# ERROR BOUNDS FOR EULER APPROXIMATION OF LINEAR-QUADRATIC CONTROL PROBLEMS WITH BANG-BANG SOLUTIONS

WALTER ALT, ROBERT BAIER, MATTHIAS GERDTS, AND FRANK LEMPIO

Abstract. We analyze the Euler discretization to a class of linear-quadratic optimal control problems. First we show convergence of order $h$ for the optimal values, where $h$ is the mesh size. Under the additional assumption that the optimal control has bang-bang structure we show that the discrete and the continuous controls coincide except on a set of measure $O(\sqrt{h})$. Under a slightly stronger assumption on the smoothness of the coefficients of the system equation we obtain an error estimate of order $O(h)$.

## 1. Introduction

We consider the following linear-quadratic control problem:

$$
\begin{aligned}
\text{(OQ)} \qquad & \min f(x,u) \\
& \text{s.t.} \\
& \dot{x}(t) = A(t)x(t) + B(t)u(t) \quad \forall' t \in [0,T]\,, \\
& x(0) = a\,, \\
& u(t) \in U \qquad\qquad\qquad\quad\ \ \forall' t \in [0,T]\,,
\end{aligned}
$$

where $f$ is a linear-quadratic cost functional defined by

$$
\begin{aligned}
f(x,u) =& \frac{1}{2} x(T)^{\mathsf{T}} Q x(T) + q^{\mathsf{T}} x(T) \\
&+ \int_0^T \tfrac{1}{2} x(t)^{\mathsf{T}} W(t) x(t) + w(t)^{\mathsf{T}} x(t) + r(t)^{\mathsf{T}} u(t) dt.
\end{aligned}
$$

Here, $u(t) \in \mathbb{R}^m$ is the control, and $x(t) \in \mathbb{R}^n$ is the state of a system at time $t$. Further $Q$ is a symmetric and positive semidefinite $n \times n$-matrix, $q \in \mathbb{R}^n$, and the functions $W\colon [0,T] \to \mathbb{R}^{n \times n}$, $w\colon [0,T] \to \mathbb{R}^n$, $r\colon [0,T] \to \mathbb{R}^m$, $A\colon [0,T] \to \mathbb{R}^{n \times n}$ $B\colon [0,T] \to R^{n \times m}$ are Lipschitz continuous. The matrices $W(t)$ are assumed to be symmetric and positive semidefinite, and the set $U \subset \mathbb{R}^m$ is defined by lower and upper bounds, i.e.,

$$U = \{u \in \mathbb{R}^m \mid b_l \le u \le b_u\}$$

with $b_l, b_u \in \mathbb{R}^m$, $b_l < b_u$, where all inequalities are to be understood componentwise.

Our aim is to derive error estimates for the Euler discretization of problem (OQ). There are some papers dealing with Euler approximations to nonlinear control

---

*Date*: July 2011, compiled: July 28, 2011.

1991 *Mathematics Subject Classification.* Primary 49J15; Secondary 49M25, 49N10, 49J30.

*Key words and phrases.* linear-quadratic optimal control, bang-bang control, discretization.

problems (see e.g. [3, 13, 15, 14, 23] and the papers cited therein). The analysis in these papers is based on the assumption that the optimal control is Lipschitz continuous. Since an optimal control for (OQ) has typically bang-bang structure this assumption is not satisfied. For bang-bang controls only simple convergence results have been obtained (see e.g. [6] and the papers cited therein).

There are also a number of articles dealing with set-valued Euler's method for nonlinear differential inclusions ([12], [32], [9], [8], [7]) which prove order of convergence equal to 1 for the approximation of the reachable set. From this fact the same order of convergence can be concluded for the approximation of the state and of the optimal value (see [30]).

Veliov [31] seems to be the only paper dealing with error estimates for control problems with control appearing linearly. In contrast to problem (OQ) he considers problems with a possibly nonlinear cost functional of Mayer type. His approach is based on Runge-Kutta methods of at least third order local consistency. In a recent paper [4] we have shown that for linear control problems with an optimal control of bang-bang structure the discrete and continuous controls coincide except on a set of measure $O(h)$, where $h$ is the mesh size of the discretization. Here we extend this result to linear-quadratic control problems. The analysis in [4] is based on the fact that for linear problems the adjoint equation does not depend on the state and can therefore be solved independently. Here we use a different approach based on a second-order condition known from the stability analysis [17] of bang-bang controls (compare also [25, 24]).

For elliptic control problems an approach similar to the one presented here has been developed recently in [10]. Errors for the controls are obtained also based on a variant of a stability condition used in the context of parameter dependent control problems in Felgenhauer [17]–[20]. Another variant of these conditions has been used in [11] in the context of bang-bang solutions for parabolic control problems.

The organization of the paper is as follows. After this introduction we define in Section 2 the Euler discretization for Problem (OQ). In Section 3 we derive error estimates for the optimal values for the discretized problems. Assuming that the optimal control is of bang-bang type, we then derive in Section 4 error estimates of order $O(\sqrt{h})$ for optimal solutions of the discretized problems. In Section 5 we use slighty stronger assumptions for the problem data in order to show structural stability of the discretized controls and to improve the error estimates for the discretized solutions to order $O(h)$. Finally, we discuss a numerical example.

We use the following notation: $\mathbb{R}^n$ is the n-dimensional Euclidean space with the inner product denoted by $\langle x, y\rangle$ and the norm $|x| = \langle x, x\rangle^{1/2}$. For an $m \times n$-matrix $B$ we denote by $\|B\| = \sup_{|z|\le 1} |Bz|$ the spectral norm. For $1 \le p < \infty$ we denote by $L^p(0,T;\mathbb{R}^n)$ the Banach space of measurable vector functions $u\colon [0,T] \to \mathbb{R}^n$ with

$$\|u\|_p = \left( \int_0^T |u(t)|^p \, dt \right)^{\frac{1}{p}} < \infty\,,$$

and $L^\infty(0,T;\mathbb{R}^n)$ is the Banach space of essentially bounded vector functions with the norm

$$\|u\|_\infty = \max_{1\le i\le n} \operatorname*{ess\,sup}_{t\in[0,T]} |u_i(t)|\,.$$

By $W^1_p(0,T;\mathbb{R}^n)$ we denote the Sobolev spaces of absolutely continuous functions

$$W^1_p(0,T;\mathbb{R}^n) = \{x \in L^p(0,T;\mathbb{R}^n) \mid \dot{x} \in L^p(0,T;\mathbb{R}^n)\}$$

with

$$\|x\|_{1,p} = \left(|x(0)|^p + \|\dot{x}\|_p^p\right)^{\frac{1}{p}}$$

for $1 \le p < \infty$ and

$$\|x\|_{1,\infty} = \max\{|x(0)|, \|\dot{x}\|_\infty\}\,.$$

We define $X = X_1 \times X_2$, $X_1 = W^1_\infty(0,T;\mathbb{R}^n)$, $X_2 = L^\infty(0,T;\mathbb{R}^m)$, and we denote by

$$\mathcal{U} = \{u \in X_2 \mid u(t) \in U\, \forall' t \in [0,T]\}$$

the set of admissible controls, and by

$$\mathcal{F} = \{(x,u) \in X_1 \times X_2 \mid u \in \mathcal{U},\ \dot{x}(t) = A(t)x(t) + B(t)u(t)\, \forall' t \in [0,T],\ x(0) = a\}$$

the feasible set of (OQ).

**Definition 1.1.** A pair $(x^*,u^*) \in \mathcal{F}$ is called a *minimizer for Problem* (OQ), if $f(x^*,u^*) \le f(x,u)$ for all $(x,u) \in \mathcal{F}$, and a *strict minimizer*, if $f(x^*,u^*) < f(x,u)$ for all $(x,u) \in \mathcal{F}$, $(x,u) \neq (x^*,u^*)$. ◇

Since the feasible set $\mathcal{F}$ is nonempty, closed, convex and bounded, and the cost functional is convex and continuous, a minimizer $(x^*,u^*) \in W^1_2(0,T;\mathbb{R}^n) \times L^2(0,T;\mathbb{R}^m)$ of this problem exists (see e.g. Ekeland/Temam [16], Chap. II, Proposition 1.2), and since $\mathcal{U}$ is bounded we have $(x^*,u^*) \in X = W^1_\infty(0,T;\mathbb{R}^n) \times L^\infty(0,T;\mathbb{R}^m)$. Moreover, the cost functional is Lipschitz continuous on $\mathcal{F}$, i.e., there is a constant $L_f$ such that

$$|f(x,u) - f(z,v)| \le L_f\left(\|x-z\|_\infty + \|u-v\|_1\right) \quad \forall (x,u),(z,v) \in \mathcal{F}. \tag{1.1}$$

An immediate consequence of the compactness of $U$, the Lipschitz continuity of $A$ and $B$ as well as the solution formula for linear differential equations, is the existence of a constant $K$ such that for any feasible control $u \in \mathcal{U}$ and the associated solution $x$ of the system equation we have with some constant $L_x$

$$\|x\|_{1,\infty} \le L_x. \tag{1.2}$$

This estimate shows that the feasible trajectories are uniformly Lipschitz with Lipschitz modulus $L_x$.

Let $(x^*,u^*) \in \mathcal{F}$ be a minimizer of (OQ). Then there exists a function $\lambda \in W^1_\infty(0,T;\mathbb{R}^n)$ such that the adjoint equation

$$-\dot{\lambda}(t) = A(t)^{\mathsf{T}}\lambda(t) + W(t)x^*(t) + w(t)\ \forall' t \in [0,T]\,, \quad \lambda(T) = Qx^*(T) + q, \tag{1.3}$$

and the minimum principle

$$[r(t)^{\mathsf{T}} + \lambda(t)^{\mathsf{T}}B(t)](u - u^*(t)) \ge 0 \quad \forall u \in U \tag{1.4}$$

hold for a.a. $t \in [0,T]$. Denoting by

$$\sigma(t) := r(t) + B(t)^{\mathsf{T}}\lambda(t) \tag{1.5}$$

the *switching function*, it is well-known that (1.4) implies for $i \in \{1,\dots,m\}$

$$u_i^*(t) = \begin{cases} b_{l,i}, & \text{if } \sigma_i(t) > 0, \\ b_{u,i}, & \text{if } \sigma_i(t) < 0, \\ \text{undetermined}, & \text{if } \sigma_i(t) = 0. \end{cases} \tag{1.6}$$

*Remark.* Since $\lambda$ satisfies the adjoint equation and $W$, $w$, $r$, $A$, $B$ are Lipschitz continuous, $\dot{\lambda}$ is bounded and hence $\lambda$ is Lipschitz continuous, which implies that $\sigma$ is also Lipschitz continuous. $\diamond$

## 2. Euler Approximation

Given a natural number $N$, let $h_N = T/N$ be the mesh size. We approximate the space $X_2$ of controls by functions in the subspace $X_{2,N} \subset X_2$ of piecewise constant functions represented by their values $u(t_j) = u_j$ at the gridpoints $jh_N$, $j = 0, 1, \ldots, N-1$. Further, we approximate state and adjoint state variables by functions in the subspace $X_{1,N} \subset X_1$ of continuous, piecewise linear functions represented by their values $x(t_j) = x_j$, $\lambda(t_j) = \lambda_j$ at the gridpoints $jh_N$, $j = 0, 1, \ldots, N$. Then the Euler discretization of (OQ) is given by

$$\begin{aligned}
(\mathrm{OQ})_N \quad & \min_{(x,u)\in X_{1,N}\times X_{2,N}} \quad f_N(x,u) \\
& \text{s.t.} \\
& x_{j+1} = x_j + h_N\,[A(t_j)x_j + B(t_j)u_j],\ j = 0, 1, \ldots, N-1\,, \\
& x_0 = a\,, \\
& u_j \in U, \qquad\qquad\qquad j = 0, 1, \ldots, N-1\,,
\end{aligned}$$

where $f_N$ is the linear-quadratic cost functional defined by

$$f_N(x,u) = \frac{1}{2}x_N^\mathsf{T} Q x_N + q^\mathsf{T} x_N + h_N \sum_{j=0}^{N-1}\left[\frac{1}{2}x_j^\mathsf{T} W(t_j)x_j + w(t_j)^\mathsf{T} x_j + r(t_j)^\mathsf{T} u_j\right].$$

By $\mathcal{F}_N$ we denote the feasible set of $(\mathrm{OQ})_N$.

**Definition 2.1.** A pair $(x_h^*, u_h^*) \in \mathcal{F}_N$ is called a *minimizer* $(\mathrm{OQ})_N$, if $f_N(x_h^*, u_h^*) \le f_N(x_h, u_h)$ for all $(x_h, u_h) \in \mathcal{F}_N$, and a *strict minimizer*, if $f_N(x_h^*, u_h^*) < f_N(x_h, u_h)$ for all $(x_h, u_h) \in \mathcal{F}_N$, $(x_h, u_h) \ne (x_h^*, u_h^*)$. $\diamond$

Again, since $U$ is compact there exists a constant $L_x$ independent of $N$ such that for any feasible control $u_h \in \mathcal{U}$ and the associated solution $x$ of the discrete system equation seen as a continuous, piecewise linear function we have

$$|\dot{x}_h(t)| \le L_x \quad \forall' t \in [0,T]\,, \tag{2.1}$$

which shows that the discrete feasible trajectories are uniformly Lipschitz with Lipschitz modulus $L_x$ independent from $h_N$, where w.l.o.g. $L_x$ is the same constant as in (1.2).

Compactness of $U$ further implies that Problem $(\mathrm{OQ})_N$ has a solution $(x_h^*, u_h^*)$, and for any solution there exists a continuous, piecewise linear multiplier $\lambda_h \in X_{1,N}$ such that the discrete adjoint equation

$$-\frac{\lambda_{h,j+1} - \lambda_{h,j}}{h_N} = A(t_j)^\mathsf{T}\lambda_{h,j+1} + W(t_j)x_{h,j}^* + w(t_j)\,, \quad j = 0, \ldots, N-1\,, \tag{2.2}$$

with end condition

$$\lambda_{h,N} = Q x_{h,N}^* + q, \tag{2.3}$$

and the discrete minimum principle

$$(r(t_j) + \lambda_{h,j+1}^\mathsf{T} B(t_j))(u - u_{h,j}^*) \ge 0 \quad \forall u \in U\,, \quad j = 0, \ldots, N-1\,, \tag{2.4}$$

are satisfied.

By $\sigma_h\colon [0,t_{N-1}] \to \mathbb{R}^m$ we denote the discrete switching function, the continuous and piecewise linear function defined by the values

$$\sigma_h(t_j) := r(t_j) + B(t_j)^{\mathsf{T}}\lambda_{h,j+1}, \quad j = 0,\dots,N-1. \tag{2.5}$$

From (2.4) we obtain for $i = 1,\dots,m$, $j = 0,\dots,N-1$,

$$u^*_{h,i}(t_j) = \begin{cases} b_{l,i}, & \text{if } \sigma_{h,i}(t_j) > 0,\\ b_{u,i}, & \text{if } \sigma_{h,i}(t_j) < 0,\\ \text{undetermined}, & \text{if } \sigma_{h,i}(t_j) = 0. \end{cases} \tag{2.6}$$

## 3. Error Estimates for Optimal Values

Without assuming a special structure of the optimal controls we can derive error estimates of order 1 for the optimal values. To this end we need some auxiliary results. For a function $z\colon [0,T] \to \mathbb{R}$ of bounded variation and $s_1, s_2 \in [0,T]$, $s_1 < s_2$, we denote by $\mathsf{V}_{s_1}^{s_2} z$ the total variation of $z$ on $[s_1,s_2]$.

**Lemma 3.1.** *Suppose that* $(x,u) \in \mathcal{F}$ *and* $u$ *has bounded variation. Then there esists* $(x_h,u_h) \in \mathcal{F}_N$ *such that*

$$\|u-u_h\|_1 \le h_N \mathsf{V}_0^T u, \quad \|u-u_h\|_2 \le \sqrt{h_N}\mathsf{V}_0^T u, \tag{3.1}$$

*and*

$$\|x_h - x\|_\infty \le c_1\, h_N \mathsf{V}_0^T \dot{x} \le (c_2 + c_3 \mathsf{V}_0^T u) h_N, \tag{3.2}$$

*where* $c_1$, $c_2$, $c_3$ *are constants independent of* $N$. ◇

*Proof.* Let $u_h$ be the piecewise constant function defined by the values $u(t_j)$, $j = 0,\dots,N-1$. Then $u_h \in \mathcal{U}$. Since for $s \in [t_j,t_{j+1}]$

$$|u(s) - u(t_j)| \le |u(t_{j+1}) - u(s)| + |u(s) - u(t_j)| \le \mathsf{V}_{t_j}^{t_{j+1}} u,$$

we have

$$\|u-u_h\|_1 = \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} |u(s) - u(t_j)|\, ds \le \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \mathsf{V}_{t_j}^{t_{j+1}} u \le h_N \mathsf{V}_0^T u,$$

which shows the first estimate in (3.1). For the $L^2$-norm we have

$$\begin{aligned} \|u-u_h\|_2^2 &= \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} |u(s) - u(t_j)|^2\, ds \le \sum_{j=0}^{N-1} h_N \left(\mathsf{V}_{t_j}^{t_{j+1}} u\right)^2 \\ &\le \mathsf{V}_0^T u \sum_{j=0}^{N-1} h_N \mathsf{V}_{t_j}^{t_{j+1}} u = h_N \left(\mathsf{V}_0^T u\right)^2, \end{aligned}$$

which shows the second estimate in (3.1).

Let $x_h$ be the solution of the discrete system equation of $(\mathrm{OQ})_N$ for $u = u_h$. Then $(x_h,u_h) \in \mathcal{F}_N$ and $x_h$ is the Euler approximation of $x$. Since $u$ has bounded variation and $x$ is the solution of the system equation, $\dot{x}$ has bounded variation. By Sendov/Popov [28, Theorem 6.1] (see also [28, (7) on p. 10]) this implies

$$\max_{1\le j\le N} |x_h(t_j) - x(t_j)| \le 2T\exp(T\|A\|_\infty) h_N \mathsf{V}_0^T \dot{x}\,. \tag{3.3}$$

From this one easily obtains the first estimate in (3.2) (compare [4], Lemma 2.2). The variation of $\dot{x}$ can be estimated by the variation of the right hand side of the system equation. If we denote by $L_A$, resp. $L_B$, the Lipschitz modulus of $A(\cdot)$, resp. $B(\cdot)$, then a simple calculation shows that for $t, s \in [0, T]$

$$\begin{aligned}|\dot{x}(t) - \dot{x}(s)| \le{}& L_A \|x\|_\infty |t-s| + \|A(\cdot)\|_\infty |x(t) - x(s)| \\ &+ L_B \|u\|_\infty |t-s| + \|A(\cdot)\|_\infty |u(t) - u(s)|\,.\end{aligned}$$

By (1.2) and the boundedness of $U$ we further obtain with some constants $L_x$, $L_u$ independent of $N$

$$\mathsf{V}_0^T \dot{x} \le (L_A \|x\|_\infty + L_x \|A(\cdot)\|_\infty + L_B L_u) T + \|A(\cdot)\|_\infty \mathsf{V}_0^T u\,,$$

which implies the second estimate in (3.2). □

*Remark.* In many applications the optimal control $u^*$ is a piecewise Lipschitz continuous function. In this case $u^*$ has bounded variation. ◇

**Lemma 3.2.** *Suppose that* $(x_h, u_h) \in \mathcal{F}_N$. *Then there exists a function* $z$, *such that* $(z, u_h) \in \mathcal{F}$ *and*

$$\|z - x_h\|_\infty \le c\, h_N \tag{3.4}$$

*with a constant* $c$ *independent of* $N$ *and the choice of* $(x_h, u_h) \in \mathcal{F}_N$. ◇

*Proof.* By assumption $u_h \in \mathcal{U}$. Let $z$ be the solution of the system equation of (OQ) for $u = u_h$. Then $(z, u_h) \in \mathcal{F}$ and $x_h$ solves the differential equation (remember that $u_h(t) = u_h(t_j)$ for $t \in ]t_j, t_{j+1}[$)

$$\dot{x}_h = A(t_j) x_h(t_j) + B(t_j) u_h(t_j) = A(t) x_h(t) + B(t) u_h(t) + y(t) \quad \forall' t \in [0, T],$$

where

$$y(t) = A(t_j) x_h(t_j) - A(t) x_h(t) + (B(t_j) - B(t)) u_h(t), \quad t \in [t_j, t_{j+1}[\,.$$

Since $u_h$ is bounded and $y(t_j) = 0$, the functions $A$, $B$, are Lipschitz-continuous and the feasible trajectories are Lipschitz uniformly with respect to $h_N$ by (2.1), it follows that

$$|y(t)| \le c_1 h_N \quad \forall' t \in [0, T]$$

with a constant $c_1$ independent of $N$ and the choice of $(x_h, u_h)$. This together with $\dot{x}_h(t) - \dot{z}(t) = y(t)$ implies

$$|x_h(t) - z(t)| \le \int_0^t |\dot{x}_h(s) - \dot{z}(s)| ds = \int_0^t |y(s)| ds \le c_1 T h_N$$

for $t \in [t_j, t_{j+1}[$ which proves (3.4). □

**Lemma 3.3.** *Suppose that* $(x_h, u_h) \in \mathcal{F}_N$. *Then*

$$|f(x_h, u_h) - f_N(x_h, u_h)| \le c\, h_N \tag{3.5}$$

*with a constant* $c$ *independent of* $N$ *and the choice of* $(x_h, u_h) \in \mathcal{F}_N$. ◇

*Proof.* It follows from (2.1) and the boundedness of $U$ that there are constants $c_x$, $c_u$ independent of $N$ such that

$$\|x_h\|_\infty \le c_x, \ \|u_h\|_\infty \le c_u \quad \forall (x_h, u_h) \in \mathcal{F}_N\,. \tag{3.6}$$

By the definition of $f$ and $f_N$ we have

$$f(x_h,u_h)-f_N(x_h,u_h)=\sum_{j=0}^{N-1}\int_{t_j}^{t_{j+1}}\left[\frac{1}{2}I_1(t)+I_2(t)+I_3(t)\right]dt\,, \tag{3.7}$$

where

$$\begin{aligned}I_1(t)&=x_h(t)^{\mathsf T}W(t)x_h(t)-x_h(t_j)^{\mathsf T}W(t_j)x_h(t_j)\,,\\ I_2(t)&=w(t)^{\mathsf T}x_h(t)-w(t_j)^{\mathsf T}x_h(t_j)\,,\\ I_3(t)&=r(t)^{\mathsf T}u_h(t)-r(t_j)^{\mathsf T}u_h(t_j)=(r(t)-r(t_j))^{\mathsf T}u_h(t_j)\end{aligned}$$

for $t\in[t_j,t_{j+1}[$. Since

$$\begin{aligned}I_1(t)&=x_h(t)^{\mathsf T}W(t)x_h(t)-x_h(t_j)^{\mathsf T}W(t)x_h(t_j)\\ &\quad+x_h(t_j)^{\mathsf T}W(t)x_h(t_j)-x_h(t_j)^{\mathsf T}W(t_j)x_h(t_j)\\ &=(x_h(t)+x_h(t_j))^{\mathsf T}\,W(t)\,(x_h(t)-x_h(t_j))+x_h(t_j)^{\mathsf T}\,(W(t)-W(t_j))\,x_h(t_j)\,,\end{aligned}$$

we get by (2.1) and (3.6)

$$|I_1(t)|\le 2c_x\|W(t)\|\,L_xh_N+c_x^2L_wh_N\,,$$

where $L_w$ is the Lipschitz modulus of $W$. Similar results can be easily obtained for $I_2(t)$ and $I_3(t)$. Together with (3.7) this implies the assertion. □

We can now derive an estimate for the optimal values of solutions. By approximation results for reachable sets (see [12, 30, 31]), the assumption on the bounded variation of the optimal control in the following theorem could be weakened by demanding only bounded variation and Lipschitz continuity of a corresponding set-valued right-hand side. To avoid additional notations, we include a direct proof for the simpler result needed here (compare [1]).

**Theorem 3.4.** *Let $(x^*,u^*)\in\mathcal F$ be a solution of* (OQ) *such that $u^*$ has bounded variation. Then for any solution $(x_h^*,u_h^*)\in\mathcal F_N$ of $(\mathrm{OQ})_N$ we have*

$$|f_N(x_h^*,u_h^*)-f(x^*,u^*)|\le c\,h_N\quad\forall t\in[0,T] \tag{3.8}$$

*with a constant $c$ independent of $N$ and the choice of $x_h^*,u_h^*$.* ◇

*Proof.* By Lemma 3.1 and the boundedness of $\bigvee_0^T u^*$ there exists $(x_h,u_h)\in\mathcal F_N$ such that

$$\|x_h-x^*\|_\infty\le c_1\,h_N,\quad\|u_h-u^*\|_1\le c_2\,h_N, \tag{3.9}$$

where $c_1$, $c_2$ are constants independent of $N$. Let $(x_h^*,u_h^*)\in\mathcal F_N$ be any solution of $(\mathrm{OQ})_N$. Since $f_N(x_h^*,u_h^*)\le f_N(x_h,u_h)$ we obtain

$$0\le f_N(x_h,u_h)-f_N(x_h^*,u_h^*)=f_N(x_h,u_h)-f(x^*,u^*)+f(x^*,u^*)-f_N(x_h^*,u_h^*),$$

and therefore

$$\begin{aligned}f_N(x_h^*,u_h^*)-f(x^*,u^*)&\le f_N(x_h,u_h)-f(x^*,u^*)\\ &\le f_N(x_h,u_h)-f(x_h,u_h)+f(x_h,u_h)-f(x^*,u^*).\end{aligned}$$

By (3.5), (1.1) and (3.9) this implies

$$f_N(x_h^*,u_h^*)-f(x^*,u^*)\le c_3h_N+L_f(c_1+c_2)\,h_N \tag{3.10}$$

with a constant $c_3$ independent of $N$ and of $x_h,u_h$.

On the other hand, by Lemma 3.2 there exists $z^*$ such that $(z^*, u_h^*) \in \mathcal{F}$ and

$$\|z^* - x_h^*\|_\infty \le c_4\, h_N, \tag{3.11}$$

where $c_4$ is a constant independent of $N$ and the choice of $x_h^*, u_h^*$. Since $f(x^*, u^*) \le f(z^*, u_h^*)$ we obtain

$$0 \le f(z^*, u_h^*) - f(x^*, u^*) = f(z^*, u_h^*) - f_N(x_h^*, u_h^*) + f_N(x_h^*, u_h^*) - f(x^*, u^*),$$

and therefore

$$\begin{aligned} f(x^*, u^*) - f_N(x_h^*, u_h^*) &\le f(z^*, u_h^*) - f_N(x_h^*, u_h^*) \\ &\le f(z^*, u_h^*) - f(x_h^*, u_h^*) + f(x_h^*, u_h^*) - f_N(x_h^*, u_h^*). \end{aligned}$$

By (3.5), (1.1) and (3.11) this implies

$$f(x^*, u^*) - f_N(x_h^*, u_h^*) \le L_f c_4 h_N + c_3 h_N.$$

Together with (3.10) we obtain (3.8). □

*Remark.* The constant $c$ in (3.8) depends on the variation of $u^*$, but is independent of $N$. Since we assume in the following that $\bigvee_0^T u^*$ is bounded, we suppress the explicit dependence of constants on $\bigvee_0^T u^*$. ◇

## 4. Error estimates for bang-bang solutions

**4.1. A lower minorant for minimal values.** The convergence analysis of Euler discretizations is usually based on a second-order optimality condition (compare e.g. [15], [23]). We show in the following that for Problem (OQ) a similar condition holds, if the optimal control is of bang-bang type. To this end we assume that (compare [17]–[20], [4])

(A1) There exists a solution $(x^*, u^*) \in \mathcal{F}$ of (OQ) such that the set $\Sigma$ of zeros of the components $\sigma_i$, $i = 1, \ldots, m$, of the switching function $\sigma$ defined by (1.5) is finite and $0, T \notin \Sigma$, i.e., $\Sigma = \{s_1, \ldots, s_l\}$ with $0 < s_1 < \ldots < s_l < T$.

*Remark.* If $0, T \notin \Sigma$ then $s_1 > t_1$ and $s_l < t_{N-1}$ for sufficiently large $N$. Assumption (A1) implies bounded variation of $u^*$. ◇

Let $\mathcal{I}(s_j) := \{1 \le i \le m : \sigma_i(s_j) = 0\}$ be the set of active indices for the components of the switching function. In order to get a bang-type structure for the discrete optimal controls we need an additional assumption:

(A2) There exist $\bar\sigma > 0$, $\bar\tau > 0$ such that

$$|\sigma_i(\tau)| \ge \bar\sigma |\tau - s_j|$$

for all $j \in \{1, \ldots, l\}$, $i \in \mathcal{I}(s_j)$, and all $\tau \in [s_j - \bar\tau, s_j + \bar\tau]$, and

$$\sigma_i(s_j - \bar\tau)\sigma_i(s_j + \bar\tau) < 0,$$

i.e., $\sigma_i$ changes sign in $s_j$.

Assumptions (A1)–(A2) imply uniqueness of the optimal control $u^*$ (see the remark following (4.12)).

For $0 < \delta \le \bar\tau$ we define

$$I(\delta) = \bigcup_{1 \le j \le l} [s_j - \delta, s_j + \delta]. \tag{4.1}$$

Let $i \in \{1, \ldots, m\}$ be arbitrary, and let

$$\Sigma_i = \{\tau_1, \ldots, \tau_{l_i}\} \subset \Sigma \quad \text{with} \quad 0 < \tau_1 < \ldots < \tau_{l_i} < T$$

be the set of zeros of $\sigma_i$ and

$$I_-(\delta) = \bigcup_{j=1,\ldots,l_i} \left[\tau_j - \bar{\tau}, \tau_j + \bar{\tau}\right], \quad I_+(\delta) = [0,T] \setminus I_-(\delta). \tag{4.2}$$

Since $\sigma_i$ is Lipschitz there exists

$$0 < \sigma_{i,\min} = \min_{t \in [0,T] \setminus I_+(\bar{\tau})} \left|\sigma_i(t)\right|. \tag{4.3}$$

We choose $0 < \bar{\delta} \leq \bar{\tau}$ such that

$$\bar{\delta}\bar{\sigma} \leq \min_{1 \leq i \leq m} \sigma_{i,\min}. \tag{4.4}$$

Then by (A2) for any $0 < \delta \leq \bar{\delta}$ and arbitrary $i \in \{1, \ldots, m\}$ we have

$$|\sigma_i(t)| \geq \delta\bar{\sigma} \quad \forall t \in [0,T] \setminus I(\delta). \tag{4.5}$$

The following result is extracted from the proof of Lemma 3.3 in Felgenhauer [17] and forms an important tool for the forthcoming analysis. For the reader's convenience the proof is included.

**Lemma 4.1.** *Let $(x^*, u^*)$ be a minimizer for Problem* (OQ)*, and let the switching function $\sigma$ be defined by* (1.5)*. If Assumptions* (A1)–(A2) *are satisfied, then there are constants $\alpha, \gamma, \bar{\delta} > 0$ such that for any feasible pair $(x, u)$*

$$\int_0^T \sigma(t)^\mathsf{T}(u(t) - u^*(t))dt \geq \alpha \left\|u - u^*\right\|_1^2 \tag{4.6}$$

*if $\|u - u^*\|_1 \leq 2\gamma\bar{\delta}$, and*

$$\int_0^T \sigma(t)^\mathsf{T}(u(t) - u^*(t))dt \geq \alpha \left\|u - u^*\right\|_1 \tag{4.7}$$

*if $\|u - u^*\|_1 > 2\gamma\bar{\delta}$.* ◇

*Proof.* Let $(x, u) \in \mathcal{F}$ be arbitrary. Since by the minimum principle (1.4) the signs of $\sigma_i(t)$ and $u_i(t) - u_i^*(t)$ coincide it follows from (4.5) that

$$\begin{aligned} J &= \int_0^T \sigma(t)^\mathsf{T}(u(t) - u^*(t))dt \geq \int_{[0,T]\setminus I(\delta)} \sigma(t)^\mathsf{T}(u(t) - u^*(t))dt \\ &= \int_{[0,T]\setminus I(\delta)} \sum_{i=1}^m |\sigma_i(t)||u_i(t) - u_i^*(t)|\, dt \geq \delta\bar{\sigma} \sum_{i=1}^m \int_{[0,T]\setminus I(\delta)} |u_i(t) - u_i^*(t)|\, dt. \end{aligned}$$

Since for $1 \leq i \leq m$,

$$|u_i(t) - u_i^*(t)| \leq b_{u,i} - b_{l,i} \quad \forall t \in [0,T],$$

we have

$$\sum_{i=1}^m \int_{I(\delta)} |u_i(t) - u_i^*(t)|\, dt \leq \gamma\delta,$$

where $\gamma = 2lm \max_{1 \leq i \leq m}(b_{u,i} - b_{l,i})$, so that

$$J \geq \delta\bar{\sigma}(\|u - u^*\|_1 - \gamma\delta). \tag{4.8}$$

We choose $\delta = \min\{\bar{\delta}, \frac{1}{2\gamma}\|u - u^*\|_1\}$. If $\delta = \bar{\delta}$, i.e. if $\|u - u^*\|_1 > 2\gamma\bar{\delta}$, we obtain

$$J \geq \frac{\bar{\delta}}{2}\bar{\sigma}\|u - u^*\|_1 .$$

If $\delta = \frac{1}{2\gamma}\|u - u^*\|_1$ (note that in this case $\delta$ depends on $u$ and is not a constant), i.e. if $\|u - u^*\|_1 \le 2\gamma\bar{\delta}$, we obtain

$$J \ge \frac{\bar{\sigma}}{4\gamma}\|u - u^*\|_1^2,$$

which proves the assertion. $\square$

Lemma 4.1 implies a quadratic minorant for the minimal values of Problem (OQ) in a sufficiently small $L^1$-neighbourhood, and a linear minorant outside this neighborhood.

**Theorem 4.2.** *Let* $(x^*, u^*)$ *be a minimizer for Problem* (OQ). *If Assumptions* (A1)–(A2) *are satisfied, then there are constants* $\alpha, \gamma, \bar{\delta} > 0$ *such that for any feasible pair* $(x, u)$

$$f(x,u) - f(x^*,u^*) \ge \alpha\, \|u - u^*\|_1^2 \tag{4.9}$$

*if* $\|u - u^*\|_1 \le 2\gamma\bar{\delta}$, *and*

$$f(x,u) - f(x^*,u^*) \ge \alpha\, \|u - u^*\|_1 \tag{4.10}$$

*if* $\|u - u^*\|_1 > 2\gamma\bar{\delta}$. ◇

*Proof.* Let $(x, u)$ be feasible for problem (OQ), let $(x^*, u^*)$ be optimal, and let $\lambda$ be the adjoint state. Defining $z = x - x^*$, $v = u - u^*$ we have

$$\begin{aligned} f(x,u) - f(x^*,u^*) &= (Qx^*(T) + q)^\mathsf{T} z(T) + \frac{1}{2} z(T)^\mathsf{T} Q z(T) \\ &\quad + \int_0^T (x^*(t)^\mathsf{T} W(t) + w(t)^\mathsf{T}) z(t) + r(t)^\mathsf{T} v(t)\, dt + \frac{1}{2}\int_0^T z(t)^\mathsf{T} W(t) z(t)\, dt \\ &\ge (Qx^*(T) + q)^\mathsf{T} z(T) + \int_0^T (x^*(t)^\mathsf{T} W(t) + w(t)^\mathsf{T}) z(t) + r(t)^\mathsf{T} v(t)\, dt, \end{aligned}$$

since $Q$ and $W(\cdot)$ are positive semidefinite. From $\lambda(T) = Qx^*(T) + q$ it follows

$$f(x,u) - f(x^*,u^*) \ge \lambda(T)^\mathsf{T} z(T) + \int_0^T (x^*(t)^\mathsf{T} W(t) + w(t)^\mathsf{T}) z(t) + r(t)^\mathsf{T} v(t)\, dt.$$

Since $z(0) = 0$ we further obtain

$$\begin{aligned} f(x,u) - f(x^*,u^*) &\ge \int_0^T (x^*(t)^\mathsf{T} W(t) + w(t)^\mathsf{T}) z(t) + r(t)^\mathsf{T} v(t)\, dt + \lambda(T)^\mathsf{T} z(T) \\ &= \int_0^T (x^*(t)^\mathsf{T} W(t) + w(t)^\mathsf{T}) z(t) + r(t)^\mathsf{T} v(t)\, dt \\ &\quad + \int_0^T \dot{z}(t)^\mathsf{T} \lambda(t)\, dt + \int_0^T z(t)^\mathsf{T} \dot{\lambda}(t)\, dt\,. \end{aligned}$$

Since $\dot z(t) = A(t)z(t) + B(t)v(t)$ and $\lambda$ solves the adjoint equation, this implies

$$\begin{aligned} f(x,u) - f(x^*,u^*) &= \int_0^T (x^*(t)^\mathsf{T} W(t) + w(t)^\mathsf{T}) z(t) + r(t)^\mathsf{T} v(t)\, dt \\ &\quad + \int_0^T [A(t)z(t) + B(t)v(t)]^\mathsf{T} \lambda(t)\, dt \\ &\quad - \int_0^T z(t)^\mathsf{T} \left[A(t)^\mathsf{T}\lambda(t) + W(t)x^*(t) + w(t)\right]\, dt \\ &= \int_0^T [\lambda(t)^\mathsf{T} B(t) + r(t)^\mathsf{T}] v(t) dt = \int_0^T \sigma(t)^\mathsf{T} v(t) dt. \end{aligned}$$

The assertion now follows from Lemma 4.1. $\square$

Since $x^*$ solves the state equation for $u^*$ and $x$ solves the state equation for $u$, we have

$$\dot x(t) - \dot x^*(t) = A(t)(x(t) - x^*(t)) + B(t)(u(t) - u^*(t)) \quad \forall' t \in [0,T],$$

and $x(0) - x^*(0) = 0$. This implies

$$\|x - x^*\|_{1,1} \le c\,\|u - u^*\|_1$$

with some constant $c$. Together with (4.9), (4.10) we obtain with some constant $\tilde\alpha > 0$

$$f(x,u) - f(x^*,u^*) \ge \tilde\alpha(\|u - u^*\|_1^2 + \|x - x^*\|_{1,1}^2) \tag{4.11}$$

for any feasible pair $(x,u)$ with $\|u - u^*\|_1 \le 2\gamma\bar\delta$, and

$$f(x,u) - f(x^*,u^*) \ge \tilde\alpha(\|u - u^*\|_1 + \|x - x^*\|_{1,1}) \tag{4.12}$$

for any feasible pair $(x,u)$ with $\|u - u^*\|_1 > 2\gamma\bar\delta$.

*Remark.* (compare [17], Theorem 2.2) These estimates also imply uniqueness of the solution of (OQ). If $(x,u) \in \mathcal{F}$ is an arbitrary solution of (OQ), then $f(x,u) = f(x^*,u^*)$. By (4.11) resp. (4.12) we then obtain $(x,u) = (x^*,u^*)$. $\diamond$

**4.2. Hölder type error estimates.** Based on the estimate (4.11) for the optimal values we now prove error estimates for the optimal controls. To this end we proceed similar to [2] (compare also [26]) and prove Hölder type error estimates first.

As above we denote by $(x^*,u^*)$ a solution of Problem (OQ) and by $(x_h^*,u_h^*)$ a solution of Problem $(\mathrm{OQ})_N$. Suppose that Assumptions (A1), (A2) are satisfied. Let $z^*$ be the solution of the system equation for $u = u_h^*$. Then $(z^*,u_h^*) \in \mathcal{F}$ and by Lemma 3.2

$$\|z^* - x_h^*\| \le c_1 h_N \tag{4.13}$$

with a constant $c_1$ independent of $N$. By (4.11) and (4.12) we have with some constant $\tilde\alpha$ independent of $N$

$$f(z^*,u_h^*) - f(x^*,u^*) \ge \tilde\alpha(\|u_h^* - u^*\|_1^2 + \|z^* - x^*\|_{1,1}^2) \tag{4.14}$$

if $\|u_h^* - u^*\|_1 \le 2\gamma\bar\delta$, and

$$f(z^*,u_h^*) - f(x^*,u^*) \ge \tilde\alpha(\|u_h^* - u^*\|_1 + \|x - x^*\|_{1,1}) \tag{4.15}$$

if $\|u_h^* - u^*\|_1 > 2\gamma\bar{\delta}$. As in the proof of Lemma 3.1 let $\hat{u}_h$ be the piecewise constant function defined by the values $\hat{u}_h(t_j) = u^*(t_j)$, $j = 0, \dots, N-1$. Then $\hat{u}_h \in \mathcal{U}$, and by (3.1)

$$\|u^* - \hat{u}_h\|_1 \le h_N \mathsf{V}_0^T u^*. \tag{4.16}$$

Let $\hat{x}_h$ be the solution of the discrete system equation of $(\mathrm{OQ})_N$ for $u_j = \hat{u}_{h,j}$. Then $(\hat{x}_h, \hat{u}_h) \in \mathcal{F}_N$, hence $f(\hat{x}_h, \hat{u}_h) \ge f(x_h^*, u_h^*)$, and (see (3.3))

$$\max_{1 \le j \le N} |\hat{x}_h(t_j) - x^*(t_j)| \le 2T \exp(T\|A\|_\infty) h_N \mathsf{V}_0^T \dot{x}^*. \tag{4.17}$$

Estimating $\mathsf{V}_0^T \dot{x}^*$ according to the proof of Lemma 3.1 and using the boundedness of $\mathsf{V}_0^T u^*$ this implies (compare [4], Lemma 2.2)

$$\|x^* - \hat{x}_h\|_\infty \le c_2\, h_N \tag{4.18}$$

with a constant $c_2$ independent of $N$. Now using (1.1), (4.18), (4.16) the left hand side of (4.14), (4.15) can be estimated by

$$\begin{aligned} f(z^*, u_h^*) - f(x^*, u^*) &= f(z^*, u_h^*) - f(x_h^*, u_h^*) + f(x_h^*, u_h^*) - f(x^*, u^*) \\ &\le f(z^*, u_h^*) - f(x_h^*, u_h^*) + f(\hat{x}_h, \hat{u}_h) - f(x^*, u^*) \le c_3\, L_f h_N \end{aligned}$$

with a constant $c_3$ independent of $N$. By (4.14), (4.15) this implies

$$\|u_h^* - u^*\|_1 \le c_4 \, \max\{h_N, h_N^{\frac{1}{2}}\}$$

with a constant $c_4$ independent of $N$. Therefore, if $N$ is sufficiently large, we have $\|u_h^* - u^*\|_1 \le 2\gamma\bar{\delta}$, and by (4.14) we finally obtain the following result, where $\lambda_h$ denotes the continuous, piecewise linear function defined by $\lambda_h(t_j) = \lambda_{h,j}$.

**Theorem 4.3.** *Let $(x^*, u^*)$ be a solution of Problem* (OQ) *for which Assumptions* (A1), (A2) *are satisfied. Then for sufficiently large $N$ any minimizer $(x_h^*, u_h^*)$ of Problem $(\mathrm{OQ})_N$ can be estimated by*

$$\|u_h^* - u^*\|_1 \le c_u h_N^{\frac{1}{2}}, \quad \|x_h^* - x^*\|_\infty \le c_x h_N^{\frac{1}{2}}, \tag{4.19}$$

*further, the associated multipliers can be estimated by*

$$\|\lambda_h - \lambda\|_\infty \le c_\lambda h_N^{\frac{1}{2}} \tag{4.20}$$

*with constants $c_u$, $c_x$, $c_\lambda$ independendent of $N$.*

*Proof.* It remains to show (4.20). To this end we prove that for sufficiently large $N$

$$\|\lambda_h - \lambda\|_\infty \le c_\lambda \left(h_N + |x_h^*(T) - x^*(T)|\right)$$

with a constant $c_\lambda$ independent of $N$. We denote by $\Phi$ the matrix function forming the fundamental solution of the adjoint system

$$-\dot{\Phi}(t) = A(t)^{\mathsf{T}} \Phi(t) \quad \forall t \in [0, T], \quad \Phi(T) = I.$$

Further we denote by $\mu_h$ the solution of the adjoint equation

$$-\dot{\mu}(t) = A(t)^{\mathsf{T}} \mu(t) + W(t) x^*(t) + w(t) \quad \forall' t \in [0, T] \tag{4.21}$$

with end condition

$$\mu_h(T) = Q x_h^*(T) + q. \tag{4.22}$$

Then we have

$$\mu_h(t) - \lambda(t) = \Phi(t) Q (x_h^*(T) - x^*(T)).$$

This implies

$$\|\mu_h - \lambda\|_\infty \le c_1 |x_h^*(T) - x^*(T)| \tag{4.23}$$

with some constant $c_1$ independent of $N$. Furthermore, we have

$$\begin{aligned}\dot\mu_h(t) = &- A(t)^\mathsf{T}\Phi(t)(Qx_h^*(T)+q) - A(t)^\mathsf{T}\Phi(t)\int_t^T \Phi(s)^{-1}\left[W(s)x^*(s)+w(s)\right]ds\\ &- W(t)x^*(t) - w(t).\end{aligned}$$

With $c_2 = \mathsf{V}_0^T(-A(\cdot)^\mathsf{T}\Phi(\cdot))$ and

$$c_3 = \mathsf{V}_0^T\left[-A(\cdot)^\mathsf{T}\Phi(\cdot)\int_\cdot^T \Phi(s)^{-1}\left[W(s)x^*(s)+w(s)\right]ds - W(\cdot)x^*(\cdot) - w(\cdot)\right]$$

we have

$$\mathsf{V}_0^T\dot\mu_h \le c_2|Qx_h^*(T)+q| + c_3.$$

Together with (3.6) it follows, that $\dot\mu_h$ has bounded variation uniformly with respect to $h$. By [28, 1.3 (7) and Theorem 6.1] this implies that

$$\max_{0\le j\le N} |\nu_h(t_j) - \mu_h(t_j)| \le 2T\exp(T\|A\|_\infty)h_N \mathsf{V}_0^T\dot\mu_h, \tag{4.24}$$

where $\nu_h$ is the Euler discretization of equation (4.21) with end condition (4.22). Further, it can be easily shown that for sufficiently large $N$

$$\max_{0\le j\le N}|\lambda_h(t_j) - \nu_h(t_j)| \le c_\nu\, h_N$$

holds with a constant $c_\nu$ independent of $N$. Together with (4.23) and (4.24) this implies

$$\max_{0\le j\le N}|\lambda_h(t_j) - \lambda(t_j)| \le c_4|x_h^*(T) - x^*(T)| + c_1\, h_N$$

with a constant $c_4$ independent of $N$. The assertion now easily follows (see e.g. the proof of Lemma 2.2 in [4]). □

Theorem 4.3 immediately implies an error estimate for the switching function. For the simple proof we refer to [4], Theorem 2.3.

**Corollary 4.4.** *Let the assumptions of Theorem* 4.3 *be satified. Further let* $\sigma$ *be defined by* (1.5), *and let* $\sigma_h$ *be defined by* (2.5). *Then for sufficiently large* $N$

$$\max_{t\in[0,t_{N-1}]}|\sigma_h(t) - \sigma(t)| \le c_\sigma\, h_N^{\frac12} \tag{4.25}$$

*with a constant* $c_\sigma$ *independent of* $N$. ◇

We now show that the discrete optimal controls are bang-bang except on a set of measure $\le \kappa h_N^{\frac12}$ with a constant $\kappa$ independent of $N$. To this end we use the following result. A proof for $\beta = 1$ can be found in [4].

**Theorem 4.5.** *Let Assumptions* (A1), (A2) *be satisfied, and suppose that for sufficiently large* $N$

$$\max_{t\in[0,t_{N-1}]}|\sigma_h(t) - \sigma(t)| \le c_\sigma\, h_N^{\beta} \tag{4.26}$$

*with a constant* $c_\sigma$ *independent of* $N$ *and* $\beta > 0$. *Then there exists a constant* $\tilde\kappa$ *independent of* $N$ *such that for sufficiently large* $N$ *any discrete optimal control* $u_h^*$ *coincides with* $u^*$ *except on a set of measure* $\le \tilde\kappa h_N^{\beta}$. ◇

*Proof.* Let $i \in \{1, \ldots, m\}$ be arbitrary, and let $\sigma_{i,\min}$ be defined by (4.3). Then (4.26) implies

$$|\sigma_{h,i}(t)| \geq |\sigma_i(t)| - c_\sigma h_N^\beta \geq \sigma_{i,\min} - c_\sigma h_N^\beta \quad \forall t \in I_+(\bar{\tau}),$$

where $I_+(\bar{\tau})$ is defined by (4.2). This shows that

$$|\sigma_{h,i}(t)| \geq \frac{1}{2}\sigma_{i,\min} > 0 \quad \forall t \in I_+(\bar{\tau}),$$

if we choose $N$ sufficiently large such that

$$h_N^\beta \leq \frac{\sigma_{i,\min}}{2c_\sigma}.$$

For $\tau \in [\tau_j - \bar{\tau}, \tau_j + \bar{\tau}]$, $j \in \{1, \ldots, l_i\}$, it follows by (A2) and (4.26) that

$$|\sigma_{h,i}(\tau)| \geq |\sigma_i(\tau)| - c_\sigma h_N^\beta \geq \bar{\sigma}|\tau - \tau_j| - c_\sigma h_N^\beta.$$

Therefore, $\sigma_{h,i}(\tau) \neq 0$ if

$$|\tau - \tau_j| > \frac{c_\sigma}{\bar{\sigma}} h_N^\beta.$$

Let $\iota \in \{1, \ldots, N-1\}$ with $\tau_j \in [t_\iota, t_{\iota+1}]$. We choose $k \in \mathbb{N}$ to be the smallest number such that

$$k > \frac{c_\sigma}{\bar{\sigma}} h_N^{\beta-1}.$$

Then

$$\begin{aligned} t_{\iota+k+1} - \tau_j &\geq t_{\iota+k+1} - t_{\iota+1} = k h_N > \frac{c_\sigma}{\bar{\sigma}} h_N^\beta, \\ \tau_j - t_{\iota-k} &\geq t_\iota - t_{\iota-k} = k h_N > \frac{c_\sigma}{\bar{\sigma}} h_N^\beta, \end{aligned}$$

and

$$\frac{c_\sigma}{\bar{\sigma}} h_N^{\beta-1} < k \leq \frac{c_\sigma}{\bar{\sigma}} h_N^{\beta-1} + 1.$$

Defining $k_j^+ := \iota + k + 1$, $k_j^- := \iota - k$, we have

$$t_{k_j^+} - t_{k_j^-} = (2k+1) h_N \leq \Big(2\frac{c_\sigma}{\bar{\sigma}} h_N^{\beta-1} + 3\Big) h_N = \Big(2\frac{c_\sigma}{\bar{\sigma}} + 3 h_N^{1-\beta}\Big) h_N^\beta$$

and therefore

$$t_{k_j^+} - t_{k_j^-} \leq \Big(2\frac{c_\sigma}{\bar{\sigma}} + 3T^{1-\beta}\Big) h_N^\beta = \kappa\, h_N^\beta \tag{4.27}$$

with a constant $\kappa$ independent of $N$. For sufficently large $N$ we then have

$$[t_{k_j^-}, t_{k_j^+}] \subset [\tau_j - \bar{\tau}, \tau_j + \bar{\tau}],$$

and it follows from (4.2) that $|\sigma_{h,i}(t)| > 0$ on $[t_{k_j^+}, \tau_j + \bar{\tau}]$ and on $[\tau_j - \bar{\tau}, t_{k_j^-}]$. Thus, defining

$$I_- := \bigcup_{j=1,\ldots,l_i} [t_{k_j^-}, t_{k_j^+}] \subset I_-(\bar{\tau}), \quad I_+ := [0, T] \setminus I_- \supset I_+(\bar{\tau}),$$

we have shown that

$$|\sigma_{h,i}(t)| > 0 \quad \forall t \in I_+. \tag{4.28}$$

By (2.6) this implies for any discrete optimal control $u_h^*$ that

$$u_{h,i}^*(t) = u_i^*(t) \quad \forall t \in I_+,$$

i.e. the continuous and the discrete optimal control coincide on $I_+$. Since the measure of $I_-$ is bounded by

$$\tilde{\kappa} = \kappa \sum_{i=1}^{m} l_i,$$

the theorem is proved. □

By Corollary 4.4 and Theorem 4.5 applied with $\beta = \frac{1}{2}$ we immediately obtain the following result.

**Theorem 4.6.** *Let Assumptions* (A1), (A2) *be satisfied. Then there exists a constant* $\tilde{\kappa}$ *independent of* $N$ *such that for sufficiently large* $N$ *any discrete optimal control* $u_h^*$ *coincides with* $u^*$ *except on a set of measure* $\leq \tilde{\kappa} h_N^{\frac{1}{2}}$. ◇

## 5. Structural stability and improved error estimates

Let again $i \in \{1, \ldots, m\}$ be arbitrary and let $\tau_j \in \Sigma_i$ be a zero of $\sigma_i$. Then by (4.28), $\sigma_{h,i}$ has no zero in $I_+$ and at least one zero in $[t_{k_j^-}, t_{k_j^+}]$. We show that this zero is unique, i.e. $u_h^*$ has the same structure as $u^*$, if we we replace Assumption (A2) by the following slightly stronger assumption:

(A3) The matrix function $B$ is differentiable, $\dot{B}$ is Lipschitz continuous, and there exists $\bar{\sigma} > 0$ such that

$$\min_{1 \leq j \leq l} \min_{i \in \mathcal{I}(s_j)} \{|\dot{\sigma}_i(s_j)|\} \geq 2\bar{\sigma}\,.$$

Since $\lambda$ satisfies the adjoint equation, $\dot{\lambda}$ is Lipschitz continuous, and therefore, if (A3) holds, $\dot{\sigma}$ is also Lipschitz continuous. Therefore, $\bar{\tau} > 0$ can be chosen such that for all $i \in \mathcal{I}(s_j)$

$$|\dot{\sigma}_i(\tau)| \geq \bar{\sigma} \text{ on } [s_j - \bar{\tau}, s_j + \bar{\tau}] \quad \forall i \in \mathcal{I}(s_j), \tag{5.1}$$

which shows that Assumption (A3) implies (A2).

The function $\sigma_h$ defined by (2.5) is differentiable on $]t_j, t_{j+1}[\,$, $j = 0, \ldots, N-1$. For $t = t_j$ we define $\dot{\sigma}_h(t) = \frac{1}{h_N}(\sigma_h(t_{j+1}) - \sigma_h(t_j))$, $j = 0, \ldots, N-1$. Based on Assumption (A3) one easily obtains an error estimate for the derivative of the switching function. The proof is almost identical to that of Theorem 2.6 in [4] and hence omitted.

**Theorem 5.1.** *Let Assumptions* (A1), (A3) *be satisfied. Let* $\sigma$ *be defined by* (1.5), *and let* $\sigma_h$ *be defined by* (2.5). *Then for sufficiently large* $N$

$$|\dot{\sigma}_h(t) - \dot{\sigma}(t)|_\infty \leq \tilde{c}_\sigma\, h_N^{\frac{1}{2}} \quad \forall t \in [0, t_{N-1}] \tag{5.2}$$

*with a constant* $\tilde{c}_\sigma$ *independent of* $N$. ◇

We now show that the error estimates of the last section can be improved, if (A3) holds. To this end let $i \in \mathcal{I}(s_j)$, i.e. $\sigma_i(s_j) = 0$. From (5.1) and (5.2) we obtain for sufficiently large $N$

$$|\dot{\sigma}_{h,i}(\tau)| \geq \tfrac{1}{2}\bar{\sigma} \text{ on } [s_j - \bar{\tau}, s_j + \bar{\tau}]. \tag{5.3}$$

This implies that $\sigma_{h,i}$ is strictly increasing or decreasing on $[s_j - \bar{\tau}, s_j + \bar{\tau}]$. Since $\sigma_{h,i}(s_j - \bar{\tau})\sigma_{h,i}(s_j + \bar{\tau}) \neq 0$, it follows that $\sigma_{h,i}$ has exactly one zero $s_{h,j}$ in $[s_j - \bar{\tau}, s_j + \bar{\tau}]$. This shows that $\sigma_h$ has the same structure as $\sigma$ (finitely many isolated zeros of its components). Note that this does not imply uniqueness of the discrete optimal controls, since it may happen that one of the zeros is a discretization point and therefore $\sigma_{h,i}(t_j) = 0$ for some $i, j$.

By (4.28) it further follows that $s_{h,j} \in [t_{k_j^-}, t_{k_j^+}]$, and by (4.27) we get the error estimate

$$|s_j - s_{h,j}| \le \kappa h_N^{\frac{1}{2}}, \quad j = 1, \dots, l\,, \tag{5.4}$$

for the zeros of the components of $\sigma$ and $\sigma_h$.

**Theorem 5.2.** *Let Assumptions* (A1), (A3) *be satisfied. Then for sufficiently large $N$ the discrete switching function $\sigma_h$ has the same structure as $\sigma$, i.e., the components of $\sigma_h$ have $l$ zeros, and the error estimates* (5.4) *hold with a constant $\kappa$ independent of $N$.* ◇

We assume that Assumptions (A1), (A3) are satisfied, so that, as shown above, $\sigma_h$ has the same structure as $\sigma$. Let $(x^*, u^*)$ be the optimal solution for Problem (OQ) and $(x_h^*, u_h^*)$ an optimal solution for Problem $(\mathrm{OQ})_N$. As in (4.1) we define for $0 < \delta \le \bar{\tau}$

$$I_h(\delta) = \bigcup_{1 \le j \le l} [s_{h,j} - \delta, s_{h,j} + \delta]. \tag{5.5}$$

Let $i \in \{1, \dots, m\}$ be arbitrary, and let

$$\Sigma_{h,i} = \{\tau_{h,1}, \dots, \tau_{h,l_i}\} \quad \text{with} \quad 0 < \tau_{h,1} < \ldots < \tau_{h,l_i} < T$$

be the set of zeros of $\sigma_{h,i}$ and

$$I_{h,-}(\delta) = \bigcup_{j=1,\dots,l_i} [\tau_{h,j} - \bar{\tau}, \tau_{h,j} + \bar{\tau}]\,, \quad I_{h,+}(\delta) = [0,T] \setminus I_{h,-}(\delta). \tag{5.6}$$

Since $\sigma_{h,i}$ is Lipschitz, there exists

$$0 < \sigma_{h,i,\min} = \min_{t \in [0,T] \setminus I_{h,+}(\bar{\tau})} |\sigma_{h,i}(t)|\,. \tag{5.7}$$

It then follows from the continuity of $\sigma_{h,i}$, (4.3) and Corollary 4.4 that for sufficiently large $N$

$$|\sigma_{h,i}(t)| \ge \sigma_{h,i,\min} \ge \frac{1}{2}\sigma_{i,\min} \quad \forall t \in [0,T] \setminus I_{h,+}(\bar{\tau})\,.$$

Moreover, by (5.3) and the Lipschitz continuity of $\dot{\sigma}_i$ we have

$$|\dot{\sigma}_{h,i}(\tau)| \ge \tfrac{1}{2}\bar{\sigma} \text{ on } [s_{h,j} - \bar{\tau}, s_{h,j} + \bar{\tau}]$$

for all $s_{h,j} \in \Sigma_{h,i}$, which implies that for $0 < \delta \le \bar{\tau}$

$$|\sigma_{h,i}(t)| \ge \frac{1}{2}\bar{\sigma}\delta \quad \forall t \notin [s_{h,j} - \delta, s_{h,j} + \delta]\,. \tag{5.8}$$

We choose $0 < \bar{\delta} \le \bar{\tau}$ such that

$$\bar{\delta}\bar{\sigma} \le \min_{1 \le i \le m} \sigma_{i,\min}.$$

Note that $\bar{\delta}$ is independent of $N$. Then it follows that for $0 \le \delta \le \bar{\delta}$

$$|\sigma_{h,i}(t)| \ge \frac{1}{2}\bar{\sigma}\delta \quad \forall t \in [0,T] \setminus I_h(\delta)\,.$$

For $\tau_\iota \in \Sigma_{h,i}$ we define

$$k_1 = k_1(\iota) = \max\{j \mid t_j \le \tau_\iota - \delta\}\,, \quad k_2 = k_2(\iota) = \min\{j \mid t_j \ge \tau_\iota + \delta\}\,.$$

Then $t_{k_2} - t_{k_1} \leq 2(\delta + h_N)$. Further we define

$$I_i(\delta) = \bigcup_{\iota=1,\dots,l_i} \{0 \leq j \leq N \mid k_1(\iota) \leq j \leq k_2(\iota)\}$$

and

$$D_i(\delta) = \bigcup_{\iota=1,\dots,l_i} [t_{k_1(\iota)}, t_{k_2(\iota)}] \,.$$

Then the measure $m(D_i(\delta))$ can be estimated by $m(D_i(\delta)) \leq 2l_i(\delta + h_N)$, and since $[t_{k_1(\iota)}, t_{k_2(\iota)}] \supset [\tau_\iota - \delta, \tau_\iota + \delta]$ we have by (5.8)

$$|\sigma_{h,i}(t)| \geq \frac{1}{2}\bar{\sigma}\delta \quad \forall t \in [0,T] \setminus D_i(\delta).$$

Now let $(x,u) \in \mathcal{F}_N$ and $i \in \{1,\dots,m\}$ be arbitrary. Then the discrete minimum principle (2.4) implies that the signs of $\sigma_{h,i}(t)$ and $u_i(t) - u^*_{h,i}(t)$ coincide at the grid points. Therefore

$$\begin{aligned} J_{N,i} &= h_N \sum_{j=0}^{N-1} \sigma_{h,i}(t_j)(u_i(t_j) - u^*_{h,i}(t_j)) \geq h_N \sum_{j \notin I_i(\delta)} \sigma_{h,i}(t_j)(u_i(t_j) - u^*_{h,i}(t_j)) \\ &\geq \frac{1}{4}\delta\bar{\sigma}h_N \sum_{j \notin I_i(\delta)} |u_i(t_j) - u^*_{h,i}(t_j)|. \end{aligned}$$

Moreover, we have

$$\begin{aligned} h_N \sum_{j \in I_i(\delta)} |u_i(t_j) - u^*_{h,i}(t_j)| &\leq \max_{1 \leq i \leq m} (b_{u,i} - b_{l,i}) \sum_{j \in I_i(\delta)} h_N \\ &\leq \max_{1 \leq i \leq m} (b_{u,i} - b_{l,i})\, m(D_i(\delta)) \leq \gamma_i(\delta + h_N), \end{aligned}$$

where

$$\gamma_i = 2l_i \max_{1 \leq i \leq m} (b_{u,i} - b_{l,i}) \,.$$

Therefore

$$J_{N,i} \geq \frac{1}{4}\delta\bar{\sigma}h_N \sum_{j=0}^{N-1} |u_i(t_j) - u^*_{h,i}(t_j)| - \frac{1}{4}\delta\bar{\sigma}\gamma_i(\delta + h_N)$$

and

$$\begin{aligned} J_N &= \sum_{i=1}^{m} J_{N,i} = h_N \sum_{j=0}^{N-1} \sigma_h(t_j)^{\mathsf{T}}(u(t_j) - u^*_h(t_j)) \\ &= h_N \sum_{i=1}^{m}\sum_{j=0}^{N-1} \sigma_{h,i}(t_j)(u_i(t_j) - u^*_{h,i}(t_j)) \\ &\geq \frac{1}{4}\delta\bar{\sigma}h_N \sum_{i=1}^{m}\sum_{j=0}^{N-1} |u_i(t_j) - u^*_{h,i}(t_j)| - \frac{1}{4}\delta\bar{\sigma}\sum_{i=1}^{m}\gamma_i(\delta + h_N)\,. \end{aligned} \tag{5.9}$$

Defining $\gamma = \sum_{i=1}^m \gamma_i$ we obtain

$$J_N \geq \frac{1}{4}\delta\bar{\sigma}\left(\|u - u^*_h\|_1 - \gamma(\delta + h_N)\right). \tag{5.10}$$

We now make the special choice $u = \hat{u}_h \in X_{2,N}$, where $\hat{u}$ is defined by the values $\hat{u}_h(t_j) = u^*(t_j)$, $j = 0, \dots, N-1$ (compare the proof of Lemma 3.1). Then we have $\hat{u}_h \in \mathcal{U}$, and by (3.1)

$$\|u^* - \hat{u}_h\|_1 \le h_N \mathsf{V}_0^T u^*.$$

Together with (4.19) this implies

$$\|u_h^* - \hat{u}_h\|_1 \le \|u_h^* - u^*\|_1 + \|u^* - \hat{u}_h\|_1 \le c_u h_N^{\frac{1}{2}} + h_N \mathsf{V}_0^T u^* \le \bar{\delta}$$

for sufficiently large $N$. With

$$\delta = \frac{1}{2\gamma}\|\hat{u}_h - u_h^*\|_1$$

we obtain from (5.10)

$$\begin{aligned} J_N &\ge \frac{\bar{\sigma}}{8\gamma}\|\hat{u}_h - u_h^*\|_1 \left(\|\hat{u}_h - u_h^*\|_1 - \frac{1}{2}\|\hat{u}_h - u_h^*\|_1 - \gamma h_N\right) \\ &= \frac{\bar{\sigma}}{16\gamma}\|\hat{u}_h - u_h^*\|_1 \left(\|\hat{u}_h - u_h^*\|_1 - 2\gamma h_N\right) . \end{aligned}$$

Now we consider two cases. If

$$\|\hat{u}_h - u_h^*\|_1 \le 4\gamma h_N \, , \tag{5.11}$$

we have a discrete error estimate of order 1. Otherwise we have

$$\|\hat{u}_h - u_h^*\|_1 - 2\gamma h_N > 2\gamma h_N > \frac{1}{2}\|\hat{u}_h - u_h^*\|_1$$

and therefore

$$J_N \ge \frac{\bar{\sigma}}{32\gamma}\|\hat{u}_h - u_h^*\|_1^2. \tag{5.12}$$

We can now adapt known proof techniques (see e.g. [27, 22, 5]) to derive an upper bound for $J_N$. By Assumption (A1) the optimal control $u^*$ is piecewise continuous. Therefore the minimum principle (1.4) holds for all $t \in [0, T]$ (see e.g. [21]). With $t = t_j$ and $u = u_h^*(t_j)$ we obtain

$$\sigma(t_j)^\mathsf{T}(u_h^*(t_j) - u^*(t_j)) = \sigma(t_j)^\mathsf{T}(u_h^*(t_j) - \hat{u}_h(t_j)) \ge 0\, , \quad j = 0, \dots, N-1\, .$$

Together with (5.9) we obtain

$$\begin{aligned} J_N &\le h_N \sum_{j=0}^{N-1} (\sigma_h(t_j) - \sigma(t_j))^\mathsf{T}(\hat{u}_h(t_j) - u_h^*(t_j)) \\ &= h_N \sum_{j=0}^{N-1} (\lambda_h(t_{j+1}) - \lambda(t_j))^\mathsf{T} B(t_j)(\hat{u}_h(t_j) - u_h^*(t_j)) \\ &= J_{N,1} + J_{N,2}\, , \end{aligned}$$

where

$$\begin{aligned} J_{N,1} &= h_N \sum_{j=0}^{N-1} (\lambda(t_{j+1}) - \lambda(t_j))^\mathsf{T} B(t_j)(\hat{u}_h(t_j) - u_h^*(t_j)), \\ J_{N,2} &= h_N \sum_{j=0}^{N-1} (\lambda_h(t_{j+1}) - \lambda(t_{j+1}))^\mathsf{T} B(t_j)(\hat{u}_h(t_j) - u_h^*(t_j))\, . \end{aligned}$$

The term $J_{N,1}$ can be estimated by

$$J_{N,1} \le h_N^2 L_\lambda \|B\| \sum_{j=0}^{N-1} |\hat{u}_h(t_j) - u_h^*(t_j)| = h_N L_\lambda \|B\| \, \|\hat{u}_h - u_h^*\|_1. \tag{5.13}$$

In order to estimate $J_{N,2}$ let $z_h$ be the solution of the discrete system equation for $u = \hat{u}_h$, i.e.,

$$z_h(t_{j+1}) = z_h(t_j) + h_N \left[A(t_j) z_h(t_j) + B(t_j)\hat{u}_h(t_j)\right] , \quad j = 0, 1, \ldots, N-1 \, ,$$

with initial condition $z_h(0) = a$, and let $\mu_h$ be the solution of the associated discrete adjoint equation, i.e.,

$$-\frac{\mu_h(t_{j+1}) - \mu_h(t_j)}{h_N} = A(t_j)^{\mathsf{T}} \mu_h(t_{j+1}) + W(t_j) z_h(t_j) + w(t_j)$$

for $j = 0, \ldots, N-1$ with end condition

$$\mu_h(T) = Q z_h(T) + q. \tag{5.14}$$

Then

$$\|z_h - x^*\|_\infty \le c\, h_N \, , \quad \|\mu_h - \lambda\|_\infty \le c\, h_N \, , \tag{5.15}$$

where $c$ is a constant independent of $N$ (compare the proof of Lemma 3.1), and $J_{N,2} = J_{N,3} + J_{N,4}$ with

$$J_{N,3} = h_N \sum_{j=0}^{N-1} (\lambda_h(t_{j+1}) - \mu_h(t_{j+1}))^{\mathsf{T}} B(t_j)(\hat{u}_h(t_j) - u_h^*(t_j)) \, ,$$
$$J_{N,4} = h_N \sum_{j=0}^{N-1} (\mu_h(t_{j+1}) - \lambda(t_{j+1}))^{\mathsf{T}} B(t_j)(\hat{u}_h(t_j) - u_h^*(t_j)) \, .$$

Using (5.15), $J_{N,4}$ can be estimated by

$$J_{N,4} \le c\, h_N^2 \|B\| \sum_{j=0}^{N-1} |\hat{u}_h(t_j) - u_h^*(t_j)| = c\, h_N \|B\| \, \|\hat{u}_h - u_h^*\|_1. \tag{5.16}$$

We now show $J_{N,3} \le 0$. By the definition of $z_h$ we have

$$\begin{aligned} h_N B(t_j)(\hat{u}_h(t_j) - u_h^*(t_j)) = & - h_N A(t_j)(z_h(t_j) - x_h^*(t_j)) \\ & + z_h(t_{j+1}) - z_h(t_j) - (x_h^*(t_{j+1}) - x_h^*(t_j)) \end{aligned}$$

for $j = 0, \ldots, N-1$. Using this, the term $J_{N,3}$ can be written in the form

$$\begin{aligned} J_{N,3} = & -h_N \sum_{j=0}^{N-1} (\lambda_h(t_{j+1}) - \mu_h(t_{j+1}))^{\mathsf{T}} A(t_j)(z_h(t_j) - x_h^*(t_j)) \\ & + \sum_{j=0}^{N-1} (\lambda_h(t_{j+1}) - \mu_h(t_{j+1}))^{\mathsf{T}} \left[z_h(t_{j+1}) - z_h(t_j) - (x_h^*(t_{j+1}) - x_h^*(t_j))\right] . \end{aligned}$$

By the definition of $\mu_h$ we have

$$\begin{aligned} -h_N A(t_j)^{\mathsf{T}} (\lambda_h(t_{j+1}) - \mu_h(t_{j+1})) = & \lambda_h(t_{j+1}) - \lambda_h(t_j) - (\mu_h(t_{j+1}) - \mu_h(t_j)) \\ & + h_N W(t_j)(x_h^*(t_j) - z_h(t_j)) \end{aligned}$$

for $j = 0, \dots, N-1$. Using this, we further obtain

$$\begin{aligned} J_{N,3} = & \sum_{j=0}^{N-1} \left[\lambda_h(t_{j+1}) - \lambda_h(t_j) - (\mu_h(t_{j+1}) - \mu_h(t_j))\right]^{\mathsf{T}} (z_h(t_j) - x_h^*(t_j)) \\ & + h_N \sum_{j=0}^{N-1} (x_h^*(t_j) - z_h(t_j))^{\mathsf{T}} W(t_j)(z_h(t_j) - x_h^*(t_j)) \\ & + \sum_{j=0}^{N-1} (\lambda_h(t_{j+1}) - \mu_h(t_{j+1}))^{\mathsf{T}} \left[z_h(t_{j+1}) - z_h(t_j) - (x_h^*(t_{j+1}) - x_h^*(t_j))\right]. \end{aligned}$$

By the end conditions (2.3) and (5.14) this implies

$$\begin{aligned} J_{N,3} &= (\lambda_h(T) - \mu_h(T))^{\mathsf{T}} (z_h(T) - x_h^*(T)) \\ &\quad + h_N \sum_{j=0}^{N-1} (x_h^*(t_j) - z_h(t_j))^{\mathsf{T}} W(t_j)(z_h(t_j) - x_h^*(t_j)) \\ &= (x_h^*(T) - z_h(T))^{\mathsf{T}} Q (z_h(T) - x_h^*(T)) \\ &\quad + h_N \sum_{j=0}^{N-1} (x_h^*(t_j) - z_h(t_j))^{\mathsf{T}} W(t_j)(z_h(t_j) - x_h^*(t_j)). \end{aligned}$$

Since the matrices $W(t_j)$, $j = 0, \dots, N$, and $Q$ are positive semidefinite, this shows $J_{N,3} \le 0$. Together with (5.13) we obtain

$$J_N = J_{N,1} + J_{N,2} = J_{N,1} + J_{N,3} + J_{N,4} \le J_{N,1} + J_{N,4} \le \tilde{c}\, h_N \, \|\hat{u}_h - u_h^*\|_1 \tag{5.17}$$

with some constant $\tilde{c}$ independent of $N$. We can now state a first order error estimate for the discrete solutions improving the results of Theorem 4.3 under the stronger assumption (A3).

**Theorem 5.3.** *Let $(x^*, u^*)$ be a solution of Problem* (OQ) *for which Assumptions* (A1), (A3) *are satisfied. Then for sufficiently large $N$ any minimizer $(x_h^*, u_h^*)$ of Problem* $(\text{OQ})_N$ *can be estimated by*

$$\|u_h^* - u^*\|_1 \le c_u h_N, \quad \|x_h^* - x^*\|_\infty \le c_x h_N, \tag{5.18}$$

*further, the associated multipliers can be estimated by*

$$\|\lambda_h - \lambda\|_\infty \le c_\lambda h_N \tag{5.19}$$

*with constants $c_u$, $c_x$, $c_\lambda$ independendent of $N$.*

*Proof.* If (5.11) holds, then by (3.1) we have

$$\|u_h^* - u^*\|_1 \le \|u_h^* - \hat{u}_h\|_1 + \|\hat{u}_h - u^*\|_1 \le 4\gamma h_N + h_N \mathsf{V}_0^T u^*,$$

i.e., the estimate (5.18) is satisfied with $c_u = 4\gamma + \mathsf{V}_0^T u^*$. Otherwise it follows from (5.12) and (5.17) that

$$\|\hat{u}_h - u_h^*\|_1^2 \le \frac{32\gamma}{\bar{\sigma}} J_N \le \frac{32\gamma}{\bar{\sigma}} h_N \tilde{c}\, \|\hat{u}_h - u_h^*\|_1 .$$

Dividing both sides by $\|\hat{u}_h - u_h^*\|_1$ it follows that the estimate (5.18) is satisfied with $c_u = \frac{32\gamma}{\bar{\sigma}}\tilde{c} + \mathsf{V}_0^T u^*$. The estimates for $x_h^*$ and $\lambda_h$ can now be derived as in the proof of Theorem 4.3. □

Theorem 5.3 immediately implies a first order error estimate for the switching function (compare Corollary 4.4).

**Corollary 5.4.** *Let the assumptions of Theorem* 4.3 *be satified. Further let* $\sigma$ *be defined by* (1.5), *and let* $\sigma_h$ *be defined by* (2.5). *Then for sufficiently large* $N$

$$\max_{t\in[0,t_{N-1}]} |\sigma_h(t)-\sigma(t)| \le c_\sigma\, h_N$$

*with a constant* $c_\sigma$ *independent of* $N$. ◇

Analogously to Theorems 4.6 and 5.2, applying Theorem 4.5 with $\beta = 1$ we finally obtain the following result.

**Theorem 5.5.** *Let Assumptions* (A1), (A3) *be satisfied. Then there exists a constant* $\tilde\kappa$ *independent of* $N$ *such that for sufficiently large* $N$ *any discrete optimal control* $u_h^*$ *coincides with* $u^*$ *except on a set of measure* $\le \tilde\kappa h_N$. *Moreover, the error estimates*

$$|s_j - s_{h,j}| \le \kappa h_N, \quad j = 1,\dots,l\,, \tag{5.20}$$

*hold for the zeros of the components of* $\sigma$ *and* $\sigma_h$ *with a constant* $\kappa$ *independent of* $N$. ◇

## 6. Numerical results

**Example 6.1** (Rocket car)**.**

$$\text{(OQ1)}\quad \begin{aligned} &\min \frac{1}{2}(x_1(5)^2 + x_2(5)^2)\\ &\text{s.t.}\\ &\dot x_1(t) = x_2(t), \quad \dot x_2(t) = u(t) \qquad \forall' t\in[0,5]\,,\\ &x_1(0) = 6\,,\ x_2(0) = 1,\\ &-1 \le u(t) \le 1 \qquad\qquad\qquad\ \ \forall' t\in[0,5]\,. \end{aligned}$$

The optimal control is

$$u^*(t) = \begin{cases} -1, & 0\le t<\tau,\\ +1, & \tau<t\le 5, \end{cases}$$

where $\tau$ is computed in the following. From the system equations we obtain

$$x_2^*(t) = \begin{cases} -t+1, & 0\le t\le\tau,\\ t-2\tau+1, & \tau\le t\le 5, \end{cases}$$

and

$$x_1^*(t) = \begin{cases} -\frac12 t^2 + t + 6, & 0\le t\le\tau,\\ \frac12 t^2 - 2\tau t + t + \tau^2 + 6, & \tau\le t\le 5. \end{cases}$$

Especially we obtain $x_1^*(5) = \tau^2 - 10\tau + 23.5$, $x_2^*(5) = -2\tau + 6$. Since

$$A(t) = \begin{pmatrix} 0 & 1\\ 0 & 0\end{pmatrix}, \quad B(t) = \begin{pmatrix} 0\\ 1\end{pmatrix},$$

the adjoint equations are

$$\dot\lambda_1(t) = 0\,, \quad \lambda_1(5) = x_1^*(5),$$

and

$$-\dot\lambda_2(t) = \lambda_1(t)\,, \quad \lambda_2(5) = x_2^*(5)$$

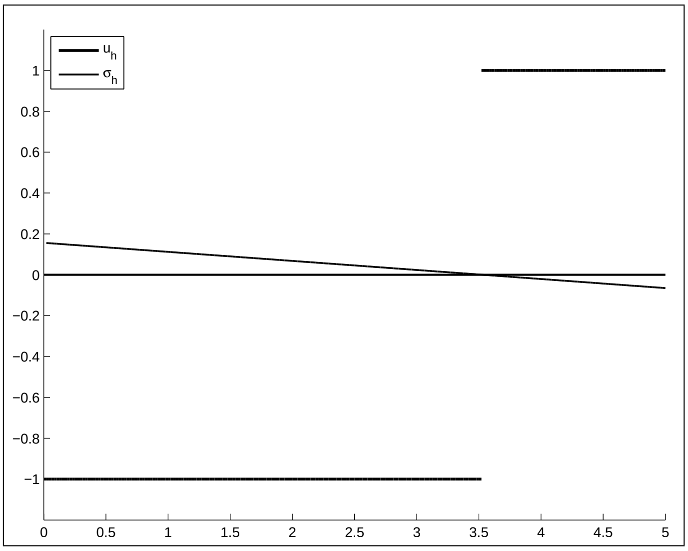

Figure 1. Rocket car

with the solutions

$$\lambda_1(t) \equiv \tau^2 - 10\tau + 23.5$$

and

$$\lambda_2(t) = -(\tau^2 - 10\tau + 23.5)t + 5\tau^2 - 52\tau + 123.5\,.$$

Since $\tau$ is a zero of $\sigma(t) = \lambda_2(t)$ we must have $\sigma(\tau) = \lambda_2(\tau) = 0$, i.e.,

$$-\tau^3 + 15\tau^2 - 75.5\tau + 123.5 = 0.$$

This implies $\tau \approx 3.5174292$.

Fig. 1 shows the discrete optimal control $u_h^*$ and the discrete switching function $\sigma_h$ for $N = 240$.

Table 1. Rocket car

| $N$ | lower bound | upper bound |
|---|---|---|
| 10 | 4.0 | 4.5 |
| 20 | 3.75 | 4.0 |
| 50 | 3.5 | 3.6 |
| 100 | 3.5 | 3.55 |
| 200 | 3.525 | 3.55 |
| 400 | 3.5125 | 3.5250 |

Table 1 shows the bounds of the discretization interval, where the discrete switching function changes sign, for different values of $N$. The results confirm the error estimates (5.20).

**Acknowledgement.** This work together with [4] was partially supported by the Hausdorff Research Institute for Mathematics, Bonn, within the HIM Junior Semester Program "Computational Mathematics" from February to April 2008. The authors would like to thank Ursula Felgenhauer for very helpful discussions, which improved the results of Sections 3 and 4.

## References

[1] W. Alt, "On the approximation of infinite optimization problems with an application to optimal control problems," *Appl. Math. Optim.* **12** (1984), 15–27.

[2] W. Alt, "Local stability of solutions to differentiable optimization problems in Banach spaces," *J. Optim. Theory Appl.* **70** (1991), 443–466.

[3] W. Alt, "Discretization and Mesh-Independence of Newton's Method for Generalized Equations," *in: A. V. Fiacco (ed.), Mathematical Programming with Data Perturbations V, Lecture Notes in Pure and Applied Mathematics 195, Marcel Dekker, 1997, 1–30.*

[4] W. Alt, R. Baier, M. Gerdts, F. Lempio, "Approximations of linear control problems with bang-bang solutions," *Optimization*, DOI: 10.1080/02331934.2011.568619 (2011).

[5] W. Alt, N. Bräutigam, "Finite-Difference discretizations of quadratic control problems governed by ordinary elliptic differential equations," *Comp. Optim. Appl.* **43** (2009), 133–150.

[6] W. Alt, U. Mackenroth, "Convergence of finite element approximations to state constrained convex parabolic boundary control problems," *SIAM J. Control Optim.* **27** (1989), 718–736.

[7] R. Baier, I. A. Chahma, F. Lempio, "Stability and convergence of Euler's method for state-constrained differential inclusions," *SIAM J. Optim.* **18** (2007), 1004–1026.

[8] W. J. Beyn, J. Rieger, "Numerical fixed grid methods for differential inclusions," *Computing* **81** (2007), 91–106.

[9] I. A. Chahma, "Set-valued discrete approximation of state-constrained differential inclusions," *Bayreuth. Math. Schr.* **67** (2003), 3–162.

[10] K. Deckelnick, M. Hinze, "A note on the approximation of elliptic control problems with bang-bang controls," *Comp. Optim. Appl.*, DOI: 10.1007/s10589-010-9365-z (2010).

[11] V. Dhamo, F. Tröltzsch, "Some aspects of reachability for parabolic boundary control problems with control constraints," *Comp. Optim. Appl.*, DOI: 10.1007/s10589-009-9310-1 (2010).

[12] A. L. Dontchev, E. M. Farkhi, "Error estimates for discretized differential inclusions," *Computing* **41** (1989), 349–358.

[13] A. L. Dontchev, W. W. Hager, "Lipschitzian stability in nonlinear control and optimization," *SIAM J. Control Optim.* **31** (1993), 569–603.

[14] A. L. Dontchev, W. W. Hager, "The Euler approximation in state constrained optimal control," *Math. Comp.* **70** (2001), 173–203.

[15] A. L. Dontchev, W. W. Hager, K. Malanowski, "Error bounds for Euler approximation of a state and control constrained optimal control problem," *Numer. Funct. Anal. Optim.* **21** (2000), 653–682.

[16] I. Ekeland, R. Temam, "Convex Analysis and Variational Problems," *North Holland, Amsterdam–Oxford, 1976.*

[17] U. Felgenhauer, "On stability of bang-bang type controls," *SIAM J. Control Optim.* **41** (2003), 1843–1867.

[18] U. Felgenhauer, "The shooting approach in analyzing bang-bang extremals with simultaneous control switches," *Control Cybernet.* **37** (2008), 307–327.

[19] U. Felgenhauer, "Directional sensitivity differentials for parametric bang-bang control problems," *in: I. Lirkov et al. (ed.), Lecture Notes Comp. Sci., Vol. 5910, Springer 2010, 264–271.*

[20] U. Felgenhauer, L. Poggiolini, G. Stefani, "Optimality and stability result for bang-bang optimal controls with simple and double switch behaviour," *Control Cybernet.* **38** (2009), 1305–1325.

[21] M. R. Hestenes, "Calculus of Variations and Optimal Control theory," *Robert E. Krieger Publ. Co., 1980.*

[22] M. Hinze, "A Variational Discretization Concept in Control Constrained Optimization: The Linear-Quadratic Case," *Comp. Optim. Appl.* **30** (2005), 45–61.

[23] K. Malanowski, C. Büskens, H. Maurer, "Convergence of Approximations to Nonlinear Optimal Control Problems," *Mathematical Programming with Data Perturbations V* (1997), 253–284.

[24] H. Maurer, C. Büskens, J.-H. R. Kim, C. Y. Kaya, "Optimization methods for the verification of second order sufficient conditions for bang-bang controls," *Optimal Control Appl. Methods* **26** (2005), 129–156.

[25] H. Maurer, N. P. Osmolovskii, "Second order sufficient conditions for time optimal bang-bang control," *SIAM J. Control Optim.* **42** (2004), 2239–2263.

[26] P. Merino, F. Tröltzsch, B. Vexler, "Error estimates for the finite element approximation of a semilinear elliptic control problem with state constraints and finite dimensional control space," *ESAIM, Math. Model. Numer. Anal.* **44***, no. 1* (2010), 167 – 188.

[27] C. Meyer, A. Rösch, "Superconvergence Properties of Optimal Control Problems", *SIAM J. Contr. Optim.* **43** (2004), 970–985.

[28] B. Sendov, V. A. Popov, "The Averaged Moduli of Smoothness," *Wiley-Interscience, 1988.*

[29] J. Stoer, R. Bulirsch, "Introduction to Numerical Analysis," *Springer, 1980.*

[30] V. M. Veliov, "On the time-discretization of control systems," *SIAM J. Control Optim.* **35***, no. 5* (1997), 1470–1486.

[31] V. M. Veliov, "Error analysis of discrete approximations to bang-bang optimal control problems: the linear case," *Control Cybernet.* **34** (2005), 967–982.

[32] P. R. Wolenski, "The exponential formula for the reachable set of a Lipschitz differential inclusion," *SIAM J. Control Optim.* **28** (1990), 1148–1161.

Institut für Angewandte Mathematik, Friedrich-Schiller-Universität Jena, 07740 Jena, Germany
*E-mail address*: `walter.alt@uni-jena.de`

Mathematisches Institut, Universität Bayreuth, 95440 Bayreuth, Germany
*E-mail address*: `robert.baier@uni-bayreuth.de`

Institut für Mathematik und Rechneranwendung, Fakultät für Luft- und Raumfahrttechnik, Universität der Bundeswehr, München, Germany
*E-mail address*: `matthias.gerdts@unibw.de`

Mathematisches Institut, Universität Bayreuth, 95440 Bayreuth, Germany
*E-mail address*: `frank.lempio@uni-bayreuth.de`