text{ex}} + \Delta i_k^{\text{ex}}$ ;

25            **Ende**

26        **Ende**

27    **Ende**

28    Bestimme  $\vec{u}_k \in \mathbb{R}^n$ ,  $\vec{\eta}_k \in \mathbb{R}^n$  als Lösung von

$$\begin{aligned} \mathbf{H}\vec{u}_k + \mathbf{A}^\top \vec{\eta}_k &= -\vec{f}, \\ \mathbf{A}\vec{u}_k &= \vec{b} \quad \text{in } \mathcal{A}_k^d, \\ \vec{\eta}_k &= \vec{0} \quad \text{in } \mathcal{I}_k^d; \end{aligned}$$

29 **Ende**

**Algorithmus 2 :** PDASSPbasic

### 3.1.4. Verfeinerter PDASSP Algorithmus

Um den PDASSPbasic Algorithmus 2 für allgemeine Probleme aus Kapitel 2.1 anwenden zu können, müssen noch diverse Erweiterungen vorgenommen werden.

Der PDASSP Algorithmus wurde bisher nur für Beispiele mit aktiven Mengen  $\mathcal{A}^d$  bestehend aus einer Zusammenhangskomponente formuliert. Unter einer Zusammenhangskomponente verstehe man in diesem Fall maximale Teilmengen von  $\mathcal{A}^d$ , die aus aufeinanderfolgenden Indizes bestehen. Beispielsweise bestünde die aktive Menge  $\mathcal{A}^d = \{4, 5, 6, 7, 9, 10, 13, 15, 16, 17\}$  aus den 4 Zusammenhangskomponenten

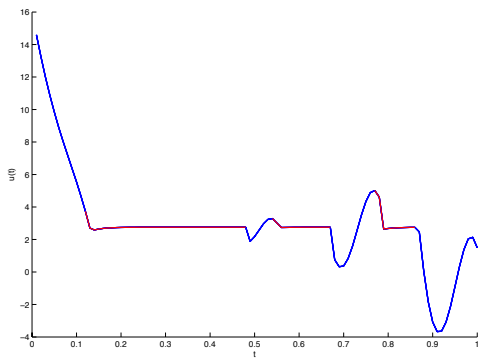
$$\mathcal{A}^{d,1} = \{4, 5, 6, 7\}, \quad \mathcal{A}^{d,2} = \{9, 10\}, \quad \mathcal{A}^{d,3} = \{13\}, \quad \mathcal{A}^{d,4} = \{15, 16, 17\}.$$

In Anlehnung an die Schreibweise in diesem Beispiel seien in der Folge  $m_k$  Zusammenhangskomponenten in der  $k$ -ten Iteration eines PDASSP Algorithmus notiert als  $\mathcal{A}_k^{d,1}, \dots, \mathcal{A}_k^{d,m_k}$ . Das kleinste Element in den Zusammenhangskomponenten sei jeweils als  $i_k^{\text{en},1}, \dots, i_k^{\text{en},m_k}$  bezeichnet, die größten Elemente als  $i_k^{\text{ex},1}, \dots, i_k^{\text{ex},m_k}$ . Um im Algorithmus 2 Fälle mit mehreren Zusammenhangskomponenten abbilden zu können, müssen nach dem Stopkriterium in der  $k$ -ten Iteration des Algorithmus die  $m_k$  Zusammenhangskomponenten der aktuellen Schätzung der aktiven Menge sowie die zugehörigen Indizes  $i_k^{\text{en},l}, i_k^{\text{ex},l}$  und die Anzahlen  $n_k^{l,\text{PDAS}} = |\mathcal{A}_k^{d,l}|$  für  $l = 1, \dots, m_k$  bestimmt werden. Im Anschluss führt man die Ideen zur Einbindung einer Schaltpunktsuche (Schritte 12 bis 26) in einer Schleife über die  $m_k$  Zusammenhangskomponenten aus. Dabei sei darauf hingewiesen, dass sich während der verschiedenen Iterationen des PDASSP Algorithmus die Anzahl der Zusammenhangskomponenten ändern kann. In Folge darf man beispielsweise nicht wie in Schritt 9 des Algorithmus 2 eine Abfrage der Form  $i_k^{\text{en},l} = i_{k-1}^{\text{en},l} + 1$  und  $i_{k-1}^{\text{en},l} = i_{k-2}^{\text{en},l} + 1$  vornehmen. Vielmehr muss für  $i_k^{\text{en},l} \in \mathcal{A}_k^{\text{en}} := \{i_k^{\text{en},1}, \dots, i_k^{\text{en},m_k}\}$  und  $i_k^{\text{ex},l} \in \mathcal{A}_k^{\text{ex}} := \{i_k^{\text{ex},1}, \dots, i_k^{\text{ex},m_k}\}$  die Abfrage lauten

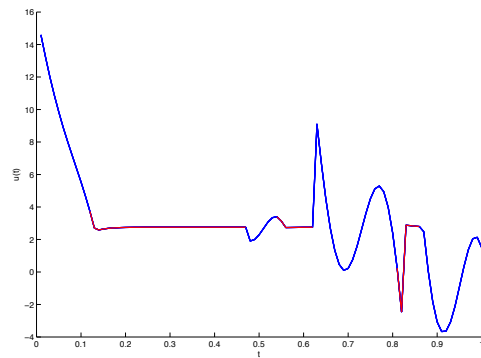
$$\begin{aligned} i_k^{\text{en},l} - 1 &\in \mathcal{A}_{k-1}^{\text{en}}, & i_k^{\text{en},l} - 2 &\in \mathcal{A}_{k-2}^{\text{en}}, \\ i_k^{\text{ex},l} + 1 &\in \mathcal{A}_{k-1}^{\text{ex}}, & i_k^{\text{ex},l} + 2 &\in \mathcal{A}_{k-2}^{\text{ex}}. \end{aligned}$$

Ein großes Problem des PDASSPbasic Algorithmus ist, dass eine Konvergenz bisher nicht gesichert ist, obwohl die Annahme 3.1.3 getroffen wurde. Nimmt man eine Verkleinerung der aktiven Menge auch mit dem Vorschlag aus einer Schaltpunktsuche vor, so könnten Schleifen auftreten. Als Veranschaulichung sei die Sequenz an Steuerungen in den verschiedenen Iterationen des PDASSPbasic für das Problem 2 aus Kapitel 3.2.2 mit mehreren aktiven Mengen betrachtet (Abbildung 3.9). In Iteration 20 hat man bereits eine gute Schätzung für die Topologie der aktiven Menge. Die Stetigkeit der Steuerung ist aber noch nicht gegeben. Die Vorschläge aus der Schaltpunktsuche verschieben in Iteration 21 und Iteration 22 die Absprungpunkte teilweise so stark, dass im darauffolgenden Schritt 23 mit der Vergrößerungsstrategie der PDAS die ersten beiden Zusammenhangskomponenten wieder zu einer großen Zusammenhangskomponente vereint werden. In der nächsten Iteration wird mit der PDAS Strategie die richtige Topologie wieder erkannt. Dieses Verhalten wiederholt sich.

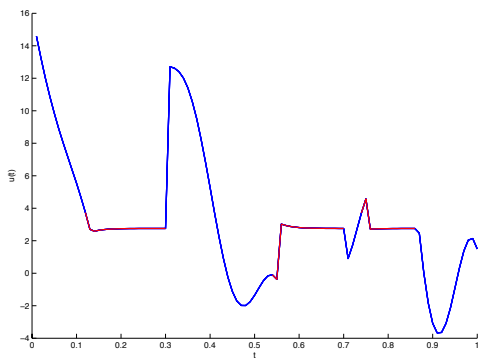
### 3. Algorithmen und numerische Versuche



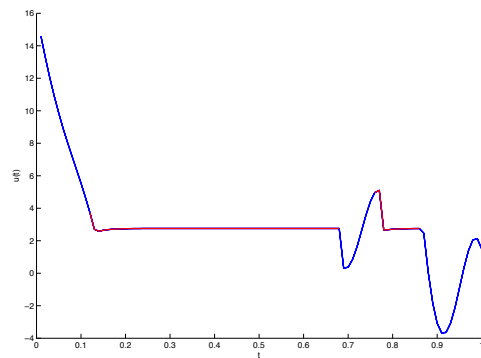
(a)  $u_{20}^T$  in der Ausgangssituation



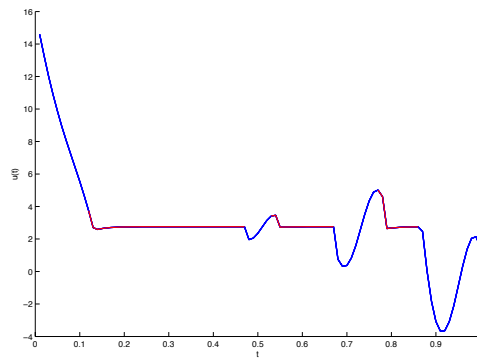
(b)  $u_{21}^T$  nach teilweiser Verringerung der aktiven Menge mit Sekantenverfahren



(c)  $u_{22}^T$  nach teilweiser Verringerung der aktiven Menge mit Sekantenverfahren



(d)  $u_{23}^T$  nach Veränderung der aktiven Menge gemäß PDAS



(e)  $u_{24}^T$  nach Veränderung der aktiven Menge gemäß PDAS

Abbildung 3.9.: Problem 2: Steuerung  $u_k^T$  nach  $k$  Iterationen der PDAS; Lösungen auf aktiver (rot) und inaktiver Menge (blau)

Es muss deshalb sichergestellt werden, dass im Zweifelsfall der PDASSP Algorithmus mit zunehmender Anzahl an Iterationen auf den PDAS Algorithmus umschwenkt, um von dessen Konvergenz zu profitieren. Dabei sollte dies so geschehen, dass wennmöglich eine gitterunabhängige Konvergenz nicht beeinträchtigt wird. Im Rahmen dieser Arbeit

wurde das folgende Vorgehen gewählt: Die maximal mögliche Schrittweite für die  $l$ -te Zusammenhangskomponente

$$\left\lfloor \frac{n_k^{l,\text{PDAS}}}{2} \right\rfloor$$

soll mit zunehmender Iteration reduziert werden, so dass ab einer gewissen Iteration in Schritt 17 und Schritt 23 von Algorithmus 2 die Schrittweite 1 gewählt wird. Der Algorithmus würde in Folge analog zu einem PDAS Algorithmus verlaufen und gemäß Annahme 3.1.3 konvergieren. Man setze deshalb in der  $k$ -ten Iteration als maximal mögliche Schrittweite für die  $l$ -te Zusammenhangskomponente

$$\left\lfloor n_k^{l,\text{PDAS}} \cdot \left(\frac{1}{2}\right)^{\sqrt{1+\frac{k-4}{10}}} \right\rfloor. \quad (3.9)$$

Dabei beachte man, dass eine Korrektur mit dem Sekantenverfahren frühestens in der Iteration  $k = 4$  vorgenommen werden kann. Die Anzahl der nötigen Iterationen des um diese Korrektur erweiterten PDASSPbasic Algorithmus im Vergleich zu einer PDAS Strategie sind in Tabelle 3.4 aufgeführt.

Anzahl an Zeitschritten $n$	100	200	400	800
PDASSPbasic erweitert	63	109	133	128
PDAS	8	14	27	53

Tabelle 3.4.: Anzahl an Iterationen für Problem 2

Die soeben vorgeschlagene Strategie führt zwar zur Konvergenz des Verfahrens unter der Annahme 3.1.3, es stellt sich bei Betrachtung der Abbildung 3.9 jedoch die Frage, ob es tatsächlich sinnvoll ist, den Vorschlag nach einem Schritt des Sekantenverfahrens ohne zusätzliche Bedingung bzw. Korrektur zu akzeptieren. Offensichtlich ist der Sprung der Steuerung an einigen Stellen nach Durchführung eines Sekantenschritts in Iteration 21 und 22 im Vergleich zur Iteration vorher deutlich größer. Außerdem hat der Sprung das Vorzeichen gewechselt. Im Rahmen eines Sekantenverfahrens würde man deshalb (exemplarisch für Aufsprungpunkte ausgeführt) mit einer Linesearch die Verschiebung  $\Delta i_k^{\text{en},l}$  ausgehend von dem Aufsprungpunkt vor der Durchführung des Sekantenschritts ( $=: i_{k,\text{alt}}^{\text{en},l}$ ) so lange anpassen, bis für den neuen Aufsprungpunkt  $i_{k,\text{neu}}^{\text{en},l} = i_{k,\text{alt}}^{\text{en},l} + \Delta i_k^{\text{en},l}$  die Bedingung

$$||\vec{u}_k(i_{k,\text{neu}}^{\text{en},l})|| < ||\vec{u}_{k-1}(i_{k,\text{alt}}^{\text{en},l})|| \quad (3.10)$$

eingehalten ist. An dieser Stelle sei darauf hingewiesen, dass wegen der möglichen Veränderung der Anzahl an Zusammenhangskomponenten von Schritt  $k-1$  zu Schritt  $k$  die Beziehung  $i_{k,\text{alt}}^{\text{en},l} = i_{k-1}^{\text{en},l}$  nicht eingehalten sein muss. Insgesamt erscheint im Rahmen des Ansatzes hier aber eine Linesearch als zu aufwendig. Folgende Variationen sind stattdessen angedacht worden, um den Vorschlag aus einem Sekantenschritt des PDASSPbasic Algorithmus gegebenenfalls zu korrigieren:

1. Verwende für den jeweiligen Auf-/Absprungpunkt den Vorschlag aus der Schaltpunktsuche nur, wenn sich dabei das Vorzeichen des Sprunges der Steuerung nicht ändert. Führe ansonsten die Veränderung um einen Index gemäß PDAS Algorithmus aus. Dahinter steckt die Idee, dass die Schaltpunktsuche nur zur beschleunigten Verringerung der aktiven Menge führen soll. Ist der Schritt so groß, dass ein Vorzeichensprung auftritt, würde im nächsten Schritt die aktive Menge mit dem Kriterium aus der PDAS Strategie wieder vergrößert. Dies soll mit diesem Ansatz vermieden werden.
2. Falls die Bedingung (3.10) verletzt ist, streiche den Vorschlag aus dem Sekantenverfahren für den jeweiligen Auf-/Absprungpunkt und verringere stattdessen dort die aktive Menge um einen Index gemäß Vorschlag aus PDAS.

Die bisherigen Vorschläge beziehen in die Berechnung eines neuen Vorschlags nicht den bereits berechneten Wert des Sprunges an der Stelle, die korrigiert werden soll, ein. Insofern bleibt Information ungenutzt. Dieses Manko soll mit den folgenden beiden Vorschlägen behoben werden. Dahinter steckt die Idee, dass, falls der Sprung nach einem Schritt aus dem Sekantenverfahren sein Vorzeichen wechselt, die „Nullstelle“ zwischen dem neuen und dem alten Vorschlag liegen muss.

3. Falls nach einem Sekantenschritt beispielsweise für einen Aufsprungpunkt ein Vorzeichenwechsel stattfindet, also die Bedingung  $[\vec{u}_k(i_{k,\text{neu}}^{\text{en},l})] \cdot [\vec{u}_{k-1}(i_{k,\text{alt}}^{\text{en},l})] < 0$  gilt, so führe einen weiteren Schritt eines Sekantenverfahrens von der Iteration  $k$  aus durch. Runde die erhaltene neue Schätzung für den Schaltpunkt auf den nächsten Zeitpunkt im Zeitgitter.
4. Falls nach einem Sekantenschritt beispielsweise für einen Aufsprungpunkt die Bedingung  $[\vec{u}_k(i_{k,\text{neu}}^{\text{en},l})] \cdot [\vec{u}_{k-1}(i_{k,\text{alt}}^{\text{en},l})] < 0$  erfüllt ist, so verwende als neuen Vorschlag für einen Schaltpunkt die Nullstelle der quadratischen Approximation durch die Datenwerte  $[\vec{u}_k(i_{k,\text{neu}}^{\text{en},l})]$ ,  $[\vec{u}_{k-1}(i_{k,\text{alt}}^{\text{en},l})]$ ,  $[\vec{u}_{k-2}(i_{k,\text{alt}}^{\text{en},l} - 1)]$ .

Die nötigen Iterationen für die verschiedenen Versionen getestet am Problem 2 aus Kapitel 3.2.2 sind in der Tabelle 3.5 aufgeführt. Für den Vergleich wurde jeweils die Anzahl der Lösungen des Optimalitätssystems gezählt, um verworfene Vorschläge auch zu berücksichtigen.

Korrekturvorschlag	1	2	3	4
PDASSPbasic bei $n = 100$	8	11	7	7
PDASSPbasic bei $n = 200$	15	12	8	11
PDASSPbasic bei $n = 400$	26	22	10	12
PDASSPbasic bei $n = 800$	52	33	13	14

Tabelle 3.5.: Anzahl an Lösungen des Optimalitätssystems für Problem 2 mit  $n$  Zeitschritten

Variante 1 ist eine sehr konservative Strategie. Entsprechend hoch sind auch die Anzahl an Lösungen des Gleichungssystems. Der Algorithmus verhält sich nahezu wie eine PDAS Strategie (vgl. Iterationsanzahl in Tabelle 3.4). In Variante 2 werden die Vorschläge aus einer Sekantenstrategie häufiger akzeptiert. Dies wirkt sich positiv aus. Die Variante 3 mit einem weiteren Sekantenschritt liefert ein nahezu gitterunabhängiges Konvergenzverhalten bei zusätzlich niedriger Anzahl an Iterationen. Der Mehraufwand einer quadratischen Approximation in Variante 4 lohnt sich nicht. Da sich an anderen Testbeispielen ein ähnliches Verhalten zeigte, wurde für den abschließenden PDASSP Algorithmus ein Korrekturschritt gemäß Variante 3 gewählt.

Die Gesamtidee des PDASSP mit Schaltpunktsuche und Korrekturschritt ist zusammenfassend in der Skizze 3.10 für den Fall einer Zusammenhangskomponente an einem Aufsprungpunkt dargestellt.

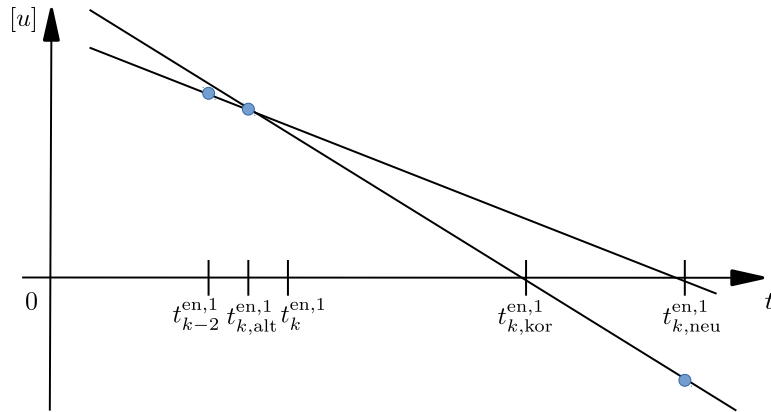


Abbildung 3.10.: Prinzipskizze zur Bestimmung eines neuen Aufsprungpunktes mit Sekantenverfahren und Korrekturschritt

Die Situation sei so gewählt, dass in der Iteration  $k$  ein Schritt eines Sekantenverfahrens ausgeführt wurde, der den neuen Aufsprungpunkt  $i_{k,neu}^{en,1}$  zur Folge hatte. Der nicht verwendete Vorschlag in Schritt  $k$  aus der PDAS Strategie ist durch  $i_k^{en,1}$  gegeben. Dieser hängt mit dem Aufsprungpunkt aus der Iteration  $k-1$  über die Beziehung  $i_{k,alt}^{en,1} = i_{k-1}^{en,1} = i_k^{en,1} - 1$  zusammen. Es ergibt sich  $i_{k,neu}^{en,1} = i_{k,alt}^{en,1} + \Delta i_k^{en,1} = i_k^{en,1} - 1 + \Delta i_k^{en,1}$ . Außerdem muss für den Aufsprungpunkt zur Iteration  $k-2$  gelten  $i_{k-2}^{en,1} = i_k^{en,1} - 2$ . Der aus dem weiteren Sekantenschritt erhaltene korrigierte Index sei als  $i_{k,kor}^{en,1}$  bezeichnet. Mit den zu den Indizes zugehörigen Zeitpunkten ergibt sich die Prinzipskizze 3.10.

Der finale PDASSP Algorithmus ist als Algorithmus 3 aufgeführt. Er beinhaltet eine Schaltpunktsuche (Algorithmus 4) und einen Korrekturschritt (Algorithmus 5).

**Daten :** Matrizen und Vektoren  $\mathbf{H}$ ,  $\vec{f}$ ,  $\mathbf{A}$ ,  $\vec{b}$  des Problems (3.7), maximale Iterationsanzahl  $k_{\max}$

**Ergebnis :** Lösung  $\vec{u}$  von (3.7), Multiplikator  $\vec{\eta}$ , Iterationsanzahl  $k$

- 1 Setze  $\mathcal{A}_1^d = \emptyset$ ,  $\vec{\eta}_1 = \vec{0}$  und  $k = 1$ ;
- 2 Bestimme  $\vec{u}_1 \in \mathbb{R}^n$  als Lösung von  $\mathbf{H}\vec{u}_1 = -\vec{f}$ ;
- 3 **solange**  $k \leq k_{\max}$  **tue**
- 4     Setze  $k = k + 1$  ;
- 5     Bestimme neue aktive und inaktive Menge
 
$$\mathcal{A}_k^d = \{i \mid \vec{\eta}_k + \mathbf{A}\vec{u}_k - \vec{b} > \vec{0}\}, \quad \mathcal{I}_k^d = \{i \mid \vec{\eta}_k + \mathbf{A}\vec{u}_k - \vec{b} \leq \vec{0}\};$$
- 6     **wenn**  $\mathcal{A}_k^d = \mathcal{A}_{k-1}^d$  **dann**
- 7         Setze  $k = k - 1$  ;
- 8         **STOP**;
- 9     **Ende**
- 10     Bestimme alle  $m_k$  Zusammenhangskomponenten  $\mathcal{A}_k^{d,l}$  von  $\mathcal{A}_k^d$  sowie  $\mathcal{A}_k^{\text{en}}, \mathcal{A}_k^{\text{ex}}$ ;
- 11     **wenn**  $k \geq 4$  **dann**
- 12         **Schaltpunktsuche:** Überprüfe für jede Zusammenhangskomponente, ob Vorschlag für  $\mathcal{A}_k^d$  mit einem Schritt eines Sekantenverfahrens verändert werden kann; führe diesen Schritt aus und passe  $\mathcal{A}_k^d$  und  $\mathcal{I}_k^d$  an: Algorithmus 4
- 13     **Ende**
- 14     Bestimme  $\vec{u}_k \in \mathbb{R}^n$ ,  $\vec{\eta}_k \in \mathbb{R}^n$  als Lösung von
 
$$\begin{aligned} \mathbf{H}\vec{u}_k + \mathbf{A}^\top \vec{\eta}_k &= -\vec{f}, \\ \mathbf{A}\vec{u}_k &= \vec{b} \quad \text{in } \mathcal{A}_k^d, \\ \vec{\eta}_k &= \vec{0} \quad \text{in } \mathcal{I}_k^d; \end{aligned}$$
- 15     **wenn**  $k \geq 4$  **und**  $\mathcal{A}_k^d$  **in** *Schaltpunktsuche* **verändert** **dann**
- 16         **Korrekturschritt:** Führe für jede Zusammenhangskomponente gegebenenfalls weiteren Schritt eines Sekantenverfahrens aus und bestimme neue Schätzung der aktiven Menge  $\tilde{\mathcal{A}}^d$ : Algorithmus 5
- 17         **wenn**  $\tilde{\mathcal{A}}^d \neq \mathcal{A}_k^d$  **dann**
- 18             Setze  $k = k + 1$ ;
- 19             Setze  $\mathcal{A}_k^d = \tilde{\mathcal{A}}^d$  und bestimme  $\mathcal{I}_k^d$ ;
- 20             Bestimme  $\vec{u}_k \in \mathbb{R}^n$ ,  $\vec{\eta}_k \in \mathbb{R}^n$  als Lösung von
 
$$\begin{aligned} \mathbf{H}\vec{u}_k + \mathbf{A}^\top \vec{\eta}_k &= -\vec{f}, \\ \mathbf{A}\vec{u}_k &= \vec{b} \quad \text{in } \mathcal{A}_k^d, \\ \vec{\eta}_k &= \vec{0} \quad \text{in } \mathcal{I}_k^d; \end{aligned}$$
- 21         **Ende**
- 22     **Ende**
- 23 **Ende**

Algorithmus 3 : PDASSP



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17	<p><b>Daten :</b> Aktive Menge <math>\mathcal{A}_k^d</math> aus PDAS mit <math>m_k</math> Zusammenhangskomponenten <math>\mathcal{A}_k^{d,l}</math> sowie Auf-/Absprungpunkte <math>\mathcal{A}_k^{\text{en}}, \mathcal{A}_k^{\text{ex}}</math>, alte Auf-/Absprungpunkte <math>\mathcal{A}_{k-1}^{\text{en}}, \mathcal{A}_{k-1}^{\text{ex}}, \mathcal{A}_{k-2}^{\text{en}}, \mathcal{A}_{k-2}^{\text{ex}}</math>, alte Steuerungsvektoren <math>\vec{u}_{k-1}, \vec{u}_{k-2}</math>, Iterationszähler <math>k</math></p> <p><b>Ergebnis :</b> Neue Schätzung für <math>\mathcal{A}_k^d</math> und <math>\mathcal{I}_k^d</math> mit gegebenenfalls neuen Auf-/Absprungpunkten <math>i_{k,\text{neu}}^{\text{en},l}, i_{k,\text{neu}}^{\text{ex},l}</math>, Verschiebungen <math>\Delta i_k^{\text{en},l}, \Delta i_k^{\text{ex},l}</math></p> <p><b>1</b> für <math>l = 1</math> bis <math>m_k</math> <b>tue</b></p> <p><b>2</b> Bestimme <math>n_k^{l,\text{PDAS}} =  \mathcal{A}_k^{d,l} </math> und setze <math>\Delta i_k^{\text{en},l} = \Delta i_k^{\text{ex},l} = 0</math> ;</p> <p><b>3</b> wenn <math>n_k^{l,\text{PDAS}} \geq 4</math> <b>dann</b></p> <p style="padding-left: 20px;"><b>4</b> wenn <math>i_k^{\text{en},l} - 1 \in \mathcal{A}_{k-1}^{\text{en}}</math> und <math>i_k^{\text{en},l} - 2 \in \mathcal{A}_{k-2}^{\text{en}}</math> <b>dann</b></p> <p style="padding-left: 40px;"><b>5</b> Setze <math>\Delta t_k^{\text{en},l} = -\frac{\tau}{[\vec{u}_{k-1}(i_k^{\text{en},l}-1)] - [\vec{u}_{k-2}(i_k^{\text{en},l}-2)]} \cdot [\vec{u}_{k-1}(i_k^{\text{en},l}-1)]</math>;</p> <p style="padding-left: 40px;"><b>6</b> Berechne <math>\Delta i_k^{\text{en},l} = r \left( \frac{\Delta t_k^{\text{en},l}}{\tau} \right)</math>;</p> <p style="padding-left: 40px;"><b>7</b> Korrigiere <math>\Delta i_k^{\text{en},l} = \max\{1, \min\{\Delta i_k^{\text{en},l}, \left\lfloor n_k^{l,\text{PDAS}} \cdot \left(\frac{1}{2}\right)^{\sqrt{1+\frac{k-4}{10}}} \right\rfloor\}\}</math>;</p> <p style="padding-left: 40px;"><b>8</b> Passe <math>\mathcal{A}_k^{d,l}</math> und <math>\mathcal{I}_k^{d,l}</math> an, so dass <math>i_{k,\text{neu}}^{\text{en},l} = i_k^{\text{en},l} - 1 + \Delta i_k^{\text{en},l}</math>;</p> <p style="padding-left: 20px;"><b>9</b> <b>Ende</b></p> <p style="padding-left: 20px;"><b>10</b> wenn <math>i_k^{\text{ex},l} + 1 \in \mathcal{A}_{k-1}^{\text{ex}}</math> und <math>i_k^{\text{ex},l} + 2 \in \mathcal{A}_{k-2}^{\text{ex}}</math> <b>dann</b></p> <p style="padding-left: 40px;"><b>11</b> Setze <math>\Delta t_k^{\text{ex},l} = -\frac{-\tau}{[\vec{u}_{k-1}(i_k^{\text{ex},l}+1)] - [\vec{u}_{k-2}(i_k^{\text{ex},l}+2)]} \cdot [\vec{u}_{k-1}(i_k^{\text{ex},l}+1)]</math>;</p> <p style="padding-left: 40px;"><b>12</b> Berechne <math>\Delta i_k^{\text{ex},l} = r \left( \frac{\Delta t_k^{\text{ex},l}}{\tau} \right)</math>;</p> <p style="padding-left: 40px;"><b>13</b> Korrigiere <math>\Delta i_k^{\text{ex},l} = \min\{-1, \max\{\Delta i_k^{\text{ex},l}, -\left\lfloor n_k^{l,\text{PDAS}} \cdot \left(\frac{1}{2}\right)^{\sqrt{1+\frac{k-4}{10}}} \right\rfloor\}\}</math>;</p> <p style="padding-left: 40px;"><b>14</b> Passe <math>\mathcal{A}_k^{d,l}</math> und <math>\mathcal{I}_k^{d,l}</math> an, so dass <math>i_{k,\text{neu}}^{\text{ex},l} = i_k^{\text{ex},l} + 1 + \Delta i_k^{\text{ex},l}</math>;</p> <p style="padding-left: 20px;"><b>15</b> <b>Ende</b></p> <p><b>16</b> <b>Ende</b></p> <p><b>17</b> <b>Ende</b></p>
---	---

Algorithmus 4 : Schaltpunktsuche

Eine Korrektur muss nur dann stattfinden, wenn in der Schaltpunktsuche eine Verschiebung um mehr als einen Index stattgefunden hat. Eine Verschiebung um einen Index entspricht genau dem Vorschlag aus dem PDAS Verfahren. Diese Vorschläge sollen nicht korrigiert werden.

<p><b>Daten :</b> Aus Schaltpunktsuche: Aktive Menge <math>\mathcal{A}_k^d</math> mit <math>m_k</math>  Zusammenhangskomponenten <math>\mathcal{A}_k^{d,l}</math>, veränderte Auf-/Absprungpunkte  aus Sekantenverfahren <math>i_{k,neu}^{en,l}</math>, <math>i_{k,neu}^{ex,l}</math>, Verschiebungen <math>\Delta i_k^{en,l}</math>, <math>\Delta i_k^{ex,l}</math></p> <p><b>Ergebnis :</b> Korrigierte Schätzung der aktiven Menge <math>\tilde{\mathcal{A}}^d</math></p> <pre> 1 Setze <math>\tilde{\mathcal{A}}^d = \mathcal{A}_k^d</math>; 2 für <math>l = 1</math> bis <math>m_k</math> tue 3   wenn <math>\Delta i_k^{en,l} &gt; 1</math> dann 4     Setze <math>i_{k,alt}^{en,l} = i_{k,neu}^{en,l} - \Delta i_k^{en,l}</math>; 5     wenn <math>[\vec{u}_k(i_{k,neu}^{en,l})] \cdot [\vec{u}_{k-1}(i_{k,alt}^{en,l})] &lt; 0</math> dann 6       Setze <math>\Delta t_{k,kor}^{en,l} = -\frac{\tau \cdot \Delta i_k^{en,l}}{[\vec{u}_k(i_{k,neu}^{en,l})] - [\vec{u}_{k-1}(i_{k,alt}^{en,l})]} \cdot [\vec{u}_k(i_{k,neu}^{en,l})]</math>; 7       Berechne <math>\Delta i_{k,kor}^{en,l} = r \left( \frac{\Delta t_{k,kor}^{en,l}}{\tau} \right)</math>; 8       Passe <math>\tilde{\mathcal{A}}^d</math> an, so dass <math>i_{k,kor}^{en,l} = i_{k,neu}^{en,l} + \Delta i_{k,kor}^{en,l}</math>; 9     Ende 10  Ende 11  wenn <math>\Delta i_k^{ex,l} &lt; -1</math> dann 12    Setze <math>i_{k,alt}^{ex,l} = i_{k,neu}^{ex,l} - \Delta i_k^{ex,l}</math>; 13    wenn <math>[\vec{u}_k(i_{k,neu}^{ex,l})] \cdot [\vec{u}_{k-1}(i_{k,alt}^{ex,l})] &lt; 0</math> dann 14      Setze <math>\Delta t_{k,kor}^{ex,l} = -\frac{\tau \cdot \Delta i_k^{ex,l}}{[\vec{u}_k(i_{k,neu}^{ex,l})] - [\vec{u}_{k-1}(i_{k,alt}^{ex,l})]} \cdot [\vec{u}_k(i_{k,neu}^{ex,l})]</math>; 15      Berechne <math>\Delta i_{k,kor}^{ex,l} = r \left( \frac{\Delta t_{k,kor}^{ex,l}}{\tau} \right)</math>; 16      Passe <math>\tilde{\mathcal{A}}^d</math> an, so dass <math>i_{k,kor}^{ex,l} = i_{k,neu}^{ex,l} + \Delta i_{k,kor}^{ex,l}</math>; 17    Ende 18  Ende 19 Ende </pre>
--

Algorithmus 5 : Korrektur Schaltpunktsuche

### 3.2. Numerische Versuche

In diesem Kapitel sollen einige Testbeispiele und die mit numerischen Berechnungen erhaltenen Resultate dargestellt werden. Die in dem Modellproblem 2.1 auftretenden Daten werden unter verschiedenen Gesichtspunkten gewählt. Hiermit soll zum einen die Leistungsfähigkeit des in Kapitel 3.1.4 eingeführten PDASSP Algorithmus getestet und bewertet werden. Zum anderen dienen die Beispiele dazu, die theoretischen Resultate aus Kapitel 2.1 numerisch nachzuvollziehen und die Schärfe der Aussagen einzuschätzen.

### 3.2.1. Implementierung

Bevor in dem nächsten Abschnitt auf die diversen Beispiele und Ergebnisse eingegangen wird, soll in diesem Abschnitt dargestellt werden, wie das Problem (3.7) aus Kapitel 3.1.2 aufgestellt und einige Details des PDASSP Algorithmus umgesetzt werden können. Dabei soll auch eine Erweiterung für zeitlich nicht (stückweise) konstante Datenfunktionen  $\delta$  und  $e$  aufgeführt werden.

Um die Matrizen und Vektoren  $\mathbf{H}$ ,  $\vec{f}$ ,  $\mathbf{A}$  und  $\vec{b}$  zu bestimmen, müssen die Funktionen  $y_{\text{init}}^\tau$  und  $y_i^\tau \in W_\tau^0$  für  $i = 1, \dots, n$  berechnet werden. Hierzu wird die Zustandsgleichung zunächst umformuliert. Testet man die Zustandsgleichung in der semidiskretisierten Form (3.3) mit Testfunktionen  $\varepsilon_j \cdot v \in W_\tau^0$ , wobei  $\varepsilon_j \in U_\tau^0$  die in Kapitel 3.1.2 eingeführte Indikatorfunktion für das Intervall  $I_j$  und  $v$  eine beliebige Funktion aus  $H^1(\Omega)$  seien, so muss  $y^\tau = \hat{S}^\tau(u^\tau) \in W_\tau^0$  die folgenden Bedingungen erfüllen:

- für  $j = 2, \dots, n$ :

$$\int_{I_j} \langle Ay^\tau, v \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt + \int_\Omega y^{\tau, j} v dx = \int_\Omega y^{\tau, j-1} v dx + \int_{I_j} u^\tau \int_\Omega ev dx dt, \quad (3.11a)$$

- für  $j = 1$ :

$$\int_{I_1} \langle Ay^\tau, v \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt + \int_\Omega y^{\tau, 1} v dx = \int_\Omega y_0 v dx + \int_{I_1} u^\tau \int_\Omega ev dx dt. \quad (3.11b)$$

Der jeweils erste Summand lässt sich wegen  $y^\tau \in W_\tau^0$  für  $j \in \{1, \dots, n\}$  umformulieren in

$$\int_{I_j} \langle Ay^\tau, v \rangle_{H^1(\Omega)^*, H^1(\Omega)} dt = \tau \langle Ay^{\tau, j}, v \rangle_{H^1(\Omega)^*, H^1(\Omega)}.$$

Approximiert man eine Funktion  $e \in L^\infty(0, T; L^2(\Omega))$  durch eine in der Zeit stückweise konstante Funktion  $e^\tau \in E_\tau^0$ , indem man  $e^\tau|_{I_j} = e(t_j)$  setzt, so lässt sich der jeweils letzte Summand für  $j \in \{1, \dots, n\}$  unter Verwendung von  $u^\tau = \sum_{j=1}^n u_j \cdot \varepsilon_j$  schreiben als

$$\int_{I_j} u^\tau \int_\Omega e^\tau v dx dt = \tau u_j \int_\Omega e(t_j) v dx.$$

Insgesamt ergibt sich ein implizites Euler-Verfahren zur Bestimmung von  $y^\tau \in W_\tau^0$ , wobei in jeder Iteration die Lösung  $y^{\tau, j}$  über ein Variationsproblem gegeben ist.

Um die einzelnen Variationsprobleme näherungsweise zu lösen, soll ein finite Elemente Ansatz verfolgt werden. Ausgehend von einem regulären Gitter für  $\bar{\Omega}$  mit Gitterweite  $h$  sei der Raum der stetigen, stückweise linearen Funktionen mit nodaler Basis  $\{\Phi_1, \dots, \Phi_m\}$  eingeführt. Eine Approximation von  $y^{\tau, j}$  sei gegeben durch  $y^{\tau, h, j} := \sum_{k=1}^m y_{j, k} \cdot \Phi_k$  mit  $y_{j, k} \in \mathbb{R}$ . Durch Testen der Gleichungen (3.11) mit den Basisfunktionen  $\Phi_k$  für  $k = 1, \dots, m$  ergibt sich mit der Steifigkeitsmatrix  $\mathbf{K} \in \mathbb{R}^{m \times m}$  mit Einträgen  $\mathbf{k}_{k, l} = \int_\Omega \nabla \Phi_k \nabla \Phi_l dx$ , der Massenmatrix  $\mathbf{M} \in \mathbb{R}^{m \times m}$  mit Einträgen  $\mathbf{m}_{k, l} = \int_\Omega \Phi_k \Phi_l dx$ , der Randmassenmatrix  $\mathbf{Q} \in \mathbb{R}^{m \times m}$  mit  $\mathbf{q}_{k, l} = \int_\Gamma \Phi_k \Phi_l ds(x)$ , dem Vektor  $\vec{F}_j \in \mathbb{R}^m$  mit  $f_{j, k} = \int_\Omega e(t_j) \Phi_k dx$  und  $\vec{y}_0 \in \mathbb{R}^m$  mit  $y_{0k} = \int_\Omega y_0 \Phi_k dx$  das implizite Euler-Schema

- für  $j = 2, \dots, n$ :

$$(\tau(\mathbf{K} + \mathbf{Q}) + \mathbf{M}) \vec{y}_j = \mathbf{M} \vec{y}_{j-1} + \tau u_j \vec{F}_j,$$

- für  $j = 1$ :

$$(\tau(\mathbf{K} + \mathbf{Q}) + \mathbf{M}) \vec{y}_1 = \vec{y}_0 + \tau u_1 \vec{F}_1.$$

Für die Bestimmung der im Ort diskretisierten Approximationen von  $y_{\text{init}}^\tau$  und  $y_i^\tau = S^\tau(\varepsilon_i)$  muss jeweils das obige Schema angewandt werden und für jeden Zeitschritt  $j = 1, \dots, n$  die Koeffizientenvektoren  $\vec{y}_{\text{init},j} \in \mathbb{R}^m$  und  $\vec{y}_{i,j} \in \mathbb{R}^m$  bestimmt werden. Dies kann simultan erfolgen. Setzt man

$$\mathbf{Y}_0 := (\vec{y}_0, \vec{0}, \dots, \vec{0}) \in \mathbb{R}^{m \times (n+1)}$$

und für  $j \in \{1, \dots, n\}$

$$\mathbf{Y}_j := (\vec{y}_{\text{init},j}, \vec{y}_{1,j}, \dots, \vec{y}_{n,j}) \in \mathbb{R}^{m \times (n+1)}, \quad \mathbf{F}_j := (\vec{0}, \dots, \underbrace{\vec{F}_j}_{\text{Spalte } j+1}, \vec{0}) \in \mathbb{R}^{m \times (n+1)},$$

so löst man

- für  $j = 2, \dots, n$ :

$$(\tau(\mathbf{K} + \mathbf{Q}) + \mathbf{M}) \mathbf{Y}_j = \mathbf{M} \mathbf{Y}_{j-1} + \tau u_j \mathbf{F}_j,$$

- für  $j = 1$ :

$$(\tau(\mathbf{K} + \mathbf{Q}) + \mathbf{M}) \mathbf{Y}_1 = \mathbf{Y}_0 + \tau u_1 \mathbf{F}_1.$$

Dabei gilt zu beachten, dass im Zeitschritt  $j$  nur die Spalten 1 bis  $j + 1$  der Matrix  $\mathbf{Y}_j$  berechnet werden müssen, da die Funktionen  $y_r^\tau$  mit  $r > j$  für Zeiten  $t \leq t_j$  ohnehin noch den Wert 0 besitzen. Die nötige Assemblierung der Matrizen und Vektoren  $\mathbf{K}$ ,  $\mathbf{Q}$ ,  $\mathbf{M}$ ,  $\vec{y}_0$  und  $\vec{F}_j$  erfolgt im Rahmen der Arbeit mit der Software Comsol Multiphysics 3.5.

Mit diesen Vorarbeiten sollen nun die Zustandsbeschränkung und das Zielfunktional für das diskretisierte Problem (3.7) aufgestellt werden. Um Beschränkungen mit zeitabhängigen Funktionen  $\delta$  behandeln zu können, sei für diese eine Approximation mit stückweise konstanten Funktionen durch  $\delta^\tau \in E_\tau^0$  durchgeführt, indem man  $\delta^\tau|_{I_j} = \delta(t_j)$  setzt. Die Zustandsbeschränkung kann somit geschrieben werden als

$$\int_{\Omega} y^{\tau,j} \delta(t_j) dx - y_{\max} \leq 0, \quad j = 1, \dots, n.$$

Die Matrix  $\mathbf{A}$  und der Vektor  $\vec{b}$  sind folglich für  $j \in \{1, \dots, n\}$  gegeben durch

$$\mathbf{a}_{j,\cdot} = \vec{\delta}_j^{\top} \cdot (\vec{y}_{1,j}, \dots, \vec{y}_{n,j}), \quad b_j = -\vec{\delta}_j^{\top} \cdot \vec{y}_{\text{init},j} + y_{\max}, \quad (3.12)$$

wobei der Vektor  $\vec{\delta}_j \in \mathbb{R}^m$  bestimmt wird als  $\delta_{jk} = \int_{\Omega} \delta(t_j) \Phi_k \, dx$ ,  $k = 1, \dots, m$ . Die Matrix  $\mathbf{H} \in \mathbb{R}^{n \times n}$  im Zielfunktional ist gegeben durch

$$\mathbf{h}_{i,j} = \iint_Q y_i^\tau y_j^\tau \, dx dt + \lambda \int_I \varepsilon_i \varepsilon_j \, dt.$$

Da es sich bei den Funktionen  $\varepsilon_i$  um die Indikatorfunktionen für das Intervall  $i$  handelt, hat man

$$\mathbf{d}_{i,j} := \int_I \varepsilon_i \varepsilon_j \, dt = \begin{cases} \tau & \text{für } i = j, \\ 0 & \text{für } i \neq j. \end{cases}$$

Verwendet man, dass die Funktionen  $y_i^\tau$  stückweise konstant in der Zeit sind, ergibt sich für den ersten Summanden mit der oben eingeführten Approximation im Ort und der zugehörigen Massenmatrix  $\mathbf{M}$

$$\iint_Q y_i^\tau y_j^\tau \, dx dt = \tau \sum_{l=1}^n \int_{\Omega} y_i^{\tau,l} y_j^{\tau,l} \, dx \approx \tau \sum_{l=1}^n \vec{y}_{i,l}^\top \mathbf{M} \vec{y}_{j,l}.$$

Alles in allem berechnet man die in Ort und Zeit diskretisierte Version der Matrix  $\mathbf{H}$  als

$$\mathbf{h}_{i,j} = \tau \sum_{l=1}^n \vec{y}_{i,l}^\top \mathbf{M} \vec{y}_{j,l} + \lambda \mathbf{d}_{i,j}.$$

Für den Vektor  $\vec{f} \in \mathbb{R}^n$  muss jeweils

$$f_i = \iint_Q y_i^\tau (y_{\text{init}}^\tau - y_d) \, dx dt = \iint_Q y_i^\tau y_{\text{init}}^\tau - y_i^\tau y_d \, dx dt$$

bestimmt werden. Für den ersten Summanden hat man mit der Ortsdiskretisierung und der zugehörigen Massenmatrix  $\mathbf{M}$

$$\iint_Q y_i^\tau y_{\text{init}}^\tau \, dx dt = \tau \sum_{l=1}^n \int_{\Omega} y_i^{\tau,l} y_{\text{init}}^{\tau,l} \, dx \approx \tau \sum_{l=1}^n \vec{y}_{i,l}^\top \mathbf{M} \vec{y}_{\text{init},l}.$$

Für die Berechnung des zweiten Summanden nähert man  $y_d$  durch eine stückweise konstante Funktion  $y_d^\tau \in E_\tau^0$  an, wobei man  $y_d^\tau|_{I_i} = y_d(t_i)$  setzt. Basierend auf der gewählten Ortsdiskretisierung ergibt sich mit der Definition von  $\vec{y}_{d,i} \in \mathbb{R}^m$  mit  $y_{d,i_k} = \int_{\Omega} y_d(t_i) \Phi_k \, dx$ ,  $k = 1, \dots, m$  die Näherung

$$\iint_Q y_i^\tau y_d \, dx dt = \tau \sum_{l=1}^n \int_{\Omega} y_i^{\tau,l} y_d \, dx \approx \tau \sum_{l=1}^n \vec{y}_{i,l}^\top \vec{y}_{d,l}.$$

Insgesamt:

$$f_i = \tau \sum_{l=1}^n \vec{y}_{i,l}^\top \mathbf{M} \vec{y}_{\text{init},l} - \tau \sum_{l=1}^n \vec{y}_{i,l}^\top \vec{y}_{d,l} = \tau \sum_{l=1}^n \vec{y}_{i,l}^\top (\mathbf{M} \vec{y}_{\text{init},l} - \vec{y}_{d,l}).$$

Um unnötige Schleifen zu vermeiden, baut man die Matrizen  $\mathbf{A}$ ,  $\mathbf{H}$  und die Vektoren  $\vec{b}$ ,  $\vec{f}$  am besten während der Iterationen des Euler-Schemas auf, wobei man beachtet, dass im Schritt  $l$  des Verfahrens nur die Funktionen  $y_i^t$  mit  $i \leq l$  Beiträge liefern. Das Vorgehen für  $\mathbf{A}$  und  $\vec{b}$  ergibt sich kanonisch aus (3.12). Für die Matrix  $\mathbf{H}$  und den Vektor  $\vec{f}$  im Zielfunktional geht man inkrementell vor. Ausgehend von  $\mathbf{H} = 0 \in \mathbb{R}^{n \times n}$  und  $\vec{f} = \vec{0} \in \mathbb{R}^n$  in einem vorbereitenden Schritt setzt man nach der Berechnung der Zustände im  $l$ -ten Schritt des Euler-Verfahrens

$$\begin{aligned} \mathbf{h}_{1\dots l,1\dots l} &= \mathbf{h}_{1\dots l,1\dots l} + \tau(\vec{y}_{1,l}, \dots, \vec{y}_{l,l})^\top \mathbf{M}(\vec{y}_{1,l}, \dots, \vec{y}_{l,l}), \\ \mathbf{h}_{i,l} &= \mathbf{h}_{i,l} + \lambda\tau \end{aligned}$$

und

$$f_{1\dots l} = f_{1\dots l} + \tau(\vec{y}_{1,l}, \dots, \vec{y}_{l,l})^\top (\mathbf{M}\vec{y}_{\text{mit},l} - \vec{y}_{d,l}).$$

Abschließend seien noch zwei Punkte zum PDASSP Algorithmus und dessen Ergebnissen aufgeführt. Die in den verschiedenen Schritten eines PDASSP Algorithmus zu lösenden Optimalitätssysteme wurden im Rahmen der Arbeit wie folgt behandelt. Falls die gegebene aktive Menge im  $k$ -ten Schritt  $\mathcal{A}_k^d$  nicht leer ist und aus  $n_k$  Elementen besteht, so sei  $\mathbf{I}_{\mathcal{A}_k^d}$  die Matrix der Dimension  $n_k \times n$ , die aus der Matrix  $\mathbf{A}$  die Zeilen auswählt, die zu der Menge der aktiven Indizes gehören. Das Gleichungssystem

$$\begin{bmatrix} \mathbf{H} & \mathbf{A}^\top \cdot \mathbf{I}_{\mathcal{A}_k^d}^\top \\ \mathbf{I}_{\mathcal{A}_k^d} \cdot \mathbf{A} & \end{bmatrix} \cdot \begin{pmatrix} \vec{u}_k \\ \vec{\eta}_k \end{pmatrix} = \begin{pmatrix} -\vec{f} \\ \mathbf{I}_{\mathcal{A}_k^d} \cdot \vec{b} \end{pmatrix}$$

wird mit `mldivide` gelöst. Der Vektor  $\vec{\eta}_k \in \mathbb{R}^{n_k}$  wird anschließend zu dem Lagrange-Multiplikator  $\vec{\eta}_k \in \mathbb{R}^n$  vervollständigt, indem man  $(\vec{\eta}_k)_i = 0$  setzt für  $i \in \mathcal{I}_k^d$ .

Hinsichtlich des Zusammenhangs des Lagrange-Multiplikators des diskretisierten Problems (3.7),  $\vec{\eta}$ , und der Approximation der Zeitableitung der Funktion  $\mu \in NBV_T(0, T)$ ,  $\mu_t^t$ , sei nochmals auf die nötige Skalierung mit  $\frac{1}{\tau}$  gemäß Bemerkung 3.1.1 in Kapitel 3.1.2 hingewiesen. In Bezug auf die Bezeichnung von  $\mu_t$  als Lagrange-Multiplikator beachte man Bemerkung 2.1.3 in Kapitel 2.1.1.

### 3.2.2. Modellprobleme und Resultate

Im Rahmen der numerischen Versuche werden die Iterationen des PDASSP Algorithmus, des PDAS Algorithmus und die in der Matlabroutine `quadprog` verwendete aktive-Mengen-Strategie nach Gill, Murray und Wright [25] miteinander verglichen. In diesem Kontext bezeichne man als Anzahl an Iterationen die Anzahl der Lösungen eines Optimalitätssystems. Die Schritte in der Zeitdiskretisierung der Probleme und damit die Anzahl der Freiheitsgrade variieren zwischen  $n = 100$  und  $n = 800$  Zeitschritte. Die Optimierungsalgorithmen PDAS (Algorithmus 1) und PDASSP (Algorithmus 3) werden mit der Startschätzung  $\mathcal{A}_1^d = \emptyset$  gestartet, die Optimierungsroutine `quadprog` mit  $u_0 = 0$ . Als maximal mögliche Anzahl an Iterationen wurde 500 festgelegt. Die Anzahl an Iterationen wird jeweils in Tabellenform aufgetragen. Beim Erreichen der Obergrenze gestoppte Testläufe werden mit 500\* markiert. Die Parameter im PDASSP Algorithmus sind für alle Probleme wie in Kapitel 3.1.4 beschrieben gewählt. Die aus dem PDASSP Algorithmus

erhaltenen optimalen Lösungen für  $n = 800$  Zeitschritte werden jeweils dargestellt. Falls keine Konvergenz innerhalb von 500 Iterationen eintritt, werden entweder die Schätzung für die optimalen Größen aus dem letzten Iterationsschritt des PDASSP Algorithmus angeführt oder gegebenenfalls die Lösung aus einem anderen, konvergierenden Verfahren. Auch wenn es sich eigentlich um stückweise konstante Funktionen handelt, so wurde aus Darstellungsgründen dennoch ein Linienplot verwendet. Die Werte der Funktionen  $u$  und  $g$  werden jeweils nur auf der aktiven bzw. inaktiven Menge mit einer durchgehenden Linie verbunden. Beim Übergang zwischen aktiver und inaktiver Menge wird auf die Verbindung verzichtet. Bei der Darstellung des Multiplikators  $\mu_t$  wird ein durchgehender Linienplot gewählt, um approximierte Diracmaße besser kenntlich zu machen. In Zeitpunkten, in denen die Zustandsbeschränkung aktiv ist, ist der Graph rot eingefärbt. Zu inaktiven Zeitpunkten blau. Aus Gründen der Übersichtlichkeit wird auf eine Indizierung der dargestellten stückweise konstanten Funktionen mit Zeitdiskretisierungsparameter  $\tau$  und Iterationsindex  $k$  (vgl. Kapitel 3.1.2–3.1.4) verzichtet.

Für alle Probleme sei  $\Omega = (0, 1) \times (0, 1)$  gewählt. Der Zeithorizont sei gegeben durch  $I = [0, 1]$ . Als Anfangswert sei  $y_0 \equiv 0$  gesetzt. Für die Approximation der Funktionen im Ort mit Lagrange-Elementen erster Ordnung sei allen Berechnungen ein Gitter mit 1031 Freiheitsgraden zugrunde gelegt.

**Problem 1, Standardproblem mit einer aktiven Menge** Für  $(x, y) \in \Omega$  und  $t \in I$  seien die Daten gegeben durch:

- $e(x, y, t) = \exp(-(x - 0.5)^2 - (y - 0.5)^2)$ ,
- $\delta(x, y, t) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{1}{2} \frac{(x-0.5)^2 + (y-0.5)^2}{\sigma^2}\right)$  mit  $\sigma = 0.1$ ,
- $y_d(x, y, t) = 1$ ,  $y_{\max}(t) = 0.9$  sowie  $\lambda = 0.01$ .

Dieses Beispiel wurde im Kapitel 3.1.3 bei der Motivation des PDASSP Algorithmus verwendet. Alle Datenfunktionen sind sehr regulär. Die aktive Menge besteht nur aus einem aktiven Intervall. Das Beispiel ist insofern sehr gutmütig, als dass im Rahmen des PDAS(SP) Algorithmus die korrekte Topologie bereits nach der ersten Iteration des Algorithmus gefunden wird und anschließend nur noch Verkleinerungen der aktiven Menge auftreten. An den Ergebnissen in Tabelle 3.12b wird ersichtlich, dass der PDASSP Algorithmus wie gewünscht keine Gitterabhängigkeit zeigt, während sich die Anzahl der Iterationen bei den anderen beiden Algorithmen ungefähr verdoppelt, wenn die Anzahl der Zeitintervalle verdoppelt wird. Auch von der Gesamtanzahl an Iterationen ist der PDASSP Algorithmus den anderen Algorithmen vorzuziehen.

Die optimale Steuerung  $\bar{u}$  dieses Problems ist, wie in Kapitel 2.1.2 bewiesen, stetig und die Funktion  $\bar{g} = \int_{\Omega} \bar{y}(x, y, \cdot) \delta(x, y, \cdot) d(x, y) - y_{\max}$  in der Zustandsbeschränkung glatt. Der Multiplikator  $\mu_t$ , sowie  $\bar{g}$  und  $\bar{u}$  scheinen separat auf den aktiven und inaktiven Teilintervallen noch deutlich höhere Regularität zu besitzen, als in den Aussagen in Kapitel 2.1.2 ausgeführt (Abbildungen 3.11, 3.12a). Beim Übergang von inaktiver zu aktiver Menge und umgekehrt sind aber auch deutlich Knicks in der Steuerung erkennbar. Insofern sind trotz der hohen Regularität der Daten global auf  $I$  keine Stetigkeiten höherer Ableitungsordnungen zu erwarten, also nur  $\bar{u} \in C([0, 1])$  und  $\bar{g} \in C^1([0, 1])$ . Der

### 3. Algorithmen und numerische Versuche

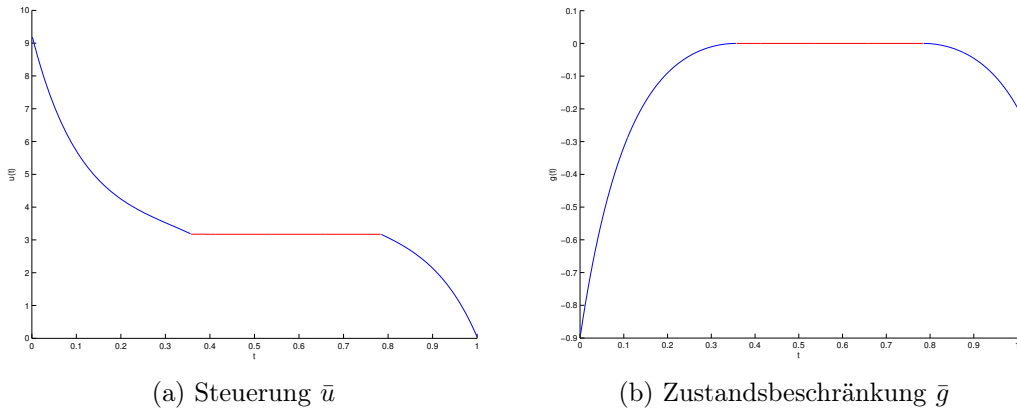


Abbildung 3.11.: Problem 1: Optimale Steuerung  $\bar{u}$  und Funktion in der Zustandsbeschränkung  $\bar{g}$  am Optimum

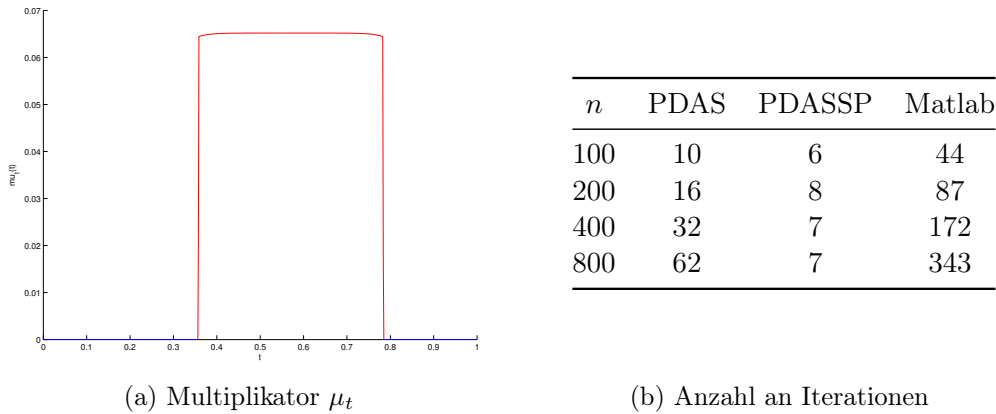


Abbildung 3.12.: Problem 1: Lagrange Multiplikator  $\mu_t$  und benötigte Anzahl an Iterationen in Abhängigkeit der Anzahl der Zeitschritte  $n$

Verlauf der Funktion  $\mu_t$  indiziert ebenfalls die bewiesene Stetigkeit von  $\mu$  auf  $[0, 1]$  und die Unstetigkeit von  $\mu_t$ . Diese Ergebnisse decken sich mit den Resultaten aus der optimalen Steuerung gewöhnlicher Differentialgleichungen [44], wenn man berücksichtigt, dass es sich bei der Zustandsbeschränkung in diesem Problem um eine Zustandsbeschränkung erster Ordnung handelt (Kapitel 2.1.3).

**Problem 2: Problem mit mehreren aktiven Mengen** Für  $(x, y) \in \Omega$  und  $t \in I$  seien die Daten gegeben durch:

- $e(x, y, t) = 1$ ,
- $\delta(x, y, t) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{1}{2} \frac{(x-0.5)^2 + (y-0.5)^2}{\sigma^2}\right)$  mit  $\sigma = 0.1$ ,
- $y_d(x, y, t) = 2 - 3t \sin(9\pi t)$ ,  $y_{\max}(t) = 0.9$  sowie  $\lambda = 0.01$ .



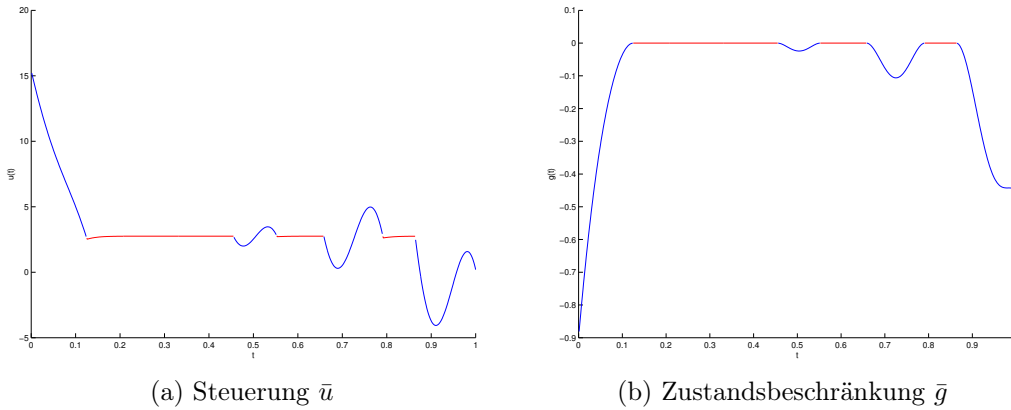


Abbildung 3.13.: Problem 2: Optimale Steuerung  $\bar{u}$  und Funktion in der Zustandsbeschränkung  $\bar{g}$  am Optimum

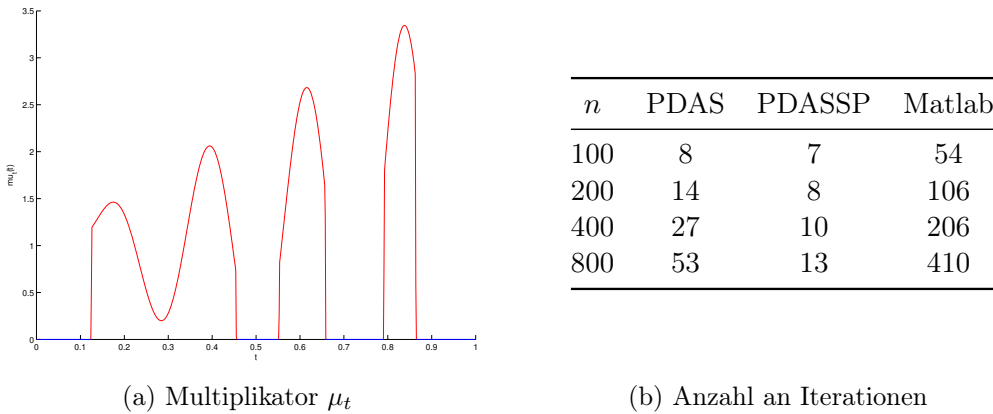


Abbildung 3.14.: Problem 2: Lagrange Multiplikator  $\mu_t$  und benötigte Anzahl an Iterationen in Abhängigkeit der Anzahl der Zeitschritte  $n$

Im Unterschied zum ersten Problem treten mehrere aktive Mengen auf. Außerdem zeigt das Problem deutlich mehr Dynamik, wobei jedoch die sehr hohe Glattheit der Funktionen beibehalten ist. Dieses Beispiel wurde in Kapitel 3.1.4 eingesetzt, um den Korrekturschritt des PDASSP Algorithmus zu motivieren. Im Vergleich zu Problem 1 muss im PDASSP Algorithmus mehrmals ein Korrekturschritt durchgeführt werden. Von Iteration zu Iteration treten Topologieänderungen der aktiven Menge und Vergrößerungen und Verkleinerungen dieser auf.

Die Anzahl der Iterationen des PDASSP Algorithmus ist bei diesem Beispiel nicht konstant (Tabelle 3.14b). Bei Verringerung der Zeitschrittweite nimmt die Anzahl der Iterationsschritte leicht zu. Die Gesamtanzahl an Iterationen ist jedoch auch hier bei zunehmender Anzahl an Freiheitsgraden deutlich geringer als bei den anderen beiden Algorithmen.

Hinsichtlich der Regularität der Lösungen verhält sich dieses Beispiel wie Problem 1 (Abbildungen 3.13, 3.14a), da die Regularität der Datenfunktionen unverändert sehr

hoch ist.

**Problem 3: Problem mit in Zeitrichtung unstetiger Funktion  $e$**  Für  $(x, y) \in \Omega$  und  $t \in I$  seien die Daten gegeben durch:

- $e(x, y, t) = 1 + \mathbf{1}_{t \in [0, \frac{1}{5}]}(t) + \mathbf{1}_{t \in [\frac{2}{5}, \frac{3}{5}]}(t) + \mathbf{1}_{t \in [\frac{4}{5}, \frac{5}{5}]}(t)$ ,
- mit  $\mathbf{1}_{t \in [t_1, t_2]}(t) = \begin{cases} 1 & \text{für } t_1 \leq t \leq t_2, \\ 0 & \text{sonst,} \end{cases}$
- $\delta(x, y, t) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{1}{2} \frac{(x-0.5)^2 + (y-0.5)^2}{\sigma^2}\right)$  mit  $\sigma = 0.1$ ,
- $y_d(x, y, t) = 2$ ,  $y_{\max}(t) = 1$  sowie  $\lambda = 0.1$ .

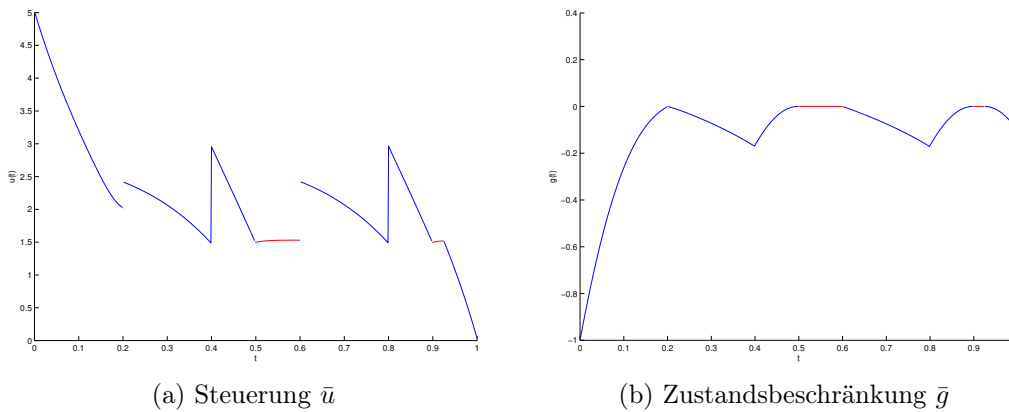
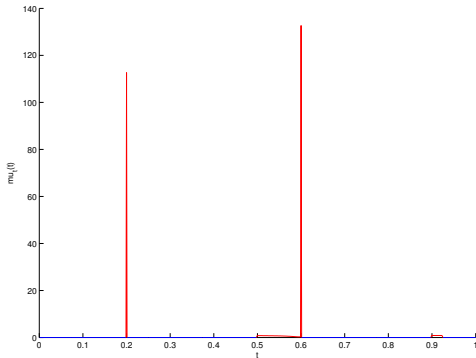


Abbildung 3.15.: Problem 3: Optimale Steuerung  $\bar{u}$  und Funktion in der Zustandsbeschränkung  $\bar{g}$  am Optimum

Mit diesem Beispiel soll untersucht werden, welche Regularität die Lösungen am Optimum zeigen, wenn  $e$  in Zeitrichtung nur der Mindestanforderung  $e \in L^\infty(0, T; L^2(\Omega))$  genügt und nicht die Bedingung  $e \in C([0, T]; L^2(\Omega))$  erfüllt (Kapitel 2.1.2). Die anderen Datenfunktionen sind wieder sehr regulär gewählt.

Berechnet man die Lösungen mit der PDAS Strategie, so zeigt sich, dass die optimale Steuerung  $\bar{u}$  bereits für dieses einfach konstruierte Beispiel nicht stetig ist (Abbildung 3.15). Die Bedingung  $e \in C([0, T]; L^2(\Omega))$  ist also eine notwendige Bedingung für die Stetigkeit der optimalen Steuerung und die weitergehenden Resultate. Dies wird auch bei Betrachtung des Lagrange Multiplikators in Abbildung 3.16a deutlich. Zum Zeitpunkt  $t = 0.2$  und  $t = 0.6$ , also Zeiten, zu denen die Zustandsbeschränkung aktiv ist und  $e$  unstetig ist, sind punktweise hohe Spitzen in der Darstellung von  $\mu_t$  erkennbar. Dies ist ein Hinweis darauf, dass  $\mu$  neben einem stetigen Anteil noch einen diskreten Anteil besitzt.

Interessanterweise konvergiert trotz der Unstetigkeit der optimalen Lösung  $\bar{u}$  der PDASSP Algorithmus in diesem Beispiel auch schneller als die PDAS Strategie (Tabelle 3.16b). Dies lässt sich dadurch erklären, dass der PDASSP Algorithmus an Auf-



(a) Multiplikator  $\mu_t$

$n$	PDAS	PDASSP	Matlab
100	21	13	19
200	41	16	31
400	81	26	56
800	161	46	105

(b) Anzahl an Iterationen

Abbildung 3.16.: Problem 3: Lagrange Multiplikator  $\mu_t$  und benötigte Anzahl an Iterationen in Abhängigkeit der Anzahl der Zeitschritte  $n$

und Absprungpunkten der aktiven Menge, bei denen auch am Optimum eine Stetigkeit der Steuerung vorliegt, schneller gute Schätzungen liefert. Gleichzeitig ist aber an Unstetigkeitsstellen eigentlich damit zu rechnen, dass die im PDASSP Algorithmus verfolgte Strategie, Stetigkeit der Steuerung zu erzeugen, keine Vorteile bietet sondern eher irreführend ist. Es sei aber an dieser Stelle nochmals darauf hingewiesen, dass aufgrund der Bedingung (3.9) der PDASSP Algorithmus bei genügend hoher Anzahl an Iterationen immer auf die PDAS Strategie umschwenken wird. Ein Beispiel für dieses Verhalten ist bei  $n = 800$  in der Anzahl der Iterationen des PDASSP Algorithmus ersichtlich. Die Anzahl der Iterationen verdoppelt sich im Vergleich zu  $n = 400$  nahezu. Dies hat damit zu tun, dass der PDASSP Algorithmus Probleme hat, den Absprungpunkt zum Zeitpunkt  $t = 0.6$  zu lokalisieren. Es finden viele Vergrößerungen und Verkleinerungen des aktiven Intervalls rund um  $t = 0.6$  statt, bis sich der Algorithmus wie der PDAS Algorithmus der Lösung nähert. In Folge zeigt sich eine Gitterabhängigkeit. Erwähnenswert ist auch, dass sich bei diesem Problemtyp die aktive-Mengen-Strategie in `quadprog` stets besser schlägt als die PDAS Strategie.

**Problem 4: Problem mit absolut stetiger Funktion  $e$**  Für  $(x, y) \in \Omega$  und  $t \in I$  seien die Daten gegeben durch:

- $e(x, y, t) = 1 + \alpha_{t \in [0, \frac{1}{5}]}(t) + \alpha_{t \in [\frac{2}{5}, \frac{3}{5}]}(t) + \alpha_{t \in [\frac{4}{5}, \frac{5}{5}]}(t),$

$$\text{mit } \alpha_{t \in [t_1, t_2]}(t) = \begin{cases} (t - t_1)^{\frac{101}{200}} & \text{für } t_1 \leq t < 0.5 \cdot (t_1 + t_2), \\ (t_2 - t)^{\frac{101}{200}} & \text{für } 0.5 \cdot (t_1 + t_2) \leq t \leq t_2, \\ 0 & \text{sonst,} \end{cases}$$

- $\delta(x, y, t) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{1}{2} \frac{(x-0.5)^2 + (y-0.5)^2}{\sigma^2}\right)$  mit  $\sigma = 0.1,$

- $y_d(x, y, t) = 2.5, y_{\max}(t) = 1$  sowie  $\lambda = 0.1.$

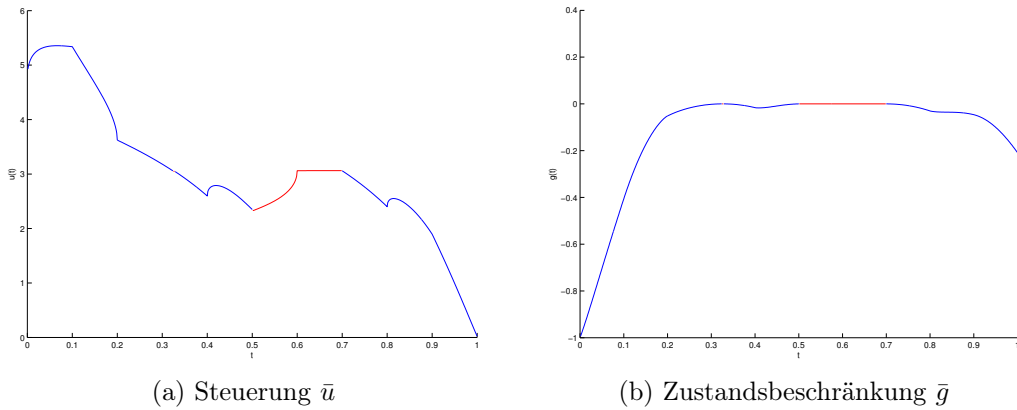


Abbildung 3.17.: Problem 4: Optimale Steuerung  $\bar{u}$  und Funktion in der Zustandsbeschränkung  $\bar{g}$  am Optimum

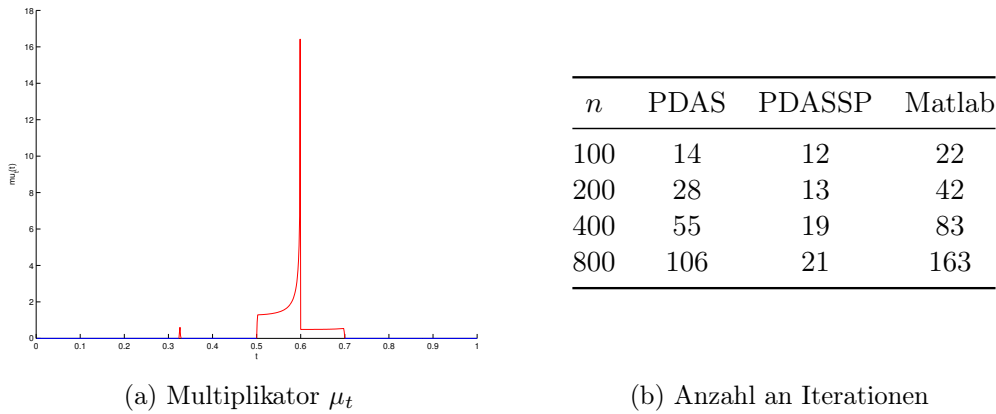


Abbildung 3.18.: Problem 4: Lagrange Multiplikator  $\mu_t$  und benötigte Anzahl an Iterationen in Abhängigkeit der Anzahl der Zeitschritte  $n$

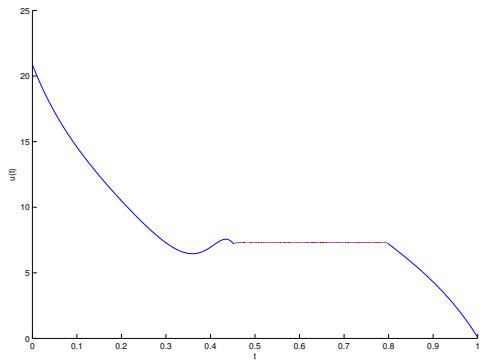
Dieses Problem ist von der Struktur möglichst nahe an Problem 3 angelehnt. Allerdings ist die Funktion  $e$  gerade so gewählt, dass die Bedingung  $e \in W(0, T)$  aus Annahme 2.1.3 erfüllt ist. Die Funktion  $e$  ist in Zeitrichtung dabei noch so irregulär, dass die Bedingung nahezu scharf erfüllt ist. Entsprechend der theoretischen Aussagen aus dem Kapitel 2.1.2 wirkt sich die erhöhte Regularität von  $e$  in der Stetigkeit der optimalen Steuerung aus. Das Beispiel unterscheidet sich aber insofern von den regulären Problemen 1 oder 2, als dass der Multiplikator (Abbildung 3.18a) nun keine sehr hohe Regularität mehr auf aktiven Teilintervallen aufweist. Die Singularität in  $t = 0.6$  müsste gemäß Satz 2.1.6 gerade so sein, dass  $\mu_t \in L^1(0, 1)$  gilt. Die optimale Steuerung ist absolut stetig aber nicht Lipschitz-stetig (Abbildung 3.17). Insofern scheinen die Aussagen aus Kapitel 2.1.2 scharf zu sein und die höheren Regularitätsergebnisse in den Problemen 1 und 2 tatsächlich den deutlich erhöhten Regularitäten der Daten geschuldet zu sein.

Bezüglich der Anzahl an Iterationen in Tabelle 3.18b verhält sich Problem 4 ähnlich zu Problem 2. Der PDASSP Algorithmus kann von der Stetigkeit der Steuerung am Opti-

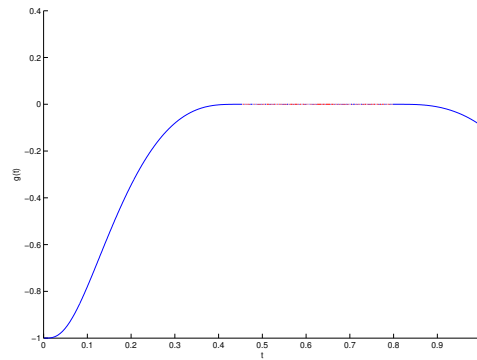
mum profitieren. Im Unterschied zu den anderen beiden Algorithmen nimmt die Anzahl der Iterationen bei feinerer Diskretisierung nur moderat zu. Auch von der absoluten Anzahl an Iterationen ist der PDASSP Algorithmus den anderen Verfahren vorzuziehen.

**Problem 5: Problem mit**  $\int_{\Omega} e(x, y, t)\delta(x, y, t)d(x, y) = 0$  **für alle**  $t \in [0, 1]$  Für  $(x, y) \in \Omega$  und  $t \in I$  seien die Daten gegeben durch:

- $e(x, y, t) = 4(x - 0.5) \cdot \mathbf{1}_{x \in [0.5, 1]}(x)$ ,
- $\delta(x, y, t) = 6 \cdot \mathbf{1}_{x \in [0, 0.25]}(x)$ ,
- $y_d(x, y, t) = 3 - t$ ,  $y_{\max}(t) = 1$  sowie  $\lambda = 0.01$ .

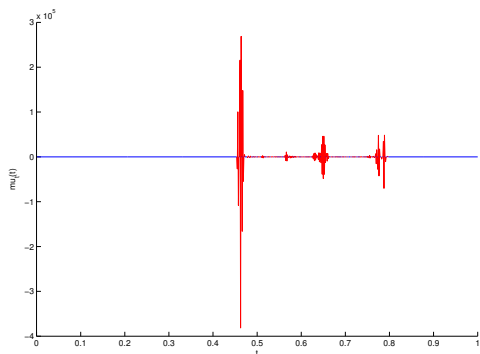


(a) Steuerung  $u$



(b) Zustandsbeschränkung  $g$

Abbildung 3.19.: Problem 5: Steuerung  $u$  und Funktion in der Zustandsbeschränkung  $g$  nach 500 Iterationen des PDASSP Algorithmus



(a) Multiplikator  $\mu_t$

$n$	PDAS	PDASSP	Matlab
100	30	31	75
200	500*	500*	438
400	500*	500*	500*
800	500*	500*	500*

(b) Anzahl an Iterationen

Abbildung 3.20.: Problem 5: Schätzung für den Lagrange Multiplikator  $\mu_t$  nach 500 Iterationen im PDASSP Algorithmus und Anzahl an Iterationen in Abhängigkeit der Anzahl der Zeitschritte  $n$  (500\*: Abbruch nach maximaler Anzahl an Iterationen)

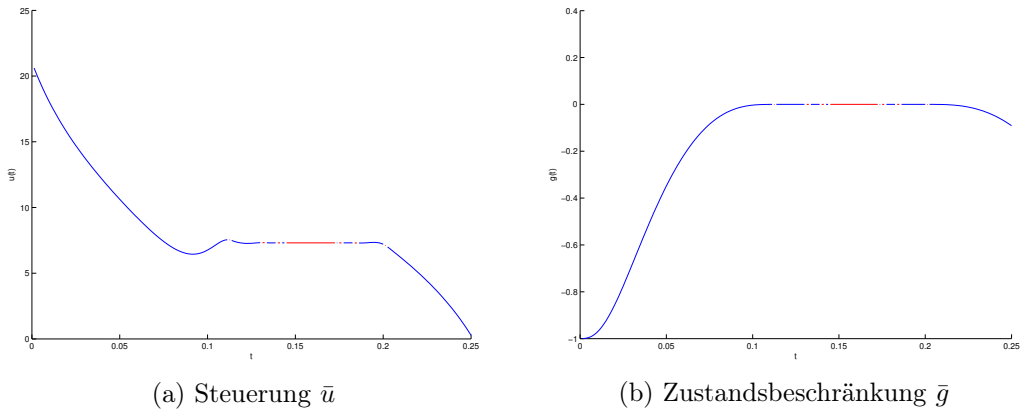


Abbildung 3.21.: Problem 5: Optimale Steuerung  $\bar{u}$  und Funktion in der Zustandsbeschränkung  $\bar{g}$  am Optimum für  $n = 200$

Eine weitere interessante Frage ist, wie sich die Algorithmen verhalten, wenn die Daten gerade so gewählt sind, dass  $\int_{\Omega} e(x, y, t)\delta(x, y, t)d(x, y) = 0$  für alle  $t \in [0, 1]$  gilt. Laut der Aussagen in Kapitel 2.1.4 müsste die optimale Steuerung für die gewählten Daten absolut stetig sein, es ist aber kein Rückschluss auf die Regularität von  $\mu$  möglich.

Beim Test zeigt sich, dass keines der verwendeten Verfahren bei feiner Diskretisierung im Rahmen der vorgegebenen maximalen Anzahl von Iterationen konvergieren kann (Tabelle 3.20b). Die aktive-Mengen-Strategie in Matlab konvergiert zumindest noch bei  $n = 200$ . Betrachtet man das Verhalten des Multiplikators während der Iterationen und nach der maximalen Anzahl an Iterationen (Abbildung 3.20a), so zeigt sich, dass dieser lokal sehr starke Schwankungen aufweist, obwohl die aktuellen Schätzungen der Steuerung  $u$  sehr regulär erscheinen (Abbildung 3.19). Dies hängt vermutlich mit der fehlenden Kopplung der Variablen in der Gleichung

$$0 = \lambda\bar{u} + q_e - \underbrace{\mu \int_{\Omega} \delta e \, dx}_{=0} \quad \text{in } I$$

aus dem Optimalitätssystem zusammen.

In der Darstellung 3.21 der optimalen Steuerung, die mit dem Algorithmus `quadprog` für  $n = 200$  gewonnen wurde, erkennt man trotz der Konvergenzprobleme der Algorithmen die bewiesene Stetigkeit der Steuerung am Optimum. Darüber hinaus scheint für dieses Problem sogar die Zeitableitung  $\frac{d}{dt}\bar{u}$  stetig zu sein. An dieser Stelle beachte man, dass wegen der Bedingung  $\int_{\Omega} e(x, y, t)\delta(x, y, t)d(x, y) = 0$  für alle  $t \in [0, 1]$  die Ordnung der Zustandsbeschränkung im Unterschied zu den vorangehenden Beispielen größer eins ist.

**Problem 6: Problem mit**  $\int_{\Omega} e(x, y, t)\delta(x, y, t)d(x, y) \neq 0$  **für alle**  $t \in [0, 1]$  Für  $(x, y) \in \Omega$  und  $t \in I$  seien die Daten gegeben durch:

- $e(x, y, t) = \frac{4}{3}(x - 0.25) \cdot \mathbf{1}_{x \in [0.25, 1]}(x)$ ,

- $\delta(x, y, t) = 2 \cdot \mathbf{1}_{x \in [0, 0.75]}(x)$ ,
- $y_d(x, y, t) = 3 - t$ ,  $y_{\max}(t) = 1$  sowie  $\lambda = 0.01$ .

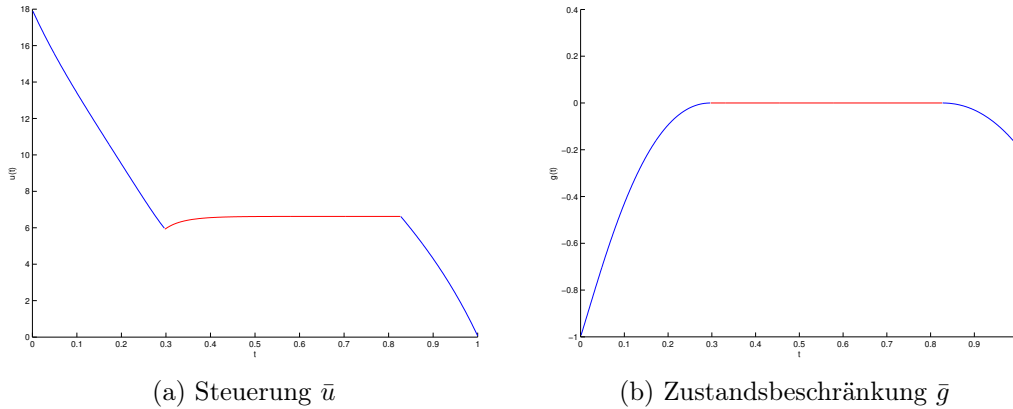


Abbildung 3.22.: Problem 6: Optimale Steuerung  $\bar{u}$  und Funktion in der Zustandsbeschränkung  $\bar{g}$  am Optimum

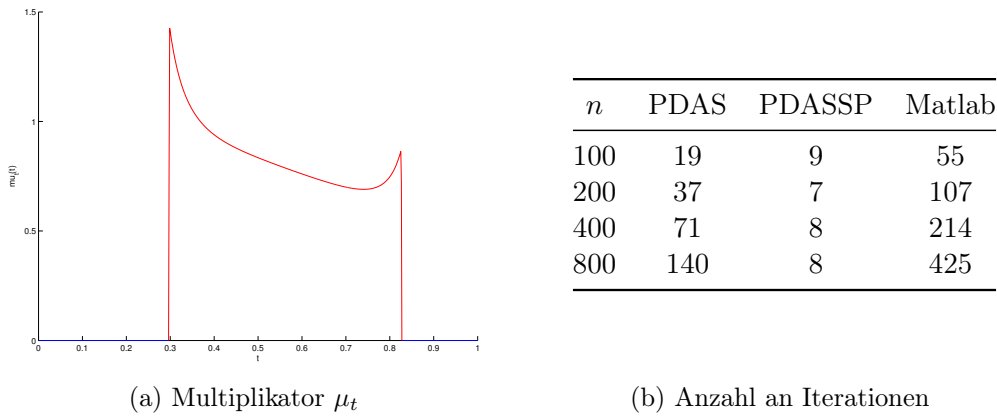


Abbildung 3.23.: Problem 6: Lagrange Multiplikator  $\mu_t$  und benötigte Anzahl an Iterationen in Abhängigkeit der Anzahl der Zeitschritte  $n$

Dieses Beispiel dient dem Beleg, dass die in Problem 5 auftretenden Konvergenzprobleme tatsächlich mit der Bedingung

$$\int_{\Omega} e(x, y, t) \delta(x, y, t) d(x, y) = 0 \quad \text{für alle } t \in [0, 1] \quad (3.13)$$

zusammenhängen und nicht der im Vergleich zu den vorherigen Problemen niedrigen Regularität der Datenfunktionen im Ort geschuldet sind. An dieser Stelle sei darauf hingewiesen, dass die Ortsregularität der Daten für die theoretischen Aussagen in Kapitel 2.1.4 immer noch hinreichend sind. Die Datenfunktionen werden für dieses Beispiel konkret so gewählt, dass sie von der Struktur mit dem vorangehenden Problem vergleichbar sind, die Bedingung  $\int_{\Omega} e(x, y, t) \delta(x, y, t) d(x, y) \neq 0$  für alle  $t \in [0, 1]$  jedoch einhalten.

Wie in Problem 1 beobachtet stellt sich eine hohe Regularität der Größen am Optimum heraus. Auch die Vorzüge des PDASSP Algorithmus im Konvergenzverhalten kommen unmittelbar zum Tragen. Das Verfahren ist gitterunabhängig (Tabelle 3.23b). Insofern scheinen die Probleme in Beispiel 5 tatsächlich von der Bedingung (3.13) und der damit verbundenen Änderung der Ordnung der Zustandsbeschränkung her zu rühren.

In den bisherigen Problemen wurde stets eine hohe Regularität der Datenfunktion  $\delta$  in Zeitrichtung vorgegeben. In den abschließenden drei Beispielen wird es deshalb darum gehen, die Abhängigkeit der optimalen Lösungen von dieser Einflussgröße darzustellen.

**Problem 7:**  $\int_{\Omega} e(x, y, t)\delta(x, y, t)d(x, y) = 0$  für alle  $t \in [0, 1]$  Für  $(x, y) \in \Omega$  und  $t \in I$  seien die Daten gegeben durch:

- $e(x, y, t) = 3(x - 0.5) \cdot \mathbf{1}_{x \in [0.5, 1]}(x) \cdot (1 + \alpha_{t \in [0, \frac{1}{5}]}(t) + \alpha_{t \in [\frac{2}{5}, \frac{3}{5}]}(t) + \alpha_{t \in [\frac{4}{5}, \frac{5}{5}]}(t)),$
- $\delta(x, y, t) = 6 \cdot \mathbf{1}_{x \in [0, 0.25]}(x) \cdot (1 + \beta_{t \in [0, \frac{1}{5}]}(t) + \beta_{t \in [\frac{2}{5}, \frac{3}{5}]}(t) + \beta_{t \in [\frac{4}{5}, \frac{5}{5}]}(t)),$

$$\text{mit } \beta_{t \in [t_1, t_2]}(t) = \begin{cases} \sqrt{t - t_1} & \text{für } t_1 \leq t < 0.5 \cdot (t_1 + t_2), \\ \sqrt{t_2 - t} & \text{für } 0.5 \cdot (t_1 + t_2) \leq t \leq t_2, \\ 0 & \text{sonst,} \end{cases}$$

- $y_d(x, y, t) = 8 - t, y_{\max}(t) = 1$  sowie  $\lambda = 0.01.$

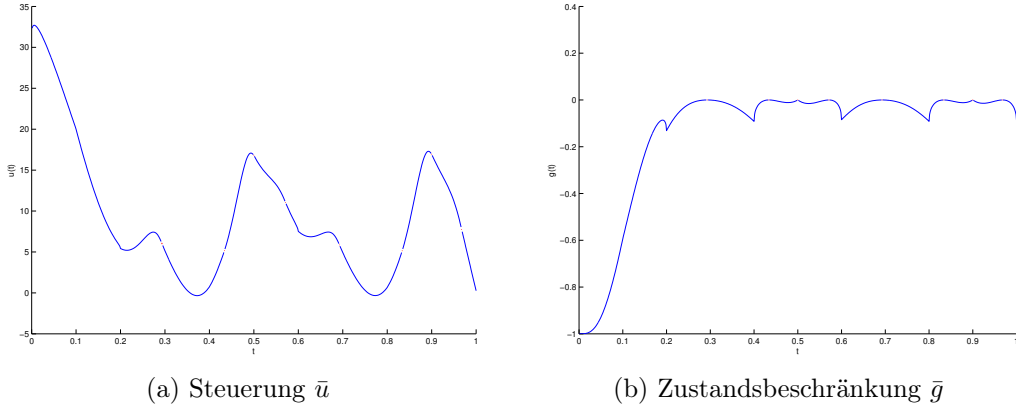
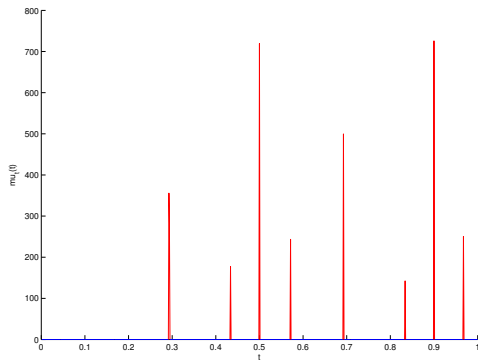


Abbildung 3.24.: Problem 7: Optimale Steuerung  $\bar{u}$  und Funktion in der Zustandsbeschränkung  $\bar{g}$  am Optimum

Wie in Problem 5 sind  $e$  und  $\delta$  so konstruiert, dass  $\int_{\Omega} e(x, y, t)\delta(x, y, t)d(x, y) = 0$  für alle  $t \in [0, 1]$  gilt. Die Zustandsbeschränkung besitzt somit eine Ordnung größer eins. Da außerdem  $e, \delta \in AC^{1,1}([0, T]; L^2(\Omega))$  und  $e \in W(0, T)$  erfüllt sind, ist Korollar 2.1.4 anwendbar und liefert für diese Vorgaben die absolute Stetigkeit der optimalen Steuerung.

Auch bei diesem Beispiel treten die aus Problem 5 bekannten Konvergenzprobleme der PDAS(SP) Algorithmen auf (Tabelle 3.25b). Die aktive-Mengen-Strategie in `quadprog`





(a) Multiplikator  $\mu_t$

$n$	PDAS	PDASSP	Matlab
100	32	32	19
200	500*	500*	32
400	500*	500*	54
800	500*	500*	100

(b) Anzahl an Iterationen

Abbildung 3.25.: Problem 7: Lagrange Multiplikator  $\mu_t$  und benötigte Anzahl an Iterationen in Abhängigkeit der Anzahl der Zeitschritte  $n$  (500\*: Abbruch nach maximaler Anzahl an Iterationen)

liefert dagegen für jede Anzahl an Zeitintervallen eine Lösung mit relativ wenigen Iterationen, wobei jedoch auch bei diesem Problem Gitterabhängigkeit auftritt.

Bezüglich der Schärfe der in Kapitel 2.1.4 getätigten Aussagen lohnt es sich, die optimalen Lösungen aus `quadprog` näher zu betrachten. Die optimale Steuerung ist wie zu erwarten stetig (Abbildung 3.24a). Wie in Problem 5 hat man sogar die Stetigkeit der Ableitung  $\frac{d}{dt}\bar{u}$ . Die Funktion  $\mu$  scheint allerdings nicht absolut stetig zu sein (Abbildung 3.25a). Vielmehr scheint  $\mu$  nur aus einem diskreten Anteil zu bestehen und die Zustandsbeschränkung nur punktweise aktiv zu werden. An diesen Zeitpunkten, z.B.  $t = 0.5$ , springt dann auch die Zeitableitung der Funktion  $\bar{g}$  (Abbildung 3.24b). Diese Beobachtung passt insofern gerade zu den Aussagen in Kapitel 2.1.4, da Korollar 2.1.4 nur die Stetigkeit der optimalen Steuerung garantiert und für weiterführende Aussagen bezüglich  $\frac{d}{dt}\bar{g}$  und  $\mu$  die nötigen Bedingungen in Satz 2.1.8 nicht eingehalten sind.

**Problem 8** Für  $(x, y) \in \Omega$  und  $t \in I$  seien die Daten gegeben durch:

- $e(x, y, t) = (x - 0.25) \cdot \mathbf{1}_{x \in [0.25, 1]}(x) \cdot (1 + \alpha_{t \in [0, \frac{1}{5}]}(t) + \alpha_{t \in [\frac{2}{5}, \frac{3}{5}]}(t) + \alpha_{t \in [\frac{4}{5}, \frac{5}{5}]}(t)),$
- $\delta(x, y, t) = 2 \cdot \mathbf{1}_{x \in [0, 0.75]}(x) \cdot (1 + \beta_{t \in [0, \frac{1}{5}]}(t) + \beta_{t \in [\frac{2}{5}, \frac{3}{5}]}(t) + \beta_{t \in [\frac{4}{5}, \frac{5}{5}]}(t)),$
- $y_d(x, y, t) = 8 - t, y_{\max}(t) = 1$  sowie  $\lambda = 0.01.$

In Problem 8 ist bei gleich bleibender Regularität der Daten in Zeitrichtung die Bedingung  $\int_{\Omega} e(x, y, t)\delta(x, y, t)d(x, y) = 0$  für alle  $t \in [0, 1]$  nicht mehr eingehalten. Insofern kann mit den Resultaten aus Kapitel 2.1.4 nicht mehr auf die Stetigkeit der optimalen Steuerung geschlossen werden.

In den numerischen Ergebnissen in Abbildung 3.26 zeigt sich auch tatsächlich, dass die optimale Steuerung in den Zeitpunkten  $t = 0.5$  und  $t = 0.9$  unstetig ist. Das Verhalten von  $\mu_t$  in Abbildung 3.27a indiziert, dass die Funktion  $\mu$  zu diesen Zeitpunkten Unstetigkeiten besitzt und nicht absolut stetig ist auf  $[0, 1].$

### 3. Algorithmen und numerische Versuche

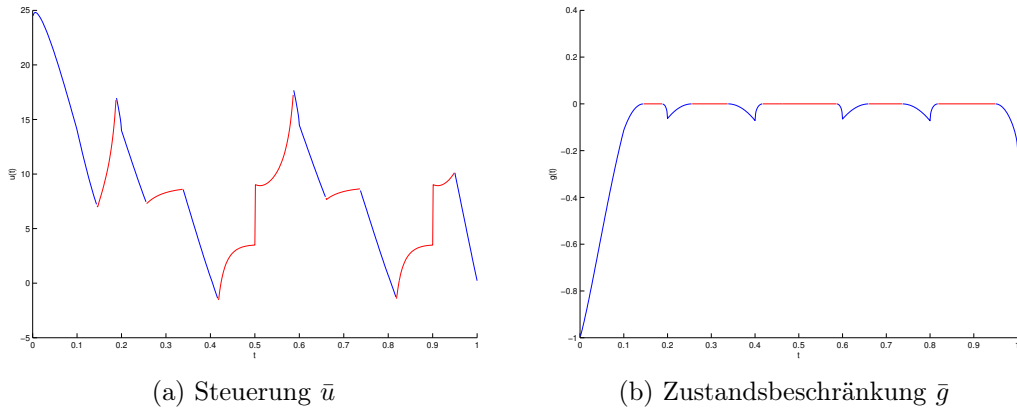


Abbildung 3.26.: Problem 8: Optimale Steuerung  $\bar{u}$  und Funktion in der Zustandsbeschränkung  $\bar{g}$  am Optimum

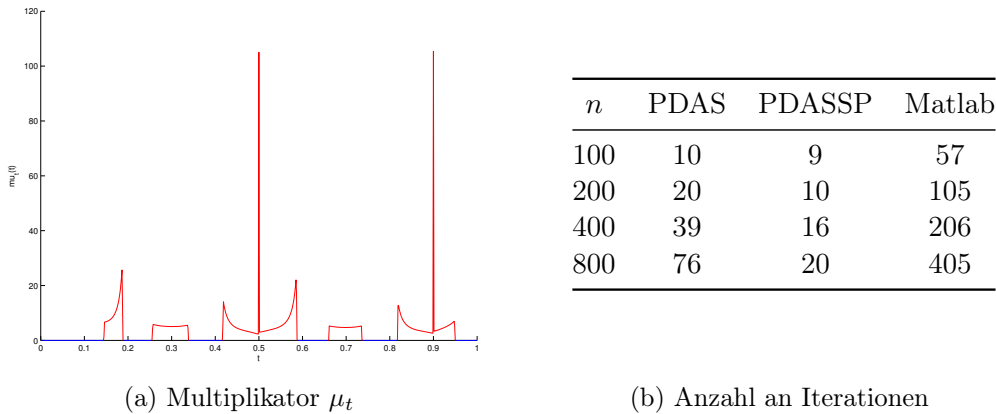


Abbildung 3.27.: Problem 8: Lagrange Multiplikator  $\mu_t$  und benötigte Anzahl an Iterationen in Abhängigkeit der Anzahl der Zeitschritte  $n$

Das auch in diesem Fall überraschend gute Konvergenzverhalten des PDASSP Algorithmus (Tabelle 3.27b) lässt sich gut dadurch erklären, dass die Unstetigkeiten der optimalen Steuerung im Inneren der aktiven Mengen auftreten und nicht am Auf- oder Absprungpunkt dieser. Insofern ist das Bestreben des PDASSP Algorithmus, an diesen Stellen Unstetigkeiten zu reduzieren ein zielführendes Vorgehen. Im Rückblick auf die Probleme 2 und 4 ist erwähnenswert, dass der PDASSP Algorithmus ein stark vergleichbares Konvergenzverhalten zeigt: Leicht ansteigende Anzahl an Iterationen bei Verfeinerung der Diskretisierung bei im Vergleich zu den anderen Algorithmen aber insgesamt niedrigen Iterationszahlen.

**Problem 9** Für  $(x, y) \in \Omega$  und  $t \in I$  seien die Daten gegeben durch:

- $e(x, y, t) = (x - 0.25) \cdot \mathbf{1}_{x \in [0.25, 1]}(x) \cdot (1 + \alpha_{t \in [0, \frac{1}{5}]}(t) + \alpha_{t \in [\frac{2}{5}, \frac{3}{5}]}(t) + \alpha_{t \in [\frac{4}{5}, \frac{5}{5}]}(t)),$

- $\delta(x, y, t) = 2 \cdot \mathbf{1}_{x \in [0, 0.75]}(x) \cdot (1 + \hat{\mathbf{1}}_{t \in [0, \frac{1}{5}]}(t) + \hat{\mathbf{1}}_{t \in [\frac{2}{5}, \frac{3}{5}]}(t) + \hat{\mathbf{1}}_{t \in [\frac{4}{5}, \frac{5}{5}]}(t)),$   
mit  $\hat{\mathbf{1}}_{t \in [\frac{k-1}{5}, \frac{k}{5}]}(t) = \begin{cases} \frac{1}{2} \sin((t - \frac{k}{20}) \cdot 10\pi) + \frac{1}{2} & \text{für } \frac{k-1}{5} \leq t < \frac{2k-1}{10}, \\ 10^4 \cdot (t - \frac{k-1}{5})^2 \cdot (t - \frac{k}{5})^2 & \text{für } \frac{2k-1}{10} \leq t \leq \frac{k}{5}, \\ 0 & \text{sonst,} \end{cases}$  für  $k \in \mathbb{N},$
- $y_d(x, y, t) = 8 - t, y_{\max}(t) = 1$  sowie  $\lambda = 0.01.$

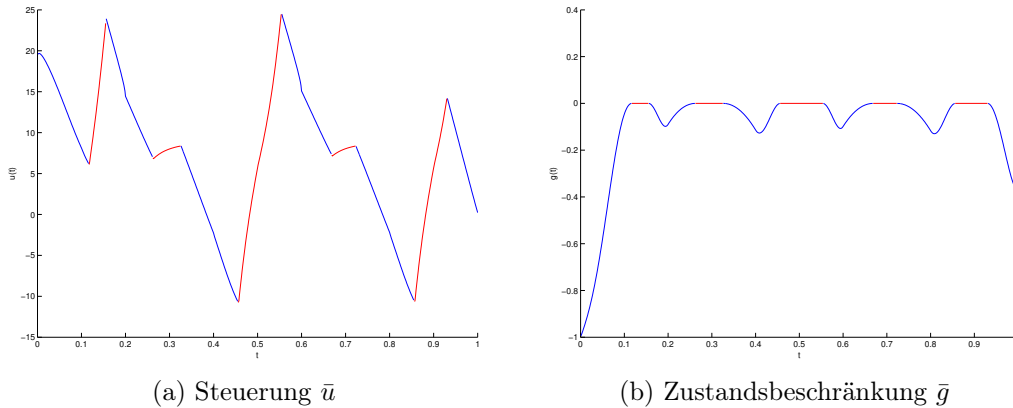


Abbildung 3.28.: Problem 9: Optimale Steuerung  $\bar{u}$  und Funktion in der Zustandsbeschränkung  $\bar{g}$  am Optimum

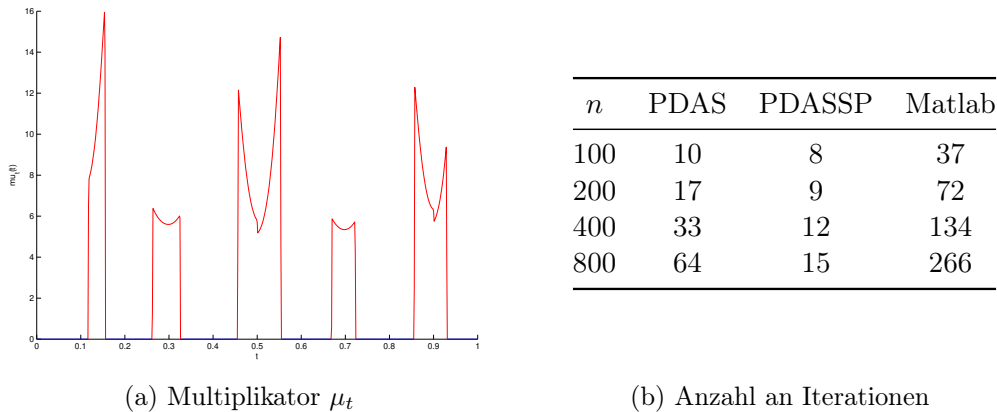


Abbildung 3.29.: Problem 9: Lagrange Multiplikator  $\mu_t$  und benötigte Anzahl an Iterationen in Abhängigkeit der Anzahl der Zeitschritte  $n$

Dieses Problem wurde im Hinblick der Ergebnisse der beiden vorangehenden Beispiele so konstruiert, dass man den in Satz 2.1.8 geforderten Voraussetzungen sehr nahe kommt und diese erfüllt: Für die erste Zeitableitung gilt  $\delta_t \in C([0, T]; L^\infty(\Omega))$ , für die zweite Zeitableitung hat man aber keine Stetigkeit mehr. Die Bedingung

$$\int_{\Omega} e(x, y, t) \delta(x, y, t) d(x, y) \neq 0$$

ist für alle  $t \in [0, 1]$  eingehalten.

Entsprechend zeigt sich am Optimum die Stetigkeit der Steuerung und die höhere Regularität der anderen Größen (Abbildungen 3.28, 3.29a). Auf aktiven Intervallen hat man in diesem Fall sogar eine Lipschitz-Stetigkeit von  $\mu$ , wenngleich in diesem Problem im Vergleich zu Problem 4 die Zeitpunkte, in denen die Zeitableitung von  $e$  singularär ist, nicht innerhalb der aktiven Menge liegen. Das Verhalten der optimalen Steuerung in den Zeitpunkten  $t = 0.2$  und  $t = 0.6$  indiziert, dass  $\bar{u}$  absolutstetig aber nicht Lipschitz-stetig ist.

Der PDASSP Algorithmus kann auch bei diesem Problem seine Stärke ausspielen (Tabelle 3.29b). Allerdings zeigt er kein vollkommen gitterunabhängiges Verhalten. Eine leichte Zunahme der Anzahl an Iterationen ist festzustellen.

**Gesamtfazit** Der PDASSP Algorithmus weist bei regulären Problemen (Probleme 1, 2, 4, 6 und 9) hinsichtlich der absoluten Anzahl an Iterationen und hinsichtlich der Veränderung dieser bei einer Zunahme der Zeitschritte ein deutlich besseres Konvergenzverhalten auf als der PDAS Algorithmus und die in Matlab `quadprog` implementierte aktive-Mengen-Strategie. Der PDASSP Algorithmus zeigt bei den Problemen 1 und 6 sogar Gitterunabhängigkeit. Bei regulären Beispielen mit großer Dynamik (Probleme 2, 4 und 9) treten moderate Zunahmen der Anzahlen an Iterationen auf. Die Anzahl der nötigen Iterationen verdoppelt sich etwa, wenn die Anzahl der Zeitschritte verachtfacht wird. In Problemen, in denen die Daten so gewählt sind, dass die Bedingung

$$\int_{\Omega} e(x, y, t) \delta(x, y, t) d(x, y) = 0 \text{ für alle } t \in [0, 1]$$

gilt, konvergiert der PDASSP Algorithmus häufig nicht innerhalb der vorgegebenen Maximalanzahl an Iterationen. Während der PDAS Algorithmus auch dieses Verhalten zeigt, erscheint der Algorithmus in `quadprog` deutlich robuster und liefert insbesondere bei Problem 7 Lösungen mit einer moderaten Anzahl an Iterationen.

Überraschend gut ist das Verhalten des PDASSP Algorithmus bei Problemen, deren optimale Steuerung auch Unstetigkeiten aufweist (Probleme 3 und 8). Hier schlägt sich der Algorithmus von der absoluten Anzahl an Iterationen stets besser als die anderen beiden Varianten. Der Vorteil ist vermutlich der Tatsache geschuldet, dass Unstetigkeiten in der optimalen Steuerung nur dann zum Tragen kommen, wenn sich die Unstetigkeiten gerade an Auf- oder Absprungpunkten einer aktiven Menge befinden. In diesen Fällen ist eine Schaltpunktsuche eigentlich alles andere als zielführend. In Problem 3 zeigt sich aber, dass mit den in Kapitel 3.1.4 gewählten Parametern ein Umschwenken des PDASSP Algorithmus auf den PDAS Algorithmus gelingt und Konvergenz erhalten werden kann.

Darüber hinaus ist festzuhalten, dass basierend auf den numerischen Ergebnissen in diesem Kapitel viele der im Kapitel 2.1 getätigten Aussagen scharf zu sein scheinen, wenn man diese auf dem Gesamtintervall  $I = [0, 1]$  betrachtet. Aussagen mit höherer Regularität der Funktionen auf aktiven/inaktiven Teilintervallen scheinen jedoch noch möglich, wenn eine hohe Glattheit der Datenfunktionen gegeben ist (Probleme 1, 2 oder 6).

Hinsichtlich der Ordnung der Zustandsbeschränkungen ist erwähnenswert, dass die in den numerischen Versuchen erhaltenen Ergebnisse zu den Vorhersagen aus der Theorie

der optimalen Steuerung gewöhnlicher Differentialgleichungen passen [44]. Für weitere Untersuchungen wäre es sehr interessant, ob die Konvergenzprobleme der PDAS(SP) Algorithmen auch an anderen Problemen mit Ordnung größer eins festgestellt werden können.



## 4. Optimale Steuerung beim Laserstrahlschmelzen

Mit dem Laserstrahlschmelzen ist es möglich, komplexe metallische Bauteile generativ zu fertigen. Hierzu werden iterativ dünne Schichten Metallpulver aufgetragen und an den gewünschten Stellen mit einem Laser aufgeschmolzen. Im Rahmen dieser Arbeit wird eine Strategie betrachtet, bei der die aufzuschmelzenden Bereiche zunächst in gleichmäßige Segmente, sogenannte Islands, in Schachbrettform unterteilt und nach einer gewissen randomisierten Reihenfolge belichtet werden.

Will man diesen Prozess des Laserstrahlschmelzens, auch LaserCUSING® genannt, optimieren, so sind verschiedene Ansatzpunkte denkbar. Zum einen könnte man versuchen, die Reihenfolge, mit der die verschiedenen Islands abgelaufen werden, so zu planen, dass sich eine möglichst gleichmäßige Temperaturverteilung im Werkstück einstellt und mögliche Verzüge reduziert werden [36]. Weiteres Optimierungspotential ist beim Belichten innerhalb eines einzelnen Islands vorhanden. Um dies nachvollziehen zu können, wird der Standardprozess für das Belichten eines Islands im nächsten Absatz vorgestellt.

Ein Standard Island besteht aus einem Quadrat mit Kantenlänge  $5\text{ mm}$ . Um das Pulver innerhalb dieses Islands aufzuschmelzen, verfährt der Laser in der Standardstrategie in Mäanderform mit konstanter Laserleistung und vorgegebener Geschwindigkeit über das Island (Abbildung 4.1). Die einzelnen geradlinig verlaufenden Abschnitte im Polygonzug werden als Scanvektoren bezeichnet. Beim Belichten mit den Standardparametern des Herstellers beträgt die Schmelzspurbreite  $130\ \mu\text{m}$  bis  $150\ \mu\text{m}$ . Da ein gewisses Überlappen der Schmelzspuren gewünscht ist, setzt man als Abstand parallel zueinander verlaufender Scanvektoren (Hatchabstand) ca.  $105\ \mu\text{m}$ , so dass man in Summe pro Island auf 48 parallel zueinander verlaufende reguläre Scanvektoren und 47 Scanvektoren für die Mäanderquerfahrt kommt. Wegen der Ausdehnung des Laserfokus beginnen und enden die regulären Scanvektoren  $52,5\ \mu\text{m}$  innerhalb des Islands.

Die Schmelzspuren nach dem Belichtungsvorgang eines Islands mit dieser Standardstrategie sind in der Abbildung 4.2 dargestellt. Es ergeben sich in den Umkehrpunkten sichtbare Erhöhungen, Nasenbildung tritt auf (Abbildung 4.2a). Außerdem wird anhand der Mikroskopaufnahme (Abbildung 4.2b) ersichtlich, dass die Schmelzspuren zwar im Inneren eines Islands sehr gleichmäßig ausgeprägt sind, am Rand des Bauteils aber systematische Abweichungen zeigen. Die Tatsache, dass genau jede zweite Schmelzspur am Rand deutlich stärker ausgeprägt ist als die benachbarte, ist ein starkes Indiz für ein grundlegendes Problem mit dieser Belichtungsstrategie: beim Fahren im Mäander und zurück ins Innere des Islands wird systematisch zu viel Energie eingetragen.

In dieser Arbeit soll deshalb untersucht werden, inwiefern durch Berechnung optimaler Strategien zur Anpassung der Leistung des Lasers ein gleichmäßigeres Schmelzbild erzeugt werden kann. Die mäanderförmige Bahn und die Geschwindigkeit, mit der sich

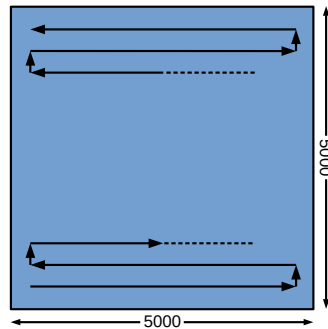
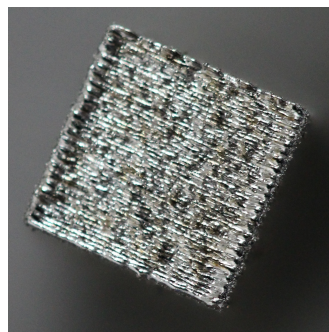
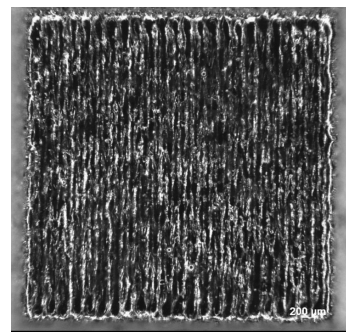


Abbildung 4.1.: Prinzipskizze zur Belichtung eines Standardislands mit Längenangaben in  $\mu m$



(a) Makroaufnahme



(b) Mikroskopaufnahme

Abbildung 4.2.: Bauteil produziert mit Standardverfahren

der Laser entlang der Bahn bewegt, sollen hierbei im Vergleich zum Standardverfahren unverändert bleiben.

Abschließend sei noch kurz das Vorgehen am Rand eines Bauteils dargestellt. Befindet sich ein Island am Rand eines Bauteils, so würde man nach der Belichtung des Inneren eines Islands noch eine Konturlinienfahrt parallel zu der Bauteilkante vornehmen, um die Oberflächenqualität am Rande des Bauteils zu verbessern. Da die Schmelzspur bei der Konturfahrt eine gewisse Breite besitzt, wird das Netz der Islands  $30 \mu m$  vom Rand des Bauteils eingerückt.

### 4.1. Modell

Das Laserstrahlschmelzen ist aus Modellierungssicht ein sehr komplexer und hochdynamischer Vorgang. In der Literatur wird versucht, diesen Prozess auf verschiedenen Skalen zu modellieren: Von der Mikroebene im Bereich einzelner Schmelzspuren bis hin zur Makroebene des Bauprozesses des gesamten Bauteils. Exemplarisch sei hier die Arbeit von Branner [13] genannt. Nach der Kenntnis des Autors liegt der Fokus bei der Modellierung des Schmelzvorgangs auf Mikroebene bisher auf einer Beschreibung im regulären Inneren eines Islands. Auch die Erfahrungswerte des Projektpartners beziehen sich bisher



überwiegend auf die Situation im Inneren eines Islands (die Parameter für Geschwindigkeit und Leistung des Lasers werden bisher so gewählt, dass dort gute Ergebnisse erzielt werden). Einzig in Ilin u. a. [37] wird das Verhalten am Rand eines Bauteils untersucht. Ein Modell zum Verhalten in den Umkehrpunkten des Mäanders ist dem Autor nicht bekannt.

Im Rahmen der Arbeit wurde deshalb das folgende Modell zur Simulation und Optimierung des Laserstrahlschmelzens entwickelt: Aufgrund der Rechenkapazität war es nicht möglich, die Belichtung eines kompletten Islands zu simulieren. Stattdessen wurde ein  $1,5\text{ mm}$  hoher,  $0,735\text{ mm}$  breiter und  $5,06\text{ mm}$  langer, quaderförmiger Körper betrachtet, der rechts und links an ein Pulverbett der Breite  $1\text{ mm}$  angrenzt. Der obere Teil des Körpers besteht aus einer  $30\text{ }\mu\text{m}$  starken Pulverschicht, die im Rahmen des Produktionsprozesses teilweise aufgeschmolzen wird.

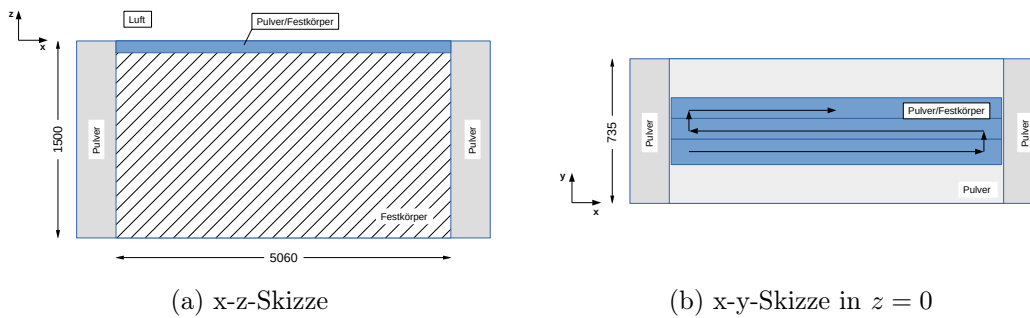


Abbildung 4.3.: Prinzipskizzen des Modells mit Längenangaben in  $\mu\text{m}$

Die obere Pulverschicht wird vom Laser in x-y-Richtung mäandrierend mit einer Geschwindigkeit von  $800\frac{\text{mm}}{\text{s}}$  und einer Leistung von  $180\text{ W}$  abgefahren. Der Laser startet hierzu  $82,5\text{ }\mu\text{m}$  vom linken Rand des Bauteils entfernt und bewegt sich mit konstanter Geschwindigkeit geradlinig, bis er  $82,5\text{ }\mu\text{m}$  vom gegenüberliegenden Rand entfernt ist. Dort fährt er auf einer Länge von  $105\text{ }\mu\text{m}$  quer zum Bauteilrand und anschließend wieder geradlinig zurück ins Bauteilinnere. Die Länge des Bauteils ist mit  $5,06\text{ mm}$  so gewählt, dass es den Maßen in den späteren Baureihen entspricht: die Länge eines Standardislands zuzüglich jeweils  $30\text{ }\mu\text{m}$  für die Konturfahrt (vgl. Kapitel 4.5).

Die Temperatur  $T$  sowohl im Festkörper als auch im Pulver in dem Zeitraum  $[t_0, t_f]$  wird mit Hilfe der instationären Wärmeleitungsgleichung und zugehörigen Randbedingungen beschrieben:

$$\begin{aligned} \rho c_p(s) \frac{\partial T}{\partial t} - \nabla \cdot (k(s) \nabla T) &= p\phi(s) && \text{in } \Omega \times (t_0, t_f), \\ k(s) \frac{\partial T}{\partial n} &= -\alpha(T - 293\text{K}) && \text{auf } \Gamma \times (t_0, t_f), \\ T(t_0) &= T_{\text{init}}, \\ s &= \text{Position des Lasers.} \end{aligned}$$

Die Parameter in den Randbedingungen sind überwiegend so gewählt, dass perfekte Wärmeleitung auftritt ( $\alpha = 10^6\frac{\text{W}}{\text{m}^2\text{K}}$ ). Nur an der oberen Grenze des Werkstücks beim

Übergang zur Luft ist  $\alpha = 50 \frac{W}{m^2K}$  gemäß Branner [13]. Der Energieeintrag durch den Laser pro Zeiteinheit ( $p \cdot \phi$ , Leistung  $p$ , Verteilungsfunktion  $\phi$ ) wird mit einem Goldak Doppellellipsoidmodell nachgebildet [26]. Die zugehörigen Modellparameter beim Abfahren eines regulären Scanvektors sind gegeben durch  $a_r = 240 \mu m$ ,  $a_f = b_f = b_r = 60 \mu m$ ,  $c_f = c_r = 150 \mu m$  und  $w_f = 0,4$ ,  $w_r = 1,6$ . Bei der Mäanderquerfahrt bildet sich jedoch kein langgezogener Ellipsoid aus. Stattdessen wird eine kreisförmige Form gewählt ( $a_r = a_f$ ), die sich zu Beginn des neuen regulären Scanvektors mit der Bewegung des Lasers wieder gleichmäßig zu dem langgezogenen Ellipsoid aufweitet. Als Werkstoff wird der Stahl 1.4404 betrachtet, da zu diesem Material bereits einige wissenschaftliche Voruntersuchungen und Daten vorhanden sind [37]. Im Unterschied zu einer exakten Beschreibung der Temperatur im Werkstück mit temperaturabhängigen Materialparametern – und damit mit einer nichtlinearen Gleichung – musste, um das spätere Optimierungsverfahren anwenden zu können, ein lineares Ersatzmodell entwickelt werden. Dabei kann man sich zunutze machen, dass hohe Temperaturen und Übergänge von Pulver zu Feststoff nur unmittelbar in der Nähe des Lasers auftreten. Die Materialparameter werden infolgedessen nicht in Abhängigkeit der Temperatur, sondern in Abhängigkeit der Position des Lasers variiert. Konkret werden im linearen Ersatzmodell die folgenden (temperaturunabhängigen) Parameter verwendet:

Physikalische Größe	Pulver	Feststoff
Wärmeleitfähigkeit $k$ in $W/(mK)$	0,19	20
Dichte $\rho$ in $kg/m^3$	5000	7850
spezifische Wärmekapazität $c_p$ in $J/(kgK)$	475	475

Tabelle 4.1.: Materialparameter in Pulver und Feststoff

Der Übergang von Pulver zu Feststoff wird in Richtung der Bewegung des Lasers mit Hilfe einer geglätteten Heavisidefunktion modelliert, die so zentriert ist, dass der Übergang in einem Abstand von  $a_f$  (vgl. Goldak Modell) vor der aktuellen Laserposition beginnt und bei dieser abgeschlossen ist. Die Umwandlung der Materialparameter senkrecht zur Bewegungsrichtung findet auf einer Gesamtschmelzspurbreite von  $130 \mu m$  statt. Um die latente Wärme, die sich in einer Singularität der Wärmekapazität  $c_p$  im Bereich der Schmelztemperatur widerspiegelt [37], zu simulieren, wird ebenfalls eine Approximation vorgenommen, die statt von der Temperatur von der Position des Lasers abhängig ist. Die Wärmekapazität wird analog der Verteilungsfunktion  $\phi$  an der aktuellen Position des Lasers auf maximal  $4000 \frac{J}{kgK}$  erhöht.

Für die späteren Untersuchungen ist es von Interesse, eine Kenngröße für die Temperaturen in der Nähe der aktuellen Position  $s$  des Lasers zu besitzen. Es bietet sich an, die Temperatur mit einer radialsymmetrischen Gaußverteilung, die um die Position des Lasers zentriert ist, zu mitteln. Konkret wählt man als Kenngröße:

$$T_s = \int_{\Omega} T \cdot \delta(s) dV, \text{ mit } \delta(s) = \frac{1}{(2\pi\sigma^2)^{\frac{3}{2}}} \cdot e^{-\frac{1}{2} \frac{(x-s_x)^2 + (y-s_y)^2 + (z-0)^2}{\sigma^2}},$$

mit der Laserposition  $s = (s_x, s_y)$  und  $\sigma = 30 \mu m$ . Grundsätzlich besteht beim hochdynamischen Prozess des Laserschmelzens das Problem, dass die realen Temperaturen

im Werkstück in unmittelbarer Nähe der aktuellen Laserposition nicht gemessen werden können. Insofern ist es schwer, das gewählte lineare Ersatzmodell mit der dynamischen Anpassung der verschiedenen Modellgrößen direkt zu verifizieren. Es lässt sich nur eine Bewertung der aus dem Modell resultierenden optimierten Strategien am realen Bauteil vornehmen. Dabei zeigen sich durchaus vielversprechende Ergebnisse (Kapitel 4.5.2).

## 4.2. Simulation des Modells

Das oben dargestellte Modell zum Laserstrahlschmelzen wurde in Comsol Multiphysics 3.5 in Kombination mit Matlab2010 implementiert. Die Ortsdiskretisierung der Variablen wurde mit einem FEM Ansatz mit Lagrange-Elementen zweiter Ordnung vorgenommen, um den auf einen sehr kleinen Bereich fokussierten Energieeintrag durch den Laser möglichst exakt widerzuspiegeln. Mittels Gittergenerator wurde ein dreidimensionales Netz aus 19123 Tetraedern erstellt, so dass sich 29347 Freiheitsgrade ergaben. Für die Berechnung des zeitabhängigen Verlaufs der Modellgrößen wurde der in Comsol Multiphysics implementierte Standardlöser *femtime* verwendet.

Im Rahmen der Arbeit wurde zunächst die Situation betrachtet, bei der der Laser nur einen regulären Scanvektor im Inneren des Quaders abfährt, sich also geradlinig von einem Rand des Werkstücks zum gegenüberliegenden Rand im Inneren bewegt.

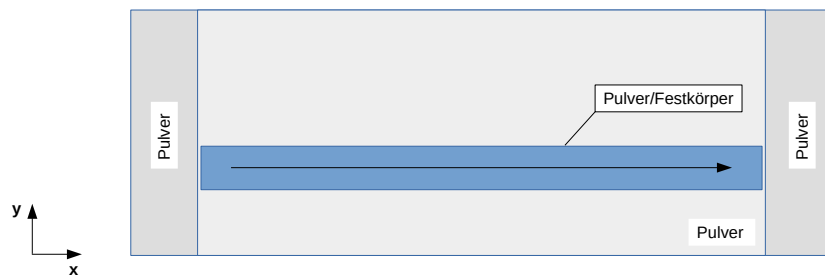


Abbildung 4.4.: Prinzipskizze zur Simulation eines Scanvektors

Start- und Endpunkt des Scanvektors sind  $82,5 \mu\text{m}$  vom Bauteilrand entfernt. Der Abstand in y-Richtung beträgt  $262,5 \mu\text{m}$  zum unteren Bauteilrand, um sicher zu stellen, dass keine unmittelbaren Randeefekte auftreten.

Bei der konkreten Wahl der Parameter wie im Modell 4.1 stellt man fest, dass die Kenngröße  $T_s$ , also die Temperatur in der Nähe des Lasers, nahezu konstant bleibt, auch wenn man sich dem rechten Rand des Bauteils und damit der angrenzenden Pulverschicht nähert. Die vom Projektpartner verwendeten Abstände zu den späteren Bauteilgrenzen sind so groß, dass das angrenzende Pulverbett noch keinen signifikanten Einfluss auf die Kenngröße hat. Infolgedessen muss eine Optimierung des Energieeintrags durch den Laser erst ab Beginn der Mäanderquerfahrt erfolgen.

### 4.3. Optimierung der Leistung des Lasers

Jede Optimierung zur Entwicklung angepasster Steuerungsstrategien startet a priori mit einer Simulation des Abfahrens des ersten regulären Scanvektors mittels *femtime*, wie in Kapitel 4.2 beschrieben. Wertet man die Temperaturkenngröße  $T_s$  (vgl. Kapitel 4.1) am Ende des Simulationsvorganges aus, so lässt sich diese Kenngröße als Referenz  $c_{melt}$  für die folgenden Scanvektoren mit optimierter Leistung heranziehen. Um das Aufschmelzen des Pulvers an der aktuellen Position des Lasers zu garantieren, darf im Rahmen der Optimierung die Laserleistung nur soweit abgesenkt werden, dass zu jedem Zeitpunkt noch die Schmelzbedingung

$$T_s \geq c_{melt}$$

eingehalten ist.

Ab der Mäanderfahrt wird deshalb das folgende, in Diskussion mit dem Projektpartner entstandene, zustandsbeschränkte Optimalsteuerungsproblem zur Berechnung optimaler Leistungssteuerungen  $u$  betrachtet:

$$\min J(T, u) = \frac{1}{2} \int_{t_0}^{t_f} \int_{B \subset \Omega} (\nabla T)^2 \, dV dt + \frac{\lambda}{2} \int_{t_0}^{t_f} u^2 \, dt$$

unter

$$\begin{aligned} \rho c_p \frac{\partial T}{\partial t} - \nabla \cdot (k \nabla T) &= u \phi && \text{in } \Omega \times (t_0, t_f), \\ k \frac{\partial T}{\partial n} &= -\alpha(T - 293) && \text{auf } \Gamma \times (t_0, t_f), \\ T(t_0) &= T_{init}, \\ \int_{\Omega} T \delta \, dV &\geq c_{melt} && \text{in } [t_0, t_f], \\ u_{min} &\leq u \leq u_{max} && \text{in } [t_0, t_f]. \end{aligned}$$

Dabei stellt  $T_{init}$  gerade die Temperaturverteilung nach Ablauf der Simulation der ersten Bahn und die Parameter  $u_{min} = 10 \, W$  und  $u_{max} = 180 \, W$  die minimal bzw. maximal mögliche Laserleistung dar. Ziel ist es, möglichst gleichmäßige Temperaturen in einem Bereich  $B$  am Rande des Werkstücks bei der Mäanderfahrt zu erreichen und dennoch das Pulver aufzuschmelzen. Infolgedessen wird im Unterschied zu den Betrachtungen in Teil 2 versucht, die Norm des Gradienten  $\nabla T$  statt die Norm von  $T$  zu minimieren. Der zweite Summand im Zielfunktional dient der Regularisierung des Problems ( $\lambda = 10^{-7}$ ).

Die Formulierung des Modells mittels linearer Wärmeleitungsgleichung ermöglicht es, das Optimalsteuerungsproblem analog zu dem Vorgehen in Kapitel 3.1.2 in eine reduzierte, endlichdimensionale Optimierungsaufgabe mit  $n$  Freiheitsgraden umzuformulieren. Als Approximation der zeitkontinuierlichen Steuerung sollte man in diesem Kontext aber  $u \approx \sum_{i=1}^n u_i \cdot u_{max} \cdot \varepsilon_i$  mit  $\vec{u} \in \mathbb{R}^n$  und den Indikatorfunktionen für das  $i$ -te Zeitintervall  $\varepsilon_i$  verwenden. Insofern lassen sich die Werte in  $\vec{u}$  als prozentuale Veränderung der

maximalen Laserleistung interpretieren. Die zugehörigen Steuerungsbeschränkungen lauten  $\frac{u_{\min}}{u_{\max}} \leq u_i \leq 1$  für  $i = 1, \dots, n$ . Die Definition der zum Aufstellen des quadratischen Problems nötigen Matrizen sind gemäß des veränderten Zielfunktionalis anzupassen. Es entsteht ein quadratische Problem der Form

$$\min_{\vec{u} \in \mathbb{R}^n} \frac{1}{2} \vec{u}^\top \mathbf{H} \vec{u} + \vec{f}^\top u \quad (4.2a)$$

unter den Nebenbedingungen

$$\mathbf{A} \vec{u} \leq \vec{b} \quad \text{und} \quad \frac{u_{\min}}{u_{\max}} \leq u_i \leq 1 \quad \text{für } i = 1, \dots, n. \quad (4.2b)$$

Um die Größen  $\mathbf{H}$ ,  $\vec{f}$ ,  $\mathbf{A}$  und  $\vec{b}$  möglichst effizient zu berechnen, geht man wie in Kapitel 3.2.1 beschrieben vor.

Auf den Maschinen des Projektpartners können mit dem in Kapitel 4.4 beschriebenen Ansatz nur stückweise konstante Steuerungen umgesetzt werden. Dabei ist die Anzahl der möglichen und betrachteten Freiheitsgrade  $k$  deutlich kleiner als die hinsichtlich guter Simulationsergebnisse gewählte Anzahl an Zeitschritten  $n$ . Das quadratische Problem ist zusätzlich um  $k$  Gleichungsbedingungen der Form

$$u_1 = u_2 = \dots = u_{n_1}, u_{n_1+1} = \dots = u_{n_2}, \dots, u_{n_{k-1}+1} = \dots = u_{n-1} = u_n \quad (4.2c)$$

zu erweitern. Die Indizes  $1 < n_1 < n_2 < \dots < n_{k-1} < n$  können unter Berücksichtigung gewisser technischer Vorgaben frei gewählt werden. Beispielsweise würde man für eine Steuerungssequenz aus 4 verschiedenen Leistungsparametern, bei der die Zeitpunkte der Anpassung der Laserleistung äquidistant verteilt sein sollen,  $n_1 \approx \frac{n}{4}$ ,  $n_2 \approx \frac{n}{2}$  und  $n_3 \approx \frac{3n}{4}$  setzen. Hierbei gilt ein wichtiger Punkt zu beachten: Bei einer Variation dieser zusätzlichen Gleichungsbedingungen an die Steuerung müssen die im quadratischen Problem (4.2) auftretenden und aufwendig zu berechnenden Größen  $\mathbf{H}$ ,  $\vec{f}$ ,  $\mathbf{A}$  und  $\vec{b}$  für die unterschiedlichen Steuerungsbeschränkungen nicht neu aufgestellt werden. So können sehr effizient unterschiedlichste Steuerungsstrategien berechnet werden, wenn nur die Zeitintervalle, auf denen die Laserleistung konstant sein soll, variiert werden. Dies hat sich im Rahmen der Entwicklung von Strategien für die realen Bauteile in Kapitel 4.5 als sehr hilfreich erwiesen.

Für die konkrete Optimierung der Laserleistung beim Abfahren der Bahnen in Mäanderform wird zur örtlichen Diskretisierung das gleiche Gitter und die selben FEM Ansatzfunktionen wie in der Simulation in Kapitel 4.2 verwendet. Der Hatchabstand parallel verlaufender Scanvektoren beträgt  $105 \mu m$ . Die zeitliche Diskretisierung erfolgt in Teilintervalle der Länge  $0,004375 ms$ , so dass für die Querfahrt im Mäander gerade 30 Zeitintervalle benötigt werden. Je nach Problem wird in der Optimierung maximal eine zweimalige Folge von Querfahrt im Mäander und Abfahren eines regulären Scanvektors betrachtet. Die resultierenden quadratischen Optimierungsprobleme haben maximal 2858 Freiheitsgrade ( $\vec{u} \in \mathbb{R}^{2858}$ ), wobei typischerweise viele Freiheitsgrade wegen der Gleichungsrestriktionen an die Steuerung eliminiert werden können.

Um die resultierenden quadratischen Probleme zu lösen, wurde die Optimierungsroutine *quadprog* in Matlab genutzt. Der in dem Kapitel 3.1.4 entwickelte PDASSP Algorithmus konnte aus den folgenden Gründen nicht verwendet werden:

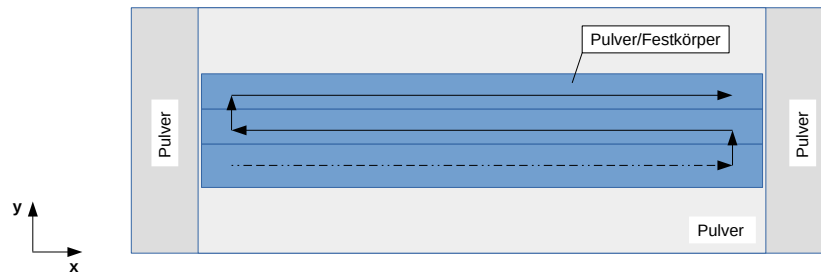


Abbildung 4.5.: Prinzipskizze zur Optimierung auf maximalem Zeithorizont mit vorausgehender Simulation (gestrichelte Linie)

1. Die theoretischen Resultate in Teil 2 dieser Arbeit wurden für ein Tracking-Type-Zielfunktional hergeleitet. Durch eine Betrachtung des Gradienten im Zielfunktional handelt man sich eine gewisse Irregularität ein. Die Auswirkung dieser Irregularität wäre entsprechend zu prüfen.
2. Der PDASSP Algorithmus in Kapitel 3.1.4 wurde als Erweiterung einer primal-dualen aktiven-Mengen-Strategie für eine einseitige Zustandsbeschränkung entwickelt. In Kunisch und Rösch [40], Ito und Kunisch [38] und Hintermüller [33] findet sich die primal-duale aktive-Mengen-Strategie für den Fall beidseitiger Steuerungsbeschränkungen. Eine Kombination von Steuerungsbeschränkungen und Zustandsbeschränkungen wurde nach Kenntnis des Autors noch nicht betrachtet. Bei numerischen Tests zeigte sich auch konkret das Problem, dass sich die Schätzungen für die verschiedenen aktiven Mengen während der Iterationen überschneiden können, so dass ein überbestimmtes, nicht lösbares Gleichungssystem entsteht.
3. Die Erweiterung um die zusätzlichen Gleichungsbeschränkungen an den Steuerungsvektor (4.2c) könnten nicht berücksichtigt werden.

#### 4.4. Umsetzung der Strategien auf der Anlage

Zur Umsetzung der berechneten Steuerungsstrategien für Produktion realer Bauteile auf den LaserCUSING<sup>®</sup> Anlagen ging man wie folgt vor: Für den ersten Scanvektor in jedem Island werden die Standardparameterwerte verwendet. Die weiteren Folgen aus Mäanderquerfahrt und regulärem Scanvektor in einem Island werden wiederholt mit der gleichen, angepassten Strategie belichtet. Um dieses Vorgehen auf den Anlagen zu realisieren, wurde die Softwareschnittstelle im CLI-Format [19] erweitert. In diesem Format werden die Maschinenparameter, wie z.B. Laserleistung und Geschwindigkeit, zu Beginn der Datei festgelegt und anschließend die abzufahrenden Scanvektoren notiert. Durch die Erweiterung ist es möglich, jedem einzelnen Scanvektor eine zugehörige Leistung oder Geschwindigkeit zuzuordnen. Dies bietet zwar eine Vielzahl von Möglichkeiten, führt aber umgekehrt zu deutlich umfangreicheren CLI Dateien und zu starken Redundanzen. Die Dateigröße nimmt signifikant zu, was u.a. Probleme beim Einlesen der Dateien nach sich zieht und umgekehrt die Anzahl der Scanvektoren und die möglichen

Bauteilgrößen und Steuerungsstrategien begrenzt. Des Weiteren könnte die Umsetzung einer Leistungssteuerung über die Softwareschnittstelle auch dafür verantwortlich sein, dass das Umschalten zwischen verschiedenen Leistungsstufen nicht sauber funktioniert. Näheres zu den Auswirkungen findet man bei der Auswertung der verschiedenen Produktionsergebnisse in den Kapiteln 4.5.1 und 4.5.2.

## 4.5. Strategien, Testbauteile und Auswertung

Im Rahmen dieser Arbeit werden verschiedene Ansätze für eine Anpassung der Laserleistung entwickelt und getestet. Zur Herleitung der Strategien legt man zunächst die Anzahl der möglichen Schaltvorgänge für die Leistung des Lasers und die Zeitintervalle, in denen die Leistung konstant ist, fest. Für diese gegebene Konfiguration wird das zugehörige Optimierungsproblem (4.2) gemäß Kapitel 4.3 gelöst und die angepassten Werte für die Leistung des Lasers berechnet.

### 4.5.1. Bauteile 1

In einer ersten Versuchsreihe zum Testen der Leistungssteuerung wurden  $1,5\text{ cm}$  hohe quaderförmige Bauteile aus dem Material 1.4404 gebaut, die aus einem einzelnen Island oder aus zwei aneinandergereihten Islands bestehen. Jedes der Bauteile besitzt einen  $450 \cdot 30\ \mu\text{m} = 1,35\text{ cm}$  hohen Sockel (450 Schichten, Schichthöhe  $30\ \mu\text{m}$ ), der im Standardverfahren hergestellt wird. Die Islands im Sockel werden zunächst mäanderförmig mit einer konstanten Leistung von  $180\text{ W}$  und einer Geschwindigkeit von  $800\text{ mm/s}$  abgefahren. Abschließend wird die Kontur mit gleicher Leistung und erhöhter Geschwindigkeit von  $1600\text{ mm/s}$  belichtet. Die letzten 50 Schichten eines Bauteils werden nach der jeweils untersuchten, angepassten Strategie aufgebaut. Da gerade die Effekte am Rand eines Bauteils, wie z.B. des Verhalten der Schmelzspur, von Interesse sind, wird bei den oberen 50 Schichten stets auf eine Konturfahrt verzichtet.

In der ersten Versuchsreihe wurden die folgenden 3 Bauteilvariationen untersucht:

**Bauteil 1a: Quader mit quadratischer Grundfläche:** Die Grundfläche besteht aus einem Quadrat mit Kantenlänge  $5060\ \mu\text{m}$ . Diese Maße ergeben sich aus der Kantenlänge eines Standardislands  $5\text{ mm}$  und einer Kontur (im Sockel) von jeweils  $30\ \mu\text{m}$ .

**Bauteil 1b: Quader mit rechteckiger Grundfläche  $10060\ \mu\text{m} \times 5060\ \mu\text{m}$ :** Zwei Islands werden nebeneinander positioniert und (im Sockel) mit einer gemeinsamen Kontur versehen. Hiermit soll untersucht werden, inwiefern sich ein bereits belichtetes Island auf die Qualität der Schmelzspur im Nachbarisland auswirkt. Es stellt sich die Frage, ob es signifikante Unterschiede zwischen einem inneren Rand (zwei Islands treffen aufeinander) und einem äußeren Bauteilrand (Übergang Festkörper zu Pulverbett) gibt.

**Bauteil 1c: Quader mit rechteckiger Grundfläche  $10120\ \mu\text{m} \times 5060\ \mu\text{m}$ :** Zwei Quader mit quadratischer Grundfläche ( $5060\ \mu\text{m} \times 5060\ \mu\text{m}$ ) werden direkt nebeneinander positioniert. Bei der Belichtung in den obersten 50 Schichten werden aber nur

jeweils Standardislands der Größe  $50\text{ mm} \times 50\text{ mm}$  abgefahren. Es finden keinerlei Konturfahrten statt, so dass sich ein kleiner Spalt von  $60\text{ }\mu\text{m}$  Breite zwischen den Islands befindet. Anhand dieses Bauteils kann leicht untersucht werden, ob vom Laser tatsächlich nur das angedachte Island aufgeschmolzen wird oder ob die Schmelze auch in den Bereich außerhalb der Islands fließt.

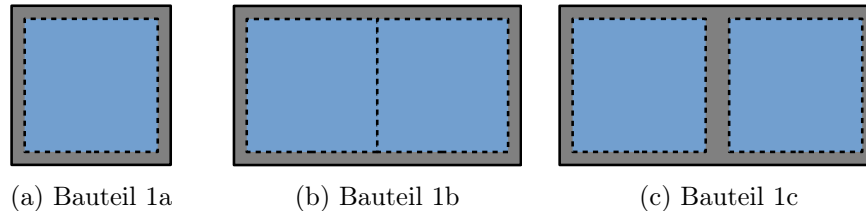


Abbildung 4.6.: Grundflächen der quaderförmigen Bauteile

Da bisher noch keinerlei Erfahrung mit der Möglichkeit der Leistungsanpassung über die Softwareschnittstelle vorhanden war, sollten in der ersten Testreihe nur maximal vier Änderungen der Leistung pro abgefahrenem Scanvektor vorgenommen werden. Die getesteten Strategien waren:

1. Standardverfahren: Abfahren eines Islands in Mäanderform bei konstanter Laserleistung  $180\text{ W}$ .
2. Verfahren mit Laserabschaltung: Abfahren der Scanvektoren grundsätzlich mit konstanter Leistung  $180\text{ W}$ , aber Abschalten des Lasers bei Querfahrt im Mäander.
3. Verfahren mit Leistungssteuerung: Jeweils der erste Scanvektor in einem Island wird mit der Standardleistung  $180\text{ W}$  des Lasers belichtet. Die Laserleistung wird ab der Querfahrt im Mäander bis zum erneuten Erreichen der Mitte des Islands angepasst. Hierzu wird diese Strecke äquidistant in zwei oder drei Teile geteilt und für diese Teilstücke ein niedrigerer, optimierter Wert für die Leistung des Lasers in der CLI Datei hinterlegt. Der Rest eines Scanvektors, von der Mitte bis zum neuen Startpunkt einer Mäanderfahrt (Rand des Bauteils), wird mit dem Standardparameter zurückgelegt, so dass sich insgesamt drei oder vier unterschiedliche Leistungsstufen ergeben. Das Vorgehen wird über das gesamte Island hinweg wiederholt.

Des Weiteren konnten die Belichtungsstrategien noch dadurch variiert werden, dass entweder jede Schicht identisch belichtet wurde (z.B. überwiegend nur x-Richtung) oder die Vorzugsrichtung, wie im Standardverfahren üblich, von Schicht zu Schicht gewechselt wurde (alternierend x-/y-Richtung).

Die optimalen Werte im Verfahren mit Leistungssteuerung wurden gewonnen, indem man das Optimierungsproblem (4.2) ab der ersten Mäanderquerfahrt bis zur Mitte des rückläufigen Scanvektors für zwei bzw. drei Freiheitsgrade mit einer äquidistanten Verteilung löst. Es ergeben sich die angepassten Leistungsstufen in Tabelle 4.2.

Jedes der obigen Bauteile wurde mit ausgewählten Belichtungsstrategien kombiniert und, um stichhaltige Aussagen treffen zu können, mehrfach gebaut. Anschließend wurden

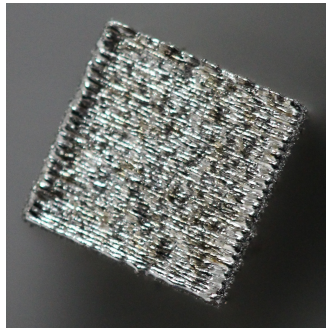


Teilintervall	1	2	3
Strategie mit 4 Leistungsstufen	123	130	144
Strategie mit 3 Leistungsstufen	123	144	

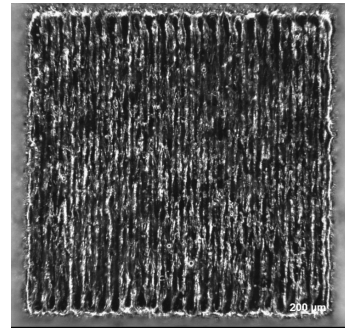
Tabelle 4.2.: Optimierte Laserleistung in  $W$ 

Makroaufnahmen der Bauteile erstellt und ausgewählte Werkstücke mit einem Mikroskop untersucht. Im Folgenden sollen die wichtigsten Ergebnisse hieraus dargestellt werden:

Beim Standardverfahren kommen, wie zu Beginn des Kapitels erwähnt, die Probleme der Nasenbildung und der nicht gleichmäßigen Ausprägung der Schmelzspuren zum Vorschein. Der Vollständigkeit halber seien an dieser Stelle nochmals die zugehörigen Aufnahmen in Abbildung 4.7 aufgeführt.



(a) Makroaufnahme



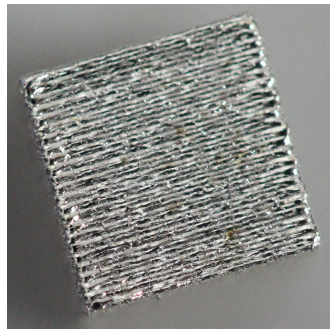
(b) Mikroskopaufnahme

Abbildung 4.7.: Bauteil 1a mit Standardverfahren und alternierender Belichtung

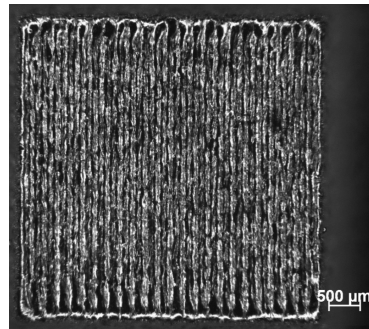
Das teilweise beim Kooperationspartner bereits zur Erreichung besserer Qualität eingesetzte Verfahren mit Laserabschaltung zeigt zwar weniger Nasenbildung, weist aber immer noch das Problem auf, dass beim Zurückfahren vom Rand des Bauteils ins Innere zu viel Energie eingetragen wird. Auch hier tritt am Rand jede zweite Schmelzspur deutlich stärker hervor (Abbildung 4.8).

Die ausgewählten Strategien mit optimierter Leistungsabsenkung (drei oder vier Leistungsstufen) vermeiden erfolgreich die oben genannten Probleme. Nasenbildung am Rand tritt kaum noch auf und die Schmelzspuren am Rand sind gleichmäßig ausgeprägt (Abbildung 4.9).

Insofern scheint die Leistungsabsenkung die Oberfläche des Werkstücks deutlich zu verbessern. Bei der Betrachtung der Schmelzspuren im Inneren der Island erkennt man jedoch deutlich mehr Ungleichmäßigkeiten im Vergleich zu den Standardverfahren. Die Ursache dieser Unsauberkeiten wird am besten in Mikroskopaufnahmen eines Werkstücks sichtbar, bei dem nicht wie üblich die Belichtungsrichtung alterniert wurde, sondern stets in eine Vorzugsrichtung belichtet wurde (Abbildung 4.10). An den Stellen, an denen sich die Leistung des Lasers ändert, treten große Unsauberkeiten auf. Diese sind insbesondere in der Mitte des Islands deutlich sichtbar, da dort nach der gewählten Strategie in jedem Scanvektor eine Änderung der Leistung auftritt. In der weiteren Entwicklung

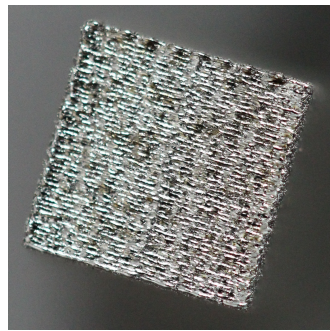


(a) Makroaufnahme

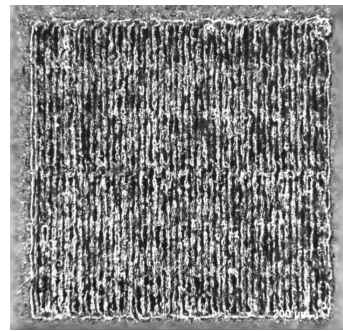


(b) Mikroskopaufnahme

Abbildung 4.8.: Bauteil 1a mit Verfahren mit Leistungsabschaltung und alternierender Belichtung



(a) Makroaufnahme



(b) Mikroskopaufnahme

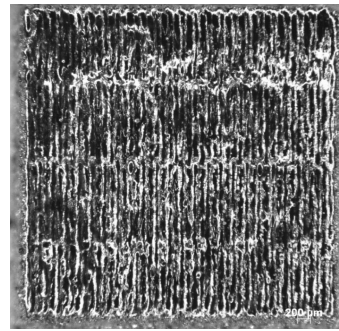
Abbildung 4.9.: Bauteil 1a mit Verfahren mit vier Leistungsstufen und alternierender Belichtung

einer Belichtungsstrategie sollte deshalb darauf geachtet werden, dass bei den hin- und rückläufigen Scanvektoren die Umschaltpunkte der Leistung nicht mehr zusammenfallen. Zum anderen könnte man vermuten, dass die Unsauberkeiten beim Erhöhen der Laserleistung durch zu große Leistungsänderungen entstehen.

Wie oben bereits aufgeführt, wurden auch Bauteile mit quaderförmiger Grundfläche untersucht. Betrachtet man die Ergebnisse zu den Bauteilen 1b in der Abbildung 4.11, so werden für das Standardverfahren wieder die typischen Probleme ersichtlich: Es treten die Nasenbildung und die ungleichmäßigen Schmelzspuren in den Umkehrpunkten auf. Man erkennt aber auch, dass es diesbezüglich keinen signifikanten Unterschied zwischen einem äußeren Rand (Übergang Pulverbett Werkstück) und einem inneren Rand (Aneinandertreffen mit benachbartem Island) eines Islands gibt. Gleiches Verhalten zeigte sich im Übrigen auch bei den Verfahren mit Leistungssteuerung. Dies ist ein starkes Indiz dafür, dass die Erhöhung der Temperatur in den Umkehrpunkten und die damit verursachten Unsauberkeiten in der Oberfläche fast ausschließlich von der Tatsache her rühren, dass der Laser umkehrt und somit lokal in den Umkehrpunkten den schon bei der Hinfahrt erhitzten Bereich noch stärker erwärmt. Die isolierende Wirkung des Pul-

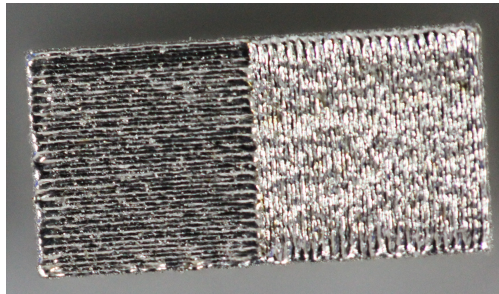


(a) Mikroskopaufnahme 1

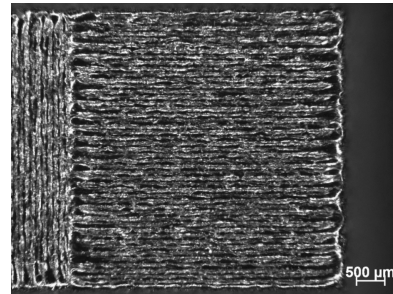


(b) Mikroskopaufnahme 2

Abbildung 4.10.: Bauteil 1a bei dauerhafter Belichtung in x-Richtung mit Verfahren mit drei Leistungsstufen



(a) Makroaufnahme



(b) Mikroskopaufnahme

Abbildung 4.11.: Bauteil 1b mit Standardverfahren

verbetts, was ja an einem äußeren Rand gegeben ist, spielt eine vernachlässigbare Rolle. Dies bestätigt die im Rahmen der Simulation erhaltenen Ergebnisse. Als Konsequenz wird im Fortgang dieser Arbeit eine Unterscheidung nach äußerem und innerem Rand nicht mehr getroffen. Die neu entwickelten Strategien sollen bei beiden Rändern identisch angewendet werden.

Die Tatsache, dass die Qualität der Ausprägung eines Islandrandes bei den Strategien mit Leistungsanpassung deutlich höher ist als beim Standardverfahren, wird bereits makroskopisch mit den Bauteilen 1c in Abbildung 4.12 sichtbar. Bei den Verfahren mit angepasster Leistung ist der schmale Spalt zwischen den beiden Islands wie gewünscht vorhanden. Beim Standardverfahren hingegen fließt die Schmelze auch in den Spalt, so dass dieser nicht mehr ersichtlich ist.

#### 4.5.2. Bauteile 2

Basierend auf den Ergebnissen der Baureihe 1 sollen in dieser Arbeit weitere Verfahren mit Leistungssteuerung entwickelt werden. Ziel ist es dabei vor allem, eine gute Qualität am Rande des Bauteils zu sichern und zusätzlich die Unsauberkeiten der Schmelzspuren im Inneren zu vermeiden. Die grundsätzliche Idee hierzu ist, entweder die Anzahl der

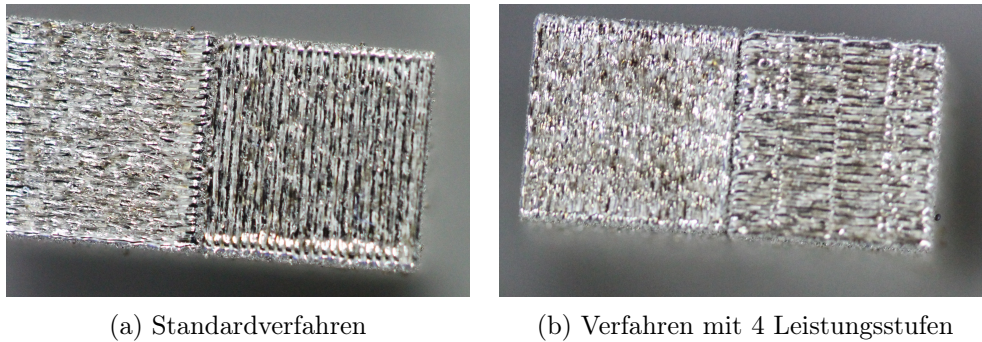


Abbildung 4.12.: Makroaufnahmen zu Bauteil 1c

Leistungsstufen zu erhöhen, um die Sprünge in den Leistungswerten möglichst gering zu halten, oder mit wenigen Schaltzeitpunkten im Inneren auszukommen. Es wurden deshalb die folgenden Verfahren anhand der Produktion von Bauteilen der Form 1a (Kapitel 4.5.1) untersucht.

1. Verfahren mit vielen Leistungsstufen: Beginnend ab einer Mäanderquerfahrt bis einschließlich des kompletten Abfahrens eines Scanvektors finden viele (mindestens fünf) Leistungsanpassungen statt. Dabei sollen im Vergleich zur Baureihe 1 die Positionen der Leistungsänderungen nicht mehr äquidistant entlang eines Scanvektors verteilt werden. Stattdessen werden sie basierend auf einer Betrachtung des Verhaltens der Temperaturkenngröße  $T_s$  beim Durchlaufen der Scanvektoren mit konstanter Leistung gewählt. Dies soll gewährleisten, dass zum einen Leistungsänderungen bei direkt aufeinanderfolgenden Scanvektoren nicht mehr an der selben Stelle auftreten (wie z.B. in der Islandmitte bei Baureihe 1) und zum anderen die Leistungsänderungen möglichst gleichmäßig und gering ausfallen.
2. Verfahren mit zwei Leistungsstufen: Beginnend ab einer Mäanderquerfahrt bis einschließlich des kompletten Abfahrens eines Scanvektors wird die Leistung des Lasers nur zu einem Zeitpunkt geändert. Die beiden Leistungswerte auf den resultierenden Teilstücken sollen aus einer Optimierung hervorgehen. Hierbei sind zwei Varianten denkbar.
  - a) Für die Mäanderquerfahrt und für die komplette Fahrt entlang eines regulären Scanvektors werden zwei separate Werte hinterlegt. Der Umschaltzeitpunkt ist zum Ende der Mäanderquerfahrt. Diese Variante ist dem Verfahren mit Leistungsabschaltung aus Testreihe 1 ähnlich, unterscheidet sich aber durch die optimierten Leistungswerte auf den Teilstücken.
  - b) Der Umschaltzeitpunkt wird nicht zum Ende der Mäanderquerfahrt, sondern zu einem Zeitpunkt gesetzt, bei der sich der Laser bereits leicht innerhalb des regulären Scanvektors der Rückfahrt befindet. Diese Idee basiert auf den Simulationsergebnissen zum Verhalten der Schmelzkenngröße  $T_s$  während der Mäanderfahrt und den Betrachtungen zu den Schmelzspuren in Kapitel 4.5.1. Übermäßig hohe Temperaturen und breite Schmelzspuren treten nicht nur

bei der Mäanderquerfahrt, sondern auch noch auf einem Teil der Rückfahrt des Lasers auf. Entsprechend sollte in einem ersten Teil der Rückfahrt die Leistung des Lasers noch reduziert sein.

Die optimalen Werte für die verschiedenen Verfahren wurden gewonnen, indem das Optimierungsproblem (4.2) für eine Folge von Querfahrt im Mäander und Abfahren eines regulären Scanvektors gelöst wurde. Für einige Strategien wurden auch zwei derartige Folgen untersucht. Aus diesen Betrachtungen ergeben sich zwei mögliche Leistungsanpassungsstrategien. Die zur ersten Folge aus Mäanderquerfahrt und vollständiger Rückfahrt gehörigen Werte und die für die zweite Sequenz berechneten Parameterwerte. Von diesen beiden Varianten wurde a priori eine Version als zu testenden Strategie für das komplette Island ausgewählt. Die Produktion der Bauteile und die Umsetzung der berechneten Leistungssteuerungsstrategien erfolgte wie in der ersten Baureihe. Zur Untersuchung wurden neben Makro- und Mikroskopaufnahmen auch Schliiffbilder herangezogen. Im Folgenden wird auf die wichtigsten Ergebnisse eingegangen.

Exemplarisch für die Verfahren, die aus mehreren, nicht äquidistant verteilten Leistungsstufen bestehen, seien an dieser Stelle für eine Strategie mit 5 Parameterwerten die Laserleistung und die Positionen, zu denen ein Umschalten stattfinden muss, aufgeführt (Tabelle 4.3). Die Angabe der Positionen erfolgt durch Angabe des Anteils der von der Beginn der Mäanderquerfahrt bis zur Schaltposition zurückgelegten Strecke  $d$  im Vergleich zur Gesamtstrecke  $d_{\text{ges}}$ , bestehend aus Mäanderquerfahrt und rückläufigem Scanvektor.

Schaltpunkt	1	2	3	4	5
$\frac{d}{d_{\text{ges}}}$ in %	0	2,08	15,7	33,3	58,0
Leistung in $W$	113	120	128	145	171

Tabelle 4.3.: Optimierte Laserleistung sowie relative Schaltpositionen für die Strategie mit 5 Parameterwerten

Die Ergebnisse dieser Strategie und von Strategien mit 11 und 26 Leistungsstufen sind in Abbildung 4.13 dargestellt.

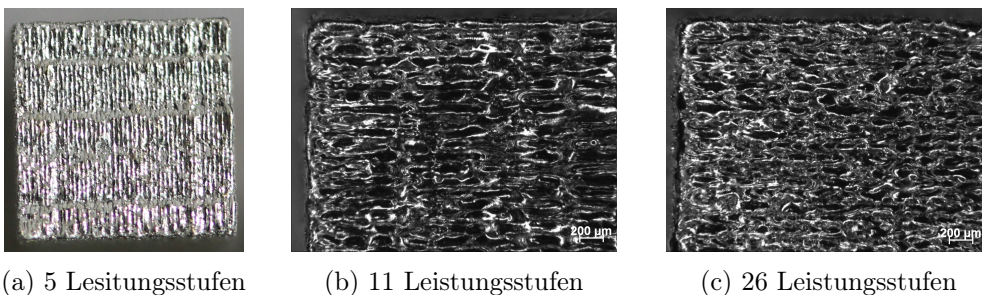


Abbildung 4.13.: Verfahren mit mehreren Leistungsstufen

Anhand der Aufnahmen stellt man fest, dass der gewünschte Effekt eines gleichmäßigen Schmelzbildes nicht eintritt. Es treten bei dem Verfahren mit fünf Leistungsan-

passungen immer noch deutlich erkennbare Stufen in der Makroaufnahme auf. An dieser Stelle könnte man zunächst noch die Vermutung äußern, dass dies nach wie vor daran liegt, dass die Leistungsänderungen zu groß ausfallen. Durch das Erhöhen der Anzahl der Leistungsstufen auf 11 und vor allem 26 werden die Unterschiede in den Leistungswerten allerdings so gering, dass diese Ursache hierfür nicht mehr in Frage kommt. Nichtsdestotrotz zeigt sich in den Mikroskopaufnahmen zu den Verfahren mit mehr Leistungsanpassungen ebenfalls keine gleichmäßige Schmelzspur. Ganz im Gegenteil tritt mit zunehmender Anzahl an Stellen, an denen die Leistung geändert wird, ein immer unruhigeres Schmelzbild auf. Hieraus lässt sich die Vermutung äußern, dass es noch ein grundlegendes Problem bei der Anpassung der Laserleistung mittels Software-schnittstelle gibt. Die Anpassungen der Laserleistung laufen noch nicht wie gewünscht ab. Möglicherweise treten Latenzzeiten beim Umschalten auf. Um diese Tatsache genauer zu untersuchen, sind jedoch sehr aufwendige Maßnahmen nötig, die im Rahmen der Arbeit nicht mehr geleistet werden konnten.

Für die Verfahren mit zwei optimierten Leistungsstufen sind die Ergebnisse der Optimierung für zwei Folgen aus Mäanderquerfahrt und regulärem Scanvektor in den nachfolgenden Tabellen 4.4 und 4.5 dargestellt.

Schaltzeitpunkt	1	2
$\frac{d}{d_{ges}}$ in %	0	2,08
Leistung in $W$ für erste Folge	125	168
Leistung in $W$ für zweite Folge	64	167

Tabelle 4.4.: Verfahren 2a mit Umschaltzeitpunkt am Ende der Mäanderquerfahrt

Schaltzeitpunkt	1	2
$\frac{d}{d_{ges}}$ in %	0	8,40
Leistung in $W$ für erste Folge	139	168
Leistung in $W$ für zweite Folge	106	168

Tabelle 4.5.: Verfahren 2b mit Umschaltzeitpunkt innerhalb des regulären Scanvektors

Da sich die erhaltenen Resultate im Rahmen der verwendeten Untersuchungstechniken vielmehr durch die grundsätzliche Strategie als durch die Parameterwerte für die erste bzw. zweite Folge aus Mäanderquerfahrt und vollständiger Rückfahrt unterscheiden, wird in den folgenden Darstellungen auf diese Subunterscheidung verzichtet.

Anhand der Aufnahmen in den Abbildungen 4.14 und 4.15 wird deutlich ersichtlich, dass durch Anpassung der Laserleistung bei der Mäanderquerfahrt der Effekt der Nasenbildung im Vergleich zum Standardverfahren reduziert werden kann. Bei Betrachten der Ergebnisse des Verfahrens 2a mit Leistungsreduktion bei der Mäanderquerfahrt zeigt sich aber auch, dass wie beim Standardverfahren in der Nähe des Randes für eine gewisse Strecke jede zweite Schmelzspur stärker ausgeprägt ist. Es wird beim Zurückfahren noch zu viel Energie eingetragen. Dieser Effekt kann durch das Verfahren 2b, das eine Leistungsreduktion auch auf einem kurzen Stück der Rückfahrt vorsieht, erfolgreich umgan-

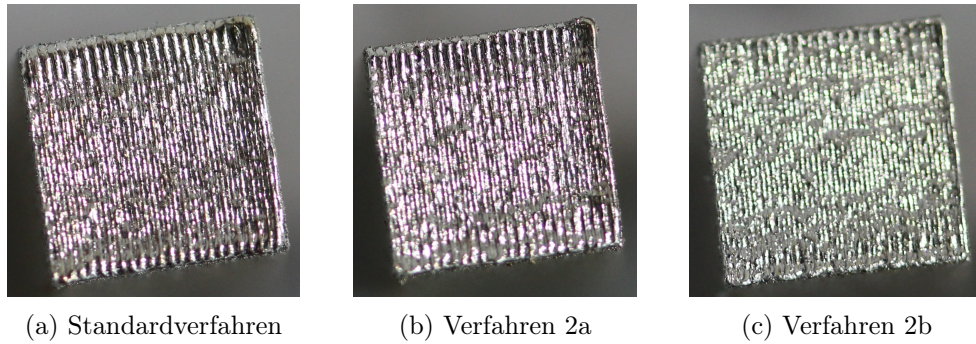


Abbildung 4.14.: Makroaufnahmen

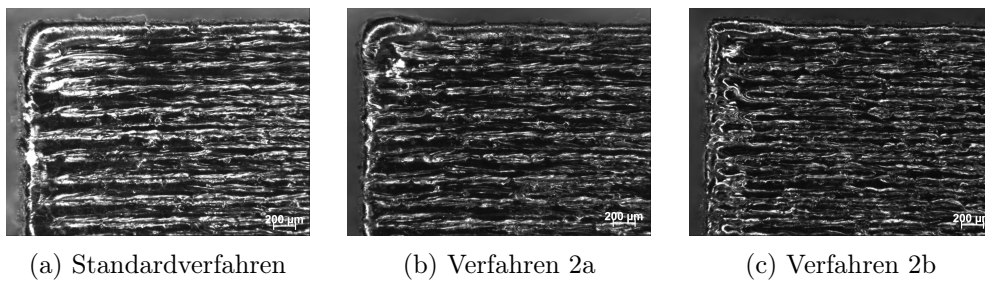


Abbildung 4.15.: Mikroskopaufnahmen

gen werden. Die Schmelzspuren sind auf der kompletten Inselfläche nahezu gleichmäßig ausgeprägt.

Zur abschließenden Bewertung der entwickelten Strategien, insbesondere der vielversprechenden Strategie 2b, ist es wichtig, dass die mit den Verfahren produzierten Bauteile noch die gewünschte Dichte aufweisen. Dies lässt sich anhand von Schliffbildern überprüfen (Abbildung 4.16). Das Standardverfahren weist wie gewünscht nur eine geringe

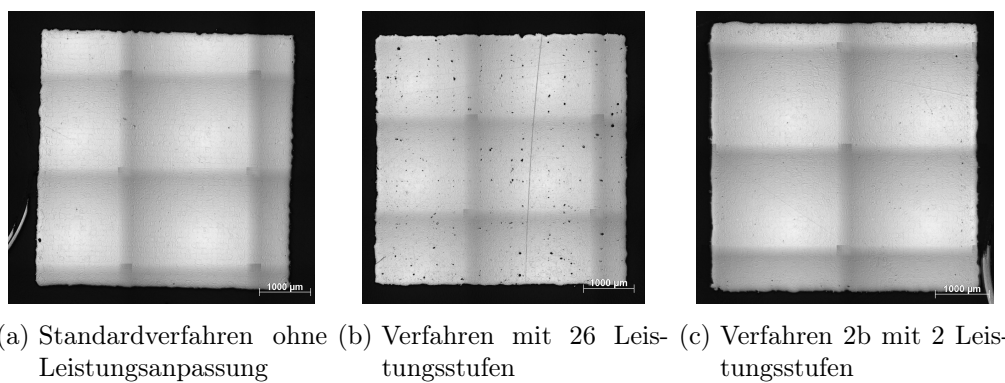


Abbildung 4.16.: Schliffbilder

Porosität auf, was sich an der geringen Anzahl an schwarzen Punkten im Schliffbild zeigt. Beim Verfahren mit 26 Leistungsänderungen führt das unruhige Schmelzbild (Abbildung

4.13c) auch zu einer sehr hohen Porosität. Die Strategie 2b mit zwei Leistungsstufen und einem Umschaltzeitpunkt im Inneren zeigt im Rahmen dieser Betrachtung eine vergleichbar niedrige Porosität wie das Standardverfahren.

Insofern konnte im Rahmen dieser Arbeit ein einfaches, gut umsetzbares Verfahren entwickelt werden, dass eine deutlich bessere Rand- und Oberflächenqualität bei vergleichbarer Porosität garantiert.



# A. Anhang

**Satz A.0.1.** Seien  $[0, T] \subset \mathbb{R}$  ein beschränktes Intervall,  $V \subseteq H \subseteq V^*$  ein Gelfand-Dreier und  $a: V \times V \rightarrow \mathbb{R}$  eine beschränkte Bilinearform, die die Gardingsche Ungleichung erfüllt, also:

$$\begin{aligned} |a(u, v)| &\leq \alpha_0 \|v\|_V \|\tilde{v}\|_V \quad \forall v, \tilde{v} \in V, \\ a(v, v) &\geq \beta \|v\|_V^2 - \beta_0 \|v\|_H^2 \quad \forall v \in V \end{aligned}$$

mit Konstanten  $\alpha_0 \geq 0, \beta > 0, \beta_0 \geq 0$ . Sei  $A: V \rightarrow V^*$  der durch

$$\langle Av, \tilde{v} \rangle_{V^*, V} = a(v, \tilde{v}) \quad \forall v, \tilde{v} \in H^1(\Omega)$$

definierte zugehörige linear beschränkte Operator. Dann lassen sich für die Evolutionsgleichung

$$y'(t) + Ay(t) = f(t) \cdot G(t) \quad f.\ddot{u}. \text{ in } [0, T], \quad (\text{A.1a})$$

$$y(0) = y_0 \quad (\text{A.1b})$$

mit Funktionen  $f: [0, T] \rightarrow \mathbb{R}, G: [0, T] \rightarrow V^*, g: [0, T] \rightarrow H$  mit

$$\langle G(t), v \rangle_{V^*, V} := \int_{\Omega} g(t)v \, dx \quad \forall v \in V$$

und  $y_0 \in H$  die folgenden Aussagen treffen:

- a) Die Evolutionsgleichung besitzt für jedes  $f \in L^2(0, T), g \in L^\infty(0, T; H)$  eine eindeutige Lösung

$$y \in W_{V, V^*}(0, T) := \{\hat{y} \in L^2(0, T; V) \mid \hat{y}' \in L^2(0, T; V^*)\}.$$

- b) Gilt  $f \in BV(0, T), g \in W_{V, V^*}(0, T)$  und  $y_0 \in D_H(A) := \{v \in V \mid Av \in H\}$ , so besitzt die eindeutige Lösung der Evolutionsgleichung die erhöhte Regularität

$$y \in L^2(0, T; V) \quad \text{mit} \quad y' \in L^2(0, T; V) \cap BV(0, T; V^*).$$

Außerdem gilt  $Ay \in W_{V, V^*}(0, T)$  als Lösung der Evolutionsgleichung

$$(Ay)'(t) + A(Ay)(t) = f(t) \cdot Ag(t) \quad f.\ddot{u}. \text{ in } [0, T], \quad (\text{A.2a})$$

$$Ay(0) = Ay_0. \quad (\text{A.2b})$$

**Bemerkung A.0.1.** Die Aussage,  $Ay \in W_{V, V^*}(0, T)$  als Lösung einer Evolutionsgleichung, ist wie folgt zu verstehen: Fast überall in  $[0, T]$  wird  $Ay$  von einer Funktion  $f_{Ay} \in W_{V, V^*}(0, T)$  erzeugt, also  $\langle Ay(t), v \rangle_{V^*, V} = \int_{\Omega} f_{Ay}(t)v \, dx$  für alle  $v \in V$ .

*Beweis.* Zum Beweis der Teilaussage a) verwendet man, dass nach der Definition von  $G$  für  $g \in L^\infty(0, T; H)$  die rechte Seite der Evolutionsgleichung (A.1),  $f \cdot G$ , wegen

$$\|fG\|_{L^2(0, T; V^*)}^2 \leq \int_0^T f(t)^2 \cdot \|g(t)\|_H^2 dt \leq \|g\|_{L^\infty(0, T; H)}^2 \|f\|_{L^2(0, T)}^2$$

in  $L^2(0, T; V^*)$  liegt. Die Existenz-, Eindeutigkeits- und Regularitätsaussage für die Lösung  $y$  der Evolutionsgleichung folgt für  $y_0 \in H$  unmittelbar aus klassischen Resultaten in Gajewski, Gröger und Zacharias [23] oder Emmrich [21]. Für den Beweis der Aussage b) soll eine additive Zerlegung der Lösung  $y = y_1 + y_2$  vorgenommen werden. Zunächst setzt man  $y_1 : [0, T] \rightarrow V$  als

$$y_1(t) := \int_0^t f(\tau)g(\tau) d\tau.$$

Da jede Funktion  $f$  beschränkter Variation auf  $[0, T]$  beschränkt ist und  $g$  in  $L^2(0, T; V)$  liegt, hat man  $f \cdot g \in L^2(0, T; V)$  und somit  $y_1 \in AC^{1,2}([0, T]; V)$  mit

$$y_1'(t) = f(t)g(t) \quad \text{f.ü. in } [0, T].$$

Dabei bezeichnet  $AC^{1,2}([0, T]; X)$  für einen Banachraum  $X$  den Raum aller absolut stetigen Funktionen  $[0, T] \rightarrow X$ , die eine distributionelle Ableitung in  $L^2(0, T; X)$  besitzen. Des Weiteren kann nach Barbu und Precupanu [4] jedes  $g \in W_{V, V^*}(0, T)$  wegen  $V \subseteq H \subseteq V^*$  insbesondere als eine absolut stetige Funktion mit Bild in  $V^*$  aufgefasst werden,  $g \in AC^{1,2}([0, T]; V^*)$ , so dass  $f \cdot g \in BV(0, T; V^*)$  für  $f \in BV(0, T)$  gilt. Man hat also:

$$y_1 \in L^2(0, T; V) \text{ und } y_1' \in L^2(0, T; V) \cap BV(0, T; V^*). \quad (\text{A.3})$$

Aus der Definition von  $y_1$  folgt mit dem gewählten Ansatz, dass  $y_2$  für  $g \in L^2(0, T; V)$  fast überall in  $[0, T]$  die Gleichung

$$(y_1 + y_2)'(t) + A(y_1 + y_2)(t) = f(t) \cdot G(t) \quad (\text{A.4a})$$

$$y_1'(t) \stackrel{= f(t)g(t)}{\Leftrightarrow} y_2'(t) + Ay_2(t) = -Ay_1(t) \quad (\text{A.4b})$$

$$= - \int_0^t f(\tau)Ag(\tau) d\tau \quad (\text{A.4c})$$

erfüllt, wobei die rechte Seite (mit der analogen Begründung wie oben für  $f \cdot Ag$  statt  $f \cdot g$ ) dem Raum  $AC^{1,2}([0, T]; V^*)$  angehört. Des Weiteren hat man

$$y_2(0) = y_0 - \underbrace{y_1(0)}_{=0} = y_0 \in D_H(A).$$

Somit gilt  $y_2 \in W_{V, V^*}(0, T)$  als Lösung einer Evolutionsgleichung mit den entsprechend regulären Daten. Es stellt sich nun die Frage, inwiefern man auch für die Ableitung  $y_2'$  erhöhte Regularität herleiten kann. Die Regularität der rechten Seite legt nahe, dass man die nach der Zeit differenzierte Evolutionsgleichung betrachten kann und somit  $y_2'$  wieder

selbst Lösung einer Evolutionsgleichung ist, wenn man auch die erhöhte Regularität des Anfangswertes zur Verfügung hat. Das genaue Vorgehen ist den Ausarbeitungen in Renardy und Rogers [50, S. 387 f.] bzw. Wloka [55, S. 393 ff.] nachempfunden.

Man definiere sich zunächst  $\nu$  als Lösung der nach der Zeit differenzierten Gleichung, also als Lösung von

$$\nu'(t) + A\nu(t) = -f(t)Ag(t) \quad \text{f.ü. in } [0, T], \quad (\text{A.5a})$$

$$\nu(0) = -Ay_0. \quad (\text{A.5b})$$

Wegen  $y_0 \in D_H(A)$  hat man für den Anfangswert  $-Ay_0 \in H$ . Zudem liegt die rechte Seite  $f \cdot Ag$  in  $L^2(0, T; V^*)$ , so dass für die eindeutige Lösung  $\nu$  gilt:  $\nu \in W_{V, V^*}(0, T)$ . Setzt man

$$w(t) := y_0 + \int_0^t v(\tau) \, d\tau,$$

so erhält man  $w \in AC^{1,2}([0, T]; V)$  und  $w'(t) = \nu(t)$  fast überall in  $[0, T]$  sowie  $w(0) = y_0$ . Außerdem gilt mit der Gleichung (A.5) fast überall in  $[0, T]$ :

$$\begin{aligned} Aw(t) &= Ay_0 + \int_0^t Av(\tau) \, d\tau \\ &= Ay_0 + \int_0^t -f(\tau)Ag(\tau) - \nu'(\tau) \, d\tau \\ &= -\nu(t) + \int_0^t -f(\tau)Ag(\tau) \, d\tau \\ &= -w'(t) - \int_0^t f(\tau)Ag(\tau) \, d\tau. \end{aligned}$$

Somit erfüllen  $y_2$  und  $w$  die gleiche Evolutionsgleichung. Wegen der Eindeutigkeit der Lösung dieser Gleichung folgt  $y_2 = w$  und somit  $y_2' = \nu$ . Insgesamt ergibt sich

$$y_2 \in L^2(0, T; V) \text{ und } y_2' \in W_{V, V^*}(0, T),$$

also insbesondere  $y_2' \in AC^{1,2}([0, T]; V^*)$  [4], so dass mit dem Regularitätsresultat (A.3) für die eindeutige Lösung  $y$  der Evolutionsgleichung (A.1) gelten muss:

$$y = y_1 + y_2 \in L^2(0, T; V) \text{ und } y' \in L^2(0, T; V) \cap BV(0, T; V^*).$$

Bei Betrachten der Beziehung (A.4) erkennt man, dass die Funktion  $y_2$  mit  $Ay$  in Zusammenhang steht:

$$y_2'(t) = -A(y_1(t) + y_2(t)) = -Ay(t) \quad \text{f.ü. in } [0, T].$$

Also ist  $Ay$  wegen  $y_2' = \nu$  mit Gleichung (A.5) als eindeutige Lösung von

$$\begin{aligned} (Ay)'(t) + A(Ay)(t) &= f(t)Ag(t) \quad \text{f.ü. in } [0, T], \\ Ay(0) &= Ay_0 \end{aligned}$$

gegeben und besitzt die Regularität  $Ay \in W_{V, V^*}(0, T)$ . □

**Bemerkung A.0.2.** Die Regularität von  $Ay$  im vorhergehenden Satz ließe sich aufgrund der Struktur der Gleichung (A.2) noch beliebig weiter erhöhen, wenn man die erhöhten Regularitätsbedingungen an  $y_0$  und  $g$  sukzessive für  $Ay_0$ ,  $Ag$  usw. fordern würde.

**Satz A.0.2.** Seien  $[0, T] \subset \mathbb{R}$  ein beschränktes Intervall,  $X$  ein Banachraum und

$$A: D(A) \subset X \rightarrow X$$

ein abgeschlossener, dicht definierter Operator mit Definitionsmenge  $D(A)$  versehen mit der Graphennorm. Außerdem besitze  $A$  maximale parabolische  $L^2(0, T; X)$  Regularität. Zu  $f \in L^2(0, T)$  und  $g \in L^\infty(0, T; X)$  betrachte man das parabolische Problem

$$\begin{aligned} \frac{\partial y}{\partial t}(t) + Ay(t) &= f(t) \cdot g(t) \quad f.\ddot{u}. \text{ in } [0, T], \\ y(0) &= 0. \end{aligned}$$

a) Dieses Problem besitzt eine eindeutige Lösung

$$y \in AC^{1,2}([0, T]; X) \cap L^2(0, T; D(A)).$$

b) Gilt  $f \in BV(0, T)$  und  $g \in AC^{1,2}([0, T]; X) \cap L^2(0, T; D(A))$ , so besitzt die eindeutige Lösung des parabolischen Problems die erhöhte Regularität

$$y \in L^2(0, T; D(A)) \quad \text{mit} \quad y_t \in L^2(0, T; D(A)) \cap BV(0, T; X).$$

Außerdem gilt  $Ay \in AC^{1,2}([0, T]; X) \cap L^2(0, T; D(A))$  als Lösung des Problems

$$\begin{aligned} \frac{\partial(Ay)}{\partial t}(t) + A(Ay)(t) &= f(t) \cdot Ag(t) \quad f.\ddot{u}. \text{ in } [0, T], \\ Ay(0) &= 0. \end{aligned}$$

*Beweis.* Teilaussage a) folgt unmittelbar mit der maximalen parabolischen  $L^2(0, T; X)$  Regularität des Operators  $A$ , da  $f \cdot g \in L^2(0, T; X)$  für die gegebenen Regularitäten von  $f$  und  $g$  gilt. Der Beweis der Aussage b) verläuft unter Anpassung der verschiedenen Räume und unter Beachtung von  $y_0 = 0$  analog zum Beweis in Satz A.0.1, weshalb die Ausführungen an dieser Stelle nicht mehr dargestellt werden. An einigen Stellen muss man statt mit der Stetigkeit des Operators  $A$  mit der Abgeschlossenheit des Operators  $A$  argumentieren.  $\square$

**Lemma A.0.1.** Sei  $V \subseteq H \subseteq V^*$  ein Gelfand-Dreier. Die Räume  $W_{V,H}(0, T)$  und  $W_{H,V^*}(0, T)$  seien definiert durch

$$\begin{aligned} W_{V,H}(0, T) &:= \{f \in L^2(0, T; V) \mid f' \in L^2(0, T; H)\}, \\ W_{H,V^*}(0, T) &:= \{g \in L^2(0, T; H) \mid g' \in L^2(0, T; V^*)\} \end{aligned}$$

und versehen mit den Normen

$$\begin{aligned} \|f\|_{W_{V,H}(0,T)} &:= \|f\|_{L^2(0,T;V)} + \|f'\|_{L^2(0,T;H)}, \\ \|g\|_{W_{H,V^*}(0,T)} &:= \|g\|_{L^2(0,T;H)} + \|g'\|_{L^2(0,T;V^*)}. \end{aligned}$$

Sei  $f \in W_{V,H}(0, T)$  und  $g \in W_{H,V^*}(0, T)$ . Für die Funktion  $[0, T] \rightarrow \mathbb{R}, t \mapsto (f(t), g(t))_H$  lassen sich dann die folgenden Aussagen herleiten:

---

a)  $(f, g)_H \in C([0, T])$ ,

b)  $\|(f, g)_H\|_{C([0, T])} \leq c \cdot \|f\|_{W_{V, H}(0, T)} \cdot \|g\|_{W_{H, V^*}(0, T)}$  mit  $c > 0$ ,

c) Formel der partiellen Integration für  $s, t \in [0, T]$ :

$$(f(t), g(t))_H = (f(s), g(s))_H + \int_s^t (f'(\tau), g(\tau))_H + \langle g'(\tau), f(\tau) \rangle_{V^*, V} d\tau.$$

*Beweis.* Die Räume  $W_{V, H}(0, T)$  und  $W_{H, V^*}(0, T)$  sind nach Roubicek [51, Ch. 7] Banachräume. Des Weiteren liegt  $C^1([0, T]; H)$  dicht in  $W_{H, V^*}(0, T)$  [51, Ch. 7]. In einem ersten Schritt soll gezeigt werden, dass die Teilaussage b) für beliebige Funktionen  $f \in W_{V, H}(0, T)$  und  $\tilde{g} \in C^1([0, T]; H)$  gilt, also:

$$\|(f, \tilde{g})_H\|_{C([0, T])} \leq c \cdot \|f\|_{W_{V, H}(0, T)} \cdot \|\tilde{g}\|_{W_{H, V^*}(0, T)} \quad \text{mit } c > 0.$$

Hierfür zerlege man mit einer Funktion  $\Phi \in C^1([0, T])$  mit  $\Phi(0) = 0, \Phi(T) = 1$  die Funktion  $\tilde{g}$  additiv

$$\tilde{g} = \underbrace{\Phi \cdot \tilde{g}}_{=: g_1} + \underbrace{(1 - \Phi) \cdot \tilde{g}}_{=: g_2}.$$

Jede Funktion  $f \in W_{V, H}(0, T)$  kann insbesondere als Funktion in  $AC^{1,2}([0, T]; H)$  aufgefasst werden. Mit der Formel der partiellen Integration und den Eigenschaften der Funktion  $\Phi$  folgt deshalb für die beiden Summanden für  $t \in [0, T]$ :

$$\begin{aligned} (f(t), g_1(t))_H &= (f(0), g_1(0))_H + \int_0^t (f'(\tau), g_1(\tau))_H + (g_1'(\tau), f(\tau))_H d\tau \\ &= \int_0^t \Phi'(\tau) (f(\tau), \tilde{g}(\tau))_H + \Phi(\tau) [(f'(\tau), \tilde{g}(\tau))_H + (g_1'(\tau), f(\tau))_H] d\tau \\ &= \int_0^t \Phi'(\tau) (f(\tau), \tilde{g}(\tau))_H + \Phi(\tau) [(f'(\tau), \tilde{g}(\tau))_H + \langle g_1'(\tau), f(\tau) \rangle_{V^*, V}] d\tau \end{aligned}$$

und

$$\begin{aligned} (f(t), g_2(t))_H &= \\ &= (f(T), g_2(T))_H - \int_t^T (f'(\tau), g_2(\tau))_H + (g_2'(\tau), f(\tau))_H d\tau \\ &= - \int_t^T -\Phi'(\tau) (f(\tau), \tilde{g}(\tau))_H + (1 - \Phi(\tau)) [(f'(\tau), \tilde{g}(\tau))_H + (g_2'(\tau), f(\tau))_H] d\tau \\ &= - \int_t^T -\Phi'(\tau) (f(\tau), \tilde{g}(\tau))_H + (1 - \Phi(\tau)) [(f'(\tau), \tilde{g}(\tau))_H + \langle g_2'(\tau), f(\tau) \rangle_{V^*, V}] d\tau. \end{aligned}$$

Hieraus ergibt sich mit einer Konstanten  $c > 0$  für beliebiges  $t \in [0, T]$

$$\begin{aligned}
|(f(t), \tilde{g}(t))_H| &= \\
&= |(f(t), g_1(t))_H + (f(t), g_1(t))_H| \\
&= \left| \int_0^T \Phi'(\tau) (f(\tau), \tilde{g}(\tau))_H + \Phi(\tau) \left[ (f'(\tau), \tilde{g}(\tau))_H + \langle \tilde{g}'(\tau), f(\tau) \rangle_{V^*, V} \right] \right. \\
&\quad \left. - \int_t^T (f'(\tau), g(\tau))_H + \langle \tilde{g}'(\tau), f(\tau) \rangle_{V^*, V} \, d\tau \right| \\
&\leq \max_{\tau \in [0, T]} |\Phi'(\tau)| \int_0^T |(f(\tau), \tilde{g}(\tau))_H| \, d\tau + \\
&\quad \left( \max_{\tau \in [0, T]} |\Phi(\tau)| + 1 \right) \int_0^T \left| (f'(\tau), \tilde{g}(\tau))_H + \langle \tilde{g}'(\tau), f(\tau) \rangle_{V^*, V} \right| \, d\tau \\
&\leq \max_{\tau \in [0, T]} |\Phi'(\tau)| \|f\|_{L^2(0, T; H)} \|\tilde{g}\|_{L^2(0, T; H)} + \left( \max_{\tau \in [0, T]} |\Phi(\tau)| + 1 \right) \cdot \\
&\quad \left( \|\tilde{g}'\|_{L^2(0, T; H)} \|f\|_{L^2(0, T; H)} + \|f'\|_{L^2(0, T; V^*)} \|\tilde{g}\|_{L^2(0, T; V)} \right) \\
&\leq c \|f\|_{W_{V, H}(0, T)} \|\tilde{g}\|_{W_{H, V^*}(0, T)}.
\end{aligned}$$

Für beliebige  $f \in W_{V, H}(0, T)$  und  $\tilde{g} \in C^1([0, T]; H)$  hat man somit mit  $c > 0$

$$\|(f, \tilde{g})_H\|_{C([0, T])} \leq c \cdot \|f\|_{W_{V, H}(0, T)} \cdot \|\tilde{g}\|_{W_{H, V^*}(0, T)}. \quad (\text{A.8})$$

Mit diesem Resultat lässt sich die Teilaussage a) herleiten. Sei hierzu  $g \in W_{H, V^*}(0, T)$  beliebig. Da  $C^1([0, T]; H)$  dicht in  $W_{H, V^*}(0, T)$  liegt, existiert eine Folge

$$(g_n) \in C^1([0, T]; H) \quad \text{mit} \quad \lim_{n \rightarrow \infty} \|g_n - g\|_{W_{H, V^*}(0, T)} = 0.$$

Damit ist  $g_n$  insbesondere eine Cauchy-Folge in  $W_{H, V^*}(0, T)$  und mit (A.8) gilt für beliebiges  $f \in W_{V, H}(0, T)$  und  $k, n \in \mathbb{N}$ :

$$\|(f, g_k)_H - (f, g_n)_H\|_{C([0, T])} \leq c \cdot \|f\|_{W_{V, H}(0, T)} \cdot \|v_k - v_n\|_{W_{H, V^*}(0, T)}.$$

$((f, g_n)_H)$  ist somit eine Cauchyfolge in  $C([0, T])$ . Da  $C([0, T])$  vollständig ist, konvergiert  $((f, g_n)_H)$  in  $C([0, T])$  gegen ein  $w \in C([0, T])$ . Andererseits hat man:

$$\begin{aligned}
\int_0^T |(f(\tau), g(\tau))_H - (f(\tau), g_n(\tau))_H| \, d\tau &= \int_0^T |(f(\tau), g(\tau) - g_n(\tau))_H| \, d\tau \\
&\leq \int_0^T \|f(\tau)\|_H \|g(\tau) - g_n(\tau)\|_H \, d\tau \\
&\leq \|f\|_{L^2(0, T; H)} \cdot \|g - g_n\|_{L^2(0, T; H)} \\
&\leq \|f\|_{L^2(0, T; H)} \cdot \|g - g_n\|_{W_{V, H}(0, T)} \xrightarrow{n \rightarrow \infty} 0,
\end{aligned}$$

also die Konvergenz von  $(f, g_n)_H$  in  $L^1(0, T)$  gegen  $(f, g)_H$ . Wegen der Eindeutigkeit des Grenzwertes findet sich in der Äquivalenzklasse von  $(f, g)_H$  dann aber der stetige Repräsentant  $w$ .

Die Teilaussage b) ergibt sich mit Abschätzung (A.8) unmittelbar wegen

$$\begin{aligned}
\|(f, g)_H\|_{C([0, T])} &\leq \|(f, g)_H - (f, g_n)_H\|_{C([0, T])} + \|(f, g_n)_H\|_{C([0, T])} \\
&\leq \|(f, g)_H - (f, g_n)_H\|_{C([0, T])} + c \cdot \|f\|_{W_{V, H}(0, T)} \cdot \|g_n\|_{W_{H, V^*}(0, T)} \\
&\leq \underbrace{\|w - (f, g_n)_H\|_{C([0, T])}}_{\rightarrow 0, n \rightarrow \infty} + c \cdot \|f\|_{W_{V, H}(0, T)} \cdot \underbrace{\|g_n - g\|_{W_{H, V^*}(0, T)}}_{\rightarrow 0, n \rightarrow \infty} \\
&\quad + c \cdot \|f\|_{W_{V, H}(0, T)} \cdot \|g\|_{W_{H, V^*}(0, T)}
\end{aligned}$$

Die Formel der partiellen Integration in c) leitet man aus der Formel der partiellen Integration für Funktionen in  $AC^{1,2}([0, T]; H)$  durch Grenzübergang her. Für jedes Folgenglied  $g_n \in C^1([0, T]; H)$  und  $f \in W_{V, H}(0, T)$  hat man für  $s, t \in [0, T]$

$$\begin{aligned}
(f(t), g_n(t))_H &= (f(s), g_n(s))_H + \int_s^t (f'(\tau), g_n(\tau))_H + (g'_n(\tau), f(\tau))_H \, d\tau \\
&= (f(s), g_n(s))_H + \int_s^t (f'(\tau), g(\tau))_H + \langle g'_n(\tau), f(\tau) \rangle_{V^*, V} \, d\tau.
\end{aligned}$$

Da  $\lim_{n \rightarrow \infty} \|(f, g_n)_H - (f, g)_H\|_{C([0, T])} = 0$ , gilt insbesondere für alle  $\tau \in [0, T]$

$$\lim_{n \rightarrow \infty} |(f(\tau), g_n(\tau))_H - (f(\tau), g(\tau))_H| = 0.$$

Für die Konvergenz des Integralausdrucks betrachte man die beiden Summanden getrennt. Für den ersten Summanden gilt für  $s, t \in [0, T]$

$$\begin{aligned}
\left| \int_s^t (f'(\tau), g_n(\tau))_H \, d\tau - \int_s^t (f'(\tau), g(\tau))_H \, d\tau \right| &\leq \\
&\leq \int_s^t \|f'(\tau)\|_H \|g_n(\tau) - g(\tau)\|_H \, d\tau \\
&\leq \|f'\|_{L^2(0, T; H)} \cdot \|g_n - g\|_{L^2(0, T; H)} \\
&\leq \|f\|_{W_{V, H}(0, T)} \cdot \|g_n - g\|_{W_{H, V^*}(0, T)} \xrightarrow{n \rightarrow \infty} 0.
\end{aligned}$$

Analoges Vorgehen ergibt für den zweiten Summanden

$$\begin{aligned}
\left| \int_s^t \langle g'_n(\tau), f(\tau) \rangle_{V^*, V} \, d\tau - \int_s^t \langle g'(\tau), f(\tau) \rangle_{V^*, V} \, d\tau \right| &\leq \\
&\leq \int_s^t \|g'_n(\tau) - g'(\tau)\|_{V^*} \|f(\tau)\|_V \, d\tau \\
&\leq \|g'_n - g'\|_{L^2(0, T; V^*)} \cdot \|f\|_{L^2(0, T; V)} \\
&\leq \|g_n - g\|_{W_{H, V^*}(0, T)} \cdot \|f\|_{W_{V, H}(0, T)} \xrightarrow{n \rightarrow \infty} 0.
\end{aligned}$$

□





# Literatur

- [1] H.W. Alt. *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*. Springer Lehrbuch. Springer, 2002.
- [2] Walter Alt. *Nichtlineare Optimierung: Eine Einführung in Theorie, Verfahren und Anwendungen*. Vieweg+Teubner Verlag, 2011.
- [3] H. Amann. *Linear and Quasilinear Parabolic Problems: Abstract linear theory*. Agents and Actions Supplements. Springer, 1995.
- [4] Viorel Barbu und Theodor Precupanu. *Convexity and optimization in Banach spaces*. Mathematics and its applications (D. Reidel Publishing Company).: East European series. Editura Academiei, 1978.
- [5] Maïtine Bergounioux, Kazufumi Ito und Karl Kunisch. „Primal-Dual Strategy for Constrained Optimal Control Problems“. In: *SIAM Journal on Control and Optimization* 37.4 (1999), S. 1176–1194.
- [6] Maïtine Bergounioux und Karl Kunisch. „Primal-Dual Strategy for State-Constrained Optimal Control Problems“. In: *Computational Optimization and Applications* 22.2 (2002), S. 193–224.
- [7] M. Bergounioux u. a. „A Comparison of a Moreau–Yosida-Based Active Set Strategy and Interior Point Methods for Constrained Optimal Control Problems“. In: *SIAM Journal on Optimization* 11.2 (2000), S. 495–521.
- [8] P. Berkmann und H. J. Pesch. „Abort landing in windshear: Optimal control problem with third-order state constraint and varied switching structure“. In: *Journal of Optimization Theory and Applications* 85.1 (1995), S. 21–57.
- [9] J. Frédéric Bonnans und Audrey Hermant. „No-gap second-order optimality conditions for optimal control problems with a single state constraint and control“. In: *Mathematical Programming* 117.1 (2009), S. 21–50.
- [10] J. Frédéric Bonnans und Audrey Hermant. „Well-Posedness of the Shooting Algorithm for State Constrained Optimal Control Problems with a Single Constraint and Control“. In: *SIAM Journal on Control and Optimization* 46.4 (2007), S. 1398–1430.
- [11] J. Frédéric Bonnans und Pascal Jaisson. „Optimal Control of a Parabolic Equation with Time-Dependent State Constraints“. In: *SIAM Journal on Control and Optimization* 48.7 (2010), S. 4550–4571.
- [12] Joseph Frédéric Bonnans und Audrey Hermant. „Stability and sensitivity analysis for optimal control problems with a first-order state constraint and application to continuation methods“. In: *ESAIM: COCV* 14.4 (2008), S. 825–863.

- [13] Gregor Branner. „Modellierung transienter Effekte in der Struktursimulation von Schichtbauverfahren“. Dissertation. München: Technische Universität München, 2010.
- [14] A.E. Bryson, W.F. Denham und S.E. Dreyfus. „Optimal programming problems with inequality constraints I: necessary conditions for extremal solutions“. In: *AIAA Journal* 1.11 (1963).
- [15] Arthur E. Bryson und Yu-Chi Ho. *Applied Optimal Control: Optimization, Estimation and Control*. Halsted Press book'. Taylor & Francis, 1975.
- [16] R. Bulirsch, F. Montrone und H. J. Pesch. „Abort landing in the presence of windshear as a minimax optimal control problem, part 1: Necessary conditions“. In: *Journal of Optimization Theory and Applications* 70.1 (1991), S. 1–23.
- [17] R. Bulirsch, F. Montrone und H. J. Pesch. „Abort landing in the presence of windshear as a minimax optimal control problem, part 2: Multiple shooting and homotopy“. In: *Journal of Optimization Theory and Applications* 70.2 (1991), S. 223–254.
- [18] Eduardo Casas. „Pontryagin’s Principle for State-Constrained Boundary Control Problems of Semilinear Parabolic Equations“. In: *SIAM Journal on Control and Optimization* 35.4 (1997), S. 1297–1327.
- [19] *Common Layer Interface (CLI): Version 2.0*. [https://www.forwiss.uni-passau.de/~welisch/papers/cli\\_format.html](https://www.forwiss.uni-passau.de/~welisch/papers/cli_format.html). Zugriff am 17. April 2014.
- [20] J. de Los Reyes u. a. „Optimality conditions for state-constrained PDE control problems with time-dependent controls“. eng. In: *Control and Cybernetics* 37.1 (2008), S. 5–38.
- [21] Etienne Emmrich. *Gewöhnliche und Operator-Differentialgleichungen: Eine integrierte Einführung in Randwertprobleme und Evolutionsgleichungen für Studierende*. Vieweg+Teubner Verlag, 2013.
- [22] Michael Frey. „Shape Calculus Applied to State-Constrained Elliptic Optimal Control Problems“. Diss. Bayreuth: Universität Bayreuth, 2012.
- [23] H. Gajewski, K. Gröger und K. Zacharias. *Nichtlineare Operatorgleichungen und Operatordifferential-Gleichungen*. Mathematische Lehrbücher und Monographien: Mathematische Monographien. Akademie-Verlag, 1974.
- [24] Matthias Gerds. *Optimal Control of ODEs and DAEs*. Berlin, Boston: de Gruyter, 2011.
- [25] Philip E. Gill, Walter Murray und Margaret H. Wright. *Practical optimization*. London, UK: Academic Press, 1981.
- [26] John Goldak, Aditya Chakravarti und Malcolm Bibby. „A new finite element model for welding heat sources“. In: *Metallurgical Transactions B* 15.2 (Juni 1984), S. 299–305.
- [27] H. Goldberg und F. Tröltzsch. „Second-Order Sufficient Optimality Conditions for a Class of Nonlinear Parabolic Boundary Control Problems“. In: *SIAM Journal on Control and Optimization* 31.4 (1993), S. 1007–1025.

- 
- [28] J.A. Griepentrog, H.C. Kaiser und J. Rehberg. „Heat kernel and resolvent properties for second order elliptic differential operators with general boundary conditions on  $L_p$ “. In: *Advances in Mathematical Sciences and Applications* 11 (2001), S. 87–112.
- [29] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Society for Industrial and Applied Mathematics, 2011.
- [30] William W. Hager und George D. Ianculescu. „Dual Approximations in Optimal Control“. In: *SIAM Journal on Control and Optimization* 22.3 (1984), S. 423–465.
- [31] Richard F. Hartl, Suresh P. Sethi und Raymond G. Vickson. „A Survey of the Maximum Principles for Optimal Control Problems with State Constraints“. In: *SIAM Rev.* 37.2 (Juni 1995), S. 181–218.
- [32] M. Hintermüller und K. Kunisch. „Feasible and Noninterior Path-Following in Constrained Minimization with Low Multiplier Regularity“. In: *SIAM Journal on Control and Optimization* 45.4 (2006), S. 1198–1221.
- [33] Michael Hintermüller. „A Primal-dual Active Set Algorithm for Bilaterally Control Constrained Optimal Control Problems“. In: *Quarterly of Applied Mathematics* 61.1 (2003), S. 131–160.
- [34] Michael Hintermüller und Karl Kunisch. „Path-following Methods for a Class of Constrained Minimization Problems in Function Space“. In: *SIAM Journal on Optimization* 17.1 (2006), S. 159–187.
- [35] Michael Hinze u. a. *Optimization with PDE Constraints*. Mathematical Modelling: Theory and Applications. Springer Netherlands, 2009.
- [36] Isabella Hoffmann, Sascha Kurz und Jörg Rambau. *The Maximum Scatter TSP on a Regular Grid*. Preprint. to appear in Proceedings of OR 2015. University of Bayreuth, 2015.
- [37] Alexander Ilin u. a. „Computer Aided Optimisation of the Thermal Management During Laser Beam Melting Process“. In: *Physics Procedia* 56 (2014), S. 390–399.
- [38] Kazufumi Ito und Karl Kunisch. „The Primal-Dual Active Set Method for Nonlinear Optimal Control Problems with Bilateral Constraints“. In: *SIAM Journal on Control and Optimization* 43.1 (2004), S. 357–376.
- [39] D.H. Jacobson, M.M. Lele und J.L. Speyer. „New necessary conditions of optimality for control problems with state-variable inequality constraints“. In: *Journal of Mathematical Analysis and Applications* 35.2 (1971), S. 255–284.
- [40] Karl Kunisch und Arnd Rösch. „Primal-Dual Active Set Strategy for a General Class of Constrained Optimal Control Problems“. In: *SIAM Journal on Optimization* 13.2 (2002), S. 321–334.
- [41] Irena Lasiecka. „State constrained control problems for parabolic systems: Regularity of optimal solutions“. In: *Applied Mathematics and Optimization* 6.1 (1980), S. 1–29.
- [42] J.L. Lions. *Control of Distributed Singular Systems*. Bordas, 1985.

- [43] David G. Luenberger. *Optimization by Vector Space Methods*. Professional Series. Wiley, 1969.
- [44] Helmut Maurer. „On the minimum principle for optimal control problems with state constraints“. In: *Schriftenreihe des Rechenzentrums* 41 (Universität Münster 1979).
- [45] Dominik Meidner, Rolf Rannacher und Boris Vexler. „A Priori Error Estimates for Finite Element Discretizations of Parabolic Optimization Problems with Pointwise State Constraints in Time“. In: *SIAM Journal on Control and Optimization* 49.5 (2011), S. 1961–1997.
- [46] Chr. Meyer, U. Prüfert und F. Tröltzsch. „On two numerical methods for state-constrained elliptic control problems“. In: *Optimization Methods and Software* 22.6 (2007), S. 871–899.
- [47] Christian Meyer, Arnd Rösch und Fredi Tröltzsch. „Optimal Control of PDEs with Regularized Pointwise State Constraints“. In: *Computational Optimization and Applications* 33.2 (2006), S. 209–228.
- [48] I.P. Natanson. *Theorie der Funktionen einer reellen Veränderlichen*. Mathematische Lehrbücher und Monographien. Akademie-Verlag, 1969.
- [49] Hans Josef Pesch. „A Practical Guide to the Solution of Real-Life Optimal Control Problems“. In: *Control and Cybernetics* 23.1 (1994), S. 7–60.
- [50] M. Renardy und R.C. Rogers. *An Introduction to Partial Differential Equations*. Texts in Applied Mathematics. Springer New York, 2004.
- [51] Tomas Roubicek. *Nonlinear Partial Differential Equations with Applications*. International Series of Numerical Mathematics. Birkhäuser Basel, 2013.
- [52] Anton Schiela. „An interior point method in function space for the efficient solution of state constrained optimal control problems“. In: *Mathematical Programming Series A* 138.1-2 (2013), S. 83–114.
- [53] Anton Schiela. „Barrier Methods for Optimal Control Problems with State Constraints“. In: *SIAM J. on Optimization* 20.2 (Juli 2009), S. 1002–1031.
- [54] Fredi Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen: Theorie, Verfahren und Anwendungen*. Vieweg, 2005.
- [55] Joseph Wloka. *Partial Differential Equations*. Cambridge University Press, 1987.

## Publikationen

- [56] Simon Bechmann und Michael Frey. *Regularisierungsmethoden für Optimalsteuerungsprobleme*. Bd. 80. Bayreuther mathematische Schriften. Bayreuth: Math. Inst. der Univ. Bayreuth, 2008.
- [57] Simon Bechmann u. a. „On Some New Adjoining Approaches for State-Constrained Elliptic Optimal Control Problems“. In: *AIP Conference Proceedings* 1168.1 (2009), S. 1322–1323.
- [58] Hans Josef Pesch, Simon Bechmann und Jan Eric Wurst. „Bang-bang and singular controls in optimal control problems with partial differential equations“. In: *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. Dez. 2012, S. 7671–7678.



# Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet habe.

Weiterhin erkläre ich, dass ich die Hilfe von gewerblichen Promotionsberatern bzw. -vermittlern oder ähnlichen Dienstleistern weder bisher in Anspruch genommen habe, noch künftig in Anspruch nehmen werde.

Zusätzlich erkläre ich hiermit, dass ich keinerlei frühere Promotionsversuche unternommen habe.

Bayreuth, 19. Dezember 2016

Simon Bechmann