

Brandenburgische Technische Universität Cottbus-Senftenberg

Optimalsteuerungsprobleme mit Totalvariation



Masterarbeit

zur Erlangung des akademischen Grades

Master of Science (M.Sc.)

vorgelegt von

JONAS MARKO

Tag der Einreichung: 04.10.2021

unter der Aufsicht und Beratung von

PROF. GERD WACHSMUTH (Fachgebiet Optimale Steuerung)

DOI: 10.26127/BTUOpen-6071

Inhaltsverzeichnis

1	Motivation	2
2	Grundlagen zu Funktionen beschränkter Variation	6
2.1	Die klassische Totalvariation einer Funktion	6
2.2	Variation von Funktionen in Lebesgueräumen	13
3	Problemstellung	19
4	Existenzresultate	27
4.1	Stetigkeitseigenschaften des Lösungsoperators	28
4.2	Stetigkeitseigenschaften des Zielfunktional	34
5	Differenzierbarkeitseigenschaften des Zielfunktional	41
5.1	Differenzierbarkeit des Lösungsoperators	41
5.2	Differentiale der Terme im Zielfunktional	49
6	Lösungsalgorithmus	54
6.1	Ein Proximalgradientenverfahren ohne Konvexität	55
6.2	Diskretisierung der Problemstellung	66
6.3	Anwendung des Bellmannschen Optimalitätsprinzips	70
6.4	Algorithmen zur Lösungsapproximation	74
7	Numerische Resultate	77
8	Zusammenfassung und Fazit	83

Notation

X^N	N -faches kartesisches Produkt des Raums X , es gilt $x = (x_1, \dots, x_N) \in X^N$, wenn $x_i \in X \forall i \in \{1, \dots, N\}$. Ausgestattet mit der Norm $\ x\ _{X^N} := \sqrt{\sum_{i=1}^N \ x_i\ _X^2}$.
$X^{N \times N}$	Raum aller Matrizen $A = (a_{ij})_{i,j=1}^N$, sodass $a_{ij} \in X \forall (i, j) \in \{1, \dots, N\}^2$ gilt.
X^*	Raum aller stetigen, linearen Funktionale von X nach \mathbb{R} .
$\ x\ _2$	Euklidische Vektornorm, falls $x \in \mathbb{R}^N$ bzw. Spektralnorm, falls $x \in \mathbb{R}^{N \times N}$.
$\ A\ _F$	Frobeniusnorm für Matrix $A \in \mathbb{R}^{N \times N}$.
$\langle x, y \rangle_2$	Standardskalarprodukt zwischen Vektoren $x, y \in \mathbb{R}^N$.
$\langle x, y \rangle_X$	Dualitätsprodukt mit $x \in X^*$, $y \in X$.
$x_k \xrightarrow{*} x$	Schwach-* Konvergenz, es gilt $x_k \xrightarrow{*} x \Leftrightarrow \langle x_k, y \rangle_X \rightarrow \langle x, y \rangle_X \forall y \in X$.
$(\nabla_x f)h$	Gateaux-Differential von f bezüglich x in Richtung h .
$\pm x$	Addieren und anschließendes Subtrahieren des Terms x , also $\pm x \hat{=} +x - x$.
$C_C^k((a, b))$	Raum der k -mal differenzierbaren, stetigen Funktionen mit kompaktem Träger in (a, b) .
$C_0((a, b))$	Abschluss von $C_C^0((a, b))$ bezüglich der Supremumsnorm.
$L^p((a, b))$	Lebesgueraum der p -fach integrierbaren Äquivalenzklassen von Funktionen auf (a, b) .
$L^\infty((a, b))$	Lebesgueraum aller Äquivalenzklassen wesentlich beschränkter Funktionen auf (a, b) .
$H^1((a, b))$	Hilbertraum aller Äquivalenzklassen schwach differenzierbarer Funktionen aus $L^2((a, b))$ mit schwacher Ableitung in $L^2((a, b))$.

Kapitel 1

Motivation

In der optimalen Steuerung wird nicht nur die Zielstellung verfolgt, eine Steuerungsfunktion so zu wählen, dass eine davon abhängige Zustandsfunktion gewisse günstige Eigenschaften besitzt, auch ist es oft von Interesse, bestimmte qualitative Eigenschaften von der Steuerungsfunktion selbst zu erhalten. Als eine solche Eigenschaft wird in dieser Arbeit die Totalvariation der Steuerung untersucht.

Die Totalvariation einer Funktion ist ein Maß dafür, wie stark diese variiert, sie sagt also grob gesprochen etwas über das Änderungsverhalten der Funktion über einem bestimmten Intervall aus, ähnlich zur Ableitung. Der fundamentale Unterschied ist hierbei allerdings, dass die Totalvariation auch für nicht-differenzierbare Funktionen gebildet werden kann. In vielen praktischen Anwendungen sind die Möglichkeiten, mit einer Steuerung auf ein System einzuwirken nur sehr begrenzt, man stelle sich hier als Beispiel ein Gerät vor, welches entweder ein- oder ausgeschaltet werden kann. Die Beschränkung einer mathematischen Modellierung auf glatte Steuerungsfunktionen scheint im Hinblick auf solche Probleme also in vielen Fällen nicht sinnvoll zu sein.

Gerade in der Signal- und Bildverarbeitung wird die Totalvariation oft als eine Art Regularisierungsterm in der Zielfunktion verwendet, um einerseits die Existenz einer Lösung garantieren zu können und andererseits Lösungen mit scharfen Kanten zu erhalten, siehe auch [1, 2, 3, 4]. Oft kommt es auch vor, dass die Lösung eines Optimalsteuerungsproblems mehrmals in kurzer Zeit zwischen verschiedenen Werten schaltet, was die praktische Applikation des theoretisch gewonnenen Steuerungsverhaltens erschwert und das Bestrafen eines solchen Verhaltens in der Zielfunktion motiviert. Auch wenn in der Praxis die Umsetzung des Steuerungsverhaltens nicht zum Problem wird, ist es trotzdem des Öfteren nicht erwünscht, häufig zwischen verschiedenen Steuerwerten zu springen, da abhängig von dem betrachteten Anwendungsfall eine Änderung in der Steuerung mit einem Energieaufwand oder dem Verschleiß ei-

ner Komponente einhergehen kann.

Um die Nützlichkeit der Totalvariation als Term in der Zielfunktion eines Optimalsteuerungsproblems zu verdeutlichen, betrachten wir zwei Beispiele.

Beispiel 1.1. (Temperatursteuerung einer Herdplatte)

Ein Koch bzw. eine Köchin bereitet in einem Topf oder einer Pfanne ein Gericht zu. Die Temperatur der Pfanne bzw. des Topfs (von nun an allgemein als Kochgeschirr bezeichnet) zum Zeitpunkt $t \in (0, \mathcal{T})$ sei durch $T_1(t)$ modelliert, die Temperatur der Herdplatte durch $T_2(t)$, jeweils in Grad Celsius.

Die Kochgeschirrtemperatur ändert sich proportional zum Temperaturunterschied zwischen dem Kochgeschirr und der Herdplatte bzw. der Umgebung. Diese Änderung kann also mit zwei Konstanten $\alpha_1 > 0$ und $\alpha_2 > 0$ durch die Gleichung

$$T_1'(t) = \alpha_1(T_2(t) - T_1(t)) - \alpha_2(T_1(t) - T_Z)$$

modelliert werden, T_Z steht hier für die Zimmer- bzw. Umgebungstemperatur. Wir nehmen an, dass zum Zeitpunkt $t = 0$ die Herdplatte und das Kochgeschirr noch nicht erhitzt wurden, es gilt also $T_1(0) = T_2(0) = T_Z$.

Die Temperatur der Herdplatte kann durch den Benutzer mit einem Regler, Knöpfen oder über ein Touchpad durch die Einstellung einer Stufe $u(t) \in \{0, 0.5, 1, \dots, u_{max}\}$ im Zeitpunkt t gesteuert werden. Wir nehmen an, dass die leere Herdplatte sich bei der Einstellung $u(t)$ bis zu einer Maximaltemperatur von $100 \cdot u(t)^\circ\text{C}$ erwärmt, andere aus den Stufen resultierende Maximaltemperaturen können als $(c_1 u(t) + c_2)^\circ\text{C}$ mit $c_1 > 0$, $c_2 \geq 0$ ebenso betrachtet und im folgenden Modell verwendet werden.

Die Stufe bestimmt also die Heizleistung, sodass die Herdplatte sich über einen gewissen Zeitraum erhitzt und $T_2(t)$ an $100 \cdot u(t)^\circ\text{C}$ angenähert wird. Die Temperaturänderung der Herdplatte ist außerdem proportional zum Temperaturunterschied zwischen dieser und dem Kochgeschirr bzw. der Umgebung, was die Modellierung

$$T_2'(t) = \gamma_1(100u(t) - T_2(t)) + \gamma_2(T_1(t) - T_2(t)) - \gamma_3(T_2(t) - T_Z)$$

mit drei Konstanten $\gamma_1, \gamma_2, \gamma_3 > 0$ motiviert.

Der Koch bzw. die Köchin verfolgt nun die Zielstellung, die Temperatur des Kochgeschirrs so zu steuern, dass sie über dem Intervall $(0, \mathcal{T})$ so nahe wie

möglich an einer Zielkurve $T_d(t) \in L^2((a, b))$ liegt. Zudem ist es auch wünschenswert, die Steuerung u über $(0, \mathcal{T})$ klein zu halten, um Energie zu sparen. Dies führt auf die Problemstellung

$$\text{Minimiere } \frac{1}{2} \int_0^{\mathcal{T}} (T_1(t) - T_d(t))^2 dt + \beta \int_0^{\mathcal{T}} |u(t)| dt$$

mit einem Parameter $\beta > 0$, der die Gewichtung der beiden Zielstellungen modelliert. Nun kann es vorkommen, dass die resultierende optimale Steuerung oft zwischen den Werten $\{0, \dots, u_{max}\}$ springt, sodass der Koch bzw. die Köchin mithilfe einer Stoppuhr akribisch den Herd steuern muss. Eine einfachere Steuerung, die es erlaubt, weitere Aufgaben nebenbei zu erledigen, ist hier also sicherlich von Vorteil, sodass Sprünge bzw. Variationen in u bestraft werden sollten.

Beispiel 1.2. (Lotka-Volterra Fishing Problem)

Wir beobachten die Beeinflussung eines Systems aus Raub- und Beutetieren mit einer Fischereiflotte, die mithilfe von Schleppnetzen in einem abgeschlossenen System, beispielsweise einem See oder Binnenmeer, mit jeder Ausfahrt einen Anteil $\theta_1 > 0$ an Raub- und einen Anteil $\theta_2 > 0$ an Beutetieren aus dem System entnimmt. Die Dynamik eines solchen Systems kann mithilfe der Lotka-Volterra Gleichungen durch

$$\begin{aligned} x'(t) &= \alpha x(t) - \beta x(t)z(t) - \theta_1 x(t)u(t) \\ z'(t) &= \gamma x(t)z(t) - \delta z(t) - \theta_2 z(t)u(t) \\ x(0) &= x_0, \quad z(0) = z_0 \end{aligned}$$

modelliert werden, wobei $x(t)$ die Beutetierpopulation und $z(t)$ die Raubtierpopulation zum Zeitpunkt $t \in (0, T)$ darstellt. Die Steuerung $u(t) \in \{0, 1\}$ nimmt den Wert 1 an, wenn die Fischereiflotte zum Zeitpunkt t fischt und 0, wenn dies nicht der Fall ist.

Hierbei liegen die Grundannahmen vor, dass sich die Beutepopulation exponentiell mit dem Anteil $\alpha > 0$ vermehrt, wenn diese nicht der Verminderung durch die Raubtiere unterliegt. Der Anteil dieser Verminderung wurde als proportional zur Häufigkeit des Aufeinandertreffens der Beute- und Raubtiere mit dem Parameter $\beta > 0$ angenommen. Jene Häufigkeit ist auch als proportional zum Wachstum der Raubtierpopulation mit dem Parameter $\gamma > 0$ modelliert. Schließlich wird dieses Wachstum durch die Populationsgröße der Raubtiere mit

der Rate $\delta > 0$ verringert.

Ziel des Unterfanges der Fischereiflotte ist es, ihre Ausfahrten so zu planen, dass die Populationen der Raub- und Beutetiere so nah wie möglich an zwei vorgegebenen Zielpopulationen $x_d(t) \in L^2((a, b))$ und $z_d(t) \in L^2((a, b))$ liegen. Dies kann mithilfe der Zielstellung

$$\text{Minimiere } \int_0^T (x(t) - x_d(t))^2 + (z(t) - z_d(t))^2 dt$$

erreicht werden. Bei dieser Problemstellung ist es nicht nur aufgrund des logistischen Aufwands einer Ausfahrt der Fischereiflotte wünschenswert, stärkere Variationen der Steuerung zu verhindern, man betrachte auch das folgende Szenario:

Wie in [5, Abschnitt 4.1] setzen wir $x_d(t) = z_d(t) := 1$ und die Parameter seien gemäß

$$(\alpha, \beta, \gamma, \delta) = (1, 1, 1, 1), (\theta_1, \theta_2) = (0.4, 0.2), T = 12, x_0 = 0.5, z_0 = 0.7,$$

gewählt. In [5, Abschnitt 4.2] ist erläutert, dass dieses Problem keine optimale Lösung besitzt, die optimale Lösung kann allerdings beliebig genau approximiert werden, wenn u oft genug zwischen 0 und 1 hin- und herschaltet. Zwei solcher Approximationen sind in Abbildung 1.1 dargestellt. Wünschenswert wäre hier, eine Strategie für die Fischereiflotte zu finden, bei der die Anzahl an Ausfahrten reduziert wird, unter geringer Steigerung des Zielfunktionswerts.

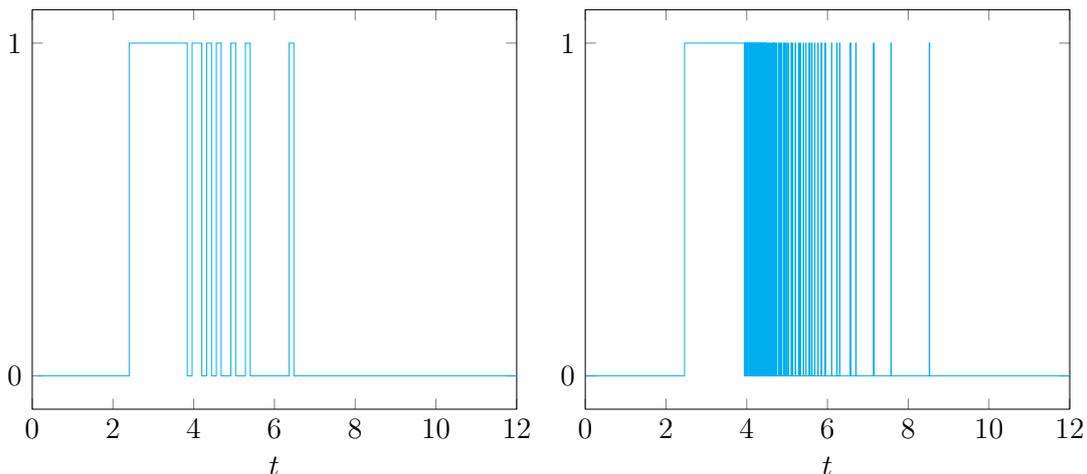


Abbildung 1.1: Lösungsapproximationen zu Beispiel 1.2 bei Diskretisierung der Zeitachse durch 100 bzw. 1000 Gitterpunkte

Kapitel 2

Grundlagen zu Funktionen beschränkter Variation

Um den Überlegungen dieser Arbeit folgen zu können, sind grundlegende Kenntnisse der Funktionalanalysis, Maßtheorie sowie Distributionstheorie vorausgesetzt. Zudem werden einige Erkenntnisse zu Funktionen beschränkter Variation benötigt. Dieses Kapitel soll einen Einstieg in die Thematik bieten.

2.1 Die klassische Totalvariation einer Funktion

Wir starten mit der klassischen Definition der Totalvariation einer Funktion auf einem Intervall.

Definition 2.1. Seien $a, b \in \mathbb{R}$ mit $a < b$ und $\mathcal{P}([a, b])$ die Menge aller Partitionen $\{t_i\}_{i=0}^n$ von $[a, b]$ mit

$$a = t_0 < t_1 < \cdots < t_n = b.$$

Für eine Funktion $u: [a, b] \rightarrow \mathbb{R}$ bezeichnen wir mit

$$\text{Var}(u, [a, b]) := \sup \left\{ \sum_{i=1}^n |u(t_i) - u(t_{i-1})|, \{t_i\}_{i=0}^n \in \mathcal{P}([a, b]) \right\}$$

die Totalvariation von u über $[a, b]$. Außerdem sei

$$\text{Var}(u, (a, b)) := \sup \left\{ \sum_{i=1}^n |u(t_i) - u(t_{i-1})| : n \geq 2, a < t_0 < \cdots < t_n < b \right\}$$

die Totalvariation von u über (a, b) . Gilt für $\Omega = (a, b)$ oder $\Omega = [a, b]$ $\text{Var}(u, \Omega) < \infty$, so sagen wir, u ist von beschränkter Variation über Ω . Die Menge aller Funktionen beschränkter Variation über Ω sei durch $\text{BV}_p(\Omega)$ notiert.

Die Bezeichnung BV_p soll hierbei verdeutlichen, dass die Zugehörigkeit einer Funktion zu $\text{BV}_p(\Omega)$ von den punktwisen Funktionswerten von u abhängt. Wir legen zunächst den Fokus auf $\text{BV}_p([a, b])$. Das folgende Beispiel zeigt, dass viele Funktionen in $\text{BV}_p([a, b])$ liegen.

Beispiel 2.2. Jede monotone und jede Lipschitzstetige Funktion auf $[a, b]$ liegt in $\text{BV}_p([a, b])$. Um Ersteres einzusehen, betrachte man eine monoton steigende Funktion u , die Begründung für eine monoton fallende Funktion \tilde{u} funktioniert analog bei Betrachtung von $u = -\tilde{u}$. Bei einer monoton steigenden Funktion können die Beträge in der Definition der Totalvariation weggelassen werden. Übrig bleibt eine Teleskopsumme, sodass schließlich

$$\text{Var}(u, [a, b]) = u(b) - u(a) < \infty$$

folgt.

Im Falle der Lipschitzstetigkeit einer Funktion u mit Lipschitzkonstante $L > 0$ gilt die Abschätzung

$$\text{Var}(u, [a, b]) \leq \sup \left\{ \sum_{i=1}^n L|t_i - t_{i-1}|, \{t_i\}_{i=0}^n \in \mathcal{P}([a, b]) \right\} = L(b - a) < \infty,$$

da erneut eine Teleskopsumme auftritt.

Außerdem ist $u(t) := \int_a^t \varphi(s) \, ds$ mit $\varphi: [a, b] \rightarrow \mathbb{R}$, $\varphi \in L^1([a, b])$ wegen

$$\begin{aligned} \sum_{i=1}^n |u(t_i) - u(t_{i-1})| &= \sum_{i=1}^n \left| \int_a^{t_i} \varphi(s) \, ds - \int_a^{t_{i-1}} \varphi(s) \, ds \right| = \sum_{i=1}^n \left| \int_{t_{i-1}}^{t_i} \varphi(s) \, ds \right| \\ &\leq \sum_{i=1}^n \int_{t_{i-1}}^{t_i} |\varphi(s)| \, ds = \int_a^b |\varphi(s)| \, ds < \infty, \quad \text{da } \varphi \in L^1([a, b]), \end{aligned}$$

in $\text{BV}_p([a, b])$.

Neben der gesamten Variation kann man auch Größen für die Variation nach oben bzw. nach unten definieren, mit diesen lassen sich einige nützliche Eigenschaften formulieren.

Definition 2.3. Wir nennen

$$\text{Var}^+(u, [a, b]) := \sup \left\{ \sum_{i=1}^n \max(u(t_i) - u(t_{i-1}), 0), \{t_i\}_{i=0}^n \in \mathcal{P}([a, b]) \right\}$$

die positive und

$$\text{Var}^-(u, [a, b]) := \sup \left\{ - \sum_{i=1}^n \min(u(t_i) - u(t_{i-1}), 0), \{t_i\}_{i=0}^n \in \mathcal{P}([a, b]) \right\}$$

die negative Variation von u über $[a, b]$.

Nun wollen wir ein paar Eigenschaften herleiten, die das Rechnen mit der Variation einer Funktion vereinfachen.

Satz 2.4. Für $a, b, c \in \mathbb{R}$, $a < c < b$, gilt

$$\text{Var}(u, [a, b]) = \text{Var}(u, [a, c]) + \text{Var}(u, [c, b]).$$

Beweis. Wir zeigen zunächst $\text{Var}(u, [a, b]) \leq \text{Var}(u, [a, c]) + \text{Var}(u, [c, b])$. Für $\{t_i\}_{i=0}^n \in \mathcal{P}([a, b])$ bezeichnen wir mit c_- den größten Index aus $\{1, \dots, n\}$, sodass $t_{c_-} \leq c$ und mit c_+ den kleinsten Index aus $\{1, \dots, n\}$, sodass $t_{c_+} \geq c$. Es gilt entweder $c_- + 1 = c_+$ oder $c_- = c_+$. Mit diesen Setzungen kann die Summe in der Definition der Totalvariation für eine beliebige gegebene Partition $\{t_i\}_{i=0}^n \in \mathcal{P}([a, b])$ auch folgendermaßen geschrieben werden:

$$\begin{aligned} \sum_{i=1}^n |u(t_i) - u(t_{i-1})| &= \sum_{i=1}^{c_-} |u(t_i) - u(t_{i-1})| + |u(t_{c_+}) - u(t_{c_-})| + \sum_{i=c_++1}^n |u(t_i) - u(t_{i-1})| \\ &= \sum_{i=1}^{c_-} |u(t_i) - u(t_{i-1})| + |u(t_{c_+}) - u(c) + u(c) - u(t_{c_-})| + \sum_{i=c_++1}^n |u(t_i) - u(t_{i-1})| \\ &\leq \sum_{i=1}^{c_-} |u(t_i) - u(t_{i-1})| + |u(t_{c_+}) - u(c)| + |u(c) - u(t_{c_-})| + \sum_{i=c_++1}^n |u(t_i) - u(t_{i-1})| \end{aligned}$$

Wegen dieser Abschätzung gilt auch

$$\begin{aligned}
\text{Var}(u, [a, b]) &\leq \sup \left\{ \sum_{i=1}^{c_-} |u(t_i) - u(t_{i-1})| + |u(t_{c_+}) - u(c)| + |u(c) - u(t_{c_-})| \right. \\
&\quad \left. + \sum_{i=c_++1}^n |u(t_i) - u(t_{i-1})|, \{t_i\}_{i=0}^n \in \mathcal{P}([a, b]) \right\} \\
&\leq \sup \left\{ \sum_{i=1}^{c_-} |u(t_i) - u(t_{i-1})| + |u(c) - u(t_{c_-})|, \{t_i\}_{i=0}^n \in \mathcal{P}([a, b]) \right\} \\
&\quad + \sup \left\{ |u(t_{c_+}) - u(c)| + \sum_{i=c_++1}^n |u(t_i) - u(t_{i-1})|, \{t_i\}_{i=0}^n \in \mathcal{P}([a, b]) \right\} \\
&\leq \text{Var}(u, [a, c]) + \text{Var}(u, [c, b]),
\end{aligned}$$

da $\{a, t_1, \dots, t_{c_-}, c\}$ eine Partition von $[a, c]$ und $\{c, t_{c_+}, \dots, t_{n-1}, b\}$ eine Partition von $[c, b]$ ist.

Um die umgekehrte Relation zu zeigen, betrachten wir drei Folgen von Partitionen $M^k = \{M_i^k\}_{i=0}^{m_k}$, $N^k = \{N_i^k\}_{i=0}^{n_k}$ und $P^k = \{P_i^k\}_{i=0}^{p_k}$ von $[a, c]$, $[c, b]$ und $[a, b]$, sodass

$$\begin{aligned}
\lim_{k \rightarrow \infty} \sum_{i=1}^{m_k} |u(M_i^k) - u(M_{i-1}^k)| &= \text{Var}(u, [a, c]), \\
\lim_{k \rightarrow \infty} \sum_{i=1}^{n_k} |u(N_i^k) - u(N_{i-1}^k)| &= \text{Var}(u, [c, b]), \\
\lim_{k \rightarrow \infty} \sum_{i=1}^{p_k} |u(P_i^k) - u(P_{i-1}^k)| &= \text{Var}(u, [a, b]).
\end{aligned}$$

Wir definieren für jedes $k \in \mathbb{N}_0$ die Partition $Q^k = M^k \cup N^k$ von $[a, b]$, es gilt also

$$Q^k = \{a = M_0^k, M_1^k, \dots, M_{m_k}^k = c, N_1^k, \dots, N_{n_k}^k = b\} =: \{Q_0^k, \dots, Q_{q_k}^k\}.$$

Es folgt wegen $N_0^k = M_{m_k}^k$

$$\begin{aligned}
\text{Var}(u, [a, c]) + \text{Var}(u, [c, b]) &= \lim_{k \rightarrow \infty} \sum_{i=1}^{m_k} |u(M_i^k) - u(M_{i-1}^k)| + \sum_{i=1}^{n_k} |u(N_i^k) - u(N_{i-1}^k)| \\
&= \lim_{k \rightarrow \infty} \sum_{i=1}^{q_k} |u(Q_i^k) - u(Q_{i-1}^k)| \\
&\leq \lim_{k \rightarrow \infty} \sum_{i=1}^{p_k} |u(P_i^k) - u(P_{i-1}^k)| = \text{Var}(u, [a, b]),
\end{aligned}$$

da P^k die Folge von Partitionen ist, für die das Supremum der Summen angenommen wird. \square

Mit analogem Vorgehen im Beweis lässt sich dieses Resultat auch für die positive und negative Variation zeigen.

Korollar 2.5. Für $a, b, c \in \mathbb{R}$, $a < c < b$, gilt

$$\begin{aligned}\text{Var}^+(u, [a, b]) &= \text{Var}^+(u, [a, c]) + \text{Var}^+(u, [c, b]) \\ \text{Var}^-(u, [a, b]) &= \text{Var}^-(u, [a, c]) + \text{Var}^-(u, [c, b]).\end{aligned}$$

Die positive und negative Variation einer Funktion können genutzt werden, um weitere nützliche Resultate zu zeigen. Die folgenden Sätze zeigen die Zusammenhänge der positiven und negativen Variation mit u und $\text{Var}(u, [a, b])$ auf.

Satz 2.6. Für $u \in \text{BV}_p([a, b])$ gilt

$$\begin{aligned}\text{Var}(u, [a, b]) &= \text{Var}^+(u, [a, b]) + \text{Var}^-(u, [a, b]) \\ u(b) - u(a) &= \text{Var}^+(u, [a, b]) - \text{Var}^-(u, [a, b]).\end{aligned}$$

Beweis. Seien $M^k = \{M_i^k\}_{i=0}^{m_k}$, $N^k = \{N_i^k\}_{i=0}^{n_k}$ und $P^k = \{P_i^k\}_{i=0}^{p_k}$ Folgen von Partitionen von $[a, b]$, sodass

$$\begin{aligned}\lim_{k \rightarrow \infty} \sum_{i=1}^{m_k} |u(M_i^k) - u(M_{i-1}^k)| &= \text{Var}(u, [a, b]) \\ \lim_{k \rightarrow \infty} \sum_{i=1}^{n_k} \max(u(N_i^k) - u(N_{i-1}^k), 0) &= \text{Var}^+(u, [a, b]), \\ - \lim_{k \rightarrow \infty} \sum_{i=1}^{p_k} \min(u(P_i^k) - u(P_{i-1}^k), 0) &= \text{Var}^-(u, [a, b])\end{aligned}$$

gilt. Sei außerdem $Q^k = \{Q_i^k\}_{i=0}^{q_k}$ eine gemeinsame Verfeinerung der Partitionen M^k , N^k und P^k , es gelte also $M^k \subset Q^k$, $N^k \subset Q^k$ und $P^k \subset Q^k$ für alle $k \in \mathbb{N}$. Die positive, negative und totale Variation von u lassen sich auch durch den Grenzwert der Summen über den Partitionen Q^k darstellen, da durch die Dreiecksungleichung die Summen beim Hinzufügen von weiteren Stützpunkten nicht kleiner und durch die Supremumsbildung über alle Partitionen nicht größer als die entsprechenden Variationen werden können.

Wir definieren die Indexmengen

$$I_+^{Q^k} := \{i \in \{1, \dots, q_k\} \mid u(Q_i^k) \geq u(Q_{i-1}^k)\}$$

$$I_-^{Q^k} := \{i \in \{1, \dots, q_k\} \mid u(Q_i^k) < u(Q_{i-1}^k)\}$$

und folgern

$$\begin{aligned} \text{Var}(u, [a, b]) &= \lim_{k \rightarrow \infty} \sum_{i=1}^{m_k} |u(M_i^k) - u(M_{i-1}^k)| \\ &= \lim_{k \rightarrow \infty} \sum_{i=1}^{q_k} |u(Q_i^k) - u(Q_{i-1}^k)| \\ &= \lim_{k \rightarrow \infty} \left(\sum_{i \in I_+^{Q^k}} (u(Q_i^k) - u(Q_{i-1}^k)) - \sum_{i \in I_-^{Q^k}} (u(Q_i^k) - u(Q_{i-1}^k)) \right) \\ &= \lim_{k \rightarrow \infty} \left(\sum_{i=1}^{q_k} \max(u(Q_i^k) - u(Q_{i-1}^k), 0) - \sum_{i=1}^{q_k} \min(u(Q_i^k) - u(Q_{i-1}^k), 0) \right) \\ &= \lim_{k \rightarrow \infty} \left(\sum_{i=1}^{q_k} \max(u(Q_i^k) - u(Q_{i-1}^k), 0) - \lim_{k \rightarrow \infty} \sum_{i=1}^{q_k} \min(u(Q_i^k) - u(Q_{i-1}^k), 0) \right) \\ &= \text{Var}^+(u, [a, b]) + \text{Var}^-(u, [a, b]). \end{aligned}$$

Mit der Folge der Verfeinerungen Q^k können die positive und negative Variation als

$$\text{Var}^+(u, [a, b]) = \lim_{k \rightarrow \infty} \sum_{i \in I_+^{Q^k}} (u(Q_i^k) - u(Q_{i-1}^k))$$

$$\text{Var}^-(u, [a, b]) = - \lim_{k \rightarrow \infty} \sum_{i \in I_-^{Q^k}} (u(Q_i^k) - u(Q_{i-1}^k))$$

geschrieben werden. Wegen $I_+^{Q^k} \cup I_-^{Q^k} = \{1, \dots, q_k\}$ und $I_+^{Q^k} \cap I_-^{Q^k} = \emptyset$ folgt

$$\begin{aligned} \text{Var}^+(u, [a, b]) - \text{Var}^-(u, [a, b]) &= \lim_{k \rightarrow \infty} \sum_{i=1}^{q_k} (u(Q_i^k) - u(Q_{i-1}^k)) \\ &= \lim_{k \rightarrow \infty} u(Q_{q_k}^k) - u(Q_0^k) \\ &= \lim_{k \rightarrow \infty} u(b) - u(a) = u(b) - u(a). \end{aligned}$$

□

Nun kann u als Differenz der positiven und negativen Variation umgeschrieben werden.

Satz 2.7. Sei $u \in BV_p([a, b])$ gegeben. Die Funktionen $\text{Var}_u^+(t) := \text{Var}^+(u, [a, t])$ und $\text{Var}_u^-(t) := \text{Var}^-(u, [a, t])$ sind monoton wachsend und u kann als $u(t) = u(a) + \text{Var}_u^+(t) - \text{Var}_u^-(t)$ geschrieben werden.

Beweis. Die Monotonie von Var_u^+ und Var_u^- folgt aus der Tatsache, dass für $s, t \in [a, b]$, $s < t$ mit Korollar 2.5 die Beziehungen

$$\begin{aligned}\text{Var}_u^+(s) &= \text{Var}_u^+(t) - \text{Var}^+(u, [s, t]) \leq \text{Var}_u^+(t) \\ \text{Var}_u^-(s) &= \text{Var}_u^-(t) - \text{Var}^-(u, [s, t]) \leq \text{Var}_u^-(t)\end{aligned}$$

gelten.

Mit Satz 2.6 folgt

$$u(a) + \text{Var}_u^+(t) - \text{Var}_u^-(t) = u(a) + u(t) - u(a) = u(t).$$

□

Schließlich erhalten wir mit der Zerlegung ein nützliches Resultat, welches die Existenz von links- und rechtsseitigen Grenzwerten für Funktionen beschränkter Variation sichert.

Satz 2.8. Eine Funktion $u \in BV_p([a, b])$ hat höchstens abzählbar viele Unstetigkeitsstellen und in jedem $t \in (a, b)$ einen linksseitigen sowie in jedem $t \in [a, b)$ einen rechtsseitigen Grenzwert.

Beweis. Sei U die Menge der Unstetigkeitsstellen von u und seien U_+, U_- die Mengen der Unstetigkeitsstellen von Var_u^+ und Var_u^- . Es gilt $U \subset U_+ \cup U_-$, denn wegen der Darstellung $u(t) = u(a) + \text{Var}_u^+(t) - \text{Var}_u^-(t)$ muss $t_0 \in U$ auch in U_+ oder U_- liegen, sonst wäre die rechte Seite stetig in t_0 . Demnach genügt es zu zeigen, dass U_+ und U_- höchstens abzählbar sind. Hierfür kann die Monotonie von Var_u^+ und Var_u^- benutzt werden: Nach [6, Satz 4.30] haben monotone Funktionen auf Intervallen höchstens abzählbar viele Unstetigkeitsstellen.

Die Existenz eines links- und rechtsseitigen Grenzwerts in jedem $t \in (a, b)$ kann mit [6, Satz 4.29] begründet werden, laut Aussage des Satzes haben Var_u^+ und Var_u^- als monotone Funktionen für alle $t \in (a, b)$ einen links- und rechtsseitigen Grenzwert in t , dies gilt deswegen auch für u als Differenz der Funktionen. Wir müssen also nur noch zeigen, dass u in a einen rechtsseitigen Grenzwert und in b einen linksseitigen Grenzwert besitzt.

Dazu definieren wir die Funktion $\tilde{u} : [a - 1, b + 1] \rightarrow \mathbb{R}$ gemäß

$$\tilde{u}(t) := \begin{cases} u(t), & \text{wenn } t \in [a, b] \\ 0, & \text{sonst} \end{cases}.$$

Mit Satz 2.4 kann man schnell einsehen, dass wegen

$$\begin{aligned} \text{Var}(\tilde{u}, [a - 1, b + 1]) &= \text{Var}(\tilde{u}, [a - 1, a]) + \text{Var}(\tilde{u}, [a, b]) + \text{Var}(\tilde{u}, [b, b + 1]) \\ &= |u(a)| + \text{Var}(\tilde{u}, [a, b]) + |u(b)| < \infty \end{aligned}$$

$\tilde{u} \in \text{BV}_p([a, b])$ gilt. Damit können die bisherigen Erkenntnisse des Beweises benutzt werden, um zu begründen, dass \tilde{u} in a einen rechtsseitigen und in b einen linksseitigen Grenzwert besitzt. Aufgrund der Definition von \tilde{u} muss dies also auch für u gelten. \square

2.2 Variation von Funktionen in Lebesgueräumen

Da wir im Hinblick auf Optimalsteuerungsprobleme mit beschränkter Variation mit Funktionen in Lebesgueräumen arbeiten wollen, ist es nicht sinnvoll, eine Definition der Variation zu benutzen, die von den punktwisen Funktionswerten abhängt. Um diese Problematik zu umgehen, führen wir den folgenden Begriff ein.

Definition 2.9. Die essentielle Variation einer Funktion $u : [a, b] \rightarrow \mathbb{R}$ ist gegeben durch

$$\text{eVar}(u, (a, b)) := \inf \{ \text{Var}(v, (a, b)) : v = u \text{ fast überall in } (a, b) \}.$$

Die essentielle Variation einer Funktion $u \in L^1((a, b))$ ist analog definiert.

Für die essentielle Variation von Funktionen $u \in L^1((a, b))$ kann eine weitere Darstellung gefunden werden, die in vielen Fällen praktischer ist. Aus [7, Theorem 3.27] ist zu entnehmen, dass die folgende Definition äquivalent zu Definition 2.9 für Funktionen $u \in L^1((a, b))$ ist.

Definition 2.10. Für $u \in L^1((a, b))$ gilt

$$\text{eVar}(u, (a, b)) := \sup \left\{ \int_a^b u(t) \varphi'(t) dt : \varphi \in C_c^1((a, b)), \|\varphi\|_{L^\infty((a, b))} \leq 1 \right\}$$

Moderne Definitionen der Variation benutzen die distributionelle Ableitung einer Funktion u . Diese kann mit einem Maß identifiziert werden, welches wir im Folgenden als Du bezeichnen. An dieser Stelle sei bemerkt, dass auch für ein signiertes Maß $\mu: \mathfrak{A} \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$, wobei \mathfrak{A} eine σ -Algebra ist, die Totalvariation über eine Menge A durch die Vorschrift

$$|\mu|(A) := \sup \left\{ \sum_{j=1}^{\infty} |\mu(A_j)| : A_j \in \mathfrak{A} \ (j \geq 1) \text{ disjunkt}, A = \bigcup_{j=1}^{\infty} A_j \right\}$$

definiert werden kann (vgl. [8, Kapitel VII, §1, Abschnitt 3]).

Wir wollen nun einen weiteren Raum von Funktionen beschränkter Variation definieren, der sich auf Funktionen aus $L^1((a, b))$ einschränkt.

Definition 2.11. Der Raum $BV((a, b))$ beinhaltet alle Funktionen $u \in L^1((a, b))$, die eine der folgenden (äquivalenten) Bedingungen erfüllen:

- i) $e\text{Var}(u, (a, b)) < \infty$.
- ii) Die distributionelle Ableitung von u ist darstellbar als endliches Radon-Maß Du in (a, b) , sodass gilt

$$\int_a^b u(t)\varphi'(t) dt = - \int_a^b \varphi dDu \quad \forall \varphi \in C_c^\infty((a, b)).$$

Die Äquivalenz der beiden Bedingungen sowie die Gültigkeit von

$$|Du|((a, b)) = e\text{Var}(u, (a, b))$$

folgt erneut aus [7, Abschnitt 3.1, S. 119-120].

Um die im vorigen Abschnitt hergeleiteten Erkenntnisse, die auf der Variation über dem abgeschlossenen Intervall $[a, b]$ basierten, auf Funktionen in $BV((a, b))$, die auf dem offenen Intervall (a, b) definiert sind zu übertragen, wird der nächste Satz benötigt.

Satz 2.12. Für $u \in BV_p((a, b))$ gilt

$$e\text{Var}(u, (a, b)) = \inf \{ \text{Var}(v, [a, b]) : v = u \text{ fast überall in } (a, b) \}.$$

Beweis. Wir halten zunächst die Ungleichung

$$\text{Var}(v, (a, b)) \leq \sup \left\{ |v(t_0) - u(a)| + \sum_{i=1}^n |v(t_i) - v(t_{i-1})| + |v(b) - v(t_n)| : \right. \\ \left. n \geq 2, a < t_0 < \dots < t_n < b \right\} \leq \text{Var}(v, [a, b])$$

für eine beliebige Funktion $v: [a, b] \rightarrow \mathbb{R}$ fest, mit der wir die Abschätzung

$$\text{eVar}(u, (a, b)) \leq \inf \{ \text{Var}(v, [a, b]) : v = u \text{ fast überall in } (a, b) \}$$

erhalten. Sei $\tilde{u}: [a, b] \rightarrow \mathbb{R}$ die Funktion, die gemäß $\tilde{u}(t) = u(t)$ für $t \in (a, b)$ sowie

$$\tilde{u}(a) := \lim_{t \downarrow a} u(t), \quad \tilde{u}(b) := \lim_{t \uparrow b} u(t) \quad (2.1)$$

definiert ist. Die Existenz dieser Grenzwerte ist mit Satz 2.8 gegeben. Sei $M^k = \{M_i^k\}_{i=0}^{m_k}$ eine Folge von Partitionen, sodass

$$\lim_{k \rightarrow \infty} \sum_{i=1}^{m_k} |\tilde{u}(M_i^k) - \tilde{u}(M_{i-1}^k)| = \text{Var}(\tilde{u}, [a, b])$$

gilt. Mit der Dreiecksungleichung und der Definition von $\text{Var}(\tilde{u}, (a, b))$ sowie der rechtsseitigen Stetigkeit von \tilde{u} in a bzw. der linksseitigen Stetigkeit in b finden wir nun die folgende Abschätzung:

$$\begin{aligned} \text{Var}(\tilde{u}, [a, b]) &= \lim_{k \rightarrow \infty} \left| \tilde{u}(M_1^k) - \tilde{u}(a) \right| + \sum_{i=2}^{m_k-1} \left| \tilde{u}(M_i^k) - \tilde{u}(M_{i-1}^k) \right| + \left| \tilde{u}(b) - \tilde{u}(M_{m_k-1}^k) \right| \\ &\leq \lim_{k \rightarrow \infty} \left| \tilde{u} \left(a + \frac{M_1^k - a}{k} \right) - \tilde{u}(a) \right| + \left| \tilde{u}(M_1^k) - \tilde{u} \left(a + \frac{M_1^k - a}{k} \right) \right| \\ &\quad + \sum_{i=2}^{m_k-1} \left| \tilde{u}(M_i^k) - \tilde{u}(M_{i-1}^k) \right| + \left| \tilde{u} \left(b - \frac{b - M_{m_k-1}^k}{k} \right) - \tilde{u}(M_{m_k-1}^k) \right| \\ &\quad + \left| \tilde{u}(b) - \tilde{u} \left(b - \frac{b - M_{m_k-1}^k}{k} \right) \right| \\ &\leq \text{Var}(\tilde{u}, (a, b)) + \lim_{k \rightarrow \infty} \left| \tilde{u} \left(a + \frac{M_1^k - a}{k} \right) - \tilde{u}(a) \right| \\ &\quad + \left| \tilde{u}(b) - \tilde{u} \left(b - \frac{b - M_{m_k-1}^k}{k} \right) \right| = \text{Var}(\tilde{u}, (a, b)). \end{aligned} \quad (2.2)$$

Es gilt also $\text{Var}(\tilde{u}, [a, b]) = \text{Var}(\tilde{u}, (a, b)) = \text{Var}(u, (a, b))$.

Sei nun $\{v_k\}_{k \in \mathbb{N}}$ eine Folge von Funktionen mit den Eigenschaften $v_k = u$ fast überall

in (a, b) sowie

$$\lim_{k \rightarrow \infty} \text{Var}(v_k, (a, b)) = \text{eVar}(u, (a, b)).$$

Wenn nun \tilde{v}_k die stetigen Fortsetzungen auf $[a, b]$ von v_k für alle $k \in \mathbb{N}$ sind (in dem Sinne, das $\tilde{v}_k(a)$ und $\tilde{v}_k(b)$ wie in (2.1) definiert sind), können wir wegen $\tilde{v}_k = u$ fast überall auf (a, b) und nach vorigem Vorgehen die Abschätzung

$$\begin{aligned} \lim_{k \rightarrow \infty} \text{Var}(v_k, (a, b)) &= \lim_{k \rightarrow \infty} \text{Var}(\tilde{v}_k, [a, b]) \\ &\geq \inf\{\text{Var}(v, [a, b]) : v = u \text{ fast überall in } (a, b)\} \end{aligned}$$

formulieren, aus der schließlich die Aussage folgt. \square

Nun kann aus [7, Abschnitt 3.2, S. 136] entnommen werden, dass zu $u \in L^1((a, b))$ ein Repräsentant \bar{u} aus der Äquivalenzklasse von u existiert, für den das Infimum aus Definition 2.9 angenommen wird, es gilt also $\text{eVar}(u, (a, b)) = \text{Var}(\bar{u}, (a, b))$. Ist nun u ein Element von $\text{BV}((a, b))$, so folgt

$$\text{Var}(\bar{u}, (a, b)) = \text{eVar}(u, (a, b)) < \infty$$

und damit $\bar{u} \in \text{BV}_p((a, b))$. Jetzt kann Satz 2.12 angewendet werden, um die Gleichheit

$$\text{eVar}(u, (a, b)) = \text{eVar}(\bar{u}, (a, b)) = \text{Var}(\bar{u}, (a, b)) = \text{Var}(\tilde{u}, [a, b]) \quad (2.3)$$

zu erhalten, wobei \tilde{u} die Funktion ist, die man erhält, wenn \bar{u} stetig auf $[a, b]$ fortgesetzt wird, \tilde{u} ist also definiert durch

$$\tilde{u}(t) := \begin{cases} \lim_{s \downarrow a} \bar{u}(s), & \text{wenn } t = a \\ \bar{u}(t), & \text{wenn } t \in (a, b) \\ \lim_{s \uparrow b} \bar{u}(s), & \text{wenn } t = b. \end{cases}$$

Die Gleichheit

$$\text{Var}(\bar{u}, (a, b)) = \text{Var}(\tilde{u}, [a, b])$$

erhält man dabei analog zu (2.2). Einen solchen Repräsentanten \bar{u} nennen wir im Folgenden einen guten Repräsentanten von u , die Fortsetzung \tilde{u} von \bar{u} bezeichnen wir als gute Fortsetzung von \bar{u} . Wenn wir eine Funktion $u \in \text{BV}((a, b))$ mit der guten Fortsetzung \tilde{u} eines guten Repräsentanten identifizieren, so gilt wegen (2.3) $\tilde{u} \in \text{BV}_p([a, b])$ und damit können die in Abschnitt 2.1 hergeleiteten Ergebnisse auf u angewendet werden.

Der Raum $\text{BV}((a, b))$ kann nun mit einer Norm ausgestattet werden. Wir definieren

diese wie folgt:

$$\|u\|_{\text{BV}((a,b))} := \|u\|_{L^1((a,b))} + |Du|((a,b)).$$

Schließlich werden im Folgenden noch einige Sätze genannt, die in dieser Arbeit eine Rolle spielen werden.

Satz 2.13. $\text{BV}((a,b))$ ist isometrisch isomorph zum Dualraum eines separablen Banachraums.

Beweis. Siehe [7, Remark 3.12]. □

Satz 2.14. Für eine Folge $\{u_k\}_{k \in \mathbb{N}} \subset \text{BV}((a,b))$ gilt genau dann $u_k \xrightarrow{*} u$ in $\text{BV}((a,b))$, wenn $u_k \rightarrow u$ in $L^1((a,b))$ gilt und eine der folgenden Bedingungen erfüllt ist:

- a) $\{u_k\}_{k \in \mathbb{N}}$ ist in $\text{BV}((a,b))$ beschränkt.
- b) $Du_k \xrightarrow{*} Du$, also

$$\lim_{k \rightarrow \infty} \int_a^b \varphi \, dDu_k = \int_a^b \varphi \, dDu \quad \forall \varphi \in C_0((a,b)).$$

Beweis. Siehe [7, Definition 3.11, Proposition 3.13]. □

Satz 2.15. $\text{BV}((a,b))$ ist stetig in $L^\infty((a,b))$ und kompakt in $L^p((a,b))$ für alle $p \in [1, \infty)$ eingebettet.

Beweis. Siehe [7, Corollary 3.49]. □

Satz 2.16. Aus $u_k \xrightarrow{*} u$ in $\text{BV}((a,b))$ für eine Folge $\{u_k\}_{k \in \mathbb{N}} \subset \text{BV}((a,b))$ folgt $u_k \rightarrow u$ in $L^p((a,b))$ für alle $p \in [1, \infty)$.

Beweis. Mit Satz 2.14 wissen wir, dass $\|u_k - u\|_{L^1((a,b))} \rightarrow 0$ gilt und $\{u_k\}_{k \in \mathbb{N}}$ in $\text{BV}((a,b))$ beschränkt ist. Demnach ist auch die Folge $\{u_k - u\}_{k \in \mathbb{N}}$ beschränkt in $\text{BV}((a,b))$. Mit Satz 2.15 können wir schlussfolgern, dass die Beschränktheit auch bezüglich des $L^\infty((a,b))$ erfüllt ist, das heißt, es gibt ein $C > 0$, sodass $\|u_k - u\|_{L^\infty((a,b))} \leq C < \infty$ gilt. Nun können wir die Interpolationsungleichung (vgl. [9, Appendix B.2.h.]) für die Rechnung

$$\|u_k - u\|_{L^p((a,b))} \leq \|u_k - u\|_{L^1((a,b))}^{1/p} \|u_k - u\|_{L^\infty((a,b))}^{1-1/p} \rightarrow 0$$

nutzen, woraus die Aussage des Satzes folgt. \square

Satz 2.17. Sei $u \in \text{BV}((a, b))$. Dann existiert ein eindeutiges $c \in \mathbb{R}$, sodass

$$u^l(t) := c + Du((a, t)), \quad u^r(t) := c + Du((a, t])$$

der linksseitig und rechtsseitig stetige Repräsentant von u ist. Weiterhin sind u^l und u^r gute Repräsentanten von u . Außerdem ist der Grenzwert $\lim_{t \downarrow a} \bar{u}(t)$, wobei \bar{u} ein guter Repräsentant von u ist, unabhängig von der Wahl des guten Repräsentanten und entspricht c .

Beweis. Mit [7, Theorem 3.28 a)] erhalten wir die Existenz eines eindeutigen (aber noch unbekanntem) $c \in \mathbb{R}$ aus dem Satz. Zudem ist zu erkennen, dass u^r und u^l gute Repräsentanten sind. Der Grenzwert $\lim_{t \downarrow a} \bar{u}(t)$ existiert für jeden guten Repräsentanten \bar{u} von u , da bei Betrachtung der guten Fortsetzung \tilde{u} von \bar{u} auf $[a, b]$ aus $u \in \text{BV}((a, b))$ wegen $e\text{Var}(u, (a, b)) = \text{Var}(\tilde{u}, [a, b])$ $\tilde{u} \in \text{BV}_p([a, b])$ geschlussfolgert und somit Satz 2.8 angewendet werden kann.

Es muss noch gezeigt werden, dass der Grenzwert $\lim_{t \downarrow a} \bar{u}(t)$ nicht von der Wahl des guten Repräsentanten abhängt. Seien also u_1 und u_2 gute Repräsentanten von u . Nach Definition eines guten Repräsentanten gilt $u_1 = u = u_2$ fast überall, das heißt, es gibt eine höchstens abzählbare Menge $M \subset (a, b)$ mit $u_1(t) \neq u_2(t)$ für alle $t \in M$. Wir wählen nun eine Folge $\{t_k\}_{k \in \mathbb{N}} \subset (a, b) \setminus M$ mit $t_k \rightarrow a$. Es folgt

$$\lim_{k \rightarrow \infty} u_1(t_k) = \lim_{k \rightarrow \infty} u_2(t_k).$$

Da nach Satz 2.8 der linksseitige Grenzwert in a für die stetigen Fortsetzungen \tilde{u}_1 und \tilde{u}_2 von u_1 und u_2 existiert, muss dieser bei beiden Funktionen gleich sein, da dies unter Betrachtung der Folge $\{t_k\}_{k \in \mathbb{N}}$ schon der Fall ist. Demnach muss auch

$$\lim_{t \downarrow a} u_1(t) = \lim_{t \downarrow a} u_2(t)$$

gelten.

Da u^l ein guter Repräsentant von u ist, kann wegen

$$\lim_{t \downarrow a} u^l(t) = c + \lim_{t \downarrow a} Du((a, t)) = c$$

bestätigt werden, dass der Wert von c durch den im Satz benannten Grenzwert bestimmt werden kann. \square

Kapitel 3

Problemstellung

In dieser Arbeit werden Optimalsteuerungsprobleme der Form

$$\begin{aligned} \text{Minimiere } & \frac{1}{2} \int_a^b (y(t) - r(t))^\top Q(t) (y(t) - r(t)) dt + \beta_1 |Du|((a, b)) + \beta_2 \Psi_\kappa(u) \\ \text{sodass } & y'(t) = F(t, u(t), y(t)) \text{ f.ü. auf } (a, b) \\ & y(a) = y_0 \\ & u \in U_{ad} := \{u \in \text{BV}((a, b)) : u(t) \in \{\nu_1, \dots, \nu_d\} \subset \mathbb{R} \text{ f.ü. auf } (a, b)\} \end{aligned} \tag{P}$$

mit $a, b \in \mathbb{R}$, $a < b$, $F: \mathbb{R} \times \mathbb{R} \times \mathbb{R}^N \rightarrow \mathbb{R}^N$, $Q \in L^\infty((a, b))^{N \times N}$, $r \in L^2((a, b))^N$, den Funktionen $\Psi_\kappa: \text{BV}((a, b)) \rightarrow \mathbb{R}$ für $\kappa \in \{1, \dots, 4\}$ sowie $y_0 \in \mathbb{R}^N$, $\beta_1, \beta_2 \in \mathbb{R}$ mit $\beta_1 > 0$ und $\beta_2 > 0$ betrachtet. Die reellen Zahlen ν_1, \dots, ν_d seien dabei aufsteigend der Größe nach sortiert und paarweise verschieden, es gelte also $\nu_i < \nu_j$ für $i < j$.

Die Funktion Ψ_κ wird hierbei je nach gewünschtem Verhalten einer optimalen Steuerung als eine der folgenden Funktionen gewählt:

$$\begin{aligned} \text{(i)} \quad \Psi_1(u) &:= 0 & \text{(ii)} \quad \Psi_2(u) &:= \left(\int_a^b u(t) dt \right)^2 \\ \text{(iii)} \quad \Psi_3(u) &:= |u(a)| & \text{(iv)} \quad \Psi_4(u) &:= \|u\|_{L^1((a, b))}. \end{aligned}$$

Die Wahl $\kappa = 2$ begünstigt Funktionen, die im Mittelwert nahe an der Null liegen, während mit $\kappa = 4$ betragsmäßig kleine Funktionen bevorzugt werden. Im Falle von $\kappa = 3$ muss wegen $u \in L^1((a, b))$ noch spezifiziert werden, dass der Ausdruck $u(a)$ als rechtsseitiger Grenzwert $\lim_{t \downarrow a} \bar{u}(t)$ für einen guten Repräsentanten \bar{u} von u zu interpretieren ist. Die Wohldefiniertheit des Ausdrucks kann mithilfe von Satz 2.17 bestätigt werden. Da in Anwendungsfällen die Steuerung häufig zum Start eines betrachteten Zeitintervalls, also im Zeitpunkt a , noch nicht aktiv ist (und damit je nach Modellierung den Wert 0 annehmen könnte, vgl. Beispiel 1.2), kann der Wert

$|u(a)|$ für mit $\kappa \neq 3$ berechnete Lösungen bereits groß sein, womit implizit eine Variation geschah, die nicht vom Term $|Du|((a, b))$ bestraft wurde. In solch einem Fall eignet sich die Option $\kappa = 3$.

Die Wahl des Zielfunktional erlaubt lineare und quadratische Terme bezüglich des Zustands y und ermöglicht die Forderung einiger Qualitätsmerkmale an die Steuerung, je nach Wahl von β_1 , β_2 und κ . Wird beispielsweise die Zielstellung verfolgt, mit dem Zustand y einen Zielzustand $y_d \in L^2((a, b))^N$ so gut wie möglich zu approximieren, so kann dies mit der Setzung

$$Q(t) = I_N := \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}, \quad r(t) = y_d(t)$$

erreicht werden, in diesem Szenario folgt nämlich

$$\begin{aligned} \int_a^b (y(t) - r(t))^\top Q(t) (y(t) - r(t)) dt &= \int_a^b (y(t) - y_d(t))^\top (y(t) - y_d(t)) dt \\ &= \int_a^b \sum_{i=1}^N (y_i(t) - y_{d_i}(t))^2 dt \\ &= \sum_{i=1}^N \|y_i - y_{d_i}\|_{L^2((a,b))}^2 = \|y - y_d\|_{L^2((a,b))^N}^2, \end{aligned}$$

wobei y_i und y_{d_i} für die i -ten Komponenten von y und y_d stehen. Mit dieser Vorgehensweise lässt sich auch das Beispiel 1.2 in der Form (P) mit dem Regularisierungsterm $\beta_1|Du|((a, b))$ modellieren:

Beispiel 3.1. (Lotka-Volterra Fishing Problem)

Wir modellieren die Problemstellung aus Beispiel 1.2 mit der neuen Zielfunktion

$$J(u) := \frac{1}{2} \int_0^T (y_1(t) - x_d(t))^2 + (y_2(t) - z_d(t))^2 dt + \beta_1|Du|((0, T))$$

als eine (P) entsprechende Problemstellung mit den Setzungen

$$(a, b) = (0, T), \quad N = 2, \quad Q(t) = I_2, \quad \kappa = 1, \quad d = 2, \quad (\nu_1, \nu_2) = (0, 1)$$

$$r(t) = \begin{pmatrix} x_d(t) \\ z_d(t) \end{pmatrix}, \quad y_0 = \begin{pmatrix} x_0 \\ z_0 \end{pmatrix},$$

$$F(t, u(t), y(t)) = \begin{pmatrix} \alpha y_1(t) - \beta y_1(t)y_2(t) - \theta_1 y_1(t)u(t) \\ \gamma y_1(t)y_2(t) - \delta y_2(t) - \theta_2 y_2(t)u(t) \end{pmatrix},$$

die Bedingungen $Q \in L^\infty((a, b))^{N \times N}$ und $r \in L^2((a, b))^N$ sind somit erfüllt.

Auch Beispiel 1.1 kann in der Form von (P) formuliert werden:

Beispiel 3.2. (Temperatursteuerung einer Herdplatte)

Wir setzen

$$(a, b) = (0, \mathcal{T}), \quad N = 2, \quad \kappa = 4, \quad \{\nu_1, \dots, \nu_d\} = \{0, 0.5, \dots, u_{max}\}, \quad \beta_2 = \beta,$$

$$Q(t) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad r(t) = \begin{pmatrix} T_d(t) \\ 0 \end{pmatrix}, \quad y_0 = \begin{pmatrix} T_Z \\ T_Z \end{pmatrix},$$

$$F(t, u(t), y(t)) = \begin{pmatrix} \alpha_1(y_2(t) - y_1(t)) - \alpha_2(y_1(t) - T_Z) \\ \gamma_1(100u(t) - y_2(t)) + \gamma_2(y_1(t) - y_2(t)) - \gamma_3(y_2(t) - T_Z) \end{pmatrix},$$

somit sind erneut $Q \in L^\infty((a, b))^{N \times N}$ und $r \in L^2((a, b))^N$ erfüllt.

Für die Untersuchungen dieser Arbeit sind weitere Anforderungen an einige der in (P) vorkommenden Terme nötig, wir fordern also die folgenden Voraussetzungen an Probleme der Form (P):

- (V1) Die Matrix Q sei positiv semidefinit und symmetrisch für alle $t \in [a, b]$. Außerdem sei F messbar bezüglich des ersten Arguments und Lipschitzstetig bezüglich des zweiten und dritten Arguments mit den Konstanten $L_1 > 0$ und $L_2 > 0$ in dem Sinne, dass die Gleichungen

$$\begin{aligned} \|F(t, v_1, y(t)) - F(t, v_2, y(t))\|_2 &\leq L_1 |v_1 - v_2| \\ \|F(t, u(t), z_1) - F(t, u(t), z_2)\|_2 &\leq L_2 \|z_1 - z_2\|_2 \end{aligned} \tag{3.1}$$

für alle $t \in (a, b)$, $v_1, v_2 \in \mathbb{R}$ und $z_1, z_2 \in \mathbb{R}^N$ erfüllt sind. Im Falle von $\kappa = 3$ gelte $\beta_1 \geq \beta_2$.

- (V2) Es existieren die partiellen Ableitungen nach u und y_i für $i = 1, \dots, N$ in allen Komponenten von F . Wir bezeichnen die Jacobimatrizen, die aus der

Differentiation von F nach u bzw. y entstehen, als F_u bzw. F_y . Außerdem gelte $t \mapsto F_u(t, u(t), y(t)) \in L^\infty((a, b))^N$ und $t \mapsto F_y(t, u(t), y(t)) \in L^\infty((a, b))^{N \times N}$.

(V3) Seien F_u und F_y Lipschitzstetig bezüglich derer zweiten und dritten Argumente, das heißt, es gelte

$$\begin{aligned} \|F_u(t, v_1, y(t)) - F_u(t, v_2, y(t))\|_2 &\leq L_3|v_1 - v_2| \\ \|F_u(t, u(t), z_1) - F_u(t, u(t), z_2)\|_2 &\leq L_4\|z_1 - z_2\|_2 \\ \|F_y(t, v_1, y(t)) - F_y(t, v_2, y(t))\|_2 &\leq L_5|v_1 - v_2| \\ \|F_y(t, u(t), z_1) - F_y(t, u(t), z_2)\|_2 &\leq L_6\|z_1 - z_2\|_2 \end{aligned} \tag{3.2}$$

für alle $t \in (a, b)$, $v_1, v_2 \in \mathbb{R}$, $z_1, z_2 \in \mathbb{R}^N$ und Konstanten $L_3 > 0$, $L_4 > 0$, $L_5 > 0$, $L_6 > 0$.

Die Voraussetzungen (V1) werden bei den Beweisen zur Existenz einer Lösung und die Voraussetzungen (V2) bei der Betrachtung der Differenzierbarkeitseigenschaften des Zielfunktional verwendet. (V3) wird benötigt, um gute Eigenschaften für den Lösungsalgorithmus nachweisen zu können.

Wir versuchen nun, die Gültigkeit der Voraussetzungen für die Problemstellungen 3.1 und 3.2 zu zeigen und beginnen mit Zweiterem.

Beispiel 3.3. (Fortsetzung von Beispiel 3.2)

Wir zeigen zuerst die Voraussetzungen aus (V1). Q ist offensichtlich symmetrisch und wegen $x^\top Qx = x_1^2$ für $x \in \mathbb{R}^2$ positiv semidefinit. Wir finden die Darstellung

$$F(t, u(t), y(t)) = \begin{pmatrix} -\alpha_1 - \alpha_2 & \alpha_1 \\ \gamma_2 & -\gamma_1 - \gamma_2 - \gamma_3 \end{pmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} + \begin{pmatrix} \alpha_2 T_Z \\ 100\gamma_1 u(t) + \gamma_3 T_Z \end{pmatrix},$$

mit

$$\|F(t, v_1, y(t)) - F(t, v_2, y(t))\|_2 = \left\| \begin{pmatrix} 0 \\ 100\gamma_1(v_1 - v_2) \end{pmatrix} \right\|_2 = 100\gamma_1|v_1 - v_2|$$

für $v_1, v_2 \in \mathbb{R}$ und

$$\begin{aligned}
\|F(t, u(t), z_1) - F(t, u(t), z_2)\|_2^2 &= \left\| \begin{pmatrix} -\alpha_1 - \alpha_2 & \alpha_1 \\ \gamma_2 & -\gamma_1 - \gamma_2 - \gamma_3 \end{pmatrix} (z_1 - z_2) \right\|_2^2 \\
&\leq \left\| \begin{pmatrix} -\alpha_1 - \alpha_2 & \alpha_1 \\ \gamma_2 & -\gamma_1 - \gamma_2 - \gamma_3 \end{pmatrix} \right\|_2^2 \|z_1 - z_2\|_2^2 \\
&\leq \left\| \begin{pmatrix} -\alpha_1 - \alpha_2 & \alpha_1 \\ \gamma_2 & -\gamma_1 - \gamma_2 - \gamma_3 \end{pmatrix} \right\|_F^2 \|z_1 - z_2\|_2^2 \\
&= ((\alpha_1 + \alpha_2)^2 + \alpha_1^2 + \gamma_2^2 + (\gamma_1 + \gamma_2 + \gamma_3)^2) \|z_1 - z_2\|_2^2
\end{aligned}$$

für $z_1, z_2 \in \mathbb{R}^2$ ist (V1) gezeigt. Die Jacobimatrizen F_u und F_y sind gemäß

$$F_u(t, u(t), y(t)) = \begin{pmatrix} 0 \\ 100\gamma_1 \end{pmatrix}, \quad F_y = \begin{pmatrix} -\alpha_1 - \alpha_2 & \alpha_1 \\ \gamma_2 & -\gamma_1 - \gamma_2 - \gamma_3 \end{pmatrix}$$

gegeben und liegen in $L^\infty((a, b))^N$ bzw. $L^\infty((a, b))^{N \times N}$, womit (V2) erfüllt ist. (V3) gilt ebenso, da die Matrizen F_u und F_y nicht von u bzw. y abhängen.

Beispiel 3.4. (Fortsetzung von Beispiel 3.1)

Aufgrund der Nichtlinearitäten in F erhalten wir für das Lotka-Volterra Fishing Problem keine globale Lipschitzstetigkeit bezüglich des zweiten und dritten Arguments von F . Da die meisten Sätze in den Abschnitten 4, 5 und 6 auf dieser Eigenschaft beruhen, lassen sich keine Vorhersagen oder Informationen über die Qualität einer aus dem Lösungsalgorithmus gewonnenen Approximation gewinnen. Da allerdings auch das Problem aus Beispiel 1.2 keine optimale Lösung besitzt und mit einer Diskretisierung nur „gute“ Lösungen (solche, die nahe an der optimalen Lösung des relaxierten Problems liegen) berechnet werden können, lohnt es sich, das Problem auch in unserem Szenario mit dem zusätzlichen Summanden $\beta_1 |Du|((0, T))$ zu betrachten und die daraus entstehenden Lösungen mit denen ohne Regularisierungsterm zu vergleichen.

Wir zeigen trotzdem die Gültigkeit der übrigen Voraussetzungen und berechnen für (V2) die Jacobimatrizen durch

$$F_u(t, u(t), y(t)) = \begin{pmatrix} -\theta_1 y_1(t) \\ -\theta_2 y_2(t) \end{pmatrix}$$

und

$$F_y(t, u(t), y(t)) = \begin{pmatrix} \alpha - \beta y_2(t) - \theta_1 u(t) & -\beta y_1(t) \\ \gamma y_2(t) & \gamma y_1(t) - \delta - \theta_2 u(t) \end{pmatrix}.$$

Wegen der Einbettungen $H^1((a, b))^N \hookrightarrow L^\infty((a, b))^N$ (siehe [10, Satz 8.9]) und $BV((a, b)) \hookrightarrow L^\infty((a, b))$ (siehe Satz 2.15) gelten mit $y \in H^1((a, b))^2$ und $u \in BV((a, b))$ $y_1, y_2, u \in L^\infty((a, b))$ und damit auch $F_u \in L^\infty((a, b))^2$ und $F_y \in L^\infty((a, b))^{2 \times 2}$.

Da in F_u keine von u abhängigen Terme vorkommen, ist die Lipschitzstetigkeit von F_u bezüglich u gegeben. Mit $v, w \in \mathbb{R}^2$ erhalten wir nun aus

$$\begin{aligned} \|F_u(t, u(t), v) - F_u(t, u(t), w)\|_2^2 &= \left\| \begin{pmatrix} -\theta_1(v_1 - w_1) \\ -\theta_2(v_2 - w_2) \end{pmatrix} \right\|_2^2 \\ &= \left\| \begin{pmatrix} -\theta_1 & 0 \\ 0 & -\theta_2 \end{pmatrix} \begin{pmatrix} v_1 - w_1 \\ v_2 - w_2 \end{pmatrix} \right\|_2^2 \\ &\leq \left\| \begin{pmatrix} -\theta_1 & 0 \\ 0 & -\theta_2 \end{pmatrix} \right\|_2^2 \|v - w\|_2^2 \\ &= \max(\theta_1^2, \theta_2^2) \|v - w\|_2^2 \end{aligned}$$

die Lipschitzstetigkeit von F_u bezüglich y . Die Lipschitzbedingungen von F_y erhalten wir mit $\eta_1, \eta_2 \in \mathbb{R}$ und $v, w \in \mathbb{R}^2$ aus

$$\begin{aligned} \|F_y(t, \eta_1, y(t)) - F_y(t, \eta_2, y(t))\|_2^2 &= \left\| \begin{pmatrix} -\theta_1(\eta_1 - \eta_2) & 0 \\ 0 & -\theta_2(\eta_1 - \eta_2) \end{pmatrix} \right\|_2^2 \\ &\leq \left\| \begin{pmatrix} -\theta_1 & 0 \\ 0 & -\theta_2 \end{pmatrix} \right\|_2^2 \left\| \begin{pmatrix} \eta_1 - \eta_2 \\ \eta_1 - \eta_2 \end{pmatrix} \right\|_2^2 \\ &= 2 \max(\theta_1^2, \theta_2^2) |\eta_1 - \eta_2|^2 \end{aligned}$$

und

$$\begin{aligned} \|F_y(t, u(t), v) - F_y(t, u(t), w)\|_2^2 &\leq \|F_y(t, u(t), v) - F_y(t, u(t), w)\|_F^2 \\ &= \left\| \begin{pmatrix} \beta(w_2 - v_2) & \beta(w_1 - v_1) \\ \gamma(v_2 - w_2) & \gamma(v_1 - w_1) \end{pmatrix} \right\|_F^2 \\ &= \beta^2 [(w_2 - v_2)^2 + (w_1 - v_1)^2] + \gamma^2 [(v_2 - w_2)^2 + (v_1 - w_1)^2] \\ &= (\beta^2 + \gamma^2) [(v_1 - w_1)^2 + (v_2 - w_2)^2] = (\beta^2 + \gamma^2) \|v - w\|_2^2. \end{aligned}$$

Für die weitere Untersuchung der Problemstellung bietet es sich an, (P) mit dem Lösungsoperator $S: L^2((a, b)) \rightarrow H^1((a, b))^N$ auszustatten, der mit der Festlegung $S(u) := y$ eine Steuerung u (bzw. allgemeiner eine Funktion in $L^2((a, b))$) auf die Zustandsfunktion und Lösung des Anfangswertproblems

$$\begin{aligned} y'(t) &= F(t, u(t), y(t)) \quad \text{f.ü. auf } (a, b) \\ y(a) &= y_0 \end{aligned} \tag{AWP}$$

abbildet. S erhält Argumente aus dem Raum $L^2((a, b))$, da ein späteres Lösungsverfahren die Minimierung eines veränderten Zielfunktional mit Argumenten aus diesem Raum anstreben wird. Problem (P) ist also äquivalent zu

$$\begin{aligned} \text{Minimiere } J_\kappa(u) &:= \frac{1}{2} \int_a^b (S(u)(t) - r(t))^\top Q(t) (S(u)(t) - r(t)) dt \\ &\quad + \beta_1 |Du|((a, b)) + \beta_2 \Psi_\kappa(u) \\ \text{sodass } u &\in U_{ad} \end{aligned}$$

mit der Funktion $J_\kappa: BV((a, b)) \rightarrow \mathbb{R}$, wenn S wohldefiniert ist, also für alle $u \in L^2((a, b))$ eine eindeutige Lösung für (AWP) existiert. Der Satz von Picard-Lindelöf ist hier allerdings nicht anwendbar, da $(t, y) \mapsto F(t, u(t), y(t))$ für beliebige $u \in L^2((a, b))$ im Allgemeinen nicht stetig bezüglich t ist. Um also die gewünschte Existenz zu garantieren, definieren wir den Operator $G: H^1((a, b))^N \rightarrow L^2((a, b))^N$ für festes $u \in L^2((a, b))$ durch

$$Gy(t) := F(t, u(t), y(t)).$$

Dabei bildet G wegen $y' \in L^2((a, b))^N$ in eben diesen Raum ab. Somit kann (AWP) als Operatordifferentialgleichung

$$y'(t) - Gy(t) = 0$$

aufgefasst werden. Passende Existenzresultate erhalten wir mit [11, S.165, 166]: Aufgrund der Lipschitzstetigkeit von F bezüglich des dritten Arguments schlussfolgern

wir

$$\begin{aligned}
\|Gy_1 - Gy_2\|_{L^2((a,b))^N}^2 &= \|F(\cdot, u(\cdot), y_1(\cdot)) - F(\cdot, u(\cdot), y_2(\cdot))\|_{L^2((a,b))^N}^2 \\
&= \int_a^b \|F(t, u(t), y_1(t)) - F(t, u(t), y_2(t))\|_2^2 dt \\
&\leq \int_a^b L_2^2 \|y_1(t) - y_2(t)\|_2^2 dt \\
&= L_2^2 \|y_1 - y_2\|_{L^2((a,b))^N}^2
\end{aligned}$$

mithilfe der Beobachtung, dass für $y \in L^2((a, b))^N$

$$\|y\|_{L^2((a,b))^N}^2 = \sum_{i=1}^N \|y_i\|_{L^2((a,b))}^2 = \int_a^b \sum_{i=1}^N y_i(t)^2 dt = \int_a^b \|y(t)\|_2^2 dt$$

gilt. Außerdem ist G ein Volterra-Operator nach Definition aus [11], denn wenn für Funktionen $y_1, y_2 \in H^1((a, b))^N$ für ein $s \in (a, b)$ $y_1(t) = y_2(t)$ für fast alle $t \in (a, s]$ gilt, dann folgt

$$Gy_1(t) = F(t, u(t), y_1(t)) = F(t, u(t), y_2(t)) = Gy_2(t)$$

für fast alle $t \in (a, s]$. Mit diesen Erkenntnissen liefert [11, Kap. 5, Satz 1.3] die Existenz einer eindeutigen Lösung y von (AWP).

Kapitel 4

Existenzresultate

Damit eine Untersuchung der Problemstellung sinnvoll ist, muss zunächst nachgewiesen werden, dass eine Lösung für (P) existiert. Ein solcher Existenzbeweis kann wie im Folgenden formuliert werden.

Satz 4.1. Sei (P) wie in Kapitel 3 gegeben, sodass die Voraussetzung (V1) erfüllt ist. Wenn J_κ schwach-* folgenunterhalbstetig ist, existiert eine Lösung für (P).

Beweis. Wir setzen $j := \inf\{J_\kappa(u) : u \in U_{ad}\}$ und wählen eine minimierende Folge $\{u_k\}_{k \in \mathbb{N}} \subset U_{ad}$, sodass $J_\kappa(u_k) \rightarrow j$ gilt.

Zunächst zeigen wir die Beschränktheit von $\{u_k\}_{k \in \mathbb{N}}$ in $BV((a, b))$. Wir finden die Abschätzung

$$\|u_k\|_{L^1((a,b))} = \int_a^b |u_k(t)| dt \leq \int_a^b \max(|\nu_1|, |\nu_d|) dt = (b-a) \max(|\nu_1|, |\nu_d|) =: C_1.$$

Da alle Summanden im Zielfunktional positiv sind und $J_\kappa(u_k)$ eine konvergente Folge ist, gilt außerdem

$$|Du_k|((a, b)) \leq J_\kappa(u_k) \leq C_2$$

für eine Konstante $C_2 > 0$. Wir können also wegen

$$\|u_k\|_{BV((a,b))} \leq C_1 + C_2 < \infty$$

die Beschränktheit von $\{u_k\}_{k \in \mathbb{N}}$ folgern. Da $BV((a, b))$ isometrisch isomorph zum Dualraum eines separablen Banachraums ist (vgl. Satz 2.13), folgt aus der Beschränktheit der Folge $\{u_k\}_{k \in \mathbb{N}}$ die Existenz einer schwach-* konvergenten Teilfolge $\{u_{k_l}\}_{l \in \mathbb{N}}$ mit $u_{k_l} \xrightarrow{*} \bar{u} \in BV((a, b))$. Da J_κ schwach-* folgenunterhalbstetig ist, kann

die Abschätzung

$$j \leq J_\kappa(\bar{u}) \leq \liminf_{l \rightarrow \infty} J_\kappa(u_{k_l}) = \lim_{k \rightarrow \infty} J_\kappa(u_k) = j$$

getätigt werden, falls $\bar{u} \in U_{ad}$.

Mit Satz 2.14 können wir aus $u_{k_l} \xrightarrow{*} \bar{u}$ in $BV((a, b))$ die Gültigkeit von $u_{k_l} \rightarrow \bar{u}$ in $L^1((a, b))$ schließen. Somit gibt es eine Teilfolge $\{u_m\}_{m \in \mathbb{N}} \subset \{u_{k_l}\}_{l \in \mathbb{N}}$, für die $u_m(t) \rightarrow \bar{u}(t)$ für fast alle $t \in (a, b)$ gilt. Wegen $u_m(t) \in \{\nu_1, \dots, \nu_d\}$ f.ü. in (a, b) ist damit auch $\bar{u}(t) \in \{\nu_1, \dots, \nu_d\}$ f.ü. in (a, b) erfüllt, mit $\bar{u} \in BV((a, b))$ schlussfolgern wir also $\bar{u} \in U_{ad}$. \square

Da nur die Existenz einer Lösung unter der Bedingung der schwach-* Folgenunterhalbstetigkeit des Zielfunktional gezeigt wurde, bleibt noch zu zeigen, dass J_κ für jede der vier Funktionen Ψ_κ ebendiese Eigenschaft besitzt. In den folgenden Abschnitten wird die Zielfunktion genauer auf diese Eigenschaft untersucht.

4.1 Stetigkeitseigenschaften des Lösungsoperators

Um die schwach-* Folgenunterhalbstetigkeit von J_κ zu erhalten, ist es nötig, gewisse Stetigkeitseigenschaften für den Lösungsoperator S nachweisen zu können. In diesem Abschnitt wird gezeigt, dass S schwach-* folgenstetig ist. Dazu ist das nächste Lemma hilfreich.

Lemma 4.2. Für eine Familie von Funktionen $f_i \in L^2((a, b))$, $i \in \{1, \dots, N\}$ gilt die Abschätzung

$$\sum_{i=1}^N \left(\int_a^b f_i(x) dx \right)^2 \leq \left(\int_a^b \sqrt{\sum_{i=1}^N f_i(x)^2} dx \right)^2.$$

Beweis. Sei $F_i := \int_a^b f_i(x) dx$, $\vec{F} := (F_1, \dots, F_N)$ und $\vec{f}(x) := (f_1(x), \dots, f_N(x))$. Wegen der Cauchy-Schwarz Ungleichung gilt

$$\|\vec{F}\|_2 = \frac{\langle \vec{F}, \vec{F} \rangle_2}{\|\vec{F}\|_2} = \int_a^b \frac{\langle \vec{F}, \vec{f}(x) \rangle_2}{\|\vec{F}\|_2} dx \leq \int_a^b \frac{\|\vec{F}\|_2 \|\vec{f}(x)\|_2}{\|\vec{F}\|_2} dx = \int_a^b \|\vec{f}(x)\|_2 dx,$$

beidseitiges Quadrieren liefert die Aussage. \square

Grundlage des Beweises der schwach-* Folgenstetigkeit von S ist die folgende Abschätzung, die mithilfe der Grönwall-Ungleichung gewonnen wird.

Lemma 4.3. Seien $y_1, y_2 \in H^1((a, b))^N$ Lösungen von (AWP) zu den Steuerungen $u_1, u_2 \in L^2((a, b))$. Dann gilt die Abschätzung

$$\|y_1'(t) - y_2'(t)\|_2 \leq L_1 L_2 \int_a^t |u_1(\tau) - u_2(\tau)| e^{L_2(t-\tau)} d\tau + L_1 |u_1(t) - u_2(t)|$$

für alle $t \in (a, b)$.

Beweis. Zunächst benutzen wir die Lipschitzstetigkeit von F für die Abschätzung

$$\begin{aligned} \|y_1'(t) - y_2'(t)\|_2 &= \|F(t, u_1(t), y_1(t)) - F(t, u_1(t), y_2(t)) - F(t, u_2(t), y_2(t)) \\ &\quad + F(t, u_1(t), y_2(t))\|_2 \\ &\leq \|F(t, u_1(t), y_1(t)) - F(t, u_1(t), y_2(t))\|_2 \\ &\quad + \|F(t, u_1(t), y_2(t)) - F(t, u_2(t), y_2(t))\|_2 \\ &\leq L_2 \|y_1(t) - y_2(t)\|_2 + L_1 |u_1(t) - u_2(t)|. \end{aligned} \tag{4.1}$$

Mit Lemma 4.2 folgt die Rechnung

$$\begin{aligned} \|y_1(t) - y_2(t)\|_2 &= \sqrt{\sum_{i=1}^N (y_{1_i}(t) - y_{2_i}(t))^2} \\ &= \sqrt{\sum_{i=1}^N \left(y_{0_i} + \int_a^t y_{1_i}'(\tau) d\tau - y_{0_i} - \int_a^t y_{2_i}'(\tau) d\tau \right)^2} \\ &= \sqrt{\sum_{i=1}^N \left(\int_a^t y_{1_i}'(\tau) - y_{2_i}'(\tau) d\tau \right)^2} \\ &\leq \left| \int_a^t \sqrt{\sum_{i=1}^N (y_{1_i}'(\tau) - y_{2_i}'(\tau))^2} d\tau \right| \\ &= \int_a^t \|y_1'(\tau) - y_2'(\tau)\|_2 d\tau, \end{aligned} \tag{4.2}$$

wobei y_{0_i}, y_{1_i} bzw. y_{2_i} die i -te Komponente des jeweiligen Vektors referenziert. Wir erhalten mit (4.1)

$$\|y_1'(t) - y_2'(t)\|_2 \leq L_2 \int_a^t \|y_1'(\tau) - y_2'(\tau)\|_2 d\tau + L_1 |u_1(t) - u_2(t)|,$$

unter Anwendung der Grönwall-Ungleichung (vgl. [9, Appendix B.2.k.]) folgt schließ-

lich die Aussage:

$$\|y'_1(t) - y'_2(t)\|_2 \leq L_1 L_2 \int_a^t |u_1(\tau) - u_2(\tau)| e^{L_2(t-\tau)} d\tau + L_1 |u_1(t) - u_2(t)|.$$

□

Nun kann das Hauptresultat dieses Abschnitts formuliert werden.

Satz 4.4. Der Lösungsoperator S ist schwach-* folgenstetig von $BV((a, b))$ nach $H^1((a, b))^N$, das heißt, aus $u_k \xrightarrow{*} u$ für eine Folge $\{u_k\}_{k \in \mathbb{N}} \subset BV((a, b))$ folgt $S(u_k) \rightarrow S(u)$ bezüglich der H^1 -Norm.

Beweis. Sei $\{u_k\}_{k \in \mathbb{N}} \subset BV((a, b))$ eine Folge mit $u_k \xrightarrow{*} u$, woraus nach Satz 2.14 $u_k \rightarrow u$ bezüglich der L^1 -Norm folgt. Es ist zu zeigen, dass $y_k := S(u_k)$ gegen $y := S(u)$ bezüglich der H^1 -Norm konvergiert.

Aus Lemma 4.3 können wir die Gültigkeit der Ungleichung

$$\|y'_k(t) - y'(t)\|_2 \leq L_1 L_2 \int_a^t |u_k(\tau) - u(\tau)| e^{L_2(t-\tau)} d\tau + L_1 |u_k(t) - u(t)| \quad (4.3)$$

entnehmen. Diese nutzen wir für die Abschätzung

$$\begin{aligned} \|y'_k - y'\|_{L^2((a,b))^N}^2 &= \int_a^b \|y'_k(t) - y'(t)\|_2^2 dt \\ &\leq \int_a^b \left(L_1 L_2 \int_a^t |u_k(\tau) - u(\tau)| e^{L_2(t-\tau)} d\tau + L_1 |u_k(t) - u(t)| \right)^2 dt \\ &\leq \int_a^b \left(L_1 L_2 \int_a^b |u_k(\tau) - u(\tau)| e^{L_2(t-\tau)} d\tau + L_1 |u_k(t) - u(t)| \right)^2 dt \\ &= L_1^2 \int_a^b \left[L_2^2 \left(\int_a^b |u_k(\tau) - u(\tau)| e^{L_2(t-\tau)} d\tau \right)^2 \right. \\ &\quad \left. + 2L_2 \left(\int_a^b |u_k(\tau) - u(\tau)| e^{L_2(t-\tau)} d\tau \right) |u_k(t) - u(t)| \right. \\ &\quad \left. + |u_k(t) - u(t)|^2 \right] dt. \end{aligned} \quad (4.4)$$

Wir betrachten die einzelnen Summanden aus (4.4). Mit der Hölderungleichung (vgl. [8, Kapitel VI, §1, Abschnitt 2]) schreiben wir zunächst

$$\int_a^b \left(\int_a^b |u_k(\tau) - u(\tau)| e^{L_2(t-\tau)} d\tau \right)^2 dt \leq \int_a^b \|u_k - u\|_{L^1((a,b))}^2 \|e^{L_2(t-\cdot)}\|_{L^\infty((a,b))}^2 dt.$$

Aus

$$\|e^{L_2(t-\cdot)}\|_{L^\infty((a,b))} = \sup_{\tau \in [a,b]} |e^{L_2(t-\tau)}| = e^{L_2(t-a)}$$

folgt

$$\int_a^b \|u_k - u\|_{L^1((a,b))}^2 \|e^{L_2(t-\cdot)}\|_{L^\infty((a,b))}^2 dt = \|u_k - u\|_{L^1((a,b))}^2 \frac{e^{2L_2(b-a)} - 1}{2L_2} \rightarrow 0.$$

Ähnliches Vorgehen liefert die Abschätzung

$$\begin{aligned} & \int_a^b \left(\int_a^b |u_k(\tau) - u(\tau)| e^{L_2(t-\tau)} d\tau \right) |u_k(t) - u(t)| dt \\ & \leq \|u_k - u\|_{L^1((a,b))} \int_a^b e^{L_2(t-a)} |u_k(t) - u(t)| dt \\ & \leq \|u_k - u\|_{L^1((a,b))}^2 \|e^{L_2(\cdot-a)}\|_{L^\infty((a,b))} = \|u_k - u\|_{L^1((a,b))}^2 e^{L_2(b-a)} \rightarrow 0. \end{aligned}$$

Schließlich ist noch

$$\int_a^b |u_k(t) - u(t)|^2 dt = \|u_k - u\|_{L^2((a,b))}^2 \rightarrow 0$$

zu zeigen. Diese Aussage folgt direkt aus Satz 2.16.

Aus den drei Abschätzungen schlussfolgern wir $\|y'_k - y'\|_{L^2((a,b))^N}^2 \rightarrow 0$. Mit dieser Kenntnis, der Rechnung (4.2) und der Jensenschen Ungleichung (vgl. [8, Kapitel VI, §1, Abschnitt 1]) folgt

$$\begin{aligned} \|y_k - y\|_{L^2((a,b))^N}^2 &= \int_a^b \|y_k(t) - y(t)\|_2^2 dt \\ &\stackrel{(4.2)}{\leq} \int_a^b \left(\int_a^t \|y'_k(\tau) - y'(\tau)\|_2 d\tau \right)^2 dt \\ &\leq \int_a^b (b-a)^2 \left(\frac{1}{b-a} \int_a^b \|y'_k(\tau) - y'(\tau)\|_2 d\tau \right)^2 dt \quad (4.5) \\ &\leq (b-a) \int_a^b \int_a^b \|y'_k(\tau) - y'(\tau)\|_2^2 d\tau dt \\ &= (b-a) \int_a^b \|y'_k - y'\|_{L^2((a,b))^N}^2 dt \\ &= (b-a)^2 \|y'_k - y'\|_{L^2((a,b))^N}^2 \rightarrow 0 \end{aligned}$$

und somit die Aussage des Satzes. \square

Mithilfe der letzten Abschätzung des Beweises lassen sich nun auch die folgenden Aussagen beweisen:

Lemma 4.5. Sind $y_1, y_2 \in H^1((a, b))^N$ zwei Lösungen von (AWP) zu den Steuerungen $u_1, u_2 \in L^2((a, b))$, so gilt die Abschätzung

$$\|y_1 - y_2\|_{L^2((a, b))^N}^2 \leq \hat{C} \|u_1 - u_2\|_{L^2((a, b))}^2$$

mit

$$\hat{C} = (b - a)^2 L_1^2 \left(\frac{1}{2} (\cosh(2L_2(b - a)) - 1) + \sqrt{2 \cosh(2L_2(b - a)) - 2} + 1 \right).$$

Beweis. Mit Lemma 4.3 und den Argumenten des Beweises von Satz 4.4 ist zu erkennen, dass die Ungleichung

$$\begin{aligned} \|y_1 - y_2\|_{L^2((a, b))^N}^2 &\leq (b - a)^2 \|y_1' - y_2'\|_{L^2((a, b))^N}^2 \\ &\leq (b - a)^2 L_1^2 \int_a^b \left[L_2^2 \left(\int_a^b |u_1(\tau) - u_2(\tau)| e^{L_2(t-\tau)} d\tau \right)^2 \right. \\ &\quad \left. + 2L_2 \left(\int_a^b |u_1(\tau) - u_2(\tau)| e^{L_2(t-\tau)} d\tau \right) |u_1(t) - u_2(t)| \right. \\ &\quad \left. + |u_1(t) - u_2(t)|^2 \right] dt \end{aligned}$$

erfüllt ist. Wir schätzen die ersten beiden Summanden ab. Die Hölderungleichung liefert

$$\begin{aligned} \int_a^b \left(\int_a^b |u_1(\tau) - u_2(\tau)| e^{L_2(t-\tau)} d\tau \right)^2 dt &\leq \int_a^b \|u_1 - u_2\|_{L^2((a, b))}^2 \|e^{L_2(t-\cdot)}\|_{L^2((a, b))}^2 dt \\ &= \|u_1 - u_2\|_{L^2((a, b))}^2 \int_a^b \int_a^b e^{2L_2(t-\tau)} d\tau dt \\ &= \frac{\|u_1 - u_2\|_{L^2((a, b))}^2}{2L_2} \int_a^b e^{2L_2(t-a)} - e^{2L_2(t-b)} dt \\ &= \frac{\|u_1 - u_2\|_{L^2((a, b))}^2}{4L_2^2} (e^{2L_2(b-a)} + e^{-2L_2(b-a)} - 2) \end{aligned}$$

und

$$\begin{aligned}
& \int_a^b \left(\int_a^b |u_1(\tau) - u_2(\tau)| e^{L_2(t-\tau)} d\tau \right) |u_1(t) - u_2(t)| dt \\
& \leq \|u_1 - u_2\|_{L^2((a,b))} \int_a^b \|e^{L_2(t-\cdot)}\|_{L^2((a,b))} |u_1(t) - u_2(t)| dt \\
& = \frac{\|u_1 - u_2\|_{L^2((a,b))}}{\sqrt{2L_2}} \int_a^b \sqrt{e^{2L_2(t-a)} - e^{2L_2(t-b)}} |u_1(t) - u_2(t)| dt \\
& \leq \frac{\|u_1 - u_2\|_{L^2((a,b))}^2}{\sqrt{2L_2}} \left\| \sqrt{e^{2L_2(\cdot-a)} - e^{2L_2(\cdot-b)}} \right\|_{L^2((a,b))} \\
& = \frac{\|u_1 - u_2\|_{L^2((a,b))}^2}{2L_2} \sqrt{e^{2L_2(b-a)} + e^{-2L_2(b-a)} - 2}.
\end{aligned}$$

Ein Zusammenfügen der Ausdrücke resultiert in der Abschätzung

$$\begin{aligned}
\|y_1 - y_2\|_{L^2((a,b))^N}^2 & \leq (b-a)^2 L_1^2 \left(\frac{1}{4} \left(e^{2L_2(b-a)} + e^{-2L_2(b-a)} - 2 \right) \right. \\
& \quad \left. + \sqrt{e^{2L_2(b-a)} + e^{-2L_2(b-a)} - 2} + 1 \right) \|u_1 - u_2\|_{L^2((a,b))}^2 \\
& = (b-a)^2 L_1^2 \left(\frac{1}{2} (\cosh(2L_2(b-a)) - 1) \right. \\
& \quad \left. + \sqrt{2 \cosh(2L_2(b-a)) - 2} + 1 \right) \|u_1 - u_2\|_{L^2((a,b))}^2.
\end{aligned}$$

□

Lemma 4.6. Sei $y \in H^1((a,b))^N$ die Lösung von (AWP) bezüglich der Steuerung $u \in U_{ad}$. Dann gibt es eine von u unabhängige Konstante $\tilde{C} > 0$, sodass

$$\|y\|_{L^2((a,b))^N} \leq \tilde{C}$$

erfüllt ist.

Beweis. Sei y_{ν_1} die Lösung von (AWP) zur Steuerung $\tilde{u} := \nu_1$. Mit Lemma 4.5 erhalten wir die Abschätzung

$$\begin{aligned}
\|y - y_{\nu_1}\|_{L^2((a,b))^N}^2 & \leq \hat{C} \|u - \nu_1\|_{L^2((a,b))}^2 \leq \hat{C} (\|u\|_{L^2((a,b))} + \|\nu_1\|_{L^2((a,b))})^2 \\
& = \hat{C} (\|u\|_{L^2((a,b))} + \sqrt{b-a} |\nu_1|)^2.
\end{aligned}$$

Wegen

$$\|u\|_{L^2((a,b))}^2 \leq \int_a^b \max(|\nu_1|, |\nu_d|)^2 dt = (b-a) \max(|\nu_1|, |\nu_d|)^2$$

und mit der umgekehrten Dreiecksungleichung erhalten wir schließlich

$$\begin{aligned} \|y\|_{L^2((a,b))^N} - \|y_{\nu_1}\|_{L^2((a,b))^N} &\leq \|y - y_{\nu_1}\|_{L^2((a,b))^N} \\ &\leq \sqrt{\hat{C}(b-a)}(\max(|\nu_1|, |\nu_d|) + |\nu_1|) \\ \Leftrightarrow \|y\|_{L^2((a,b))^N} &\leq \|y_{\nu_1}\|_{L^2((a,b))^N} + \sqrt{\hat{C}(b-a)}(\max(|\nu_1|, |\nu_d|) + |\nu_1|) =: \tilde{C}. \end{aligned}$$

□

4.2 Stetigkeitseigenschaften des Zielfunktional

Mit der schwach-* Folgenstetigkeit von S können nun Stetigkeitseigenschaften des Zielfunktional gefolgert werden. Da die Summe schwach-* folgenunterhalbstetiger Funktionen ebenso schwach-* folgenunterhalbstetig ist, genügt es, diese Eigenschaft für die einzelnen Summanden des Zielfunktional zu zeigen.

Sei zunächst $\{u_k\}_{k \in \mathbb{N}} \in \text{BV}((a,b))$ eine Folge mit $u_k \xrightarrow{*} u$. Wir beginnen mit dem Term

$$\begin{aligned} u &\mapsto \frac{1}{2} \int_a^b (S(u)(t) - r(t))^\top Q(t) (S(u)(t) - r(t)) dt \\ &= \frac{1}{2} \left(\int_a^b S(u)(t)^\top Q(t) S(u)(t) dt - 2 \int_a^b r(t)^\top Q(t) S(u)(t) dt \right. \\ &\quad \left. + \int_a^b r(t)^\top Q(t) r(t) dt \right). \end{aligned}$$

Die ersten zwei Summanden werden im Folgenden einzeln betrachtet, der letzte Summand spielt als Konstante keine Rolle für die Stetigkeitseigenschaften der obigen Funktion.

Zur Untersuchung des ersten Summanden definieren wir die Bilinearform

$$a : L^\infty((a,b))^N \times L^\infty((a,b))^N \rightarrow \mathbb{R}, \quad a(x,y) := \int_a^b x(t)^\top Q(t) y(t) dt.$$

Mithilfe der Hölderungleichung gelingt die Abschätzung

$$\begin{aligned} \int_a^b x(t)^\top Q(t) y(t) dt &= \int_a^b \langle x(t), Q(t) y(t) \rangle_2 dt \leq \int_a^b \|x(t)\|_2 \|Q(t) y(t)\|_2 dt \\ &\leq \int_a^b \|x(t)\|_2 \|Q(t)\|_2 \|y(t)\|_2 dt \\ &\leq \| \|x(\cdot)\|_2 \|_{L^\infty((a,b))} \| \|y(\cdot)\|_2 \|_{L^\infty((a,b))} \| \|Q(\cdot)\|_2 \|_{L^1((a,b))} \\ &\leq \|x\|_{L^\infty((a,b))^N} \|y\|_{L^\infty((a,b))^N} \| \|Q(\cdot)\|_2 \|_{L^1((a,b))}, \end{aligned}$$

wobei die letzte Umformung mit

$$\begin{aligned} \| \|x(\cdot)\|_2 \|_{L^\infty((a,b))} &= \operatorname{ess\,sup}_{t \in (a,b)} \sqrt{\sum_{i=1}^N x_i(t)^2} = \sqrt{\operatorname{ess\,sup}_{t \in (a,b)} \sum_{i=1}^N x_i(t)^2} \leq \sqrt{\sum_{i=1}^N \operatorname{ess\,sup}_{t \in (a,b)} x_i(t)^2} \\ &= \sqrt{\sum_{i=1}^N \left(\operatorname{ess\,sup}_{t \in (a,b)} |x_i(t)| \right)^2} = \sqrt{\sum_{i=1}^N \|x_i\|_{L^\infty((a,b))}^2} = \|x\|_{L^\infty((a,b))}^N \end{aligned} \quad (4.6)$$

begründet werden kann. Da die Ungleichung

$$\|Q(t)\|_2 \leq \|Q(t)\|_F = \sqrt{\sum_{i=1}^N \sum_{j=1}^N Q_{ij}(t)^2} \leq \sqrt{\left(\sum_{i=1}^N \sum_{j=1}^N |Q_{ij}(t)| \right)^2} = \sum_{i=1}^N \sum_{j=1}^N |Q_{ij}(t)| \quad (4.7)$$

gilt, können wir wegen $Q \in L^\infty((a,b))^{N \times N}$

$$\int_a^b \sum_{i=1}^N \sum_{j=1}^N |Q_{ij}(t)| \, dt \leq \sum_{i=1}^N \sum_{j=1}^N \int_a^b \|Q_{ij}\|_{L^\infty((a,b))} \, dt = (b-a) \sum_{i=1}^N \sum_{j=1}^N \|Q_{ij}\|_{L^\infty((a,b))}$$

schließen, somit ist $\|Q(\cdot)\|_2 \in L^1((a,b))$ und demnach a beschränkt und stetig.

Wegen $H^1((a,b))^N \hookrightarrow L^\infty((a,b))^N$ wissen wir, dass $S(u_k), S(u) \in L^\infty((a,b))^N$ gilt. Außerdem gibt es eine Konstante $C > 0$, sodass die Abschätzung $\|S(u_k) - S(u)\|_{L^\infty((a,b))^N} \leq C \|S(u_k) - S(u)\|_{H^1((a,b))^N}$ erfüllt ist. Da aus Satz 4.4 $\|S(u_k) - S(u)\|_{H^1((a,b))^N} \rightarrow 0$ folgt, muss $S(u_k)$ auch bezüglich der L^∞ -Norm gegen $S(u)$ konvergieren. Mit der Stetigkeit von a bezüglich beider Argumente schlussfolgern wir schließlich

$$a(S(u_k), S(u_k)) \rightarrow a(S(u), S(u)),$$

also ist der erste Summand schwach-* folgenstetig.

Da die Abbildung

$$v \mapsto \int_a^b \nu(t)^\top v(t) \, dt \quad (4.8)$$

für $v, \nu \in L^2((a,b))^N$ ein Element von $(L^2((a,b))^N)^*$ ist, sieht man mit der Setzung $\nu(t) = r(t)^\top Q(t)$ schnell, dass der zweite Summand wie in (4.8) geschrieben werden kann und somit stetig ist. Aus $S(u_k) \rightarrow S(u)$ in $L^2((a,b))^N$ folgt nun die schwach-* Folgenstetigkeit des Summanden.

Nun betrachten wir die Funktion $u \mapsto |Du|((a,b))$. Nach Satz 2.14, b) folgt

$$\int_a^b \varphi \, dDu = \lim_{k \rightarrow \infty} \int_a^b \varphi \, dDu_k = \liminf_{k \rightarrow \infty} \int_a^b \varphi \, dDu_k \quad \forall \varphi \in C_0((a,b)).$$

Mit der Setzung $C_{0,1} := \{\varphi \in C_0((a,b)) : \|\varphi\|_{L^\infty((a,b))} \leq 1\}$ können wir also

$$\begin{aligned} |Du|((a,b)) &= \sup_{\varphi \in C_{0,1}} \int_a^b \varphi \, dDu = \sup_{\varphi \in C_{0,1}} \liminf_{k \rightarrow \infty} \int_a^b \varphi \, dDu_k \\ &\leq \sup_{\varphi \in C_{0,1}} \liminf_{k \rightarrow \infty} \|\varphi\|_{L^\infty((a,b))} |Du_k|((a,b)) \\ &= \liminf_{k \rightarrow \infty} |Du_k|((a,b)) \end{aligned}$$

schreiben, womit die schwach-* Folgenunterhalbstetigkeit gezeigt ist.

Schließlich zeigen wir die Eigenschaft für die Funktionen Ψ_κ , $\kappa \in \{1, \dots, 4\}$. Für Ψ_1 ist nichts zu zeigen, für Ψ_4 folgt die schwach-* Folgenstetigkeit mit Anwendung der umgekehrten Dreiecksungleichung direkt aus Satz 2.14:

$$\left| \|u_k\|_{L^1((a,b))} - \|u\|_{L^1((a,b))} \right| \leq \|u_k - u\|_{L^1((a,b))} \rightarrow 0.$$

Um die schwach-* Folgenstetigkeit für Ψ_2 herzuleiten, benutzen wir die Dreiecksungleichung für die Abschätzung

$$\begin{aligned} \left| \left(\int_a^b u_k(t) \, dt \right)^2 - \left(\int_a^b u(t) \, dt \right)^2 \right| &= \left| \int_a^b u_k(t) \, dt \left(\int_a^b u_k(t) \, dt - \int_a^b u(t) \, dt \right) \right. \\ &\quad \left. - \int_a^b u(t) \, dt \left(\int_a^b u(t) \, dt - \int_a^b u_k(t) \, dt \right) \right| \\ &\leq \int_a^b |u_k(t)| \, dt \int_a^b |u_k(t) - u(t)| \, dt \\ &\quad + \int_a^b |u(t)| \, dt \int_a^b |u(t) - u_k(t)| \, dt \\ &= (\|u_k\|_{L^1((a,b))} + \|u\|_{L^1((a,b))}) \|u_k - u\|_{L^1((a,b))}. \end{aligned}$$

Aus Satz 2.14 folgt nun $u_k \rightarrow u$ in $L^1((a,b))$, also $\|u_k - u\|_{L^1((a,b))} \rightarrow 0$. Wegen der L^1 -Konvergenz muss $\{u_k\}_{k \in \mathbb{N}}$ in $L^1((a,b))$ beschränkt sein, womit es einen Faktor $C > 0$ gibt, sodass $\|u_k\|_{L^1((a,b))} + \|u\|_{L^1((a,b))} \leq C$ gilt. Es folgt

$$\left| \left(\int_a^b u_k(t) \, dt \right)^2 - \left(\int_a^b u(t) \, dt \right)^2 \right| \leq C \|u_k - u\|_{L^1((a,b))} \rightarrow 0$$

und damit die Aussage.

Im Falle von Ψ_3 muss man anders vorgehen, da die Funktion $u \mapsto |u(a)|$ im Allgemeinen nicht schwach-* folgenunterhalbstetig ist. Man betrachte dazu das folgende Beispiel:

Beispiel 4.7. Es sei $\{u_k\}_{k \in \mathbb{N}} \subset \text{BV}((0, 1))$ eine Funktionenfolge, welche durch

$$u_k(t) := \begin{cases} 1, & \text{wenn } t \geq 1/k \\ 0, & \text{wenn } t < 1/k \end{cases}$$

definiert ist. Wegen

$$\int_0^1 |u_k(t) - 1| dt = \int_0^{1/k} 1 dt = 1/k \rightarrow 0$$

gilt $u_k \rightarrow 1$ im L^1 , demnach ist $\{u_k\}_{k \in \mathbb{N}}$ auch beschränkt in $L^1((0, 1))$. Außerdem ist u_k für jedes $k \in \mathbb{N}$ konstant bis auf einen Sprung der Höhe 1, es gilt also $|Du_k|((0, 1)) = 1 \forall k \in \mathbb{N}$. Somit ist $\{u_k\}_{k \in \mathbb{N}}$ beschränkt in $\text{BV}((0, 1))$, mit Satz 2.14 folgt also $u_k \xrightarrow{*} u = 1$. Allerdings ist mit der Rechnung

$$|u(a)| = 1 > \liminf_{k \rightarrow \infty} 0 = \liminf_{k \rightarrow \infty} |u_k(a)|$$

Ψ_3 nicht schwach-* folgenunterhalbstetig.

Um dieses Problem zu umgehen, zeigen wir die Eigenschaft für die Funktion $u \mapsto |Du|((a, b)) + |u(a)|$. Dazu definieren wir die Funktionenfolge $\hat{u}_k : (a - 1, b) \rightarrow \mathbb{R}$ gemäß

$$\hat{u}_k(t) := \begin{cases} u_k(t), & \text{wenn } t \in (a, b) \\ 0, & \text{sonst.} \end{cases}$$

Wir wählen für jedes $k \in \mathbb{N}$ die gute Fortsetzung eines guten Repräsentanten von \hat{u}_k durch die Funktion \tilde{u}_k so, dass $\tilde{u}_k(a) = 0$ erfüllt ist. Dann gilt

$$\begin{aligned} |D\hat{u}_k|((a - 1, b)) &= \text{Var}(\tilde{u}_k, [a - 1, b]) = \text{Var}(\tilde{u}_k, [a - 1, a]) + \text{Var}(\tilde{u}_k, [a, b]) \\ &= \text{Var}(\tilde{u}_k, [a, b]) \end{aligned}$$

wegen $\tilde{u}_k(t) = 0$ für $t \in [a - 1, a]$. Wir betrachten nun für ein beliebiges k den Term $\text{Var}(\tilde{u}_k, [a, b])$ und zeigen die Gültigkeit von $\text{Var}(\tilde{u}_k, [a, b]) = |u_k(a)| + |Du_k|((a, b))$, wobei $u_k(a)$ durch $\lim_{t \downarrow a} u_k(t)$ gegeben ist. Sei dazu v_k die durch

$$v_k(t) := \begin{cases} \tilde{u}_k(t) & \text{für } t \in (a, b) \\ u_k(a) & \text{für } t = a \end{cases}$$

auf $[a, b]$ definierte Funktion. Da \tilde{u}_k eine gute Fortsetzung eines guten Repräsentanten von \hat{u}_k ist, ist nach Definition v_k eine gute Fortsetzung eines guten Repräsentanten

von u_k , es gilt also $|Du_k|((a, b)) = \text{Var}(v_k, (a, b)) = \text{Var}(v_k, [a, b])$. Wir definieren die Folge $\{M_i^l\}_{i=0}^{m_l}$, für die

$$\lim_{l \rightarrow \infty} \sum_{i=1}^{m_l} |\tilde{u}_k(M_i^l) - \tilde{u}_k(M_{i-1}^l)| = \text{Var}(\tilde{u}_k, [a, b])$$

erfüllt ist. Zunächst schätzen wir in die eine Richtung ab und erhalten

$$\begin{aligned} & \lim_{l \rightarrow \infty} \sum_{i=1}^{m_l} |\tilde{u}_k(M_i^l) - \tilde{u}_k(M_{i-1}^l)| = \lim_{l \rightarrow \infty} |\tilde{u}_k(M_1^l) - \tilde{u}_k(a)| + \sum_{i=2}^{m_l} |\tilde{u}_k(M_i^l) - \tilde{u}_k(M_{i-1}^l)| \\ &= \lim_{l \rightarrow \infty} |\tilde{u}_k(M_1^l)| + \sum_{i=2}^{m_l} |\tilde{u}_k(M_i^l) - \tilde{u}_k(M_{i-1}^l)| \leq \lim_{l \rightarrow \infty} \left| \tilde{u}_k \left(a + \frac{M_1^l - a}{l} \right) \right| \\ & \quad + \left| \tilde{u}_k(M_1^l) - \tilde{u}_k \left(a + \frac{M_1^l - a}{l} \right) \right| + \sum_{i=2}^{m_l-1} |\tilde{u}_k(M_i^l) - \tilde{u}_k(M_{i-1}^l)| \\ & \quad + \left| \tilde{u}_k \left(b - \frac{b - M_{m_l-1}^l}{l} \right) - \tilde{u}_k(M_{m_l-1}^l) \right| + \left| \tilde{u}_k(b) - \tilde{u}_k \left(b - \frac{b - M_{m_l-1}^l}{l} \right) \right| \\ & \leq \lim_{l \rightarrow \infty} \left| \tilde{u}_k \left(a + \frac{M_1^l - a}{l} \right) \right| + \left| \tilde{u}_k(b) - \tilde{u}_k \left(b - \frac{b - M_{m_l-1}^l}{l} \right) \right| + \text{Var}(v_k, (a, b)) \\ & = |u_k(a)| + \text{Var}(v_k, (a, b)), \end{aligned}$$

wegen $\tilde{u}_k = v_k$ in $(a, b]$ und da \tilde{u}_k auf (a, b) ein guter Repräsentant von u_k ist.

Für die andere Richtung definieren wir die Folge von Partitionen $\{P_i^l\}_{i=0}^{p_l} \in 2^{(a, b)}$, für die

$$\lim_{l \rightarrow \infty} \sum_{i=1}^{p_l} |v_k(P_i^l) - v_k(P_{i-1}^l)| = \text{Var}(v_k, (a, b))$$

erfüllt ist. Da wir $|u_k(a)|$ auch als $\lim_{l \rightarrow \infty} \left| \tilde{u}_k \left(a + \frac{P_0^l - a}{l} \right) \right|$ darstellen können, kann die Abschätzung

$$\begin{aligned} |u_k(a)| + \text{Var}(v_k, (a, b)) &= \lim_{l \rightarrow \infty} \left| \tilde{u}_k \left(a + \frac{P_0^l - a}{l} \right) \right| + \sum_{i=1}^{p_l} |v_k(P_i^l) - v_k(P_{i-1}^l)| \\ &= \lim_{l \rightarrow \infty} \left| \tilde{u}_k \left(a + \frac{P_0^l - a}{l} \right) - \tilde{u}_k(a) \right| + \sum_{i=1}^{p_l} |\tilde{u}_k(P_i^l) - \tilde{u}_k(P_{i-1}^l)| \\ &\leq \lim_{l \rightarrow \infty} \left| \tilde{u}_k \left(a + \frac{P_0^l - a}{l} \right) - \tilde{u}_k(a) \right| + \sum_{i=1}^{p_l} |\tilde{u}_k(P_i^l) - \tilde{u}_k(P_{i-1}^l)| \\ & \quad + |\tilde{u}_k(b) - \tilde{u}_k(P_{p_l}^l)| \leq \text{Var}(\tilde{u}_k, [a, b]) \end{aligned}$$

durchgeführt werden. Es folgt also

$$|D\hat{u}_k|((a-1, b)) = |u_k(a)| + |Du_k|((a, b)) < \infty \quad \forall k \in \mathbb{N},$$

da $u_k \in \text{BV}((a, b)) \quad \forall k \in \mathbb{N}$ gilt. Damit ist $\hat{u}_k \in \text{BV}((a-1, b))$ und da nach vorigem Vorgehen bekannt ist, dass auch die Abbildung $u \mapsto |Du|((a-1, b))$ schwach-* folgenunterhalbstetig ist, muss nur noch

$$\hat{u}_k \xrightarrow{*} \hat{u} := \begin{cases} u(t), & \text{wenn } t \in (a, b) \\ 0, & \text{sonst} \end{cases}$$

gezeigt werden, um die schwach-* Folgenunterhalbstetigkeit von $u \mapsto |Du|((a, b)) + |u(a)|$ zu erhalten. Nach Satz 2.14 muss also nachgewiesen werden, dass $\hat{u}_k \rightarrow \hat{u}$ in $L^1((a-1, b))$ sowie $\|\hat{u}_k\|_{\text{BV}((a-1, b))} \leq C \quad \forall k \in \mathbb{N}$ erfüllt sind.

Ersteres folgt direkt aus $u_k \rightarrow u$ in $L^1((a, b))$, weil $\hat{u}_k(t) = \hat{u}(t)$ für $t \in (a-1, a]$ gilt. Aus dieser Konvergenz folgt auch die Beschränktheit der Folge im L^1 . Die Beschränktheit des Ausdrucks $|D\hat{u}_k|((a-1, b)) = |u_k(a)| + |Du_k|((a, b))$ folgt daraus, dass einerseits $|Du_k|((a, b))$ nach Satz 2.14, a) wegen $u_k \xrightarrow{*} u$ beschränkt sein muss, andererseits kann mithilfe von Satz 2.17 die Beschränktheit von $|u_k(a)|$ bestätigt werden: Mit der Darstellung

$$u_k^l(t) = u_k(a) + Du_k((a, t)) \quad \text{bzw.} \quad u_k(a) = u_k^l(t) - Du_k((a, t))$$

wobei u_k^l der linksseitig stetige Repräsentant von u ist, folgt die Abschätzung

$$|u_k(a)| \leq |u_k^l(t)| + |Du_k((a, t))| \leq |u_k^l(t)| + |Du_k|((a, b)).$$

Aus dieser Ungleichung kann die Rechnung

$$\begin{aligned} (b-a)|u_k(a)| &= \int_a^b |u_k(a)| \, dt \leq \|u_k\|_{L^1((a, b))} + (b-a)|Du_k|((a, b)) \\ \Leftrightarrow |u_k(a)| &\leq \frac{1}{b-a} \|u_k\|_{L^1((a, b))} + |Du_k|((a, b)) \end{aligned}$$

aufgestellt werden, die Beschränktheit der Terme auf der rechten Seite wurde schon im Vorigen erläutert.

Nun muss noch gezeigt werden, dass aus der schwach-* Folgenunterhalbstetigkeit von $u \mapsto |Du|((a, b)) + |u(a)|$ auch jene von $u \mapsto \beta_1 |Du|((a, b)) + \beta_2 |u(a)|$ folgt. Mit der aus (V1) gewonnenen Bedingung $\beta_1 \geq \beta_2$ ist dies aufgrund der schwach-*

Folgenunterhalbstetigkeit von $|Du|((a, b))$ und

$$u \mapsto (\beta_1 - \beta_2)|Du|((a, b)) + \beta_2|Du|((a, b)) + \beta_2|u(a)| = \beta_1|Du|((a, b)) + \beta_2|u(a)|$$

der Fall. Damit ist auch die schwach-* Folgenunterhalbstetigkeit des Zielfunktional
im Falle von $\kappa = 3$ gesichert.

Kapitel 5

Differenzierbarkeitseigenschaften des Zielfunktional

In den nachfolgenden Kapiteln soll ein Lösungsalgorithmus in Form eines Splitting-Verfahrens hergeleitet werden, der Probleme der Form

$$\begin{aligned} \text{Minimiere } J_\kappa(u) &= f_\kappa(u) + g_\kappa(u) \\ \text{sodass } u &\in U_{ad} \end{aligned} \tag{P-Split}$$

behandelt, wobei $f_\kappa : L^2((a, b)) \rightarrow \mathbb{R}$ Gateaux-differenzierbar ist. Unter der Gateaux-Differenzierbarkeit eines Operators $f : X \rightarrow Y$ in $u \in X$ verstehen wir die Existenz des Grenzwerts

$$\lim_{\tau \downarrow 0} \frac{f(u + \tau h) - f(u)}{\tau} =: \nabla_u f(u)h$$

für alle $h \in X$. Außerdem sei der als Gateaux-Differential bezeichnete Operator $\nabla_u f(u) : X \rightarrow Y$ linear und stetig.

In diesem Kapitel werden dazu die Summanden des Zielfunktional auf Gateaux-Differenzierbarkeit untersucht und wenn möglich die zugehörigen Gateaux-Ableitungen berechnet. Wir gehen dazu ähnlich wie in Kapitel 4 vor und untersuchen zunächst den Lösungsoperator S .

5.1 Differenzierbarkeit des Lösungsoperators

In diesem Abschnitt schreiben wir $S(u)(t) = y(t, u)$, um zu verdeutlichen, dass y als Lösung von (AWP) von der Wahl von u abhängt. Wir nehmen zunächst die Gateaux-Differenzierbarkeit von S an und leiten damit eine Formel zur Berechnung des Gateaux-Differentials von S her. Anschließend berechnen wir mithilfe der adjungierten Gleichung die Ableitungen der Terme, in denen S enthalten ist.

Wir definieren zunächst die Funktion $z(t) := [S'(u)h](t)$ für eine Richtung $h \in L^2((a, b))$ und leiten diese nach t ab. Angenommen, die Ableitungen $\frac{d}{dt}(\nabla_u S(u)h)(t)$ und $[\nabla_u(\frac{d}{dt}S(u)(t))h](t)$ existieren und sind stetig. Dann gilt nach dem Satz von Schwarz

$$\frac{d}{dt}(\nabla_u S(u)h)(t) = \left[\nabla_u \left(\frac{d}{dt}S(u)(t) \right) h \right] (t).$$

Mit der Kettenregel folgt

$$\begin{aligned} \frac{d}{dt}z(t) &= \frac{d}{dt}[S'(u)h](t) = \frac{d}{dt}[\nabla_u y(t, u)h](t) = \nabla_u \left[\frac{d}{dt}y(t, u) \right] h(t) \\ &= \nabla_u F(t, u(t), y(t, u))h(t) = \frac{d}{d\tau}|_{\tau=0} F(t, u(t) + \tau h(t), y(t, u + \tau h)) \\ &= F_u(t, u(t), y(t, u))h(t) + F_y(t, u(t), y(t, u)) \frac{d}{d\tau}|_{\tau=0} y(t, u + \tau h) \\ &= F_u(t, u(t), y(t, u))h(t) + F_y(t, u(t), y(t, u))[S'(u)h](t) \\ &= F_u(t, u(t), y(t, u))h(t) + F_y(t, u(t), y(t, u))z(t), \end{aligned}$$

wobei F_u als die Vektorfunktion zu verstehen ist, die man erhält, wenn man F komponentenweise nach dem zweiten Argument ableitet und F_y als die Jacobimatrix, die durch Differentiation von F nach dem dritten Argument entsteht (vgl. Voraussetzung (V2)).

Wegen $z(a) = [\nabla_u y(a, u)h](t) = [(\nabla_u y_0)h](t) = 0$ erhalten wir nun das neue Anfangswertproblem

$$\begin{aligned} z'(t) &= F_u(t, u(t), y(t, u))h(t) + F_y(t, u(t), y(t, u))z(t) \quad \text{f.ü. auf } (a, b) \\ z(a) &= 0, \end{aligned} \quad (S'\text{-AWP})$$

dessen Lösung z als Gateaux-Differential von S in Richtung h interpretiert werden kann. Die Existenz der eindeutigen Lösung z kann wie in Kapitel 4 mit [11, Kap. 5, Satz 1.3] begründet werden, die dazu nötige Lipschitzstetigkeit der rechten Seite $\hat{F}(t, u(t), z(t)) := F_u(t, u(t), y(t, u))h(t) + F_y(t, u(t), y(t, u))z(t)$ bezüglich des dritten

Arguments erhält man mit der Hölderungleichung aus der Rechnung

$$\begin{aligned}
\|\hat{F}(\cdot, u(\cdot), z_1(\cdot)) - \hat{F}(\cdot, u(\cdot), z_2(\cdot))\|_{L^2((a,b))^N}^2 &= \|F_y(\cdot, u(\cdot), y(\cdot, u))(z_1 - z_2)\|_{L^2((a,b))^N}^2 \\
&= \int_a^b \|F_y(t, u(t), y(t, u))(z_1(t) - z_2(t))\|_2^2 dt \\
&\leq \int_a^b \|F_y(t, u(t), y(t, u))\|_2^2 \|z_1(t) - z_2(t)\|_2^2 dt \\
&\leq \|\|F_y(\cdot, u(\cdot), y(\cdot, u))\|_2^2\|_{L^\infty((a,b))} \int_a^b \|z_1(t) - z_2(t)\|_2^2 dt \\
&= \|\|F_y(\cdot, u(\cdot), y(\cdot, u))\|_2^2\|_{L^\infty((a,b))} \|z_1 - z_2\|_{L^2((a,b))^N}^2,
\end{aligned}$$

wobei wegen

$$\begin{aligned}
\|F_y(t, u(t), y(t, u))\|_2 &= \max_{\|v\|_2=1} \|F_y(t, u(t), y(t, u))v\|_2 \\
&= \max_{\|v\|_2=1} \lim_{\tau \downarrow 0} \left\| \frac{F(t, u(t), y(t, u) + \tau v) - F(t, u(t), y(t, u))}{\tau} \right\|_2 \\
&\leq \max_{\|v\|_2=1} \lim_{\tau \downarrow 0} \frac{L_2}{\tau} \|y(t, u) + \tau v - y(t, u)\|_2 = L_2
\end{aligned} \tag{5.1}$$

der Term $\|\|F_y(\cdot, u(\cdot), y(\cdot, u))\|_2^2\|_{L^\infty((a,b))}$ kleiner als unendlich ist.

Bei der Herleitung von (S' -AWP) haben wir einige noch nicht bewiesene Annahmen genutzt. Mit dem nächsten Satz wird gezeigt, dass z tatsächlich als Gateaux-Differential von S aufgefasst werden kann.

Satz 5.1. Der Lösungsoperator S ist Gateaux-differenzierbar und für alle Richtungen $h \in L^2((a, b))$ gilt $S'(u)h = z$, wobei z die Lösung von (S' -AWP) ist.

Beweis. Sei $y_\tau = S(u + \tau h)$. Wir zeigen zunächst, dass

$$\zeta_\tau(t) := \frac{y_\tau(t) - y(t)}{\tau} - z(t) \rightarrow 0$$

für $\tau \rightarrow 0$ gilt. Dazu leiten wir ζ_τ ab und erhalten die Gleichung

$$\begin{aligned}
\zeta'_\tau(t) &= \frac{1}{\tau}(y'_\tau(t) - y'(t)) - z'(t) \\
&= \frac{1}{\tau}(F(t, u(t) + \tau h(t), y_\tau(t)) - F(t, u(t), y(t)) - \tau F_u(t, u(t), y(t))h(t) \\
&\quad - \tau F_y(t, u(t), y(t))z(t)) \pm F_y(t, u(t), y(t)) \left(\frac{y_\tau(t) - y(t)}{\tau} - z(t) \right) \\
&= F_y(t, u(t), y(t))\zeta_\tau(t) + \frac{1}{\tau}(F(t, u(t) + \tau h(t), y_\tau(t)) - F(t, u(t), y(t)) \\
&\quad - \tau F_u(t, u(t), y(t))h(t) - F_y(t, u(t), y(t))(y_\tau(t) - y(t))) \\
&=: F_y(t, u(t), y(t))\zeta_\tau(t) + \phi_\tau(t) =: \tilde{F}(t, \phi_\tau(t), \zeta_\tau(t)),
\end{aligned}$$

zusammen mit $\zeta_\tau(a) = \frac{y_0 - y_0}{\tau} - z(a) = 0$ entsteht ein Anfangswertproblem. Dieses besitzt aufgrund der Linearität bezüglich ζ_τ mit analoger Begründung zu (S' -AWP) eine eindeutige Lösung. Sind $\zeta_1, \zeta_2 \in H^1((a, b))$ die Lösungen von

$$\zeta'_1(t) = \tilde{F}(t, \varphi_1(t), \zeta_1(t)) \quad \text{bzw.} \quad \zeta'_2(t) = \tilde{F}(t, \varphi_2(t), \zeta_2(t)), \quad \zeta_1(a) = \zeta_2(a) = 0,$$

mit zwei Funktionen $\varphi_1, \varphi_2 \in L^2((a, b))^N$, so gilt mit einer Rechnung analog zu (4.5)

$$\|\zeta_1 - \zeta_2\|_{H^1((a, b))^N}^2 \leq ((b - a)^2 + 1) \|\zeta'_1 - \zeta'_2\|_{L^2((a, b))^N}^2. \quad (5.2)$$

Wir schätzen nun den Term $\|\zeta'_1 - \zeta'_2\|_{L^2((a, b))^N}$ ähnlich wie in Abschnitt 4.1 ab und beobachten dazu zunächst, dass $\|F_u(t, u(t), y(t))\|_2 \leq C_1$ für eine Konstante $C_1 > 0$ wegen

$$\|F_u(t, u(t), y(t))\|_2 \leq \|F_u(\cdot, u(\cdot), y(\cdot))\|_{L^\infty((a, b))^N}$$

und $\|F_y(t, u(t), y(t))\|_2 \leq L_2$ (vgl. (5.1)) gilt. Nun erhalten wir die Abschätzung

$$\begin{aligned}
\|\zeta'_1(t) - \zeta'_2(t)\|_2 &= \|F_y(t, u(t), y(t))(\zeta_1(t) - \zeta_2(t)) + \varphi_1(t) - \varphi_2(t)\|_2 \\
&\leq \|F_y(t, u(t), y(t))\|_2 \|\zeta_1(t) - \zeta_2(t)\|_2 + \|\varphi_1(t) - \varphi_2(t)\|_2 \\
&\leq L_2 \|\zeta_1(t) - \zeta_2(t)\|_2 + \|\varphi_1(t) - \varphi_2(t)\|_2 \\
&\stackrel{(4.2)}{\leq} L_2 \int_a^t \|\zeta'_1(s) - \zeta'_2(s)\|_2 ds + \|\varphi_1(t) - \varphi_2(t)\|_2,
\end{aligned}$$

eine Anwendung der Grönwall-Ungleichung liefert also

$$\|\zeta'_1(t) - \zeta'_2(t)\|_2 \leq L_2 \int_a^t \|\varphi_1(s) - \varphi_2(s)\|_2 e^{L_2(t-s)} ds + \|\varphi_1(t) - \varphi_2(t)\|_2.$$

Wie in (4.4) erhalten wir nun die Abschätzung

$$\begin{aligned}
\|\zeta'_1 - \zeta'_2\|_{L^2((a,b))^N}^2 &= \int_a^b \|\zeta'_1(t) - \zeta'_2(t)\|_2^2 dt \\
&\leq \int_a^b \left[L_2^2 \left(\int_a^b \|\varphi_1(s) - \varphi_2(s)\|_2 e^{L_2(t-s)} ds \right)^2 \right. \\
&\quad + L_2 \int_a^b \|\varphi_1(s) - \varphi_2(s)\|_2 e^{L_2(t-s)} ds \|\varphi_1(t) - \varphi_2(t)\|_2 \\
&\quad \left. + \|\varphi_1(t) - \varphi_2(t)\|_2^2 \right] dt. \tag{5.3}
\end{aligned}$$

Nach analogem Vorgehen zum Beweis von Satz 4.4 folgen nun die Rechnungen

$$\int_a^b \left(\int_a^b \|\varphi_1(s) - \varphi_2(s)\|_2 e^{L_2(t-s)} ds \right)^2 dt \leq C_2 \|\|\varphi_1(\cdot) - \varphi_2(\cdot)\|_2\|_{L^1((a,b))}^2$$

und

$$\int_a^b \int_a^b \|\varphi_1(s) - \varphi_2(s)\|_2 e^{L_2(t-s)} ds \|\varphi_1(t) - \varphi_2(t)\|_2 dt \leq C_3 \|\|\varphi_1(\cdot) - \varphi_2(\cdot)\|_2\|_{L^1((a,b))}^2$$

mit

$$C_2 := \frac{e^{2L_2(b-a)} - 1}{2L_2} \quad \text{und} \quad C_3 := e^{L_2(b-a)},$$

der dritte Summand entspricht $\|\varphi_1 - \varphi_2\|_{L^2((a,b))^N}^2$. Aufgrund von

$$\begin{aligned}
\|\|\varphi_1(\cdot) - \varphi_2(\cdot)\|_2\|_{L^1((a,b))}^2 &\leq (b-a) \|\|\varphi_1(\cdot) - \varphi_2(\cdot)\|_2\|_{L^2((a,b))}^2 \\
&= (b-a) \|\varphi_1 - \varphi_2\|_{L^2((a,b))^N}^2
\end{aligned}$$

gewinnen wir schließlich die Ungleichung

$$\begin{aligned}
\|\zeta'_1 - \zeta'_2\|_{L^2((a,b))^N}^2 &\leq (b-a) \left(L_2^2 C_2 + L_2 C_3 + \frac{1}{b-a} \right) \|\varphi_1 - \varphi_2\|_{L^2((a,b))^N}^2 \\
&=: C_4 \|\varphi_1 - \varphi_2\|_{L^2((a,b))^N}^2.
\end{aligned}$$

Nun beobachten wir, dass die Lösung von

$$\zeta'_0(t) = \tilde{F}(t, 0, \zeta_0(t)) = F_y(t, u(t), y(t)) \zeta_0(t)$$

mit $\zeta_0(a) = 0$ durch $\zeta_0(t) = 0$ gegeben ist und wenden die gewonnene Ungleichung auf (5.2) an:

$$\|\zeta_\tau\|_{H^1((a,b))^N}^2 = \|\zeta_\tau - \zeta_0\|_{H^1((a,b))^N}^2 \leq ((b-a)^2 + 1) \|\zeta'_\tau - \zeta'_0\|_{L^2((a,b))^N}^2 \leq C_5 \|\phi_\tau\|_{L^2((a,b))^N}^2$$

mit $C_5 := C_4((b-a)^2 + 1)$.

Es bleibt, $\lim_{\tau \downarrow 0} \|\phi_\tau\|_{L^2((a,b))^N}^2 = 0$ zu zeigen. Dazu wollen wir den Satz der majorierten Konvergenz auf die Funktion $\|\phi_\tau(t)\|_2$ anwenden, um

$$\|\phi_\tau\|_{L^2((a,b))^N}^2 = \int_a^b \|\phi_\tau(t)\|_2^2 dt = \|\|\phi_\tau(\cdot)\|_2\|_{L^2((a,b))}^2 \rightarrow 0$$

zu erhalten. Wir müssen also die folgenden Aussagen beweisen:

- i) Es gibt eine Funktion $g \in L^2((a,b))$ mit $\|\phi_\tau(t)\|_2 \leq g(t)$ für fast alle $t \in (a,b)$.
- ii) $\|\phi_\tau(t)\|_2$ konvergiert punktweise fast überall gegen 0.

Dazu ist zunächst die Beobachtung hilfreich, dass analog zur Rechnung (4.2) die Abschätzung

$$\|y_\tau(t) - y(t)\|_2 \leq \int_a^t \|y'_\tau(s) - y'(s)\|_2 ds \leq \int_a^b \|y'_\tau(s) - y'(s)\|_2 ds$$

erfüllt ist, mit Anwendung von Lemma 4.3 auf den Term im Integral und mit der Hölderungleichung ergibt sich schließlich

$$\begin{aligned} \|y_\tau(t) - y(t)\|_2 &\leq \int_a^b \left(L_1 L_2 \tau \int_a^s |h(x)| e^{L_2(s-x)} dx + L_1 \tau |h(s)| \right) ds \\ &\leq L_1 L_2 \tau \|h\|_{L^1((a,b))} \int_a^b e^{L_2(s-a)} ds + L_1 \tau \|h\|_{L^1((a,b))} \\ &= \tau (L_1 e^{L_2(b-a)} - L_1 + L_1) \|h\|_{L^1((a,b))} = \tau L_1 e^{L_2(b-a)} \|h\|_{L^1((a,b))}. \end{aligned} \tag{5.4}$$

Wir beginnen nun mit dem Beweis von i). Bei Betrachtung der ersten beiden Terme in ϕ_τ erinnern wir uns an die Lipschitzstetigkeit von F und gewinnen die Abschätzung

$$\begin{aligned} &\frac{1}{\tau} \|F(t, u(t) + \tau h(t), y_\tau(t)) - F(t, u(t), y(t))\|_2 \\ &= \frac{1}{\tau} \|F(t, u(t) + \tau h(t), y_\tau(t)) \pm F(t, u(t) + \tau h(t), y(t)) - F(t, u(t), y(t))\|_2 \\ &\leq \frac{L_2}{\tau} \|y_\tau(t) - y(t)\|_2 + \frac{L_1}{\tau} |u(t) + \tau h(t) - u(t)| \\ &\stackrel{(5.4)}{\leq} L_1 L_2 e^{L_2(b-a)} \|h\|_{L^1((a,b))} + L_1 |h(t)|, \end{aligned}$$

wegen $h \in L^2((a,b))$ ist der entstehende Ausdruck ebenso eine Funktion aus $L^2((a,b))$.

Für den dritten Term gilt

$$\|F_u(t, u(t), y(t))h(t)\|_2 = \|F_u(t, u(t), y(t))\|_2 |h(t)| \leq C_1 |h(t)|.$$

Der vierte Term lässt sich durch

$$\begin{aligned} \left\| F_y(t, u(t), y(t)) \frac{y_\tau(t) - y(t)}{\tau} \right\|_2 &\leq \|F_y(t, u(t), y(t))\|_2 \frac{1}{\tau} \|y_\tau(t) - y(t)\|_2 \\ &\stackrel{(5.4)}{\leq} L_2 L_1 e^{L_2(b-a)} \|h\|_{L^1((a,b))} =: C_6. \end{aligned}$$

abschätzen. Zusammenfassend kann also die Funktion g aus i) mit der Funktionsvorschrift

$$g(t) := L_1 L_2 e^{L_2(b-a)} \|h\|_{L^1((a,b))} + C_6 + (L_1 + C_1) |h(t)|$$

als Majorante gewählt werden.

Um ii) zu zeigen, führen wir eine Taylorentwicklung 1. Grades für die Funktion $\hat{F}(v, w) := F(t, u(t) + v, y(t) + w)$ für ein festes $t \in (a, b)$ in der Entwicklungsstelle $(0, 0)$ durch. Wir finden also die Darstellung

$$T_{\hat{F}}^1((v, w), 0) = F(t, u(t), y(t)) + \begin{pmatrix} F_u(t, u(t), y(t)) & F_y(t, u(t), y(t)) \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix}$$

mit dem Restglied

$$R_{\hat{F}}^1((v, w), 0) = \mathfrak{o}(\|(v, w)\|_2) = \mathfrak{o}(|v| + \|w\|_2).$$

Nun wählen wir die Setzungen $v := \tau h(t)$, $w := y_\tau(t) - y(t)$ und erreichen die Darstellung

$$\begin{aligned} \hat{F}(v, w) &= F(t, u(t) + \tau h(t), y_\tau(t)) = F(t, u(t), y(t)) + \tau F_u(t, u(t), y(t))h(t) \\ &\quad + F_y(t, u(t), y(t))(y_\tau(t) - y(t)) + \mathfrak{o}(\tau |h(t)| + \|y_\tau(t) - y(t)\|_2), \end{aligned}$$

eingesetzt in ϕ_τ ergibt dies also für ein $t \in (a, b)$

$$\begin{aligned}
\phi_\tau(t) &= \frac{\mathfrak{o}(\tau|h(t)| + \|y_\tau(t) - y(t)\|_2)}{\tau} \\
&= \frac{\mathfrak{o}(\tau|h(t)| + \|y_\tau(t) - y(t)\|_2)}{\tau|h(t)| + \|y_\tau(t) - y(t)\|_2} \frac{\tau|h(t)| + \|y_\tau(t) - y(t)\|_2}{\tau} \\
&\leq \frac{\mathfrak{o}(\tau|h(t)| + \|y_\tau(t) - y(t)\|_2)}{\tau|h(t)| + \|y_\tau(t) - y(t)\|_2} (|h(t)| + L_1 e^{L_2(b-a)} \|h\|_{L^1((a,b))}) \\
&=: C_7 \frac{\mathfrak{o}(\tau|h(t)| + \|y_\tau(t) - y(t)\|_2)}{\tau|h(t)| + \|y_\tau(t) - y(t)\|_2} \rightarrow 0
\end{aligned}$$

für $\tau \rightarrow 0$, somit ist die Gültigkeit von ii) gezeigt.

Wir erhalten also die Gleichheit $S'(u)h = z$. Da das Anfangswertproblem (S' -AWP) linear bezüglich h ist, ist auch z linear bezüglich h , wir müssen also nur noch die Beschränktheit von z bezüglich h zeigen, um die Gateaux-Differenzierbarkeit zu erhalten. Erneut kann Rechnung (4.2) nachgeahmt werden, um die Gültigkeit von

$$\|z\|_{H^1((a,b))^N}^2 \leq ((b-a)^2 + 1) \|z'\|_{L^2((a,b))^N}^2$$

zu erhalten. Anschließend finden wir mit Lemma 4.2 die Abschätzung

$$\begin{aligned}
\|z'(t)\|_2 &\leq \|F_u(t, u(t), y(t))\|_2 |h(t)| + \|F_y(t, u(t), y(t))\|_2 \int_a^t \|z'(s)\|_2 ds \\
&\leq C_1 |h(t)| + L_2 \int_a^t \|z'(s)\|_2 ds,
\end{aligned}$$

die Grönwall-Ungleichung liefert uns

$$\|z'(t)\|_2 \leq C_1 |h(t)| + C_1 L_2 \int_a^t |h(s)| e^{L_2(t-s)} ds.$$

Erneutes Vorgehen analog zum Beweis von Satz 4.4 liefert die Abschätzung

$$\begin{aligned}
\|z'\|_{L^2((a,b))^N}^2 &\leq C_1^2 L_2^2 \frac{e^{2L_2(b-a)} - 1}{2L_2} \|h\|_{L^1((a,b))}^2 + 2C_1^2 L_2 e^{L_2(b-a)} \|h\|_{L^1((a,b))}^2 \\
&\quad + C_1^2 \|h\|_{L^2((a,b))}^2 \\
&\leq (b-a) \left(C_1^2 L_2 \frac{e^{2L_2(b-a)} - 1}{2} + 2C_1^2 L_2 e^{L_2(b-a)} + \frac{C_1^2}{b-a} \right) \|h\|_{L^2((a,b))}^2 \\
&=: C_8 \|h\|_{L^2((a,b))}^2.
\end{aligned}$$

Zusammenfassend ist wegen

$$\|z\|_{H^1((a,b))^N} \leq ((b-a)^2 + 1) \sqrt{C_8} \|h\|_{L^2((a,b))}$$

mit z das Gateaux-Differential von S an der Stelle u in Richtung h gefunden. \square

5.2 Differentiale der Terme im Zielfunktional

In diesem Abschnitt werden die Gateaux-Differentiale verschiedener Summanden des Zielfunktional berechnet bzw. charakterisiert. Für die Untersuchung des von S abhängigen Terms im Zielfunktional definieren wir ein weiteres Anfangswertproblem gemäß

$$\begin{aligned} -p'(t) &= F_y(t, u(t), y(t, u))^\top p(t) + Q(t)(y(t, u) - r(t)) \quad \text{f.ü. auf } (a, b) \\ p(b) &= 0. \end{aligned} \quad (\text{adj-AWP})$$

Die Existenz der eindeutigen Lösung p lässt sich analog zu der von (S' -AWP) zeigen. Wir werden sehen, dass p hilfreich bei der Berechnung des Gateaux-Differentials der Zielfunktion ist. Dazu benötigen wir das nächste Lemma.

Lemma 5.2. Sei $S'(u)^*: (H^1((a, b))^N)^* \rightarrow L^2((a, b))^*$ der adjungierte Operator zu $S'(u): L^2((a, b)) \rightarrow H^1((a, b))^N$. Wir fassen $Q(y - r)$ als Funktion in $(H^1((a, b))^N)^*$ auf, sodass $\langle Q(y - r), v \rangle_{H^1((a, b))^N} := \langle Q(y - r), v \rangle_{L^2((a, b))^N}$ für alle $v \in H^1((a, b))^N$ gilt. Dann folgt

$$\langle S'(u)^* Q(y(\cdot, u) - r), h \rangle_{L^2((a, b))} = \langle F_u(\cdot, u(\cdot), y(\cdot, u))^\top p, h \rangle_{L^2((a, b))}$$

für alle $h \in L^2((a, b))$.

Beweis. In diesem Beweis schreiben wir der Übersichtlichkeit halber $\langle \cdot, \cdot \rangle$ anstelle von $\langle \cdot, \cdot \rangle_{L^2((a, b))}$ und $\langle \cdot, \cdot \rangle_N$ anstelle von $\langle \cdot, \cdot \rangle_{L^2((a, b))^N}$. Es gilt

$$\langle p, z' \rangle_N = \int_a^b p(t)^\top z'(t) dt = - \int_a^b p'(t)^\top z(t) dt = - \langle p', z \rangle_N$$

wegen $z(a) = 0$, $p(b) = 0$. Bei Betrachtung von (S' -AWP) und (adj-AWP) kann

damit die Rechnung

$$\begin{aligned}
\langle F_u(\cdot, u(\cdot), y(\cdot, u))^\top p, h \rangle &= \langle p, F_u(\cdot, u(\cdot), y(\cdot, u))h \rangle_N = \langle p, z' - F_y(\cdot, u(\cdot), y(\cdot, u))z \rangle_N \\
&= \langle p, z' \rangle_N - \langle p, F_y(\cdot, u(\cdot), y(\cdot, u))z \rangle_N \\
&= -\langle p', z \rangle_N - \langle F_y(\cdot, u(\cdot), y(\cdot, u))^\top p, z \rangle_N \\
&= \langle -p' - F_y(\cdot, u(\cdot), y(\cdot, u))^\top p, z \rangle_N \\
&= \langle Q(y(\cdot, u) - r), z \rangle_N = \langle Q(y(\cdot, u) - r), S'(u)h \rangle_N
\end{aligned}$$

durchgeführt werden. Wir definieren nun ein Funktional $\xi \in (H^1((a, b))^N)^*$ gemäß

$$\langle \xi, v \rangle_{H^1((a, b))^N} := \langle Q(y - r), v \rangle_N$$

für alle $v \in H^1((a, b))^N$. Es folgt

$$\begin{aligned}
\langle Q(y(\cdot, u) - r), S'(u)h \rangle_N &= \langle \xi, S'(u)h \rangle_{H^1((a, b))^N} = \langle S'(u)^* \xi, h \rangle \\
&= \langle S'(u)^* Q(y - r), h \rangle
\end{aligned}$$

und damit die Aussage des Satzes. □

Wir starten nun mit der Berechnung des Gateaux-Differentials des Terms

$$\hat{J}(u) := \frac{1}{2} \int_a^b (S(u)(t) - r(t))^\top Q(t) (S(u)(t) - r(t)) dt.$$

Dazu berechnen wir zunächst die Richtungsableitung von

$$\tilde{J}(y) := \frac{1}{2} \int_a^b (y(t) - r(t))^\top Q(t) (y(t) - r(t)) dt.$$

in Richtung $v \in H^1((a, b))^N$. Es gilt

$$\begin{aligned}
\tilde{J}'(y)v &= \lim_{\tau \downarrow 0} \frac{1}{\tau} \left[\frac{1}{2} \int_a^b (y(t) + \tau v(t) - r(t))^\top Q(t) (y(t) + \tau v(t) - r(t)) \right. \\
&\quad \left. - (y(t) - r(t))^\top Q(t) (y(t) - r(t)) dt \right] \\
&= \lim_{\tau \downarrow 0} \frac{1}{\tau} \left[\frac{1}{2} \tau^2 \int_a^b v(t)^\top Q(t) v(t) dt + \tau \int_a^b (y(t) - r(t))^\top Q(t) v(t) dt \right] \\
&= \int_a^b (y(t) - r(t))^\top Q(t) v(t) dt.
\end{aligned}$$

Diese Abbildung ist linear und beschränkt bezüglich v , zweiteres folgt mit der Hölderungleichung aus

$$\begin{aligned}
\left| \int_a^b (y(t) - r(t))^\top Q(t) v(t) dt \right| &\leq \| (y - r)^\top Q v \|_{L^1((a,b))} \\
&= \left\| \sum_{i=1}^N ((y - r)^\top Q)_i v_i \right\|_{L^1((a,b))} \\
&\leq \sum_{i=1}^N \| ((y - r)^\top Q)_i v_i \|_{L^1((a,b))} \\
&\leq \sum_{i=1}^N \| ((y - r)^\top Q)_i \|_{L^2((a,b))} \| v_i \|_{L^2((a,b))} \\
&\leq \left\| \begin{pmatrix} \| ((y - r)^\top Q)_1 \|_{L^2((a,b))} \\ \vdots \\ \| ((y - r)^\top Q)_N \|_{L^2((a,b))} \end{pmatrix} \right\|_2 \left\| \begin{pmatrix} \| v_1 \|_{L^2((a,b))} \\ \vdots \\ \| v_N \|_{L^2((a,b))} \end{pmatrix} \right\|_2 \\
&= \| (y - r)^\top Q \|_{L^2((a,b))^N} \| v \|_{L^2((a,b))^N} \\
&\leq \| (y - r)^\top Q \|_{L^2((a,b))^N} \| v \|_{H^1((a,b))^N},
\end{aligned}$$

wobei die vorletzte Abschätzung aus der Cauchy-Schwarzschen Ungleichung gewonnen wurde, \tilde{J} ist demnach Gateaux-differenzierbar. Wegen $H^1((a,b))^N \hookrightarrow L^\infty((a,b))^N$ gibt es ein $C > 0$, sodass

$$\| v \|_{L^\infty((a,b))^N} \leq C \| v \|_{H^1((a,b))^N},$$

gilt, mit $Q \in L^\infty(a,b)^{N \times N}$ und (4.7) folgt nun sogar aus

$$\begin{aligned}
|\tilde{J}(y+v) - \tilde{J}(y) - \tilde{J}'(y)v| &= \left| \int_a^b v(t)^\top Q(t) v(t) dt \right| \leq \int_a^b \| Q(t) \|_2 \| v(t) \|_2^2 dt \\
&\leq \| \| Q(\cdot) \|_2 \|_{L^1((a,b))} \| \| v \|_2^2 \|_{L^\infty((a,b))} \\
&\leq \| \| Q(\cdot) \|_2 \|_{L^1((a,b))} \sum_{i=1}^N \| v_i^2 \|_{L^\infty((a,b))} \\
&= \| \| Q(\cdot) \|_2 \|_{L^1((a,b))} \sum_{i=1}^N \| v_i \|_{L^\infty((a,b))}^2 \\
&= \| \| Q(\cdot) \|_2 \|_{L^1((a,b))} \| v \|_{L^\infty((a,b))^N}^2 \\
&\leq C \| \| Q(\cdot) \|_2 \|_{L^1((a,b))} \| v \|_{H^1((a,b))^N}^2 \in \mathfrak{o}(\| v \|_{H^1((a,b))^N})
\end{aligned} \tag{5.5}$$

die Frechét-Differenzierbarkeit von \tilde{J} . Damit kann die Kettenregel angewendet werden, um

$$\hat{J}'(u)h = \nabla_u \tilde{J}(S(u))h = \tilde{J}'(S(u))(S'(u)h) = \int_a^b (S(u)(t) - r(t))^\top Q(t)(S'(u)h(t)) dt$$

für $h \in L^2((a, b))$ zu erhalten.

Mit Lemma 5.2 folgt nun

$$\begin{aligned} \hat{J}'(u)h &= \langle Q(S(u) - r), S'(u)h \rangle_{L^2((a,b))^N} = \langle S'(u)^* Q(y(\cdot, u) - r), h \rangle_{L^2((a,b))} \\ &= \langle F_u(\cdot, u(\cdot), y(\cdot, u))^\top p, h \rangle_{L^2((a,b))}. \end{aligned}$$

Das Gateaux-Differential von \hat{J} in Richtung h lässt sich also mit der Lösung p von (adj-AWP) darstellen. Das hat den Vorteil, dass nicht für jede Richtung h der Term $S'(u)h$ als Lösung von (S' -AWP) berechnet werden muss, sondern nach der Bestimmung von p eine Darstellung des Gateaux-Differentials für alle Richtungen h gefunden ist.

Die Funktion $u \mapsto |Du|((a, b))$ ist nicht Gateaux-differenzierbar, dies sieht man an folgender Rechnung: Man betrachte die Richtungsableitung dieser Abbildung im Punkt 0 in eine Richtung $h \in BV((a, b))$:

$$\lim_{\tau \downarrow 0} \frac{1}{\tau} (|D(0 + \tau h)|((a, b)) - 0) = |Dh|((a, b)).$$

Diese Abbildung ist nicht linear bezüglich h , also auch nicht Gateaux-differenzierbar.

Schließlich untersuchen wir noch die Funktionen Ψ_κ , $\kappa \in \{2, \dots, 4\}$, für Ψ_1 ist nichts zu zeigen. Wir berechnen für Ψ_2 die Richtungsableitung in Richtung $h \in L^2((a, b))$.

$$\begin{aligned} \nabla_u \Psi_2(u)h &= \lim_{\tau \downarrow 0} \frac{1}{\tau} \left[\left(\int_a^b u(t) + \tau h(t) dt \right)^2 - \left(\int_a^b u(t) dt \right)^2 \right] \\ &= \lim_{\tau \downarrow 0} \frac{1}{\tau} \left[2\tau \int_a^b u(t) dt \int_a^b h(t) dt + \tau^2 \left(\int_a^b h(t) dt \right)^2 \right] \quad (5.6) \\ &= 2 \int_a^b u(t) dt \int_a^b h(t) dt = \langle 2 \int_a^b u(s) ds, h \rangle_{L^2((a,b))}. \end{aligned}$$

Diese Abbildung ist offensichtlich linear bezüglich h , deren Beschränktheit folgt mit

$$\begin{aligned} |\nabla_u \Psi_2(u)h| &\leq 2 \int_a^b |u(t)| dt \int_a^b |h(t)| dt = 2 \|u\|_{L^1((a,b))} \|h\|_{L^1((a,b))} \\ &\leq 2 \|u\|_{L^1((a,b))} \sqrt{b-a} \|h\|_{L^2((a,b))} \end{aligned}$$

wegen $u \in L^1((a, b))$. Damit beschreibt (5.6) das Gateaux-Differential von Ψ_2 .

Die Terme Ψ_3 und Ψ_4 sind nicht Gateaux-differenzierbar. Um dies einzusehen, betrachte man die Richtungsableitungen von Ψ_3 und Ψ_4 im Punkt 0 (wobei wir $h(a)$ als rechtsseitigen Grenzwert eines guten Repräsentanten in a auffassen):

$$\begin{aligned}\lim_{\tau \downarrow 0} \frac{1}{\tau} (|0 + \tau h(a)| - 0) &= |h(a)|, \\ \lim_{\tau \downarrow 0} \frac{1}{\tau} (\|0 + \tau h\|_{L^1((a,b))} - \|0\|_{L^1((a,b))}) &= \|h\|_{L^1((a,b))}.\end{aligned}$$

Diese Ausdrücke sind nicht linear bezüglich h und damit sind Ψ_3 und Ψ_4 nicht Gateaux-differenzierbar. Im späteren Splittingverfahren werden die Funktionen $|Du|((a, b))$, Ψ_3 und Ψ_4 also dem nicht Gateaux-differenzierbaren Anteil g_κ zugewiesen.

Kapitel 6

Lösungsalgorithmus

Wie schon in Kapitel 5 angemerkt wurde, wollen wir zur Lösung von (P) das Zielfunktional in einen differenzierbaren Teil f_κ und nicht-differenzierbaren Teil g_κ aufteilen, sodass das Problem

$$\begin{aligned} \text{Minimiere } J_\kappa(u) &= f_\kappa(u) + g_\kappa(u) \\ \text{sodass } u &\in U_{ad} \end{aligned} \tag{P-Split}$$

entsteht. Wir haben schon festgestellt, welche Terme im Zielfunktional Gateaux-differenzierbar sind und können somit die Funktionen f_κ und g_κ als durch die Vorschrift

$$\begin{aligned} g_\kappa(u) &= \beta_1 |Du|((a, b)) \quad \text{für } \kappa \in \{1, 2\}, \\ g_\kappa(u) &= \beta_1 |Du|((a, b)) + \beta_2 \Psi_\kappa(u) \quad \text{für } \kappa \in \{3, 4\}, \\ f_\kappa(u) &= J_\kappa(u) - g_\kappa(u) \quad \text{für } \kappa \in \{1, \dots, 4\} \end{aligned}$$

gegeben festlegen. Im Falle der Konvexität des Zielfunctionals kann ein Proximalgradientenverfahren für das Problem

$$\begin{aligned} \text{Minimiere } J_\kappa(u) &= f_\kappa(u) + g_\kappa(u) + \delta_{U_{ad}}(u) \\ \text{sodass } u &\in L^2((a, b)) \end{aligned} \tag{P-Prox}$$

mit

$$\delta_{U_{ad}}(u) := \begin{cases} 0, & \text{wenn } u \in U_{ad} \\ \infty, & \text{wenn } u \notin U_{ad} \end{cases}$$

durchgeführt werden, um einen Minimierer zu approximieren. Im Falle von $u \in L^2((a, b)) \setminus \text{BV}((a, b))$ setzen wir $|Du|((a, b)) = \infty$, woraus $g_\kappa(u) = \infty$ folgt. Da J_κ allerdings nicht konvex ist, kann zunächst nicht garantiert werden, dass die Iterationsvorschrift des Proximalgradientenverfahrens eine Iterationsfolge mit den

gewünschten Eigenschaften erzeugt. Nichtsdestotrotz wollen wir diese Iterationsvorschrift nutzen und zeigen im Folgenden, dass die entstehende Folge unter den entsprechenden Voraussetzungen tatsächlich gute Eigenschaften besitzt. Anschließend diskretisieren wir das Problem (P) und stellen den Lösungsalgorithmus unter Verwendung der speziellen Form von U_{ad} auf.

6.1 Ein Proximalgradientenverfahren ohne Konvexität

Wir definieren zuerst die Funktion $\tilde{g}_\kappa(u) := g_\kappa(u) + \delta_{U_{ad}}(u)$ für alle $\kappa \in \{1, \dots, 4\}$. Ziel dieses Abschnitts ist es, einen groben Lösungsalgorithmus in Form eines Proximalgradientenverfahrens aufzustellen und für Folgen $\{u_k\}_{k \in \mathbb{N}} \subset L^2((a, b))$, die durch dieses Verfahren erzeugt wurden, gute Konvergenzeigenschaften nachzuweisen. Dafür benötigen wir die Lipschitzstetigkeit von $\nabla_u f_\kappa \in L^2((a, b))$ in dem Sinne, dass

$$\|\nabla_u f_\kappa(u_1) - \nabla_u f_\kappa(u_2)\|_{L^2((a, b))} \leq L_f \|u_1 - u_2\|_{L^2((a, b))} \quad (6.1)$$

für ein $L_f > 0$ und zwei Steuerungen $u_1, u_2 \in U_{ad}$ erfüllt ist. Wir zeigen zunächst diese Eigenschaft und definieren dazu die Funktionen $p_1, p_2 \in H^1((a, b))^N$ als Lösungen von (adj-AWP) bezüglich der Steuerungen u_1 und u_2 .

Für $\kappa \in \{1, 3, 4\}$ schreiben wir den Term $\|\nabla_u f_\kappa(u_1) - \nabla_u f_\kappa(u_2)\|_{L^2((a, b))}$ als

$$\begin{aligned} & \|F_u(\cdot, u_1(\cdot), y_1(\cdot))^\top p_1 - F_u(\cdot, u_2(\cdot), y_2(\cdot))^\top p_2\|_{L^2((a, b))} \\ &= \|F_u(\cdot, u_1(\cdot), y_1(\cdot))^\top p_1 \pm F_u(\cdot, u_1(\cdot), y_2(\cdot))^\top p_1 \\ & \quad \pm F_u(\cdot, u_2(\cdot), y_2(\cdot))^\top p_1 - F_u(\cdot, u_2(\cdot), y_2(\cdot))^\top p_2\|_{L^2((a, b))} \\ &\leq \|(F_u(\cdot, u_1(\cdot), y_1(\cdot)) - F_u(\cdot, u_1(\cdot), y_2(\cdot)))^\top p_1\|_{L^2((a, b))} \\ & \quad + \|(F_u(\cdot, u_1(\cdot), y_2(\cdot)) - F_u(\cdot, u_2(\cdot), y_2(\cdot)))^\top p_1\|_{L^2((a, b))} \\ & \quad + \|F_u(\cdot, u_2(\cdot), y_2(\cdot))^\top (p_1 - p_2)\|_{L^2((a, b))} \\ &= \|\langle F_u(\cdot, u_1(\cdot), y_1(\cdot)) - F_u(\cdot, u_1(\cdot), y_2(\cdot)), p_1 \rangle_2\|_{L^2((a, b))} \\ & \quad + \|\langle F_u(\cdot, u_1(\cdot), y_2(\cdot)) - F_u(\cdot, u_2(\cdot), y_2(\cdot)), p_1 \rangle_2\|_{L^2((a, b))} \\ & \quad + \|\langle F_u(\cdot, u_2(\cdot), y_2(\cdot)), p_1 - p_2 \rangle_2\|_{L^2((a, b))} \end{aligned}$$

mit $y_1 = S(u_1)$ und $y_2 = S(u_2)$. Eine Anwendung der Cauchy-Schwarz Ungleichung führt nun zur Abschätzung

$$\begin{aligned} & \|\nabla_u f_\kappa(u_1) - \nabla_u f_\kappa(u_2)\|_{L^2((a,b))} \\ & \leq \left\| \|F_u(\cdot, u_1(\cdot), y_1(\cdot)) - F_u(\cdot, u_1(\cdot), y_2(\cdot))\|_2 \|p_1(\cdot)\|_2 \right\|_{L^2((a,b))} \\ & \quad + \left\| \|F_u(\cdot, u_1(\cdot), y_2(\cdot)) - F_u(\cdot, u_2(\cdot), y_2(\cdot))\|_2 \|p_1(\cdot)\|_2 \right\|_{L^2((a,b))} \\ & \quad + \left\| \|F_u(\cdot, u_2(\cdot), y_2(\cdot))\|_2 \|p_1(\cdot) - p_2(\cdot)\|_2 \right\|_{L^2((a,b))}. \end{aligned}$$

Bei der Betrachtung der drei Summanden können wir einerseits mit der Lipschitzstetigkeit von F_u und der Hölderungleichung

$$\begin{aligned} & \left\| \|F_u(\cdot, u_1(\cdot), y_1(\cdot)) - F_u(\cdot, u_1(\cdot), y_2(\cdot))\|_2 \|p_1(\cdot)\|_2 \right\|_{L^2((a,b))}^2 \\ & = \int_a^b \|F_u(t, u_1(t), y_1(t)) - F_u(t, u_1(t), y_2(t))\|_2^2 \|p_1(t)\|_2^2 dt \\ & \leq \int_a^b L_4^2 \|y_1(t) - y_2(t)\|_2^2 \|p_1(t)\|_2^2 dt \\ & \leq L_4^2 \left\| \|y_1(\cdot) - y_2(\cdot)\|_2 \right\|_{L^1((a,b))}^2 \left\| \|p_1(\cdot)\|_2 \right\|_{L^\infty((a,b))}^2 \\ & \leq L_4^2 \|p_1\|_{L^\infty((a,b))^N}^2 \|y_1 - y_2\|_{L^2((a,b))^N}^2 \end{aligned}$$

sowie

$$\begin{aligned} & \left\| \|F_u(\cdot, u_1(\cdot), y_2(\cdot)) - F_u(\cdot, u_2(\cdot), y_2(\cdot))\|_2 \|p_1(\cdot)\|_2 \right\|_{L^2((a,b))}^2 \\ & = \int_a^b \|F_u(t, u_1(t), y_2(t)) - F_u(t, u_2(t), y_2(t))\|_2^2 \|p_1(t)\|_2^2 dt \\ & \leq \int_a^b L_3^2 |u_1(t) - u_2(t)|^2 \|p_1(t)\|_2^2 dt \\ & \leq L_3^2 \|(u_1 - u_2)^2\|_{L^1((a,b))} \left\| \|p_1(\cdot)\|_2 \right\|_{L^\infty((a,b))}^2 \\ & \leq L_3^2 \|p_1\|_{L^\infty((a,b))^N}^2 \|u_1 - u_2\|_{L^2((a,b))}^2 \end{aligned}$$

festhalten, wobei die Gültigkeit von $\left\| \|p_1(\cdot)\|_2 \right\|_{L^\infty((a,b))} \leq \|p_1\|_{L^\infty((a,b))^N}$ mit ähnlichem Vorgehen wie in (5.5) begründet werden kann. Andererseits folgt die Ungleichung

$$\begin{aligned} & \left\| \|F_u(\cdot, u_2(\cdot), y_2(\cdot))\|_2 \|p_1(\cdot) - p_2(\cdot)\|_2 \right\|_{L^2((a,b))}^2 \\ & = \int_a^b \|F_u(t, u_2(t), y_2(t))\|_2^2 \|p_1(t) - p_2(t)\|_2^2 dt \\ & \leq \left\| \|F_u(\cdot, u_2(\cdot), y_2(\cdot))\|_2 \right\|_{L^\infty((a,b))}^2 \left\| \|p_1(\cdot) - p_2(\cdot)\|_2 \right\|_{L^1((a,b))}^2 \\ & = \left\| \|F_u(\cdot, u_2(\cdot), y_2(\cdot))\|_2 \right\|_{L^\infty((a,b))}^2 \|p_1 - p_2\|_{L^2((a,b))^N}^2. \end{aligned}$$

Wir schließen aus Lemma 4.5

$$\begin{aligned}
& \|\nabla_u f_\kappa(u_1) - \nabla_u f_\kappa(u_2)\|_{L^2((a,b))} \\
& \leq L_4 \|p_1\|_{L^\infty((a,b))^N} \|y_1 - y_2\|_{L^2((a,b))^N} + L_3 \|p_1\|_{L^\infty((a,b))^N} \|u_1 - u_2\|_{L^2((a,b))} \\
& \quad + \left\| \|F_u(\cdot, u_2(\cdot), y_2(\cdot))\|_2^2 \right\|_{L^\infty((a,b))}^{\frac{1}{2}} \|p_1 - p_2\|_{L^2((a,b))^N} \\
& \leq L_4 \sqrt{\widehat{C}} \|p_1\|_{L^\infty((a,b))^N} \|u_1 - u_2\|_{L^2((a,b))} + L_3 \|p_1\|_{L^\infty((a,b))^N} \|u_1 - u_2\|_{L^2((a,b))} \\
& \quad + \left\| \|F_u(\cdot, u_2(\cdot), y_2(\cdot))\|_2^2 \right\|_{L^\infty((a,b))}^{\frac{1}{2}} \|p_1 - p_2\|_{L^2((a,b))^N}.
\end{aligned}$$

Nun schätzen wir den Term $\|p_1 - p_2\|_{L^2((a,b))^N}$ weiter ab. An dieser Stelle beobachten wir, dass für $i \in \{1, \dots, N\}$ die i -te Komponente von p_1 , die wir folgend p_{1_i} nennen, wegen $p_{1_i}(b) = 0$ die Gleichung

$$p_{1_i}(b) - p_{1_i}(t) = \int_t^b p_{1_i}'(s) \, ds \quad \Leftrightarrow \quad p_{1_i}(t) = - \int_t^b p_{1_i}'(s) \, ds$$

erfüllt (und analoges gilt für p_2), damit kann die Rechnung (4.2) imitiert werden, um

$$\begin{aligned}
\|p_1(t) - p_2(t)\|_2 &= \sqrt{\sum_{i=1}^N (p_{1_i}(t) - p_{2_i}(t))^2} \\
&= \sqrt{\sum_{i=1}^N \left(\int_t^b p_{2_i}'(\tau) - p_{1_i}'(\tau) \, d\tau \right)^2} \\
&\leq \left| \int_t^b \sqrt{\sum_{i=1}^N (p_{2_i}'(\tau) - p_{1_i}'(\tau))^2} \, d\tau \right| \\
&= \int_t^b \|p_2'(\tau) - p_1'(\tau)\|_2 \, d\tau \leq \int_a^b \|p_1'(\tau) - p_2'(\tau)\|_2 \, d\tau
\end{aligned}$$

zu erhalten, mit denselben Argumenten wie in (4.5) bekommen wir schließlich die Aussage

$$\|p_1 - p_2\|_{L^2((a,b))^N}^2 \leq (b-a)^2 \|p_1' - p_2'\|_{L^2((a,b))^N}^2.$$

Um erneut mithilfe der Grönwall-Ungleichung den Term auf der rechten Seite abschätzen zu können, betrachten wir die Rechnung

$$\begin{aligned}
\|p_1'(t) - p_2'(t)\|_2 &\leq \|F_y(t, u_2(t), y_2(t))^\top p_2(t) - F_y(t, u_1(t), y_1(t))^\top p_1(t) \\
&\quad \pm F_y(t, u_2(t), y_2(t))^\top p_1(t)\|_2 + \|Q(t)\|_2 \|y_2(t) - y_1(t)\|_2 \\
&= \|F_y(t, u_2(t), y_2(t))^\top (p_2(t) - p_1(t)) + (F_y(t, u_2(t), y_2(t)) \\
&\quad - F_y(t, u_1(t), y_1(t)))^\top p_1(t)\|_2 + \|Q(t)\|_2 \|y_2(t) - y_1(t)\|_2 \\
&\leq \|F_y(t, u_2(t), y_2(t))\|_2 \|p_1(t) - p_2(t)\|_2 + \|F_y(t, u_1(t), y_1(t)) \\
&\quad - F_y(t, u_2(t), y_2(t))\|_2 \|p_1(t)\|_2 + \|Q(t)\|_2 \|y_1(t) - y_2(t)\|_2 \\
&\leq \|F_y(t, u_2(t), y_2(t))\|_2 \int_a^b \|p_1'(\tau) - p_2'(\tau)\|_2 d\tau + \|F_y(t, u_1(t), y_1(t)) \\
&\quad - F_y(t, u_2(t), y_2(t))\|_2 \|p_1(t)\|_2 + \|Q(t)\|_2 \|y_1(t) - y_2(t)\|_2.
\end{aligned} \tag{6.2}$$

Mit der Lipschitzstetigkeit von F_y erhalten wir

$$\begin{aligned}
\|F_y(t, u_1(t), y_1(t)) - F_y(t, u_2(t), y_2(t))\|_2 &= \|F_y(t, u_1(t), y_1(t)) \pm F_y(t, u_2(t), y_1(t)) \\
&\quad - F_y(t, u_2(t), y_2(t))\|_2 \\
&\leq L_5 |u_1(t) - u_2(t)| + L_6 \|y_1(t) - y_2(t)\|_2.
\end{aligned}$$

Der Übersichtlichkeit halber führen wir die Funktion

$$\alpha(t) := (L_5 |u_1(t) - u_2(t)| + L_6 \|y_1(t) - y_2(t)\|_2) \|p_1(t)\|_2 + \|Q(t)\|_2 \|y_1(t) - y_2(t)\|_2$$

ein und benutzen Rechnung (5.1), um festzustellen, dass $\|F_y(t, u_2(t), y_2(t))\|_2 \leq L_2$ gilt. Damit folgt

$$\|p_1'(t) - p_2'(t)\|_2 \leq \alpha(t) + L_2 \int_a^b \|p_1'(\tau) - p_2'(\tau)\|_2 d\tau$$

und mit der Grönwall- und Hölderungleichung erhalten wir

$$\|p_1'(t) - p_2'(t)\|_2 \leq \alpha(t) + L_2 \int_a^b \alpha(\tau) e^{L_2(t-\tau)} d\tau \leq \alpha(t) + L_2 \|\alpha\|_{L^1((a,b))} e^{L_2(t-a)}.$$

Nun schätzen wir den Term $\|\alpha\|_{L^2((a,b))}$ ab. Dazu betrachten wir zunächst die mit der Hölderungleichung gewonnene Abschätzung

$$\begin{aligned}
& \left\| \left(L_5 |u_1(\cdot) - u_2(\cdot)| + L_6 \|y_1(\cdot) - y_2(\cdot)\|_2 \right) \|p_1(\cdot)\|_2 \right\|_{L^2((a,b))}^2 \\
&= \int_a^b \left(L_5 |u_1(t) - u_2(t)| + L_6 \|y_1(t) - y_2(t)\|_2 \right)^2 \|p_1(t)\|_2^2 dt \\
&\leq \left\| \left(L_5 |u_1(\cdot) - u_2(\cdot)| + L_6 \|y_1(\cdot) - y_2(\cdot)\|_2 \right) \right\|_{L^1((a,b))} \left\| \|p_1(\cdot)\|_2 \right\|_{L^\infty((a,b))}^2 \\
&\leq \left\| L_5 |u_1(\cdot) - u_2(\cdot)| + L_6 \|y_1(\cdot) - y_2(\cdot)\|_2 \right\|_{L^2((a,b))}^2 \|p_1\|_{L^\infty((a,b))}^N,
\end{aligned}$$

unter Zuhilfenahme von Lemma 4.5 und der Dreiecksungleichung folgt

$$\begin{aligned}
& \left\| L_5 |u_1(\cdot) - u_2(\cdot)| + L_6 \|y_1(\cdot) - y_2(\cdot)\|_2 \right\|_{L^2((a,b))} \\
&\leq \|L_5 |u_1(\cdot) - u_2(\cdot)|\|_{L^2((a,b))} + \|L_6 \|y_1(\cdot) - y_2(\cdot)\|_2\|_{L^2((a,b))} \\
&= L_5 \|u_1 - u_2\|_{L^2((a,b))} + L_6 \|y_1 - y_2\|_{L^2((a,b))}^N \\
&\leq L_5 \|u_1 - u_2\|_{L^2((a,b))} + L_6 \sqrt{\hat{C}} \|u_1 - u_2\|_{L^2((a,b))} \\
&=: C_1 \|u_1 - u_2\|_{L^2((a,b))}
\end{aligned}$$

und damit

$$\begin{aligned}
& \left\| \left(L_5 |u_1(\cdot) - u_2(\cdot)| + L_6 \|y_1(\cdot) - y_2(\cdot)\|_2 \right) \|p_1(\cdot)\|_2 \right\|_{L^2((a,b))}^2 \\
&\leq C_1^2 \|p_1\|_{L^\infty((a,b))}^N \|u_1 - u_2\|_{L^2((a,b))}^2.
\end{aligned}$$

Außerdem gilt mit erneuter Anwendung der Hölderungleichung und von Lemma 4.5 die Abschätzung

$$\begin{aligned}
\| \|Q(\cdot)\|_2 \|y_1(\cdot) - y_2(\cdot)\|_2 \|_{L^2((a,b))}^2 &= \int_a^b \|Q(t)\|_2^2 \|y_1(t) - y_2(t)\|_2^2 dt \\
&\leq \| \|Q(\cdot)\|_2 \|_{L^\infty((a,b))} \| \|y_1(\cdot) - y_2(\cdot)\|_2 \|_{L^1((a,b))}^2 \\
&\leq \| \|Q(\cdot)\|_2 \|_{L^\infty((a,b))} \|y_1 - y_2\|_{L^2((a,b))}^2 \\
&\leq C_2 \hat{C} \|u_1 - u_2\|_{L^2((a,b))}^2
\end{aligned}$$

für ein $C_2 > 0$ aufgrund von $Q \in L^\infty((a,b))^{N \times N}$ und (4.7). Wir können mit analoger Rechnung zu (4.6) wegen $p_1 \in H^1((a,b))^N \hookrightarrow L^\infty((a,b))^N$ die Gültigkeit von

$$\| \|p_1(\cdot)\|_2 \|_{L^\infty((a,b))} \leq \|p_1\|_{L^\infty((a,b))}^N \leq C_3$$

begründen. Nun muss allerdings noch gezeigt werden, dass C_3 unabhängig von u_1 ist. Man betrachte dazu das nächste Lemma.

Lemma 6.1. Sei $y \in H^1((a, b))^N$ die Lösung von (AWP) und $p \in H^1((a, b))^N$ die Lösung von (adj-AWP) bezüglich der Steuerung $u \in U_{ad}$. Dann gibt es eine von u unabhängige Konstante $C_3 > 0$, sodass

$$\|p\|_{L^\infty((a, b))^N} \leq C_3$$

erfüllt ist.

Beweis. Wir zeigen

$$\|p\|_{H^1((a, b))^N} \leq C_4,$$

für eine von u unabhängige Konstante C_4 , da mit der Einbettung $H^1((a, b))^N \hookrightarrow L^\infty((a, b))^N$ eine Konstante $C_5 > 0$ existiert, sodass

$$\|p\|_{L^\infty((a, b))^N} \leq C_5 \|p\|_{H^1((a, b))^N} \leq C_4 C_5 =: C_3$$

erfüllt ist.

Sei $y_{\nu_1} \in H^1((a, b))^N$ die Lösung von (AWP) zur konstanten Steuerungsfunktion ν_1 . Außerdem sei $p_{\nu_1} \in H^1((a, b))^N$ die Lösung von (adj-AWP) bezüglich ν_1 und y_{ν_1} . Mit der umgekehrten Dreiecksungleichung und den in diesem Abschnitt gewonnenen Erkenntnissen erhalten wir die Gültigkeit von

$$\begin{aligned} \|p\|_{H^1((a, b))^N} - \|p_{\nu_1}\|_{H^1((a, b))^N} &\leq \|p - p_{\nu_1}\|_{H^1((a, b))^N} \\ &\leq \sqrt{(b-a)^2 + 1} \|p' - p'_{\nu_1}\|_{L^2((a, b))^N} \\ &\leq \sqrt{(b-a)^2 + 1} (\|p'\|_{L^2((a, b))^N} + \|p'_{\nu_1}\|_{L^2((a, b))^N}) \\ \Leftrightarrow \|p\|_{H^1((a, b))^N} &\leq \sqrt{(b-a)^2 + 1} (\|p'\|_{L^2((a, b))^N} + \|p'_{\nu_1}\|_{L^2((a, b))^N}) + \|p_{\nu_1}\|_{H^1((a, b))^N}. \end{aligned}$$

Wir müssen also noch $\|p'\|_{L^2((a, b))^N}$ gegen eine von u unabhängige Konstante abschätzen. Dazu beobachten wir mit der Kenntnis von $\|F_y(t, u(t), y(t))\|_2 \leq L_2$ und der Betrachtung von Lemma 4.2 bzw. der Rechnung (4.2), dass auf

$$\begin{aligned} \|p'(t)\|_2 &\leq \|F_y(t, u(t), y(t))\|_2 \|p(t)\|_2 + \|Q(t)\|_2 (\|y(t)\|_2 + \|r(t)\|_2) \\ &\leq L_2 \int_a^b \|p'(s)\|_2 ds + \|Q(t)\|_2 (\|y(t)\|_2 + \|r(t)\|_2) \\ &=: L_2 \int_a^b \|p'(s)\|_2 ds + \gamma(t) \end{aligned}$$

wieder die Grönwall-Ungleichung angewendet werden kann. Wir erhalten mit der

Hölderungleichung

$$\|p'(t)\|_2 \leq L_2 \int_a^b \gamma(s) e^{L_2(t-s)} ds + \gamma(t) \leq L_2 e^{L_2(t-a)} \|\gamma\|_{L^1((a,b))} + \gamma(t).$$

Nun schließen wir

$$\begin{aligned} \|p'\|_{L^2((a,b))^N}^2 &= \int_a^b \|p'(t)\|_2^2 dt \leq \int_a^b (L_2 e^{L_2(t-a)} \|\gamma\|_{L^1((a,b))} + \gamma(t))^2 dt \\ &= \|L_2 e^{L_2(\cdot-a)} \|\gamma\|_{L^1((a,b))} + \gamma(\cdot)\|_{L^2((a,b))}^2 \\ &\leq (\|L_2 e^{L_2(\cdot-a)} \|\gamma\|_{L^1((a,b))}\|_{L^2((a,b))} + \|\gamma\|_{L^2((a,b))})^2 \\ &= \left(\sqrt{\frac{L_2}{2} e^{2L_2(b-a)} - 1} \|\gamma\|_{L^1((a,b))} + \|\gamma\|_{L^2((a,b))} \right)^2 \end{aligned}$$

und schätzen $\|\gamma\|_{L^2((a,b))}$ weiter ab.

Es gilt

$$\begin{aligned} \|\gamma\|_{L^2((a,b))}^2 &= \int_a^b \|Q(t)\|_2^2 (\|y(t)\|_2 + \|r(t)\|_2)^2 dt \\ &\leq \| \|Q(\cdot)\|_2^2 \|_{L^\infty((a,b))} \|(\|y(\cdot)\|_2 + \|r(\cdot)\|_2)^2\|_{L^1((a,b))} \\ &= \| \|Q(\cdot)\|_2^2 \|_{L^\infty((a,b))} \| \|y(\cdot)\|_2 + \|r(\cdot)\|_2 \|_{L^2((a,b))}^2 \\ &\leq \| \|Q(\cdot)\|_2^2 \|_{L^\infty((a,b))} (\| \|y(\cdot)\|_2 \|_{L^2((a,b))} + \| \|r(\cdot)\|_2 \|_{L^2((a,b))})^2. \end{aligned}$$

Aufgrund von $r \in L^2((a,b))^N$ und $\| \|r(\cdot)\|_2 \|_{L^2((a,b))} = \|r\|_{L^2((a,b))^N}$ sowie mit Lemma 4.6 folgt nun

$$\begin{aligned} \|\gamma\|_{L^2((a,b))}^2 &\leq \| \|Q(\cdot)\|_2^2 \|_{L^\infty((a,b))} (\|y\|_{L^2((a,b))^N} + \|r\|_{L^2((a,b))^N})^2 \\ &\leq \| \|Q(\cdot)\|_2^2 \|_{L^\infty((a,b))} (\tilde{C} + \|r\|_{L^2((a,b))^N})^2 =: C_6^2. \end{aligned}$$

Mit

$$\begin{aligned} \|p'\|_{L^2((a,b))^N} &\leq \sqrt{(b-a) \left(\frac{L_2}{2} e^{2L_2(b-a)} - 1 \right)} \|\gamma\|_{L^2((a,b))} + \|\gamma\|_{L^2((a,b))} \\ &\leq \left(\sqrt{(b-a) \left(\frac{L_2}{2} e^{2L_2(b-a)} - 1 \right)} + 1 \right) C_6 \end{aligned}$$

erhalten wir schließlich die Aussage des Lemmas. \square

Zusammenfassend schreiben wir also nach Anwendung der Dreiecksungleichung

$$\|\alpha\|_{L^2((a,b))} \leq C_1 C_3 \|u_1 - u_2\|_{L^2((a,b))} + \sqrt{C_2 \tilde{C}} \|u_1 - u_2\|_{L^2((a,b))} =: C_7 \|u_1 - u_2\|_{L^2((a,b))}.$$

Wir kehren zurück zum Term $\|p'_1 - p'_2\|_{L^2((a,b))^N}$.

$$\begin{aligned}\|p'_1 - p'_2\|_{L^2((a,b))^N}^2 &= \int_a^b \|p'_1(t) - p'_2(t)\|_2^2 dt \leq \int_a^b (\alpha(t) + L_2\|\alpha\|_{L^1((a,b))}e^{L_2(t-a)})^2 dt \\ &= \|\alpha(\cdot) + L_2\|\alpha\|_{L^1((a,b))}e^{L_2(\cdot-a)}\|_{L^2((a,b))}^2.\end{aligned}$$

Jetzt folgern aus den bisherigen Erkenntnissen

$$\begin{aligned}\|\alpha(\cdot) + L_2\|\alpha\|_{L^1((a,b))}e^{L_2(\cdot-a)}\|_{L^2((a,b))} &\leq \|\alpha\|_{L^2((a,b))} + L_2\|\alpha\|_{L^1((a,b))}\|e^{L_2(\cdot-a)}\|_{L^2((a,b))} \\ &\leq \|\alpha\|_{L^2((a,b))} + L_2\sqrt{b-a}\|\alpha\|_{L^2((a,b))}\sqrt{\frac{e^{2L_2(b-a)} - 1}{2L_2}} \\ &=: C_8\|\alpha\|_{L^2((a,b))} \leq C_7C_8\|u_1 - u_2\|_{L^2((a,b))},\end{aligned}$$

woraus wir

$$\|p_1 - p_2\|_{L^2((a,b))^N} \leq (b-a)C_7C_8\|u_1 - u_2\|_{L^2((a,b))}$$

und damit auch

$$\begin{aligned}\|\nabla_u f_\kappa(u_1) - \nabla_u f_\kappa(u_2)\|_{L^2((a,b))} &\leq C_3(L_4\sqrt{\hat{C}} + L_3)\|u_1 - u_2\|_{L^2((a,b))} \\ &+ \|\|F_u(\cdot, u_2(\cdot), y_2(\cdot))\|_2^2\|_{L^\infty((a,b))}^{\frac{1}{2}}(b-a)C_7C_8\|u_1 - u_2\|_{L^2((a,b))}\end{aligned}$$

erhalten. Wegen

$$\begin{aligned}\|F_u(t, u_2(t), y_2(t))\|_2 &= \lim_{\tau \downarrow 0} \frac{1}{\tau} \|F(t, u_2(t) + \tau, y_2(t)) - F(t, u_2(t), y_2(t))\|_2 \\ &\leq \lim_{\tau \downarrow 0} \frac{1}{\tau} L_1 |u_2(t) + \tau - u_2(t)| = L_1\end{aligned}$$

folgt

$$\|\|F_u(\cdot, u_2(\cdot), y_2(\cdot))\|_2^2\|_{L^\infty((a,b))}^{\frac{1}{2}} \leq \|L_1\|_{L^\infty((a,b))}^{\frac{1}{2}} = L_1,$$

sodass wir schließlich

$$\|\nabla_u f_\kappa(u_1) - \nabla_u f_\kappa(u_2)\|_{L^2((a,b))} \leq C_9\|u_1 - u_2\|_{L^2((a,b))}$$

mit

$$C_9 := C_3(L_4\sqrt{\hat{C}} + L_3) + L_1(b-a)C_7C_8$$

schreiben können.

Im Falle von $\kappa = 2$ gilt

$$\begin{aligned}
& \|\nabla_u f_\kappa(u_1) - \nabla_u f_\kappa(u_2)\|_{L^2((a,b))} \\
& \leq \|F_u(\cdot, u_1(\cdot), y_1(\cdot))^\top p_1 - F_u(\cdot, u_2(\cdot), y_2(\cdot))^\top p_2\|_{L^2((a,b))} \\
& \quad + \left\| 2\beta_2 \left(\int_a^b u_1(s) \, ds - \int_a^b u_2(s) \, ds \right) \right\|_{L^2((a,b))} \\
& \leq C_9 \|u_1 - u_2\|_{L^2((a,b))} + 2\beta_2 \left\| \int_a^b u_1(s) - u_2(s) \, ds \right\|_{L^2((a,b))} \\
& = C_9 \|u_1 - u_2\|_{L^2((a,b))} + 2\beta_2 \left| \int_a^b u_1(s) - u_2(s) \, ds \right| \|1\|_{L^2((a,b))} \\
& \leq C_9 \|u_1 - u_2\|_{L^2((a,b))} + 2\beta_2 \sqrt{b-a} \|u_1 - u_2\|_{L^1((a,b))} \\
& \leq C_9 \|u_1 - u_2\|_{L^2((a,b))} + 2\beta_2(b-a) \|u_1 - u_2\|_{L^2((a,b))} \\
& =: C_{10} \|u_1 - u_2\|_{L^2((a,b))}.
\end{aligned}$$

Damit ist die Lipschitzstetigkeit von $\nabla_u f_\kappa$ für alle $\kappa \in \{1, \dots, 4\}$ gezeigt. Mit dieser Voraussetzung lässt sich ein Proximalgradientenverfahren für (P) aufstellen.

Wir betrachten nun Algorithmus 1, der inklusive der Abnahmebedingung (6.4) aus

Algorithmus 1 Proximalgradientenverfahren mit Abnahmebedingung

Eingabe:

Gateaux-differenzierbare Funktion $f_\kappa: L^2((a,b)) \rightarrow \mathbb{R}$ und nicht-differenzierbare

Funktion $\tilde{g}_\kappa: L^2((a,b)) \rightarrow \overline{\mathbb{R}}$

Startfunktion $u_0 \in U_{ad}$

Parameter $\eta > 0, k = 0$

Ausgabe:

Folge $\{u_k\}_{k \in \mathbb{N}} \subset L^2((a,b))$

1: Wähle $\tau_k > 0$ so, dass die globale Lösung u_{k+1} von

$$\min_{u \in L^2((a,b))} f_\kappa(u_k) + \nabla_u f_\kappa(u_k)(u - u_k) + \frac{\tau_k}{2} \|u - u_k\|_{L^2((a,b))}^2 + \tilde{g}_\kappa(u) \quad (6.3)$$

die Bedingung

$$\eta \|u_{k+1} - u_k\|_{L^2((a,b))}^2 \leq f_\kappa(u_k) + \tilde{g}_\kappa(u_k) - (f_\kappa(u_{k+1}) + \tilde{g}_\kappa(u_{k+1})) \quad (6.4)$$

erfüllt.

2: Setze $k := k + 1$ und gehe zu Schritt 1.

[12, Algorithm 3.21] stammt und beginnen damit, Konvergenzeigenschaften für die von Algorithmus 1 erzeugte Iterationsfolge nachzuweisen.

Satz 6.2. Sei $\{u_k\}_{k \in \mathbb{N}}$ eine von Algorithmus 1 erzeugte Folge. Dann gilt:

- i) Die Folgen $\{u_k\}_{k \in \mathbb{N}}$ und $\{\nabla_u f_\kappa(u_k)\}_{k \in \mathbb{N}}$ sind in $L^2((a, b))$ beschränkt.
- ii) Die Folge $\{f_\kappa(u_k) + \tilde{g}_\kappa(u_k)\}_{k \in \mathbb{N}}$ ist monoton fallend und konvergent.
- iii) $\|u_{k+1} - u_k\|_{L^2((a, b))} \rightarrow 0$.

Beweis. In diesem Beweis gehen wir an vielen Stellen wie im Beweis von [12, Theorem 3.13] vor. Die Abnahmebedingung (6.4) kann umgestellt werden zu

$$f_\kappa(u_{k+1}) + \tilde{g}_\kappa(u_{k+1}) \leq f_\kappa(u_k) + \tilde{g}_\kappa(u_k) - \eta \|u_{k+1} - u_k\|_{L^2((a, b))}^2, \quad (6.5)$$

woraus man schließen kann, dass die Folge $\{f_\kappa(u_k) + \tilde{g}_\kappa(u_k)\}_{k \in \mathbb{N}}$ monoton fallend ist. Wegen $J_\kappa(u_k) = f_\kappa(u_k) + g_\kappa(u_k) \geq 0$ für alle $k \in \mathbb{N}$ ist die Folge auch von unten beschränkt und somit konvergent, womit ii) gezeigt ist.

Wäre u_1 nicht in U_{ad} , so wäre u_1 kein Minimierer von (6.3) wegen $\tilde{g}_\kappa(u_1) = \infty$ und $\tilde{g}_\kappa(u_0) < \infty$. Induktiv muss demnach $u_k \in U_{ad}$ für alle $k \in \mathbb{N}$ gelten. Einerseits erhalten wir nun wie im Beweis von Satz 4.1 die Abschätzung

$$\|u_k\|_{L^1((a, b))} = \int_a^b |u_k(t)| dt \leq \int_a^b \max(|\nu_1|, |\nu_d|) dt = (b - a) \max(|\nu_1|, |\nu_d|) =: C_1.$$

Andererseits beobachten wir, dass wegen der Konvergenz von $\{f_\kappa(u_k) + \tilde{g}_\kappa(u_k)\}_{k \in \mathbb{N}}$ auch die Folge $\{\tilde{g}_\kappa(u_k)\}_{k \in \mathbb{N}}$ konvergent und damit beschränkt ist. Daraus können wir die Gültigkeit von

$$|Du_k|((a, b)) \leq g_\kappa(u_k) \leq \tilde{g}_\kappa(u_k) \leq C_2$$

für eine Konstante $C_2 > 0$ schlussfolgern. Mit $\|u_k\|_{\text{BV}((a, b))} \leq C_1 + C_2$ folgt die Beschränktheit von $\{u_k\}_{k \in \mathbb{N}}$ in $\text{BV}((a, b))$ und wegen $\text{BV}((a, b)) \hookrightarrow L^2((a, b))$ (vgl. Satz 2.15) auch jene in $L^2((a, b))$. Die Beschränktheit von $\{\nabla_u f_\kappa(u_k)\}_{k \in \mathbb{N}}$ kann mit der Lipschitzstetigkeit von $\nabla_u f_\kappa$ und wegen $u_k \in U_{ad} \forall k \in \mathbb{N}$ durch

$$\begin{aligned} \|\nabla_u f_\kappa(u_k)\|_{L^2((a, b))} - \|\nabla_u f_\kappa(\nu_1)\|_{L^2((a, b))} &\leq \|\nabla_u f_\kappa(u_k) - \nabla_u f_\kappa(\nu_1)\|_{L^2((a, b))} \\ &\leq L_f \|u_k - \nu_1\|_{L^2((a, b))} \\ &\leq L_f \|u_k\|_{L^2((a, b))} + L_f \|\nu_1\|_{L^2((a, b))} \\ &= L_f \|u_k\|_{L^2((a, b))} + L_f \sqrt{b - a} |\nu_1| \\ \Leftrightarrow \|\nabla_u f_\kappa(u_k)\|_{L^2((a, b))} &\leq \|\nabla_u f_\kappa(\nu_1)\|_{L^2((a, b))} + L_f \|u_k\|_{L^2((a, b))} + L_f \sqrt{b - a} |\nu_1| \end{aligned}$$

bestätigt werden, womit nun i) gezeigt wurde.

Um schließlich iii) zu zeigen, summieren wir (6.5) über $k = 1, \dots, n$ für ein $n \in \mathbb{N}$.

$$\begin{aligned} \sum_{k=1}^n \left(f_\kappa(u_{k+1}) + \tilde{g}_\kappa(u_{k+1}) + \eta \|u_{k+1} - u_k\|_{L^2((a,b))}^2 \right) &\leq \sum_{k=1}^n f_\kappa(u_k) + \tilde{g}_\kappa(u_k) \\ \Leftrightarrow f_\kappa(u_{n+1}) + \tilde{g}_\kappa(u_{n+1}) + \eta \sum_{k=1}^n \|u_{k+1} - u_k\|_{L^2((a,b))}^2 &\leq f_\kappa(u_1) + \tilde{g}_\kappa(u_1). \end{aligned}$$

Der Grenzübergang $n \rightarrow \infty$ liefert nun die Ungleichung

$$\lim_{n \rightarrow \infty} \left(f_\kappa(u_{n+1}) + \tilde{g}_\kappa(u_{n+1}) \right) + \eta \sum_{k=1}^{\infty} \|u_{k+1} - u_k\|_{L^2((a,b))}^2 \leq f_\kappa(u_1) + \tilde{g}_\kappa(u_1),$$

für deren Gültigkeit die Reihe $\sum_{k=1}^{\infty} \|u_{k+1} - u_k\|_{L^2((a,b))}^2$ konvergieren muss, es muss also $\|u_{k+1} - u_k\|_{L^2((a,b))} \rightarrow 0$ und damit iii) gelten. \square

Man beachte hierbei, dass mit Kenntnis der Lipschitzkonstante L_f analog zum Beweisvorgehen in [12, Theorem 3.13] die Ungleichung

$$f_\kappa(u_{k+1}) + \tilde{g}_\kappa(u_{k+1}) \leq f_\kappa(u_k) + \tilde{g}_\kappa(u_k) - \frac{\tau_k - L_f}{2} \|u_{k+1} - u_k\|_{L^2((a,b))}^2$$

bewiesen werden kann, mit der Wahl $\tau_k \geq L_f \forall k \in \mathbb{N}$ wird also die Abnahmebedingung (6.4) für die Resultate von Satz 6.2 nicht benötigt. Außerdem folgt aus dieser Ungleichung, dass mit der Wahl $\tau_k = 2\eta + L_f \forall k \in \mathbb{N}$ die Abnahmebedingung stets erfüllt ist. Da allerdings eine schnelle Konvergenz des Algorithmus gewünscht ist, sollte τ_k so klein wie möglich gewählt werden, um große Sprünge in den Iterierten u_k nicht zu sehr zu bestrafen. Deswegen werden wir die folgende Strategie zur Wahl von τ_k wählen:

- i) Prüfe für ein initial ausgewähltes $\tau^0 > 0$ und $\theta \in (0, 1)$, ob mit $\tau_k = \tau^0 \theta^{-i}$ für $i = 0, 1, 2, \dots$ (6.4) erfüllt ist, angefangen mit $i = 0$. Gehe zu ii), sobald ein solches i gefunden ist.
- ii) Wenn τ^0 schon (6.4) erfüllt, teste zusätzlich, ob mit $\tau_k = \tau^0 \theta^i$ für $i = 1, 2, \dots$ (6.4) erfüllt ist. Stoppe, sobald dies für ein i nicht mehr der Fall ist.

Um Algorithmus 1 zu implementieren, wird zunächst eine Strategie zur Lösung des Problems (6.3) benötigt. Wir wollen (6.3) umformen und definieren dazu

$$v_k := u_k - \frac{1}{\tau_k} \nabla_u f_\kappa(u_k) \in L^2((a, b)).$$

In Anbetracht der Rechnung

$$\begin{aligned}
\frac{\tau_k}{2} \|u - v_k\|_{L^2((a,b))}^2 &= \frac{\tau_k}{2} \|u\|_{L^2((a,b))}^2 - \tau_k \langle u, u_k - \frac{1}{\tau_k} \nabla_u f_\kappa(u_k) \rangle_{L^2((a,b))} \\
&\quad + \frac{\tau_k}{2} \left\| u_k - \frac{1}{\tau_k} \nabla_u f_\kappa(u_k) \right\|_{L^2((a,b))}^2 \\
&= \frac{\tau_k}{2} \|u\|_{L^2((a,b))}^2 - \tau_k \langle u, u_k \rangle_{L^2((a,b))} + \langle u, \nabla_u f_\kappa(u_k) \rangle_{L^2((a,b))} \\
&\quad + \frac{\tau_k}{2} \|u_k\|_{L^2((a,b))}^2 - \langle u_k, \nabla_u f_\kappa(u_k) \rangle_{L^2((a,b))} + \frac{1}{2\tau_k} \|\nabla_u f_\kappa(u_k)\|_{L^2((a,b))}^2 \\
&= \frac{\tau_k}{2} \|u - u_k\|_{L^2((a,b))}^2 + \langle \nabla_u f_\kappa(u_k), u - u_k \rangle_{L^2((a,b))} + \frac{1}{2\tau_k} \|\nabla_u f_\kappa(u_k)\|_{L^2((a,b))}^2
\end{aligned}$$

ist (6.3) äquivalent zur Problemstellung

$$\min_{u \in L^2((a,b))} f_\kappa(u_k) + \frac{\tau_k}{2} \|u - v_k\|_{L^2((a,b))}^2 - \frac{1}{2\tau_k} \|\nabla_u f_\kappa(u_k)\|_{L^2((a,b))}^2 + \tilde{g}_\kappa(u). \quad (6.6)$$

Aufgrund der Form von U_{ad} und dem Term $\tilde{g}_\kappa = g_\kappa + \delta_{U_{ad}}$ müssen wir zur Lösung von (6.6) nur stückweise konstante Funktionen mit Werten in $\{\nu_1, \dots, \nu_d\}$ f.ü. berücksichtigen, da $\delta_{U_{ad}}$ andere Funktionen auf ∞ abbildet. Außerdem können die konstanten Summanden in (6.6) weggelassen werden, da wir nur am Minimierer u interessiert sind. Diese Überlegungen führen auf die Problemstellung

$$\min_{u \in U_{ad}} \frac{\tau_k}{2} \|u - v_k\|_{L^2((a,b))}^2 + g_\kappa(u), \quad (6.7)$$

eine Lösungsstrategie für dieses Problem wird in den folgenden Abschnitten erläutert.

6.2 Diskretisierung der Problemstellung

Zur Implementation des Algorithmus mit einer Programmiersprache ist zunächst eine Diskretisierung nötig, sodass wir als Lösung eine stückweise konstante Steuerung auf einem Gitter berechnen können. Ein solches Gitter über dem Intervall $[a, b]$ sei als Partition $\{t_i\}_{i=0}^n \in \mathcal{P}([a, b])$ gegeben. Der Einfachheit halber seien die Gitterpunkte t_i äquidistant verteilt, das heißt, es gelte $t_i - t_{i-1} = \Delta t \forall i \in \{1, \dots, n\}$. Das Vorgehen in diesem und den folgenden Abschnitten kann auch ohne Schwierigkeiten auf andere Gitterweiten angepasst werden.

Nun wollen wir die Problemstellung (P) über dem Gitter $\{t_i\}_{i=1}^n$ diskretisieren, wir betrachten also die beteiligten Funktionen und Terme im Zielfunktional. Dazu nehmen wir die Stetigkeit von r_j und $Q_{j,l}$ sowie die Existenz eines linksseitigen Grenzwerts in $t_0 = a$ für alle $j, l \in \{1, \dots, N\}$ an und bezeichnen mit \bar{u} den linksseitig

stetigen Repräsentanten einer Funktion $u \in \text{BV}((a, b))$ (vgl. Satz 2.17). Mit diesen Festlegungen können die Bezeichnungen

$$r_j^i := r_j(t_i), \quad Q_{j,l}^i := Q_{j,l}(t_i), \quad u^i := \bar{u}(t_i) \text{ für } j, l \in \{1, \dots, N\}, \quad i \in \{0, \dots, n-1\}$$

eingeführt werden. Die Werte $r_j(t_i)$ und $Q_{j,l}(t_i)$, $i \in \{1, \dots, n-1\}$ sind hierbei als die Funktionswerte der stetigen Repräsentanten von $r_j \in L^2((a, b))$ und $Q_{j,l} \in L^\infty((a, b))$ zu verstehen, wobei wir die Interpretationen $r_j(t_0) = \lim_{t \downarrow a} r_j(t)$, $Q_{j,l}(t_0) = \lim_{t \downarrow a} Q_{j,l}(t)$ und $\bar{u}(t_0) = \lim_{t \downarrow a} \bar{u}(t)$ verwenden.

Die Funktion y als Lösung von (AWP) zur Steuerung u lässt sich im diskreten Szenario als Matrix

$$Y(u) := \begin{pmatrix} y_1^0 & \dots & y_1^n \\ \vdots & & \vdots \\ y_N^0 & \dots & y_N^n \end{pmatrix} \text{ mit } \begin{pmatrix} y_1^0 \\ \vdots \\ y_N^0 \end{pmatrix} := y_0$$

darstellen, wobei diese Werte nach Anwendung eines numerischen Verfahrens auf (AWP) entstanden sind, y_j^i steht hier für die j -te Komponente der Lösung im Zeitpunkt t_i . Wenn beispielsweise das explizite Eulerverfahren verwendet wird, so wird Y spaltenweise durch die Iterationsvorschrift

$$y^i := \begin{pmatrix} y_1^i \\ \vdots \\ y_N^i \end{pmatrix} = \begin{pmatrix} y_1^{i-1} \\ \vdots \\ y_N^{i-1} \end{pmatrix} + \Delta t \cdot F(t_{i-1}, u^{i-1}, y^{i-1}) \text{ für } i = 1, \dots, n$$

gebildet, andere Verfahren können ohne Probleme ebenso verwendet werden. Die Abhängigkeiten der Komponenten der Matrix $Y(u)$ von u wurden hier und werden im Folgenden der Übersichtlichkeit halber weggelassen.

Da wir in den obigen Setzungen die Werte u^i als Funktionswerte eines linksseitig stetigen Repräsentanten einer Steuerung u verstanden haben, macht es Sinn, bei der Diskretisierung der von u abhängigen Terme für $t \in [t_{i-1}, t_i)$ mit einem $i \in \{1, \dots, n\}$ den Funktionswert $u(t)$ als durch u^{i-1} gegeben zu verstehen, analoges gelte für $y_j(t)$, $r_j(t)$ und $Q_{j,l}(t)$, $j, l \in \{1, \dots, N\}$.

Jetzt wollen wir die Terme im Zielfunktional diskret approximieren. Es gilt

$$\begin{aligned}
& \int_a^b (y(t) - r(t))^\top Q(t)(y(t) - r(t)) dt \\
&= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \sum_{j=1}^n \sum_{l=1}^n (y_j(t) - r_j(t))(y_l(t) - r_l(t)) Q_{j,l}(t) dt \\
&\approx \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \sum_{j=1}^n \sum_{l=1}^n (y_j^{i-1} - r_j^{i-1})(y_l^{i-1} - r_l^{i-1}) Q_{j,l}^{i-1} dt \\
&= \Delta t \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n (y_j^{i-1} - r_j^{i-1})(y_l^{i-1} - r_l^{i-1}) Q_{j,l}^{i-1}.
\end{aligned}$$

Ψ_2 und Ψ_4 lassen sich mit derselben Approximation des Integrals in diskreter Form als

$$\Psi_2(u) \approx \left(\Delta t \sum_{i=1}^n u^{i-1} \right)^2, \quad \Psi_4(u) \approx \Delta t \sum_{i=1}^n |u^{i-1}|$$

darstellen, Ψ_3 kann sogar durch $\Psi_3 = |u^0|$ exakt ins diskrete Szenario übertragen werden, da $u^0 = \lim_{t \downarrow a} \bar{u}(t)$ festgelegt wurde. Weil mit unserer Diskretisierung die Funktion u als konstant in jedem Intervall $[t_{i-1}, t_i)$ aufgefasst wird, wird das Supremum in der Definition der Totalvariation 2.1 mit der Partition $\{t_i\}_{i=0}^n$ angenommen, was die Setzung

$$|Du|((a, b)) \approx \sum_{i=1}^{n-1} |u^i - u^{i-1}|$$

motiviert. Zudem führen wir mit \bar{J}_κ die diskretisierte Zielfunktion ein, die nach den vorigen Erkenntnissen durch

$$\bar{J}_1(u, Y(u)) := \frac{\Delta t}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n (y_j^{i-1} - r_j^{i-1})(y_l^{i-1} - r_l^{i-1}) Q_{j,l}^{i-1} + \beta_1 \sum_{i=1}^{n-1} |u^i - u^{i-1}|,$$

$$\bar{J}_2(u, Y(u)) := \bar{J}_1(u, Y(u)) + \beta_2 \left(\Delta t \sum_{i=1}^n u^{i-1} \right)^2, \quad \bar{J}_3(u) := \bar{J}_1(u, Y(u)) + \beta_2 |u^0|,$$

$$\bar{J}_4(u, Y(u)) := \bar{J}_1(u, Y(u)) + \beta_2 \Delta t \sum_{i=1}^n |u^{i-1}|.$$

gegeben ist.

Als nächstes diskretisieren wir noch die Gateaux-Ableitung von $f_\kappa(u)$ für eine Funk-

tion $u \in \text{BV}((a, b))$. Dazu verstehen wir unter

$$P(u) := \begin{pmatrix} p_1^0 & \dots & p_1^{n-1} & 0 \\ \vdots & & \vdots & \vdots \\ p_N^0 & \dots & p_N^{n-1} & 0 \end{pmatrix}$$

die durch ein numerisches Verfahren gewonnene Lösung von (adj-AWP), wobei p_j^i wieder für die j -te Komponente der Lösung im Gitterpunkt t_i steht. Die Abhängigkeiten der Komponenten der Matrix $P(u)$ von u werden wieder für eine einfachere Notation weggelassen. Erneut zeigen wir am Beispiel des expliziten Eulerverfahrens, in diesem Fall rückwärts in der Zeit betrachtet, die iterative Bestimmung der Spalten von $P(u)$, die mit den Setzungen $p^i := (p_1^i, \dots, p_N^i)^\top$ und $r^i := (r_1^i, \dots, r_N^i)^\top$ gemäß

$$\begin{aligned} p^i &= p^{i+1} - \Delta t \left(-F_y(t_{i+1}, u^{i+1}, y^{i+1})^\top p^{i+1} - \begin{pmatrix} Q_{1,1}^{i+1} & \dots & Q_{1,N}^{i+1} \\ \vdots & & \vdots \\ Q_{N,1}^{i+1} & \dots & Q_{N,N}^{i+1} \end{pmatrix} (y^{i+1} - r^{i+1}) \right) \\ &= p^{i+1} + \Delta t \left(F_y(t_{i+1}, u^{i+1}, y^{i+1})^\top p^{i+1} + \begin{pmatrix} Q_{1,1}^{i+1} & \dots & Q_{1,N}^{i+1} \\ \vdots & & \vdots \\ Q_{N,1}^{i+1} & \dots & Q_{N,N}^{i+1} \end{pmatrix} (y^{i+1} - r^{i+1}) \right) \end{aligned}$$

für $i = n-1, \dots, 0$ geschieht. Für die Berechnung von p^{n-1} werden mit dieser Formel die Werte u^n, r_j^n sowie $Q_{j,l}^n$ für $j, l \in \{1, \dots, N\}$ benötigt, die bisher noch undefiniert sind. Dazu nehmen wir nun die rechtsseitige Stetigkeit von r_j und $Q_{j,l}$ in b für alle $j, l \in \{1, \dots, N\}$ an und führen die Setzungen

$$u^n := u^{n-1}, \quad r_j^n := \lim_{t \uparrow b} r_j(t), \quad Q_{j,l}^n := \lim_{t \uparrow b} Q_{j,l}(t)$$

ein. Falls die gegebenen Funktionen r_j und $Q_{j,l}$ nicht linksseitig oder rechtsseitig stetig in a bzw. b sind, kann man auch die Setzungen

$$r_j^0 := r_j(a + \epsilon), \quad r_j^n := Q_{j,l}(b - \epsilon), \quad Q_{j,l}^0 := Q_{j,l}(a + \epsilon), \quad Q_{j,l}^n := Q_{j,l}(b - \epsilon)$$

für ein $\epsilon < \Delta t$ wählen.

Nun erhalten wir für $\kappa \in \{1, 3, 4\}$ die diskrete Approximation von $\nabla_u f_\kappa(u) \in L^2((a, b))$ durch den als

$$\nabla f^i(\kappa, u) := F_u(t_i, u^i, y^i)^\top p^i$$

definierten Vektor, im Falle von $\kappa = 2$ sei wiederum

$$\nabla f^i(2, u) := F_u(t_i, u^i, y^i)^\top p^i + 2\Delta t \sum_{j=1}^n u^{j-1}$$

für $i \in \{0, \dots, n-1\}$. Mit dieser Kenntnis setzen wir schließlich

$$v^i := u^i - \frac{1}{\tau_k} \nabla f^i(\kappa, u), \quad i \in \{0, \dots, n-1\}, \quad (6.8)$$

um auch die Funktion v_k aus (6.7) in diskreter Form ausdrücken zu können.

Schließlich brauchen wir noch die Abnahmebedingung (6.4) im diskreten Fall. Diese lässt sich als

$$\eta \Delta t \sum_{i=1}^n (u_{k+1}^{i-1} - u_k^{i-1})^2 \leq \bar{J}_\kappa(u_k, Y(u_k)) - \bar{J}_\kappa(u_{k+1}, Y(u_{k+1})) \quad (6.9)$$

schreiben.

6.3 Anwendung des Bellmannschen Optimalitätsprinzips

Um einen globalen Minimierer von (6.7) zu erhalten, nutzen wir das Bellmansche Optimalitätsprinzip. Dieses beschreibt die Eigenschaft einer optimalen Lösung, dass unabhängig von der ersten, initialen Entscheidung die weiteren Entscheidungen eine optimale Lösung bezüglich des aus der initialen Entscheidung resultierenden Zustands bilden müssen. Einen Einblick in das Thema bietet [13].

Auf die Problemstellung (6.7) bezogen bedeutet das Prinzip also das Folgende: Die erste Entscheidung hat mit unserer Diskretisierung die Form $u^0 = \nu_{m_0}$, $m_0 \in \{1, \dots, d\}$. Nun können wir uns für jedes $m_0 \in \{1, \dots, d\}$ überlegen, in welcher Form die übrigen Entscheidungen getroffen werden sollten, also welche Werte u^1, \dots, u^{n-1} annehmen müssen, damit die entstehende Funktion u die Aufgabe

$$\min_{u \in U_{ad}} \frac{\tau_k \Delta t}{2} \sum_{i=2}^n (u^{i-1} - v^{i-1})^2 + \frac{\tau_k \Delta t}{2} (\nu_{m_0} - v^0)^2 + \beta_1 \sum_{i=2}^{n-1} |u^i - u^{i-1}| + \beta_1 |u^1 - \nu_{m_0}| \quad (6.10)$$

bei gegebenem Vektor $v = (v^0, \dots, v^n)^\top$ im Falle von $\kappa \in \{1, 2\}$ löst. Die Bestimmung von $u^n = u(b)$ ist dabei nicht notwendig, da dieser Punktwert die Terme $\|u - v_k\|_{L^2((a,b))}^2$, $|Du|((a,b))$ und $\Psi_\kappa(u)$ nicht beeinflusst. Wir erhalten also ein wei-

teres Optimierungsproblem, welches wir erneut mit dem Bellmannschen Optimalitätsprinzip untersuchen können und so für jede der möglichen initialen Entscheidungen $u^1 = \nu_{m_1}$, $m_1 \in \{1, \dots, d\}$ eine optimale Setzung der Werte u^2, \dots, u^{n-1} finden wollen. Diese Argumentation kann nun fortgesetzt werden, bis wir bei der Entscheidung über den Wert von u^{n-1} angelangen. Für jedes $m \in \{1, \dots, d\}$ berechnen wir also erneut obiger Argumentation folgend die optimale Lösung von

$$\min_{u \in U_{ad}} \frac{\tau_k \Delta t}{2} (u^{n-1} - v^{n-1})^2 + \frac{\tau_k \Delta t}{2} (\nu_m - v^{n-1})^2 + \beta_1 |u^{n-1} - \nu_m| \quad (6.11)$$

Der konstante Term $\frac{\tau_k \Delta t}{2} (\nu_m - v^{n-1})^2$ kann weggelassen werden und da wir nur einen Punktwert $u^{n-1} \in \{\nu_1, \dots, \nu_d\}$ suchen, können wir das Problem auch zu

$$\min_{u^{n-1} \in \{\nu_1, \dots, \nu_d\}} \frac{\tau_k \Delta t}{2} (u^{n-1} - v^{n-1})^2 + \beta_1 |u^{n-1} - \nu_m|.$$

bzw.

$$\min_{j \in \{1, \dots, d\}} \frac{\tau_k \Delta t}{2} (\nu_j - v^{n-1})^2 + \beta_1 |\nu_j - \nu_m|. \quad (6.12)$$

vereinfachen. Der Minimierer kann schließlich durch das Vergleichen der entstehenden Zielfunktionswerte für die verschiedenen $j \in \{1, \dots, d\}$ bestimmt werden.

Was wir mit dem Lösen von (6.12) für ein m erhalten, ist die optimale Setzung von u^{n-1} in dem Fall, dass u^{n-2} den Wert ν_m annimmt. In der Praxis ist diese Rechnung tatsächlich der erste Schritt, den wir bei der Suche einer Lösungsapproximation für (6.7) durchführen. Das ganze geschieht d -mal, einmal für jeden Wert $u^{n-2} = \nu_m$, $m \in \{1, \dots, d\}$. Die entstehenden d Lösungen der Aufgabe (6.11) werden nun in einer Matrix $U \in \mathbb{R}^{d \times (n-1)}$ gespeichert. Speziell setzen wir $U_{m,n-1} = j(m)$ für $m \in \{1, \dots, d\}$, wobei $j(m) \in \{1, \dots, d\}$ derjenige Index ist, der (6.12) zu festgehaltenem m löst. Außerdem speichern wir in der Matrix $\Phi \in \mathbb{R}^{d \times (n-1)}$ die Zielfunktionswerte von (6.12) für jede Setzung von $m \in \{1, \dots, d\}$, sodass

$$\Phi_{m,n-1} = \frac{\tau_k \Delta t}{2} (\nu_{U_{m,n-1}} - v^{n-1})^2 + \beta_1 |\nu_{U_{m,n-1}} - \nu_m| \quad (6.13)$$

gilt.

Diese Vorgehensweise setzen wir nun fort und suchen für alle $m \in \{1, \dots, d\}$ eine Strategie zur Wahl der Werte u^{n-2} , u^{n-1} , wenn $u^{n-3} = \nu_m$ gilt. Dies führt auf die Aufgabenstellung

$$\min_{u^{n-2}, u^{n-1} \in \{\nu_1, \dots, \nu_d\}} \frac{\tau_k \Delta t}{2} \sum_{i=n-1}^n (u^{i-1} - v^{i-1})^2 + \beta_1 |u^{n-1} - u^{n-2}| + \beta_1 |u^{n-2} - \nu_m|.$$

beziehungsweise

$$\min_{u^{n-2} \in \{\nu_1, \dots, \nu_d\}} \left[\min_{u^{n-1} \in \{\nu_1, \dots, \nu_d\}} \left(\frac{\tau_k \Delta t}{2} (u^{n-1} - v^{n-1})^2 + \beta_1 |u^{n-1} - u^{n-2}| \right) + \frac{\tau_k \Delta t}{2} (u^{n-2} - v^{n-2})^2 + \beta_1 |u^{n-2} - \nu_m| \right]. \quad (6.14)$$

Nun kann folgendermaßen vorgegangen werden: Wir probieren wieder die Setzungen $u^{n-2} = \nu_j$ für jedes $j \in \{1, \dots, d\}$ durch. Aus dem vorangegangenen Vorgehen wissen wir schon, dass für ein jedes j die Wahl $u^{n-1} = \nu_{U_{j,n-1}}$ (6.12) minimiert, wenn $u^{n-2} = \nu_j$ festgehalten wird. Die Problemstellung (6.12) ist dabei gerade als innere Minimierung in (6.14) enthalten und deren optimaler Zielfunktionswert ist für jede Wahl von $u^{n-2} = \nu_j$, $j \in \{1, \dots, d\}$ in $\Phi_{j,n-1}$ gespeichert. Wir können also die Paare $(\nu_j, \nu_{U_{j,n-1}})$ für $j \in \{1, \dots, d\}$ in (6.14) einsetzen, die d entstehenden Lösungen vergleichen und mit $U_{m,n-2} = j(m)$ für jedes $m \in \{1, \dots, d\}$ den Wert $j(m)$ speichern, für den das Paar $(\nu_{j(m)}, \nu_{U_{j(m),n-1}})$ den niedrigsten Zielfunktionswert in (6.14) realisiert. Den Zielfunktionswert von (6.14) bei der Wahl $u^{n-3} = \nu_m$ halten wir wieder in der Matrix Φ mit den Setzungen

$$\Phi_{m,n-2} = \frac{\tau_k \Delta t}{2} (\nu_{U_{m,n-2}} - v^{n-2})^2 + \beta_1 |\nu_{U_{m,n-2}} - \nu_m| + \Phi_{(U_{m,n-2}),n-1}$$

für $m = 1, \dots, d$ fest. Mit diesen Erkenntnissen und Definitionen modifizieren wir das Problem (6.14) um, sodass

$$\min_{j \in \{1, \dots, d\}} \frac{\tau_k \Delta t}{2} (\nu_j - v^{n-2})^2 + \beta_1 |\nu_j - \nu_m| + \Phi_{j,n-1} \quad (6.15)$$

entsteht.

Nun betrachten wir das Vorgehen allgemeiner und bestimmen eine Strategie zur Füllung der Spalte $U_{m,\iota}$ für ein $\iota \in \{1, \dots, n-2\}$ und alle $m \in \{1, \dots, d\}$. Nach der vorangegangenen Motivation gilt es, die Aufgabenstellung

$$\min_{j \in \{1, \dots, d\}} \frac{\tau_k \Delta t}{2} (\nu_j - v^\iota)^2 + \beta_1 |\nu_j - \nu_m| + \Phi_{j,\iota+1}. \quad (6.16)$$

zu lösen, den Löser $j(m)$ in U durch $U_{m,\iota} = j(m)$ zu speichern und die Spalte ι in Φ mit den Setzungen

$$\Phi_{m,\iota} = \frac{\tau_k \Delta t}{2} (\nu_{U_{m,\iota}} - v^\iota)^2 + \beta_1 |\nu_{U_{m,\iota}} - \nu_m| + \Phi_{(U_{m,\iota}),\iota+1} \quad (6.17)$$

zu aktualisieren. Diese Vorgehensweise wird für $\iota = n - 2, \dots, 1$ absteigend wiederholt, sodass schließlich die Werte $U_{m,1}$ und $\Phi_{m,1}$, $m \in \{1, \dots, d\}$ zur Verfügung stehen. Wir stellen fest, dass der optimale Zielfunktionswert des Problems (6.10) in dem Falle, dass u^0 als ν_m festgehalten wird, als

$$J(m) := \Phi_{m,1} + \frac{\tau_k \Delta t}{2} (\nu_m - v^0)^2 \quad (6.18)$$

geschrieben werden kann. Jetzt können die Werte $J(m)$ für $m = 1, \dots, d$ verglichen werden. Der Index $m_0 \in \{1, \dots, d\}$, der $J(m)$ minimiert, markiert die bestmögliche initiale Entscheidung für u^0 , somit gewinnen wir den Wert der gesuchten Steuerung im ersten Zeitschritt durch die Setzung $u^0 = \nu_{m_0}$. Um nun aus dieser Information den Rest der Steuerung zu erschließen, nutzen wir die in der Matrix U gespeicherten Daten: In $U_{m_0,1}$ ist der Index m_1 gespeichert, für den die Wahl $u^1 = \nu_{m_1}$ im Szenario $u^0 = \nu_{m_0}$ optimal ist, mit $m_2 = U_{m_1,2}$ erhalten wir die Setzung $u^2 = \nu_{m_2}$ und im Zeitschritt t_ι , $\iota \in \{1, \dots, n - 1\}$ folgt schließlich die Setzung $u^\iota = \nu_{m_\iota}$ mit $m_\iota = U_{m_{\iota-1}, \iota}$.

Das beschriebene Vorgehen wurde bisher nur für $\kappa \in \{1, 2\}$ betrachtet. Im Falle von $\kappa = 3$ kann es einfach modifiziert werden: Da nur der Term u^0 in Ψ_3 einwirkt, ersetzen wir (6.18) durch

$$J(m) := \Phi_{m,1} + \frac{\tau_k \Delta t}{2} (\nu_m - v^0)^2 + \beta_2 |\nu_m|. \quad (6.19)$$

Für $\kappa = 4$ fügen wir einen zusätzlichen Summanden in die Optimierungsprobleme, aus denen die u^i , $i \in \{1, \dots, n - 1\}$ berechnet werden, ein. So muss das Problem (6.12), mit dessen Hilfe die $n - 1$ -te Spalte von U bestimmt wird, im Falle von $\kappa = 4$

$$\min_{j \in \{1, \dots, d\}} \frac{\tau_k \Delta t}{2} (\nu_j - v^{n-1})^2 + \beta_1 |\nu_j - \nu_m| + \beta_2 \Delta t |\nu_j| \quad (6.20)$$

lauten, die Speicherung des Zielfunktionswerts muss natürlich auch durch die Setzung

$$\Phi_{m,n-1} = \frac{\tau_k \Delta t}{2} (\nu_{U_{m,n-1}} - v^{n-1})^2 + \beta_1 |\nu_{U_{m,n-1}} - \nu_m| + \beta_2 \Delta t |\nu_{U_{m,n-1}}| \quad (6.21)$$

angepasst werden. Für die Bestimmung der Spalte $\iota \in \{1, \dots, n - 2\}$ von U muss (6.16) durch

$$\min_{j \in \{1, \dots, d\}} \frac{\tau_k \Delta t}{2} (\nu_j - v^\iota)^2 + \beta_1 |\nu_j - \nu_m| + \beta_2 \Delta t |\nu_j| + \Phi_{j, \iota+1}. \quad (6.22)$$

ersetzt werden, für die Setzung der entsprechenden Spalte von Φ wird die Vorschrift

$$\Phi_{m,\iota} = \frac{\tau_k \Delta t}{2} (\nu_{U_{m,\iota}} - v^\iota)^2 + \beta_1 |\nu_{U_{m,\iota}} - \nu_m| + \beta_2 \Delta t |\nu_{U_{m,n-1}}| + \Phi_{(U_{m,\iota}),\iota+1} \quad (6.23)$$

verwendet. Schließlich ist auch eine Änderung von (6.18) zu

$$J(m) := \Phi_{m,1} + \frac{\tau_k \Delta t}{2} (\nu_m - v^0)^2 + \beta_2 \Delta t |\nu_m| \quad (6.24)$$

nötig.

6.4 Algorithmen zur Lösungsapproximation

Die in Abschnitt 6.3 hergeleitete Methodik wird in Algorithmus 2 festgehalten. Der Algorithmus dient zur Lösung des diskretisierten Problems zu (6.3), genauer genommen zur Lösung der diskreten Form des äquivalenten Problems (6.7) und hat eine Laufzeit von $\mathcal{O}(d^2 n)$.

Nun wollen wir auch die in Algorithmus 1 beschriebene Vorgehensweise im diskreten Szenario darstellen. Natürlich kann in der Praxis keine unendliche Folge $\{u_k\}_{k \in \mathbb{N}}$ berechnet werden. Es wird also zunächst eine Abbruchbedingung benötigt. Dazu verwenden wir zwei Parameter $k_{max} \in \mathbb{N}$ und $\delta > 0$ und stoppen den Algorithmus, wenn die Iterationszahl k_{max} erreicht wird, oder die Iterierten sich nur noch sehr wenig verändern. Zweiteres modellieren wir mit der Bedingung

$$|\bar{J}_\kappa(u_{k+1}, Y) - \bar{J}_\kappa(u_k, Y)| \leq \delta. \quad (6.25)$$

Außerdem führen wir den Parameter $i_{max} \in \mathbb{N}$ ein, der die Anzahl der Schrittweitentests begrenzen soll.

Algorithmus 3 hält nun die Vorgehensweise von Algorithmus 1 im diskreten Szenario mit der in Abschnitt 6.1 beschriebenen Schrittweitensteuerung fest.

Algorithmus 2 Diskretisiertes Lösungsverfahren nach Bellman-Prinzip

Eingabe:Parameter $\tau_k > 0$, $\kappa \in \{1, \dots, 4\}$, $n \in \mathbb{N}$

Problem Daten aus (P)

Vektor $v = (v^0, \dots, v^{n-1})^\top \in \mathbb{R}^n$ **Ausgabe:**Vektor $u = (u^0, \dots, u^{n-1})^\top \in \mathbb{R}^n$

```
1: function BELLMAN-SOLVER( $\tau_k, v, \kappa, n$ )
2:   Initialisiere  $U = 0 \in \mathbb{R}^{d \times (n-1)}$ ,  $\Phi = 0 \in \mathbb{R}^{d \times (n-1)}$ . Berechne  $\Delta t = \frac{b-a}{n}$ .
3:   for  $\iota = n - 1, \dots, 1$  do
4:     for  $m = 1, \dots, d$  do
5:       if  $\kappa \in \{1, 2, 3\}$  then
6:         if  $\iota = n - 1$  then
7:           Finde  $j \in \{1, \dots, d\}$  als Lösung von (6.12).
8:           Setze  $U_{m,n-1} = j$  und bestimme  $\Phi_{m,n-1}$  gemäß Formel (6.13).
9:         else
10:          Finde  $j \in \{1, \dots, d\}$  als Lösung von (6.16).
11:          Setze  $U_{m,\iota} = j$  und bestimme  $\Phi_{m,\iota}$  gemäß Formel (6.17).
12:        end if
13:      else
14:        if  $\iota = n - 1$  then
15:          Finde  $j \in \{1, \dots, d\}$  als Lösung von (6.20).
16:          Setze  $U_{m,n-1} = j$  und bestimme  $\Phi_{m,n-1}$  gemäß Formel (6.21).
17:        else
18:          Finde  $j \in \{1, \dots, d\}$  als Lösung von (6.22).
19:          Setze  $U_{m,\iota} = j$  und bestimme  $\Phi_{m,\iota}$  gemäß Formel (6.23).
20:        end if
21:      end if
22:    end for
23:  end for
24:
25:  if  $\kappa \in \{1, 2\}$  then
26:    Finde  $m_0 \in \{1, \dots, d\}$  als Minimierer von (6.18)
27:  else if  $\kappa = 3$  then
28:    Finde  $m_0 \in \{1, \dots, d\}$  als Minimierer von (6.19)
29:  else
30:    Finde  $m_0 \in \{1, \dots, d\}$  als Minimierer von (6.24)
31:  end if
32:  Setze  $u^0 = \nu_{m_0}$ .
33:  for  $\iota = 1, \dots, n - 1$  do
34:    Setze  $m_\iota = U_{m_{\iota-1}, \iota}$  und  $u^\iota = \nu_{m_\iota}$ .
35:  end for
36: return  $u = (u^0, \dots, u^{n-1})^\top$ .
37: end function
```

Algorithmus 3 Diskretisiertes Proximalgradientenverfahren zur Lösung von (P)

Eingabe:

Parameter $\eta > 0$, $\tau^0 > 0$, $\theta > 0$, $\delta > 0$, $k_{max} \in \mathbb{N}$, $i_{max} \in \mathbb{N}$, $\kappa \in \{1, \dots, 4\}$, $n \in \mathbb{N}$
Problemdaten aus (P) und (V2)
Startvektor $u_0 = (u_0^1, \dots, u_0^{n-1})^\top \in \mathbb{R}^n$

Ausgabe:

Vektor $u \in \mathbb{R}^n$

```
1: function PROX-GRAD( $\eta, \tau^0, \theta, \delta, k_{max}, i_{max}, \kappa, n$ )
2:   Setze  $k := 0$  und  $\Delta t = \frac{b-a}{n}$ .
3:   while Abbruchbedingung (6.25) nicht erfüllt und  $k < k_{max}$  do
4:     Berechne  $Y(u_k)$  und  $P(u_k)$  durch numerische Verfahren.
5:      $\tau_k \leftarrow \tau^0$ 
6:     Berechne Vektor  $v = (v^0, \dots, v^{n-1})$  gemäß (6.8).
7:     Setze  $i := 1$ .
8:      $u_{k+1} \leftarrow$  Bellman-Solver( $\tau_k, v, \kappa, n$ )
9:     while Abnahmebedingung (6.9) nicht erfüllt und  $i < i_{max}$  do
10:       $\tau_k \leftarrow \tau^0 \theta^{-i}$ 
11:      Aktualisiere  $v$  gemäß (6.8) mit  $\tau_k$ .
12:       $u_{k+1} \leftarrow$  Bellman-Solver( $\tau_k, v, \kappa, n$ )
13:       $i \leftarrow i + 1$ 
14:     end while
15:     if  $i = 1$  then
16:       Setze  $\tilde{u}_1 := u_{k+1}$ 
17:       while Abnahmebedingung (6.9) mit  $u_{k+1} \hat{=} \tilde{u}_i$  erfüllt und  $i < i_{max}$  do
18:         $\tau_k \leftarrow \tau^0 \theta^i$ 
19:        Aktualisiere  $v$  gemäß (6.8) mit  $\tau_k$ 
20:         $i \leftarrow i + 1$ 
21:         $\tilde{u}_i :=$  Bellman-Solver( $\tau_k, v, \kappa, n$ )
22:       end while
23:        $u_{k+1} \leftarrow \tilde{u}_{i-1}$ 
24:     end if
25:      $k \leftarrow k + 1$ 
26:   end while
27: return  $u = (u_k^0, \dots, u_k^{n-1})^\top$ .
28: end function
```

Kapitel 7

Numerische Resultate

Um die in Abschnitt 6.4 hergeleiteten Approximationsalgorithmen zu testen, wenden wir diese auf die Problemstellungen 3.2 und 3.1 an. In diesem Abschnitt wählen wir die für Algorithmus 3 nötigen Parameter gemäß

$$\eta := 10^{-6}, \quad \tau^0 := 0.01, \quad \theta := 0.5, \quad \delta = 10^{-10}, \quad k_{max} := 100, \quad i_{max} := 30.$$

Wir beginnen mit der Untersuchung von Problem 3.2 und betrachten das Szenario eines Elektroherds mit den Steuerungsoptionen $u(t) \in \{0, 0.5, \dots, 4\}$ auf dem Zeitintervall $(0, 60)$, was wir als Zeitraum über 60 Minuten interpretieren. Gewünscht sei eine Annäherung an die Kurve

$$T_d(t) := \begin{cases} 250 & \text{für } t \leq 10, \\ 250e^{\frac{1}{4}(10-t)} & \text{für } t > 10, \end{cases}$$

die als scharfes Anbraten bei einer Temperatur von 250° und anschließendem Garen bei fallender Temperatur verstanden werden kann. Die verbleibenden Parameter seien durch

$$T_Z := 22, \quad (\alpha_1, \alpha_2) := (1.5, 0.7), \quad (\gamma_1, \gamma_2, \gamma_3) := (4, 0.1, 0.5)$$

gegeben. Wir setzen zunächst $\beta_1 = \beta_2 := 0.5$ für eine ausgeglichene Gewichtung zwischen den Forderungen der Zielfunktion und diskretisieren das Intervall $(0, 60)$ durch $n = 3000$ Gitterpunkte. Der Startvektor u_0 besitze in jeder Komponente einen Zufallswert zwischen 0 und 4. Die Forderung der Ganzzahligkeit ist für u_0 noch nicht nötig, da nach jeder Anwendung von Algorithmus 2 ein ganzzahliger Vektor entsteht.

Mit diesen Setzungen erhalten wir aus Algorithmus 3 die in Abbildung 7.1 dargestellte Steuerungsfunktion u_1 und Temperaturfunktion $T_1(t)$ des Kochgeschirrs nach

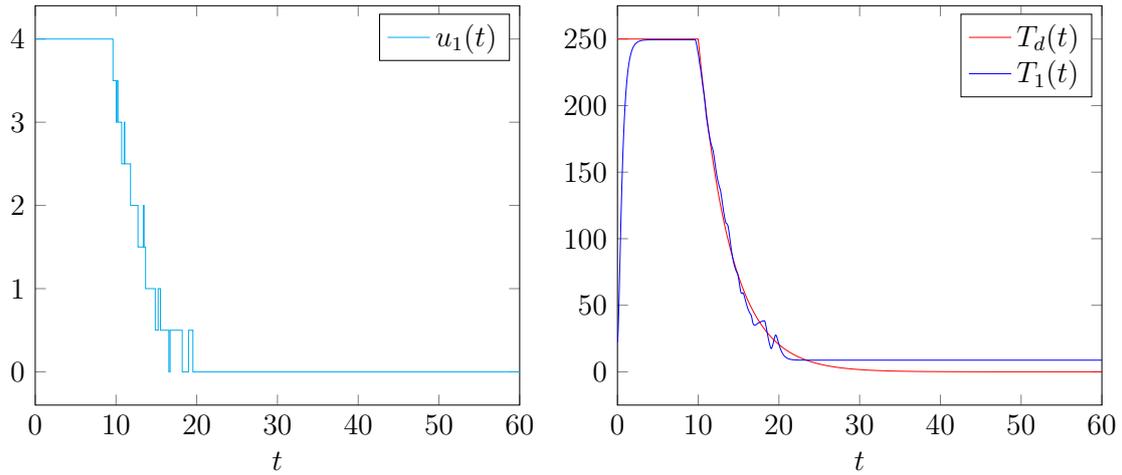


Abbildung 7.1: Lösungsapproximation zu Beispiel 3.2 mit $\beta_1 = \beta_2 := 0.5$ und Zielfunktionswert $J_4(u_1) = 12394.28$.

einer Laufzeit von 35.33s und $k = 19$ Iterationen der äußeren Schleife. Die Funktion u_1 variiert noch an einigen Stellen recht stark. Um die Wirkung des Regularisierungsterms $|Du|((a, b))$ in der Problemstellung (P) zu verdeutlichen, erhöhen wir den Einfluss dieses Summanden auf die Zielfunktion. In Abbildung 7.2 und 7.3 sind die Lösungsapproximationen u_2 und u_3 der Problemstellung 3.2 mit den gleichen Parametern wie zuvor bezüglich der Setzungen $\beta_1 := 5$ bzw. $\beta_1 := 15$ dargestellt.

Es ist zu erkennen, dass die Steuerungen u_2 und u_3 seltener zwischen den Werten in

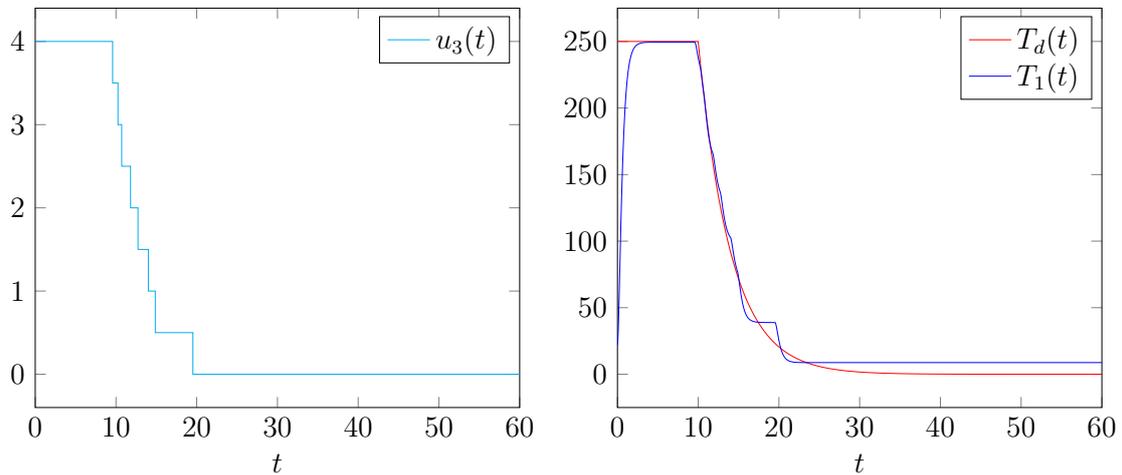


Abbildung 7.2: Lösungsapproximation zu Beispiel 3.2 mit $\beta_1 := 5$, $\beta_2 := 0.5$ und Zielfunktionswert $J_4(u_2) = 12449.42$.

$\{0, 0.5, \dots, 4\}$ springen und sich die Kurve $T_1(t)$ trotzdem noch recht nahe an $T_d(t)$ anschmiegt, auch wenn die Ungenauigkeiten von T_1 im Intervall $(15, 20)$ im Falle von u_3 größer geworden sind.

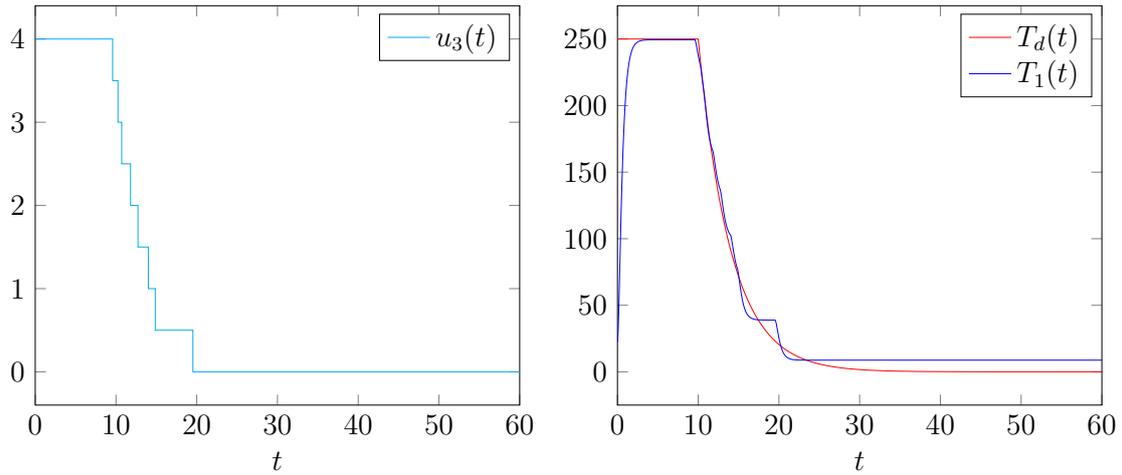


Abbildung 7.3: Lösungsapproximation zu Beispiel 3.2 mit $\beta_1 := 15$, $\beta_2 := 0.5$ und Zielfunktionswert $J_4(u_3) = 12605.12$.

Weiterhin wollen wir nun bei gleichbleibenden Parametern und $\beta_1 = \beta_2 := 0.5$ die Funktion T_1 an

$$T_d(t) := 100 \sin\left(\frac{1}{5}t\right) + 100$$

durch geschickte Wahl von u annähern. Dieses eher chaotische Kochverhalten dient weniger einer praktischen Anwendung, sondern soll viel mehr die Eigenschaften und Qualität der vom Algorithmus gewonnenen Lösungsapproximationen aufzeigen. Algorithmus 3 liefert die in Abbildung 7.4 dargestellten Ergebnisse nach einer Rechenzeit von 11.44s und $k = 7$ Iterationen der äußeren Schleife.

Wir bemerken erneut, dass die Steuerung sehr oft variiert und erhöhen β_1 auf 22,

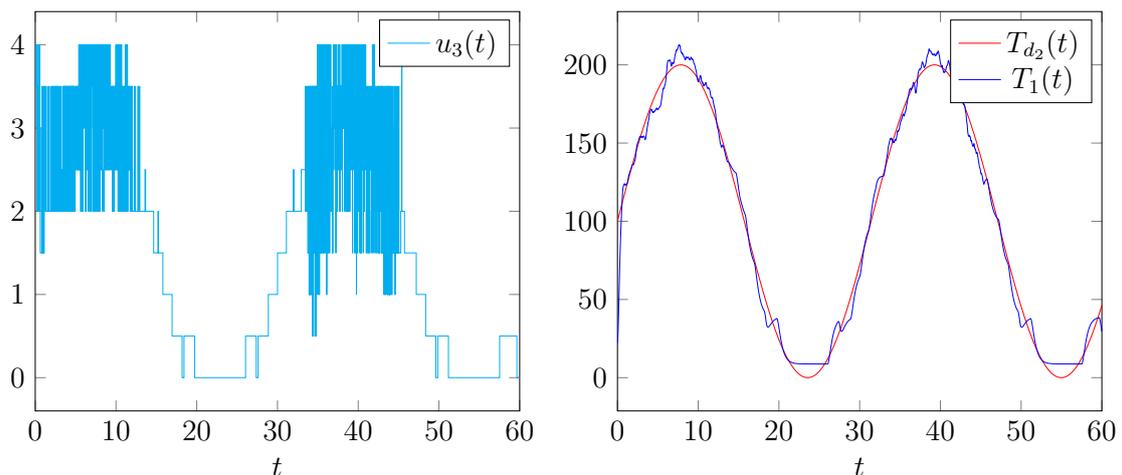


Abbildung 7.4: Lösungsapproximation zu Beispiel 3.2 mit $\beta_1 = \beta_2 := 0.5$ und Zielfunktionswert $J_4(u_4) = 2435.58$.

sodass die in Abbildung 7.5 dargestellte Steuerung u_5 erzeugt wird. Erstaunlich ist

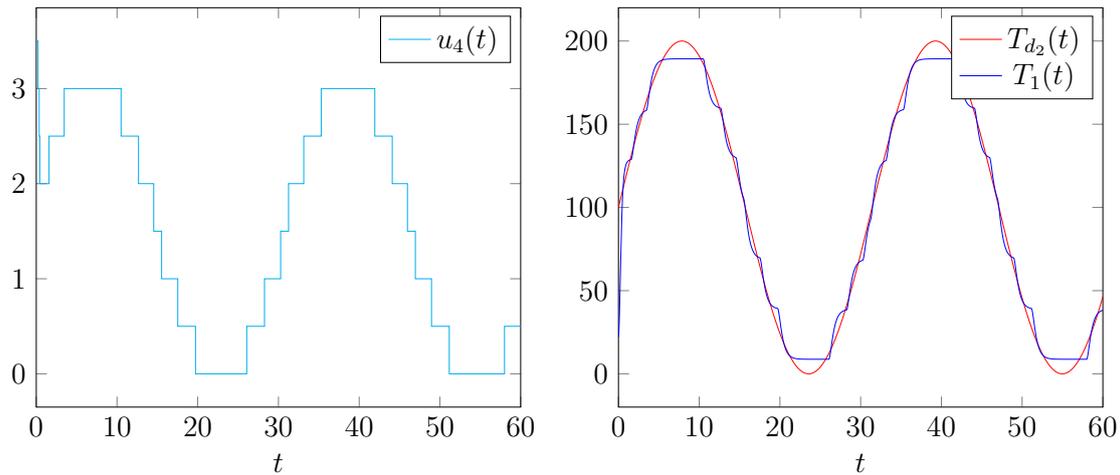


Abbildung 7.5: Lösungsapproximation zu Beispiel 3.2 mit $\beta_1 := 22$, $\beta_2 := 0.5$ und Zielfunktionswert $J_4(u_5) = 1946.81$.

hier, dass trotz Erhöhung von β_1 der Zielfunktionswert gesunken ist. Für niedrige Werte von β_1 ist die Qualität der Lösungsapproximation u stärker von dem zufällig gewählten Startvektor abhängig: In verschiedenen Durchläufen konnten bei $\beta_1 = 0.5$ Zielfunktionswerte zwischen 1880 und 2600 zu den zufälligen Startvektoren beobachtet werden, während im Falle von $\beta_1 = 22$ die Zielfunktionswerte stets 1946.81 betragen bzw. die Lösungsapproximationen zu jedem Zufallsstartvektor u_0 identisch sind. Dieser Effekt wird vor allem bei der Setzung $T_d(t) := 100$ deutlich: In den Abbildungen 7.6 und 7.7 sind zwei Lösungsapproximationen, die durch Algorithmus 3 erzeugt wurden, zu $\beta_1 := 0.5$ bzw. $\beta_1 := 20$ dargestellt.

Potentiell können also bessere Lösungsapproximationen u mit kleineren Werten

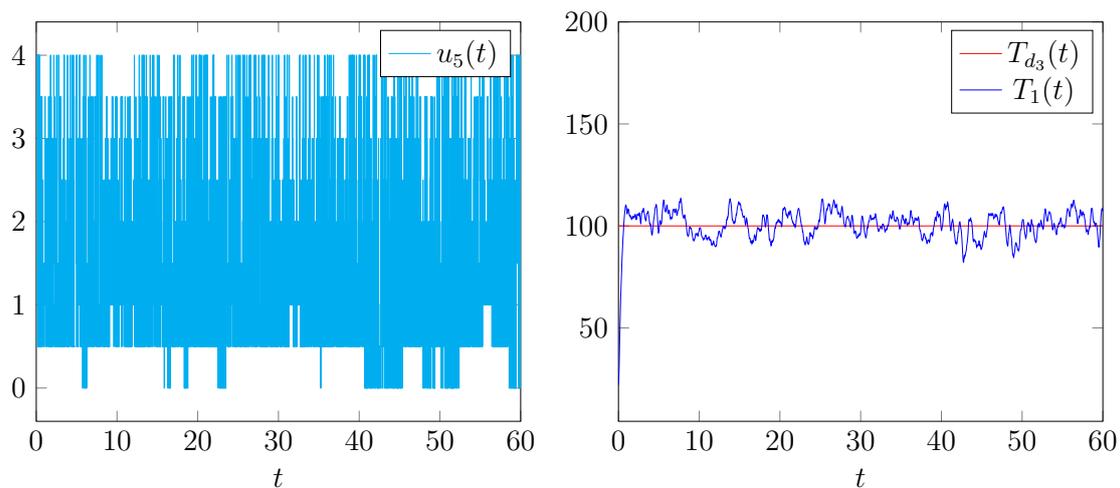


Abbildung 7.6: Lösungsapproximation zu Beispiel 3.2 mit $\beta_1 := 0.5$, $\beta_2 := 0.5$ und Zielfunktionswert $J_4(u_6) = 3656.05$.

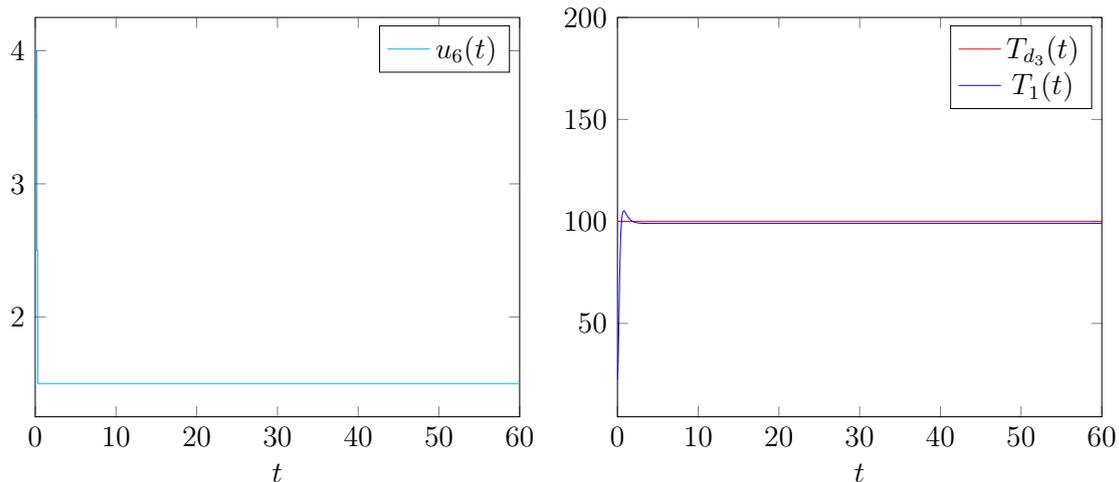


Abbildung 7.7: Lösungsapproximation zu Beispiel 3.2 mit $\beta_1 := 20$, $\beta_2 := 0.5$ und Zielfunktionswert $J_4(u_7) = 721.64$.

von β_1 gewonnen werden, praktisch gibt es in solch einem Fall allerdings viele „gute“ Lösungen, in denen der Algorithmus schließlich terminiert. Höhere Werte von β_1 erlauben nicht so viele Freiheitsgrade bei der Bestimmung von u , da Variationen stärker bestraft werden und sich somit nur in wenigen Fällen eine Änderung von u lohnt, sodass auch nur eine geringe Anzahl an Funktionen als günstige Lösungsapproximationen in Frage kommen. Der Summand $\beta_1 |Du|((a, b))$ im Zielfunktional sorgt bei großem β_1 also nicht nur für eine einfachere Steuerung, sondern auch für eine zuverlässig gute Lösungsapproximation.

Nun wollen wir auch den Algorithmus auf das Lotka-Volterra Fishing Problem mit Parametern wie in Beispiel 1.2 anwenden. Wir diskretisieren das Intervall $(0, 12)$ mit $n = 1000$ Gitterpunkten und gewinnen nach einer Laufzeit von 1.14s und $k = 2$ Durchläufen der äußeren Schleife mit $\beta_1 := 0.5$ die in Abbildung 7.8 dargestellte Lösungsapproximation.

Die Zustandsfunktionen $x(t)$ und $z(t)$ modellieren eher schlecht das gewünschte Verhalten, auch der Zielfunktionswert ist relativ groß: Der Optimalwert des relaxierten Problems ohne Totalvariation ist 0.67204 (in [5, Abschnitt 4.2] wird der Optimalwert 1.34408 angegeben, in (P) befindet sich allerdings noch der Term $\frac{1}{2}$ vor dem entsprechenden Integral). Auch eine Erhöhung von β_1 hilft hier im Gegensatz zum vorigen Beispiel nicht, mit der Setzung $\beta_1 := 3$ erhalten wir beispielsweise die in Abbildung 7.9 dargestellte Approximation.

Man könnte in der nicht gegebenen Lipschitzstetigkeit der rechten Seite F von (AWP) eine Begründung für diesen Umstand suchen, allerdings sind wie schon in Beispiel 3.4 begründet die übrigen Voraussetzungen erfüllt und die zu den Steue-

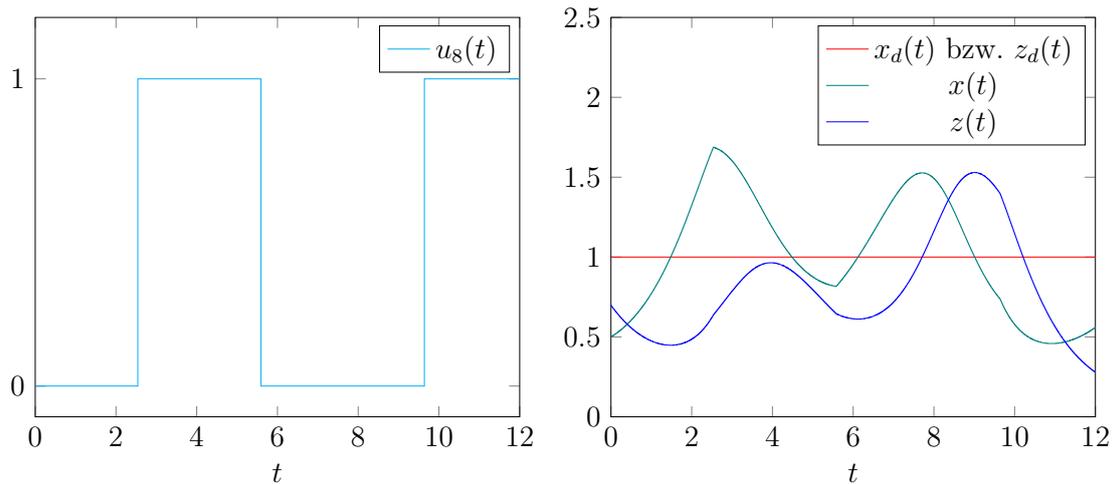


Abbildung 7.8: Lösungsapproximation zu Beispiel 3.1 mit $\beta_1 := 0.5$ und Zielfunktionswert $J_1(u_8) = 3.18$.

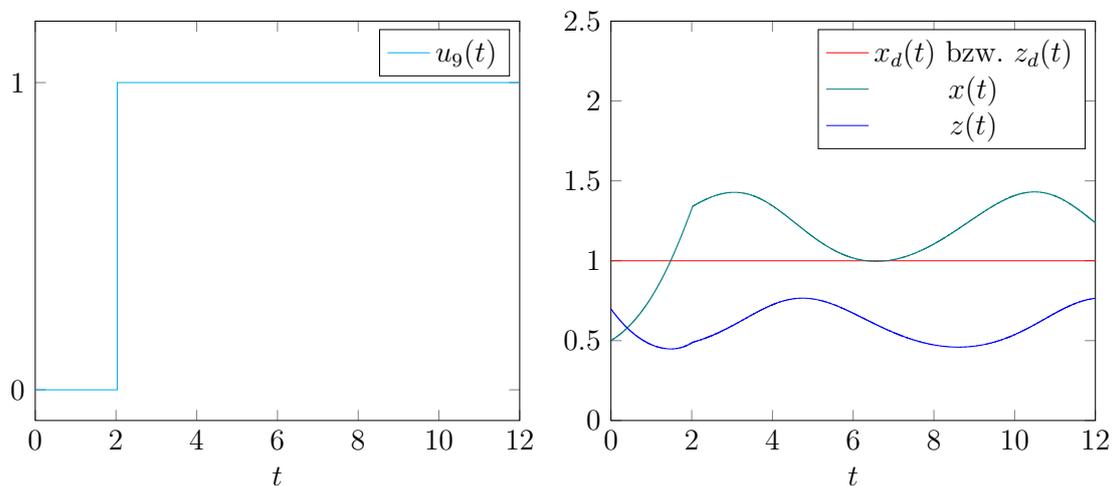


Abbildung 7.9: Lösungsapproximation zu Beispiel 3.1 mit $\beta_1 := 3$ und Zielfunktionswert $J_1(u_9) = 4.58$.

rungen gehörenden Zustände sind während der Iterationen beschränkt geblieben, sodass man hier mit lokaler Lipschitzstetigkeit ähnliche Existenzresultate herbeizuarbeiten könnte. Problematisch am Lotka-Volterra Fishing Problem ist eher die fehlende Linearität von F , woraus wir schließen, dass der Algorithmus 3 für komplexere Probleme weniger geeignet sein könnte.

Alle Ergebnisse dieses Kapitels wurden auf einem Lenovo Thinkpad T560 mit einem Intel Core i5-6300U @ 2.4GHz Prozessor und 8GB Arbeitsspeicher in Matlab Version 9.8.0.1538580 (R2020a) Update 6 erhoben.

Kapitel 8

Zusammenfassung und Fazit

In dieser Arbeit wurden Probleme der Form (P) untersucht. Nach der Herleitung einiger nützlicher Eigenschaften zu Funktionen beschränkter Variation in Kapitel 2 und der Präzisierung der Aufgabenstellung sowie Formulierung der nötigen Voraussetzungen in Kapitel 3 wurde die Wohldefiniertheit des Problems, also die Existenz einer Lösung der zugehörigen Anfangswertaufgabe (AWP) sowie des Problems (P) an sich, in den Kapiteln 3 und 4 begründet. Um Ableitungsinformationen der Terme des Zielfunktional zu erhalten, wurde in Kapitel 5 die Differenzierbarkeit der Summanden sondiert und die Gateaux-Ableitungen der differenzierbaren Terme berechnet. Schließlich gelang es in Kapitel 6, einen Lösungsalgorithmus basierend auf dem Proximalgradientenverfahren und mit dem Bellmannschen Optimalitätsprinzip zu entwickeln, gute Eigenschaften für diesen zu nachzuweisen und das Vorgehen ins diskrete Szenario zu übertragen, um numerische Lösungsapproximationen berechnen zu können. Einige numerische Experimente und deren Resultate sind in Kapitel 7 zu finden.

Die untersuchte Problemstellung (P) beschreibt eine große Problemklasse von wohldefinierten Aufgabenstellungen mit quadratischen und linearen Termen bezüglich des Zustands im Zielfunktional unter den Forderungen von (V1). Wie wir anhand des Lotka-Volterra Fishing Problems 3.1 gesehen haben, ist die Forderung nach der Lipschitzstetigkeit der rechten Seite der Differentialgleichung recht einschränkend. Hier lohnt sich die Untersuchung, inwiefern eine lokale Lipschitzstetigkeit genügt, um ähnliche Existenzresultate herleiten zu können, die die Wohldefiniertheit der Problemstellung implizieren. Allerdings wurden noch viele andere Erkenntnisse in den Kapiteln 4, 5 und 6 mithilfe der globalen Lipschitzstetigkeit bewiesen, sodass die Umgestaltung einiger Beweise dieser Arbeit unvermeidlich wäre.

Der Algorithmus 2 zur Durchführung des Proximalgradientenschritts in der inneren Schleife des Lösungsalgorithmus 3 zeichnet sich vor allem durch seine Schnelligkeit

aus. Da dessen Laufzeit linear in der Anzahl der Gitterpunkte n einer Diskretisierung ist, können vor allem Probleme mit einer geringen Anzahl an Steuerungsoptionen, beispielsweise solche mit einer binären Steuerungsfunktion, sehr schnell gelöst werden. Wenn das betrachtete Problem „schön“ genug ist und die Voraussetzungen (V1)-(V3) erfüllt sind, erzeugt Algorithmus 3 gerade für größere Werte von β_1 gute Lösungsapproximationen mit einer geringen Anzahl an Sprüngen in der Steuerung. Wir haben allerdings auch gesehen, dass bei Nichtlinearitäten wie in Problem 3.1 Schwierigkeiten entstehen können, sodass man in solch einem Fall keine guten Lösungsapproximationen von Algorithmus 3 erwarten kann. Dies motiviert, gerade für komplexere Probleme mit weniger schönen Voraussetzungen nach weiteren, geeigneteren Lösungsverfahren zu suchen.

In vielen Anwendungen treten mehrere Sachverhalte auf, die man mit unterschiedlichen Steuerungsfunktionen kontrollieren kann. Insofern ist eine Verallgemeinerung der in dieser Arbeit betrachteten Problemstellung auf das Szenario von mehrdimensionalen Steuerungsfunktionen sinnvoll. Eine theoretische Grundlage bieten hier die Erkenntnisse von L. Ambrosio, N. Fusco und D. Pallara aus [7], die schon in Kapitel 2 oft verwendet wurden.

Literatur

- [1] Laurent Condat. “A Direct Algorithm for 1-D Total Variation Denoising”. In: *IEEE Signal Processing Letters* 20.11 (2013), S. 1054–1057.
- [2] Tony F. Chan und Chiu-Kwong Wong. “Total variation blind deconvolution”. In: *IEEE Transactions on Image Processing* 7.3 (1998), S. 370–375.
- [3] Antonin Chambolle und Pierre-Louis Lions. “Image recovery via total variation minimization and related problems”. In: *Numerische Mathematik* 76.2 (1997), S. 167–188.
- [4] Kristian Bredies, Karl Kunisch und Thomas Pock. “Total Generalized Variation”. In: *SIAM Journal on Imaging Sciences* 3.3 (2010), S. 492–526.
- [5] Sebastian Sager. *On the Integration of Optimization Approaches for Mixed-Integer Nonlinear Optimal Control*. University of Heidelberg. Habilitation. 2011.
- [6] Walter Rudin. *Analysis*. 4., verbesserte Auflage. München: Oldenbourg Verlag, 2009.
- [7] Luigi Ambrosio, Nicola Fusco und Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford mathematical monographs. Oxford: Clarendon Press, 2006.
- [8] Jürgen Elstrodt. *Maß- und Integrationstheorie*. Achte, erweiterte und aktualisierte Auflage. Berlin: Springer Spektrum, 2018.
- [9] Lawrence C. Evans. *Partial differential equations*. 2. ed. Bd. 19. Graduate studies in mathematics. Providence, RI: American Math. Soc, 2010.
- [10] Hans W. Alt. *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*. 6., überarb. Aufl. Berlin und Heidelberg: Springer, 2012.
- [11] Herbert Gajewski, Konrad Gröger und Klaus Zacharias. “Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen”. In: *Mathematische Nachrichten* 67.22 (1975).

- [12] Daniel Wachsmuth. “Iterative Hard-Thresholding Applied to Optimal Control Problems with $L^0(\Omega)$ Control Cost”. In: *SIAM Journal on Control and Optimization* 57.2 (2019), S. 854–879.
- [13] Moshe Sniedovich. “A new look at Bellman’s principle of optimality”. In: *Journal of Optimization Theory and Applications* 49.1 (1986), S. 161–176.