

Error estimates for linear-quadratic control problems with control constraints

A. RÖSCH *

An abstract linear-quadratic optimal control problem with pointwise control constraints is investigated. This paper is concerned with the discretization of the control by piecewise linear functions. Under the assumption that the optimal control and the optimal adjoint state are Lipschitz continuous and piecewise of class C^2 an approximation of order $h^{3/2}$ is proved for the solution of the control discretized problem with respect to the solution of the continuous one. The theoretical results are illustrated by numerical tests.

Keywords: Linear-quadratic optimal control problems, error estimates, numerical approximation, control constraints.

2000 Mathematical Subject Classification: 49N10, 49M05, 65K10

1 Introduction

The paper is concerned with the discretization of the following abstract linear-quadratic control problem

$$(P) \begin{cases} \min & J(u) = \frac{1}{2}\|y - y_d\|_Y^2 + \frac{\nu}{2}\|u\|_U^2 \\ \text{subject to} & y = Su \\ \text{and } u \in C & = \{u \in U \mid a \leq u \leq b \text{ a.e. on } [0, T]\}, \end{cases}$$

where $U = L^2(0, T)$, Y is a Hilbert space and $S : U \rightarrow Y$ is a linear and continuous operator. We assume $\nu > 0$ and a, b are real numbers with $a > b$. We will investigate the effects caused by the discretization of the control. Therefore we consider the control discretized counterpart of (P)

$$(P_h) \begin{cases} \min & J(u) = \frac{1}{2}\|y - y_d\|_Y^2 + \frac{\nu}{2}\|u\|_U^2 \\ \text{subject to} & y = Su \\ \text{and } u \in C_h & = \{u \in U_h \mid a \leq u \leq b \text{ a.e. on } [0, T]\}, \end{cases}$$

*Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria, arnd.roesch@oeaw.ac.at

where is *no* discretization in the state space Y . Here, the space U_h is an a priori chosen finite dimensional space. The parameter h is connected to the discretization of the control.

The operator S can be interpreted as a solution operator for a partial differential equation. This setting fits with the boundary control of an one-dimensional parabolic equation or the boundary control of a two-dimensional elliptic equation.

In this paper we discuss a space U_h of *piecewise linear functions on a uniform grid*. The case of piecewise constant functions is well investigated, we refer to Falk [7], Geveci [8], Malanowski [12] and Arada, Casas, and Tröltzsch [1]. The authors show for different examples the convergence order h .

The convergence order h is optimal if the optimal control is only an element of H^1 . If the optimal solution is more regular, then the convergence order of piecewise constant functions can not improve, due to the fact that the best approximation of a smooth function by piecewise constant functions has only convergence order h . Under reasonable assumptions we prove in the main result that the error between the solutions of (P) and (P_h) for piecewise linear functions is of order $h^{3/2}$. To get this result, we assume that the optimal control is Lipschitz continuous and piecewise of class C^2 . This assumption holds for several examples in literature, we refer to Arada, Raymond, and Tröltzsch [2], Tröltzsch and Mittelmann [15] for parabolic problems, and Bergounioux, Ito, and Kunisch [3] for elliptic problems.

The topic of the paper is well investigated for optimal control problems for ODEs with control constraints. We refer here to Hager [9], [10], Malanowski, Büskens, and Maurer [13], and Dontchev, Hager, and Veliov [6].

Some preprints and techreports are published after the first submission of this paper, which are close related to this one: Casas and Tröltzsch [5], Casas, Mateos and Tröltzsch [4], Hinze [11], Rösch [16], Meyer and Rösch [14].

We describe the discretized space U_h in the following form: Let $h = T/n$, $t_i = i \cdot h$ and e_i ($i = 0..n$) be the functions

$$e_i = \begin{cases} (x - t_{i-1})/h & \text{if } x \in [t_{i-1}, t_i) \\ (t_{i+1} - x)/h & \text{if } x \in [t_i, t_{i+1}] \\ 0 & \text{otherwise.} \end{cases}$$

Now, the space U_h is defined by

$$U_h = \{u \in U \mid u = \sum_{i=0}^n u_i e_i\}.$$

The condition, $u \in C_h$ with

$$C_h = \{u \in U_h \mid a \leq u \leq b \text{ a.e. on } (0, T)\}$$

can be expressed in the form

$$a \leq u_i \leq b \quad \forall i = 0..n.$$

Thus, (P_h) can be equivalently written as

$$(P'_h) \left\{ \begin{array}{l} \min \quad J(u) = \frac{1}{2} \|y - y_d\|_Y^2 + \frac{\nu}{2} \|u\|_U^2 \\ \text{subject to } y = Su \\ u = \sum_{i=0}^n u_i e_i \\ \text{and } a \leq u_i \leq b \quad \forall i = 0..n \end{array} \right.$$

in this case.

The paper is organized as follows: In section 2 we formulate the optimality conditions and state the main result. Section 3 contains several auxiliary results. The proof of the error estimate is presented in section 4. The paper ends with numerical tests in section 5.

2 Optimality conditions and error estimates

The introduced minimization problems (P) and (P_h) are strictly convex. Therefore, the first-order optimality conditions are sufficient for optimality.

Lemma 1 *The necessary and sufficient first-order optimality conditions for the unique solution \bar{u} of P and the unique solution u_h of P_h are given by*

$$(S^*(S\bar{u} - y_d) + \nu\bar{u}, u - \bar{u})_U \geq 0 \quad \forall u \in C, \quad (2.1)$$

$$(S^*(Su_h - y_d) + \nu u_h, u - u_h)_U \geq 0 \quad \forall u \in C_h, \quad (2.2)$$

where $S^* : Y \rightarrow U$ denotes the adjoint operator.

This result is quite standard, hence we drop the proof.

We point out that in (P_h) only the control is discretized. The main concern of this paper is to investigate the influence of the a priori chosen finite dimensional space U_h on the approximation behaviour. Clearly, for applications we have to discretize the state equation, too. In the next lemma, we sketch a simple possibility to estimate the approximation error of the fully discretized problem. However, a tuning of the control and state discretization can improve the approximation behaviour.

Usually, the necessary and sufficient optimality condition (2.2) is exploited to compute numerical solutions of (P_h) . Here, we assume that we can solve the state equation and the adjoint equation with a certain over-all accuracy $\tilde{\delta}$. This setting covers for instance black box solver for both equations.

Lemma 2 Let \tilde{u}_h be the unique solution of

$$(S^*(S\tilde{u}_h - y_d) + \tilde{f} + \nu\tilde{u}_h, u - \tilde{u}_h)_U \geq 0 \quad \forall u \in C_h,$$

and

$$\|\tilde{f}\|_U \leq \tilde{\delta}.$$

Then the error estimate

$$\|u_h - \tilde{u}_h\|_U \leq \frac{\tilde{\delta}}{\nu}.$$

holds true.

Proof: We test the variational inequality for \tilde{u}_h with u_h and (2.2) with \tilde{u}_h

$$\begin{aligned} (S^*(Su_h - y_d) + \nu u_h, \tilde{u}_h - u_h)_U &\geq 0 \\ (S^*(S\tilde{u}_h - y_d) + \tilde{f} + \nu\tilde{u}_h, u_h - \tilde{u}_h)_U &\geq 0. \end{aligned}$$

We add these two inequalities and obtain

$$(S^*S(\tilde{u}_h - u_h) + \tilde{f} + \nu(\tilde{u}_h - u_h), u_h - \tilde{u}_h)_U \geq 0.$$

We continue by

$$\nu\|u_h - \tilde{u}_h\|_U^2 + \|S(\tilde{u}_h - u_h)\|_Y^2 \leq \|\tilde{f}\|_U\|u_h - \tilde{u}_h\|_U$$

which implies easily the assertion. \square

The over-all discretization error can be estimated by the sum of two errors. The first error is the approximation error between (P) and (P_h) . The second one is generated by the discretization of the state equation and can be estimated by Lemma 2. Next, we discuss the connection between (P) and (P_h) in detail. Precisely, we investigate the discretized problem (P'_h) for piecewise linear functions on a uniform grid.

Assumption (A): The optimal control \bar{u} of (P) is Lipschitz continuous and piecewise of class C^2 , that means there exists a *finite* number of points $t^k \in [0, T]$ ($k = 0..N$, $t^0 = 0$, $t^N = T$) such that $\bar{u} \in C^2[t^{k-1}, t^k]$ for all $k = 1..N$.

The expression $\bar{p} = S^*(S\bar{u} - y_d)$ is called *adjoint state*. For the adjoint state we need a similar assumption.

Assumption (B): The adjoint state $\bar{p} = S^*(S\bar{u} - y_d)$ is Lipschitz continuous and piecewise of class C^2 , that means there exists a *finite* number of points $t^k \in [0, T]$ ($k = 0..N$, $t^0 = 0$, $t^N = T$) such that $\bar{p} \in C^2[t^{k-1}, t^k]$ for all $k = 1..N$.

In general it is difficult to show that the optimal control and the optimal adjoint state fulfil these assumptions. Nevertheless, they are fulfilled in several examples in literature, see [2], [15], and [3]. These assumptions reflect the best case in constrained control of PDEs. It is not necessary for the following theory, that the points t^k from (A) and (B) coincide. However, the necessary optimality conditions (especially the projection formula) imply a strong correlation between these points.

Theorem 1 Assume that the optimal control \bar{u} of (P) and the associated adjoint state $\bar{p} = S^*(S\bar{u} - y_d)$ fulfil assumptions (A) and (B). Furthermore, let u_h be the solution of (P'_h). Then the estimate

$$\|\bar{u} - u_h\|_U \leq c_1 \cdot h^{3/2} \quad (2.3)$$

with a positive constant c_1 holds true.

The proof of Theorem 1 is contained in section 4.

3 Auxiliary results

First, we introduce the linear interpolate v of the solution \bar{u} of (P)

$$v = \sum_{i=0}^n \bar{u}(t_i) e_i.$$

Lemma 3 Under assumption (A) we have

$$\|\bar{u} - v\|_U \leq c_3 \cdot h^{3/2} \quad (3.1)$$

Proof: We group our intervals $[t_i, t_{i+1}]$ in two classes. In the first class I_1 of these intervals the function \bar{u} is of class C^2 . Since (A) is true, there exist a constant C_2 independent of the discretization with

$$C_2 = \max_{k=1..N} \|u''\|_{C[t^{k-1}, t^k]}.$$

In the second class I_2 the function \bar{u} is only Lipschitz. We denote the (global) Lipschitz constant by L . Because of assumption (A) class I_2 contains at most $N - 1$ intervals. Using $T = n \cdot h$, we estimate

$$\begin{aligned} \|\bar{u} - v\|_U^2 &= \sum_{i=1}^N \int_{t_{i-1}}^{t_i} |\bar{u}(t) - v(t)|^2 dt \\ &= \sum_{I_1} \int_{t_{i-1}}^{t_i} |\bar{u}(t) - v(t)|^2 dt + \sum_{I_2} \int_{t_{i-1}}^{t_i} |\bar{u}(t) - v(t)|^2 dt \\ &\leq \sum_{I_1} \frac{1}{64} C_2^2 h^4 \cdot h + \sum_{I_2} \frac{1}{4} L^2 h^2 \cdot h \\ &\leq n \frac{1}{64} C_2^2 h^4 \cdot h + \frac{1}{4} L^2 \cdot (N - 1) \cdot h^3 \\ &\leq T \frac{1}{64} C_2^2 h^4 + \frac{1}{4} L^2 \cdot (N - 1) \cdot h^3. \end{aligned}$$

The number N does not depend on h . Therefore the assertion is true. \square

Using $h < T$, the constant c_3 can be simply estimated by

$$c_3 \leq \sqrt{T^2 \frac{1}{64} C_2^2 + \frac{1}{4} L^2 \cdot (N-1)} \leq \frac{T}{8} C_2 + \frac{L}{2} \sqrt{N-1}.$$

For sufficiently fine discretization the Lipschitz part dominates the error estimate. Hence, for a sufficiently fine discretization we find an estimate

$$c_3 \leq L \sqrt{N-1}.$$

For the proof of the main result we need the following auxiliary result.

Lemma 4 *Assume that the optimal control \bar{u} of (P) fulfils assumption (A). Furthermore, let u_h be the solution of (P'_h) . Then the estimate*

$$(S^*(S\bar{u} - y_d) + \nu u_h, v - u_h)_U + c_4 \cdot h^{3/2} \|v - u_h\| \geq 0 \quad (3.2)$$

holds true where the interpolate of \bar{u} is again denoted by v . Moreover, the constant c_4 is given by

$$c_4 = c_3 \|S^* S\|_{\mathcal{L}(U)}.$$

Proof: The interpolate v belongs to C_h . Therefore we can apply (2.2) with $u = v$ and obtain

$$(S^*(S u_h - y_d) + \nu u_h, v - u_h)_U \geq 0. \quad (3.3)$$

Using Lemma 3, we get

$$\begin{aligned} (S^* S(\bar{u} - u_h), v - u_h)_U &= (S^* S(\bar{u} - v), v - u_h)_U + (S^* S(v - u_h), v - u_h)_U \\ &= (S^* S(\bar{u} - v), v - u_h)_U + (S(v - u_h), S(v - u_h))_Y \\ &\geq -c_3 \|S^* S\|_{\mathcal{L}(U)} \cdot h^{3/2} \|v - u_h\|_U. \end{aligned}$$

Combining this inequality with (3.3), we obtain

$$\begin{aligned} (S^*(S\bar{u} - y_d) + \nu u_h, v - u_h)_U &= (S^*(S u_h - y_d) + \nu u_h, v - u_h)_U \\ &\quad + (S^* S(\bar{u} - u_h), v - u_h)_U \\ &\geq -c_3 \|S^* S\|_{\mathcal{L}(U)} \cdot h^{3/2} \|v - u_h\|_U \end{aligned}$$

that implies the assertion. \square

For our further investigations we need the representation of different functions in coordinates. We start with the solution u_h of (P'_h)

$$u_h = \sum_{i=0}^n \zeta_i e_i \quad \text{with } \zeta_i = u_h(t_i)$$

and the interpolate v of the solution \bar{u} of (P)

$$v = \sum_{i=0}^n \eta_i e_i \quad \text{with } \eta_i = \bar{u}(t_i).$$

Furthermore we need a representation of the adjoint state $\bar{p} = S^*(S\bar{u} - y_d)$. This function does not belong to U_h . Therefore we use an interpolation

$$\bar{p} := S^*(S\bar{u} - y_d) = \sum_{i=0}^n \phi_i e_i + \delta_p \quad \text{with } \phi_i = \bar{p}(t_i).$$

We denote by ϕ, ζ, η the $n + 1$ -dimensional vectors with the coordinates ϕ_i, ζ_i, η_i .

Remark 1 *Under assumption (B) we get an estimate for the interpolation error δ_p*

$$\|\delta_p\|_U \leq c_5 \cdot h^{3/2} \quad (3.4)$$

with a positive constant c_5 because of Lemma 3. The constant c_5 can be similar estimated like c_3 . Only the Lipschitz constant L and the constant c_2 have to be replaced by the corresponding values for \bar{p} .

For convenience we introduce the matrix G of the scalar products

$$G = \begin{pmatrix} (e_0, e_0) & \dots & (e_N, e_0) \\ \dots & \dots & \dots \\ (e_0, e_N) & \dots & (e_N, e_N) \end{pmatrix}$$

which has here the form

$$G = \frac{h}{6} \cdot \begin{pmatrix} 2 & 1 & & 0 \\ 1 & 4 & 1 & \\ & \dots & \dots & \dots \\ & & 1 & 4 & 1 \\ 0 & & & 1 & 2 \end{pmatrix} \quad (3.5)$$

Remark 2 *Using the matrix G , the scalar product in $U_h \subset U$ can be expressed by a scalar product in R^{n+1} , for instance*

$$\|u_h\|_U^2 = (u_h, u_h)_U = (G\zeta, \zeta)_{R^{n+1}}.$$

Remark 3 *Using Gershgorin discs, we can estimate the U -norm for an element $u_h \in U_h$ by the Euklidean norm in R^{n+1} , that means*

$$\frac{1}{6}h\|\zeta\|_{R^{n+1}}^2 \leq \|u_h\|_U^2 = (u_h, u_h)_U = (G\zeta, \zeta)_{R^{n+1}} \leq h\|\zeta\|_{R^{n+1}}^2.$$

Lemma 5 *Assume that the assumptions (A) and (B) hold. Then we have*

$$(G(\phi + \nu\zeta), \eta - \zeta)_{R^{n+1}} + c_6 \cdot h^{3/2} \|v - u_h\|_U \geq 0 \quad (3.6)$$

with

$$c_6 = c_4 + c_5.$$

Proof: We can write inequality (3.2) in the form

$$(G(\phi + \nu\zeta), \eta - \zeta)_{R^{n+1}} + (\delta_p, v - u_h)_U + c_4 \cdot h^{3/2} \|v - u_h\|_U \geq 0.$$

Now Remark 1 implies

$$(G(\phi + \nu\zeta), \eta - \zeta)_{R^{n+1}} + (c_4 + c_5) \cdot h^{3/2} \|v - u_h\|_U \geq 0.$$

□

Until now we proved several properties of the solution of (P'_h) . The next lemma contains an important inequality for the interpolate v .

Lemma 6 *Let $D \in R^{n+1 \times n+1}$ be a diagonal matrix with positive entries d_i . If the assumptions (A) and (B) are fulfilled, then the inequality*

$$(D(\phi + \nu\eta), \zeta - \eta)_{R^{n+1}} \geq 0 \quad (3.7)$$

holds true.

Proof: Inserting $\bar{p} = S^*(S\bar{u} - y_d)$ in inequality (2.1), we obtain

$$(\bar{p} + \nu\bar{u}, u - \bar{u})_U \geq 0 \quad \forall u \in C.$$

This inequality implies

$$\bar{u}(t) = \Pi_{[a,b]} \left(-\frac{1}{\nu} \bar{p}(t, 1) \right)$$

with the projection operator

$$\Pi_{[a,b]}(r) = \begin{cases} a & \text{for } r < a \\ r & \text{for } r \in [a, b] \\ b & \text{for } r > b. \end{cases}$$

almost everywhere, see [12]. Since \bar{p} (Assumption (B)) and \bar{u} (Assumption (A)) are continuous, this holds pointwise everywhere. A simple discussion shows, that this implies

$$(\bar{p}(t) + \nu\bar{u}(t)) \cdot (w - \bar{u}(t)) \geq 0$$

for all $t \in [0, T]$ and $w \in [a, b]$.

Setting $u = u_h$, we get in the grid points

$$(\bar{p}(t_i) + \nu \bar{u}(t_i)) \cdot (u_h(t_i) - \bar{u}(t_i)) \geq 0 \quad \text{for } i = 0..n.$$

Using $\zeta_i = u_h(t_i)$, $\eta_i = \bar{u}(t_i)$, and $\phi_i = \bar{p}(t_i)$, we can write this inequality in the form

$$(\phi_i + \nu \eta_i) \cdot (\zeta_i - \eta_i) \geq 0 \quad \text{for } i = 0..n.$$

We multiply these $n + 1$ inequalities by positive weights d_i . Then we add all inequalities. We end up with

$$(D(\phi + \nu \eta), \zeta - \eta)_{R^{n+1}} \geq 0$$

where D is the diagonal matrix with the entries d_i . □

4 Error estimates

The proof of Theorem 1 base on the inequalities (3.6) and (3.7). We only need a suitable choice of the diagonal matrix D . The difference of the matrices G and D plays an important role in this proof. It is desirable that this difference is singular. Therefore we choose the following matrix

$$D = \frac{h}{6} \cdot \begin{pmatrix} 3 & & & 0 \\ & 6 & & \\ & & \ddots & \\ 0 & & & 6 \\ & & & & 3 \end{pmatrix} \quad (4.1)$$

Consequently, we obtain for the difference $G - D$

$$G - D = \frac{h}{6} \cdot \begin{pmatrix} -1 & 1 & & & 0 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ 0 & & & 1 & -1 \end{pmatrix} \quad (4.2)$$

Lemma 7 *If the assumptions (A) and (B) are fulfilled, then the estimates*

$$\|(G - D)\eta\|_{R^{n+1}} \leq c_7 h^2 \quad (4.3)$$

$$\|(G - D)\phi\|_{R^{n+1}} \leq c_8 h^2 \quad (4.4)$$

hold true.

Proof: We discuss the vector $z = (G - D)\eta$ in detail. There are three possibilities:

1. The components z_0 and z_n :

We find the inequalities

$$|z_0| = \frac{h}{6} |\eta_1 - \eta_0| \leq L \cdot \frac{h^2}{6}$$

and

$$|z_n| = \frac{h}{6} |\eta_{n-1} - \eta_n| \leq L \cdot \frac{h^2}{6}$$

where L again denotes the Lipschitz constant of \bar{u} .

2. Components i such that \bar{u} is only Lipschitz in $[t_{i-1}, t_{i+1}]$:

Here we obtain

$$|z_i| = \frac{h}{6} |\eta_{i-1} - 2\eta_i + \eta_{i+1}| \leq 2L \cdot \frac{h^2}{6}.$$

3. Components i such that \bar{u} is of class C^2 in $[t_{i-1}, t_{i+1}]$:

In this case we have

$$|z_i| = \frac{h}{6} |\eta_{i-1} - 2\eta_i + \eta_{i+1}| \leq C_2 \cdot \frac{h^3}{6}$$

with the constant C_2 introduced in Lemma 3.

Again we group the components in two classes. Class J_1 is associated to the cases 1 and 2 and contains the Lipschitz part. Case 3 with the C^2 -part is represented in J_2 . We obtain

$$\begin{aligned} \|(G - D)\eta\|_{R^{n+1}}^2 &= \sum_{i=0}^n |z_i|^2 \\ &= \sum_{J_1} |z_i|^2 + \sum_{J_2} |z_i|^2 \\ &\leq (2N + 2) \cdot L^2 \cdot \frac{h^4}{9} + n \cdot C_2^2 \cdot \frac{h^6}{36} \\ &\leq c_7^2 h^4. \end{aligned}$$

Analogue, we can prove the other inequality. □

The constant c_7 can be estimated using $h < T$

$$c_7 \leq \sqrt{\frac{2N + 2}{9} \cdot L^2 + C_2^2} \cdot \frac{T^2}{36} \leq \frac{L}{3} \sqrt{2N + 2} + C_2 \cdot \frac{T}{6}.$$

Again, the Lipschitz part dominates for sufficiently fine discretization. The constant c_8 can be estimated by the same method. Here, the quantities L and C_2 have to be replaced by the corresponding quantities for \bar{p} .

We are now able to prove the main result.

Proof of Theorem 1: First, we recall the inequalities (3.6) and (3.7)

$$\begin{aligned} (G(\phi + \nu\zeta), \eta - \zeta)_{R^{n+1}} + c_6 \cdot h^{3/2} \|v - u_h\|_U &\geq 0 \\ (D(\phi + \nu\eta), \zeta - \eta)_{R^{n+1}} &\geq 0. \end{aligned}$$

Next, we add these two inequalities

$$((G - D)\phi, \eta - \zeta)_{R^{n+1}} - \nu(D\eta, \eta - \zeta)_{R^{n+1}} + \nu(G\zeta, \eta - \zeta)_{R^{n+1}} + c_6 \cdot h^{3/2} \|v - u_h\|_U \geq 0$$

and write this inequality in the form

$$\begin{aligned} ((G - D)\phi, \eta - \zeta)_{R^{n+1}} + \nu((G - D)\eta, \eta - \zeta)_{R^{n+1}} \\ + \nu(G(\zeta - \eta), \eta - \zeta)_{R^{n+1}} + c_6 \cdot h^{3/2} \|v - u_h\|_U &\geq 0 \end{aligned} \quad (4.5)$$

According to Remark 2 we have

$$\nu(G(\zeta - \eta), \eta - \zeta)_{R^{n+1}} = -\nu \|u_h - v\|_U^2.$$

We insert this inequality in (4.5) and obtain

$$\nu \|u_h - v\|_U^2 \leq ((G - D)\phi, \eta - \zeta)_{R^{n+1}} + \nu((G - D)\eta, \eta - \zeta)_{R^{n+1}} + c_6 \cdot h^{3/2} \|v - u_h\|_U.$$

Lemma 7 implies

$$\begin{aligned} ((G - D)\phi, \eta - \zeta)_{R^{n+1}} &\leq c_7 h^2 \|\eta - \zeta\|_{R^{n+1}}, \\ ((G - D)\eta, \eta - \zeta)_{R^{n+1}} &\leq c_8 h^2 \|\eta - \zeta\|_{R^{n+1}}. \end{aligned}$$

Due to Remark 3, we obtain

$$\begin{aligned} ((G - D)\phi, \eta - \zeta)_{R^{n+1}} &\leq \sqrt{6} c_7 h^{3/2} \|v - u_h\|_U, \\ ((G - D)\eta, \eta - \zeta)_{R^{n+1}} &\leq \sqrt{6} c_8 h^{3/2} \|v - u_h\|_U. \end{aligned}$$

Therefore, we get

$$\|u_h - v\|_U \leq (\sqrt{6}(c_7 + c_8) + c_6) \cdot h^{3/2}.$$

From this inequality and (3.1) we imply

$$\|u_h - \bar{u}\|_U \leq \|u_h - v\|_U + \|v - \bar{u}\|_U \leq c_1 \cdot h^{3/2}$$

which is exactly the assertion. The constant c_1 is given by

$$c_1 = \sqrt{6}(c_7 + c_8) + c_6 + c_3.$$

□

Recalling the estimates of the constants c_3 , c_6 , c_7 , and c_8 , we find that the constant c_1 depends (linear) on the Lipschitz constant L , the bound of the second derivative C_2 of \bar{u} and the corresponding quantities L_p and $C_{2,p}$ of \bar{p} :

$$\begin{aligned} c_1 \leq & (1 + \|S^*S\|_{\mathcal{L}(U)})\left(\frac{T}{8}C_2 + \frac{L}{2}\sqrt{N-1}\right) + \frac{T}{8}C_{2,p} \\ & + \frac{L_p}{2}\sqrt{N-1} + \sqrt{6}\left(\frac{L}{3}\sqrt{2N+2} + C_2 \cdot \frac{T}{6} + \frac{L_p}{3}\sqrt{2N+2} + C_{2,p} \cdot \frac{T}{6}\right). \end{aligned}$$

For sufficiently fine discretization we find

$$c_1 \leq (1 + \|S^*S\|_{\mathcal{L}(U)})L\sqrt{N-1} + L_p\sqrt{N-1} + \sqrt{2}L\sqrt{2N+2} + \sqrt{2}L_p\sqrt{2N+2},$$

since the Lipschitz parts dominate then the estimates.

Remark 4 *The result of Theorem 1 can be slightly generalized. Instead of $a \leq u \leq b$, we require now*

$$a(t) \leq u(t) \leq b(t),$$

where a is a convex continuous function and b is a concave continuous function with $a(t) < b(t)$ on $[0, T]$. Then the result of Theorem 1 is also valid.

5 Numerical tests

For our tests we use an example where the exact solution is known. We investigate here a simple optimal control problem for an ODE. Because of the simple structure, all numerical solutions can be calculated with an (nearly) arbitrary high accuracy. Therefore, the term $\tilde{\delta}/\nu$ from Lemma 2 is small with respect to the approximation error $\|\bar{u} - u_h\|_U$. Hence, the following numerical results reflects the investigated approximation properties.

In principal this can also be done for optimal control problems governed by PDEs. Then it needs fine discretizations of the state equation. For instance, in Test 3, the approximation error is approximately 10^{-6} . Then, the state equation and adjoint equation should be solved with an (much) higher accuracy. Corresponding numerical results can be found in [16].

Clearly, optimal control problems for ODEs can be solved by more efficient methods. Nevertheless, our examples give an impression, how piecewise linear function can improve the convergence order compared with piecewise constant functions.

In our optimal control problem we want to minimize

$$\min F(y, u) = \frac{1}{2}\|y - y_d\|_Y^2 + \frac{\nu}{2}\|u\|_U^2$$

subject to

$$\begin{aligned} \dot{y}(t) &= ay(t) + u(t) & t \in (0, T) \\ y(0) &= 0 \end{aligned}$$

and

$$|u(t)| \leq 1 \quad \text{a.e. on } (0, T).$$

In this setting we choose $U = Y = L^2(0, T)$. The adjoint equation is given by

$$\begin{aligned} -\dot{p}(t) &= ap(t) + y(t) - y_d(t), & t \in (0, T) \\ p(T) &= 0. \end{aligned}$$

The desired state is chosen as

$$y_d = \begin{cases} \frac{1}{a}(e^{at} - 1) + a\nu & \text{on } [0, \alpha) \\ \frac{1}{a}(e^{at} - 2e^{a(t-\alpha)} + 1 + \frac{1}{2}e^{a(T+t-2\alpha)} - \frac{1}{2}e^{a(T-t)}) + a\nu & \text{on } [\alpha, T] \end{cases}$$

where $\alpha \in (0, T)$. The quantities a, T, α fulfil the condition

$$a(T - \alpha) = \ln 2 \tag{5.1}$$

in our example. The optimal control is given by

$$\bar{u} = \begin{cases} 1 & \text{on } [0, \alpha) \\ e^{a(T-t)} - 1 & \text{on } [\alpha, T]. \end{cases}$$

For the state we find

$$\bar{y} = \begin{cases} \frac{1}{a}(e^{at} - 1) & \text{on } [0, \alpha) \\ \frac{1}{a}(e^{at} - 2e^{a(t-\alpha)} + 1 + \frac{1}{2}e^{a(T+t-2\alpha)} - \frac{1}{2}e^{a(T-t)}) & \text{on } [\alpha, T]. \end{cases}$$

Therefore we obtain for the difference

$$\bar{y} - y_d \equiv -a\nu.$$

Consequently we get

$$\bar{p} = -\nu(e^{a(T-t)} - 1)$$

for the adjoint state. It is easy to see that \bar{u} is optimal. It holds

$$-\frac{\bar{p}}{\nu} > 1 \quad \text{on } [0, \alpha).$$

On this interval we have $u \equiv 1$. Moreover, on $[\alpha, T]$

$$\left| -\frac{\bar{p}}{\nu} \right| \leq 1$$

is valid and the condition

$$\bar{p} + \nu \bar{u} = 0$$

is fulfilled. Therefore $(\bar{y}, \bar{u}, \bar{p})$ fulfils the necessary first-order optimality conditions. These conditions are sufficient because of the strict convexity of our optimal control problem.

Test 1: We choose the following values of the parameters

$$\nu = 0.1, \quad a = 1, \quad \alpha = 1, \quad T = 1 + \ln 2.$$

In this test we use piecewise *constant* functions. The numerical result is presented in the Table 1. The number of intervals is denoted by n .

n	$\ u - u_h\ _U$	$n \cdot \ u - u_h\ _U$
8	0.070828	0.566622
16	0.037153	0.594447
32	0.018584	0.594677
64	0.009298	0.595079
128	0.004665	0.597115
256	0.002339	0.598755

Table 1: Piecewise constant functions

Because of $n \cdot h = T$, the last column shows the linear convergence order with respect to h .

Test 2: We take the same values as in Test 1. Now we use piecewise *linear* functions. The number of intervals is again denoted by n . The relation between α and T is an irrational number. Therefore α is not a grid point. We obtained the following result (Table 2):

n	$\ u - u_h\ _U$	$n \cdot \ u - u_h\ _U$	$n^{3/2} \cdot \ u - u_h\ _U$	distance
8	0.021242	0.169930	0.480658	0.058217
16	0.008508	0.136059	0.544236	0.047604
32	0.001474	0.047176	0.266865	0.005306
64	0.001050	0.067229	0.537836	0.005306
128	0.000369	0.047201	0.534014	0.005306
256	0.000088	0.022643	0.362294	0.001308

Table 2: Piecewise linear functions

The numbers in the fourth column vary between 0.26 and 0.55. This result is in accordance with the theoretical convergence order $h^{3/2}$. Let us comment the

variations in these values. The last column shows the distance between the corner point of the exact solution and the nearest grid point. If this distance is constant, then the values in the fourth column increase. If the finer grid fits the corner essentially better, then the values in this column decrease.

A comparison between the results for piecewise constant and piecewise linear functions shows more interesting facts. For the finest discretization ($n = 256$) we got an approximation error of 0.002339 for piecewise constant functions. The approximation error was nearly the half of this value (0.001474) for the small discretization with 32 intervals by use of piecewise linear functions.

Test 3: We change the values of the parameters

$$\nu = 0.1, \quad a = 1, \quad \alpha = \ln 2, \quad T = 2 \ln 2.$$

Now the relation between α and T is $1/2$, i.e. α is a grid point. Again, we use piecewise *linear* functions. The numerical result (for this unrealistic case) is presented in Table 3:

n	$\ u - u_h\ _U$	$n \cdot \ u - u_h\ _U$	$n^{3/2} \cdot \ u - u_h\ _U$	$n^2 \cdot \ u - u_h\ _U$
8	0.001785	0.014282	0.040396	0.114256
16	0.000397	0.006349	0.025398	0.101592
32	0.000093	0.002963	0.016764	0.094831
64	0.000022	0.001427	0.011415	0.091323
128	0.000005	0.000699	0.007914	0.089533
256	0.000001	0.000346	0.005539	0.088629

Table 3: Piecewise linear functions - kink in a grid point

We obtained a quadratic approximation order. Again, the results have a high accuracy for a small number of intervals. The theoretical and numerical results suggest the use of piecewise linear functions instead of piecewise constant functions for such a class of optimal control problems. If the optimal controls have one part where the constraints are active and another part where the constraint are inactive, then these controls are not better than Lipschitz continuous. Therefore, higher splines can not improve the approximation order, if the switching points are unknown.

References

- [1] N. Arada, E. Casas, and F. Tröltzsch, (2002). Error estimates for a semilinear elliptic optimal control problem. *Computational Optimization and Approximation*, **23**, 201–229.

- [2] N. Arada, J.-P. Raymond, and F. Tröltzsch, (2002). On an augmented Lagrangian SQP method for a class of optimal control problems in Banach spaces. *Computational Optimization and Approximation*, **22**, 369–398.
- [3] M. Bergounioux, K. Ito, and K. Kunisch, (1999). Primal-dual strategy for constrained optimal control problems. *SIAM J. Control and Optimization*, **37**, 1176–1194.
- [4] E. Casas, M. Mateos, and F. Tröltzsch, (2004). Error estimates for the numerical approximation of boundary semilinear elliptic control problems. to appear in *Computational Optimization and Applications*.
- [5] E. Casas and F. Tröltzsch, (2003). Error estimates for linear-quadratic elliptic control problems. In V. Barbu et al, editor, *Analysis and Optimization of Differential Systems*, pages 89–100. Kluwer Academic Publishers, Boston.
- [6] A. L. Dontchev, W. W. Hager, and V. M. Veliov, (2000). Second-order Runge-Kutta approximations in constrained optimal control. *SIAM J. Num. Anal.*, **38**, 202–226.
- [7] R. Falk, (1973). Approximation of a class of optimal control problems with order of convergence estimates. *J. Math. Anal. Appl.*, **44**, 28–47.
- [8] T. Geveci, (1979). On the approximation of the solution of an optimal control problem governed by an elliptic equation. *R.A.I.R.O. Analyse numérique*, **13**, 313–328.
- [9] W. W. Hager, (1975). The Ritz-Trefftz method for state and control constrained optimal control problems. *SIAM J. Num. Anal.*, **12**, 854–867.
- [10] W. W. Hager, (2000). Runge-Kutta methods in optimal control and the transformed adjoint system. *Numerische Mathematik*, **87**, 247–282.
- [11] M. Hinze, (2004). A Variational Discretization Concept in Control Constrained Optimization: The Linear-Quadratic Case. to appear in *Computational Optimization and Applications*.
- [12] K. Malanowski, (1981). Convergence of approximations vs. regularity of solutions for convex, control-constrained optimal control problems. *Appl. Math. Opt.*, **8**, 69–95.
- [13] K. Malanowski, Ch. Büskens, and H. Maurer, (1998). Convergence of approximations to nonlinear optimal control problems. In *Mathematical programming with data perturbation*, pages 253–284, Marcel Dekker, New York, Basel.
- [14] C. Meyer and A. Rösch, (2004). Superconvergence properties of optimal control problems. to appear in *SIAM J. Control and Optimization*.

- [15] H. D. Mittelmann and F. Tröltzsch, (2002). Sufficient optimality in parabolic control problem. In *Proceedings of the first International Conference on Industrial and Applied Mathematics in Indian Subcontinent*, Dordrecht, Kluwer Dordrecht, The Netherlands.
- [16] A. Rösch, (2004). Error estimates for parabolic optimal control problems with control constraints. *Zeitschrift für Analysis und ihre Anwendungen*, **23**(2), 353–376.