ruhr.paD
UA Ruhr Zentrum für
partielle Differentialgleichungen

Mass lumping for the optimal control of elliptic partial differential equations

Arnd Rösch, Gerd Wachsmuth

Preprint 2016-01

Mass lumping for the optimal control of elliptic partial differential equations

Arnd Rösch* Gerd Wachsmuth †

May 9, 2016

Abstract. The finite element discretization of a control constrained elliptic optimal control problem is studied. Control and state are discretized by higher order finite elements. The inequality constraints are only posed in the Lagrange points. The computational effort is significantly reduced by a new mass lumping strategy. The main contribution is the derivation of new a priori error estimates up to order h^4. Moreover, we propose a new algorithmic strategy to obtain such highly accurate results. The theoretical findings are illustrated by numerical examples.

Key words. optimal control, control constraint, higher order finite elements, mass lumping, a priori error estimates.
AMS Subject Classifications. 49K20, 49M25, 65N30

1 Introduction

The discretization of optimal control problems by finite elements is nowadays a standard tool. A series of papers investigates a priori discretization error estimates in particular for control constrained problems. In this sense, the theory seems to be nearly completed. A closer look shows that the known approaches are limited due to regularity issues. A standard discretization with piecewise constant controls is limited to the rate of h, which is the spatial discretization parameter, see Arada et al. [2002], Falk [1973], Geveci [1979]. Piecewise linear approximations are limited to $h^{3/2}$, see Rösch [2006], Casas and Mateos [2008]. The superconvergence approach yields a numerical approximation of order h^2, Meyer and Rösch [2004].

By using adaptive mesh refinement, the approximation order in two dimensions is limited by $N^{-3/2}$ (this corresponds to h^3) because of the required number of cells close to the kinks of the optimal control, see [Schneider and Wachsmuth, 2015, Eq. (4)] for

*Faculty of Mathematics, University of Duisburg-Essen, Thea-Leymann-Straße 9, 45127 Essen, Germany; email: arnd.roesch@uni-due.de

†Faculty of Mathematics, Technische Universität Chemnitz, Reichenhainer Str. 39, 09126 Chemnitz, Germany; gerd.wachsmuth@mathematik.tu-chemnitz.de.

a discussion of this limitation and Schneider and Wachsmuth [2016] for results showing convergence of order $N^{-3/2}$.

The only known (non-adaptive) approach where the accuracy is not limited by h^2 is the variational discretization by Hinze [2005]. Of course, higher order finite elements are necessary to obtain such an accuracy. The main challenge is the numerical computation of the optimal control and the evaluation of the scalar product of the optimal control with a finite element function. This can be done exactly for piecewise linear finite elements, but the usage of higher order finite elements is computationally challenging, see Sevilla and Wachsmuth [2010].

This is the starting point of our new method which has convergence order up to h^4. We propose a new fully discrete approach with higher accuracy and low computational effort. To this aim we consider the model problem

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\|y-y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} \\ \text{s.t.} \quad & -\Delta y + y = u \quad \text{in } \Omega \\ & \frac{\partial}{\partial n} y = 0 \quad \text{on } \partial\Omega \\ \text{and} \quad & u_a \le u \le u_b \quad \text{in } \Omega. \end{aligned} \tag{P}$$

We assume that y_d is continuous and sufficiently smooth. Moreover, we assume that the domain $\Omega \subset \mathbb{R}^2$ is convex and polygonal. For simplicity of the presentation, we further assume $u_a, u_b \in \mathbb{R}$.

2 Motivation

It is well-known that the (necessary and sufficient) optimality condition for (P) is given by the projection formula

$$\bar{u}(x) = \operatorname{Proj}_{[u_a,u_b]} \frac{\bar{p}(x)}{\alpha} \quad \text{for a.a. } x \in \Omega, \tag{1}$$

where $\bar{p}$ is the (weak) solution of the adjoint equation

$$\begin{aligned} -\Delta\bar{p} + \bar{p} &= y_d - \bar{y} \quad && \text{in } \Omega, \\ \frac{\partial}{\partial n}\bar{p} &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{2}$$

Now, we consider a discretized version of (P). Let $\mathcal{T}$ be a regular triangulation of Ω. We will work with a finite element space

$$V_h = \{v_h \in C(\bar{\Omega}) : v_h|_T \in \mathcal{P}(T)\ \forall T \in \mathcal{T}\},$$

where $\mathcal{P}$ denotes a certain polynomial space of higher order. We will specify the details later. In the sequel we will use the notation v_h for the discretized function, i.e. $v_h \in V_h$, and also for the coefficient vector. Both are connected by a bijective mapping induced by the Lagrange interpolation. We denote by M and K the usual mass and stiffness matrix associated with the inner products of $L^2(\Omega)$ and $H^1(\Omega)$, respectively. A possibility to discretize the optimal control problem (**P**) would be

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\,\|y_h - I_h\, y_d\|_M^2 + \frac{\alpha}{2}\,\|u\|_M^2 \\ \text{s.t.} \quad & K\, y_h = M\, u_h \\ \text{and} \quad & u_a \le u_h \le u_b. \end{aligned} \tag{$\mathbf{P}'_h$}$$

Since V_h consists of polynomials of higher order, only a coefficient-wise interpretation of the control constraints is practicable from a computational point of view. That is, the control constraint in ($\mathbf{P}'_h$) is to be understood as $u_a \le u_h(x_L) \le u_b$ for all Lagrange-nodes x_L of the triangulation or, equivalently, as $u_a \le u_h^i \le u_b$ for all coefficients. Note that such a discretization is not conforming if the polynomial degree is larger than one. The operator $I_h : C(\bar\Omega) \to V_h$ in ($\mathbf{P}'_h$) is the (usual) nodal Lagrange interpolation.

It is easy to see that the (necessary and sufficient) optimality condition of ($\mathbf{P}'_h$) can be written as

$$\bar u_h = \operatorname{Proj}^M_{[u_a,u_b]} \frac{\bar p_h}{\alpha}, \tag{3}$$

where the discretized adjoint state $\bar p_h$ is defined by the equations

$$K\,\bar p_h = M\,(I_h\, y_d - \bar y_h), \qquad K\,\bar y_h = M\,\bar u_h.$$

Note that we have to project with respect to the inner product generated by the mass matrix M, i.e., the projection (3) is characterized by

$$(\alpha\,\bar u_h - \bar p_h)^\top M\,(u_h - \bar u_h) \ge 0 \qquad \forall u_h \in V_h : u_a \le u_h \le u_b.$$

This projection can be evaluated coefficient-wise if and only if the mass matrix M is diagonal, otherwise nonlocal effects appear.

The superconvergence approach introduced in Meyer and Rösch [2004] works with different discrete spaces for state and control. In particular, the controls are discretized by piecewise constant functions and, thus, the mass matrix becomes a diagonal matrix which is heavily exploited in the derivation of the approximation results.

We will propose here a new approach with a mass matrix M and a diagonal lumped-mass matrix M_L. The mass matrix is used in the tracking term and the lumped-mass matrix is used twice: in the right-hand side of the state equation and in the control cost term in the objective.

Hence, we propose to use

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\,\|y_h - I_h\, y_d\|_M^2 + \frac{\alpha}{2}\,\|u_h\|_{M_L}^2 \\ \text{s.t.} \quad & K\, y_h = M_L\, u_h \\ \text{and} \quad & u_a \le u_h \le u_b. \end{aligned} \tag{$\mathbf{P}_h$}$$

with a diagonal, positive semidefinite approximation M_L of the mass matrix. As in ($\mathbf{P}'_h$), the control constraints are to be understood coefficient-wise.

Mass lumping is a standard tool for the numerical solution of *time dependent* partial differential equations. Until now, only a few papers are devoted to mass lumping in optimal control. Mass lumping is used in the computation of an L^1 term, see [Wachsmuth and Wachsmuth, 2011, (4.13)] and to obtain a discrete projection formula, see [Casas et al., 2012, Lemma 3.4]. However, only convergence of order h was proved. This was improved by Pieper [2015], who also considered the lumped-mass matrix in the right-hand side of the discrete PDE, see [Pieper, 2015, Section 4.5.4] and convergence of order h^2 for a piecewise linear discretization was obtained. Similar ideas are also used for the control of ordinary differential equations, see for instance Alt et al. [2007].

It is easy to check, that the optimality conditions for $(\mathbf{P}_h)$ are given by

$$\bar{u}_h = \operatorname{Proj}_{[u_a,u_b]} \frac{\bar{p}_h}{\alpha}, \tag{4a}$$

$$K\,\bar{p}_h = M\,(I_h\,y_d - \bar{y}_h), \tag{4b}$$

$$K\,\bar{y}_h = M_L\,\bar{u}_h. \tag{4c}$$

If the diagonal M_L is not strictly positive, the solution of $(\mathbf{P}_h)$ is not unique. Indeed, entries of u_h corresponding to a zero diagonal entry of M_L do not enter the objective or the state equation in $(\mathbf{P}_h)$. We fix these entries by the optimality condition (4a). We emphasize that the projection in (4a) is to be understood coefficient-wise.

We mention that there are two possible interpretations of mass lumping. The first one is from a linear algebra point-of-view and the lumped-mass matrix is understood as a diagonal approximation of the mass matrix, by defining, e.g.,

$$(M_L)_{ii} = \sum_{j=1}^{N} M_{ij}.$$

From a numerical analysis point-of-view, mass lumping can be understood as follows. For the triangulation $\mathcal{T}$ we have the (global) Lagrange nodes x_L^i, $i = 1, \ldots, N$ and the associated (global) basis functions ϕ_i satisfy $\phi_i(x_L^j) = \delta_{ij}$. Then, we choose a quadrature formula whose nodes are exactly the Lagrange nodes and the weights ω_i are non-negative. Now, we can define a lumped-mass matrix by approximating the $L^2(\Omega)$ inner product by the quadrature formula. Indeed,

$$(M_L)_{ij} = \sum_{k=1}^{N} \phi_i(x_L^k)\,\phi_j(x_L^k)\,\omega_k = \begin{cases} \omega_i & \text{if } i = j, \\ 0 & \text{else.} \end{cases}$$

3 Basic error estimate for the control problem

In this section we will derive the basic error estimate. Here, we do not need a specific form of the finite element space V_h. Only the following three properties are required:

- The coefficient vector u_h corresponds to a nodal basis, i.e., for every coefficient u_h^i there is a Lagrange point $x_L^i \in \Omega$ with

$$u_h^i = u_h(x_L^i).$$

Moreover, the interpolation operator $I_h : C(\bar\Omega) \to V_h$ satisfies

$$(I_h\, v)^i = v(x_L^i) \qquad \forall v \in C(\bar\Omega), i = 1, \ldots, N.$$

- The control constraints in $(\mathbf{P}_h)$ are inequalities for single coefficients, i.e.,

  $$u_a \le u_h^i \le u_b$$

  for all coefficients associated with the nodal basis.
- The lumped-mass matrix M_L is diagonal and all entries are non-negative.

Lemma 3.1. *Let us assume that the above assumptions are satisfied by the finite element space V_h. Moreover, we assume $\bar u, \bar p \in C(\bar\Omega)$. Then we have*

$$\alpha\, \|I_h\, \bar u - \bar u_h\|_{M_L}^2 \le (\bar p_h - I_h\, \bar p, \bar u_h - I_h\, \bar u)_{M_L}, \tag{5}$$

where $\bar y_h, \bar u_h, \bar p_h$ are the unique solution of the optimality system (4).

In (5), we used the usual notations

$$(a, b)_{M_L} := a^\top M_L\, b \quad \text{and} \quad \|a\|_{M_L}^2 := (a, a)_{M_L}.$$

Proof. Together with (4a) the assumptions on V_h imply

$$[\alpha\, \bar u_h(x_L^i) - \bar p_h(x_L^i)]\, [v - \bar u_h(x_L^i)] \ge 0 \qquad \forall v \in [u_a, u_b].$$

We choose $v = \bar u(x_L^i)$, which is possible due to $\bar u \in C(\bar\Omega)$ and $u_a \le \bar u \le u_b$ a.e. in Ω. This yields

$$[\alpha\, \bar u_h(x_L^i) - \bar p_h(x_L^i)]\, [\bar u(x_L^i) - \bar u_h(x_L^i)] \ge 0.$$

Since $\bar u$ and $\bar p$ are continuous, the projection formula (1) holds everywhere, and we obtain

$$[\alpha\, \bar u(x_L^i) - \bar p(x_L^i)]\, [\bar u_h(x_L^i) - \bar u(x_L^i)] \ge 0,$$

since $\bar u_h(x_L^i) \in [u_a, u_b]$. Now, we weight both inequalities by the i-th diagonal entry of M_L and sum over all indices i to obtain

$$(\alpha\, I_h\, \bar u - \alpha\, \bar u_h + \bar p_h - I_h\, \bar p)^\top M_L\, (\bar u_h - I_h\, \bar u) \ge 0.$$

Here, we used $(I_h\, \bar u)^i = \bar u(x_L^i)$. This implies the assertion. □

In the following theorem, we estimate the right-hand side of (5) by approximation errors for the state and adjoint equation.

Theorem 3.2. *Under the assumptions of Lemma 3.1, the error estimate*

$$\begin{aligned}\frac{\alpha}{2}\, \|I_h\, \bar u - \bar u_h\|_{M_L}^2 + \frac12\, \|I_h\, \bar y - \bar y_h\|_M^2 + \frac12\, \|K^{-1}\, M_L\, I_h\, \bar u - \bar y_h\|_M^2 \\ \le \frac12\, \|I_h\, \bar y - K^{-1}\, M_L\, I_h\, \bar u\|_M^2 + \frac{1}{2\,\alpha}\, \|I_h\, \bar p - K^{-1}\, M\, (I_h\, y_d - I_h\, \bar y)\|_{M_L}^2\end{aligned} \tag{6}$$

is valid.

Proof. We start with the identity

$$\begin{aligned}
&\frac{1}{2}\,\|I_h\,\bar{y}-\bar{y}_h\|_M^2+\frac{1}{2}\,\|K^{-1}\,M_L\,I_h\,\bar{u}-\bar{y}_h\|_M^2\\
&\quad=\frac{1}{2}\,\|I_h\,\bar{y}-K^{-1}\,M_L\,I_h\,\bar{u}\|_M^2+(I_h\,\bar{y}-\bar{y}_h,K^{-1}\,M_L\,I_h\,\bar{u}-\bar{y}_h)_M\\
&\quad=\frac{1}{2}\,\|I_h\,\bar{y}-K^{-1}\,M_L\,I_h\,\bar{u}\|_M^2+(I_h\,\bar{y}-\bar{y}_h,K^{-1}\,M_L\,I_h\,\bar{u}-K^{-1}\,M_L\,\bar{u}_h)_M\\
&\quad=\frac{1}{2}\,\|I_h\,\bar{y}-K^{-1}\,M_L\,I_h\,\bar{u}\|_M^2+(K^{-1}\,M\,(I_h\,\bar{y}-\bar{y}_h),I_h\,\bar{u}-\bar{u}_h)_{M_L}\\
&\quad=\frac{1}{2}\,\|I_h\,\bar{y}-K^{-1}\,M_L\,I_h\,\bar{u}\|_M^2+(K^{-1}\,M\,(I_h\,\bar{y}-I_h\,y_d)+\bar{p}_h,I_h\,\bar{u}-\bar{u}_h)_{M_L}.
\end{aligned}$$

Together with the inequality (5), we obtain

$$\begin{aligned}
&\alpha\,\|I_h\,\bar{u}-\bar{u}_h\|_{M_L}^2+\frac{1}{2}\,\|I_h\,\bar{y}-\bar{y}_h\|_M^2+\frac{1}{2}\,\|K^{-1}\,M_L\,I_h\,\bar{u}-\bar{y}_h\|_M^2\\
&\quad\le\frac{1}{2}\,\|I_h\,\bar{y}-K^{-1}\,M_L\,I_h\,\bar{u}\|_M^2+(I_h\,\bar{p}-K^{-1}\,M\,(I_h\,y_d-I_h\,\bar{y}),I_h\,\bar{u}-\bar{u}_h)_{M_L}.
\end{aligned}$$

Finally, we use Young's inequality to obtain

$$\begin{aligned}
&\frac{\alpha}{2}\,\|I_h\,\bar{u}-\bar{u}_h\|_{M_L}^2+\frac{1}{2}\,\|I_h\,\bar{y}-\bar{y}_h\|_M^2+\frac{1}{2}\,\|K^{-1}\,M_L\,I_h\,\bar{u}-\bar{y}_h\|_M^2\\
&\quad\le\frac{1}{2}\,\|I_h\,\bar{y}-K^{-1}\,M_L\,I_h\,\bar{u}\|_M^2+\frac{1}{2\,\alpha}\,\|I_h\,\bar{p}-K^{-1}\,M\,(I_h\,y_d-I_h\,\bar{y})\|_{M_L}^2,
\end{aligned}$$

and this is the assertion. □

In the next section we will estimate the two terms on the right-hand side of this inequality. The first term describes the approximation error of the state equation caused by discretization and mass lumping. The second term measures the discretization error of the adjoint equation in the mass lumping norm.

We briefly interpret the terms on the left-hand side of (6). The first term is an error estimate for the approximation of the control. However, since the lumped-mass matrix is involved, only the error in the Lagrange nodes is measured. The second term is equal to $\frac{1}{2}\,\|I_h\,\bar{y}-\bar{y}_h\|_{L^2(\Omega)}^2$ and this is, up to the interpolation error, the $L^2(\Omega)$-error in the state variable.

4 Error estimates for the equations

Let us now give a precise description of the discrete spaces V_h. We require the following properties:

- The convex polygonal domain Ω is discretized exactly by the triangulation $\mathcal{T}$, i.e., we have $\Omega = \Omega_h$.

- The elements of $\mathcal{T}$ are (shape) regular in the sense of [Brenner and Scott, 2008, Definition (4.4.13)]. Hanging nodes are not allowed. There is no restriction on the size h_T of a single element $T \in \mathcal{T}$. Later we will have the mild restriction (13). We denote by $h = \max_{T\in\mathcal{T}} h_T$ the global mesh size.
- The discrete space V_h is generated by continuous and piecewise polynomial basis functions ϕ_i, $i = 1, \dots, N$, corresponding to Lagrange nodes $x_L^i \in \bar{\Omega}$. We postulate the following assumptions on the basis functions.
  - We require $\phi_i(x_L^j) = \delta_{ij}$, for $1 \leq i, j \leq N$.
  - All basis functions have nonnegative integrals, i.e., $\int_\Omega \phi_i \,\mathrm{d}x \geq 0$.
  - The basis functions build a partition of unity, i.e., $\sum_{i=1}^N \phi_i \equiv 1$ on $\bar{\Omega}$.
  - There exists two positive numbers k and k' with $P_k \subset V_h \subset P_{k'}$. Later, we will need the inequalities $k \geq 2$ and $k' \leq k+1$. Here, P_k ($P_{k'}$) is the usual space of continuous functions which are piecewise polynomials of degree at most k (k').
  - The mass matrix M of the finite element space V_h and the lumped mass matrix M_L are connected by the relation

    $$(M_L)_{ii} = \sum_{j=1}^N M_{ij}.$$

  - The quadrature rule corresponding with the lumped-mass matrix is exact for (piecewise) polynomials of degree $k + k' - 2$, that is

    $$\int_\Omega \varphi \,\mathrm{d}x = \sum_{i=1}^N \omega_i \, \varphi(x_L^i)$$

    for all $\varphi \in C(\bar{\Omega})$ for which $\varphi|_T$ is a polynomial of degree at most $k + k' - 2$ for all $T \in \mathcal{T}$. Note that this directly implies

    $$\int_\Omega \phi_i \,\mathrm{d}x = \sum_{j=1}^N \omega_j \, \phi_i(x_L^j) = \omega_i \geq 0,$$

    since the local degree of ϕ_i is at most $k' \leq k + k' - 2$. Hence, the quadrature weights ω_i are uniquely determined by the Lagrange nodes x_L^i. Moreover, we have

    $$(M_L)_{ii} = \sum_{j=1}^N M_{ij} = \sum_{j=1}^N \int_\Omega \phi_i \, \phi_j \,\mathrm{d}x = \int_\Omega \phi_i \sum_{j=1}^N \phi_j \,\mathrm{d}x = \int_\Omega \phi_i \,\mathrm{d}x = \omega_i.$$

  In Section 7 we specify two finite element spaces which satisfy these assumptions.

We denote the space of continuous, piecewise linear functions by $V_h^{(1)}$ and by $I_h^{(1)}$ the nodal interpolation operator to the space $V_h^{(1)}$.

4.1 Error estimates for the adjoint equation

In this section we will estimate the second term in (6). We start with an auxiliary result.

Lemma 4.1. *Let $v_h \in V_h$ an arbitrary function. Then we have*

$$\|v_h\|_{M_L}^2 \le c\,\|v_h\|_M^2 = c\,\|v_h\|_{L^2(\Omega)}^2. \tag{7}$$

Proof. The result follows from a simple transformation argument. Let us investigate the norm on a single element $T \in \mathcal{T}$. After transformation to the reference element we can estimate the lumped-mass matrix semi-norm by the mass matrix norm. Retransformation yields the assertion. □

Lemma 4.2. *The following estimate is valid*

$$\|I_h\,\bar{p} - K^{-1}\,M\,(I_h\,y_d - I_h\,\bar{y})\|_{M_L} \le c\,(\|I_h\,\bar{p} - \bar{p}\|_{L^2(\Omega)} + \|\bar{p} - K^{-1}\,M\,(I_h\,y_d - I_h\,\bar{y})\|_{L^2(\Omega)}). \tag{8}$$

Proof. We obtain the desired result immediately from the last lemma:

$$\begin{aligned}
&\|I_h\,\bar{p} - K^{-1}\,M\,(I_h\,y_d - I_h\,\bar{y})\|_{M_L}\\
&\quad\le c\,(\|I_h\,\bar{p} - K^{-1}\,M\,(I_h\,y_d - I_h\,\bar{y})\|_{L^2(\Omega)})\\
&\quad\le c\,(\|I_h\,\bar{p} - \bar{p}\|_{L^2(\Omega)} + \|\bar{p} - K^{-1}\,M\,(I_h\,y_d - I_h\,\bar{y})\|_{L^2(\Omega)}).
\end{aligned}$$
□

Let us remark that the second term on the right-hand side of this inequality is the usual finite element error for the adjoint equation evaluated for the optimal state $\bar{y}$. The first term represents an interpolation error.

4.2 Error estimates for the state equation

Next, we estimate the first term in (6). Let us introduce the notation

$$\tilde{y} := K^{-1}\,M_L\,I_h\,\bar{u}.$$

This auxiliary discrete state is just the Galerkin solution of the state equation in which the quadrature rule corresponding to the lumped-mass matrix is used to evaluate the right-hand side $\bar{u}$. Indeed,

$$v_h^\top K\,\tilde{y} = v_h^\top M_L\,I_h\,\bar{u} = \sum_{i=1}^{N} \omega_i\,v_h(x_L^i)\,\bar{u}(x_L^i) \approx \int_\Omega v_h\,\bar{u}\,\mathrm{d}x$$

for all $v_h \in V_h$.

The first addend on the right-hand side of (6) can be estimated by the triangle inequality

$$\begin{aligned}
\|I_h\,\bar{y} - K^{-1}\,M_L\,I_h\,\bar{u}\|_M &= \|I_h\,\bar{y} - K^{-1}\,M_L\,I_h\,\bar{u}\|_{L^2(\Omega)}\\
&\le \|I_h\,\bar{y} - \bar{y}\|_{L^2(\Omega)} + \|\bar{y} - K^{-1}\,M_L\,I_h\,\bar{u}\|_{L^2(\Omega)}.
\end{aligned}$$

Owing to the regularity of $\bar{y}$, we can estimate the interpolation error $\|I_h\,\bar{y} - \bar{y}\|_{L^2(\Omega)}$ and the second term is addressed in the following lemma.

Lemma 4.3. *The following a priori error estimate holds*

$$\begin{aligned}\|\bar y-\tilde y\|_{L^2(\Omega)} \le c\,h\left(\|\bar y - I_h\,\bar y\|_{H^1(\Omega)} + \sup_{w_h\in V_h\setminus\{0\}}\frac{(w_h,\bar u)_{L^2(\Omega)} - w_h^\top M_L\,I_h\,\bar u}{\|w_h\|_{H^1(\Omega)}}\right)\\ + c\sup_{w_h\in V_h^{(1)}\setminus\{0\}}\frac{(w_h,\bar u)_{L^2(\Omega)} - w_h^\top M_L\,I_h\,\bar u}{\|w_h\|_{H^1(\Omega)}}.\end{aligned}\tag{9}$$

Proof. Using the first Lemma of Strang, see, e.g., [Ciarlet, 1978, Thm. 4.1.1], we obtain

$$\|\bar y-\tilde y\|_{H^1(\Omega)} \le c\,\|\bar y - I_h\,\bar y\|_{H^1(\Omega)} + c\sup_{w_h\in V_h\setminus\{0\}}\frac{(w_h,\bar u)_{L^2(\Omega)} - w_h\,M_L\,I_h\,\bar u}{\|w_h\|_{H^1(\Omega)}}.$$

Now, we use the Nitsche trick to estimate $\|\bar y-\tilde y\|_{L^2(\Omega)}$. We define φ as the solution of the dual problem with right-hand side $\bar y-\tilde y$, i.e.

$$a(\varphi,v) = (\bar y-\tilde y, v)_{L^2(\Omega)} \quad \forall \varphi\in H^1(\Omega).$$

Now, we find

$$\begin{aligned}\|\bar y-\tilde y\|^2_{L^2(\Omega)} &= a(\varphi,\bar y-\tilde y)\\ &= a(\varphi - I_h^{(1)}\,\varphi,\bar y-\tilde y) + a(I_h^{(1)}\,\varphi,\bar y-\tilde y)\\ &= a(\varphi - I_h^{(1)}\,\varphi,\bar y-\tilde y) + (I_h^{(1)}\,\varphi,\bar u)_{L^2(\Omega)} - (I_h^{(1)}\,\varphi)^\top M_L\,I_h\,\bar u\\ &\le \|\varphi - I_h^{(1)}\,\varphi\|_{H^1(\Omega)}\,\|\bar y-\tilde y\|_{H^1(\Omega)}\\ &\quad + \sup_{w_h\in V_h^{(1)}\setminus\{0\}}\frac{(w_h,\bar u)_{L^2(\Omega)} - w_h\,M_L\,I_h\,\bar u}{\|w_h\|_{H^1(\Omega)}}\,\|I_h^{(1)}\,\varphi\|_{H^1(\Omega)}\\ &\le \left(h\,\|\bar y-\tilde y\|_{H^1(\Omega)} + \sup_{w_h\in V_h^{(1)}\setminus\{0\}}\frac{(w_h,\bar u)_{L^2(\Omega)} - w_h\,M_L\,I_h\,\bar u}{\|w_h\|_{H^1(\Omega)}}\right)\|\varphi\|_{H^2(\Omega)}\\ &\le \left(h\,\|\bar y-\tilde y\|_{H^1(\Omega)} + \sup_{w_h\in V_h^{(1)}\setminus\{0\}}\frac{(w_h,\bar u)_{L^2(\Omega)} - w_h\,M_L\,I_h\,\bar u}{\|w_h\|_{H^1(\Omega)}}\right)\|\bar y-\tilde y\|_{L^2(\Omega)}.\end{aligned}$$

Here, we used [Brenner and Scott, 2008, Theorem (4.4.4)] for the interpolation error and the stability of the interpolation in $H^2(\Omega)$. Together with the error estimate for the energy norm we obtain the assertion. □

This result is the key to estimate the first term in (6). Hence, the main contribution to this error term is addressed in (9). We emphasize that the terms containing the sup in (9) are just the (normalized) quadrature errors

$$(w_h,\bar u)_{L^2(\Omega)} - w_h^\top M_L\,I_h\,\bar u = \int_\Omega w_h\,\bar u\,\mathrm{d}x - \sum_{i=1}^N \omega_i\,w_h(x_L^i)\,\bar u(x_L^i).\tag{10}$$

It remains to estimate these quadrature errors. The global quadrature error can be split into elementwise error contributions.

4.3 Quadrature error on a single element

In this section we will study the error caused by the mass lumping. Let us define the order $r := k + k' - 1$. We are interested in two particular cases:

1. quadratic polynomials, i.e., $k = k' = 2$, $r = 3$,
2. enriched cubic polynomials, i.e., $k = 3$, $k' = 4$, $r = 6$.

In both cases mass lumping is known with a quadrature rule that is exact for polynomials of degree up to order $r - 1$, see Section 7. Let us discuss a single triangle T of the triangulation $\mathcal{T}$. In the following discussions, c will denote a generic constant, which will not depend on the triangle T, but only on its shape regularity.

To allow for a local analysis, the (global) quadrature formula with points x_L^i and weights ω_i is divided into quadrature formulas on each cell $T \in \mathcal{T}$. In particular, we denote the quadrature points and weights associated with the cell $T \in \mathcal{T}$ by $y_L^{T,j}$ and ω_j^T, $j = 1, \dots, N_L$. The relation with the global quadrature formula is given by the requirement

$$\sum_{i=1}^{N} v(x_L^i)\,\omega_i = \sum_{T \in \mathcal{T}} \sum_{j=1}^{N_L} v(y_L^{T,j})\,\omega_j^T \qquad \forall v \in C(\bar{\Omega}). \tag{11}$$

For $\phi \in C(T)$ and $v_h \in V_h|_T$ we define the quadrature error

$$E_T(\phi, v_h) := \Big| \int_T \phi\, v_h \,\mathrm{d}x - \sum_{j=1}^{N_L} \phi(x_L^{T,j})\, v_h(x_L^{T,j})\,\omega_j^T \Big|.$$

The sum over all elements yields the desired estimate, cf. (10) and (11).

Theorem 4.4. *There exist a constant c such that for each cell $T \in \mathcal{T}$, we have the estimates*

$$E_T(\phi, v_h) \le c\, h_T^{k+1}\, \|\phi\|_{W^{k+1,2}(T)}\, \|v_h\|_{W^{2,2}(T)} \tag{12a}$$

$$E_T(\phi, v_h) \le c\, h_T^{2}\, \|\phi\|_{L^\infty(T)}\, \|v_h\|_{L^\infty(T)} \tag{12b}$$

for $\phi \in C(T)$ and $v_h \in V_h|_T$.

We emphasize that the constant c in Theorem 4.4 does not depend on the triangulation $\mathcal{T}$, but only on its shape regularity.

Proof. By standard arguments, we obtain the estimate

$$E_T(\phi, v_h) \le c\, h_T^{r}\, \big|\phi\, v_h\big|_{W^{r,1}(T)}.$$

Next, we use the nodal interpolant $I_h\phi \in V_h|_T$. Of course we have $\phi(x_L^i) = (I_h\phi)(x_L^i)$. Hence, $I_h\phi - \phi$ is zero in the Lagrange points and these are precisely the quadrature nodes x_L^i. Consequently, we find

$$\begin{aligned}
E_T(\phi, v_h) &\le E_T(\phi - I_h\phi, v_h) + E_T(I_h\phi, v_h) \\
&= \Big| \int_T (\phi - I_h\phi)\, v_h \,\mathrm{d}x \Big| + E_T(I_h\phi, v_h) \\
&\le c\,\|\phi - I_h\phi\|_{L^2(T)}\, \|v_h\|_{L^2(T)} + c\, h_T^r\, |I_h\phi\, v_h|_{W^{r,1}(T)},
\end{aligned}$$

which implies

$$E_T(\phi, v_h) \le c\,\|\phi - I_h\phi\|_{L^2(T)}\, \|v_h\|_{L^2(T)} + c\, h_T^r\, \|I_h\phi\|_{W^{r,2}(T)}\, \|v_h\|_{W^{r,2}(T)}.$$

Since $I_h\phi, v_h$ are polynomials of degree k', all derivatives of order $k'+1, \dots, r$ are zero. Together with an inverse estimate (see [Brenner and Scott, 2008, Lemma (4.5.3)]) we get

$$\begin{aligned}
E_T(\phi, v_h) &\le c\,\|\phi - I_h\phi\|_{L^2(T)}\, \|v_h\|_{L^2(T)} + c\, h_T^r\, \|I_h\phi\|_{W^{k',2}(T)}\, \|v_h\|_{W^{k',2}(T)} \\
&\le c\,\|\phi - I_h\phi\|_{L^2(T)}\, \|v_h\|_{L^2(T)} + c\, h_T^{k+1}\, \|I_h\phi\|_{W^{k',2}(T)}\, \|v_h\|_{W^{2,2}(T)}.
\end{aligned}$$

Next we use the stability of the Lagrange interpolant, see [Brenner and Scott, 2008, Theorem (4.4.4)] (note that this requires $k' \le k+1$), and the interpolation estimate [Brenner and Scott, 2008, Theorem (4.4.4)] to obtain

$$\begin{aligned}
E_T(\phi, v_h) &\le c\,\|\phi - I_h\phi\|_{L^2(T)}\, \|v_h\|_{L^2(T)} + c\, h_T^{k+1}\, \|\phi\|_{W^{k',2}(T)}\, \|v_h\|_{W^{2,2}(T)} \\
&\le c\, h_T^{k+1}\, \|\phi\|_{W^{k+1,2}(T)}\, \|v_h\|_{L^2(T)} + c\, h_T^{k+1}\, \|\phi\|_{W^{k',2}(T)}\, \|v_h\|_{W^{2,2}(T)} \\
&\le c\, h_T^{k+1}\, \|\phi\|_{W^{k+1,2}(T)}\, \|v_h\|_{W^{2,2}(T)}.
\end{aligned}$$

This shows (12a). The estimate (12b) follows from

$$E_T(\phi, v_h) \le \Big(\int_T 1 \,\mathrm{d}x + \sum_{i=1}^N \omega_i \Big)\, \|\phi\|_{L^\infty(T)}\, \|v_h\|_{L^\infty(T)}$$

and the shape regularity of T. □

4.4 Quadrature error on the whole mesh

Now, we use the estimate from Theorem 4.4 in order to bound the quadrature error term in the estimate (9) for the state equation from Lemma 4.3.

Our idea is to work with two different mesh sizes $h_{\text{good}} \ge h_{\text{bad}}$ with

$$|\ln h_{\text{good}}| \sim |\ln h_{\text{bad}}|. \tag{13}$$

We will first derive an error estimate containing both mesh sizes. Later we will balance the error terms.

- Cells T with smooth behavior of the solution $\bar{u}$ have a diameter less than h_{good}. More precisely, we require that $\bar{u} = \bar{p}$ or $\bar{u} = u_a$ or $\bar{u} = u_b$ hold on these cells. In particular, this implies $\bar{u} \in W^{k+1,2}(T)$ if $p \in W^{k+1,2}(T)$. The set of all these cells is denoted by $\mathcal{T}_{\text{good}}$.
- The remaining elements are denoted by $\mathcal{T}_{\text{bad}}$ and have a diameter less than h_{bad}. These are cells where the optimal control has a kink. By N_{bad} we denote the numbers of cells in $\mathcal{T}_{\text{bad}}$.

Theorem 4.5. *We* assume *that the optimal adjoint state $\bar{p}$ belongs to the space $X = W^{k+1,2}(\Omega)$. Together with the above assumptions, we have*

$$\sum_{T\in\mathcal{T}} E_T(\bar{u}, w_h) \le c\,C\left(h_{\text{good}}^{k+1} + (1 + |\ln h_{\text{good}}|)^{1/2} N_{\text{bad}}\, h_{\text{bad}}^3\right)\sqrt{\sum_{T\in\mathcal{T}} \|w_h\|_{H^2(T)}^2}$$

and

$$\sum_{T\in\mathcal{T}} E_T(\bar{u}, w_h) \le c\,C\left(h_{\text{good}}^{k} + (1 + |\ln h_{\text{good}}|)^{1/2} N_{\text{bad}}\, h_{\text{bad}}^3\right)\|w_h\|_{H^1(\Omega)},$$

where $C = \max\left\{\|\bar{p}\|_X, |u_a|, |u_b|\right\}$.

Proof. On the cells belonging to $\mathcal{T}_{\text{good}}$ we use (12a) and obtain

$$\begin{aligned}
\sum_{T\in\mathcal{T}_{\text{good}}} E_T(\bar{u}, w_h) &\le c\,h_{\text{good}}^k \sum_{T\in\mathcal{T}_{\text{good}}} h_T\,\|\bar{u}\|_{W^{k+1,2}(T)}\,\|w_h\|_{W^{2,2}(T)} \\
&\le c\,h_{\text{good}}^k \sqrt{\sum_{T\in\mathcal{T}} \|\bar{u}\|_{W^{k+1,2}(T)}^2}\sqrt{\sum_{T\in\mathcal{T}} h_T^2\,\|w_h\|_{H^2(T)}^2} \qquad (14)\\
&\le c\,h_{\text{good}}^{k+1}\,\max\left\{\|\bar{p}\|_X, |u_a|, |u_b|\right\}\sqrt{\sum_{T\in\mathcal{T}} \|w_h\|_{H^2(T)}^2} \qquad (15)
\end{aligned}$$

with $X = W^{k+1,2}(\Omega)$.

Using an inverse estimate in (14), we find

$$\sum_{T\in\mathcal{T}_{\text{good}}} E_T(\bar{u}, w_h) \le c\,h_{\text{good}}^k\,\max\left\{\|\bar{p}\|_X, |u_a|, |u_b|\right\}\|w_h\|_{H^1(\Omega)}. \qquad (16)$$

Let us now investigate the second type of cells $T \in \mathcal{T}_{\text{bad}}$. We start with the identity

$$E_T(\bar{u}, w_h) = \alpha^{-1}\,E_T(\bar{p}, w_h) + E_T(\bar{u} - \alpha^{-1}\,\bar{p}, w_h).$$

The first term contains only smooth terms and can be estimated in the same way as on the first type of cells. The crucial term is the second one where we use (12b) to obtain

$$\begin{aligned}
E_T(\bar{u} - \alpha^{-1}\,\bar{p}, w_h) &\le c\,h_{\text{bad}}^2\,\|\bar{u} - \alpha^{-1}\,\bar{p}\|_{L^\infty(T)}\,\|w_h\|_{L^\infty(T)} \\
&\le c\,h_{\text{bad}}^3\,\|\bar{p}\|_{W^{1,\infty}(T)}\|w_h\|_{L^\infty(T)}. \qquad (17)
\end{aligned}$$

Here we used the Lipschitz continuity of $\bar{u}$, $\bar{p}$ and the fact $\bar{u} = \alpha^{-1}\,\bar{p}$ for at least one point in the element. This implies $\|\bar{u} - \alpha^{-1}\,\bar{p}\|_{L^\infty(T)} \le c\,h_T\,\|\bar{p}\|_{W^{1,\infty}(T)}$. Summing up the error terms we find

$$\begin{aligned}\sum_{T\in\mathcal{T}_{\text{bad}}} E_T(\bar{u} - \alpha^{-1}\,\bar{p}, w_h) &\le c\,N_{\text{bad}}\,h_{\text{bad}}^3\,\|\bar{p}\|_{W^{1,\infty}(\Omega)}\,\|w_h\|_{L^\infty(\Omega)} \\ &\le c\,N_{\text{bad}}\,h_{\text{bad}}^3\,\|\bar{p}\|_X\,\|w_h\|_{L^\infty(\Omega)},\end{aligned}$$

where we used $X \hookrightarrow W^{3,1}(\Omega) \hookrightarrow W^{1,\infty}(\Omega)$. Since our mesh parameters h_{good} and h_{bad} satisfy (13), we can use [Brenner and Scott, 2008, Lemma (4.9.2)] and find the discrete Sobolev embedding

$$\|w_h\|_{L^\infty(\Omega)} \le c\,(1 + |\ln h_{\text{good}}|)^{1/2}\|w_h\|_{H^1(\Omega)}.$$

This shows

$$\sum_{T\in\mathcal{T}_{\text{bad}}} E_T(\bar{u} - \alpha^{-1}\,\bar{p}, w_h) \le c\,N_{\text{bad}}\,(1 + |\ln h_{\text{good}}|)^{1/2}\,h_{\text{bad}}^3\,\|\bar{p}\|_X\,\|w_h\|_{H^1(\Omega)}.$$

Together with (15) and (16), we obtain the assertion. □

Plugging these estimates into (9), we obtain the following corollary.

Corollary 4.6. *Under the assumptions of Theorem 4.5, we have*

$$\|\bar{y} - \tilde{y}\|_{L^2(\Omega)} \le c\,\big(h_{\text{good}}\,\|\bar{y} - I_h\,\bar{y}\|_{H^1(\Omega)} + h_{\text{good}}^{k+1} + N_{\text{bad}}\,h_{\text{bad}}^3\,(1 + |\ln h_{\text{good}}|)^{1/2}\big),$$

where the constant c depends on $\|\bar{p}\|_X$.

Proof. The estimate follows from Lemma 4.3 and Theorem 4.5. □

5 Error estimates for the optimal control problem

In this section we will combine the results of subsection 4.4 and our main estimate (6). The number of refined cells N_{bad} plays a crucial role in Corollary 4.6. We will require

$$N_{\text{bad}} \le c\,h_{\text{bad}}^{-1},$$

which is reasonable if the kinks are a finite number of curves. A similar assumption is commonly used for error estimates of control constrained problems, see Rösch [2006], Pieper [2015].

Theorem 5.1. *Let us assume $\bar{p} \in W^{k+1,2}(\Omega)$, $N_{\text{bad}} \le c\,h_{\text{bad}}^{-1}$. Then the error estimate*

$$\begin{aligned}&\frac{\alpha}{2}\,\|I_h\,\bar{u} - \bar{u}_h\|_{M_L}^2 + \frac{1}{2}\,\|I_h\,\bar{y} - \bar{y}_h\|_M^2 + \frac{1}{2}\,\|K^{-1}\,M_L\,I_h\,\bar{u} - \bar{y}_h\|_M^2 \\ &\qquad \le c\,\big(h_{\text{good}}^{k+1} + (1 + |\ln h_{\text{good}}|)^{1/2})\,h_{\text{bad}}^2 + h_{\text{good}}\,\|\bar{y} - I_h\,\bar{y}\|_{H^1(\Omega)}\big) \qquad (18)\end{aligned}$$

is satisfied.

This statement follows immediately from Corollary 4.6 combined with (6).

Corollary 5.2. *The choice* $h_{\text{bad}} = h_{\text{good}}^2$ *gives the estimate*

$$\frac{\alpha}{2}\,\|I_h\,\bar u - \bar u_h\|_{M_L}^2 + \frac{1}{2}\,\|I_h\,\bar y - \bar y_h\|_M^2 + \frac{1}{2}\,\|K^{-1}\,M_L\,I_h\,\bar u - \bar y_h\|_M^2 \\ \le c\,(h_{\text{good}}^3 + h_{\text{good}}\,\|\bar y - I_h\,\bar y\|_{H^1(\Omega)}) \tag{19}$$

in the case of P_2*-elements* $(k = k' = 2)$ *and*

$$\frac{\alpha}{2}\,\|I_h\,\bar u - \bar u_h\|_{M_L}^2 + \frac{1}{2}\,\|I_h\,\bar y - \bar y_h\|_M^2 + \frac{1}{2}\,\|K^{-1}\,M_L\,I_h\,\bar u - \bar y_h\|_M^2 \\ \le c\,((1 + |\ln h_{\text{good}}|)^{1/2}\,h_{\text{good}}^4 + h_{\text{good}}\,\|\bar y - I_h\,\bar y\|_{H^1(\Omega)}) \tag{20}$$

in the case of enriched P_3*-elements* $(k = 3,\ k' = 4)$*.*

The logarithmic term in (19) can be dropped due to Corollary 4.6. Further, for the verification of (19), $h_{\text{bad}} = h_{\text{good}}^{3/2+\varepsilon}$, $\varepsilon > 0$, is enough, with $\varepsilon = 0$, one obtains an additional logarithmic term.

Remark 5.3. *Our assumption on the adjoint state* $\bar p$ *is quite strong since one has to expect corner singularities due to the polygonal domain. However, it is possible to combine our approach with a mesh grading at the corners of the polygon. We refer to Apel [1999, 2004] for graded meshes in combination with higher order finite elements. The adjoint state* $\bar p$ *belongs to a corresponding weighted Sobolev space of higher order. The optimal state* $\bar y$ *has the same regularity if one stays away from the kinks of the optimal control* $\bar u$*.*

Remark 5.4. *Let us analyze the interpolation error* $\|\bar y - I_h\,\bar y\|_{H^1(\Omega)}$*, which appears in the different error estimates. Here two different effects are of interest. The first effect is connected with possible corner singularities that reduce the order of the interpolation error. This effect can be again compensated by mesh grading. The second effect is caused by kinks in the optimal control. Then, the control belongs to* $H^{1.5-\varepsilon}(\Omega)$*, for any* $\varepsilon > 0$*, but not to* $H^2(\Omega)$*. Hence, we can expect (up to corner singularities)* $H^{3.5-\varepsilon}(\Omega)$*-regularity of the optimal state* $\bar y$ *but not* $H^4(\Omega)$*-regularity.*

Consequently, we have no restriction for P_2*-elements* $(k = k' = 2)$ *to obtain*

$$\frac{\alpha}{2}\,\|I_h\,\bar u - \bar u_h\|_{M_L}^2 + \frac{1}{2}\,\|I_h\,\bar y - \bar y_h\|_M^2 + \frac{1}{2}\,\|K^{-1}\,M_L\,I_h\,\bar u - \bar y_h\|_M^2 \\ \le c\,h_{\text{good}}^3 \tag{21}$$

The situation is more difficult for the enriched P_3*-elements* $(k = 3,\ k' = 4)$*. The elements containing the kink have the smaller mesh size* h_{bad}*. Consequently, in the regular case in which*

$$\|I_h\bar y - \bar y\|_{H^1(\Omega)} \le c\,h_{\text{good}}^3$$

is satisfied, Corollary 5.2 leads to the error estimate

$$\frac{\alpha}{2}\,\|I_h\,\bar u - \bar u_h\|_{M_L}^2 + \frac{1}{2}\,\|I_h\,\bar y - \bar y_h\|_M^2 + \frac{1}{2}\,\|K^{-1}\,M_L\,I_h\,\bar u - \bar y_h\|_M^2 \\ \le c\,(1 + |\ln h_{\text{good}}|)^{1/2}\,h_{\text{good}}^4. \tag{22}$$

By standard arguments, we obtain the same rates for the approximation of the adjoint states.

6 An algorithmic approach

In the last section we derived an optimal convergence order. However, it was assumed that we know in advance where the kinks are located. If the location of the kinks is unknown, there are at least two simple strategies available:

Algorithm 1

1. Compute a numerical solution for a quasi-uniform mesh with mesh size h_{good}.
2. Refine the mesh in the region where kinks of $\bar{u}$ may occur with a mesh size h_{bad} and compute on the new mesh an improved numerical solution.

Algorithm 2

1. Compute a numerical solution for a quasi-uniform mesh with mesh size h_{good}.
2. Repeat as often it is necessary: Refine all elements which contain a kink of $\bar{u}_h$ and possess a diameter larger then h_{bad}.

Next we will analyze **Algorithm 1** and the regular case addressed in Remark 5.4. Algorithm 2 is rather heuristically and will not be analyzed. Let us define the family of sets

$$K(\varepsilon) = \{x \in \Omega : |\bar{p} - \alpha u_a| \leq \varepsilon \text{ or } |\bar{p} - \alpha u_b| \leq \varepsilon\}.$$

for arbitrary $\varepsilon > 0$.

Let us denote the numerical solution of the first step with $(\tilde{y}_h, \tilde{u}_h, \tilde{p}_h)$. Since the mesh is quasi-uniform, we have $h_{\text{good}} \sim h_{\text{bad}}$. We can directly apply the error estimate (18) to obtain

$$\|I_h\, \bar{y} - \tilde{y}_h\|_M = \|I_h\, \bar{y} - \tilde{y}_h\|_{L^2(\Omega)} \leq c\,(1 + |\ln h_{\text{good}}|)^{1/2}\, h_{\text{good}}^2.$$

From that we get easily

$$\|\bar{y} - \tilde{y}_h\|_{L^2(\Omega)} \leq c\,(1 + |\ln h_{\text{good}}|)^{1/2}\, h_{\text{good}}^2.$$

The regularity of the optimal adjoint state and the above inequality imply

$$\|\bar{p} - \tilde{p}_h\|_{L^\infty(\Omega)} \leq c\,(1 + |\ln h_{\text{good}}|)^{1/2}\, h_{\text{good}}^2.$$

The last estimate means that the kinks of the optimal control $\bar{u}$ are contained in the set $K(c\,(1+|\ln h_{\text{good}}|)^{1/2}\, h_{\text{good}}^2)$. Let us *assume* that the area of elements containing this set

$$T_K = \bigcup_T \{T \cap K(c\,(1 + |\ln h_{\text{good}}|)^{1/2}\, h_{\text{good}}^2) \neq \emptyset\}$$

can be limited by

$$|T_K| \leq c\,\gamma\,(1 + |\ln h_{\text{good}}|)^{1/2}\, h_{\text{good}}^2)$$

with a certain positive γ. We remark that the similar property $|K(\varepsilon)| \le \gamma\,\varepsilon$ is frequently used, in particular for the approximation of bang-bang controls, see Deckelnick and Hinze [2012], Wachsmuth and Wachsmuth [2011].

This region is discretized in the second step with the mesh size h_{bad}. The number of elements needed is proportional to $(1 + |\ln h_{\text{good}}|)^{1/2}\, h_{\text{good}}^2 / h_{\text{bad}}^2$. Hence, the number of needed element is not essentially increasing if

$$(1 + |\ln h_{\text{good}}|)^{1/2}\, h_{\text{good}}^4 \le c\, h_{\text{bad}}^2.$$

For the quadratic finite elements we can choose $h_{\text{bad}} \sim h_{\text{good}}^{3/2}$. In the case of enriched cubic elements we use $h_{\text{bad}} \sim (1 + |\ln h_{\text{good}}|)^{1/4}\, h_{\text{good}}^2$.

On the refined mesh, a new finite element solution $(\bar y_h, \bar u_h, \bar p_h)$ will be computed. We can apply directly (21) and (22) to the new discrete solution to obtain

$$\begin{aligned}\frac{\alpha}{2}\,\|I_h\,\bar u - \bar u_h\|_{M_L}^2 + \frac{1}{2}\,\|I_h\,\bar y - \bar y_h\|_M^2 + \frac{1}{2}\,\|K^{-1}\,M_L\,I_h\,\bar u - \bar y_h\|_M^2 \\ \le c\,(1 + |\ln h_{\text{good}}|)^{1/2} h_{\text{good}}^3 \end{aligned} \tag{23}$$

in the case of P_2-elements ($k = k' = 2$) and

$$\begin{aligned}\frac{\alpha}{2}\,\|I_h\,\bar u - \bar u_h\|_{M_L}^2 + \frac{1}{2}\,\|I_h\,\bar y - \bar y_h\|_M^2 + \frac{1}{2}\,\|K^{-1}\,M_L\,I_h\,\bar u - \bar y_h\|_M^2 \\ \le c\,(1 + |\ln h_{\text{good}}|)\, h_{\text{good}}^4 \end{aligned} \tag{24}$$

in the case of enriched P_3-elements ($k = 3$, $k' = 4$).

7 Numerical experiments

In this section, we present a numerical example which illustrates the convergence results (21) and (22).

On the unit square $\Omega = (0,1)^2$ we consider the optimal control problem (**P**) with $\alpha = 0.05$, $y_d(x_1, x_2) = \exp x_1\, \sin(x_2)$ and $u_a = -1.5$, $u_b = 1.0$.

As described in Section 4.3, we are interested in two particular finite elements: standard P^2 elements and enriched P^3 elements as constructed in Cohen et al. [2001]. We emphasize that the lumped mass matrix M_L is singular for P^2 elements. This is not a obstruction for the analysis, since we only required that M_L is positive semidefinite. To the contrary, this is beneficial for the numeric implementation, since ($\mathbf{P}_h$) does not depend on the vertex values of u_h and, thus, we can work with less degrees of freedom for u_h.

The enriched P^3 elements consists of the standard cubic finite element space and this space is enriched by all bubble functions of degree at most four. The associated quadrature rule is exact for polynomials of degree at most 5 if the Lagrange nodes are chosen suitably, cf. [Cohen et al., 2001, Lemma 4.3].

In the numerical implementation, we used an algorithm similar to Algorithm 2 in Section 6 which is coupled with a nested iteration. We start with a coarse initial mesh and perform the following in each step:

- We set $h_{\text{good}} = \max_{T\in\mathcal{T}} \operatorname{diam}(T)$ and $h_{\text{bad}} = h_{\text{good}}^{3/2}$ (P^2 elements) or $h_{\text{bad}} = h_{\text{good}}^2$ (P^3 elements).
- If there is any cell T with $\operatorname{diam}(T) > h_{\text{bad}}$ and on which $\operatorname{Proj}_{[u_a,u_b]}(\bar{p}_h/\alpha)$ has a kink, we refine all those cells (local refinement).
- Otherwise, we refine all cells T with $\operatorname{diam}(T) > h_{\text{good}}/2$ (global refinement).

This algorithm is implemented in the finite element toolbox FEniCS, cf. Alnæs et al. [2015], by using the geometric multigrid implementation from Ospald [2012].

The computational results are shown in Figure 1 and Figure 2. Since an analytical

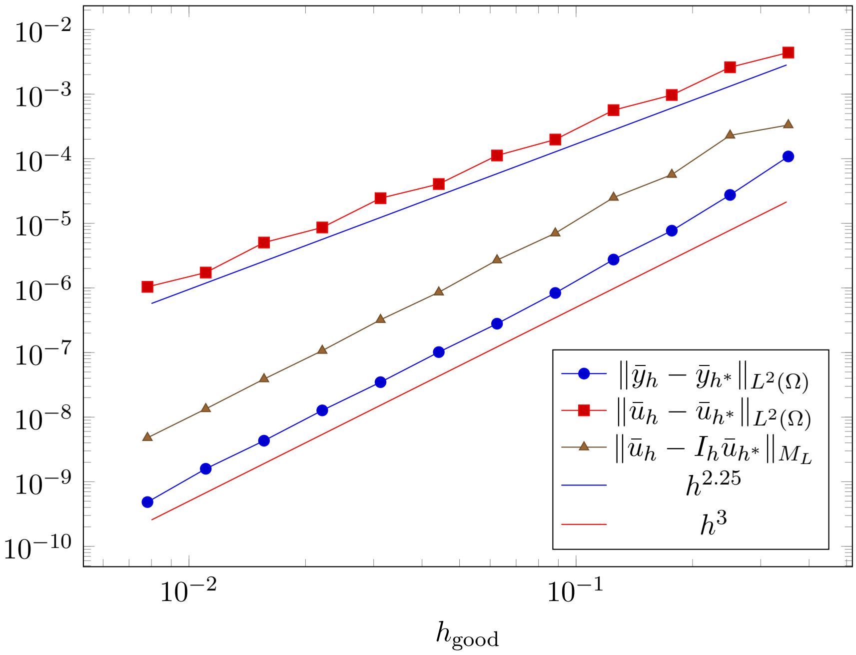

Figure 1: Errors in the control and state for the discretization with P^2 elements.

solution for the problem under consideration is not known, we used a fine grid solution as a reference for the computed errors. In fact, we used

$$\bar{y}_{h^*} = \bar{y}_{h'} \qquad \text{and} \qquad \bar{u}_{h^*} = \operatorname{Proj}_{[u_a,u_b]} \frac{\bar{p}_{h'}}{\alpha},$$

where $(\bar{u}_{h'}, \bar{y}_{h'}, \bar{p}_{h'})$ is the solution of $(\mathbf{P}_h)$ on a finer grid. As predicted in (21) and (22), we see convergence of order h^{k+1} for the errors

$$\|\bar{y}_h - \bar{y}_{h^*}\|_{L^2(\Omega)} \qquad \text{and} \qquad \|\bar{u}_h - I_h\bar{u}_{h^*}\|_{M_L}.$$

Note that the error of the control $\|\bar{u}_h - \bar{u}_{h^*}\|_{L^2(\Omega)}$ converges significantly slower. For quasiuniform meshes and P^1-elements one knows a convergence of order $h^{3/2}$, cf. Rösch

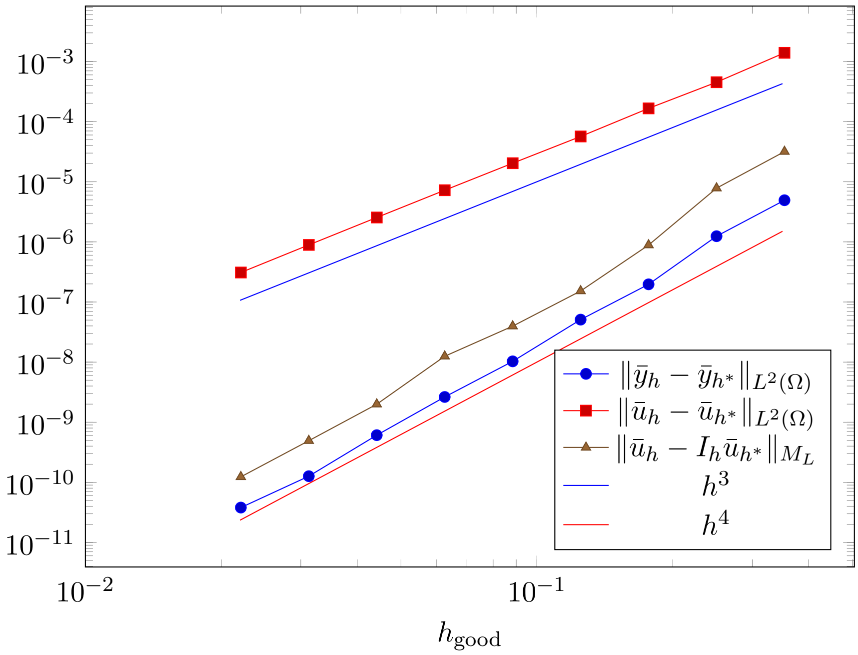

Figure 2: Errors in the control and state for the discretization with enriched P^3 elements.

[2006], Casas and Mateos [2008]. Clearly, these techniques can be extended to more general meshes and larger classes of finite elements. However, the best approximation on our mesh is of order $h_{\text{bad}}^{3/2}$ due to the presence of the kink. Exactly this order is observed in our numerical tests. Because of the coupling of h_{good} and h_{bad}, we obtain $h_{\text{bad}}^{3/2} = (h_{\text{good}}^{3/2})^{3/2} = h_{\text{good}}^{2.25}$ for the case of P^2 elements and $h_{\text{bad}}^{3/2} = (h_{\text{good}}^{2})^{3/2} = h_{\text{good}}^{3}$ for the case of enriched P^3 elements. This is essentially smaller as the convergence order of our new method (3 respectively 4).

References

Martin S. Alnæs, Jan Blechta, Johan Hake, August Johansson, Benjamin Kehlet, Anders Logg, Chris Richardson, Johannes Ring, Marie E. Rognes, and Garth N. Wells. The fenics project version 1.5. *Archive of Numerical Software*, 3(100), 2015. doi: 10.11588/ans.2015.100.20553.

Walter Alt, Nils Bräutigam, and Arnd Rösch. Error estimates for finite element approximations of elliptic control problems. *Discussiones Mathematicae. Differential Inclusions, Control and Optimization*, 27(1):7–22, 2007.

Th. Apel. *Anisotropic finite elements: Local estimates and applications*. Advances in Numerical Mathematics. Teubner, Stuttgart, 1999.

Th. Apel. Interpolation in h-version finite element spaces. In E. Stein, R. de Borst, and T. J. R. Hughes, editors, *Encyclopedia of Computational Mechanics*, volume 1 Fundamentals, pages 55–72. Wiley, Chichester, 2004.

N. Arada, E. Casas, and F. Tröltzsch. Error estimates for a semilinear elliptic optimal control problem. *Computional Optimization and Applications*, 23:201–229, 2002.

S. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods.* Springer, New York, third edition, 2008.

Eduardo Casas and Mariano Mateos. Error estimates for the numerical approximation of Neumann control problems. *Computational Optimization and Applications*, 39(3): 265–295, 2008.

Eduardo Casas, Roland Herzog, and Gerd Wachsmuth. Approximation of sparse controls in semilinear equations by piecewise linear functions. *Numerische Mathematik*, 122(4): 645–669, 2012. ISSN 0029-599X. doi: 10.1007/s00211-012-0475-7.

Philippe G. Ciarlet. *The finite element method for elliptic problems.* North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978. ISBN 0-444-85028-7. Studies in Mathematics and its Applications, Vol. 4.

G. Cohen, P. Joly, J. E. Roberts, and N. Tordjman. Higher order triangular finite elements with mass lumping for the wave equation. *SIAM Journal on Numerical Analysis*, 38 (6):2047–2078, 2001. ISSN 0036-1429. doi: 10.1137/S0036142997329554.

Klaus Deckelnick and Michael Hinze. A note on the approximation of elliptic control problems with bang-bang controls. *Computational Optimization and Applications. An International Journal*, 51(2):931–939, 2012. ISSN 0926-6003. doi: 10.1007/s10589-010-9365-z.

R. Falk. Approximation of a class of optimal control problems with order of convergence estimates. *J. Math. Anal. Appl.*, 44:28–47, 1973.

T. Geveci. On the approximation of the solution of an optimal control problem governed by an elliptic equation. *R.A.I.R.O. Analyse numeriqué*, 13:313–328, 1979.

Michael Hinze. A variational discretization concept in control constrained optimization: the linear-quadratic case. *Computational Optimization and Applications*, 30(1):45–61, 2005.

Christian Meyer and Arnd Rösch. Superconvergence properties of optimal control problems. *SIAM Journal on Control and Optimization*, 43(3):970–985, 2004.

F. Ospald. Implementation of a geometric multigrid method for FEniCS and its application. Diploma thesis, Technische Universität Chemnitz, Germany, 2012.

Konstantin Pieper. *Finite element discretization and efficient numerical solution of elliptic and parabolic sparse control problems.* PhD thesis, Technische Universität München, Germany, 2015.

A. Rösch. Error estimates for linear-quadratic control problems with control constraints. *Optimization Methods and Software*, 21(1):121–134, 2006.

René Schneider and Gerd Wachsmuth. Achieving optimal convergence order for fem in control constrained optimal control problems. *Proceedings in Applied Mathematics and Mechanics (PAMM)*, 15(1):731–734, 2015.

René Schneider and Gerd Wachsmuth. A-posteriori error estimation for control-constrained, linear-quadratic optimal control problems. *SIAM Journal on Numerical Analysis*, 54(2):1169–1192, 2016. doi: 10.1137/15M1020460.

D. Sevilla and D. Wachsmuth. Polynomial integration on regions defined by a triangle and a conic. In *ISSAC 2010—Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation*, pages 163–170. ACM, New York, 2010. doi: 10.1145/1837934.1837968.

Gerd Wachsmuth and Daniel Wachsmuth. Convergence and regularization results for optimal control problems with sparsity functional. *ESAIM. Control, Optimisation and Calculus of Variations*, 17(3):858–886, 2011. ISSN 1292-8119. doi: 10.1051/cocv/2010027.