Optimal Boundary Control of Nonlinear Hyperbolic Conservation Laws by On/Off-Switching with State constraints *

Johann M. Schmitt   Stefan Ulbrich

July 31, 2017

Abstract

This paper studies the optimal control of hyperbolic conservation laws by On/Off-Switching with state constraints. In the considered optimization problem the switching times between the on-modes and off-modes are the control variables, which have to be chosen such that the corresponding entropy solution of the hyperbolic conservation law satisfies pointwise state constraints, and a given general tracking-type functional is minimal. Using the recently developed sensitivity- and adjoint calculus by [15], we will derive necessary optimality conditions for the state constrained optimization problem. In addition, we will use Moreau-Yosida regularization for the algorithmic treatment of the pointwise state constraints. Hereby, we will prove convergence of the optimal controls and weak convergence of the corresponding Lagrange multiplier estimates of the regularized problems.

1 Introduction

We consider optimal control problems of the form

$$\min J(y(\sigma)) := \int_a^b \psi(y(\bar{t}, x; \sigma), y_d(x))\ \mathrm{d}x \tag{P}$$

where $\psi \in C^{1,1}_{\mathrm{loc}}(\mathbb{R}^2)$, $y_d \in BV_{loc}(\mathbb{R}) \cap L^\infty(\mathbb{R})$ is the desired state,

$$\sigma = (\sigma^0_{\mathrm{on}}, \sigma^1_{\mathrm{off}}, \sigma^1_{\mathrm{on}}, \ldots, \sigma^{n_\sigma}_{\mathrm{on}}, \sigma^{n_\sigma+1}_{\mathrm{off}}) \in \Sigma$$

are the switching times where

$$\Sigma := \left\{ \nu \in \mathbb{R}^{2(n_\sigma+1)}\ :\ 0 = \nu_1 < \nu_2 < \cdots < \nu_{2n_\sigma+1} < \nu_{2n_\sigma+2} = T \right\}, \tag{1.1}$$

*Supported by German Research Foundation (DFG) within the collaborative research center TRR 154 "Mathematical Modelling, Simulation and Optimization using the Example of Gas Networks"

y is the entropy solution of the OOSP

$$\begin{aligned}
y_t + f(y)_x &= g(\cdot, y), && \text{on } \Omega_{\text{on},i+1}, && i = 0, \dots, n_\sigma, && \text{(1.2a)}\\
(y_j)_t + f(y_j)_x &= g(\cdot, y_j), && \text{on } \Omega^j_{\text{off},i}, && i = 1, \dots, n_\sigma,\ j = 1, 2, && \text{(1.2b)}\\
y(0, \cdot) &= u_I, && \text{on } I, && && \text{(1.2c)}\\
y(\sigma^i_{\text{on}}, \cdot)\big|_{I_j} &= y_j(\sigma^i_{\text{on}}-, \cdot), && \text{on } I_j, && i = 1, \dots, n_\sigma,\ j = 1, 2, && \text{(1.2d)}\\
y_j(\sigma^i_{\text{off}}, \cdot) &= y(\sigma^i_{\text{off}}-, \cdot)\big|_{I_j}, && \text{on } I_j, && i = 1, \dots, n_\sigma,\ j = 1, 2, && \text{(1.2e)}\\
y_1(\cdot, 0-) &= 0, && \text{on } [\sigma^i_{\text{off}}, \sigma^i_{on}], && i = 1, \dots, n_\sigma, && \text{(1.2f)}\\
y_2(\cdot, 0+) &= 1, && \text{on } [\sigma^i_{\text{off}}, \sigma^i_{on}], && i = 1, \dots, n_\sigma, && \text{(1.2g)}
\end{aligned}$$

and the following constraints are satified:

$$\sigma \in \Sigma_{ad} \tag{1.3}$$

$$y(\bar{t}, x) \leq \bar{y}\,(x) \quad \forall x \in [a, b] \tag{1.4}$$

Hereby, we consider $I = \mathbb{R}$ with an on/off-switching device at $x = 0$. The full time interval $[0, T]$ is divided into on-phases $\left]\sigma^i_{\text{on}}, \sigma^{i+1}_{\text{off}}\right[$ and off-phases $\left]\sigma^i_{\text{off}}, \sigma^i_{\text{on}}\right[$. As described in [15], during on-phases the solution y of the OOSP coincides with the entropy of the initial-value problems (IVP) on $\Omega_{\text{on},i+1} := I \times \left]\sigma^i_{\text{on}}, \sigma^{i+1}_{\text{off}}\right[$ with initial data given in (1.2c) and (1.2e) in the sense of [9]. During the off-phases, $I_1 := (-\infty, 0]$ and $I_2 := [0, \infty)$ are separated, such that two initial-boundary value problems (IBVP) with initial data (1.2d) and boundary data (1.2f)-(1.2g) have to be solved on $\Omega^j_{\text{off},i} := I_j \times \left]\sigma^i_{\text{off}}, \sigma^i_{\text{on}}\right[,\ j = 1, 2$, respectively. The solutions of (1.2b)-(1.2d) has to be understood in the sense of [9], and the boundary conditions in the sense of [1]. Hence, through the choice of the switching time vector $\sigma \in \Sigma$, the corresponding entropy solution can be controlled until some fixed time $T > 0$. The analysis of IVPs can be found in [17], whereby the analysis of IBVPs as well as OOSPs of the form above is carried out in [14].

Remark 1.1. *In the following, we will use the same notion of a solution for the OOSP* (1.2a)-(1.2g) *as introduced in [15].*

Before we start the analysis of the OOSP, we will first state some basic assumptions. The first one coincides with assumption $(A1)$ in [15].

Assumption 1. *We assume that* $\bar{t} \in (\sigma^{n_\sigma}_{\text{on}}, T)$ *for all* $\sigma \in \Sigma_{ad}$, $\bar{y} \in C^1(\mathbb{R})$ *and* $u_I \in \mathrm{PC}^1(\mathbb{R}; x_1, \dots, x_{n_x})$, $0 \leq u_I \leq 1$, *where* $\mathrm{PC}^1(\mathbb{R}; x_1, \dots, x_{n_x})$ *denotes the space of piecewise* C^1-*functions with possible discontinuities at* $-\infty < x_1 < \dots < x_{n_x} < \infty$. *Furthermore, we assume that* $x_1, \dots, x_{n_x}$ *are all discontinuities of* u_I. *Additionally, let the flux function satisfy* $f \in C^2_{\text{loc}}(\mathbb{R})$, $f(0) = f(1) = 0$ *and* $f'' \geq m_{f''}$ *for some* $m_{f''} > 0$. *We further assume that the source term satisfies* $g \in C\left([0, T]; C^1(\mathbb{R} \times [0, 1])\right)$ *and for all* $(t, x) \in [0, T] \times \mathbb{R}$ *the following holds true*

$$g(t, x, y) \geq 0 \text{ for all } y \leq 0, \quad g(t, x, y) \leq 0 \text{ for all } y \geq 1 \tag{1.5}$$

Finally, let the source term g *be equal to zero on* $[-\varepsilon_g, \varepsilon_g]$ *for some* $\varepsilon_g > 0$.

Assumption 2. *We assume that* $\Sigma_{\text{ad}} \subset \Sigma$ *is a closed, convex and nonempty set in* $[0,T]^{2(n_\sigma+1)}$.

As stated in the following Lemma, under Assumption 1, for all $\sigma \in \Sigma$ there exists a unique solution y of the OOSP. This existence and uniqueness result can be found e.g. in [15]:

Theorem 1.1 (Existence and uniqueness for on/off-switching problems)**.** *Let Assumption 1 hold and* $u_I \in \mathrm{PC}^1(\mathbb{R}; x_1, \ldots, x_{n_x})$*, where* $0 \le u_I \le 1$*. Then for every* $\sigma \in \Sigma$ *there exists a unique entropy solution* $y = y(\sigma) \in \mathrm{L}^\infty(\Omega_T)$ *of the OOSP* (1.2a)-(1.2g)*. After a possible modification on a set of measure zero it even holds* $y \in \mathrm{C}([0,T]; \mathrm{L}^1_{\text{loc}}(\mathbb{R}))$ *and* $y(t,\cdot) \in \mathrm{BV}_{\text{loc}}(\mathbb{R})$ *for all* $t \in [0,T]$*. For almost all* $(x,t) \in \Omega_T$ *the solution of the OOSP satisfies* $y(t,x) \in [0,1]$*. Furthermore, there exists a constant* $L_\Sigma > 0$ *such that for every* $\tilde{\sigma}, \hat{\sigma} \in \Sigma_{\text{ad}}$ *and all* $t \in [0,T]$ *the following holds:*

$$\|y(t,\cdot;\tilde{\sigma}) - y(t,\cdot;\hat{\sigma})\|_{1,\text{loc}} \le L_\Sigma \|\tilde{\sigma} - \hat{\sigma}\| .$$

Proof. See e.g. the proof of [15, Cor.3.1]. □

Corollary 1.1. *Let assumption 1 hold and*

$$y \in L^\infty(\Omega_T) \cap C([0,T]; L^1_{loc}(\mathbb{R}))$$

be an entropy solution of the OOSP. Then the one sided limits $y(t,x-)$ *and* $y(t,x+)$ *exist for all* $t \in [0,T]$ *and all* $x \in \mathbb{R}$*. Furthermore, if* $x \neq 0$ *or* $(t,x) \in \Omega_{g,i+1}$*, it holds that*

$$y(t,x-) \ge y(t,x+) \tag{1.6}$$

Proof. By Theorem 2.1 it holds that $y(t,\cdot) \in \mathrm{BV}_{\text{loc}}(\mathbb{R})$ for all $t \in [0,T]$. This implies that the one-sided limits $y(t,x-)$ and $y(t,x+)$ exist for all $t \in [0,T]$ and all $x \in \mathbb{R}$, see e.g. [2], [4]. The second assertion is based on the results in [12] and can be found for example in [15]. □

Remark 1.2. *In what follows, we choose a pointwise representative*

$$y \in L^\infty(\Omega_T) \cap C([0,T]; L^1_{loc}(\mathbb{R})),$$

that coincides with $y(t,x-)$ *for* $(t,x) \notin \left]\sigma^i_{\text{off}}, \sigma^i_{\text{on}}\right] \times \{0\}$*.*

As explained e.g. in [17], using the results of Theorem 1.1 and the fact that $\psi \in C^{1,1}_{\text{loc}}(\mathbb{R}^2)$ and $y_d \in BV([a,b])$, one can prove the following result:

Corollary 1.2. *Let Assumption 1 hold and* $u_I \in \mathrm{PC}^1(\mathbb{R}; x_1, \ldots, x_{n_x})$, $0 \le u_I \le 1$*. Then the function* $\Sigma \ni \sigma \mapsto J(y(\sigma)) \in \mathbb{R}$ *is Lipschitz continuous.*

1.1 Existence of global optimal solutions

In the following theorem we will prove existence of a global optimal solution of (P).

Theorem 1.2. *Let Assumption 1 hold and assume that there exists* $\tilde{\sigma} \in \Sigma_{ad}$ *such that* $y(\bar{t}, x, \tilde{\sigma}) \le \bar{y}(x)$ *for all* $x \in [a,b]$*. Furthermore, we assume that* $x = a$ *is a continuity point of* $y(\bar{t}, \cdot, \sigma)$ *for all* $\sigma \in \Sigma_{ad}$*. Then there exists a global optimal solution for the optimal control problem (P).*

Proof. We first consider the set

$$\tilde{\Sigma} := \{\sigma \in \Sigma_{ad} : y(\bar{t}, x, \sigma) \leq \bar{y}\,(x) \quad \forall x \in [a,b]\}$$

and prove that it is compact. By Assumption 5 the set $\tilde{\Sigma}$ is non-empty, so we can consider a sequence $(\sigma_n)_{n\in\mathbb{N}} \subseteq \tilde{\Sigma} \subseteq \Sigma_{ad}$. Since Σ_{ad} is compact, there exists a convergent subsequence, again denote by $(\sigma_n)_{n\in\mathbb{N}}$, with corresponding limit $\bar{\sigma} \in \Sigma_{ad}$. Theorem 1.1 implies that the sequence $(y(\bar{t}, \cdot, \sigma_n))_{n\in\mathbb{N}}$ converges in $L^1([a,b])$ to $y(\bar{t}, \cdot, \bar{\sigma})$. Then it is well known that there exists a convergent subsequence, again denoted by $(y(\bar{t}, \cdot, \sigma_n))_{n\in\mathbb{N}}$, that converges pointwise almost everywhere to $y(\bar{t}, \cdot, \bar{\sigma})$. We hence obtain that

$$y(\bar{t}, x, \bar{\sigma}) \leq \bar{y}\,(x) \tag{1.7}$$

holds true for almost all $x \in [a,b]$. It remains to show that (1.7) holds true for all $x \in [a,b]$. To this end, suppose that for some $\hat{x} \in [a,b]$ (1.7) is violated. We first consider the case that $\hat{x} \in (a,b]$. Since (1.7) holds true for almost all $x \in [a,b]$, we can choose a sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \nearrow \hat{x}$ for $n \to \infty$ such that for all $x_n, n \in \mathbb{N}$, (1.7) is satified. Now, using that $y(\bar{t}, x, \bar{\sigma}) = y(\bar{t}, x-, \bar{\sigma})$ by Remark 1.2, we obtain that

$$y(\bar{t}, \hat{x}, \bar{\sigma}) = \lim_{n\to\infty} y(\bar{t}, x_n, \bar{\sigma}) \leq \bar{y}(\hat{x}),$$

which is a contradiction. Now we consider the case $x = a$. As above, We can choose a sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \downarrow \hat{x}$ for $n \to \infty$ such that for all $x_n, n \in \mathbb{N}$, (1.7) is satified. Now, using that $\hat{x} = a$ is by assumption a continuity point of $y(\bar{t}, \cdot, \bar{\sigma})$, we obtain that

$$y(\bar{t}, \hat{x}, \bar{\sigma}) = \lim_{n\to\infty} y(\bar{t}, x_n, \bar{\sigma}) \leq \bar{y}(\hat{x}),$$

which is again a contradiction.

Since $y(\bar{t}, x, \sigma_n) \leq \bar{y}\,(x)$ for all $x \in [a,b]$ and all $n \in \mathbb{N}$, we obtain

$$y(\bar{t}, x, \bar{\sigma}) \leq \bar{y}\,(x) \quad x \in [a,b].$$

Thus, we can conclude that $\bar{\sigma} \in \tilde{\Sigma}$ and have proved compactness of $\tilde{\Sigma}$. We now consider a minimizing sequence $(\sigma_n)_{n\in\mathbb{N}} \subseteq \tilde{\Sigma}$, i.e.

$$\hat{J}(\sigma_n) \to \inf_{\sigma\in\tilde{\Sigma}} \hat{J}(\sigma) \quad \text{for } k \to \infty,$$

where $\hat{J}(\cdot) = J(y(\cdot))$ denotes the reduced cost functional of (P). Since $\tilde{\Sigma}$ is compact, there exists a convergent subsequence, again denoted by $(\sigma_n)_{n\in\mathbb{N}}$, with $\sigma_n \to \bar{\sigma} \in \tilde{\Sigma}$. Since $\hat{J}$ is Lipschitz-continuous w.r.t. σ by Corollary 1.2, we obtain

$$\hat{J}(\bar{\sigma}) = \inf_{\sigma\in\tilde{\Sigma}} \hat{J}(\sigma)$$

Hence, $\bar{\sigma}$ is a global minimum for (P). $\square$

2 Optimality conditions

2.1 Structure of the solution of the OOSP

In this section we will summarize some results of [15] concerning the structure of entropy solutions of the On/Off-Switching Problem (OOSP). Those results were obtained by applying Dafermos' theory of generalized characteristics, see [4], and are an extension of the results in [17]. We will use those results to exploit the nice structure of entropy solutions of the OOSP in order to derive optimality conditions. To this end, we will need the following assumption:

Assumption 3. *We assume that Assumption 1 holds. In addition, let the flux function satisfy* $f'^{-1} \in C^{2,\alpha}_{loc}(\mathbb{R})$ *for some* $\alpha \in]0,1]$. *Furthermore, we assume that the source term satisfies* $g \in C^1([0,T] \times \mathbb{R} \times [0,1])$ *and is affine linear w.r.t.* y.

Definition 1. We say that $\sigma \in \Sigma_{ad}$ satisfies the Nondegeneracy-condition (ND) if σ^i_{off} is nondegenerated for all $i \in \{1, \dots, n_\sigma\}$, $y(\bar{t}, \cdot, \sigma)$ has no shock generation points on $[a, b]$ and $y(\bar{t}, \cdot, \sigma)$ has N nondegenerated shocks

$$a < \bar{x}_1(\sigma) < \cdots < \bar{x}_N(\sigma) < b$$

that are no shock interaction points.

According to [13] nondegenerated red-switchings and nondegenerated shocks are defined as follows:

Definition 2. For a given $\sigma \in \Sigma$, we call a red-switching point σ^i_{off} nondegenerated, if the following conditions hold: 0 is a continuity point of $y(\sigma^i_{\text{off}}, \cdot; (0, \sigma^1_{\text{off}}, \sigma^1_{\text{on}}, \dots, \sigma^{i-1}_{\text{on}}, T))$ and no shock generation point. Furthermore, the corresponding genuine backward characteristic through $(0, \sigma^i_{\text{off}})$ ends in some point $(0, \theta)$ with $\theta \in (\sigma^j_{\text{off}}, \sigma^j_{\text{on}})$ for some $j < i$, or in some point $(x_0, 0)$ such that $u_I(x)$ continuously differentiable in $x = x_0$, or $(0, \sigma^i_{\text{off}})$ lies in the interior of a rarefaction created either by a green switching or by a discontinuity of the initial data u_I. Finally, σ^i_{off} has to satisfy $y(\sigma^i_{\text{off}}-, 0) \notin \{0, 1\}$.

Definition 3. A discontinuity $\bar{x}$ of $y(\bar{t}, \cdot, \sigma)$ is called nondegnerated, if it is no shock generation point and the corresponding minimal and the maximal genuine backward characteristics through $(\bar{x}, \bar{t})$ end in some point $(0, \theta)$ with $\theta \in (\sigma^j_{\text{off}}, \sigma^j_{\text{on}})$ for some $j < i$, or in some point $(x_0, 0)$ such that $u_I(x)$ continuously differentiable in $x = x_0$, or in the interior of a rarefaction wave which is created either by a green switching or by a discontinuity of the initial data u_I.

Theorem 2.1. *(a) Let Assumption 3 hold and assume in addition that* $f \in C^3_{\text{loc}}(\mathbb{R})$, $g \in C^2([0,T] \times \mathbb{R} \times [-1, 0])$ *and* $u_I \in \text{PC}^2(\mathbb{R}; x_1, \dots, x_{n_x})$, $0 \le u_I \le 1$. *In addition, let* $\sigma \in \Sigma_{ad}$ *be chosen such that all off-switchings are non-degenerated. Then for almost all* $t \in (0, T]$ *and* $-\infty < a < b < \infty$, $y(t, \cdot, \sigma)$ *has no shock generation points on* $I := [a, b]$ *and finitely many nondegenerate shocks at* $a < x_1 < \cdots < x_N < b$.

(b) Let Assumption 3 hold and σ *satisfy (ND). Then* $y(\bar{t}, \cdot, \sigma)$ *can be cut into smooth functions* y_i *which depend continuously differentiable on* σ *and are separated by points* $\tilde{x}_1, \dots, \tilde{x}_K \in (a, b)$, *for some* $K \ge N$, *which depend continuously differentiable on* σ *again. More precisely, there exists*

a neighborhood U_Σ of $\bar{\sigma}$, $\varepsilon > 0$ and continuously Fréchet-differentiable mappings

$$Y_k : (x, \sigma) \in (x_{k-1}(\bar{\sigma}) - \varepsilon, x_k(\bar{\sigma}) + \varepsilon) \times U_\Sigma \mapsto Y_k\left(\bar{t}, x, \sigma\right) \in \mathbb{R} \tag{2.1}$$

$$x_k : \sigma \in U_\Sigma \mapsto x_k\left(\sigma\right) \in \left(x_k(\bar{\sigma}) - \frac{\varepsilon}{2}, x_k(\bar{\sigma}) + \frac{\varepsilon}{2}\right) \in \mathbb{R} \tag{2.2}$$

for all $k \in \{1, \ldots, K+1\}$, where $x_0 = a$ and $x_{K+1} = b$, such that for all $\sigma \in U_\Sigma$ the solution $y\left(\bar{t}, \cdot, \sigma\right)$ is on $[a, b]$ given by:

$$y\left(\bar{t}, x, \sigma\right) = Y_1(\bar{t}, x, \sigma) \cdot 1_{[a, x_1(\sigma)]}(x) + \sum_{k=2}^{K+1} Y_k(\bar{t}, x, \sigma) \cdot 1_{(x_{k-1}(\sigma), x_k(\sigma)]}(x)$$

Finally, the mappings

$$\sigma \in U_\Sigma \mapsto Y_k\left(\bar{t}, \cdot, \sigma\right) \in C\left(I_k\right), \quad k = 1, \ldots, K$$

are continuously Fréchet differentiable, where

$$I_k := \left(x_{k-1}(\bar{\sigma}) - \varepsilon, x_k(\bar{\sigma}) + \varepsilon\right).$$

According to the results in [15], the derivatives of the mappings x_j in (2.2) *can be computed as follows: If $x_j(\sigma)$ is a discontinuity of $y(\bar{t}, \cot, \sigma)$, its derivative is given by*

$$\frac{\partial}{\partial \sigma} x_j(\sigma) \cdot \delta\sigma = \sum_{i=1}^{n_\sigma} p_{\mathrm{off}}^i \cdot \delta\sigma_{\mathrm{off}}^i + p_{\mathrm{on}}^i \cdot \delta\sigma_{\mathrm{on}}^i, \tag{2.3}$$

where

$$\begin{aligned} p_{\mathrm{on}}^i &:= \int_{f'(0)}^{f'(1)} \lim_{s \searrow \sigma_{\mathrm{on}}^i} p(s, (s - \sigma_{\mathrm{on}}^i) \cdot w) \cdot w \cdot \left(f'^{-1}\right)'(w)\ \mathrm{d}w, \\ p_{\mathrm{off}}^i &:= \left(p(\sigma_{\mathrm{off}}^i, 0+) - p(\sigma_{\mathrm{off}}^i, 0-)\right) \cdot f(y(\sigma_{\mathrm{off}}^i, 0)) \end{aligned}$$

and p is the reversible solution of the adjoint equation

$$p_t + f'(y) p_x = -g_y(\cdot, \hat{y}) p, \quad p\left(\bar{t}, \cdot\right) = p^{\bar{t}}, \tag{2.4}$$

$$p\left(\bar{t}, \cdot\right) = \mathbb{1}_{x_j(\sigma)}(x) \frac{1}{[y(\bar{t}, x_j(\sigma), \sigma)]}. \tag{2.5}$$

If $x_j(\sigma)$ lies on the boundary of a rarefaction wave created by the inital conditions, its derivative is 0*. If $x_j(\sigma)$ lies on the boundary of a rarefaction wave created by a green switching, the corresponding derivative can be computed by solving the linearized characteristic equation (3.5a)-(3.5c) in [15]. Based on this linearized characteristic equation, the formulas of the derivatives of the mappings* (2.1) *are given in Lemma 7.2.1, Lemma 7.2.3., Lemma 7.2.7 and Lemma 7.2.8 in [13] and in Lemma 5.6 in [15].*

Proof. The proof of part (a) is quite similar to the proof of Theorem 3.3.6 in [16]. For the proof of (b), see Chapter 5 in [15]. $\square$

2.2 Differentiability of the cost functional

Another important result is the continuous Fréchet-differentiability of the reduced cost functional $\hat{J}(\cdot) := J(y(\cdot))$, see also [16], [13], [15] and [14].

Theorem 2.2 (Fréchet-differentiability of the tracking-type functional)**.** *Let Assumption 3 hold and $\bar{\sigma}$ satisfy (ND). Then the mapping $\sigma \in \Sigma_{ad} \mapsto \hat{J}(\sigma) \in \mathbb{R}$ is continuously Fréchet-differentiable in $\bar{\sigma}$. The derivative in a direction $\delta\sigma \in \Sigma$ is given by:*

$$\hat{J}'(\bar{\sigma}) \cdot \delta\sigma = \sum_{i=1}^{n_\sigma} p^i_{\text{off}} \cdot \delta\sigma^i_{\text{off}} + p^i_{\text{on}} \cdot \delta\sigma^i_{\text{on}}, \tag{2.6}$$

where

$$\begin{aligned} p^i_{\text{on}} &:= \int_{f'(0)}^{f'(1)} \lim_{s \searrow \sigma^i_{\text{on}}} p(s, (s - \sigma^i_{\text{on}}) \cdot w) \cdot w \cdot \left(f'^{-1}\right)' (w) \, \mathrm{d}w, \\ p^i_{\text{off}} &:= \left(p(\sigma^i_{\text{off}}, 0+) - p(\sigma^i_{\text{off}}, 0-)\right) \cdot f(y(\sigma^i_r, 0)) \end{aligned}$$

and p is the reversible solution of the adjoint equation

$$p_t + f'(y)p_x = -g_y(\cdot, \hat{y})p, \quad p\left(\bar{t}, \cdot\right) = p^{\bar{t}}, \tag{2.7}$$

$$p\left(\bar{t}, \cdot\right) = \bar{\psi}_y(x) \tag{2.8}$$

where

$$\bar{\psi}_y(x) := \int_0^1 \psi_y\left(y(\bar{t}, x+; \bar{\sigma})\right) + \tau[y(\bar{t}, x; \bar{\sigma})], y_d(x+) + \tau[y_d(x)]) \, \mathrm{d}\tau.$$

Proof. See proof of Theorem 4.8 in [13]. □

Notation 2.1. *Since the first and the last entry in the switching time vector $\sigma = (\sigma^0_{\text{on}}, \sigma^1_{\text{off}}, \sigma^1_{\text{on}}, \ldots, \sigma^{n_\sigma}_{\text{on}}, \sigma^{n_\sigma+1}_{\text{off}}) \in \Sigma$ are fixed, i.e. $\sigma^0_{\text{on}} = 0$ and $\sigma^{n_\sigma+1}_{\text{off}} = T$, we redefine Σ as follows:*

$$\Sigma := \left\{\nu \in \mathbb{R}^{2n_\sigma} \ : \ 0 < \nu_1 < \cdots < \nu_{2n_\sigma} < T\right\}. \tag{2.9}$$

2.3 Reformulation of the state variable

In order to formulate first order necessary optimality conditions we need a constraint qualification which requires in the case of state constraints that $y\left(\bar{t}, \cdot, \sigma\right)$ is element of $C\left([a, b]\right)$ or at least of $L^\infty\left([a, b]\right)$. The problem consists in the fact that the control-to-state-mapping $\sigma \mapsto y\left(\bar{t}, \cdot, \sigma\right) \in L^\infty([a, b])$ is in general not even continuous. According to Theorem 2.1, if $\bar{\sigma} \in \Sigma_{ad}$ satisfies (ND), there exists a neighborhood $U_\Sigma \ni \bar{\sigma}$ such for all $\sigma \in U_\Sigma$ the solution $y(\bar{t}, \cdot, \sigma)|_I$ consists of $K+1$ smooth states $Y_0, \ldots, Y_{K+1}$, which are defined on U_Σ and continuously Fréchet-differentiable w.r.t. σ and are separated by K non-degenerated shock points $x_1, \ldots, x_K$, which are defined on U_Σ and depend continuously Fréchet-differentiable on σ as well. Hence, for $k = 1, \ldots, K+1$, we introduce $(y_1, \ldots, y_{K+1}, x_1, \ldots, x_K)$ as new state variables where

$$y_k\left(\lambda, \sigma\right) := Y_k\left(\bar{t}, x_{k-1}\left(\sigma\right) + \lambda\left(x_k\left(\sigma\right) - x_{k-1}\left(\sigma\right)\right), \sigma\right), \quad \sigma \in U_\Sigma,\ \lambda \in [0, 1] \tag{2.10}$$

Remark 2.1. *Due to the reparametrization in* (2.10)*, we also have to adapt the upper bound* $\bar{y} \in C^1\,[a,b])$ *of the state constraints. To this end, we first note that in terms of the new state variables the state constraints* (1.4) *read*

$$y_k\left(\lambda,\sigma\right) \le \bar{y}\left(\bar{t}, x_{k-1}\left(\sigma\right) + \lambda\left(x_k\left(\sigma\right) - x_{k-1}\left(\sigma\right)\right), \sigma\right), \quad \sigma \in U_\Sigma,\ \lambda \in [0,1] \tag{2.11}$$

Hence, we easily see that the new upper bound is given by the function

$$\bar{y}_k\left(\lambda,\sigma\right) := \bar{y}\left(x_{k-1}\left(\sigma\right) + \lambda\left(x_k\left(\sigma\right) - x_{k-1}\left(\sigma\right)\right), \sigma\right). \tag{2.12}$$

Remark 2.2. *Since the upper bound* $\bar{y}$ *is a smooth function and*

$$y(\bar{t}, x_k(\sigma)-, \sigma) \ge y(\bar{t}, x_k(\sigma)+, \sigma), \quad k = 1, \dots, K \tag{2.13}$$

by (1.6) *and the fact that* $\bar{t} \in (\sigma_{on}^{n_\sigma})$*, it obviously holds for* $k = 1, \dots, K+1$ *:*

$$\begin{aligned} & y(\bar{t}, x, \sigma) \le \bar{y}(x) && \forall x \in [a,b] \\ \iff\ & y_k(\lambda,\sigma) \le \bar{y}_k(\lambda,\sigma) && \forall \lambda \in [0,1]. \end{aligned}$$

Indeed, suppose that $y_k(\lambda,\sigma) \le \bar{y}_k(\lambda,\sigma)$ *holds for all* $\lambda \in [0,1]$ *and all* $k = 1, \dots, K+1$*. Then the inequality* $y(\bar{t}, x, \sigma) \le \bar{y}(x)$ *is obviously satisfied for all* $x \in [a,b]$*. On the other hand, if* $y(\bar{t}, x, \sigma) \le \bar{y}(x)$ *for all* $x \in [a,b]$*, it holds that* $y_k(\lambda,\sigma) \le \bar{y}_k(\lambda,\sigma)$ *for all* $\lambda \in (0,1]$*. Finally, using* (2.13) *and the smoothness of* $\bar{y}$*, one can show for all* $k = 1, \dots, K+1$

$$y_k(0,\sigma) \le y_{k-1}(1,\sigma) \le \bar{y}_{k-1}(1,\sigma) = \bar{y}_k(0,\sigma).$$

Therefore, $y_k(\lambda,\sigma) \le \bar{y}_k(\lambda,\sigma)$ *for all* $\lambda \in [0,1]$ *and* $k = 1, \dots, K+1$*.*

From the results of the previous subsection we can deduce the following regularity result concerning the new state variables:

Theorem 2.3 (Continuous differentiability of the state)**.** *Let Assumption 3 hold and* $\bar{\sigma} \in \Sigma_{ad}$ *satisfy (ND). Then there exists a neighborhood* $U_\Sigma \ni \bar{\sigma}$ *such that the corresponding mappings*

$$\sigma \in U_\Sigma \mapsto x_k(\sigma) \in \mathbb{R}, \quad k = 1, \dots, K$$

and

$$\sigma \in U_\Sigma \mapsto y_k\left(\lambda,\sigma\right) \in C\left([0,1]\right), \quad k = 1, \dots, K+1$$

are continuously Fréchet-differentiable. The derivatives of y_k*, for* $k = 1, \dots, K+1$ *are given by:*

$$\begin{aligned} \frac{\partial}{\partial\sigma} y_k\left(\lambda,\sigma\right)\delta\sigma = {} & \frac{\partial}{\partial x} Y_k\left(\bar{t}, x_{k-1}\left(\sigma\right) + \lambda\left(x_k\left(\sigma\right) - x_{k-1}\left(\sigma\right)\right), \sigma\right) \\ & \cdot \left[\lambda \cdot \frac{\partial}{\partial\sigma} x_k(\sigma) \cdot \delta\sigma + (1-\lambda) \cdot \frac{\partial}{\partial\sigma} x_{k-1}(\sigma) \cdot \delta\sigma\right] \\ & + \frac{\partial}{\partial\sigma} Y_k\left(\bar{t}, x_{k-1}\left(\sigma\right) + \lambda\left(x_k\left(\sigma\right) - x_{k-1}\left(\sigma\right)\right), \sigma\right) \cdot \delta\sigma \end{aligned} \tag{2.14}$$

2.4 First order necessary optimality conditions

In this subsection we will prove first order necessary conditions for (P). To this end, we will first consider the following problem:

$$\min_{w \in W} f(w) \quad \text{subject to} \quad G(w) \in \mathcal{K}, \quad w \in \mathcal{C} \qquad (\mathcal{P})$$

Assumption 4. *We assume that $f : W \to \mathbb{R}$, $G : W \to V$ are continuously Fréchet differentiable with Banach spaces W and V. We further assume that $\mathcal{C} \subset W$ is closed and convex and non-empty, and $\mathcal{K} \subset V$ is a closed convex cone.*

Definition 4. Consider an optimization problem in the form of $(\mathcal{P})$ and let Assumption 4 be satisfied. Then we say that Robinson's CQ is satisisfied in $\bar{w} \in W$, if and only if

$$0 \in int\left(G(\bar{w}) + G'(\bar{w})(\mathcal{C} - \bar{w}) - \mathcal{K}\right)$$

The following result is very well known:

Theorem 2.4. *[Karush-Kuhn-Tucker conditions] Let Assumption 4 be satisfied. Then for any local solution $\bar{w}$ of $(\mathcal{P})$ at which Robinson's CQ is satisfied, there exists a Lagrange multiplier $\bar{q} \in V^*$ (where V^* denotes the dual space of V) with*

$$\begin{aligned} G(w) &\in \mathcal{K} \\ \bar{q} \in \mathcal{K}^\circ := \{q \in V^* : \langle q, v\rangle_{V^*.V} &\le 0 \quad \forall v \in \mathcal{K}\} \\ \langle \bar{q}, G(\bar{w})\rangle_{V^*.V} &= 0, \\ \bar{w} \in \mathcal{W}, \quad \left\langle f'(\bar{w}) + G'(\bar{w})^* \bar{q}, w - \bar{w}\right\rangle_{W^*,W} &\ge 0 \quad \forall w \in \mathcal{C}, \end{aligned}$$

where $\mathcal{K}^\circ$ is the so-called polar cone of $\mathcal{K}$.

Proof. This result is standard and can be found e.g. in [6]. □

Next, we consider a locally optimal solution $\bar{\sigma} \in \Sigma_{ad}$ of (P) in which (ND) is satisfied. Then, according to Theorem 2.1, we can formulate (P) as a problem of the form of $(\mathcal{P})$ by setting

$$W = \mathbb{R}^{2n_\sigma}, \; V = C([0,1])^{K+1} \times \mathbb{R}^K, \tag{2.15}$$

$$G_k(\sigma) = y_k(\lambda, \sigma) - \bar{y}_k(\lambda, \sigma) \quad \forall k \in \{1, \dots, K+1\} \tag{2.16}$$

$$G_{k+K+1}(\sigma) = x_k(\sigma) \quad \forall k \in \{1, \dots, K\} \tag{2.17}$$

$$\mathcal{K} = C([0,1], (-\infty, 0])^{K+1} \times \mathbb{R}^K \tag{2.18}$$

$$\mathcal{C} = \Sigma_{ad} \tag{2.19}$$

Remark 2.3. *The mappings*

$$\sigma \mapsto G_k(\sigma), \quad k \in \{1, \dots, 2K+1\}$$

have been originally defined only on some neighborhood $U_\Sigma \subset [0,T]^{2n_\sigma}$ of $\bar{\sigma}$. In order to be able to apply the results of Theorem 2.4, we choose continuously differentiable extensions, again denoted by G_k, to $W = \mathbb{R}^{2n_\sigma}$ such that

$$\begin{aligned} G_k(\sigma)|_{\Sigma_{ad}} &= (y_k(\lambda, \sigma) - \bar{y}_k(\lambda, \sigma)) && k = 1, \dots, K+1 \\ G_{k+K+1}(\sigma)|_{\Sigma_{ad}} &= x_k(\sigma) && k = 1, \dots, K \end{aligned}$$

Lemma 2.1. *Denoting by $\mathcal{M}([0,1])$ the space of bounded Radon measures on $[0,1]$, the polar cone of $\mathcal{K}$ can be characterized as follows:*

$$q \in \mathcal{K}^{\circ} \Longleftrightarrow q = (\mu_1, \ldots, \mu_{K+1}, 0, \ldots, 0), \tag{2.20}$$

where $\mu_1, \ldots, \mu_{K+1} \in \mathcal{M}([0,1])$ are nonnegative.

We are now able to derive first order necessary optimality conditions for (P):

Theorem 2.5 (Karush-Kuhn-Tucker conditions)**.** *Let Assumption 3 be satisfied. Then for any local solution $\bar{\sigma}$ of (P) at which (ND) and Robinson's CQ are satisfied, there exist nonnegative measures $\mu_1, \ldots, \mu_{K+1} \in \mathcal{M}([0,1])$ such that:*

$$y_k(\lambda, \bar{\sigma}) \leq \bar{y}_k(\lambda, \bar{\sigma}) \quad \forall \lambda \in [0,1], \ \forall k \in \{1, \ldots, K+1\} \tag{2.21}$$

$$\sum_{k=1}^{K+1} \int_{[0,1]} (y_k(\lambda, \bar{\sigma}) - \bar{y}_k(\lambda, \bar{\sigma}))\, d\mu_k(\lambda) = 0 \tag{2.22}$$

$$\bar{\sigma} \in \Sigma_{ad} \tag{2.23}$$

$$\hat{J}'(\bar{\sigma}) \cdot (\sigma - \bar{\sigma}) + \sum_{k=1}^{K+1} \int_{[0,1]} \frac{\partial}{\partial \sigma} (y_k(\lambda, \bar{\sigma}) - \bar{y}_k(\lambda, \bar{\sigma}))(\sigma - \bar{\sigma})\, d\mu_k(\lambda) \geq 0, \quad \forall\, \sigma \in \Sigma_{ad} \tag{2.24}$$

Proof. Considering (2.15)-(2.19), according to Theorem 2.3 the statements in Assumption 4 are obviously satisfied. Hence, Theorem 2.4 and Lemma 2.1 yield the statement of the above theorem. □

Corollary 2.1. *Let $\bar{\sigma} \in \Sigma_{ad}$ be a switching time vector such that the KKT conditions are satisfied, i.e. there exist nonnegative measures $\mu_1, \ldots, \mu_{K+1} \in \mathcal{M}([0,1])$ such that (2.21), (2.22), (2.23) and (2.24) are satisfied. Then for all measurable $A, B \subset [0,1]$ it holds that*

$$\bar{y}_k(\lambda, \bar{\sigma}) - y_k(\lambda, \bar{\sigma}) > 0 \quad \forall \lambda \in A \implies \mu_k(A) = 0 \tag{2.25}$$

Proof. The first assertion (2.25) follows from (2.21) and (2.22). Indeed, suppose that

$$\bar{y}_k(\lambda, \bar{\sigma}) - y_k(\lambda, \bar{\sigma}) > 0 \quad \forall \lambda \in A$$

for some $k \in \{1, \ldots, K+1\}$ and assume that

$$\mu_k(A) < 0.$$

Then, we obtain by (2.21):

$$\begin{aligned}
&\sum_{j=1}^{K+1} \int_{[0,1]} (\bar{y}_j(\lambda, \bar{\sigma}) - y_j(\lambda, \bar{\sigma}))\, d\mu_j(\lambda) \\
&\leq \int_{[0,1]} (\bar{y}_k(\lambda, \bar{\sigma}) - y_k(\lambda, \bar{\sigma}))\, d\mu_k(\lambda) < 0
\end{aligned}$$

This is a contradiction to (2.22) and (2.25) holds. □

The following result will be crucial in the next chapter:

Lemma 2.2. *Let Assumption 1 hold and consider some $\bar{\sigma}$ in which (ND) and Robinson's CQ are satisfied. Then there exists $\tilde{\sigma} \in \Sigma_{ad}$ such that for some $\varepsilon > 0$*

$$\left(y_k(\lambda, \bar{\sigma}) - \bar{y}_k(\lambda, \bar{\sigma}) - \frac{\partial}{\partial \sigma} \left(y_k(\lambda, \bar{\sigma}) - \bar{y}_k(\lambda, \bar{\sigma}) \right) \cdot (\tilde{\sigma} - \bar{\sigma}) \right) \leq -\varepsilon, \tag{2.26}$$

for all $\lambda \in [0,1]$ and $k = 1, \ldots, K+1$.

Proof. We note that Robinson's CQ implies

$$\begin{aligned} 0 \in \operatorname{int} & \left(y(\lambda, \bar{\sigma}) - \bar{y}(\lambda, \bar{\sigma}) - \frac{\partial}{\partial \sigma} \left(y(\lambda, \bar{\sigma}) - \bar{y}(\lambda, \bar{\sigma}) \right) \cdot (\Sigma_{ad} - \bar{\sigma}) - \tilde{K} \right) \\ & \subseteq C([0,1])^{K+1} \end{aligned} \tag{2.27}$$

where $\tilde{K} := C([0,1], (-\infty, 0])^{K+1}$. From this we can directly deduce (2.26). □

Our aim is not to reformulate (2.21), (2.22) and (2.24) w.r.t. to the original state. As a first step, we will rewrite (2.21), (2.22) and (2.24) w.r.t. the mappings $Y_1, \ldots, Y_{K+1}$ which were introduced in Theorem 2.1. Recalling that we have introduced the new state variables in (2.10) by using the variable transformations

$$\begin{aligned} &\varphi_{k,\sigma} : [0,1] \to \left[x_{k-1}(\sigma), x_k(\sigma) \right] \\ &\lambda \mapsto x_{k-1}(\sigma) + \lambda \big(x_k(\sigma) - x_{k-1}(\lambda) \big), \end{aligned} \tag{2.28}$$

which is bijective where the inverse is given by

$$\begin{aligned} &\varphi_{k,\sigma}^{-1} : \left[x_{k-1}(\sigma), x_k(\sigma) \right] \to [0,1] \\ &x \mapsto \frac{x - x_{k-1}(\bar{\sigma})}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})}, \end{aligned} \tag{2.29}$$

Using (2.28) and (2.14), the optimality conditions in (2.21)-(2.24) can be written as follows

$$Y_k \left(\bar{t}, \varphi_{k,\bar{\sigma}}(\lambda), \bar{\sigma} \right) \leq \bar{y} \left(\varphi_{k,\bar{\sigma}}(\lambda) \right) \quad \forall \lambda \in [0,1], \ \forall k \in \{1, \ldots, K+1\}$$

$$\sum_{k=1}^{K+1} \int_{[0,1]} \big(Y_k \left(\bar{t}, \varphi_{k,\bar{\sigma}}(\lambda), \bar{\sigma} \right) - \bar{y} \left(\varphi_{k,\bar{\sigma}}(\lambda) \right) \big) d\mu_k (\lambda) = 0$$

$$\bar{\sigma} \in \Sigma_{ad}$$

$$\begin{aligned} &\hat{J}' (\bar{\sigma}) \cdot (\sigma - \bar{\sigma}) + \sum_{k=1}^{K+1} \int_{[0,1]} \left[\frac{\partial}{\partial x} \big(Y_k \left(\bar{t}, \varphi_{k,\bar{\sigma}}(\lambda), \bar{\sigma} \right) - \bar{y}(\varphi_{k,\bar{\sigma}}(\lambda)) \big) \right. \\ &\cdot \left(\varphi_{k,\sigma}^{-1}(\varphi_{k,\sigma}(\lambda)) \cdot \frac{\partial}{\partial \sigma} x_k(\bar{\sigma}) \cdot (\sigma - \bar{\sigma}) + \big(1 - \varphi_{k,\sigma}^{-1}(\varphi_{k,\sigma}(\lambda)) \big) \right. \\ &\left. \left. \cdot \frac{\partial}{\partial \sigma} x_{k-1}(\sigma) \cdot (\sigma - \bar{\sigma}) \right) + \frac{\partial}{\partial \sigma} \big(Y_k \left(\bar{t}, \varphi_{k,\bar{\sigma}}(\lambda), \bar{\sigma} \right) - \bar{y}(\varphi_{k,\bar{\sigma}}(\lambda)) \big) \cdot (\sigma - \bar{\sigma}) \right] d\mu_k (\lambda) \\ &\geq 0, \quad \forall \, \sigma \in \Sigma_{ad} \end{aligned}$$

Using [5, V, §3, (3.1)], one can show that there exist nonnegative measures $\bar{\mu}_k \in \mathcal{M}(I_k)$, where $I_k := [x_{k-1}(\bar{\sigma}), x_k(\bar{\sigma})]$, which are given by

$$\bar{\mu}_k(A) := \mu_k(\varphi_{k,\bar{\sigma}}^{-1}(A)) \tag{2.30}$$

for all measurable $A \subset I_k$ and all $k = 1, \ldots, K+1$ such that the following holds true

$$Y_k(\bar{t}, x, \bar{\sigma}) \leq \bar{y}(x) \quad \forall x \in I_k, \quad \forall k \in \{1, \ldots, K+1\} \tag{2.31}$$

$$\sum_{k=1}^{K+1} \int_{I_k} (Y_k(\bar{t}, x, \bar{\sigma})) - \bar{y}(x)\, d\bar{\mu}_k(x) = 0 \tag{2.32}$$

$$\bar{\sigma} \in \Sigma_{ad} \tag{2.33}$$

$$\begin{aligned}
\hat{J}'(\bar{\sigma}) \cdot (\sigma - \bar{\sigma}) + \sum_{k=1}^{K+1} \Bigg[\int_{I_k} & \frac{\partial}{\partial x} [Y_k(\bar{t}, x, \bar{\sigma}) - \bar{y}(x)] \frac{x - x_{k-1}(\bar{\sigma})}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \, d\bar{\mu}_k(x) \\
& \cdot \frac{\partial}{\partial \sigma} x_k(\bar{\sigma})(\sigma - \bar{\sigma}) \\
+ \int_{I_k} & \frac{\partial}{\partial x} [Y_k(\bar{t}, x, \bar{\sigma}) - \bar{y}(x)] \frac{x_k(\bar{\sigma}) - x}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \, d\bar{\mu}_k(x) \\
& \cdot \frac{\partial}{\partial \sigma} x_{k-1}(\bar{\sigma})(\sigma - \bar{\sigma}) \\
+ \int_{I_k} & \frac{\partial}{\partial \sigma} Y_k(\bar{t}, x, \bar{\sigma})(\sigma - \bar{\sigma}) \, d\bar{\mu}_k(x) \Bigg] \geq 0 \quad, \forall \sigma \in \Sigma_{ad}
\end{aligned} \tag{2.34}$$

In connection with Theorem 2.5, we then obtain the following result:

Theorem 2.6. *Let Assumption 3 hold and consider some local solution $\bar{\sigma}$ of (P) at which (ND) and Robinson's CQ are satisfied. Then there exist $K+1$ nonnegative measures $\bar{\mu}_k \in \mathcal{M}(I_k)$, where $I_k := [x_{k-1}(\bar{\sigma}), x_k(\bar{\sigma})]$, such that* (2.31)-(2.34) *hold true.*

Lemma 2.3. *Let Assumption 3 hold and $\bar{\sigma} \in \Sigma_{ad}$ be a switching time vector such that the KKT conditions are satisfied, i.e. there exist nonnegative measures $\mu_k \in \mathcal{M}(I_k)$, $k = 1, \ldots, K+1$ such that* (2.31), (2.32), (2.33) *and* (2.34) *are satisfied. Then we obtain for all measurable $A \subseteq (x_{k-1}(\bar{\sigma}), x_k(\bar{\sigma}))$:*

$$\int_A \frac{\partial}{\partial x} [\bar{y}(x) - Y_k(\bar{t}, x, \bar{\sigma})] \left(\frac{x - x_{k-1}(\bar{\sigma})}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \right) d\bar{\mu}_k(x) = 0 \tag{2.35}$$

$$\int_A \frac{\partial}{\partial x} [\bar{y}(x) - Y_k(\bar{t}, x, \bar{\sigma})] \left(\frac{x_k(\bar{\sigma}) - x}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \right) d\bar{\mu}_k(x) = 0 \tag{2.36}$$

Proof. For arbitrary $k \in \{1, \ldots, K+1\}$ and measurable $A \subseteq (x_{k-1}(\bar{\sigma}), x_k(\bar{\sigma}))$ we set

$$A_1 := \{x \in A : Y_k(\bar{t}, x, \bar{\sigma}) < \bar{y}(x)\}$$
$$A_2 := \{x \in A : Y_k(\bar{t}, x, \bar{\sigma}) = \bar{y}(x)\}$$

and observe that A_1 and A_2 are both measurable due to the regularity of $Y_k(\bar{t}, \cdot, \bar{\sigma})$ and $\bar{y}$. Furthermore, from (2.31) we deduce that $A = A_1 \cup A_2$ and observe that

$$\bar{\mu}_k(A_1) = 0 \tag{2.37}$$

$$\frac{\partial}{\partial x}\left[\bar{y}(x) - Y_k(\bar{t}, x, \bar{\sigma})\right]\Bigg|_{A_2} \equiv 0 \tag{2.38}$$

Using (2.31), (2.32), the first assertion can be shown as in the proof of Corollary 2.1. The second assertion follows from the continuity of $\frac{\partial}{\partial x}\left[\bar{y}(x) - Y_k(\bar{t}, x, \bar{\sigma})\right]$ and (2.31). Indeed, assume that there is some $\tilde{x} \in A_2$ with $\frac{\partial}{\partial x}\left[Y_k(\bar{t}, \tilde{x}, \bar{\sigma}) - \bar{y}(\tilde{x})\right] \neq 0$. We can w.l.o.g. assume that it holds that $\frac{\partial}{\partial x}\left[\bar{y}(x) - Y_k(\bar{t}, x, \bar{\sigma})\right] > 0$. From the continuity of $\frac{\partial}{\partial x}\left[\bar{y}(x) - Y_k(\bar{t}, x, \bar{\sigma})\right]$, one can deduce that there exists $\varepsilon > 0$ such that $[\tilde{x}, \tilde{x} + \varepsilon] \subset I_k$ and $Y_k(\bar{t}, x, \bar{\sigma}) - \bar{y}(x) > 0$ for all $x \in [\tilde{x}, \tilde{x} + \varepsilon]$. Hence (2.31) is violated on I_k which is obviously a contradiction. Therefore (2.38) holds. Recalling $A = A_1 \cup A_2$, using (2.37) -(2.38) and the fact that A was arbitrary chosen, we obtain for all measurable $A \subseteq (x_{k-1}(\bar{\sigma}), x_k(\bar{\sigma}))$:

$$\begin{aligned}
&\int_A \frac{\partial}{\partial x}\left[\bar{y}(x) - Y_k(\bar{t}, x, \bar{\sigma})\right]\left(\frac{x - x_{k-1}(\bar{\sigma})}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})}\right) d\bar{\mu}_k(x) \\
=& \int_{A_1} \frac{\partial}{\partial x}\left[\bar{y}(x) - Y_k(\bar{t}, x, \bar{\sigma})\right]\left(\frac{x_k(\bar{\sigma}) - x}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})}\right) d\bar{\mu}_k(x) \\
+& \int_{A_2} \frac{\partial}{\partial x}\left[\bar{y}(x) - Y_k(\bar{t}, x, \bar{\sigma})\right]\left(\frac{x_k(\bar{\sigma}) - x}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})}\right) d\bar{\mu}_k(x) \\
=& \, 0
\end{aligned}$$

Analogously, we can prove (2.36). □

Using this result, we can further simplify the optimality conditions in Theorem 2.6:

Corollary 2.2. *Let Assumption 1 hold and let $\bar{\sigma} \in \Sigma_{ad}$ be a switching time vector which satisfies (ND). We further assume that the KKT conditions are satisfied in $\bar{\sigma}$, i.e. there exist nonpositive measures $\mu_1, \ldots, \mu_{K+1} \in \mathcal{M}(\mathbb{R})$ such that* (2.31)*,* (2.32)*,* (2.33) *and* (2.34) *are satisfied. Then* (2.34) *can also be written as:*

$$\begin{aligned}
&\hat{J}'(\bar{\sigma}) \cdot (\sigma - \bar{\sigma}) \\
&+ \sum_{k=1}^{K+1} \left[\frac{\partial}{\partial x}\left[Y_k(\bar{t}, x_k(\bar{\sigma}), \bar{\sigma}) - \bar{y}(x_k(\bar{\sigma}))\right] \cdot \bar{\mu}_k(\{x_k(\bar{\sigma})\}) \cdot \frac{\partial}{\partial \sigma} x_k(\bar{\sigma}) \cdot (\sigma - \bar{\sigma}) \right. \\
&+ \frac{\partial}{\partial x}\left[Y_k(\bar{t}, x_{k-1}(\bar{\sigma}), \bar{\sigma}) - \bar{y}(x_{k-1}(\bar{\sigma}))\right] \cdot \bar{\mu}_k(\{x_{k-1}(\bar{\sigma})\}) \cdot \frac{\partial}{\partial \sigma} x_{k-1}(\bar{\sigma}) \cdot (\sigma - \bar{\sigma}) \\
&+ \left. \int_{I_k} \frac{\partial}{\partial \sigma} Y_k(\bar{t}, x, \bar{\sigma}) \cdot (\sigma - \bar{\sigma}) d\bar{\mu}_k(x) \right] \geq 0
\end{aligned} \tag{2.39}$$

As we will see in the next theorem, we can formulate the optimality conditions from Theorem 2.6 also in terms of the original state y:

Theorem 2.7. *Let Assumption 3 hold and consider some local solution $\bar{\sigma}$ of (P) at which (ND) and Robinson's CQ are satisfied. Then there exists a nonpositive measure $\mu \in \mathcal{M}([a,b])$ such that (2.31)-(2.34) and (2.39) are still valid, if we replace $Y_k(\bar{t},x,\bar{\sigma})$, $\frac{\partial}{\partial\sigma}Y_k(\bar{t},x,\bar{\sigma})$ and $\frac{\partial}{\partial x}Y_k(\bar{t},x,\bar{\sigma})$ by $y(\bar{t},x,\bar{\sigma})$, $\frac{\partial}{\partial\sigma}Y_k(\bar{t},x-,\bar{\sigma})$ and $\frac{\partial}{\partial x}Y_k(\bar{t},x-,\bar{\sigma})$, respectively.*

We end this chapter by stating a result that will play a key role in the following chapter:

Lemma 2.4. *Let Assumption 3 hold and consider some $\bar{\sigma}$ in which (ND) and Robinson's CQ are satisfied. Then there exists $\varepsilon > 0$ and $\tilde{\sigma} \in \Sigma_{ad}$ such that for all $x \in (x_{k-1}(\bar{\sigma}), x_k(\bar{\sigma}))$, $k = 1, \ldots, K+1$:*

$$\begin{aligned}\Bigg(&\bar{y}(x) - Y_k(\bar{t},x,\bar{\sigma}) - \frac{\partial}{\partial\sigma}Y_k(\bar{t},x,\bar{\sigma}) \cdot (\tilde{\sigma} - \bar{\sigma}) \\ &+ \frac{\partial}{\partial x}\left[\bar{y}(x) - Y_j(\bar{t},x,\bar{\sigma})\right]\left(\frac{x - x_{k-1}(\bar{\sigma})}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \cdot \frac{\partial}{\partial\sigma}x_k(\bar{\sigma})\,(\tilde{\sigma} - \bar{\sigma})\right. \\ &\left.+ \frac{x_k(\bar{\sigma}) - x}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \cdot \frac{\partial}{\partial\sigma}x_{k-1}(\bar{\sigma})\,(\tilde{\sigma} - \bar{\sigma})\right)\Bigg) \geq \varepsilon.\end{aligned} \tag{2.40}$$

Proof. This result follows directly by Lemma 2.2. $\square$

3 Moreau-Yosida Regularization

Since it is quite involved to compute a solution of the optimality system in Theorem 2.6, we will omit the state constraints and take them into account by adding a penalty function $P(y(\sigma))$ to the cost functional, which we multiply with a penalty parameter $\frac{1}{\gamma}$, where $\gamma > 0$:

$$\min_{\sigma\in\Sigma} J_\gamma(y(\sigma)) := J(y(\sigma)) + \frac{1}{2\gamma}\int_a^b \left(y(\bar{t},x,\sigma) - \bar{y}(x)\right)_+^2\, dx \qquad (P_\gamma)$$

$$\text{where} \quad y(\sigma) \text{ solves the (OOSP)}$$
$$\sigma \in \Sigma_{ad}$$

This approach is called Moreau-Yosida regularization, see for example [11], [10]. There are also other approaches, e.g. the virtual control method [8], [7].

We will work under the following assumption:

Assumption 5. *Let Assumption 3 hold and Σ_{ad} be a closed, convex and nonempty subset of*

$$\bar{\Sigma} := \{\sigma \in [\bar{\varepsilon}, \bar{t} - \bar{\varepsilon}]^{n_\sigma} \mid \quad \sigma^i_{\text{off}} + \bar{\varepsilon} \leq \sigma^i_{\text{on}} \quad \forall i = 1,\dots,n_\sigma\}$$

Furthermore, let the observation time $\bar{t} \in (0,T)$, the interval $I := [a,b]$, $a < b$, $u_I \in \mathrm{PC}^1(\mathbb{R}; x_1,\dots,x_{n_x})$ and Σ_{ad} be chosen such that for all $\sigma \in \Sigma_{ad}$ the off-switching times are nondegenerated, a and b are continuity points of $y(\bar{t},\cdot,\sigma)$, $\bar{t} \in (\sigma^{n_\sigma}_{on}, T)$ and $y(\bar{t},\cdot,\sigma)$ has on I no shock generation points and exactly N nondegenerated shock points $x_1,\dots,x_N$. In addition, we assume that those shock points are uniformly (w.r.t. $\sigma \in \Sigma_{ad}$) bounded away from a, b and from each other. We further assume that there exists a switching time vector $\sigma \in \Sigma_{ad}$ for which the corresponding state y satisfies the state constraints. Finally, let $y_d \in C^1(\mathbb{R})$.

Theorem 3.1. *Let Assumption 5 hold. Then there exist continuously Fréchet-differentiable mappings*

$$Y_k : (x,\sigma) \in (x_{k-1}(\bar{\sigma}) - \varepsilon, x_k(\bar{\sigma}) + \varepsilon) \times \Sigma_{ad} \mapsto Y_k(\bar{t},x,\sigma) \in \mathbb{R} \qquad (3.1)$$

$$x_k : \sigma \in \Sigma_{ad} \mapsto x_k(\sigma) \in [a,b] \qquad (3.2)$$

for all $k = 1,\dots,K+1$, where $x_0 = a$ and $x_{K+1} = b$, such that for all $\sigma \in \Sigma_{ad}$ the entropy solution $y(\bar{t},\cdot,\sigma)$ of the On/Off Switching Problem is on $[a,b]$ given by:

$$y(\bar{t},x,\sigma) = Y_1(\bar{t},x,\sigma)\cdot 1_{[a,x_1(\sigma)]}(x) + \sum_{k=2}^{K+1} Y_k(\bar{t},x,\sigma)\cdot 1_{(x_{k-1}(\sigma),x_k(\sigma)]}(x) \qquad (3.3)$$

Furthermore,

$$\sigma \in \Sigma_{ad} \mapsto Y_k(\cdot,\sigma) \in C(I_k) \qquad (3.4)$$

are continuously Fréchet differentiable, where $I_k := (x_{k-1}(\bar{\sigma}) - \varepsilon, x_k(\bar{\sigma}) + \varepsilon)$.

Proof. This proof can be done analogously to the proof of part (b) of Lemma 2.1. □

Theorem 3.2. *Let Assumption 5 hold. Then the cost functional of the regularized problem is Fréchet-differentiable and Lipschitz continuous w.r.t. σ. The derivative in a direction $\delta\sigma$ is given by:*

$$\frac{\partial}{\partial\sigma} J_\gamma(y(\sigma)) \cdot \delta\sigma = \sum_{i=1}^{n_\sigma} \left(p^i_{\mathrm{off}} \cdot \delta\sigma^i_{\mathrm{off}} + p^i_{\mathrm{on}} \cdot \delta\sigma^i_{\mathrm{on}} \right) \tag{3.5}$$

where

$$\begin{aligned} p^i_{\mathrm{on}} &:= \int_{f'(-1)}^{f'(0)} \lim_{s \searrow \sigma^i_{\mathrm{on}}} p(s, (s - \sigma^i_{\mathrm{on}}) \cdot w) \cdot w \cdot \left({f'}^{-1} \right)' (w) \ \mathrm{d}w, \\ p^i_{\mathrm{off}} &:= \left(p(\sigma^i_{\mathrm{off}}, 0+) - p(\sigma^i_{\mathrm{off}}, 0-) \right) \cdot f(y(\sigma^i_{\mathrm{on}}, 0)) \end{aligned}$$

and p is the reversible solution of the adjoint equation

$$p_t + f'(y) p_x = -g_y(\cdot, \hat{y}) p, \quad p\,(\bar{t}, \cdot) = p^{\bar{t}}, \tag{3.6}$$

$$p\,(\bar{t}, \cdot) = 1_{]a,b[} \bar{\psi}_y(\cdot) \tag{3.7}$$

where

$$\bar{\psi}_y(x) := \int_0^1 \left[\psi_y \left(y(\bar{t}, x+; \sigma) \right) + \tau[y(\bar{t}, x; \sigma)], y_d(x) \right) \tag{3.8}$$

$$+ \frac{1}{\gamma} \left(y(\bar{t}, x+, \sigma) + \tau[y(\bar{t}, x; \sigma)] - \bar{y}(x) \right)_+ \bigg] \ \mathrm{d}\tau. \tag{3.9}$$

Proof. The proof is similar to the proof of Theorem 2.2. ☐

Theorem 3.3. *Let Assumption 5 hold. Then for each penalty parameter γ there exists a global optimal solution σ_γ for P_γ.*

Proof. Firstly, we observe that the cost functionals of (P_γ) and (P) are both of the same structure. Since Σ_{ad} is compact, we can use the same techniques as in the proof of Theorem 1.2. ☐

Theorem 3.4. *Let Assumption 5 hold and let $(\sigma_{\gamma_k})_{k\in\mathbb{N}} \subset \Sigma_{ad}$ denote a sequence of global solutions of (P_γ) with $\lim_{k\to\infty} \gamma_k = 0$. Then there exists a subsequence, again denoted by $(\sigma_{\gamma_k})_{k\in\mathbb{N}}$, that converges strongly to a global optimal solution σ^* of (P).*

Proof. We consider a sequence of global optima of (P_{γ_k}) and prove that there exists a subsequence, that converges to a global optimal solution for (P). Since the set Σ_{ad} is compact, there exists a convergent subsequence, again denoted by $(\sigma_{\gamma_k})_{k\in\mathbb{N}}$, such that:

$$\sigma_{\gamma_k} \to \sigma^* \in \Sigma_{ad} \tag{3.10}$$

We know from that the state constrained problem (P) possesses an optimal control $\bar{\sigma}$. In the next step, we prove that $y(\sigma^*)$ fulfills the state constraints: Firstly, we note that for all $k \in \mathbb{N}$, it holds that

$$J_\gamma(y(\sigma_k)) \leq J(y(\bar{\sigma})), \tag{3.11}$$

where $\bar{\sigma}$ is a global optimum of (P), which exists by Theorem 1.2. From (3.10), (3.11) and the continuity of $J(y(\cdot))$ w.r.t. σ, we can deduce that there exists a constant $C > 0$ with

$$0 \leq \frac{1}{2\gamma} \int_a^b \left(y(\bar{t}, x, \sigma_{\gamma_k}) - \bar{y}(x)\right)_+^2 \, dx \leq J(y(\bar{\sigma})) - J(y(\sigma_{\gamma_k})) \leq C$$

$$\iff 0 \leq \int_a^b \left(y(\bar{t}, x, \sigma_{\gamma_k}) - \bar{y}(x)\right)_+^2 \, dx \leq 2 \cdot \gamma_k \cdot C \tag{3.12}$$

Using (3.12), will prove that

$$y(\bar{t}, x, \sigma^*) \leq \bar{y}(x) \quad \forall x \in [a, b]$$

We will do this via contradiction: We assume that there exists $\tilde{x} \in [a, b]$ and $\varepsilon > 0$ with

$$y(\bar{t}, \tilde{x}, \sigma^*) - \bar{y}(\tilde{x}) > \varepsilon$$

1. Case: $\tilde{x} \neq x_k(\sigma^*) \quad \forall j \in \{1, \ldots, K\}$
It holds that $\tilde{x} \in (x_{j-1}(\sigma^*), x_j(\sigma^*))$ for some $j \in \{1, \ldots, K+1\}$, or $\tilde{x} \in \{a, b\}$. If $\tilde{x} \in (x_{j-1}(\sigma^*), x_j(\sigma^*))$, there exists $\delta > 0$ with $x_{j-1}(\sigma^*) + \delta \leq \tilde{x} \leq x_j(\sigma^*) - \delta$. Since $x_j(\sigma_{\gamma_k}) \to x_j(\sigma^*)$ for all $j \in \{1, \ldots, K\}$, there exists $\bar{k} \in \mathbb{N}$ such that

$$y(\bar{t}, \tilde{x}, \sigma_{\gamma_k}) = Y_j(\bar{t}, \tilde{x}, \sigma_{\gamma_k}) \quad \forall k > \bar{k}$$

Exploiting the regularity of Y_j, we can conclude that there exists $\varepsilon_1 > 0$ with

$$Y_j(\bar{t}, x, \sigma_k) - \bar{y}(x) > \frac{\varepsilon}{2} \quad \forall k > \bar{k}, \forall x \in B_{\varepsilon_1}(\tilde{x})$$

Using (3.12), we therefore obtain

$$0 < \frac{1}{2}\varepsilon^2 \cdot \varepsilon_1 \leq \int_a^b \left(y(\bar{t}, x, \sigma_{\gamma_k}) - \bar{y}(x)\right)_+^2 \, dx \leq 2 \cdot \gamma_k \cdot C \tag{3.13}$$

Since $\gamma_k \to 0$ for $k \to \infty$, this is obviously a contradiction. If $\tilde{x} \in \{a, b\}$, using the fact that a and b are continuity points of $y(\bar{t}, \cdot, \sigma)$ for all $\sigma \in \Sigma_{ad}$, one can analogously deduce a contradiction.
2. Case: $\tilde{x} = x_j(\sigma^*)$ for some $j \in \{1, \ldots, K\}$
Since we know that

$$y(\bar{t}, \tilde{x}-, \sigma^*) = Y_j(\bar{t}, \tilde{x}, \sigma^*)$$

and

$$y(\bar{t}, \tilde{x}, \sigma^*) = Y_j(\bar{t}, \tilde{x}, \sigma^*)$$

by Remark 1.2, we can again exploit the regularity of Y_j: There exists $\varepsilon_2 > 0$ with

$$Y_j(\bar{t}, x, \sigma_{\gamma_k}) - \bar{y}(x) > \frac{\varepsilon}{2} \quad \forall k > \bar{k},\, \forall x \in \left[\tilde{x} - \varepsilon_2, \tilde{x} - \frac{\varepsilon_2}{2}\right]$$

Using (3.12), we therefore obtain

$$0 < \frac{1}{4}\varepsilon^2 \cdot \varepsilon_2 \leq \int_a^b \left(y(\bar{t}, x, \sigma_{\gamma_k}) - \bar{y}(x)\right)_+^2 \, dx \leq 2 \cdot \gamma_k \cdot C \tag{3.14}$$

Since $\gamma_k \to 0$ for $k \to \infty$, this is obviously a contradiction again. Therefore we have proven, that $y(\bar{t}, \cdot, \sigma^*)$ fulfills the state constraints.

Next, since we know that

$$J_{\gamma_k}(y(\sigma_{\gamma_k})) \leq J_{\gamma_k}(y(\bar{\sigma})) \quad \forall k \in \mathbb{N},$$

it holds that

$$J(y(\sigma^*)) \leq J(y(\bar{\sigma}))$$

As $\bar{\sigma}$ is a global optimal solution, it even holds that

$$J(y(\sigma^*)) = J(y(\bar{\sigma}))$$

and σ^* is hence a global optimal solution for (P). □

Since optimization algorithms generally are only able to find local optima of (P_γ), in the next theorem we will examine the convergence of local solutions of (P_γ) to local solutions of (P).

Theorem 3.5. *Let Assumption 5 hold and let $\bar{\sigma} \in \Sigma_{ad}$ be a local optimum of (P) in which the following condition is satisfied: It exists positive constants ε and δ such the for all $\sigma \in \Sigma_{ad}$ with $\|\sigma - \bar{\sigma}\|_2 < \varepsilon$ it holds that*

$$J(y(\bar{\sigma})) + \frac{\delta}{2}\|\sigma - \bar{\sigma}\|_2^2 \leq J(y(\sigma)) \tag{QGC}$$

Then there exists a sequence $(\sigma_{\gamma_k})_k \in \mathbb{N}$ of local solutions of (P_γ) that converges strongly to $\bar{\sigma}$.

Proof. See proof of Theorem 5.2 in [10]. □

Theorem 3.6. *Let Assumption 5 hold and consider a sequence $(\sigma_{\gamma_k})_{k\in\mathbb{N}}$ of local solutions of (P_{γ_k}) such that for all $k \in \mathbb{N}$ the following condition is satisfied: There exist positive constants ε and δ such the for all $\sigma \in \Sigma_{ad}$ with $\|\sigma - \sigma_{\gamma_k}\|_2 < \varepsilon$ it holds that*

$$J_{\gamma_k}(y(\sigma_{\gamma_k})) + \frac{\delta}{2}\|\sigma - \sigma_{\gamma_k}\|_2^2 \leq J_{\gamma_k}(y(\sigma)) \tag{QGC_{γ_k}}$$

Then there exists a subsequence again denoted by $(\sigma_{\gamma_k})_k \in \mathbb{N}$ that converges to some $\bar{\sigma} \in \Sigma_{ad}$, which is a local solution of (P).

Proof. Since Σ_{ad} is by assumption compact, we know that there is a convergent subsequence $(\sigma_{\gamma_k})_k \in \mathbb{N}$ with limit $\bar{\sigma} \in \Sigma_{ad}$. As in [3], we define the auxiliary problem

$$\begin{aligned} &\min_{\sigma\in\Sigma} J(y(\sigma)) && (P^r)\\ &\text{where} \quad y(\sigma) \text{ solves the (OOSP)} \\ &\qquad\quad\;\; y(\bar{t}, x, \sigma) \leq \bar{y}(x) \quad \forall x \in [a, b] \\ &\qquad\quad\;\; \sigma \in \Sigma_{ad}^r := \Sigma_{ad} \cap B_r(\bar{\sigma}) \end{aligned}$$

and

$$\begin{aligned}
&\min_{\sigma\in\Sigma} J_\gamma(y(\sigma)) && (P^r_\gamma)\\
&\text{where} \quad y(\sigma) \text{ solves the (OOSP)}\\
&\qquad\qquad \sigma \in \Sigma^r_{ad}
\end{aligned}$$

Analogous to the proof of Theorem 1.2, one can show that (P^r) possesses a global optimal solution for all $r > 0$. We choose $r = \frac{\varepsilon}{2}$ and denote the corresponding global optimal solution by $\bar{\sigma}$. Furthermore, (QGC_{γ_k}) yields that σ_{γ_k} is a unique global optimal solution of $(P^{\frac{\varepsilon}{2}}_{\gamma_k})$ for all k large enough. Theorem 1.2 finally yields that $\bar{\sigma} \in \Sigma_{ad}$ is a global optimal solution of $(P^{\frac{\varepsilon}{2}})$, and hence a local optimal solution of (P). □

Theorem 3.7. *Let Assumption 5 hold and consider a sequence of penalty parameter γ_k with $\lim_{k\to\infty} \gamma_k = 0$ and $\lim_{k\to\infty} \sigma_{\gamma_k} = \sigma^*$. Then it holds that*

$$\lim_{k\to\infty} y(\bar{t}, \cdot, \sigma_{\gamma_k}) = y(\bar{t}, \cdot, \sigma^*) \qquad \text{in } L^1([a, b])$$

and furthermore

$$\lim_{k\to\infty} Y_j(\bar{t}, \cdot, \sigma_{\gamma_k}) = Y_j(\bar{t}, \cdot, \sigma^*) \qquad \text{in } C^1\left([x_{j-1}(\sigma^*), x_j(\sigma^*)]\right),$$

for all $j = 1, \dots, K + 1$.

Proof. The first assertion is a simple consequence of Theorem 1.1, and the second one can be deduced by using Theorem 3.1. □

Theorem 3.8 (Necessary Optimality Conditions for P_γ)**.** *Let Assumption 5 hold, $\gamma > 0$ be arbitrary and let $\sigma_\gamma \in \Sigma_{ad}$ be an optimal solution for P_γ. Then the following holds:*

$$\frac{\partial}{\partial\sigma} J_\gamma(y(\sigma_\gamma)) \cdot (\sigma - \sigma_\gamma) \geq 0 \quad \forall \sigma \in \Sigma_{ad} \tag{3.15}$$

Proof. This result is well known in the literature and can be found for example in [6]. □

We notice that

$$\begin{aligned}
J_\gamma(y(\sigma)) =& J(y(\sigma)) + \frac{1}{2\gamma} \int_a^b \left(y(\bar{t}, x, \sigma) - \bar{y}(x)\right)_+^2 \, dx\\
=& J(y(\sigma)) + \sum_{j=1}^{K+1} z_j(\sigma) \cdot \frac{1}{2\gamma z_j(\sigma)} \int_{x_{j-1}(\sigma)}^{x_j(\sigma)} \left(Y_j(\bar{t}, x, \sigma) - \bar{y}(x)\right)_+^2 \, dx
\end{aligned}$$

where $z_j(\sigma) := (x_j(\sigma) - x_{j-1}(\sigma))$. Using this and the abbreviation $G_j(\bar{t}, x, \sigma) := (Y_j(\bar{t}, x, \bar{\sigma}) - \bar{y}(x))$, one can rewrite the derivative with respect to σ in σ_γ in the direction $(\sigma - \sigma_\gamma) := \delta\sigma$ in the following way:

$$
\begin{aligned}
&\frac{\partial}{\partial\sigma} J_\gamma(y(\sigma_\gamma)) \cdot \delta\sigma \\
&= \frac{\partial}{\partial\sigma} J(y(\sigma_\gamma)) \cdot (\sigma - \sigma_\gamma) \\
&+ \sum_{j=1}^{K+1} z_j(\sigma) \cdot \left[\frac{1}{\gamma z_j(\sigma)} \int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)} \frac{\partial}{\partial\sigma} G(\bar{t}, x, \sigma_\gamma)\delta\sigma \cdot (G_j(\bar{t}, x, \sigma_\gamma))_+ \, dx \right. \\
&+ \frac{-1}{2\gamma z_j(\sigma_\gamma)^2} \int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)} (G_j(\bar{t}, x, \sigma_\gamma))_+^2 \, dx \cdot \frac{\partial}{\partial\sigma} z_j(\sigma_\gamma) \cdot \delta\sigma \\
&+ \frac{1}{2\gamma z_j(\sigma_\gamma)} (G_j(\bar{t}, x_j(\sigma_\gamma), \sigma_\gamma))_+^2 \cdot \frac{\partial}{\partial\sigma} x_j(\sigma_\gamma) \cdot \delta\sigma \\
&\left. - \frac{1}{2\gamma z_j(\sigma_\gamma)} \left. (G_j(\bar{t}, x_{j-1}(\sigma_\gamma), \sigma_\gamma))_+^2 \right) \cdot \frac{\partial}{\partial\sigma} x_{j-1}(\sigma_\gamma) \cdot \delta\sigma \right] \\
&+ \frac{1}{2\gamma z_j(\sigma)} \int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)} (G_j(\bar{t}, x, \sigma_\gamma))_+^2 \, dx \cdot \frac{\partial}{\partial\sigma} z_j(\sigma_\gamma) \cdot \delta\sigma
\end{aligned}
$$

Using the abbreviation

$$
\begin{aligned}
r(\sigma_\gamma) :=& \frac{1}{2\gamma z_j(\sigma_\gamma)} \int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)} (G_j(\bar{t}, x, \sigma_\gamma))_+^2 \, dx \\
=& \frac{1}{2\gamma z_j(\sigma_\gamma)} \int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)} (Y_j(\bar{t}, x, \sigma_\gamma) - \bar{y}(x))_+^2 \, dx
\end{aligned}
$$

and integration by parts, we obtain

$$
\begin{aligned}
&\frac{\partial}{\partial\sigma}J_\gamma(y(\sigma_\gamma))\cdot\delta\sigma \\
&=\frac{\partial}{\partial\sigma}J(y(\sigma_\gamma))\cdot\delta\sigma \\
&+\sum_{j=1}^{K+1}\left[z_j(\sigma)\cdot\left[\frac{1}{\gamma z_j(\sigma)}\int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)}\frac{\partial}{\partial\sigma}G_j(\bar t,x,\sigma_\gamma)\cdot\delta\sigma\cdot(G_j(\bar t,x,\sigma_\gamma))_+\,dx\right.\right. \\
&+\frac{1}{2\gamma z_j(\sigma)^2}\left[\int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)}\frac{\partial}{\partial x}\left(G_j(\bar t,x,\sigma_\gamma)\right)_+^2\cdot(x-x_{j-1})\,dx\cdot\frac{\partial}{\partial\sigma}x_j(\sigma_\gamma)\cdot\delta\sigma\right. \\
&\left.\left.\int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)}\frac{\partial}{\partial x}\left(G_j(\bar t,x,\sigma_\gamma)\right)_+^2\cdot(x_j-x)\,dx\cdot\frac{\partial}{\partial\sigma}x_{j-1}(\sigma_\gamma)\cdot\delta\sigma\right] \\
&\left.+r(\sigma_\gamma)\cdot\frac{\partial}{\partial\sigma}z_j(\sigma_\gamma)\delta\sigma\right] \\
&=\frac{\partial}{\partial\sigma}J(y(\sigma_\gamma))\cdot(\sigma-\sigma_\gamma) \\
&+\sum_{j=1}^{K+1}\left[\int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)}\frac{\partial}{\partial\sigma}G_j(\bar t,x,\sigma_\gamma)\cdot\delta\sigma\cdot\frac{(G_j(\bar t,x,\sigma_\gamma))_+}{\gamma}\,dx\right. \\
&+\int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)}\frac{\partial}{\partial x}G_j(\bar t,x,\sigma_\gamma)\cdot\frac{x-x_{j-1}(\sigma_\gamma)}{z_j(\sigma_\gamma)}\cdot\frac{(G_j(\bar t,x,\sigma_\gamma))_+}{\gamma}\,dx\cdot\frac{\partial}{\partial\sigma}x_j(\sigma_\gamma)\cdot\delta\sigma \\
&+\int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)}\frac{\partial}{\partial x}G_j(\bar t,x,\sigma_\gamma)\cdot\frac{x_j(\sigma_\gamma)-x}{z_j(\sigma_\gamma)}\cdot\frac{(G_j(\bar t,x,\sigma_\gamma))_+}{\gamma}\,dx\cdot\frac{\partial}{\partial\sigma}x_{j-1}(\sigma_\gamma)\cdot\delta\sigma \\
&\quad\left.+\,r(\sigma_\gamma)\cdot\frac{\partial}{\partial\sigma}z_j(\sigma_\gamma)\delta\sigma\right]
\end{aligned}
$$

Using the above deduced facts and defining

$$
\lambda_j(x,\sigma):=\begin{cases}-\frac{(G_j(\bar t,x,\sigma_\gamma))_+}{\gamma}, & \text{for } x_{j-1}(\sigma)\le x\le x_j(\sigma)\\ 0, & \text{for } x\in[a,b]\setminus[x_j(\sigma),x_{j+1}(\sigma)]\end{cases} \tag{3.16}
$$

$$
=\begin{cases}-\frac{(Y_j(\bar t,x,\sigma)-\bar y(x))_+}{\gamma}, & \text{for } x_{j-1}(\sigma)\le x\le x_j(\sigma)\\ 0, & \text{for } x\in[a,b]\setminus[x_j(\sigma),x_{j+1}(\sigma)]\end{cases} \tag{3.17}
$$

for all $k\in\{1,\ldots,K+1\}$, the optimality conditions of Theorem 3.8 can be written in the following way:

Theorem 3.9 (Necessary optimality Conditions for P_γ)**.** *Let Assumption 5 hold, $\gamma>0$ be arbitrary and let $\sigma_\gamma\in\Sigma_{ad}$ be a local optimal solution for P_γ.*

Then for all $\sigma \in \Sigma_{ad}$ it holds that:

$$
\begin{aligned}
&\frac{\partial}{\partial \sigma} J(y(\sigma_\gamma)) \cdot (\sigma - \sigma_\gamma) \\
&+ \sum_{j=1}^{K+1} \Bigg[- \int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)} \frac{\partial}{\partial \sigma} Y_j(\bar{t}, x, \sigma_\gamma)) \cdot (\sigma - \sigma_\gamma) \cdot \lambda_j(x, \sigma_\gamma) \, dx \\
&\quad + \left(\int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)} \frac{\partial}{\partial x} \left(\bar{y}(x) - Y_j(\bar{t}, x, \sigma_\gamma)\right) \cdot \frac{x - x_{j-1}(\sigma_\gamma)}{x_j(\sigma_\gamma) - x_{j-1}(\sigma_\gamma)} \cdot \lambda_j(x, \sigma_\gamma) \, dx \right) \\
&\quad \cdot \frac{\partial}{\partial \sigma} x_j(\sigma_\gamma) \cdot (\sigma - \sigma_\gamma) \\
&\quad + \left(\int_{x_{j-1}(\sigma_\gamma)}^{x_j(\sigma_\gamma)} \frac{\partial}{\partial x} \left(\bar{y}(x) - Y_j(\bar{t}, x, \sigma_\gamma)\right) \cdot \frac{x_j(\sigma_\gamma) - x}{x_j(\sigma_\gamma) - x_{j-1}(\sigma_\gamma)} \cdot \lambda_j(x, \sigma_\gamma) \, dx \right) \\
&\quad \cdot \frac{\partial}{\partial \sigma} x_{j-1}(\sigma_\gamma) \cdot (\sigma - \sigma_\gamma) + r_j(\sigma_\gamma) \cdot (\sigma - \sigma_\gamma) \Bigg] \geq 0
\end{aligned}
\tag{3.18}
$$

Proof. This is a simple result of Theorem 3.8 and the above considerations. □

Lemma 3.1. *Let $(\sigma_{\gamma_k})_k \in \mathbb{N}$ denote a sequence of local solutions σ_{γ_k} of P_{γ_k} that converges to a local solution $\bar{\sigma}$ of P. Then it holds for all $j \in \{1, \ldots, K+1\}$:*

$$\lim_{k \to \infty} r_j(\sigma_{\gamma_k}) = 0 \tag{3.19}$$

Proof. We first note that for $k \in \mathbb{N}$ large enough

$$J_{\gamma_k}(\sigma_{\gamma_k}) \leq J_{\gamma_k}(\bar{\sigma}) = J(\bar{\sigma}) < \infty$$

is equivalent to

$$0 \leq \frac{1}{2\gamma} \int_a^b \left(y(\bar{t}, x, \sigma_{\gamma_k}) - \bar{y}(x)\right)_+^2 \, dx \leq J(y(\bar{\sigma})) - J(y(\sigma_{\gamma_k})).$$

Since $\sigma_k \to \bar{\sigma}$ for $k \to \infty$ and the continuity of $\Sigma \ni \sigma \mapsto J(y(\sigma_{\gamma_k})) \in \mathbb{R}$, we conclude that

$$\lim_{k \to \infty} \frac{1}{2\gamma} \int_a^b \left(y(\bar{t}, x, \sigma_{\gamma_k}) - \bar{y}(x)\right)_+^2 \, dx = 0$$

Since $x_{j-1}(\sigma)$ and $x_j(\sigma)$ are by assumption bounded away from each other for all $j \in \{1, \ldots, K+1\}$ and the integrand of the above integral is nonnegative, it holds for all $j = 1, \ldots, K+1$:

$$\lim_{k \to \infty} \frac{1}{2\gamma \left(x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})\right)} \int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} \left(Y_j(\bar{t}, x, \sigma_{\gamma_k}) - \bar{y}(x)\right)_+^2 \, dx = 0,$$

and hence

$$\lim_{k \to \infty} r_j(\sigma_{\gamma_k}) = 0, \quad j = 1, \ldots, K+1$$

□

In the remaining part of this section, we will consider a sequence of local solutions $(\sigma_{\gamma_k})_{k\in\mathbb{N}}$ that converges to some local solution $\bar{\sigma}$ of (P) in which a Robinson CQ is satisfied. Our aim is to show that for a further subsequence, again denoted by $(\gamma_k)_{k\in\mathbb{N}}$, the corresponding sequence of Lagrange multiplier estimates $\lambda_j(x,\sigma)$, $j \in \{1,\ldots,K+1\}$, in (3.16) converge w.r.t. the weak*-topology of Borel measures to some nonpositive measures μ_j^*. In order to prove that, we will proceed in the same way as in [7, Lemma 9, Lemma 10], i.e. we first show that the sequences $\lambda_j(x,\sigma)$ are uniformly bounded in L^1 and then apply the Banach-Alaoglu theorem. Furthermore, we will be able to show that, if in σ^* the Robinson CQ is satisfied, the necessary optimality conditions in Theorem 2.6 are satisfied by $(\sigma^*, \mu_1^*, \ldots, \mu_{K+1}^*)$.

Lemma 3.2. *Let Assumption 5 hold and consider a sequence of local solutions $(\sigma_{\gamma_k})_{k\in\mathbb{N}}$ of (P_{γ_k}) that converges towards a local solution $\bar{\sigma}$ of (P) in which the Robinson CQ is satisfied. Then the sequences $\lambda_j(\cdot,\sigma_{\gamma_k})$, $j = 1,\ldots,K+1$ are uniformly bounded in $L^1([a,b])$.*

Proof. For all $k \in \mathbb{N}$ the optimal conditions (3.18) are satisfied in σ_{γ_k}, i.e. for all $\sigma \in \Sigma_{ad}$ it holds that:

$$
\begin{aligned}
&\frac{\partial}{\partial\sigma} J(y(\sigma_{\gamma_k})) \cdot (\sigma - \sigma_{\gamma_k}) \\
&+ \sum_{j=1}^{K+1} \Bigg[\int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} -\frac{\partial}{\partial\sigma} Y_j(\bar{t}, x, \sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k}) \cdot \lambda_j(x,\sigma_{\gamma_k})\, dx \\
&+ \left(\int_{x_{j-1}(\sigma_{\gamma})}^{x_j(\sigma_{\gamma_k})} \frac{\partial}{\partial x}\left(\bar{y}(x) - Y_j(\bar{t},x,\sigma_{\gamma_k})\right) \cdot \frac{x - x_{j-1}(\sigma_{\gamma_k})}{x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})} \cdot \lambda_j(x,\sigma_{\gamma_k})\, dx \right) \\
&\cdot \frac{\partial}{\partial\sigma} x_j(\sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k}) \\
&+ \left(\int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} \frac{\partial}{\partial x}\left(\bar{y}(x) - Y_j(\bar{t},x,\sigma_{\gamma_k})\right) \cdot \frac{x_j(\sigma_{\gamma_k}) - x}{x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})} \cdot \lambda_j(x,\sigma_{\gamma_k})\, dx \right) \\
&\cdot \frac{\partial}{\partial\sigma} x_{j-1}(\sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k}) \\
&+ r_j(\sigma_{\gamma_k}) \cdot \frac{\partial}{\partial\sigma}\left(x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})\right) \cdot (\sigma - \sigma_{\gamma_k}) \Bigg] \geq 0
\end{aligned}
\tag{3.20}
$$

or equivalently

$$
\begin{aligned}
&\sum_{j=1}^{K+1} - \Bigg[\int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} \Big(-\frac{\partial}{\partial\sigma} Y_j(\bar{t},x,\sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k}) \\
&+ \frac{\partial}{\partial x}\left(\bar{y}(x) - Y_j(\bar{t},x,\sigma_{\gamma_k})\right) \cdot \bigg(\frac{x - x_{j-1}(\sigma_{\gamma_k})}{x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})} \cdot \frac{\partial}{\partial\sigma} x_j(\sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k}) \\
&+ \frac{x_j(\sigma_{\gamma_k}) - x}{x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})} \cdot \frac{\partial}{\partial\sigma} x_{j-1}(\sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k}) \bigg) \Big) \lambda_j(x,\sigma_{\gamma_k})\, dx \Bigg] \\
&\leq \frac{\partial}{\partial\sigma} J(y(\sigma_{\gamma_k})) \cdot (\sigma - \sigma_{\gamma_k}) + r(\sigma_{\gamma_k}) \cdot \frac{\partial}{\partial\sigma}\left(x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})\right) \cdot (\sigma - \sigma_{\gamma_k})
\end{aligned}
$$

Using Theorem 2.2, Lemma 3.1 and the compactness of Σ_{ad}, we conclude that the expression on the right hand side is uniformly bounded w.r.t. k and σ. Therefore, it exists a positive constant $C > 0$ with

$$\sum_{j=1}^{K+1} \Bigg[\int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} (-\frac{\partial}{\partial \sigma_k} Y_j(\bar{t}, x, \sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k})$$
$$+ \frac{\partial}{\partial x} (\bar{y}(x) - Y_j(\bar{t}, x, \sigma_{\gamma_k})) \cdot \left(\frac{x - x_{j-1}(\sigma_{\gamma_k})}{x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})} \cdot \frac{\partial}{\partial \sigma} x_j(\sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k}) \right.$$
$$\left. + \frac{x_j(\sigma_{\gamma_k}) - x}{x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})} \cdot \frac{\partial}{\partial \sigma} x_{j-1}(\sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k}) \right)) \, (-\lambda_j(x, \sigma_{\gamma_k})) \; dx \Bigg]$$
$$\leq C \tag{3.21}$$

Considering $\bar{\sigma}$, we define by $I_j \subset [x_{j-1}(\bar{\sigma}) - \varepsilon_0, x_j(\bar{\sigma}) + \varepsilon_0]$ the sets

$$I_j := \{x \in [x_{j-1}(\bar{\sigma}) - \varepsilon_0, x_j(\bar{\sigma}) + \varepsilon_0] : Y_j(\bar{t}, x, \bar{\sigma}) \geq \bar{y}(x)\} \tag{3.22}$$

for $j = 1, \dots, K+1$, where $\varepsilon_0 > 0$ is chosen small enough such that the extensions $Y_j(\bar{t}, \cdot, \bar{\sigma})$ are well defined on $[x_{j-1}(\bar{\sigma}) - \varepsilon_0, x_j(\bar{\sigma}) + \varepsilon_0]$ for all $j = 1, \dots, K+1$. Furthermore, we define the sets

$$I_{j,\varepsilon_1} = \bigcup_{x \in I_j} (x - \varepsilon_1, x + \varepsilon_1) \cap [x_{j-1}(\bar{\sigma}) - \varepsilon_0, x_j(\bar{\sigma}) + \varepsilon_0], \quad j = 1, \dots, K+1$$

for some $\varepsilon_1 > 0$. Since the sets $[x_{j-1}(\bar{\sigma}) - \varepsilon_0, x_j(\bar{\sigma}) + \varepsilon_0] \setminus I_{j,\varepsilon_1}$ are closed and bounded and hence compact and the functions $(Y_j(\bar{t}, \cdot, \bar{\sigma}) - \bar{y}(\cdot))$ are continuous on the corresponding sets, for all $\varepsilon_0 > 0$ there exists $\delta_0 > 0$ such that

$$Y_j(\bar{t}, x, \bar{\sigma}) - \bar{y}(x) \leq -\delta_0 < 0 \quad \text{for all} \quad x \in [x_{j-1}(\bar{\sigma}) - \varepsilon_0, x_j(\bar{\sigma}) + \varepsilon_0] \setminus I_{j,\varepsilon_1}.$$

Using the continuity of $Y_j(\bar{t}, x, \sigma)$ w.r.t. σ and using that $\sigma_{\gamma_k} \to \bar{\sigma}$ for $k \to \infty$, we obtain for all $j = 1, \dots, K+1$

$$Y_j(\bar{t}, x, \sigma_{\gamma_k}) - \bar{y}(x) \leq -\frac{\delta_0}{2} < 0 \quad \text{for all} \quad x \in [x_{j-1}(\sigma_{\gamma_k}), x_j(\sigma_{\gamma_k})] \setminus I_{j,\varepsilon_1}.$$

for k large enough. Recalling that $\lambda_j(\cdot, \sigma_{\gamma_k})$ is zero outside $[x_{j-1}(\sigma_{\gamma_k}), x_j(\sigma_{\gamma_k})]$, this yields

$$\lambda_j(x, \lambda_{\gamma_k}) = 0 \quad \text{for all} \quad x \in \mathbb{R} \setminus I_{j,\varepsilon_1}. \tag{3.23}$$

for all $j = 1, \dots, K+1$ if k is large enough. Our goal is to show that the sequences $\lambda_j(\cdot, \sigma_{\gamma_k})$, $j = 1, \dots, K+1$ are uniformly bounded in $L^1(\mathbb{R})$. By assumption, in $\bar{\sigma}$ the Robinson CQ is satisfied, it hence holds that

$$\Bigg(\bar{y}(x) - Y_j(\bar{t}, x, \bar{\sigma}) - \frac{\partial}{\partial \sigma} Y_j(\bar{t}, x, \bar{\sigma}) \cdot (\tilde{\sigma} - \bar{\sigma})$$
$$+ \frac{\partial}{\partial x} [\bar{y}(x) - Y_j(\bar{t}, x, \bar{\sigma})] \left(\frac{x - x_{k-1}(\bar{\sigma})}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \cdot \frac{\partial}{\partial \sigma} x_k(\bar{\sigma}) \, (\tilde{\sigma} - \bar{\sigma}) \right.$$
$$\left. + \frac{x_k(\bar{\sigma}) - x}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \cdot \frac{\partial}{\partial \sigma} x_{k-1}(\bar{\sigma}) \, (\tilde{\sigma} - \bar{\sigma}) \right) \Bigg) \geq \varepsilon_2$$

for all $x \in [x_{j-1}(\bar{\sigma}), x_j(\bar{\sigma})]$ for some $\tilde{\sigma} \in \Sigma_{ad}$. Using the continuity of the above terms w.r.t. x, one can show that there exists $\varepsilon_3 \leq \varepsilon_2$ with

$$\begin{aligned}\Bigg(&\bar{y}(x) - Y_j(\bar{t}, x, \bar{\sigma}) - \frac{\partial}{\partial \sigma} Y_j(\bar{t}, x, \bar{\sigma}) \cdot (\tilde{\sigma} - \bar{\sigma}) \\ &+ \frac{\partial}{\partial x} \left[\bar{y}(x) - Y_j(\bar{t}, x, \bar{\sigma})\right] \Bigg(\frac{x - x_{k-1}(\bar{\sigma})}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \cdot \frac{\partial}{\partial \sigma} x_k(\bar{\sigma}) \, (\tilde{\sigma} - \bar{\sigma}) \\ &+ \frac{x_k(\bar{\sigma}) - x}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \cdot \frac{\partial}{\partial \sigma} x_{k-1}(\bar{\sigma}) \, (\tilde{\sigma} - \bar{\sigma}) \Bigg) \Bigg) \geq \frac{\varepsilon_2}{2} \end{aligned} \tag{3.24}$$

for all $x \in [x_{j-1}(\bar{\sigma}) - \varepsilon_2, x_j(\bar{\sigma}) + \varepsilon_2]$.

Since it holds that $(\bar{y}(x) - Y_j(\bar{t}, x, \bar{\sigma}))\, |_{I_j \cap [x_{j-1}(\bar{\sigma}) - \varepsilon_2, x_j(\bar{\sigma}) + \varepsilon_2]} \leq 0$, we obtain by (3.24) that

$$\begin{aligned}\Bigg(&- \frac{\partial}{\partial \sigma} Y_j(\bar{t}, x, \bar{\sigma}) \cdot (\tilde{\sigma} - \bar{\sigma}) \\ &+ \frac{\partial}{\partial x} \left[\bar{y}(x) - Y_j(\bar{t}, x, \bar{\sigma})\right] \Bigg(\frac{x - x_{k-1}(\bar{\sigma})}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \cdot \frac{\partial}{\partial \sigma} x_k(\bar{\sigma}) \, (\tilde{\sigma} - \bar{\sigma}) \\ &+ \frac{x_k(\bar{\sigma}) - x}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \cdot \frac{\partial}{\partial \sigma} x_{k-1}(\bar{\sigma}) \, (\tilde{\sigma} - \bar{\sigma}) \Bigg) \Bigg) \geq \frac{\varepsilon_2}{2} \end{aligned} \tag{3.25}$$

for all $x \in I_j \cap [x_{j-1}(\bar{\sigma}) - \varepsilon_2, x_j(\bar{\sigma}) + \varepsilon_2]$ and again by continuity of the above terms

$$\begin{aligned}\Bigg(&- \frac{\partial}{\partial \sigma} Y_j(\bar{t}, x, \bar{\sigma}) \cdot (\tilde{\sigma} - \bar{\sigma}) \\ &+ \frac{\partial}{\partial x} \left[\bar{y}(x) - Y_j(\bar{t}, x, \bar{\sigma})\right] \Bigg(\frac{x - x_{k-1}(\bar{\sigma})}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \cdot \frac{\partial}{\partial \sigma} x_k(\bar{\sigma}) \, (\tilde{\sigma} - \bar{\sigma}) \\ &+ \frac{x_k(\bar{\sigma}) - x}{x_k(\bar{\sigma}) - x_{k-1}(\bar{\sigma})} \cdot \frac{\partial}{\partial \sigma} x_{k-1}(\bar{\sigma}) \, (\tilde{\sigma} - \bar{\sigma}) \Bigg) \Bigg) \geq \frac{\varepsilon_2}{4} \end{aligned} \tag{3.26}$$

for all $x \in I_{j,\varepsilon_1} \cap [x_{j-1}(\bar{\sigma}) - \varepsilon_2, x_j(\bar{\sigma}) + \varepsilon_2]$ if ε_1 is chosen small enough. Exploiting the continuity of the above terms w.r.t. σ, we further obtain that

$$\begin{aligned}\Bigg(&- \frac{\partial}{\partial \sigma} Y_j(\bar{t}, x, \sigma_{\gamma_k}) \cdot (\tilde{\sigma} - \sigma_{\gamma_k}) \\ &+ \frac{\partial}{\partial x} \left[\bar{y}(x) - Y_j(\bar{t}, x, \sigma_{\gamma_k})\right] \Bigg(\frac{x - x_{k-1}(\sigma_{\gamma_k})}{x_k(\sigma_{\gamma_k}) - x_{k-1}(\sigma_{\gamma_k})} \cdot \frac{\partial}{\partial \sigma} x_k(\sigma_{\gamma_k}) \, (\tilde{\sigma} - \sigma_{\gamma_k}) \\ &+ \frac{x_k(\sigma_{\gamma_k}) - x}{x_k(\sigma_{\gamma_k}) - x_{k-1}(\sigma_{\gamma_k})} \cdot \frac{\partial}{\partial \sigma} x_{k-1}(\sigma_{\gamma_k}) \, (\tilde{\sigma} - \sigma_{\gamma_k}) \Bigg) \Bigg) \geq \frac{\varepsilon_2}{8} \end{aligned} \tag{3.27}$$

for all $x \in I_{j,\varepsilon_1} \cap [x_{j-1}(\bar{\sigma}) - \frac{\varepsilon_2}{2}, x_j(\bar{\sigma}) + \frac{\varepsilon_2}{2}]$ if k is chosen large enough. In the following, we will use the abbreviations

$$\begin{aligned} R_j(x, \sigma) := \Bigg(&- \frac{\partial}{\partial \sigma} Y_j(\bar{t}, x, \sigma) \cdot (\tilde{\sigma} - \sigma) + \frac{\partial}{\partial x} \left[\bar{y}(x) - Y_j(\bar{t}, x, \sigma)\right] \\ &\cdot \Bigg(\frac{x - x_{k-1}(\sigma)}{x_k(\sigma) - x_{k-1}(\sigma)} \cdot \frac{\partial}{\partial \sigma} x_k(\sigma) \, (\tilde{\sigma} - \sigma) \\ &+ \frac{x_k(\sigma) - x}{x_k(\sigma) - x_{k-1}(\sigma)} \cdot \frac{\partial}{\partial \sigma} x_{k-1}(\sigma) \, (\tilde{\sigma} - \sigma) \Bigg) \Bigg) \end{aligned} \tag{3.28}$$

Now, from (3.23), we deduce that for k large enough it holds for all $j = 1, \ldots, K+1$:

$$\int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} R_j(x, \sigma_{\gamma_k})(-\lambda_j(x, \sigma_{\gamma_k}))\,\mathrm{d}x$$
$$= \int_{[x_{j-1}(\sigma_{\gamma_k}), x_j(\sigma_{\gamma_k})]\cap I_{j,\varepsilon_1}} R_j(x, \sigma_{\gamma_k})(-\lambda_j(x, \sigma_{\gamma_k}))\,\mathrm{d}x$$

Using (3.27) and the nonnegativity of $(-\lambda_j(x, \sigma_{\gamma_k}))$, we further obtain

$$\int_{[x_{j-1}(\sigma_{\gamma_k}), x_j(\sigma_{\gamma_k})]\cap I_{j,\varepsilon_1}} R_j(x, \sigma_{\gamma_k})(-\lambda_j(x, \sigma_{\gamma_k}))\,\mathrm{d}x$$
$$\geq \int_{[x_{j-1}(\sigma_{\gamma_k}), x_j(\sigma_{\gamma_k})]\cap I_{j,\varepsilon_1}} \frac{\varepsilon_2}{8}(-\lambda_j(x, \sigma_{\gamma_k}))\,\mathrm{d}x$$

Using again (3.23), we see that

$$\int_{[x_{j-1}(\sigma_{\gamma_k}), x_j(\sigma_{\gamma_k})]\cap I_{j,\varepsilon_1}} \frac{\varepsilon_2}{8}(-\lambda_j(x, \sigma_{\gamma_k}))\,\mathrm{d}x$$
$$= \int_{[x_{j-1}(\sigma_{\gamma_k}), x_j(\sigma_{\gamma_k})]\cap I_{j,\varepsilon_1}} \frac{\varepsilon_2}{8}(-\lambda_j(x, \sigma_{\gamma_k}))\,\mathrm{d}x$$
$$+ \int_{[x_{j-1}(\sigma_{\gamma_k}), x_j(\sigma_{\gamma_k})]\setminus I_{j,\varepsilon_1}} \frac{\varepsilon_2}{8}(-\lambda_j(x, \sigma_{\gamma_k}))\,\mathrm{d}x$$
$$= \int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} \frac{\varepsilon_2}{8}(-\lambda_j(x, \sigma_{\gamma_k}))\,\mathrm{d}x = \int_{[a,b]} \frac{\varepsilon_2}{8}(-\lambda_j(x, \sigma_{\gamma_k}))\,\mathrm{d}x$$

Hence, there exists $\bar{k} \in \mathbb{N}$ large enough such that for all $k \geq \bar{k}$ and all $j = 1, \ldots, K+1$ it holds that

$$\int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} R_j(x, \sigma_{\gamma_k})(-\lambda_j(x, \sigma_{\gamma_k}))\,\mathrm{d}x = \int_{[a,b]} \frac{\varepsilon_2}{8}(-\lambda_j(x, \sigma_{\gamma_k}))\,\mathrm{d}x \geq 0,$$

where the last inequality directly follows from the nonnegativity of $(-\lambda_j(x, \sigma_{\gamma_k}))$. Combining this result with (3.21) yields

$$\sum_{j=1}^{K+1} \int_{[a,b]} |\lambda_j(x, \sigma_{\gamma_k})|\ dx \leq \frac{8C}{\varepsilon_2} := \tilde{C} \quad \forall k \geq \bar{k}$$

Hence, the sequences $\lambda_j\,(x, \sigma_{\gamma_k})$, $j = 1, \ldots, K+1$, are uniformly bounded in $L^1([a,b])$.

□

Theorem 3.10. *Let Assumption 5 hold and consider a sequence $\sigma_{\gamma_k} \to \bar{\sigma}$ for $k \to \infty$, where $\bar{\sigma}$ is an optimal solution for the original problem (P). We assume that in $\bar{\sigma}$ the Robinson's CQ is satisfied. Then there exists a subsequence, again denoted by $(\sigma_{\gamma_k})_{k\in\mathbb{N}}$, such that the corresponding sequences $(\lambda_j(\cdot, \sigma_{\gamma_k}))_{k\in\mathbb{N}}$, $j = 1, \ldots, K+1$, converge w.r.t. the weak* topology in the space of Borel measures to nonpositive Borel-measures $\mu_j \in \mathcal{M}([a,b])$, $j = 1, \ldots, K+1$, such that the optimality conditions in Theorem 2.6 are satisfied in $\bar{\sigma}$, where μ_j, $j = 1, \ldots, K+1$, are the corresponding Lagrange multipliers.*

Proof. Since the Lagrange multiplier estimates are uniformly bounded w.r.t. the L^1-norm by Lemma 3.2, the Banach-Alaoglu theorem yields that there exists a subsequence, again denoted by $(\gamma_k)_{k\in\mathbb{N}}$, such that the corresponding sequences of Lagrange multiplier estimates $(\lambda_j(\cdot,\sigma_{\gamma_k}))_{k\in\mathbb{N}}$, $j = 1,\ldots,K+1$, converge w.r.t. the weak*-topology in the space of Borel measures to nonpositive Borel measures $\mu_j \in \mathcal{M}([a,b])$, $j = 1,\ldots,K+1$, respectively. It is easy to check that the optimality conditions in Theorem 2.6, i.e. (2.31), (2.32), (2.33) and (2.34) are satisfied for $\bar{\mu}_j = \mu_j$, $j = 1,\ldots,K+1$.

(2.31) as well as (2.33) can be shown as in the proof of Theorem 3.4. Concerning (2.32), we first notice that since $0 \le J_\gamma(\sigma_{\gamma_k}) \le J_\gamma(\bar{\sigma}) = J(y(\bar{\sigma}))$ and $\sigma_{\gamma_k} \to \bar{\sigma}$. Using the definitions of $\lambda_j(\cdot,\sigma)$, $j = 1,\ldots,K+1$, and $J(y(\sigma))$, one can show that

$$\begin{aligned}
0 &\le \lim_{k\to\infty} \sum_{j=1}^{K+1} \int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} \left(\bar{y}\,(x) - Y_j\,(\bar{t},x,\sigma_{\gamma_k})\right) \cdot \lambda_j(x,\sigma_{\gamma_k})\,dx \\
&\le \lim_{k\to\infty} \left(J(\bar{\sigma}) - J(\sigma_{\gamma_k})\right) = 0
\end{aligned}$$

Using this result, the definition of $\lambda_j(\cdot,\sigma)$ and the fact that

$$\left(\bar{y}\,(x) - Y_j\,(\bar{t},x,\sigma_{\gamma_k})\right) \cdot \lambda_j(x,\sigma_{\gamma_k}) \ge 0, \quad j = 1,\ldots,K+1$$

one can show that

$$\begin{aligned}
0 &= \lim_{k\to\infty} \sum_{j=1}^{K+1} \int_{x_{j-1}(\bar{\sigma})}^{x_j(\bar{\sigma})} \left(\bar{y}\,(x) - Y_j\,(\bar{t},x,\sigma_{\gamma_k})\right) \cdot \lambda_j(x,\sigma_{\gamma_k})\,dx \\
&= \lim_{k\to\infty} \sum_{j=1}^{K+1} \int_{x_{j-1}(\bar{\sigma})}^{x_j(\bar{\sigma})} \left(\bar{y}\,(x) - Y_-\,(\bar{t},x,\bar{\sigma})\right) \cdot \lambda_j(x,\sigma_{\gamma_k})\,dx \\
&\quad + \lim_{k\to\infty} \sum_{j=1}^{K+1} \int_{x_{j-1}(\bar{\sigma})}^{x_j(\bar{\sigma})} \left(Y_j\,(\bar{t},x,\bar{\sigma}) - Y_j\,(\bar{t},x,\sigma_{\gamma_k})\right) \cdot \lambda_j(x,\sigma_{\gamma_k})\,dx
\end{aligned}$$

Since, as shown in the previous lemma, $\lambda_j(\cdot,\sigma_{\gamma_k})$, $j \in \{1,\ldots,K+1\}$, are uniformly bounded in $L^1([a,b])$ and for $\varepsilon > 0$ small enough the mappings $B_\varepsilon(\bar{\sigma}) \ni \sigma \mapsto Y_j\,(\bar{t},\cdot,\sigma) \in C([x_{j-1}(\bar{\sigma}),x_j(\bar{\sigma})])$, $j \in \{1,\ldots,K+1\}$, are continuous and $\sigma_{\gamma_k} \to \bar{\sigma}$ for $k \to \infty$, we obtain

$$\lim_{k\to\infty} \sum_{j=1}^{K+1} \int_{x_{j-1}(\bar{\sigma})}^{x_j(\bar{\sigma})} \left(Y_j\,(\bar{t},x,\bar{\sigma}) - Y_j\,(\bar{t},x,\sigma_{\gamma_k})\right) \cdot \lambda_j(x,\sigma_{\gamma_k})\,dx = 0$$

Further, using the fact that $\lambda_j(\cdot,\sigma_{\gamma_k}) \to \mu_j$, $j = 1,\ldots,K+1$, w.r.t. the weak*-topology of Borel measures, we can conclude

$$\begin{aligned}
0 &= \lim_{k\to\infty} \sum_{j=1}^{K+1} \int_{x_{j-1}(\bar{\sigma})}^{x_j(\bar{\sigma})} \left(\bar{y}\,(x) - Y_j\,(\bar{t},x,\bar{\sigma})\right) \cdot \lambda_j(x,\sigma_{\gamma_k})\,dx \\
&= \lim_{k\to\infty} \sum_{j=1}^{K+1} \int_{x_{j-1}(\bar{\sigma})}^{x_j(\bar{\sigma})} \left(\bar{y}\,(x) - Y_j\,(\bar{t},x,\bar{\sigma})\right)\,\mu_j(\mathrm{d}x)
\end{aligned}$$

Hence, (2.32) is proved. Finally, we will prove (2.34): We will use the following abbreviations

$$I_1^j(\sigma, x) := \frac{\partial}{\partial x}\left(\bar{y}(x) - Y_j(\bar{t}, x, \sigma)\right) \cdot \frac{x - x_{j-1}(\sigma)}{x_j(\sigma) - x_{j-1}(\sigma)}$$
$$I_2^j(\sigma, x) := \frac{\partial}{\partial x}\left(\bar{y}(x) - Y_j(\bar{t}, x, \sigma)\right) \cdot \frac{x_j(\sigma) - x}{x_j(\sigma) - x_{j-1}(\sigma)}$$

and recall that each optimal solution σ_{γ_k} of (P_{γ_k}) satisfies for all $\sigma \in \Sigma_{ad}$:

$$\begin{aligned}
&\frac{\partial}{\partial \sigma} J(y(\sigma_{\gamma_k})) \cdot (\sigma - \sigma_{\gamma_k}) && (3.29)\\
&+ \sum_{j=1}^{K+1} \left[- \int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} \frac{\partial}{\partial \sigma} Y_j(\bar{t}, x, \bar{\sigma}) \cdot (\sigma - \sigma_{\gamma_k}) \cdot \lambda_j(x, \sigma_{\gamma_k})\, dx \right.\\
&+ \int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} I_1^j(\sigma_{\gamma_k}, x) \cdot \lambda_j(x, \sigma_{\gamma_k})\, dx \cdot \frac{\partial}{\partial \sigma} x_j(\sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k})\\
&+ \int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} I_2^j(\sigma_{\gamma_k}, x) \cdot \lambda_j(x, \sigma_{\gamma_k})\, dx \cdot \frac{\partial}{\partial \sigma} x_{j-1}(\sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k})\\
&\left. + r_j(\sigma_{\gamma_k}) \cdot \frac{\partial}{\partial \sigma}\left(x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})\right) \cdot (\sigma - \sigma_{\gamma_k}) \right] \geq 0 && (3.30)
\end{aligned}$$

We now want to analyse the above variational inequality for $k \to \infty$. We first note that $r_j(\gamma_k) \to 0$ for $k \to \infty$ for all $j \in \{1, \ldots, K+1\}$ by Lemma 3.1. This yields

$$\begin{aligned}
\lim_{k \to \infty} & \left[\frac{\partial}{\partial \sigma} J(y(\sigma_{\gamma_k})) \cdot (\sigma - \sigma_{\gamma_k}) \right.\\
&+ \sum_{j=1}^{K+1} \left[\int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} - \frac{\partial}{\partial \sigma} Y_j(\bar{t}, x, \sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k}) \cdot \lambda_j(x, \sigma_{\gamma_k})\, dx \right.\\
&+ \int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} I_1^j(\sigma_{\gamma_k}, x) \cdot \lambda_j(x, \sigma_{\gamma_k})\, dx \cdot \frac{\partial}{\partial \sigma} x_j(\sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k})\\
&\left.\left. + \int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} I_2^j(\sigma_{\gamma_k}, x) \cdot \lambda_j(x, \sigma_{\gamma_k})\, dx \cdot \frac{\partial}{\partial \sigma} x_{j-1}(\sigma_{\gamma_k}) \cdot (\sigma - \sigma_{\gamma_k}) \right]\right]\\
&\geq 0
\end{aligned}$$

Next, we want to replace the integration limit $x_j(\sigma_{\gamma_k})$ by $x_j(\bar{\sigma})$ for all $j \in \{1, \ldots, K+1\}$. To this end, we are going to use the variable transformation

$$x = x_{j-1}(\sigma_{\gamma_k}) + \frac{\tilde{x} - x_{j-1}(\bar{\sigma})}{x_j(\bar{\sigma}) - x_{j-1}(\bar{\sigma})}\big(x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})\big) \tag{3.31}$$

and obtain

$$
\begin{aligned}
&\lim_{k\to\infty}\Bigg[\frac{\partial}{\partial\sigma}J(y(\sigma_{\gamma_k}))\cdot(\sigma-\sigma_{\gamma_k})\\
&+\sum_{j=1}^{K+1}\Bigg[\int_{x_{j-1}(\bar\sigma)}^{x_j(\bar\sigma)} -\tilde{Y}_j(\bar t,\tilde x,\sigma_{\gamma_k})\cdot(\sigma-\sigma_{\gamma_k})\frac{x_j(\bar\sigma)-x_{j-1}(\bar\sigma)}{x_j(\sigma_{\gamma_k})-x_{j-1}(\sigma_{\gamma_k})}\cdot\tilde\lambda_j(\tilde x,\sigma_{\gamma_k})\,dx\\
&\quad+\int_{x_{j-1}(\bar\sigma)}^{x_j(\bar\sigma)}\tilde I_1^j(\sigma_{\gamma_k},\tilde x)\frac{x_j(\bar\sigma)-x_{j-1}(\bar\sigma)}{x_j(\sigma_{\gamma_k})-x_{j-1}(\sigma_{\gamma_k})}\cdot\tilde\lambda_j(\tilde x,\sigma_{\gamma_k})\,d\tilde x\\
&\qquad\cdot\frac{\partial}{\partial\sigma}x_j(\sigma_{\gamma_k})\cdot(\sigma-\sigma_{\gamma_k})\\
&\quad+\int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})}\tilde I_2^j(\sigma_{\gamma_k},\tilde x)\frac{x_j(\bar\sigma)-x_{j-1}(\bar\sigma)}{x_j(\sigma_{\gamma_k})-x_{j-1}(\sigma_{\gamma_k})}\cdot\tilde\lambda_j(\tilde x,\sigma_{\gamma_k})\,d\tilde x\\
&\qquad\cdot\frac{\partial}{\partial\sigma}x_{j-1}(\sigma_{\gamma_k})\cdot(\sigma-\sigma_{\gamma_k})\Bigg]\Bigg]\geq 0
\end{aligned}
\tag{3.32}
$$

where

$$
\begin{aligned}
&\tilde Y_j(\bar t,\tilde x,\sigma_{\gamma_k})\\
&:=\frac{\partial}{\partial\sigma}Y_j\left(\bar t,x_{j-1}(\sigma_{\gamma_k})+\frac{\tilde x-x_{j-1}(\bar\sigma)}{x_j(\bar\sigma)-x_{j-1}(\bar\sigma)}\big(x_j(\sigma_{\gamma_k})-x_{j-1}(\sigma_{\gamma_k})\big),\sigma_{\gamma_k}\right)\\
&\tilde I_i^j(\sigma_{\gamma_k},\tilde x):=I_1^j\left(\sigma_{\gamma_k},x_{j-1}(\sigma_{\gamma_k})+\frac{\tilde x-x_{j-1}(\bar\sigma)}{x_j(\bar\sigma)-x_{j-1}(\bar\sigma)}\big(x_j(\sigma_{\gamma_k})-x_{j-1}(\sigma_{\gamma_k})\big)\right)\\
&\tilde\lambda_j(\tilde x,\sigma_{\gamma_k}):=\lambda_j\left(x_{j-1}(\sigma_{\gamma_k})+\frac{\tilde x-x_{j-1}(\bar\sigma)}{x_j(\bar\sigma)-x_{j-1}(\bar\sigma)}\big(x_j(\sigma_{\gamma_k})-x_{j-1}(\sigma_{\gamma_k})\big),\sigma_{\gamma_k}\right)
\end{aligned}
\tag{3.33}
$$

with $j=1,\ldots,K+1$ and $i=1,2$. Since the sequence $\big(\lambda_j(x,\sigma_{\gamma_k})\big)_{k\in\mathbb{N}}$ is bounded in $L^1([a,b])$ by Lemma 3.2, the sequences $\big(\tilde\lambda_j(\cdot,\sigma_{\gamma_k})\big)_{k\in\mathbb{N}}$ are obviously also bounded in $L^1([a,b])$. By the Banach-Alaoglu theorem one can deduce that there exists another subsequence, again denoted by $(\gamma_k)_{k\in\mathbb{N}}$, such that the corresponding sequences $\big(\tilde\lambda_j(\cdot,\sigma_{\gamma_k})\big)_{k\in\mathbb{N}}$, converge w.r.t. the weak*-topology in the space of Borel measures to nonpositive Borel measures $\tilde\mu_j\in\mathcal{M}([a,b])$, $j=1,\ldots,K+1$, respectively. In addition, it holds that

$$\tilde Y_j(\bar t,\cdot,\sigma_{\gamma_k})\to\frac{\partial}{\partial\sigma}Y_j(\bar t,\cdot,\bar\sigma)\quad\text{for } k\to\infty\quad\text{in } C([x_{j-1}(\bar\sigma),x_j(\bar\sigma)])$$

$$\tilde I_i^j(\sigma_{\gamma_k},\cdot)\to I_i^j(\sigma_{\bar\gamma},\cdot)\quad\text{for } k\to\infty\quad\text{in } C([x_{j-1}(\bar\sigma),x_j(\bar\sigma)])$$

Using this result and the fact that $\frac{x_j(\bar\sigma)-x_{j-1}(\bar\sigma)}{x_j(\sigma_{\gamma_k})-x_{j-1}(\sigma_{\gamma_k})}\to 1$ for $k\to\infty$, one can show that (3.32) is equivalent to

$$\begin{aligned}J(y(\bar\sigma))\cdot(\sigma-\bar\sigma)+\sum_{j=1}^{K+1}\Bigg[&-\int_{x_{j-1}(\bar\sigma)}^{x_j(\bar\sigma)}\frac{\partial}{\partial\sigma}Y_j(\bar t,x,\bar\sigma)\cdot(\sigma-\bar\sigma)\,\mathrm{d}\tilde\mu_j(x)\\&+\int_{x_{j-1}(\bar\sigma)}^{x_j(\bar\sigma)}I_1^j(\bar\sigma,x)\,\mathrm{d}\tilde\mu_j(x)\cdot\frac{\partial}{\partial\sigma}x_j(\bar\sigma)\cdot(\sigma-\bar\sigma)\\&+\int_{x_{j-1}(\bar\sigma)}^{x_j(\bar\sigma)}I_2^j(\bar\sigma,x)\,\mathrm{d}\tilde\mu_j(x)\cdot\frac{\partial}{\partial\sigma}x_{j-1}(\bar\sigma)\cdot(\sigma-\bar\sigma)\Bigg]\ge 0\end{aligned}$$

and hence (2.34) is satisfied for $\bar\mu_j=\tilde\mu_j, j=1,\dots,K+1$. We now want to show that (2.34) is also satisfied for choice $\bar\mu_j=\mu_j, j=1,\dots,K+1$. To this end, we will show that $\tilde\mu_j=\mu_j$ in $\mathcal{M}\left([x_{j-1}(\bar\sigma),x_j(\bar\sigma)]\right)$ for all $j=1,\dots,K+1$. We consider compact intervals

$$I_{j,\varepsilon}=[x_{j-1}(\bar\sigma)-\varepsilon,x_j(\bar\sigma)+\varepsilon]$$

for some small constant $\varepsilon>0$. Let $\bar k\in\mathbb{N}$ be large enough such that the following holds true for all $k\ge\bar k$:

$$x_{j-1}(\bar\sigma)-\varepsilon\le x_{j-1}(\sigma_{\gamma_k})<x_j(\sigma_{\gamma_k})\le x_j(\bar\sigma)+\varepsilon$$

We note that the weak-* convergence of $(\lambda_j(\cdot,\sigma_{\gamma_k}))_{k\in\mathbb{N}}$ to $\mu_j\in\mathcal{M}([a,b])$, $j=1,\dots,K+1$, implies that for arbitrary $f_j\in C(I_{j,\varepsilon})$ it holds true that

$$\int_{I_{j,\varepsilon}}f_j(x)\mu_j(\mathrm{d}x)=\lim_{k\to\infty}\int_{I_{j,\varepsilon}}f(x)\lambda_j(x,\sigma_{\gamma_k})\mathrm{d}x$$

Recalling (3.16), one can show that

$$\begin{aligned}\int_{I_{j,\varepsilon}}f_j(x)\mu_j(\mathrm{d}x)&=\lim_{k\to\infty}\int_{I_{j,\varepsilon}}f_j(x)\lambda_j(x,\sigma_{\gamma_k})\mathrm{d}x\\&=\lim_{k\to\infty}\int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})}f_j(x)\lambda_j(x,\sigma_{\gamma_k})\mathrm{d}x\end{aligned}$$

Using again the variable transformation in (3.31), we further obtain

$$\begin{aligned}\int_{I_{j,\varepsilon}}f_j(x)\mu_j(\mathrm{d}x)&=\lim_{k\to\infty}\int_{I_{j,\varepsilon}}f_j(x)\lambda_j(x,\sigma_{\gamma_k})\mathrm{d}x\\&=\lim_{k\to\infty}\int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})}f_j(x)\lambda_j(x,\sigma_{\gamma_k})\mathrm{d}x\\&=\lim_{k\to\infty}\int_{x_{j-1}(\bar\sigma)}^{x_j(\bar\sigma)}\tilde f_j(\tilde x)\tilde\lambda_j(\tilde x,\sigma_{\gamma_k})\mathrm{d}\tilde x\end{aligned}$$

where

$$\begin{aligned}&\tilde f(\tilde x)=\\&f\big(x_{j-1}(\sigma_{\gamma_k})+\frac{\tilde x-x_{j-1}(\bar\sigma)}{x_j(\bar\sigma)-x_{j-1}(\bar\sigma)}\big(x_j(\sigma_{\gamma_k})-x_{j-1}(\sigma_{\gamma_k})\big)\big)\frac{x_j(\bar\sigma)-x_{j-1}(\bar\sigma)}{x_j(\sigma_{\gamma_k})-x_{j-1}(\sigma_{\gamma_k})}\end{aligned}$$

Using the definition of $\tilde{\lambda}$ in (3.33), one can finally prove that the following holds true:

$$\begin{aligned}
\int_{I_{j,\varepsilon}} f(x)\mu_j(\mathrm{d}x) &= \lim_{k\to\infty} \int_{I_{j,\varepsilon}} f_j(x)\lambda_j(x,\sigma_{\gamma_k})\mathrm{d}x \\
&= \lim_{k\to\infty} \int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} f_j(x)\lambda_j(x,\sigma_{\gamma_k})\mathrm{d}x \\
&= \lim_{k\to\infty} \int_{x_{j-1}(\bar{\sigma})}^{x_j(\bar{\sigma})} \tilde{f}_j(\tilde{x})\tilde{\lambda}_j(\tilde{x},\sigma_{\gamma_k})\mathrm{d}\tilde{x} \\
&= \lim_{k\to\infty} \int_{I_{j,\varepsilon}} \tilde{f}_j(\tilde{x})\tilde{\lambda}_j(\tilde{x},\sigma_{\gamma_k})\mathrm{d}\tilde{x}
\end{aligned}$$

Since $f_j \in C(I_{j,\varepsilon})$ and due to the transformation (3.31) and the fact that

$$\frac{x_j(\bar{\sigma}) - x_{j-1}(\bar{\sigma})}{x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})} \to 1 \quad \text{for} \quad k \to \infty,$$

we obtain that for all $j = 1, \ldots, K+1$ the term

$$f_j\big(x_{j-1}(\sigma_{\gamma_k}) + \frac{\cdot - x_{j-1}(\bar{\sigma})}{x_j(\bar{\sigma}) - x_{j-1}(\bar{\sigma})}\big(x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})\big)\big)\frac{x_j(\bar{\sigma}) - x_{j-1}(\bar{\sigma})}{x_j(\sigma_{\gamma_k}) - x_{j-1}(\sigma_{\gamma_k})}$$

converges to $f_j(\cdot)$ in $C(I_{j,\varepsilon})$. Therefore, since $\big(\tilde{\lambda}_j(\cdot,\sigma_{\gamma_k})\big)_{k\in\mathbb{N}}$ converge w.r.t. the weak*-topology in the space of Borel measures to nonpositive Borel measures $\tilde{\mu}_j \in \mathcal{M}([a,b])$ we obtain the following

$$\begin{aligned}
\int_{I_{j,\varepsilon}} f_j(x)\mu_j(\mathrm{d}x) &= \lim_{k\to\infty} \int_{I_{j,\varepsilon}} f_j(x)\lambda_j(x,\sigma_{\gamma_k})\mathrm{d}x \\
&= \lim_{k\to\infty} \int_{x_{j-1}(\sigma_{\gamma_k})}^{x_j(\sigma_{\gamma_k})} f_j(x)\lambda_j(x,\sigma_{\gamma_k})\mathrm{d}x \\
&= \lim_{k\to\infty} \int_{x_{j-1}(\bar{\sigma})}^{x_j(\bar{\sigma})} \tilde{f}_j(\tilde{x})\tilde{\lambda}_j(\tilde{x},\sigma_{\gamma_k})\mathrm{d}\tilde{x} \\
&= \lim_{k\to\infty} \int_{I_{j,\varepsilon}} \tilde{f}_j(\tilde{x})\tilde{\lambda}_j(\tilde{x},\sigma_{\gamma_k})\mathrm{d}\tilde{x} \\
&= \int_{I_{j,\varepsilon}} f_j(\tilde{x})\tilde{\mu}_j(\tilde{x})\mathrm{d}\tilde{x}
\end{aligned}$$

Since $f_j \in C(I_{j,\varepsilon})$, $j = 1, \ldots, K+1$, were arbitrary chosen it holds that $\tilde{\mu}_j = \mu_j$ in $\mathcal{M}(I_{j,\varepsilon})$ and hence (2.34) is also satisfied for $\bar{\mu}_j = \mu_j$. $\square$

References

[1] C. Bardos, A. Y. Leroux, and J. C. Nedelec, *First order quasilinear equations with boundary conditions*, Communications in Partial Differential Equations, 4 (1979), pp. 1017–1034.

[2] A. Bressan, *Hyperbolic systems of conservation laws*, vol. 20 of Oxford Lecture Ser. Math. Appl., Oxford University Press, Oxford, 2000. The one-dimensional Cauchy problem.

[3] E. Casas, *Control of an elliptic problem with pointwise state constraints*, SIAM Journal on Control and Optimization, 24 (1986), pp. 1309–1318.

[4] C. M. Dafermos, *Generalized characteristics and the structure of solutions of hyperbolic conservation laws*, Indiana Univ. Math. J., 26 (1977), pp. 1097–1119.

[5] J. Elstrodt, *Mass- und Integrationstheorie*, Springer, 2011.

[6] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, vol. 23 of Mathematical modelling, Springer, 2009.

[7] Krumbiegel, Neitzel, and Rösch, *Regularization for semilinear elliptic optimal control problems with pointwise state and control constraints*, Comput Optim Appl, (2010).

[8] K. Krumbiegel and A. Rösch, *A virtual control concept for state constrained optimal control problems*, Computational Optimization and Applications, 43 (2009), pp. 213–233.

[9] S. N. Kružkov, *First order quasilinear equations in several independent variables*, Math. USSR Sb., 10 (1970), pp. 217–243.

[10] C. Meyer and I. Yousept, *Regularization of state-constrained elliptic optimal control problems with nonlocal radiation interface conditions*, Computational Optimization and Applications, 44 (2007), pp. 183–212.

[11] I. Neitzel and F. Tröltzsch, *On convergence of regularization methods for nonlinear parabolic optimal control problems with control and state constraints*, Control and Cybernetics, (2008).

[12] O. A. Oleĭnik, *Discontinuous solutions of non-linear differential equations*, Amer. Math. Soc. Transl. (2), 26 (1963), pp. 95–172.

[13] S. Pfaff, *Optimal control of a hyperbolic conservation law on a bounded domain with switching controls*, Dissertation, Fachbereich Mathematik, Technische Universität Darmstadt, Germany, 2015.

[14] S. Pfaff and S. Ulbrich, *Optimal boundary control of nonlinear hyperbolic conservation laws with switched boundary data*, SIAM J. Control and Optimization, 53 (2015), pp. 1250–1277.

[15] ——, *Optimal control of nonlinear hyperbolic conservation laws by on/off-switching,*, tech. rep., Department of Mathematics, TU Darmstadt, 2015.

[16] S. Ulbrich, *Optimal control of nonlinear hyperbolic conservation laws with source terms*, Habilitation, Zentrum Mathematik, Technische Universität München, Germany, 2001.

[17] ——, *Adjoint-based derivative computations for the optimal control of discontinuous solutions of hyperbolic conservation laws*, Systems Control Lett., 48 (2003), pp. 313–328. Optimization and control of distributed systems.