Finite Element Methods for Elliptic Optimal Control Problems with General Tracking

SeongHee Jeong
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://repository.lsu.edu/gradschool_dissertations

Part of the Numerical Analysis and Computation Commons

Recommended Citation
https://repository.lsu.edu/gradschool_dissertations/6096

This Dissertation is brought to you for free and open access by the Graduate School at LSU Scholarly Repository. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Scholarly Repository. For more information, please contact gradetd@lsu.edu.
FINITE ELEMENT METHODS FOR ELLIPTIC
OPTIMAL CONTROL PROBLEMS WITH
GENERAL TRACKING

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Mathematics

by
SeongHee Jeong
B.S., Sookmyung Women’s University, 2013
M.S., Yonsei University, 2016
May 2023
Acknowledgments

Words cannot express my deepest gratitude to my advisor and chair of my committee Professor Susanne C. Brenner for her invaluable patience and unwavering guidance. I also could not have undertaken this journey without Professor Li-Yeng Sung who generously encouraged and inspired me. This endeavor would not have been possible without them. I was fortunate to work with Dr. Zhiyu Tan who kindly helped and inspired me. Thanks to his cubic $C^0$ interior penalty method code, I was able to conduct numerical simulations in Chapter 7.

I would like to thank my dissertation committee, Professor Hongchao Zhang and Professor Luis A. Escobar, for their insightful comments and expertise. A special thanks to National Science Foundation for the support under grants DMS-19-13035 and DMS-22-08404.

Lastly, I wholeheartedly thank my parents for their unconditional support. Their belief in me has kept my spirits and motivation high during this process. Especially, I could not have withstood the hard time without my husband Seulip who has stayed by my side at any time.
# Table of Contents

Acknowledgments ................................................................. ii

List of Tables ................................................................. v

List of Figures ............................................................... vi

Abstract ................................................................. viii

Chapter 1. Introduction ....................................................... 1
  1.1. Optimal Control Problems ........................................... 1
  1.2. Numerical Methods .................................................... 2
  1.3. Existing Work .......................................................... 2
  1.4. Our Approach .......................................................... 4
  1.5. Dissertation Organization ......................................... 6

Chapter 2. Preliminaries .................................................... 9
  2.1. Sobolev Spaces .......................................................... 9
  2.2. Elliptic Equalities ...................................................... 13
  2.3. Elliptic Variational Inequalities ..................................... 16
  2.4. Regularity .............................................................. 19

Chapter 3. Elliptic Optimal Control Problems .......................... 23
  3.1. Elliptic Optimal Control Problems with Pointwise Constraints .................................................. 23
  3.2. Elliptic Optimal Control Problem with General Tracking and Pointwise State Constraints .................................................. 26

Chapter 4. Finite Element Methods ....................................... 38
  4.1. Fundamental Theory .................................................... 38
  4.2. $C^0$ Interior Penalty Methods ....................................... 44

Chapter 5. $C^0$ Interior Penalty Methods for the Elliptic Optimal Control Problem with General Tracking and Pointwise State Constraints .................................................. 52
  5.1. Discrete Problem ....................................................... 52
  5.2. Convergence Analysis .................................................. 55

Chapter 6. Numerical Implementation .................................... 61
  6.1. A Primal-Dual Active Set Algorithm ............................... 61
  6.2. Numerical Experiments ................................................ 64

Chapter 7. Elliptic Optimal Control Problems with State and Control Constraints .................................................. 79
  7.1. Continuous Problem .................................................... 79
  7.2. Discrete Problem ....................................................... 80
  7.3. Numerical Simulations .................................................. 81
Chapter 8. Conclusion and Future Work ........................................ 86

Bibliography .................................................................................. 90

Vita ............................................................................................... 96
List of Tables

6.1 Relative error of the state versus mesh size $h$ and orders of convergence for Example 6.2.1. ................................................................. 69

6.2 Relative error of the control versus mesh size $h$ and orders of convergence for Example 6.2.1. ................................................................. 69

6.3 Error of the state versus mesh size $h$ and orders of convergence for Example 6.2.2.a. ................................................................. 71

6.4 Error of the state versus mesh size $h$ and orders of convergence for Example 6.2.2.b. ................................................................. 72

6.5 Error of the state versus mesh size $h$ and orders of convergence for Example 6.2.2.c. ................................................................. 74

6.6 Error of the state versus mesh size $h$ and orders of convergence for Example 6.2.2.d. ................................................................. 75

6.7 Relative error of the state versus mesh size $h$ and orders of convergence for Example 6.2.3. ................................................................. 77

6.8 Relative error of the control versus mesh size $h$ and orders of convergence for Example 6.2.3. ................................................................. 78

7.1 Relative error of the state versus mesh size $h$ and orders of convergence for Example 6.3.1. ................................................................. 83

7.2 Relative error of the control versus mesh size $h$ and orders of convergence for Example 6.3.1. ................................................................. 83

7.3 Relative error of the state versus mesh size $h$ and orders of convergence for Example 6.3.2 ................................................................. 84

7.4 Relative error of the control versus mesh size $h$ and orders of convergence for Example 6.3.2 ................................................................. 84
List of Figures

1.1 Mesh generation ................................................................. 2
3.1 Distributed control in Ω ......................................................... 23
4.1 linear, quadratic, and cubic Lagrange triangle ............................. 40
4.2 Degrees of freedom of the HCT element .................................... 44
4.3 The normal vector $n_e$ in average and jump across the edge $e$ .... 46
6.1 Graphs of optimal state and optimal control with $h = 2^{-5}$ for Example 6.2.1 .......................... 70
6.2 Active set with $h = 2^{-5}$ for Example 6.2.1 .................................. 70
6.3 Trackings for Example 6.2.2.a (left top), Example 6.2.2.b (right top), Example 6.2.2.c (left bottom), and Example 6.2.2.d (right bottom) .................................. 71
6.4 Graphs of optimal state and optimal control for Example 6.2.2.a .................................. 72
6.5 Active set for Example 6.2.2.a .................................................. 72
6.6 Graphs of optimal state and optimal control for Example 6.2.2.b .................................. 73
6.7 Active set for Example 6.2.2.b .................................................. 73
6.8 Graphs of optimal state and optimal control for Example 6.2.2.c .................................. 74
6.9 Active set for Example 6.2.2.c .................................................. 76
6.10 Graphs of optimal state and optimal control for Example 6.2.2.d .................................. 76
6.11 Active set for Example 6.2.2.d .................................................. 76
6.12 Tracking for Example 6.2.3 .................................................. 77
6.13 Graphs of optimal state and optimal control with $h = 2^{-7}$ for Example 6.2.3 .................................. 78
6.14 Active set with $h = 2^{-7}$ for Example 6.2.3 .................................. 78
7.1 Degrees of freedom of the cubic Hermite element ......................... 80
7.2 Tracking for Example 6.3.1 and Example 6.3.2 .................................. 82
7.3 Graphs of optimal state and optimal control with $h = 2^{-5}$ for Example 6.3.1 . . . 83
7.4 Active set with $h = 2^{-5}$ for Example 6.3.1 . . . . . . . . . . . . . . . . . . . . . . 83
7.5 Graphs of optimal state and optimal control with $h = 2^{-5}$ for Example 6.3.2 . . 85
7.6 Active set with $h = 2^{-5}$ for Example 6.3.2 . . . . . . . . . . . . . . . . . . . . . . 85
Abstract

This dissertation concerns a linear-quadratic elliptic distributed optimal control problem with pointwise state constraints in two spatial dimensions, where the cost function tracks the state at points, curves and regions of a domain.

First we explore the elliptic optimal control problem subject to pointwise control constraints. This problem is reduced into a problem that only involves the control. The solution of the reduced problem is characterized by a variational inequality. Then we introduce the elliptic optimal control problem with general tracking and pointwise state constraints. Here we reformulate the optimal control problem into a problem that only involves the state, which is equivalent to a fourth order variational inequality. We derive the Karush-Kuhn-Tucker conditions from the variational inequality and find the regularity result of the solution.

The reduced minimization problem is solved by a $C^0$ interior penalty method. The $C^0$ interior penalty methods are very effective for fourth order problems and much simpler than $C^1$ finite element methods. The discrete problem is a quadratic program with simple box constraints which can be solved efficiently by the primal-dual active set algorithm. We provide a convergence analysis and demonstrate the performance of the method through several numerical experiments.
Chapter 1. Introduction

1.1. Optimal Control Problems

Optimal control theory has been studied in the past few decades and developed rapidly into an important area of applied mathematics. There are numerous applications of this theory such as robotics, chemical processes, heat conduction, fluid flows, etc.

Optimal control problems consist of

• cost functional \( J \) to be minimized

• control variable \( u \)

• state variable \( y \)

• various constraints that have to be satisfied.

Then an optimal control problem is to find

\[
(\bar{y}, \bar{u}) = \arg\min J(y, u)
\]

subject to some constraints. There are many different types of constraints: ordinary differential equations, partial differential equations, pointwise equality, pointwise inequality constraints, etc. We focus on the optimal control problems subject to linear elliptic partial differential equation and pointwise state constraints. The elliptic partial differential equations can be employed for variety applications such as heat conduction and diffusion processes.

To show the existence of optimal controls, the optimal control problem is reformulated into a reduced optimization problem in terms of the state \( y \) or the control \( u \). For the minimization problem, we find an equivalent variational inequality, which is called first-order optimality condition. By introducing Lagrange multipliers, the variational inequality
can be reformulated and then Karush-Kuhn-Tucker conditions are obtained.

1.2. Numerical Methods

Among numerous numerical methods, the finite element method is a popular method for solving differential equations arising in mathematical modeling. The basic concept of the finite element method is to divide a domain of a problem into a collection of subdomains. We refer to individual subdomains as elements. These elements can be triangles, quadrilaterals or polygons in two dimensions, and tetrahedrons or hexahedrons in three dimensions.

![Figure 1.1. Mesh generation](image)

Finite element methods work well for complex geometries, and one can use higher order elements to get a high order of convergence.

1.3. Existing Work

Optimal control problems with pointwise control constraints have been studied with various approaches (cf. [5, 58]). However, the literature concerning pointwise state constraints is relatively small. In [13, 14, 17, 18, 23, 25, 26, 34, 35, 50, 51, 54, 60, 64, 67],...
an optimal control problem

\[
\min_{(y,u)} \frac{1}{2} \left[ \|y - y_d\|^2_{L^2(\Omega)} + \beta \|u\|^2_{L^2(\Omega)} \right].
\]

(1.1)

is considered subject to the elliptic partial differential equations and pointwise state constraints. Here \(\beta\) is a positive constant.

Interior point methods have been used to solve the optimal control problem with pointwise state constraints in [54, 67]. By introducing multipliers, the dual problem has been obtained and analyzed in [64, 51, 50]. In [35], optimal control problem was reformulated into a reduced minimization problem with only control variable. On the other hand, in [60], the reformulated reduced problem only with state variable is obtained whose optimality condition is characterized by a fourth order elliptic variational inequality. The problem where both state and control constraints exist has been discussed in [22, 37, 49, 63, 12].

There have been numerous studies to investigate the variational inequality which has played an important role in the optimal control theory (cf. [31, 32, 45, 46, 56, 57, 59, 42, 55]). For example, a reduced optimal control problem involving only state variable can be written as an equivalent elliptic variational inequality [60, 12, 13, 22, 25, 26]. This first-order optimality condition leads to the Karush-Kuhn-Tucker conditions which are the key to prove the regularity results.

To find an approximate solution to the optimal control problem, various finite element methods have been used (cf. [65]). In [3, 14, 6, 18, 36], linear finite element methods have been used. Conforming finite element methods guarantee the convergence by Galerkin orthogonality but they require many degrees of freedom if a high order prob-
lem is involved. As one of the nonconforming finite element methods, a Morley finite element method was used in [12]. Nonconforming finite element methods are simpler than conforming methods because finite element functions only need to satisfy weak continuity conditions. However, they are only low order elements which are not efficient for capturing smoothness. Also, the virtual element method was used in [23]. This method allows to use general polygonal meshes, and is more flexible on the discrete space. Finally, $C^0$ interior penalty methods were used in [13, 17, 21, 22, 25, 26] to solve the optimal control problem.

The majority of prior research has focused on the standard elliptic optimal control problem (1.1). The problem with point tracking was investigated in [3, 4, 6, 29, 30, 36]:

$$\min_{(y,u)} \frac{1}{2} \left[ \sum_j (y(p_j) - y_d)^2 + \beta \|u\|^2_{L^2(\Omega)} \right].$$

(1.2)

Note that in all these previous studies on point tracking problems, only the control constraints are considered.

In this dissertation, we consider an elliptic optimal control problem with a general cost functional and utilize the $C^0$ interior penalty method to solve the discrete minimization problem.

1.4. Our Approach

In the standard optimal control problem with pointwise state constraints, we want the state $y$ to be close to the desired state $y_d$ in an entire domain. However, we asked ourselves what if we want $y$ to be close to a desired state only in a designated part of the domain?

This led us to introduce the general cost functional tracks points, curves and re-
regions in a two dimensional domain $\Omega$. To be specific, the cost functional $J(y, u)$ is defined

$$J(y, u) = \frac{1}{2} [G(y) + \beta \| u \|_{L_2(\Omega)}^2],$$

where

$$G(y) = \sum_{j=1}^{J} (y(p_j) - y_0(p_j))^2 w_0(x_j) + \sum_{l=1}^{L} \int_{C_l} (y - y_1)^2 w_1 ds + \int_{\Omega} (y - y_2)^2 w_2 dx.$$ 

The desired states (target functions) $y_0, y_1$ and $y_2$ are defined on a set $P$ of points $\{p_1, \ldots, p_J\}$, the union $C$ of the curves $C_1, \ldots, C_L$ and the domain $\Omega$, respectively. The weight functions $w_0, w_1$ and $w_2$ allow preference for the desired states. The pointwise state constraints are given by

$$\psi_- \leq y \leq \psi_+ \text{ a.e. in } \Omega,$$

and the elliptic state equation is

$$\int_{\Omega} \nabla y \cdot \nabla z dx = \int_{\Omega} uz dx \text{ } \forall z \in H^1_0(\Omega).$$

Our goal is to find an optimal pair of state and control which minimizes $J(y, u)$ subject to the constraints. We reformulate the optimal control problem into the reduced minimization problem which involves only state variable which is equivalent to a fourth order variational inequality.

To find an approximate solution to the minimization problem, we use a $C^0$ interior penalty method. $C^0$ interior penalty methods [11, 20, 27, 43] are discontinuous Galerkin methods that can overcome the shortcomings of the classical approaches, such as $C^1$ finite elements, non-conforming finite elements, and mixed formulations. Here are some advantages of the $C^0$ interior penalty method.
• It is effective for fourth order elliptic boundary value problems.

• It uses $P_k$ Lagrange triangular finite elements for $k \geq 2$.

• It is also useful for capturing smooth solutions.

• It is much simpler than $C^1$ finite elements.

The discrete problem obtained by using $C^0$ interior penalty methods is a quadratic program with simple box constraints and will be solved by a primal-dual active set algorithm (cf. [7, 8, 9]). Numerical experiments have been conducted by using MATLAB.

1.5. Dissertation Organization

The goal of this dissertation is to construct and analyze a finite element method that solves a fourth order variational inequality obtained from an optimal control problem. This dissertation is organized as follows.

In Chapter 2 we review the properties of the Sobolev spaces and discuss weak solutions to elliptic partial differential problems. We cover the elliptic equalities and elliptic variational inequalities. Regularity results on the solutions of the second-order and fourth-order elliptic problems are discussed.

In Chapter 3 we consider elliptic optimal control problems with pointwise constraints. First we take a look at the elliptic optimal control problems with control constraints and then we present the existence and uniqueness of their solutions. First-order necessary optimality conditions are covered as well. We then move to the elliptic optimal control problems with pointwise state constraints and introduce the general cost function which tracks points, curves and regions of a domain. The elliptic optimal control problem is reduced into the minimization problem that involves only the state variable. The fourth
order variational inequality is obtained from the reformulated minimization problem and the equivalent Karush-Kuhn-Tucker (KKT) conditions are derived. The local and global regularity results for the state are provided at the end of the chapter.

In Chapter 4 we deal with finite element methods. Some background material on the theory of finite element methods are presented such as Galerkin orthogonality, Céa’s Lemma, finite element spaces etc. We review the interpolation theory and the interpolation error estimates. Section 4.2 is concerned with $C^0$ interior penalty methods for a fourth order elliptic boundary value problem. We show the well-posedness of the discrete problem. The interpolation operator and the enriching operator are discussed and error estimates are also obtained.

In Chapter 5 we apply the $C^0$ interior penalty methods to the elliptic optimal control problem with general tracking. We show well-posedness of the method and introduce the mesh-dependent norm. The error estimates in other norms are also obtained. We provide the convergence analysis for the optimal state and optimal control.

Chapter 6 contains numerical implementations for the case where pointwise state constraints exist. We consider the primal-dual active set algorithm which will be used to solve the box constrained problem obtained from the discrete KKT conditions. The details of the algorithm and the convergence result are addressed. By using $P_2$ Lagrange finite element, we run various numerical experiments in different settings. The figures of tracking, the optimal state and the optimal control are plotted for each example. Numerical simulations confirm the theoretical estimates.

In Chapter 7 we conduct numerical simulations for the optimal control problem with both state and control constraints. A modified cubic Hermite finite element method
is used to handle the control constraints. We carry out numerical experiment for the resulting $C^0$ interior penalty method. Through these numerical experiments, we expect that the optimal control problem with both state and control constraints will have similar convergence analysis to the one with only state constraints.

Finally, in Chapter 8, we summarize our work and suggest possible future work.
Chapter 2. Preliminaries

In this section, we review the basic theories, notations and definitions appearing throughout this dissertation (cf. [1, 61, 19, 44, 68]).

2.1. Sobolev Spaces

Let $\Omega$ be a domain in $\mathbb{R}^n$, where $n$ is a positive integer. For a Lebesgue measurable function $f$ in $\Omega$, we denote the Lebesgue integral of $f$ by

$$\int_{\Omega} f(x) \, dx.$$ 

We introduce the norm

$$\|f\|_{L^p(\Omega)} = \left( \int_{\Omega} |f|^p \, dx \right)^{1/p}, \quad 1 \leq p < \infty.$$ 

The space of Lebesgue measurable functions defined on $\Omega$ is denoted by

$$L^p(\Omega) = \left\{ f : \int_{\Omega} |f|^p \, dx < \infty \right\}.$$ 

Endowed with the norm, $L^p(\Omega)$ becomes a Banach space. Moreover, $L^\infty(\Omega)$ is the Banach space of all Lebesgue measurable and essentially bounded functions, equipped with the norm

$$\|f\|_{L^\infty(\Omega)} = \text{ess sup}_{x \in \Omega} |f(x)|.$$ 

Remark 2.1. For $p = 2$, the space $L^2(\Omega)$ is a Hilbert space with respect to the following inner product

$$(u, v)_{L^2(\Omega)} = \int_{\Omega} uv \, dx.$$ 

Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be an $n$-tuple of nonnegative integers $\alpha_i$. For $x \in \mathbb{R}^n$, we denote

$$D_i = \frac{\partial}{\partial x_i},$$
as a partial differential operator, and

\[ D^\alpha = D_1^{\alpha_1} \cdots D_n^{\alpha_n}, \]

as a differential operator of order \(|\alpha| = \sum_{i=1}^n \alpha_i|.

We denote \(C^k(\Omega)\), where \(k\) is a positive integer, the vector space of real-valued functions on \(\Omega\) which have continuous partial derivatives up to order \(k\). The space of \(k\)-times continuously differentiable functions with compact support in \(\Omega\) is denoted by \(C_0^k(\Omega)\), \(k \in \mathbb{N} \cup \{0, \infty\}\).

**Definition 2.1.** The set of locally integrable functions is defined by

\[ L^1_{loc}(\Omega) = \{ f : f \in L^1(K) \}, \]

where \(K\) is a subset of \(\Omega\) such that closure of \(K\) is compact.

**Definition 2.2.** If \(f \in L^1_{loc}(\Omega)\), then for any \(\phi \in C_0^\infty(\Omega)\),

\[ \int_{\Omega} f(x) D^\alpha \phi(x) \, dx = (-1)^{|\alpha|} \int_{\Omega} g(x) \phi(x) \, dx, \]

and \(g = D^\alpha f\) is called the weak derivative of \(f\).

We can extend this definition to the case with \(f \in C^{|\alpha|}(\Omega)\).

**Definition 2.3.** Suppose \(f \in L^1_{loc}(\Omega)\). For nonnegative integer \(k\), we define the norm

\[ \| f \|_{W^{k,p}(\Omega)} = \left( \sum_{|\alpha| \leq k} \int_{\Omega} |D^\alpha f(x)|^p \, dx \right)^{1/p}, \tag{2.1} \]

where \(|\alpha| \leq k\). If \(p = \infty\),

\[ \| f \|_{W^{k,\infty}(\Omega)} = \max_{|\alpha| \leq k} \| D^\alpha f \|_{L^\infty(\Omega)}. \tag{2.2} \]

The Sobolev space \(W^{k,p}(\Omega)\) is defined by

\[ W^{k,p}(\Omega) = \{ f \in L^1_{loc}(\Omega) : \| f \|_{W^{k,p}(\Omega)} < \infty \} \tag{2.3} \]
for $1 \leq p \leq \infty$.

**Definition 2.4.** Let $k$ be a nonnegative integer and $f \in W^{k,p}(\Omega)$. For $1 \leq p < \infty$, we define the Sobolev semi-norm as

$$
|f|_{W^{k,p}(\Omega)} = \left( \sum_{|\alpha|=k} \|D^\alpha f\|_{L^p(\Omega)}^p \right)^{1/p}.
$$

Also, for $p = \infty$,

$$
|f|_{W^{k,\infty}(\Omega)} = \max_{|\alpha|=k} \|D^\alpha f\|_{L^\infty(\Omega)}.
$$

**Theorem 2.2.** (cf. [1]) The Sobolev space $W^{k,p}(\Omega)$ is a Banach space.

In particular,

$$
H^k(\Omega) := W^{k,2}(\Omega)
$$

is a Hilbert space with respect to the inner product

$$(u, v)_{H^k(\Omega)} = \sum_{|\alpha| \leq k} \int_\Omega (D^\alpha u)(D^\alpha v) \, dx.$$

We write $\| \cdot \|_{H^k(\Omega)}$ and $| \cdot |_{H^k(\Omega)}$ as Hilbert space norm and semi-norm, respectively.

**Definition 2.5.** For nonnegative integer $k$, we define $H^k_0(\Omega)$ to be the closure of $C_0^\infty(\Omega)$ in $H^k(\Omega)$.

**Theorem 2.3.** (cf. [44]) Suppose $\Omega$ is an open set. Then $C^\infty(\Omega) \cap W^{k,p}(\Omega)$ is dense in $W^{k,p}(\Omega)$ for $1 \leq p < \infty$.

Let $f$ be a real-valued function. Then $f$ is called Lipschitz continuous if there exists a positive constant $C$ such that

$$
|f(x) - f(y)| \leq C|x - y| \quad \forall \, x, y \in \mathbb{R}.
$$

We say $\Omega$ is a Lipschitz domain if the boundary $\partial \Omega$ is sufficiently smooth in the sense that $\partial \Omega$ is locally the graph of a Lipschitz continuous function.
Theorem 2.4. (cf. [44]) Let $\Omega$ be any Lipschitz open subset of $\mathbb{R}^n$. Then $C^\infty(\Omega)$ is dense in $W^{k,p}(\Omega)$ for $1 \leq p < \infty$.

We will review the Sobolev embedding theorems. We begin with the following definitions.

Definition 2.6. Let $X$ and $Y$ be Banach spaces. Then $X$ is continuously embedded in $Y$ if

(i) $X \subseteq Y$,

(ii) there is a positive constant $C$ such that $\|x\|_Y \leq C \|x\|_X$ for all $x \in X$.

Definition 2.7. Let $X$ and $Y$ be Banach spaces. Then $X$ is said to be compactly embedded in $Y$ if

(i) $X$ is continuously embedded in $Y$

(ii) every bounded sequence in $X$ has a convergent subsequence in $Y$.

Theorem 2.5. (Sobolev Embedding Theorem, cf. [1]) Suppose $\Omega$ is a Lipschitz domain in $\mathbb{R}^n$.

(i) If $m > n/p$, then for $1 \leq p < \infty$,

$$W^{m,p}(\Omega) \subseteq C(\bar{\Omega}).$$

(2.4)

In particular,

$$W^{m,p}(\Omega) \subseteq L^q(\Omega) \quad \text{for} \quad p \leq q \leq \infty.$$  

(2.5)

(ii) If $m > k$ and $p < n$ such that $1/p - m/n = 1/q - k/n$, then

$$W^{m,p}(\Omega) \subseteq W^{k,q}(\Omega).$$  

(2.6)

Since the boundary $\partial \Omega$ has measure zero with respect to Lebesgue measure, a function in $W^{m,p}(\Omega)$ is equivalent to functions in $W^{m,p}(\Omega)$ which have arbitrary values on $\partial \Omega$. 

12
To clarify the boundary value of the function, we introduce the trace theorem.

**Theorem 2.6.** (Trace Theorem, cf. [44]) Let $\Omega$ be a bounded Lipschitz domain and $1 \leq p \leq \infty$. Then there exists a linear and bounded mapping $\gamma : W^{1,p}(\Omega) \to L^p(\partial \Omega)$ such that

(i) $\gamma f = f|_{\partial \Omega}$ if $f \in W^{1,p}(\Omega) \cap C(\bar{\Omega})$,

(ii) $\|\gamma f\|_{L^p(\partial \Omega)} \leq C\|f\|_{W^{1,p}(\Omega)}$ for $f \in W^{1,p}(\Omega),$

with $C$ depending only on $p$ and $\Omega$. The mapping $\gamma$ is called the trace operator.

We can define the space $H^1_0(\Omega)$ using the trace operator:

$$H^1_0(\Omega) = \{ v \in H^1(\Omega) : \gamma v = 0 \text{ on } L^2(\partial \Omega) \}.$$  

More results of trace theorem for polygonal domains can be found in [1].

We recall Friedrichs’ inequality for a zero boundary function.

**Theorem 2.7.** (Friedrichs’ Inequality, cf. [19]) Suppose $\Omega$ is a bounded subset of $\mathbb{R}^n$. For $u \in W^{m,p}(\Omega)$ with $u = 0$ on $\partial \Omega$, we have

$$\|u\|_{L^p(\Omega)} \leq (\text{diam } \Omega)^m |u|_{W^{m,p}(\Omega)}.$$  

(2.7)

**Theorem 2.8.** (Bramble-Hilbert Lemma, cf. [19]) Suppose $\Omega$ is a bounded domain in $\mathbb{R}^n$. Then for all $u \in W^{k,p}(\Omega)$, there exists a polynomial $v$ of order up to $k - 1$ such that for $0 \leq m \leq k$,

$$|u - v|_{W^{m,p}(\Omega)} \leq C(\text{diam } \Omega)^{k-m}|u|_{W^{k,p}(\Omega)},$$  

(2.8)

where $C$ depends on $k$ and $\Omega$.

**2.2. Elliptic Equalities**

Let $V$ be a Hilbert space with the inner product $(\cdot, \cdot)$. We define the bilinear form

$a(\cdot, \cdot) : V \times V \to \mathbb{R}$, and the linear and continuous functional $F : V \to \mathbb{R}$. Then the norm
associated with \( V \) is denoted by

\[
\|v\|_V = \sqrt{\langle v, v \rangle} \quad \forall v \in V.
\] (2.9)

The variational problem is written as

\[
a(u, v) = F(v) \quad \forall v \in V.
\] (2.10)

The space of all linear continuous functionals on \( V \) is denoted by \( V^* \), and hence, \( F \in V^* \).

We often use the notation \( F(v) = \langle F, v \rangle \).

**Theorem 2.9.** (Riesz Representation Theorem, cf. [44]) There exists a unique \( x \in V \) such that

\[
\langle F, w \rangle = (x, w) \quad \forall w \in V.
\] (2.11)

**Definition 2.8.** We say a bilinear form is continuous if there exists a positive constant \( C_\dagger \) such that

\[
|a(u, v)| \leq C_\dagger \|u\|_V \|v\|_V.
\] (2.12)

**Definition 2.9.** We say a bilinear form is coercive if there exists a positive constant \( C_\dagger \) such that

\[
a(u, u) \geq C_\dagger \|u\|_V^2.
\] (2.13)

**Lemma 2.10.** (Lax-Milgram, cf. [39]) Let \( V \) be a real Hilbert space, and \( a : V \times V \to \mathbb{R} \) a bilinear form. Suppose \( a(\cdot, \cdot) \) is continuous and coercive. Then for every \( F \in V^* \), the equation (2.10) has a unique solution \( u \in V \). Furthermore, there is some constant \( C > 0 \), not depending on \( F \), such that

\[
\|u\|_V \leq C\|F\|_{V^*}.
\] (2.14)
**Theorem 2.11.** (Projection Theorem, cf. [66]) Let $K$ be a nonempty closed convex subset of a Hilbert space $V$. Given any $u \in V$, there exists a unique $v^* \in K$ such that

$$
\|u - v^*\|_V = \inf_{v \in K} \|u - v\|_V. \tag{2.15}
$$

**Definition 2.10.** Let $K$ be a nonempty closed convex subset of a Hilbert space $V$ with the inner product $(\cdot, \cdot)$. Given $u \in V$, the element of $K$ closest to $u$ is denoted by $P_Ku$ and it is called the projection of $u$.

**Example 2.2.1.** Consider the following elliptic boundary value problem:

$$
\Delta u + u = f \quad \text{in } \Omega \\
u = 0 \quad \text{on } \partial \Omega.
$$

By integration by parts, we obtain

$$
\int_{\Omega} \nabla u \cdot \nabla v + uv \, dx = \int_{\Omega} f v \, dx \quad \forall v \in H^1_0(\Omega).
$$

Define the bilinear form $a(\cdot, \cdot)$ by

$$
a(v, w) = \int_{\Omega} \nabla v \cdot \nabla w + vw \, dx,
$$

and the linear functional $F$ by

$$
F(v) = \int_{\Omega} f v \, dx.
$$

Then we have the equivalent variational problem

$$
a(u, v) = F(v) \quad \forall v \in H^1_0(\Omega). \tag{2.16}
$$

It follows from the Lax-Milgram Lemma (Lemma 2.10) that (2.16) has a unique solution $u \in H^1_0(\Omega)$. 

15
2.3. Elliptic Variational Inequalities

We introduce an abstract elliptic variational inequality. Let $V$ be a Hilbert space with the inner product $(\cdot, \cdot)$, $a(\cdot, \cdot)$ be a symmetric, continuous and coercive bilinear form on $V$, $K \subset V$ be a nonempty closed convex subset, and $F : V \to \mathbb{R}$ be a bounded linear functional.

**Lemma 2.12.** Consider the following minimization problem

$$
\min_{v \in K} \left[ \frac{1}{2}a(v, v) - F(v) \right].
$$

(2.17)

If $u$ is a unique solution of (2.17), then $u$ is also the solution of the variational inequality

$$
a(u, v - u) - F(v - u) \geq 0 \quad \forall v \in K.
$$

(2.18)

Conversely, if $u$ is a unique solution of (2.18), then $u$ is also the unique solution of (2.17).

**Proof.** Suppose

$$
u = \arg\min_{v \in K} \left[ \frac{1}{2}a(v, v) - F(v) \right].
$$

Since $K$ is convex, for $t \in [0, 1]$, we have

$$
u + t(v - u) \in K \quad \forall v \in K.
$$

Define a function

$$
\phi(t) = \left[ \frac{1}{2}a(u + t(v - u), u + t(v - u)) - F(u + t(v - u)) \right].
$$

Then $\phi(t)$ attains it minimum at $t = 0$, and consequently

$$
0 \leq \phi'(t)\bigg|_{t=0} = \frac{d}{dt} \left[ \frac{1}{2}a(u + t(v - u), u + t(v - u)) - F(u + t(v - u)) \right]_{t=0}
$$

$$
= a(u, v - u) - F(v - u).
$$
Therefore, \( u \) is the solution of the variational inequality (2.18).

Now we assume that \( u \) solves the variational inequality (2.18). Note that

\[
\phi'(t) = ta(v - u, v - u) + a(u, v - u) - F(v - u).
\]

Since \( a(\cdot, \cdot) \) is coercive, for \( t \in [0, 1] \), we have

\[
\phi'(t) \geq 0 \quad \forall v \in K.
\]

Therefore \( \phi(1) \geq \phi(0) \), and it implies that

\[
u = u + t(v - u)\bigg|_{t=0}
\]

is the minimizer of the problem (2.17).

First we take a look at the special case where \( a(\cdot, \cdot) = (\cdot, \cdot) \). Then the variational problem becomes the following problem: Find \( u \in K \) such that

\[
(u, v - u) \geq \langle F, v - u \rangle \quad \forall v \in K.
\] \hspace{1cm} (2.19)

By the Riesz Representation Theorem, there exists a unique \( x \in V \) such that

\[
\langle F, w \rangle = (x, w) \quad \forall w \in V.
\] \hspace{1cm} (2.20)

Then we want to find \( u \in K \) such that

\[
(u, v - u) \geq (x, v - u) \quad \forall v \in K.
\] \hspace{1cm} (2.21)

**Lemma 2.13.** \( u = \arg\min_{v \in K} \left[ \frac{1}{2}(v, v) - (x, v) \right] \).

**Proof.** According to Lemma 2.12, the result follows.
Then
\[
\frac{1}{2}(v,v) - (x,v) = \frac{1}{2}(v,v) - (x,v) + \frac{1}{2}(x,x) - \frac{1}{2}(x,x)
\]
\[
= \frac{1}{2}(x - v, x - v) - \frac{1}{2}(x,x).
\]
Thus
\[
u = \arg\min_{v \in K} \|x - v\|^2_V = \arg\min_{v \in K} \|x - v\|_V.
\]
Therefore, \(u\) is the member of \(K\) closer to \(x\) than any other \(v \in K\).

The existence and uniqueness of the solution to the problem (2.21) follow from Theorem 2.11.

We consider the problem: Find \(u \in K\) such that
\[
a(u, v - u) \geq \langle F, v - u \rangle \quad \forall v \in K.
\]

Theorem 2.14. Let \(V\) be a Hilbert space with the inner product \((\cdot, \cdot)\), \(a(\cdot, \cdot)\) be a symmetric, continuous and coercive bilinear form on \(V\), \(K \subset V\) be a nonempty closed convex subset, and \(F : V \rightarrow \mathbb{R}\) be a bounded linear functional. The abstract variational inequality (2.22) has a unique solution.

Proof. Since there are positive constants \(C\) and \(\alpha\) such that
\[
\|v\|_a = \sqrt{a(v,v)} \leq \sqrt{C}\|v\|^2_V = \sqrt{C}\|v\|_V,
\]
\[
\|v\|_a = \sqrt{a(v,v)} \geq \sqrt{\alpha}\|v\|^2_V = \sqrt{\alpha}\|v\|_V,
\]
it implies that \(\|\cdot\|_a\) is equivalent to \(\|\cdot\|_V\). Then we can view the space \(V\) as a Hilbert space associated with the inner product \(a(\cdot, \cdot)\). According to the argument discussed above, we obtain the existence and uniqueness of the solution. \(\square\)
Remark 2.15. (cf. [56]) Consider the general case where the bilinear form $a(\cdot, \cdot)$ is not necessarily symmetric. We assume that $a(\cdot, \cdot)$ is coercive on the set $K - K$, where

$$K - K = \{v - w : v, w \in K\}.$$ 

Then given a linear functional $F \in V^*$, variational inequality (2.22) has a unique solution. The proof of uniqueness only uses the coercivity of the bilinear form $a(\cdot, \cdot)$.

Example 2.3.1. Let $f \in L^2(\Omega)$ and $\psi \in H^2(\Omega)$. Consider the following obstacle problem:

$$\min_{v \in K} \left[ \frac{1}{2} \int_\Omega \nabla v \cdot \nabla v \, dx - \int_\Omega fv \, dx \right],$$

(2.23)

where $K = \{v \in H^1_0(\Omega) : v \geq \psi \text{ a.e. in } \Omega\}$ with $\psi < 0$ on $\partial \Omega$. By Lemma 2.12, there exists a unique solution $u \in K$ of (2.23) which is also the solution of the following variational inequality:

$$\int_\Omega \nabla u \cdot \nabla (v - u) \, dx - \int_\Omega f(v - u) \, dx \geq 0 \quad \forall v \in K.$$

(2.24)

2.4. Regularity

In this section, we present the regularity results for the Poisson equation and the biharmonic problem.

2.4.1. Poisson Equation

We discuss the regularity results for the Poisson equation.

Theorem 2.16. (Poincaré-Friedrichs Inequality, cf. [39]) Let $\Omega$ be a connected bounded domain. There exists a positive constant $C$ such that

$$\|v\|_{L^2(\Omega)} \leq C \|
abla v\|_{L^2(\Omega)}$$

(2.25)

for all $v \in H^1_0(\Omega)$. 

19
We first check the existence and uniqueness of the solution for Poisson’s equation:

\[-\Delta y = f \quad \text{in } \Omega \tag{2.26}\]
\[y = 0 \quad \text{on } \partial \Omega, \tag{2.27}\]

where \( f \in L_2(\Omega) \). Multiplying Poisson’s equation by a test function \( v \in C_0^\infty(\Omega) \) and integrating over \( \Omega \), we have

\[- \int_\Omega v \Delta y \, dx = \int_\Omega f v \, dx.\]

By integration by parts, it follows that

\[\int_\Omega \nabla y \cdot \nabla v \, dx = \int_\Omega f v \, dx.\]

Since \( C_0^\infty(\Omega) \) is dense in \( H_0^1(\Omega) \), if \( y \) satisfies the weak formulation

\[\int_\Omega \nabla y \cdot \nabla v \, dx = \int_\Omega f v \, dx \quad \forall \ v \in H_0^1(\Omega), \tag{2.28}\]

then \( y \in H_0^1(\Omega) \) is a weak solution to the boundary value problem (2.26)-(2.27).

**Theorem 2.17.** (cf. [44]) If \( \Omega \) is Lipschitz continuous and bounded, then for every \( f \in L_2(\Omega) \), the problem (2.26)-(2.27) has a unique weak solution \( y \in H_0^1(\Omega) \). Moreover, there exists a constant \( C > 0 \), not depending on \( f \), such that

\[\|y\|_{H^1(\Omega)} \leq C\|f\|_{L_2(\Omega)}.\tag{2.29}\]

**Theorem 2.18.** (cf. [62]) Let \( \Omega \) be a bounded polygonal domain. Then there exists a constant \( C \) and \( \alpha \in (0,1] \) such that

\[\|y\|_{H^{1+\alpha}(\Omega)} \leq C\|f\|_{H^{\alpha-1}(\Omega)},\tag{2.30}\]

where \( C \) is depending on \( \Omega \) and \( \alpha \) is depending on the interior angles at the corners of \( \Omega \).

If the domain \( \Omega \) is convex, then we can take \( \alpha = 1 \).

For general smooth domains, when \( f \in L_2(\Omega) \), the solution \( y \in H^2(\Omega) \).
2.4.2. Biharmonic Equation

We discuss the regularity results for the biharmonic problem.

Let \( \Omega \subseteq \mathbb{R}^2 \) be a bounded domain. We now consider the biharmonic boundary problem:

\[
\Delta^2 u = f \quad \text{in } \Omega \quad (2.31)
\]
\[
u = \Delta u = 0 \quad \text{on } \partial \Omega, \quad (2.32)
\]

where \( f \in L_2(\Omega) \). Define the bilinear form

\[
a(v, w) = \int_\Omega D^2 v : D^2 w \, dx, \quad (2.33)
\]

where \( D^2 v \) is the Hessian matrix of \( v \) and \( D^2 v : D^2 w \) is the Frobenius inner product of \( D^2 v \) and \( D^2 w \). By using the Poincaré-Friedrichs inequality, the bilinear form \( a(\cdot, \cdot) \) is continuous and coercive on \( H^2(\Omega) \cap H^1_0(\Omega) \). It follows from the Lax-Milgram Lemma (cf. Lemma 2.10) that there exists a unique solution \( u \) of (2.31)-(2.32).

We present the regularity results depending on \( \Omega \).

**Theorem 2.19.** (cf. [47]) Let \( \Omega \) be a smooth domain and \( u \) be the solution of (2.31)-(2.32). Then for \( f \in L_2(\Omega) \), we have \( u \in H^4(\Omega) \) and

\[
\|u\|_{H^4(\Omega)} \leq C \|f\|_{L_2(\Omega)},
\]

where \( C \) is dependent of \( \Omega \).

**Theorem 2.20.** (cf. [62]) Let \( \Omega \) be a polygonal domain and \( u \) be the solution of (2.31)-(2.32). Then for \( f \in L_2(\Omega) \), we have \( u \in H^{2+\alpha}(\Omega) \) for some \( \alpha \in (0, 2] \), where \( \alpha \) depends on the interior angles at the corners of \( \Omega \). Moreover,

\[
\|u\|_{H^{2+\alpha}(\Omega)} \leq C \|f\|_{L_2(\Omega)},
\]
where $C$ is dependent of $\Omega$.

Remark 2.21. In Theorem 2.20, $\alpha$ can be close to 0 even when the domain is convex. If the largest interior angle of $\Omega$ is less than or equal to $\pi/2$, then $\alpha = 1$.

More details on the elliptic regularity results for second order problems and biharmonic problems can be found in [40, 47, 62].
Chapter 3. Elliptic Optimal Control Problems

3.1. Elliptic Optimal Control Problems with Pointwise Constraints

Let us consider a body to be heated or cooled in $\Omega \in \mathbb{R}^2$. The control $u$ acts as a heat source in $\Omega$, which is constant in time but depends on the location $x$, i.e. $u = u(x)$. Our goal is to choose the control so that the corresponding temperature distribution $y = y(x)$ in $\Omega$ (the state) is the best possible approximation to a desired state $y_d = y_d(x)$ in $\Omega$.

![Figure 3.1. Distributed control in $\Omega$](image)

The model problem is an elliptic optimal control problem:

$$
\min J(y, u) = \frac{1}{2} \| y(x) - y_d(x) \|^2_{L_2(\Omega)} + \frac{\beta}{2} \| u(x) \|^2_{L_2(\Omega)}, \tag{3.1}
$$

subject to

$$
-\Delta y = \lambda u \quad \text{in } \Omega \tag{3.2}
$$

$$
y = 0 \quad \text{on } \partial \Omega \tag{3.3}
$$

and

$$
u_-(x) \leq u(x) \leq u_+ (x) \quad \text{in } \quad \Omega. \tag{3.4}
$$

Here, a regularization parameter $\beta$ is a positive constant, and $\lambda = \lambda(x)$ is prescribed. Assume that

$$
u_-(x) < u_+ (x) \quad \text{on } \quad \Omega.
$$
We define the admissible set by
\[ U_{ad} = \{ u \in L^2(\Omega) : u_-(x) \leq u(x) \leq u_+(x) \text{ for almost every } x \in \Omega \}. \]

Since \( u_-(x) < u_+(x) \) on \( \Omega \), the admissible set \( U_{ad} \) is nonempty, closed and convex. It follows from Theorem 2.17, for every \( u \in U_{ad} \), there exists a unique weak solution \( y \in H^1_0(\Omega) \) to (3.2)-(3.3). In view of the estimate

\[ \| y \|_{L^2(\Omega)} \leq \| y \|_{H^1(\Omega)}, \]

the space \( H^1_0(\Omega) \) is continuously embedded in \( L^2(\Omega) \). So we have

\[ S : L^2(\Omega) \to L^2(\Omega), \quad u \mapsto y(u), \]

since \( y \) is dependent on \( u \). Note that \( S \) is a continuous linear operator. The optimal control problem (3.1)-(3.4) can be reduced as

\[ \min_{u \in U_{ad}} \tilde{J}(u) := \left[ \frac{1}{2} \| Su - y_d \|_{L^2(\Omega)}^2 + \frac{\beta}{2} \| u \|_{L^2(\Omega)}^2 \right]. \quad (3.5) \]

**Theorem 3.1.** The reduced optimal control problem (3.5) has a unique solution characterized by the variational inequality

\[ (S^*(S\bar{u} - y_d) + \beta \bar{u}, u - \bar{u})_{L^2(\Omega)} \geq 0 \quad \forall \ u \in U_{ad}, \quad (3.6) \]

where \( S^* \) is the adjoint operator of \( S \).

**Proof.** Notice that

\[
\min_{u \in U_{ad}} \left[ \frac{1}{2} \| Su - y_d \|_{L^2(\Omega)}^2 + \frac{\beta}{2} \| u \|_{L^2(\Omega)}^2 \right] \\
= \min_{u \in U_{ad}} \left[ \frac{1}{2} \| Su \|_{L^2(\Omega)}^2 + \frac{1}{2} \| y_d \|_{L^2(\Omega)}^2 \right] - (Su, y_d)_{L^2(\Omega)} + \frac{\beta}{2} \| u \|_{L^2(\Omega)}^2 \\
= \min_{u \in U_{ad}} \left[ \frac{1}{2} \| Su \|_{L^2(\Omega)}^2 + \beta \| u \|_{L^2(\Omega)}^2 \right] - (Su, y_d)_{L^2(\Omega)}. 
\]
Define the bilinear form $a(\cdot, \cdot)$ by

$$a(u, v) = (Su, Sv)_{L_2(\Omega)} + \beta(u, v)_{L_2(\Omega)}, \quad (3.7)$$

and the linear functional $F$ by

$$F(v) = (y_d, Sv)_{L_2(\Omega)}. \quad (3.8)$$

We can rewrite the reduced problem (3.5) as to find $\bar{u}$ such that

$$\bar{u} = \arg\min_{u \in U_{ad}} \left[ \frac{1}{2} a(u, u) - F(u) \right]. \quad (3.9)$$

By the Cauchy-Schwarz inequality and the boundedness of $S$,

$$|a(u, v)| = |(Su, Sv)_{L_2(\Omega)} + \beta(u, v)_{L_2(\Omega)}|$$

$$\leq ||Su||_{L_2(\Omega)}||Sv||_{L_2(\Omega)} + \beta||u||_{L_2(\Omega)}||v||_{L_2(\Omega)}$$

$$\leq C ||u||_{L_2(\Omega)}||v||_{L_2(\Omega)}.$$

Also we obtain

$$a(u, u) = ||Su||^2_{L_2(\Omega)} + ||u||^2_{L_2(\Omega)} \geq ||u||^2_{L_2(\Omega)}.$$ 

Therefore, the bilinear form $a(\cdot, \cdot)$ is symmetric, bounded and coercive. By the Cauchy-Schwarz inequality

$$|F(u)| = |(y_d, Su)_{L_2(\Omega)}| \leq ||Su||_{L_2(\Omega)}||y_d||_{L_2(\Omega)} < \infty,$$

and hence $F$ is bounded. Note that $U_{ad}$ is nonempty, closed and convex. By Lemma 2.12 and Theorem 2.14, the minimization problem (3.5) has a unique solution $\bar{u}$ characterized by the following variational inequality:

$$a(\bar{u}, u - \bar{u}) \geq F(u - \bar{u}) \quad \forall u \in U_{ad}. \quad (3.10)$$
Using (3.7) and (3.8), we have

\[
0 \leq (S\bar{u}, Su - S\bar{u})_{L^2(\Omega)} + \beta(\bar{u}, u - \bar{u})_{L^2(\Omega)} - (y_d, Su - S\bar{u})_{L^2(\Omega)} \\
= (S^*S\bar{u}, u - \bar{u})_{L^2(\Omega)} + \beta(\bar{u}, u - \bar{u})_{L^2(\Omega)} - (S^*y_d, u - \bar{u})_{L^2(\Omega)} \\
= (S^*(S\bar{u} - y_d) + \beta\bar{u}, u - \bar{u})_{L^2(\Omega)} \quad \forall u \in U_{ad}.
\]

\[\square\]

### 3.2. Elliptic Optimal Control Problem with General Tracking and Pointwise State Constraints

In this section, we introduce the elliptic optimal control problem with general tracking and pointwise state constraints. Let \( \Omega \subset \mathbb{R}^2 \) be a bounded convex polygon, \( \beta \) be a positive constant, \( \mathcal{P} = \{p_1, \ldots, p_J\} \) be a finite set of points in \( \Omega \), and \( \mathcal{C} = \bigcup_{l=1}^L \mathcal{C}_l \subset \Omega \) be the union of the curves \( \mathcal{C}_1, \ldots, \mathcal{C}_L \), where each curve is parametrized by a Lipschitz continuous function defined on \([0, 1]\). The weight functions \( w_0, w_1 \) and \( w_2 \) are bounded nonnegative Borel measurable functions defined on \( \mathcal{P}, \mathcal{C} \) and \( \Omega \) respectively. The desired states \( y_0, y_1 \) and \( y_2 \) are Borel measurable functions defined on \( \mathcal{P}, \mathcal{C} \) and \( \Omega \) such that

\[
\sum_{j=1}^J y_0(p_j)^2 w_0(p_j) + \sum_{l=1}^L \int_{\mathcal{C}_l} y_1^2 w_1 \, ds + \int_{\Omega} y_2^2 w_2 \, dx < \infty. \tag{3.10}
\]

The optimal control problem is to find

\[
(\bar{y}, \bar{u}) = \text{argmin}_{(y, u) \in K} \frac{1}{2} \left[ G(y) + \beta \|u\|_{L^2(\Omega)}^2 \right], \tag{3.11}
\]

where

\[
G(y) = \sum_{j=1}^J (y(p_j) - y_0(p_j))^2 w_0(p_j) + \sum_{l=1}^L \int_{\mathcal{C}_l} (y - y_1)^2 w_1 \, ds + \int_{\Omega} (y - y_2)^2 w_2 \, dx, \tag{3.12}
\]

26
$\mathbb{K} \subset H^1_0(\Omega) \times L^2(\Omega)$, and $(y,u) \in H^1_0(\Omega) \times L^2(\Omega)$ belongs to $\mathbb{K}$ if and only if
\[
\int_{\Omega} \nabla y \cdot \nabla z \, dx = \int_{\Omega} uz \, dx \quad \forall z \in H^1_0(\Omega),
\]
\[
\psi^- \leq y \leq \psi^+ \quad \text{a.e. in } \Omega.
\]

We assume that the functions $\psi_{\pm}$ satisfy
\[
\psi_{\pm} \in W^{3,q}(\Omega) \quad \text{for } q > 2,
\]
\[
\psi^- < \psi^+ \quad \text{on } \bar{\Omega},
\]
\[
\psi^- < 0 < \psi^+ \quad \text{on } \partial\Omega.
\]

**Remark 3.2.** Since $\Omega$ is convex, the partial differential equation constraint (3.13) implies through elliptic regularity (cf. [47, 40, 62]) that $y \in H^2(\Omega)$ if $(y,u) \in \mathbb{K}$ (cf. Theorem 2.18). Therefore the function $G$ is well-defined by the Sobolev embedding theorem $H^2(\Omega) \subset C(\bar{\Omega})$ (cf. Theorem 2.5).

**Remark 3.3.** In the case where $w_0 = w_1 = 0$ and $w_2 = 1$, the function $G$ becomes to a standard tracking function which is the same cost function as (3.1) (cf. [52, 68]), and the optimal control problem with pointwise state constraints was introduced in [33]. In the case where $w_0 \neq 0$ and $w_1 = w_2 = 0$, the problem has a point-tracking cost functional (cf. [30, 36, 29, 6, 4, 5, 3]).

**Remark 3.4.** As we already discussed in Section 3.1, the optimal control problem defined by (3.11)-(3.14) can be interpreted as a heat conduction problem (cf. [68]). However, in this case we have different desired temperature $y_0$ (resp., $y_1$ and $y_2$) at the points (resp., curves and regions). Moreover, unlike (3.4), our problem has pointwise state constraints $\psi_{\pm}$ in $\Omega$. The weights $w_0$, $w_1$ and $w_2$ allow preferences for the desired temperature.
**Remark 3.5.** We introduce the Radon measure $\nu$ on $\bar{\Omega}$ defined by

$$
\int_{\Omega} f \, d\nu = \sum_{j=1}^{J} f(p_j) w_0(p_j) + \sum_{l=1}^{L} \int_{\mathcal{E}_l} f w_1 \, ds + \int_{\Omega} f w_2 \, dx.
$$

(3.18)

Then we have

$$
G(y) = \int_{\Omega} (y - y_d)^2 \, d\nu = \|y - y_d\|_{L^2(\Omega;\nu)}^2,
$$

(3.19)

where

$$
y_d = \begin{cases}
y_0 & \text{on } \mathcal{P} \\
y_1 & \text{on } \mathcal{C} \setminus \mathcal{P} \\
y_2 & \text{on } \Omega \setminus (\mathcal{C} \cup \mathcal{P})
\end{cases}
$$

(3.20)

and the condition (3.10) becomes $\|y_d\|_{L^2(\Omega;\nu)}^2 < \infty$.

### 3.2.1. Reduced Problem

As mentioned in Remark 3.2, the constraint (3.13) implies that $y \in H^2(\Omega) \cap H^1_0(\Omega) \subset C(\bar{\Omega})$. From (3.13), integration by parts gives us that

$$
\int_{\Omega} uz \, dx = \int_{\Omega} \nabla y \cdot \nabla z \, dx = \int_{\Omega} -\Delta y z \, dx \quad \forall z \in H^1_0(\Omega).
$$

Therefore, in view of (3.19) and the relation $u = -\Delta y$, the optimal control problem defined by (3.11)-(3.14) is equivalent to the following reduced minimization problem with only state variable $y$:

Find $\bar{y} = \arg\min_{y \in K} \frac{1}{2} [\beta \|\Delta y\|^2_{L^2(\Omega)} + \|y - y_d\|^2_{L^2(\Omega;\nu)}],

(3.21)

where

$$
K = \{y \in H^2(\Omega) \cap H^1_0(\Omega) : \psi^- \leq y \leq \psi^+ \text{ in } \Omega\}.
$$

(3.22)

**Remark 3.6.** While the minimization problem (3.5) contains only control variable $u$, the reduced optimal control problem (3.21) consider only state variable $y$, which results in fourth order cost functional.
Theorem 3.7. The reduced minimization problem (3.21)-(3.22) has a unique solution.

Proof. Note that

\[
\min_{y \in K} \frac{1}{2} \left[ \beta \| \Delta y \|^2_{L_2(\Omega)} + \| y - y_d \|^2_{L_2(\Omega;\nu)} \right] = \min_{y \in K} \frac{1}{2} \left[ \beta \| \Delta y \|^2_{L_2(\Omega)} + \| y \|^2_{L_2(\Omega,\nu)} - 2 \int_{\Omega} y \, y_d \, d\nu \right].
\]

Define the bilinear form \( \tilde{a}(\cdot, \cdot) \) as

\[
\tilde{a}(y, z) = \beta \int_{\Omega} \Delta y \Delta z \, dx + \int_{\Omega} y z \, d\nu,
\]

and the linear functional \( \tilde{F} \) as

\[
\tilde{F}(y) = \int_{\Omega} y y_d \, d\nu.
\]

Then the reduced problem (3.21) can be written as

\[
\text{Find } \bar{y} = \arg \min_{y \in K} \left[ \frac{1}{2} \tilde{a}(y, y) - \tilde{F}(y) \right]. \tag{3.23}
\]

By the Cauchy-Schwarz inequality,

\[
|\tilde{a}(y, z)| = \left| \beta \int_{\Omega} \Delta y \Delta z \, dx + \int_{\Omega} y z \, d\nu \right| \leq C \left( |y|_{H^2(\Omega)} |z|_{H^2(\Omega)} + \| y \|_{L_2(\Omega;\nu)} \| z \|_{L_2(\Omega;\nu)} \right), \tag{3.24}
\]

for any \( y, z \in H^2(\Omega) \cap H^1_0(\Omega) \). According to the Sobolev embedding theorem (cf. Theorem 2.5), the second term of the right hand side of (3.24) is bounded as

\[
\| y \|_{L_2(\Omega;\nu)} \| z \|_{L_2(\Omega;\nu)} \leq C \| y \|_{H^2(\Omega)} \| z \|_{H^2(\Omega)}, \tag{3.25}
\]

and thus

\[
|\tilde{a}(y, z)| \leq C \| y \|_{H^2(\Omega) \cap H^1_0(\Omega)} \| z \|_{H^2(\Omega) \cap H^1_0(\Omega)}.
\]
for any \( y, z \in H^2(\Omega) \cap H^1_0(\Omega) \).

It also follows from (3.18) and the trace theorem that

\[
\tilde{a}(y, y) = \beta \int_\Omega \Delta y^2 \, dx + \int_\Omega y^2 \, d\nu \\
\geq C\left( |y|_{H^2(\Omega)}^2 + \|y\|_{L^2(\Omega; \nu)}^2 \right) \\
\geq C\left( |y|_{H^2(\Omega)}^2 + \|y\|_{L^2(\Omega)}^2 + |y|_{H^1(\Omega)}^2 \right) \\
\geq C\|y\|_{H^2(\Omega)}^2 \cap H^1_0(\Omega), \quad \forall \, y \in H^2(\Omega) \cap H^1_0(\Omega).
\]

By the Cauchy-Schwarz inequality and the Sobolev embedding theorem (cf. Theorem 2.5), we have

\[
|\tilde{F}(y)| = \left| \int_\Omega y y_d \, d\nu \right| \leq C\|y\|_{L^2(\Omega; \nu)}\|y_d\|_{L^2(\Omega; \nu)} < \infty,
\]

for any \( y \in H^2(\Omega) \cap H^1_0(\Omega) \).

Note that \( K \) is nonempty, closed and convex. Since the bilinear form \( \tilde{a}(\cdot, \cdot) \) is continuous and coercive, and the linear functional \( \tilde{F} \) is bounded, by Lemma 2.12 and Theorem 2.14, the minimization problem (3.23) has a unique solution. \( \square \)

3.2.2. Variational Inequality and KKT Conditions

Note that \( K \) is a nonempty closed convex subset of \( H^2(\Omega) \cap H^1_0(\Omega) \). We define

\[
E(y) = \frac{1}{2} \left[ \beta\|\Delta y\|_{L^2(\Omega)}^2 + \|y - y_d\|_{L^2(\Omega; \nu)}^2 \right].
\]

Let \( y \in K \) be arbitrary. The function \( \phi(t) = E((1 - t)\bar{y} + ty) \) is defined on \( \mathbb{R} \). Since \( \bar{y} \) is the minimizer of \( E(y) \), it follows that

\[
\phi(0) = E(\bar{y}) \leq E((1 - t)\bar{y} + ty) = \phi(t) \quad \text{for} \quad 0 \leq t \leq 1.
\]
Therefore, we have \( \phi'(0) \geq 0 \). Note that

\[
E((1 - t)\bar{y} + ty) = E(\bar{y} + t(y - \bar{y}))
\]

\[
= \frac{1}{2} \left[ \beta \int_{\Omega} (\Delta(\bar{y} + t(y - \bar{y})))^2 \, dx + \int_{\Omega} (\bar{y} + t(y - \bar{y}) - y_d)^2 \, d\nu \right]
\]

\[
= \frac{1}{2} \left[ \beta \int_{\Omega} (\Delta \bar{y} + t\Delta(y - \bar{y}))^2 \, dx + \int_{\Omega} (\bar{y} - y_d + t(y - \bar{y}))^2 \, d\nu \right].
\]

Then

\[
0 \leq \phi'(0) = \left. \frac{d}{dt} E((1 - t)\bar{y} + ty) \right|_{t=0}
\]

\[
= \left[ \beta \int_{\Omega} (\Delta \bar{y} + t\Delta(y - \bar{y})) \Delta(y - \bar{y}) \, dx + \int_{\Omega} (\bar{y} - y_d + t(y - \bar{y}))(y - \bar{y}) \, d\nu \right]_{t=0}
\]

\[
= \beta \int_{\Omega} (\Delta \bar{y})(\Delta y - \Delta \bar{y}) \, dx + \int_{\Omega} (\bar{y} - y_d)(y - \bar{y}) \, d\nu.
\]

It follows from the standard theory of calculus of variations (cf. [56, 42]) that the minimization problem defined by (3.21)-(3.22) has a unique solution \( \bar{y} \in K \) characterized by the fourth order variational inequality

\[
\beta \int_{\Omega} (\Delta \bar{y})(\Delta y - \Delta \bar{y}) \, dx + \int_{\Omega} (\bar{y} - y_d)(y - \bar{y}) \, d\nu \geq 0 \quad \forall \, y \in K.
\] (3.26)

**Definition 3.1.** The active sets \( \mathcal{A}_\pm \) are defined by

\[
\mathcal{A}_\pm = \{ x \in \Omega : \bar{y}(x) = \psi_\pm(x) \}.
\] (3.27)

**Lemma 3.8.** The fourth order variational inequality (3.26) is equivalent to the following generalized Karush-Kuhn-Tucker conditions:

\[
\beta \int_{\Omega} (\Delta \bar{y})(\Delta z) \, dx + \int_{\Omega} (\bar{y} - y_d) z \, d\nu = \int_{\Omega} z \, d\mu,
\] (3.28)
for any \( z \in H^2(\Omega) \cap H^1_0(\Omega) \), where \( \mu \) is a regular Borel measure, such that

\[
\begin{align*}
\mu &\geq 0 \quad \text{if } \bar{y} = \psi_-, \\
\mu &\leq 0 \quad \text{if } \bar{y} = \psi_+, \\
\mu &= 0 \quad \text{otherwise}.
\end{align*}
\] (3.29) (3.30) (3.31)

Proof. Let \( \mathcal{A} = \mathcal{A}_+ \cup \mathcal{A}_- \). For any \( z \in H^2(\Omega) \cap H^1_0(\Omega) \) whose support is disjoint from \( \mathcal{A} \), we have that \( y_\epsilon = \pm \epsilon z + \bar{y} \) belongs to \( K \) if \( \epsilon \) is sufficiently small. Then (3.26) implies

\[
\beta \int_\Omega (\Delta \bar{y})(\Delta z) \, dx + \int_\Omega (\bar{y} - y_d)z \, d\nu = 0.
\] (3.32)

Let \( \phi_1, \phi_2, \phi_3 \in C^\infty(\bar{\Omega}) \) be a partition of unity on \( \bar{\Omega} \) with the following properties:

(i) \( \phi_1, \phi_2, \phi_3 \geq 0 \)

(ii) \( \phi_1 + \phi_2 + \phi_3 = 1 \)

(iii) \( \phi_1 = 1 \) on an open neighborhood of \( \mathcal{A}_- \).

(iv) \( \phi_2 = 1 \) on an open neighborhood of \( \mathcal{A}_+ \).

For any nonnegative \( z \in H^2(\Omega) \cap H^1_0(\Omega) \), we have \( \phi_1 z + \bar{y} = y \in K \). By (3.26),

\[
\beta \int_\Omega (\Delta \bar{y})(\Delta (\phi_1 z)) \, dx + \int_\Omega (\bar{y} - y_d)(\phi_1 z) \, d\nu \geq 0.
\]

By the Riez Representation Theorem for nonnegative functionals (cf. [66]) that there exists a unique nonnegative Borel measure \( \mu_- \) such that

\[
\beta \int_\Omega (\Delta \bar{y})(\Delta (\phi_1 z)) \, dx + \int_\Omega (\bar{y} - y_d)(\phi_1 z) \, d\nu = \int_\Omega z \, d\mu_-.
\] (3.33)

Now, for any nonnegative \( z \in H^2(\Omega) \cap H^1_0(\Omega) \), we have \( -\phi_2 z + \bar{y} = y \in K \). By (3.26),

\[
\beta \int_\Omega (\Delta \bar{y})(\Delta (\phi_2 z)) \, dx + \int_\Omega (\bar{y} - y_d)(\phi_2 z) \, d\nu \leq 0.
\]
By the Riesz Representation Theorem for nonpositive functionals (cf. [66]) that there exists a unique nonpositive Borel measure $\mu_+$ such that
\[
\beta \int_{\Omega} (\Delta \bar{y})(\Delta (\phi_2 z)) \, dx + \int_{\Omega} (\bar{y} - y_d)(\phi_2 z) \, d\nu = \int_{\Omega} z \, d\mu_+.
\] (3.34)

From (3.32), (3.33), (3.34), and the properties of $\phi_1, \phi_2, \phi_3$, it follows that
\[
\beta \int_{\Omega} (\Delta \bar{y})(\Delta z) \, dx + \int_{\Omega} (\bar{y} - y_d)z \, d\nu = \beta \int_{\Omega} (\Delta \bar{y})(\Delta (\phi_1 z)) \, dx + \int_{\Omega} (\bar{y} - y_d)(\phi_1 z) \, d\nu
+ \beta \int_{\Omega} (\Delta \bar{y})(\Delta (\phi_2 z)) \, dx + \int_{\Omega} (\bar{y} - y_d)(\phi_2 z) \, d\nu
+ \beta \int_{\Omega} (\Delta \bar{y})(\Delta (\phi_3 z)) \, dx + \int_{\Omega} (\bar{y} - y_d)(\phi_3 z) \, d\nu
= \int_{\Omega} z \, d\mu_- + \int_{\Omega} z \, d\mu_+,
\] (3.35)

since the support of $\phi_3 z$ is disjoint from $\mathcal{A}$.

Let $\mu = \mu_+ + \mu_-$. We can rewrite (3.35) as
\[
\beta \int_{\Omega} (\Delta \bar{y})(\Delta z) \, dx + \int_{\Omega} (\bar{y} - y_d)z \, d\nu = \int_{\Omega} z \, d\mu,
\] (3.36)

for any $z \in H^2(\Omega) \cap H^1_0(\Omega)$. Then we have the followings:

If $\psi_- < \bar{y} < \psi_+$, then $\mu = 0$ by (3.32).

If $\bar{y} = \psi_-$, then $\mu \geq 0$ because $\mu_+ = 0$ and $\mu_- \geq 0$.

If $\bar{y} = \psi_+$, then $\mu \leq 0$ because $\mu_- = 0$ and $\mu_+ \leq 0$.

Therefore $\mu$ satisfies the KKT conditions (3.29)-(3.31).

**Remark 3.9.** It follows from (3.29)-(3.31) that the support of $\mu$ is the union of the active sets $\mathcal{A}_\pm$. Since $\bar{y} = 0$ on $\partial \Omega$ and $\psi_+ > 0$ on $\partial \Omega$ by the assumption (3.17), the active set
\( \mathcal{A}_+ \) is a compact subset of \( \Omega \). Similarly, \( \mathcal{A}_- \) is a compact subset of \( \Omega \). Therefore \( \mu \) is a bounded measure.

### 3.2.3. Regularity

In this section, we discuss the local and global regularity of the optimal state \( \bar{y} \) and the Borel measure \( \mu \). By introducing a new bilinear form \( \mathcal{A}(\cdot, \cdot) \), we can also rewrite (3.28) as

\[
\mathcal{A}(\bar{y}, z) - \int_\Omega y_d z \, d\nu = \int_\Omega z \, d\mu,
\]

where

\[
\mathcal{A}(y, z) = \beta a(y, z) + \int_\Omega y z \, d\nu,
\]

and (cf. [47])

\[
a(y, z) = \int_\Omega (\Delta y)(\Delta z) \, dx = \int_\Omega \nabla^2 y : \nabla^2 z \, dx \quad \forall y, z \in H^2(\Omega) \cap H^1_0(\Omega).
\]

Here \( \nabla^2 y \) is the Hessian matrix of \( y \) and \( \nabla^2 y : \nabla^2 z \) is the Frobenius inner product of \( \nabla^2 y \) and \( \nabla^2 z \).

We define the adjoint state \( \bar{p} \in L_2(\Omega) \) associated with \( \bar{y} \) by

\[
\int_\Omega \bar{p}(-\Delta z) \, dx = \int_\Omega (\bar{y} - y_d) z \, d\nu - \int_\Omega z \, d\mu \quad \forall z \in H^2(\Omega) \cap H^1_0(\Omega).
\]

Then we have (cf. [33, Theorem 1])

\[
\bar{p} \in W_0^{1,s}(\Omega) \quad \forall s < 2.
\]

**Remark 3.10.** Note that \( -\Delta \) maps \( H^2(\Omega) \) into \( L_2(\Omega) \). Since \( \mathcal{C} \cup \mathcal{P} \) and the support of \( \mu \) are compact subsets of \( \Omega \), the adjoint state \( \bar{p} \) belongs to \( H^2 \) in a neighborhood of \( \partial \Omega \) and vanishes on \( \partial \Omega \).
It follows from (3.28) and (3.40) that
\[ \bar{u} = -\Delta \bar{y} = -\left(1/\beta\right) \bar{p} \in W^{1,s}_0(\Omega) \quad \forall \, s < 2, \tag{3.42} \]
and, in view of Remark 3.10, \( \bar{u} \) belongs to \( H^2 \) in a neighborhood of \( \partial \Omega \) and \( \bar{u} = 0 \) on \( \partial \Omega \).

According to (3.42), the solution \( \bar{y} \in K \subset H^2(\Omega) \cap H^1_0(\Omega) \) of (3.21) satisfies
\[ \Delta \bar{y} = \bar{p}/\beta \text{ in } \Omega \quad \text{and} \quad \bar{y} = 0 \text{ on } \partial \Omega. \tag{3.43} \]

It then follows from (3.41), (3.43) and interior elliptic regularity (cf. [2] and [53]) that
\[ \bar{y} \in W^{3,s}_{\text{loc}}(\Omega) \quad \forall \, s < 2. \tag{3.44} \]

Moreover we can conclude from Remark 3.10, (3.43), (3.44) and the elliptic regularity theory for polygonal domains (cf. [47, 40]) that globally
\[ \bar{y} \in H^{2+\alpha}(\Omega) \tag{3.45} \]
for some \( \alpha \in (0,1) \), where the index of elliptic regularity \( \alpha \) is determined by the angles at the corners of \( \Omega \).

First, recall the Hahn decomposition: Let \( \Sigma \) be a \( \sigma \)-algebra on a set \( X \) and \( \lambda \) be a signed measure on \( \Sigma \). Then there exist a positive set \( P \) and negative set \( N \) such that
\[ P \cup N = X \quad \text{and} \quad P \cap N = \emptyset. \]

In view of (3.27) and the assumption (3.16), we have
\[ \mathcal{A}_- \cap \mathcal{A}_+ = \emptyset. \tag{3.46} \]
According to Remark 3.9, the support of $\mu$ is the union of the active sets $\mathcal{A}_\pm$, and hence the Hahn decomposition of $\mu$ is given by

$$\mu = \mu_- + \mu_+ \quad (3.47)$$

where $\mu_-$ is the finite nonnegative Borel measure (cf. (3.30)) defined by

$$\mu_-(B) = \mu(B \cap \mathcal{A}_-) \quad \text{for all Borel subsets } B \text{ of } \Omega, \quad (3.48)$$

and $\mu_+$ is the finite nonpositive Borel measure (cf. (3.29)) defined by

$$\mu_+(B) = \mu(B \cap \mathcal{A}_+) \quad \text{for all Borel subsets } B \text{ of } \Omega. \quad (3.49)$$

Since the support of $\mu_+$ (resp., $\mu_-$) is a subset of the active set $\mathcal{A}_+$ (resp., $\mathcal{A}_-$), we have

$$\int_{\Omega} (\bar{y} - \psi_+) \, d\mu_+ = 0 \quad \text{and} \quad \int_{\Omega} (\bar{y} - \psi_-) \, d\mu_- = 0. \quad (3.50)$$

Let $z \in H^2(\Omega) \cap H^1_0(\Omega)$ be arbitrary. Because of (3.46) we can construct $\phi \in C_c^{\infty}(\Omega)$ such that

- $\phi = 1$ in a neighborhood of $\mathcal{A}_+$ and
- $\phi = 0$ in a neighborhood of $\mathcal{A}_-$.

It follows from (3.18), (3.20), (3.28), (3.44), (3.47) and integration by parts that

$$\int_{\Omega} z \, d\mu_+ = \int_{\Omega} \phi z \, d\mu$$

$$= \beta \int_{\Omega} (\Delta \bar{y}) [\Delta (\phi z)] \, dx + \int_{\Omega} (\bar{y} - y_d)(\phi z) \, d\nu$$

$$= \beta \int_{\Omega} \nabla (\Delta \bar{y}) \cdot \nabla (\phi z) \, dx + \sum_{j=1}^L w_0(p_j) [\bar{y}(p_j) - y_0(p_j)] \phi(p_j) z(p_j)$$

$$+ \sum_{l=1}^L \int_{\bar{e}_l} w_1(\bar{y} - y_1)(\phi z) \, ds + \int_{\Omega} w_2(\bar{y} - y_2)(\phi z) \, dx.$$
By (3.44),
\[
\int_{\Omega} \nabla(\Delta \bar{y}) \cdot \nabla(\phi z) \, dx \leq C \| \nabla(\phi z) \|_{L^2(\Omega)} \leq C \| z \|_{H^{1+\epsilon}(\Omega)}.
\]

The trace theorem gives us that
\[
\int_{\partial \Omega} w_1(\bar{y} - y_1)(\phi z) \, ds \leq C \| \phi z \|_{L^2(\partial \Omega)} \leq C \| z \|_{H^{1+\epsilon}(\Omega)}.
\]

Finally,
\[
\phi(p_j)z(p_j) \leq C \| \phi z \|_{L^\infty(\Omega)} \leq C \| z \|_{H^{1+\epsilon}(\Omega)},
\]

by the Sobolev embedding \( H^{1+\epsilon}(\Omega) \subset C(\bar{\Omega}) \). Therefore, we have
\[
\left| \int_{\Omega} z \, d\mu_+ \right| \leq C \| z \|_{H^{1+\epsilon}(\Omega)} \quad \forall \, z \in H^2(\Omega) \cap H^1_0(\Omega)
\]
and for any \( \epsilon > 0 \).

Given any \( z \in H^{1+\epsilon}(\Omega) \), we can construct a sequence \( z_n \in H^2(\Omega) \cap H^1_0(\Omega) \) such that \( z_n \phi \) converges to \( z\phi \) in \( H^{1+\epsilon}(\Omega) \) as \( n \to \infty \). Hence we can extend the definition of the integral \( \int_{\Omega} z \, d\mu_+ \) to \( z \in H^{1+\epsilon}(\Omega) \) such that
\[
\left| \int_{\Omega} z \, d\mu_+ \right| \leq C \| z \|_{H^{1+\epsilon}(\Omega)} \quad \forall \, z \in H^{1+\epsilon}(\Omega).
\]

Similarly, the estimate (3.52) also holds if \( \mu_+ \) is replaced by \( \mu_- \).
Chapter 4. Finite Element Methods

In this chapter, we review the finite element methods (cf. [19, 39]), and in particular the \(C^0\) interior methods for fourth order boundary value problems.

4.1. Fundamental Theory

In this section, we discuss the fundamental theory of finite element methods. We will also review the interpolation error estimates and the inverse estimates of various norms on finite element spaces. Consider the following variational problem (cf. [39]): Find \(u \in V\) such that

\[
a(u, v) = F(v) \quad \forall v \in V,
\]

where \(a(\cdot, \cdot)\) is a bilinear form, \(F\) is a linear functional on \(V^*\), and \(V\) is a Hilbert space.

The Ritz-Galerkin method converts a continuous problem to a discrete problem in order to find an approximate solution which is a linear combination of basis functions. To be specific, with a finite-dimensional subspace \(V_h \subset V\), the discrete problem is to find \(u_h \in V_h\) such that

\[
a(u_h, v_h) = F(v_h) \quad \forall v_h \in V_h.
\]

Since \(V_h \subset V\), by (4.1) and (4.2), we have the Galerkin orthogonality relation:

\[
a(u - u_h, v) = 0 \quad \forall v_h \in V_h.
\]

Lemma 4.1. (Céa) (cf. [19]) Let the bilinear form \(a(\cdot, \cdot)\) be continuous and coercive on \(V\), i.e., there exist \(C < \infty\) and \(\alpha > 0\) such that

\[
|a(v, w)| \leq C\|v\|_V\|w\|_V \quad \forall v, w \in V
\]

\[
a(v, v) \geq \alpha\|v\|_V^2 \quad \forall v \in V.
\]
Suppose $u$ solves the continuous problem (4.1). Then for the discrete problem (4.2), we have

$$\|u - u_h\|_V \leq \frac{C}{\alpha} \min_{v \in V_h} \|u - v\|_V.$$  

(4.4)

**Proof.** Let $v \in V_h$ be arbitrary. By the Galerkin orthogonality, we have

$$\alpha \|u - u_h\|^2_V \leq a(u - u_h, u - u_h)$$

$$= a(u - u_h, u - v) + a(u - u_h, v - u_h)$$

$$= a(u - u_h, u - v)$$

$$\leq C \|u - u_h\|_V \|u - v\|_V.$$  

Dividing both sides by $\alpha \|u - u_h\|_V$, we have the result. \qed

**Remark 4.2.** Céa’s Lemma implies that $u_h$ is the best approximation to $u$ in the subspace $V_h$.

We need to construct a finite-dimensional subspace $V_h \subset V$ to find an approximate solution to (4.2).

**Remark 4.3.** If the finite-dimensional subspace $V_h$ is a subspace of the continuous solution space $V$, then it is called a "conforming" finite element method. If $V_h \not\subset V$, then it is called a "nonconforming" finite element method.

**Definition 4.1.** (cf. [19, Definition 3.1.1]) Let

(i) $K \subseteq \mathbb{R}^n$ be a bounded closed set with nonempty interior and piecewise smooth boundary,

(ii) $\mathcal{P}$ be a finite-dimensional space of functions on $K$, and

(iii) $\mathcal{N} = \{N_1, N_2, \ldots, N_l\}$ be a basis for the dual of $\mathcal{P}$.  

39
Then \((K, \mathcal{P}, \mathcal{N})\) is called a finite element. Also, we say \(K\) is the element, \(\mathcal{P}\) is the space of shape functions, and \(\mathcal{N}\) is the set of nodal variables.

**Definition 4.2.** (cf. [19, Definition 3.1.2]) The basis \(\{\phi_1, \phi_2, \ldots, \phi_l\}\) of \(\mathcal{P}\) such that

\[N_i(\phi_j) = \delta_{ij}\]

is called the nodal basis of \(\mathcal{P}\).

We denote \(P_k\) as the set of polynomials with degree less than or equal to \(k\). Then the dimension of \(P_k\) is computed by \((k + 1)(k + 2)/2\). For example (cf. [19]), let \(k = 2\).

![Figure 4.1. linear, quadratic, and cubic Lagrange triangle](image)

Then the shape function space \(\mathcal{P}\) is the set of polynomials with degree less than equal to 2 and \(\mathcal{N}_2 = \{N_1, \ldots, N_6\}\) where

\[N_i(v) = v(x_i) \text{ if } x_i \text{ is a vertex or a midpoint of } K.\]

The element is described in the second picture of Figure 4.1. Note that the black dot indicates the nodal variable (or degrees of freedom) evaluation at the point.

**Definition 4.3.** (cf. [19, Definition 3.3.1]) For a finite element \((K, \mathcal{P}, \mathcal{N})\), let \(\{\phi_i : 1 \leq i \leq l\} \subseteq \mathcal{P}\) be the basis dual to \(\mathcal{N}\). Then the local interpolant is defined by

\[I_K v := \sum_{i=1}^{l} N_i(v) \phi_i, \tag{4.5}\]

where \(v\) is a smooth function so that \(N_i(v)\) is defined for \(1 \leq i \leq l\).
The local interpolant $I_K$ is linear.

**Definition 4.4.** (cf. [19, Definition 3.3.11]) A triangulation of a polygonal domain $\Omega$ is a finite collection of elements $K_i$ such that

(i) $\text{int}(K_i) \cap \text{int}(K_j) = \emptyset$ for $i \neq j$

(ii) $\cup K_i = \bar{\Omega}$, and

(iii) no vertex of any triangle lies in the interior of an edge of another triangle.

If a finite collection of $K_i$ satisfies the condition (i) and (ii), then it is called a subdivision of $\Omega$.

**Definition 4.5.** (cf. [19, Definition 3.3.9]) Let $r$ be the order of the highest partial derivatives involved in the nodal variables. For $f \in C^m(\bar{\Omega})$, the global interpolant is defined by

$$I_\mathcal{T}f|_{K_i} = I_{K_i}f, \quad \forall K_i \in \mathcal{T},$$

where $\mathcal{T}$ is a subdivision of $\Omega$.

**Definition 4.6.** (cf. [19, Definition 3.3.15]) Given a subdivision $\mathcal{T}$ of $\Omega$, if $I_\mathcal{T}f \in C^r$, for any $f \in C^m(\bar{\Omega})$, then we say that $I_\mathcal{T}$ has continuity order $r$ and that the space $\mathcal{V}_\mathcal{T} = \{I_\mathcal{T}f : f \in C^m\}$ is a $C^r$ finite element space.

**Remark 4.4.** The Lagrange elements are $C^0$ elements.

**Definition 4.7.** (cf. [19, Definition 4.4.13]) For $0 < h \leq 1$, let $\{\mathcal{T}_h\}$ be a family of subdivisions of $\Omega$ such that

$$\max\{\text{diam} T : T \in \mathcal{T}_h\} \leq h \text{ diam } \Omega.$$ 

The family is said to be quasi-uniform if there exists $\rho > 0$ such that

$$\min\{\text{diam} B_T : T \in \mathcal{T}_h\} \geq \rho h \text{ diam } \Omega,$$
where $B_T$ is the largest ball contained in $T$. The family is said to be regular or non-degenerate if there exists $\rho > 0$ such that

\[
\text{diam } B_T \geq \rho \text{ diam } T.
\]  

(4.6)

**Remark 4.5.** If a family is quasi-uniform, then it is regular.

We say that two open sets $K$ and $\hat{K}$ of $\mathbb{R}^n$ are affine-equivalent if there exists an invertible affine mapping $F: \mathbb{R}^n \to \mathbb{R}^n$ such that

\[
F(\hat{x}) = B\hat{x} + b \quad \text{and} \quad K = F(\hat{K}).
\]

For affine-equivalent elements, we use the following properties:

(i) For $\hat{x} \in \hat{K}$,

\[
x = F(\hat{x}) \in K,
\]

and

(ii) For a function $\hat{v}: \hat{K} \to \mathbb{R}$,

\[
v = \hat{v} \cdot F^{-1} : K \to \mathbb{R}.
\]

With a regular family of subdivisions $\{\mathcal{T}_h\}$ of a polyhedral domain $\Omega$, the global interpolation operator $I^h : C^l(\bar{\Omega}) \to L_1(\Omega)$ is defined by

\[
I^h u|_T = I^h_T u \quad T \in \mathcal{T}_h, \ h \in (0, 1],
\]

where $I^h_T$ is the interpolation operator for the affine-equivalent element $(T, P_T, N_T)$.

Here we obtain the interpolation error estimates.

**Theorem 4.6.** (cf. [19]) Let $\{\mathcal{T}_h\}, \ 0 < h \leq 1$, be a regular family of subdivisions of $\Omega \subset \mathbb{R}^n$. Let $(K, P, N)$ be a reference element such that
(i) for any $x \in K$, the closed convex hull of $\{x\} \cup B$ is a subset of $K$.

(ii) $P_k(K) \subset \mathcal{P} \subset W^{m,\infty}(K)$, and

(iii) the nodal variables has derivatives up to order $l$.

Suppose for $m \geq 0$ and $p \in [1, \infty]$,

$$m - l - n/p > 0 \text{ when } p > 1 \quad \text{or} \quad m - l - n \geq 0 \text{ when } p = 1.$$ For all $T \in \mathcal{T}_h$, let $(T, \mathcal{P}_T, \mathcal{N}_T)$ be the affine-equivalent element. Then there exists a positive constant $C\hat{K},n,m,p,\rho$ (see (4.6)) such that for $0 \leq s \leq m$,

$$\left(\sum_{T \in \mathcal{T}_h} \left\| v - I_h v \right\|_{W^{s,p}(T)}^p \right)^{1/p} \leq C\hat{K},n,m,p,\rho h^{m-s} \left\| v \right\|_{W^{m,p}(\Omega)},$$

(4.7)

for all $v \in W^{m,p}(\Omega)$. If $p = \infty$, for $0 \leq s \leq l$,

$$\max_{T \in \mathcal{T}_h} \left\| v - I_h v \right\|_{W^{s,\infty}(T)} \leq C\hat{K},n,m,p,\rho h^{m-s-n/p} \left\| v \right\|_{W^{m,p}(\Omega)},$$

(4.8)

for all $v \in W^{m,p}(\Omega)$.

Let $K$ be a bounded domain in $\mathbb{R}^n$. For a function $v$ defined on $K$, we have $\hat{v}$ defined on $\hat{K} = \left\{ \frac{x}{h_K} : x \in K \right\}$ such that

$$\hat{v}(\hat{x}) = v(h_K \hat{x}) \quad \forall \hat{x} \in \hat{K}.$$ One can see that $v \in W^{k,r}(K)$ if and only if $\hat{v} \in W^{k,r}(\hat{K})$ and

$$\left\| \hat{v} \right\|_{W^{k,r}(\hat{K})} = h_K^{k-n/r} \left\| v \right\|_{W^{k,r}(K)}.$$ We denote $\hat{\mathcal{P}} := \{\hat{v} : v \in \mathcal{P}\}$, where $\mathcal{P}$ is a vector space of functions defined on $K$.

**Lemma 4.7.** (cf. [19, Lemma 4.5.3]) Let $\rho h \leq h_K \leq h$, $0 < h \leq 1$, and $\mathcal{P}$ be a finite dimensional subspace of $W^{l,p}(K) \cap W^{m,q}(K)$, where $p, q \in [1, \infty]$ and $0 \leq m \leq l$. Then there exists a constant $C := C_{\hat{\mathcal{P}},K,l,p,q,\rho}$ such that for all $v \in \mathcal{P}$, we have

$$\left\| v \right\|_{W^{l,p}(K)} \leq C h^{m-l+n/p-n/q} \left\| v \right\|_{W^{m,q}(K)}.$$ (4.9)
Theorem 4.8. (cf. [19, Theorem 4.5.11]) Let \( \mathcal{T}_h \}, 0 < h \leq 1, be a quasi-uniform family of subdivisions of \( \Omega \subset \mathbb{R}^n \). Let \((K, \mathcal{P}, \mathcal{N})\) be a reference finite element satisfying the assumptions in Lemma 4.7 for \( \mathcal{P}, p, q, l, m \). For all \( T \in \mathcal{T}_h \), let \((T, \mathcal{P}_T, \mathcal{N}_T)\) be the affine-equivalent element, and \( V_h = \{ v : v \text{ is measurable and } v|_T \in \mathcal{P}_T \} \). Then there exists \( C := C_{l,p,q,\rho} \) such that

\[
\left( \sum_{T \in \mathcal{T}_h} \| v \|_{W^{l,p}(T)}^p \right)^{1/p} \leq C h^{m-l+\min(0,n/p-n/q)} \left( \sum_{T \in \mathcal{T}_h} \| v \|_{W^{m,q}(T)}^q \right)^{1/p} \tag{4.10}
\]

Lastly, we take the Hsieh-Clough-Tocher (HCT) finite element as an example of \( C^1 \) conforming finite elements (cf. [38, 39]). Let \( V_h \) be the HCT finite element space. The degrees of freedom of the HCT element \( K \) (see Figure 4.2) consist of

(i) three vertices of \( K \),

(ii) first order derivatives at three vertices of \( K \), and

(iii) normal derivatives at the midpoint of three sides of \( K \).

For any \( v \in V_h \subseteq C^1(\bar{\Omega}) \), the function \( v|_T \) for \( T \in \mathcal{T}_h \) is a piecewise cubic polynomial on the three triangles in \( K \).

![Figure 4.2. Degrees of freedom of the HCT element](image)

4.2. \( C^0 \) Interior Penalty Methods

In this section we explore the basic properties of the \( C^0 \) interior penalty methods.
Let $\mathcal{T}_h$ be a simplicial triangulation of $\Omega$. We collect some notations:

- $P_k$ is the set of polynomials with degree less than or equal to $k$
- $V_h$ is the set of the vertices of $\mathcal{T}_h$
- $E_h$ is the set of the edges of $\mathcal{T}_h$
- $E_h^i$ is the set of the edges of $\mathcal{T}_h$ interior to $\Omega$
- $|e|$ is the length of an edge $e$
- $h_T$ is the diameter of $T \in \mathcal{T}_h$
- the mesh parameter $h$ equals $\max_{T \in \mathcal{T}_h} h_T$.

Let $V_h \in H^1_0(\Omega)$ be the $P_k$ $(k \geq 2)$ Lagrange finite element space associated with $\mathcal{T}_h$ (cf. [39, 19]). Note that $V_h \subseteq C(\bar{\Omega})$ but $V_h \not\subseteq C^1(\bar{\Omega})$. To construct the $C^0$ interior penalty method, we define the piecewise Sobolev space

$$H^2(\Omega, \mathcal{T}_h) = \{ v \in L^2(\Omega) : v_T = v|_T \in H^2(T) \quad \forall T \in \mathcal{T}_h \}.$$

Given a unit normal $n$ of an edge $e \in E_h^i$, the jump $[\partial v/\partial n]$ and the mean $\{\{\partial^2 v/\partial n^2\}\}$ across $e$ are defined by

$$[\partial v/\partial n] = \frac{\partial v_+}{\partial n} - \frac{\partial v_-}{\partial n} \quad \text{and} \quad \{\{\partial^2 v/\partial n^2\}\} = \frac{1}{2} \left( \frac{\partial^2 v_+}{\partial n^2} + \frac{\partial^2 v_-}{\partial n^2} \right), \quad (4.11)$$

where $v_\pm$ is the restriction of $v$ to $T_\pm$, the two triangles that share $e$ as a common edge, and the vector $n$ points from $T_-$ to $T_+$. Note that these definitions are independent of the choice of $n$, and they are well-defined on any function $v$ that is piecewise $H^2$ with respect to $\mathcal{T}_h$.

**Remark 4.9.** It follows from the Poincaré-Friedrichs inequality for piecewise $H^2$ functions in [28] that

$$\|v\|_{L^2(\Omega)} \leq C|v|_{H^2(\Omega, \mathcal{T}_h)} \quad (4.12)$$

45
Figure 4.3. The normal vector \( n_e \) in average and jump across the edge \( e \)

for all \( v \in H^1_0(\Omega) \) that is piecewise \( H^2 \) with respect to \( \mathcal{T}_h \), where the positive constant \( C \)
only depends on the shape regularity of \( \mathcal{T}_h \).

Lemma 4.10. (Discrete Sobolev Inequality) (cf. [19, Lemma 4.9.2]) For \( y_h \in V_h \),

\[
\|y_h\|_{L^\infty(\Omega)} \leq C(1 + |\ln h|)^{1/2}\|y_h\|_{H^1(\Omega)},
\]  

(4.13)

where the positive constant \( C \) is independent of \( h \).

4.2.1. Well-posedness

We apply the \( C^0 \) interior penalty method to the biharmonic problem. Let \( \Omega \subseteq \mathbb{R}^2 \)
be a bounded domain. We recall the biharmonic problem (cf. Section 2.4.2):

\[
\Delta^2 u = f \quad \text{in } \Omega \tag{2.31}
\]

\[
u = \Delta u = 0 \quad \text{on } \partial \Omega, \tag{2.32}
\]

where \( f \in L^2(\Omega) \). Then the weak formulation of the problem is to find \( u \in H^2(\Omega) \cap H^1_0(\Omega) \)
such that

\[
a(u, v) = \int_{\Omega} fv \, dx \quad \forall v \in H^2(\Omega) \cap H^1_0(\Omega), \tag{4.14}
\]

where the bilinear form \( a(\cdot, \cdot) \) is defined in (2.33). As we discussed in Section 2.4.2, the
weak problem (4.14) has a unique solution. The regularity results of the solution can be
found in Theorem 2.19 and Theorem 2.20.
The discrete problem of (4.14) is to find \( u_h \)

\[
a_h(u_h, v_h) = \int_{\Omega} f v_h \, dx \quad \forall \, v_h \in V_h,
\]

where

\[
a_h(v_h, w_h) = \sum_{T \in \mathcal{T}_h} \int_T D^2 v_h \cdot D^2 w_h \, dx + \sum_{e \in \mathcal{E}_h} \sigma h_e^{-1} \int_e \left[ \partial v_h / \partial n \right] \left[ \partial w_h / \partial n \right] \, ds
\]

\[
+ \sum_{e \in \mathcal{E}_h} \int_e \left[ \left\{ \partial^2 v_h / \partial n^2 \right\} \left[ \partial w_h / \partial n \right] + \left\{ \partial^2 w_h / \partial n^2 \right\} \left[ \partial v_h / \partial n \right] \right] \, ds,
\]

\( h_e \) is the diameter of the edge \( e \), and \( \sigma > 0 \) is a penalty parameter.

The bilinear form \( a_h(\cdot, \cdot) \) defined by (4.16) is the \( C^0 \) interior penalty bilinear form (cf. [43, 20, 11]) that approximates the bilinear form \( a(\cdot, \cdot) \) defined in (2.33). While the finite element functions are globally continuous, the normal derivatives are discontinuous across edges in \( \mathcal{E}_h \) because the finite element functions are not in \( H^2(\Omega) \). The discontinuity in normal derivatives yields the jump and the average across edges in the discrete bilinear form. Moreover through symmetrization \( \sum_{e \in \mathcal{E}_h} \int_e \left[ \left\{ \partial^2 v_h / \partial n^2 \right\} \left[ \partial w_h / \partial n \right] + \left\{ \partial^2 w_h / \partial n^2 \right\} \left[ \partial v_h / \partial n \right] \right] \, ds \) and stabilization \( \sum_{e \in \mathcal{E}_h} \sigma h_e^{-1} \int_e \left[ \partial v_h / \partial n \right] \left[ \partial w_h / \partial n \right] \, ds \), we can get the symmetric positive-definiteness (SPD) of the bilinear form \( a_h(\cdot, \cdot) \) with sufficiently large penalty parameter \( \sigma \), and thus the discrete problem preserves the SPD property of the continuous problem.

We can describe the properties of \( a_h(\cdot, \cdot) \) in terms of the norm \( \| \cdot \|_{H^2(\Omega, \mathcal{T}_h)} \) defined by

\[
|v|_{H^2(\Omega, \mathcal{T}_h)}^2 = \sum_{T \in \mathcal{T}_h} |v|_{H^2(T)}^2 + \sum_{e \in \mathcal{E}_h} h_e^{-1} \left\| \left[ \partial v / \partial n \right] \right\|_{L^2(e)}^2.
\]

To see the well-posedness of the \( C^0 \) interior penalty method, we check the coercivity and continuity of the bilinear form \( a_h(\cdot, \cdot) \). Note that for \( v_h \in V_h \),

\[
\sum_{e \in \mathcal{E}_h} h_e \left\| \left\{ \partial^2 v_h / \partial n^2 \right\} \right\|_{L^2(e)}^2 \leq C \sum_{T \in \mathcal{T}_h} |v_h|_{H^2(T)}^2
\]
by the trace theorem and the standard inverse estimates. From the Cauchy-Schwarz inequality and (4.18),

\[ \sum_e \int_e \{ \partial^2 v_h / \partial n^2 \} \{ \partial w_h / \partial n \} \ ds \]

\[ \leq \left( \sum_e h_e \{ \partial^2 v_h / \partial n^2 \}^2 \right)^{1/2} \left( \sum_e h_e^{-1} \{ \partial w_h / \partial n \}^2 \right)^{1/2} \]

\[ \leq C \left( \sum_{T \in T_h} |v_h|_{H^2(T)}^2 \right)^{1/2} \left( \sum_e h_e^{-1} \{ \partial w_h / \partial n \}^2 \right)^{1/2} \]

(4.19)

By (4.17) and (4.19), there is a positive constant \( C^\dagger \) such that

\[ |a_h(v_h, w_h)| \leq C^\dagger |v_h|_{H^2(\Omega, T_h)} |w_h|_{H^2(\Omega, T_h)} \]

for all \( v_h, w_h \in V_h \).

Also, from (4.18), (4.19) and the arithmetic-geometric mean inequality, there is a positive constant \( C_\ddagger \) such that

\[ a_h(v_h, v_h) \geq |v_h|_{H^2(T)}^2 - C \left( |v_h|_{H^2(T)}^2 \right)^{1/2} \left( \sum_e h_e^{-1} \{ \partial v_h / \partial n \}^2 \right)^{1/2} \]

\[ + \sigma \sum_e h_e^{-1} \{ \partial v_h / \partial n \}^2 \]

\[ \geq \frac{1}{2} |v_h|_{H^2(T)}^2 + \left( \sigma - \frac{C^2}{2} \right) \sum_e h_e^{-1} \{ \partial v_h / \partial n \}^2 \]

\[ \geq C_\ddagger |v_h|_{H^2(\Omega, T_h)}^2, \quad \forall v_h \in V_h, \]

(4.21)

provided \( \sigma \) is sufficiently large, which is assumed to be the case from here on.

4.2.2. Error Estimates

In this subsection, we introduce the interpolation operator and the enriching operator.
The operator $\Pi_h : H^2(\Omega) \cap H^1_0(\Omega) \to V_h$ is the Lagrange nodal interpolation operator.

$$(\Pi_h \xi)(p) = \xi(p) \quad \forall p \in V_h. \quad (4.22)$$

We have the standard interpolation error estimate (cf. [19, 39])

$$h^{-2s} \|\xi - \Pi_h \xi\|_{L^2(T)}^2 + h^{2(1-s)} \|\xi - \Pi_h \xi\|_{H^1(T)}^2 + h^{2(2-s)} \|\xi - \Pi_h \xi\|_{H^2(T)}^2 \leq C \|\xi\|_{H^s(T)}^2 \quad \forall T \in \mathcal{T}_h, \xi \in H^s(\Omega), \quad (4.23)$$

for $s > 5/2$.

We turn now to the enriching operator. Let $W_h \subset H^2(\Omega) \cap H^1_0(\Omega)$ be the Hsieh-Clough-Tocher finite element space associated with $\mathcal{T}_h$ (cf. Section 4.1). One can define an operator $E_h : V_h \to W_h$ by averaging (cf. [15, 11]) as follows

- For all $p \in V_h$,
  $$(E_h v_h)(p) = v_h(p). \quad (4.24)$$

- At an interior vertex $p$ of $\mathcal{T}_h$,
  $$(\nabla E_h v_h)(p) = \frac{1}{|\mathcal{T}_p|} \sum_{T \in \mathcal{T}_p} \nabla v_h|_T, \quad (4.25)$$

where $\mathcal{T}_p$ is the set of the elements in $\mathcal{T}_h$ that share the common vertex $p$, and $|\mathcal{T}_p|$ is the number of elements in $\mathcal{T}_p$.

- At the midpoint $m_e$ on an edge in $\mathcal{E}_h^i$,
  $$\frac{\partial E_h v_h}{\partial n}(m_e) = \frac{1}{2} \sum_{T \in \mathcal{T}_e} \frac{\partial v_h|_T}{\partial n}(m_e), \quad (4.26)$$

where $\mathcal{T}_e$ is the set of the elements in $\mathcal{T}_h$ that share the common edge $e$.  

49
We have the following estimate (cf. [11]): for all $v_h \in V_h$,

$$\sum_{T \in \mathcal{T}_h} \left( h_T^{-4} \|v_h - E_h v_h\|_{L_2(T)}^2 + h_T^{-2} \|v_h - E_h v_h\|_{H^2(T)}^2 + |E_h v_h|_{H^2(T)}^2 \right) \leq C |v_h|_{H^2(\Omega, T_h)},$$

(4.27)

where the positive constant $C$ depends only on the shape regularity of $\mathcal{T}_h$.

**Lemma 4.11.** (cf. [69]) Let $u$ be the solution of (4.14). Then

$$a_h(u, w - E_h w) \leq Ch^\alpha |u|_{H^{2+\alpha}(\Omega)} |w|_{H^2(\Omega, T_h)},$$

for any $w \in V_h$.

We have the following result.

**Theorem 4.12.** (cf. [20]) Let $u$ be the solution of (4.14) and $u_h$ be the solution of (4.15). For $f \in L_2(\Omega)$, we have

$$|u - u_h|_{H^2(\Omega, T_h)} \leq C h^\alpha \|f\|_{L_2(\Omega)},$$

(4.28)

where $\alpha \in (0, 2]$ depends on the interior angles at the corners of $\Omega$.

**Proof.** By the triangular inequality,

$$|u - u_h|_{H^2(\Omega, T_h)} \leq |u - \Pi_h u|_{H^2(\Omega, T_h)} + |\Pi_h u - u_h|_{H^2(\Omega, T_h)},$$

(4.29)

By (4.21), (4.15), and (4.20),

$$C_1 |\Pi_h u - u_h|_{H^2(\Omega, T_h)}^2 \leq a_h(\Pi_h u - u_h, \Pi_h u - u_h)$$

$$= a_h(\Pi_h u - u, \Pi_h u - u_h) + a_h(u - u_h, \Pi_h u - u_h)$$

$$= a_h(\Pi_h u - u, \Pi_h u - u_h) + [a_h(u, \Pi_h u - u_h) - (f, \Pi_h u - u_h)]$$

$$\leq C_1 |u - \Pi_h u|_{H^2(\Omega, T_h)} |\Pi_h u - u_h|_{H^2(\Omega, T_h)}$$

$$+ [a_h(u, \Pi_h u - u_h) - (f, \Pi_h u - u_h)].$$

(4.30)
Therefore,

$$|\Pi_h u - u_h|_{H^2(\Omega, \tau_h)} \leq \frac{1}{C^\dagger} \left[ C\dagger |u - \Pi_h u|_{H^2(\Omega, \tau_h)} + \frac{a_h(u, \Pi_h u - u_h) - (f, \Pi_h u - u_h)}{|\Pi_h u - u_h|_{H^2(\Omega, \tau_h)}} \right]$$

$$\leq \frac{1}{C^\dagger} \left[ C\dagger |u - \Pi_h u|_{H^2(\Omega, \tau_h)} + \sup_{w \in \mathbb{V}_h \setminus \{0\}} \frac{a_h(u, w) - (f, w)}{|w|_{H^2(\Omega, \tau_h)}} \right]. \tag{4.31}$$

Note that

$$a_h(v, w) = a(v, w) \quad \forall v, w \in H^2(\Omega) \cap H^1_0(\Omega). \tag{4.32}$$

By using (4.32), (4.14), Lemma 4.11, (4.27), and Theorem 2.20,

$$a_h(u, w) - (f, w) = a_h(u, w - E_h w) + a_h(u, E_h w) - (f, w - E_h w) - (f, E_h w)$$

$$= a_h(u, w - E_h w) - (f, w - E_h w)$$

$$\leq C \left[ h^\alpha |u|_{H^{2+\alpha}(\Omega)} |w|_{H^2(\Omega, \tau_h)} + h^2 \|f\|_{L^2(\Omega)} |w|_{H^2(\Omega, \tau_h)} \right]$$

$$\leq C (h^\alpha + h^2) \|f\|_{L^2(\Omega)} |w|_{H^2(\Omega, \tau_h)}. \tag{4.33}$$

From a standard local interpolation error estimate (cf. [25]), we have

$$|\xi - \Pi_T \xi|_{H^2(T)} \leq C h_T^\alpha |\xi|_{H^{2+\alpha}(T)}, \tag{4.34}$$

for $\xi \in H^{2+\alpha}(T)$. By combining (4.29), (4.31), (4.33), (4.34), and Theorem 2.20,

$$|u - u_h|_{H^2(\Omega, \tau_h)} \leq C \left( |u - \Pi_h u|_{H^2(\Omega, \tau_h)} + h^\alpha \|f\|_{L^2(\Omega)} \right)$$

$$\leq C (h^\alpha |u|_{H^{2+\alpha}(\Omega)} + h^\alpha \|f\|_{L^2(\Omega)})$$

$$= C h^\alpha \|f\|_{L^2(\Omega)}. \tag*{□}$$
Chapter 5. $C^0$ Interior Penalty Methods for the Elliptic Optimal Control Problem with General Tracking and Pointwise State Constraints

In this chapter, we apply the $C^0$ interior methods to the elliptic optimal control problem with general tracking and pointwise state constraints (3.21)-(3.22).

5.1. Discrete Problem

The nodal interpolation operator for the $P_1$ finite element space associated with $T_h$ is denoted by $I_h$. The discrete problem of (3.21)-(3.22) is to find

$$
\bar{y}_h = \arg\min_{y_h \in K_h} \frac{1}{2} \left[ \beta a_h(y_h, y_h) + \|y_h - y_d\|_{L^2(\Omega)}^2 \right],
$$

where

$$
K_h = \{ y_h \in V_h : I_h \psi_- \leq I_h y_h \leq I_h \psi_+ \},
$$

$a_h(\cdot, \cdot)$ is defined in (4.16), $h_e$ is the diameter of the edge $e$, and $\sigma > 0$ is a penalty parameter. Here $V_h$ is the $P_k$ ($k \geq 2$) Lagrange finite element space.

Similar to (3.26), $K_h$ is nonempty and hence the minimization problem (5.1) has a unique solution $\bar{y}_h \in K_h$ characterized by the discrete variational inequality

$$
\beta a_h(\bar{y}_h, y_h - \bar{y}_h) + \int_{\Omega} (\bar{y}_h - y_d)(y_h - \bar{y}_h) \, d\nu \geq 0 \quad \forall y_h \in K_h,
$$

which can be written as

$$
A_h(\bar{y}_h, y_h - \bar{y}_h) - \int_{\Omega} y_d(y_h - \bar{y}_h) \geq 0 \quad \forall y_h \in K_h,
$$

where the bilinear form

$$
A_h(y_h, z_h) = \beta a_h(y_h, z_h) + \int_{\Omega} y_h z_h \, d\nu
$$

approximates the bilinear form $A(\cdot, \cdot)$ in (3.38).
Recall the Lagrange nodal interpolation operator $\Pi_h : H^2(\Omega) \cap H_0^1(\Omega) \to V_h$. It follows from (3.22) and (5.2) that

$$\Pi_h \text{ maps } K \text{ into } K_h. \quad (5.6)$$

The following error estimates for $\Pi_h \bar{y}$ (cf. Section 4.2), which are based on the Bramble-Hilbert lemma (cf. [10, 41]) and the regularity estimates (3.44)-(3.45), can be found in [20, 24].

We have

$$\|\bar{y} - \Pi_h \bar{y}\|_{L^2(\Omega)} \leq C h^{2+\tau}, \quad (5.7)$$

$$|\bar{y} - \Pi_h \bar{y}|_{H^1(\Omega)} \leq C h^{1+\tau}, \quad (5.8)$$

$$\|\bar{y} - \Pi_h \bar{y}\|_{L^\infty(\Omega)} \leq C h^{1+\tau}, \quad (5.9)$$

$$\left( \sum_{T \in \mathcal{T}_h} |\bar{y} - \Pi_h \bar{y}|^2_{H^2(T)} \right)^{\frac{1}{2}} \leq C h^{\tau}, \quad (5.10)$$

where

$$\tau = \begin{cases} 
\alpha & \text{if } \mathcal{T}_h \text{ is quasi-uniform} \\
1 - \epsilon & \text{if } \mathcal{T}_h \text{ is graded around the corners of } \Omega \end{cases}, \quad (5.11)$$

where the interior angles are $>(\pi/2)$.

Here $\alpha$ is the index of elliptic regularity in (3.45) and $\epsilon$ can be any positive number.

**Remark 5.1.** Details of the graded mesh can be found in [25]. Since the singularities around the corners of $\Omega$ (with respect to $H^3$ regularity) are resolved by the graded meshes, the interpolation error estimates for $\Pi_h \bar{y}$ are determined by the interior regularity (3.44) which leads to $\tau = 1 - \epsilon$, and the constant $C$ in (5.7)-(5.10) depends on $\epsilon$. The dependence on $\epsilon$ also holds for the constants in the estimates in the rest of the dissertation where $\tau = 1 - \epsilon$. 

53
We have, by (5.8), (5.10) and the trace inequality with scaling,

\[
\left( \sum_{e \in E_h} h_e^{-1} \left\| \left( \partial \bar{y} / \partial n \right) - \left( \partial (\Pi_h \bar{y}) / \partial n \right) \right\|_{L^2(e)}^2 \right)^{\frac{1}{2}} \leq C h^\tau. \tag{5.12}
\]

Moreover it follows from (3.18) and (5.7)-(5.9) that

\[
\| \bar{y} - \Pi_h \bar{y} \|_{L^2(\Omega, \nu)} \leq C h^{1+\tau}. \tag{5.13}
\]

We also have the error estimate for the enriching operator $E_h$ (cf. Section 4.2). It follows from (3.18), (4.13) and the trace theorem that

\[
\| y_h - E_h y_h \|_{L^2(\Omega, \nu)} \leq C h(1 + | \ln h |)^{\frac{1}{2}} | y_h |_{H^2(\Omega, \mathcal{T}_h)} \quad \forall y_h \in V_h. \tag{5.14}
\]

The derivation of the following estimates that connect $\Pi_h$ and $E_h$, which reply on the regularity of $\bar{y}$ in (3.44)-(3.45) and the Bramble-Hilbert lemma, can be found in [20, 24, 21]. We have

\[
\| \bar{y} - E_h \Pi_h \bar{y} \|_{L^2(\Omega)} \leq C h^{2+\tau}, \tag{5.15}
\]

\[
| \bar{y} - E_h \Pi_h \bar{y} |_{H^1(\Omega)} \leq C h^{1+\tau}, \tag{5.16}
\]

\[
\| \bar{y} - E_h \Pi_h \bar{y} \|_{L^\infty(\Omega)} \leq C h^{1+\tau}, \tag{5.17}
\]

\[
| \bar{y} - E_h \Pi_h \bar{y} |_{H^2(\Omega)} \leq C h^\tau, \tag{5.18}
\]

and

\[
a_h(\Pi \bar{y}, y_h) - a(\bar{y}, E_h y_h) \leq C h^\tau | y_h |_{H^2(\Omega, \mathcal{T}_h)} \quad \forall y_h \in V_h, \tag{5.19}
\]

where $\tau$ is given by (5.11).

**Remark 5.2.** The estimates (5.15)-(5.18) indicate that $E_h \Pi_h$ behaves like a quasi-interpolation operator. The estimate (5.19) indicates that $E_h$ and $\Pi_h$ are approximate adjoint operators with respect to the bilinear forms $a(\cdot, \cdot)$ and $a_h(\cdot, \cdot)$. 
5.2. Convergence Analysis

We will carry out the convergence analysis for (3.21)-(3.22) in terms of the mesh-dependent norm $\| \cdot \|_h$ defined by

$$\|v\|_h^2 = \beta |v|_{H^2(\Omega; \mathcal{T}_h)}^2 + \|v\|_{L^2(\Omega; \nu)}^2. \quad (5.20)$$

In view of (4.17)-(4.21), (3.38) and (5.20), there exist positive constants $C_\sharp$ and $C_\flat$ such that

$$\mathcal{A}_h(y_h, z_h) \leq C_\sharp \|y_h\|_h \|z_h\|_h \quad \forall y_h, z_h \in V_h, \quad (5.21)$$

$$\mathcal{A}_h(y_h, y_h) \geq C_\flat \|y_h\|_h^2 \quad \forall y_h \in V_h. \quad (5.22)$$

Moreover (5.10)-(5.13) and (5.20) imply

$$\|\bar{y} - \Pi_h \bar{y}\|_h \leq C h^\tau. \quad (5.23)$$

By (5.6), we have $\Pi_h \bar{y} \in K_h$ and hence from (5.4),

$$\mathcal{A}_h(\bar{y}, \Pi_h \bar{y} - \bar{y}_h) \geq \int_\Omega y_d (\Pi_h \bar{y} - \bar{y}_h). \quad (5.24)$$

It follows from (5.22), (5.23) and (5.24) that

$$\|\bar{y} - \bar{y}_h\|_h^2 \leq 2\|\bar{y} - \Pi_h \bar{y}\|_h^2 + 2\|\Pi_h \bar{y} - \bar{y}_h\|_h^2 \leq C_1 h^{2\tau} + C_2 \mathcal{A}_h(\Pi_h \bar{y} - \bar{y}_h, \Pi_h \bar{y} - \bar{y}_h) \quad (5.25)$$

$$\leq C_1 h^{2\tau} + C_2 \left[ \mathcal{A}_h(\Pi_h \bar{y}, \Pi_h \bar{y} - \bar{y}_h) - \int_\Omega y_d (\Pi_h \bar{y} - \bar{y}_h) d\nu \right].$$

It is important to bound the second term on the right-hand side of (5.25). We will show that

$$\mathcal{A}_h(\Pi_h \bar{y}, \Pi_h \bar{y} - \bar{y}_h) - \int_\Omega y_d (\Pi_h \bar{y} - \bar{y}_h) d\nu \leq C (h^{2\tau} + h^\tau \|\Pi_h \bar{y} - \bar{y}_h\|_h), \quad (5.26)$$

and then the convergence analysis is completed as follows.
Theorem 5.3. There exists a positive constant $C$ independent of $h$ such that

$$\| \bar{y} - \bar{y}_h \|_h \leq Ch^\tau, \quad (5.27)$$

where $\tau$ is given by (5.11).

Proof. We have, by (5.23)-(5.26),

$$\| \bar{y} - \bar{y}_y \|_h^2 \leq C \left[ h^{2\tau} + h^\tau \left( \| \Pi_h \bar{y} - \bar{y}_h \|_h + \| \bar{y} - \bar{y}_h \|_h \right) \right] \leq C \left[ h^{2\tau} + h^\tau \| \bar{y} - \bar{y}_h \|_h \right].$$

From the inequality of arithmetic and geometric means,

$$\| \bar{y} - \bar{y}_y \|_h^2 \leq C \left[ h^{2\tau} + \frac{\| \bar{y} - \bar{y}_y \|_h^2 + h^{2\tau}}{2} \right],$$

and it implies that we obtain the desired result. \qed

We will establish (5.26) by reducing it to an estimate at the continuous level.

We begin with an analog of (5.19).

Lemma 5.4. There exists a positive constant $C$ independent of $h$ such that

$$\mathcal{A}_h(\Pi_h \bar{y}, y_h) - \mathcal{A}(\bar{y}, E_h y_h) \leq C h^\tau \| y_h \|_h \quad \forall y_h \in V_h. \quad (5.28)$$

Proof. According to (3.38) and (5.5), we have

$$\mathcal{A}_h(\Pi_h \bar{y}, y_h) - \mathcal{A}(\bar{y}, E_h y_h) = \beta [a_h(\Pi_h \bar{y}, y_h) - a(\bar{y}, E_h y_h)]$$

$$+ \int_{\Omega} (\Pi_h \bar{y} - \bar{y}) y_h \, dv + \int_{\Omega} \bar{y}(y_h - E_h y_h) \, dv,$$

which together with (4.12), (5.11), (5.13), (5.14) and (5.19) implies (5.28). \qed

It follows from Lemma 5.4 that

$$\mathcal{A}_h(\Pi_h \bar{y}, \Pi_h \bar{y} - \bar{y}_h) \leq \mathcal{A}(\bar{y}, E_h (\Pi_h \bar{y} - \bar{y}_h)) + C h^\tau \| \Pi_h \bar{y} - \bar{y}_h \|_h. \quad (5.29)$$
Furthermore, the Cauchy-Schwarz inequality and the estimate (5.14) imply

\[- \int_\Omega y_d (\Pi_h \bar{y} - \bar{y}_h) \, d\nu \]

\[= - \int_\Omega y_d E_h (\Pi_h \bar{y} - \bar{y}_h) \, d\nu + \int_\Omega y_d (E_h (\Pi_h \bar{y} - \bar{y}_h) - (\Pi_h \bar{y} - \bar{y}_h)) \, d\nu \]

\[\leq - \int_\Omega y_d E_h (\Pi_h \bar{y} - \bar{y}_h) \, d\nu + C \|E_h (\Pi_h \bar{y} - \bar{y}_h) - (\Pi_h \bar{y} - \bar{y}_h)\|_{L^2(\Omega;\nu)} \]

\[\leq - \int_\Omega y_d E_h (\Pi_h \bar{y} - \bar{y}_h) \, d\nu + Ch^r \|\Pi_h \bar{y} - \bar{y}_h\|_h. \quad (5.30)\]

Putting (5.29) and (5.30) together, we arrive at

\[\mathcal{A}_h (\Pi_h \bar{y}, \Pi_h \bar{y} - \bar{y}_h) - \int_\Omega y_d (\Pi_h \bar{y} - \bar{y}_h) \, d\nu \]

\[\leq [\mathcal{A}(\bar{y}, E_h (\Pi_h \bar{y} - \bar{y}_h)) - \int_\Omega y_d E_h (\Pi_h \bar{y} - \bar{y}_h) \, d\nu] + Ch^r \|\Pi_h \bar{y} - \bar{y}_h\|_h. \]

Comparing (5.26) and (5.31), we see that the proof of (5.26) has been reduced to an estimate for the first term on the right-hand side of (5.31), which does not involve the discrete bilinear form \(\mathcal{A}_h (\cdot, \cdot)\). Therefore we can use (3.37) and (3.47) to write

\[\mathcal{A}(\bar{y}, E_h (\Pi_h \bar{y} - \bar{y}_h)) - \int_\Omega y_d E_h (\Pi_h \bar{y} - \bar{y}_h) \, d\nu \]

\[= \int_\Omega E_h (\Pi_h \bar{y} - \bar{y}_h) \, d\mu_+ + \int_\Omega E_h (\Pi_h \bar{y} - \bar{y}_h) \, d\mu_. \quad (5.32)\]

We will focus on the first integral on the right-hand side of (5.32). The other integral can be bounded analogously.

In view of (3.50), we can write

\[\int_\Omega E_h (\Pi_h \bar{y} - \bar{y}_h) \, d\mu_+ = \int_\Omega (E_h \Pi_h \bar{y} - \bar{y}) \, d\mu_+ + \int_\Omega (\psi_+ - I_h \psi_+) \, d\mu_+ \]

\[+ \int_\Omega (I_h \psi_+ - I_h E_h \bar{y}_h) \, d\mu_+ + \int_\Omega (I_h E_h \bar{y}_h - E_h \bar{y}_h) \, d\mu_. \quad (5.33)\]
For the first term on the right-hand side of (5.33), we have
\[ \int_{\Omega} \left( E_h \Pi_h \bar{y} - \bar{y} \right) d\mu_+ \leq |\mu_+| \left| \left| \Pi_h \bar{y} - \bar{y} \right| \right|_{L_\infty(\Omega)} \leq Ch^{1+\tau} \tag{5.34} \]
by (5.17), and we can use a standard interpolation error estimate for \( I_h \) (cf. [39, 19]) to bound the second term by
\[ \int_{\Omega} (\psi_+ - I_h \psi_+) d\mu_+ \leq |\mu_+| \left| \left| \psi_+ - I_h \psi_+ \right| \right|_{L_\infty(\Omega)} \leq Ch^2 \tag{5.35} \]
because \( \psi_+ \) belongs to \( W^{2,\infty}(\Omega) \) by the assumption (3.15) and the Sobolev embedding \( W^{1,q}(\Omega) \subset C(\bar{\Omega}) \) for \( q > 2 \).

Since \( \mu_+ \) is nonpositive, the third term satisfies
\[ \int_{\Omega} (I_h \psi_+ - I_h E_h \bar{y}_h) d\mu_+ = \int_{\Omega} (I_h \psi_+ - I_h \bar{y}_h) d\mu_+ \leq 0 \tag{5.36} \]
by (5.2) and (4.24).

We can split the last term on the right-hand side of (5.33) as
\[ \int_{\Omega} (I_h E_h \bar{y}_h - E_h \bar{y}_h) d\mu_+ \tag{5.37} \]
\[ = \int_{\Omega} \left[ I_h (E_h \bar{y}_h - \bar{y}) - (E_h \bar{y}_h - \bar{y}) \right] d\mu_+ + \int_{\Omega} (I_h \bar{y} - \bar{y}) d\mu_+ , \]
and we have
\[ \int_{\Omega} (I_h \bar{y} - \bar{y}) d\mu_+ \leq |\mu_+| \left| \left| I_h \bar{y} - \bar{y} \right| \right|_{L_\infty(\Omega)} \leq C_\varepsilon h^{2-\varepsilon} \tag{5.38} \]
by (3.44) and a standard interpolation error estimate for \( I_h \).

Finally, it follows from (3.52), (4.27), (5.18) and a standard interpolation error esti-
mate for $I_h$ that

$$\int_\Omega [I_h(E_h\bar{y}_h - \bar{y}) - (E_h\bar{y}_h - \bar{y})] \, d\mu_+ \leq C|I_h(E_h\bar{y}_h - \bar{y}) - (E_h\bar{y}_h - \bar{y})|_{H^{1+\epsilon}(\Omega)}$$

$$\leq Ch^{1-\epsilon}|E_h\bar{y}_h - \bar{y}|_{H^2(\Omega)}$$

$$\leq Ch^{1-\epsilon}(|E_h(\bar{y}_h - \Pi_h\bar{y})|_{H^2(\Omega)} + |E_h\Pi_h\bar{y} - \bar{y}|_{H^2(\Omega)})$$

$$\leq Ch^{1-\epsilon}(\|\bar{y}_h - \Pi_h\bar{y}\|_h + h^\tau). \quad (5.39)$$

Combining (5.33)-(5.39), we have

$$\int_\Omega E_h(\Pi_h\bar{y} - \bar{y}_h) \, d\mu_+ \leq C(h^{1+\tau} + h^2 + h^{2-\epsilon} + h^{1+\tau-\epsilon} + h^{1-\epsilon}\|\Pi_h\bar{y} - \bar{y}_h\|_h)$$

and hence, in view of (5.11),

$$\int_\Omega E_h(\Pi_h\bar{y} - \bar{y}_h) \, d\mu_+ \leq C(h^{2\tau} + h^\tau\|\Pi_h\bar{y} - \bar{y}_h\|_h). \quad (5.40)$$

Similarly, we have

$$\int_\Omega E_h(\Pi_h\bar{y} - \bar{y}_h) \, d\mu_- \leq C(h^{2\tau} + h^\tau\|\Pi_h\bar{y} - \bar{y}_h\|_h). \quad (5.41)$$

The estimate (5.26) follows from (5.31), (5.40) and (5.41).

We can approximate the optimal control $\bar{u}$ by $\bar{u}_h = -\Delta_h\bar{y}_h$, where $\Delta_h$ is the piecewise defined Laplacian with respect to $T_h$. The following result is a direct consequence of (5.20) and Theorem 5.3.

**Corollary 5.5.** There exists a positive constant $C$ independent of $h$ such that

$$\|\bar{u} - \bar{u}_h\|_{L^2(\Omega)} \leq Ch^\tau.$$

We can also establish error estimates for $\bar{y} - \bar{y}_h$ in lower order norms.
Corollary 5.6. There exists a positive constant $C$ independent of $h$ such that

$$
\| \bar{y} - \bar{y}_h \|_{L^2(\Omega)} + | \bar{y} - \bar{y}_h |_{H^1(\Omega)} + \| \bar{y} - \bar{y}_h \|_{L^\infty(\Omega)} \leq C h^r.
$$

Proof. The estimates for $\| \bar{y} - \bar{y}_h \|_{L^2(\Omega)}$ and $| \bar{y} - \bar{y}_h |_{H^1(\Omega)}$ follow immediately from (5.20), Theorem 5.3 and the Poincaré-Friedrichs inequality for piecewise $H^2$ functions in [28].

Similarly the estimate for $\| \bar{y} - \bar{y}_h \|_{L^\infty(\Omega)}$ follows from (5.20), Theorem 5.3 and the Sobolev inequality for piecewise $H^2$ functions in [16].
Chapter 6. Numerical Implementation

Recall the discrete minimization problem (5.1)-(5.2): find

\[ \bar{y}_h = \arg\min_{y_h \in K_h} \frac{1}{2} \left[ \beta a_h(y_h, y_h) + \| y_h - y_d \|_{L_2(\Omega; \nu)}^2 \right], \]  

(5.1)

where

\[ K_h = \{ y_h \in V_h : I_h \psi_- \leq I_h y_h \leq I_h \psi_+ \}. \]  

(5.2)

6.1. A Primal-Dual Active Set Algorithm

In this section, we consider the matrix-vector form of the discrete problem (5.1)-(5.2). Suppose \( V_h = \text{Span}\{ \varphi_i \}_{i=1}^{N_d} \), where \( \varphi_i \)'s are \( P_2 \) basis functions corresponding to the degrees of freedom, which consists of the interior vertices and the interior midpoints of \( T_h \). Then for all \( y_h \in V_h \), we obtain

\[ y_h = \sum_{i=1}^{N_d} c_i \varphi_i, \]  

(6.1)

where \( c_i \)'s are constants, and \( N_d \) is the number of the degrees of freedom. By using (5.4) and (5.5), we can rewrite the discrete minimization problem (5.1)-(5.2) as

\[ \min_{y_h \in K_h} \left[ \frac{1}{2} A_h(y_h, y_h) - \mathcal{F}(y_h) \right], \]  

(6.2)

where

\[ \mathcal{F}(v) = \int_{\Omega} y_d v \, d\nu. \]  

(6.3)

6.1.1. Matrix-Vector Form of Discrete Minimization Problem

The matrix-vector form of (6.2) can be written as

\[ \min_{y \in K} \left[ \frac{1}{2} y^T A y - y^T f \right], \]  

(6.4)
where

$$K = \{ \mathbf{v} \in \mathbb{R}^{Nd} : \psi_{\pm}(n_i) \leq \mathbf{v}(i) \leq \psi_{+}(n_i) \text{ for } i = 1, \ldots, N_{iv} \}$$  \hspace{1cm} (6.5)$$

where $N_{iv}$ is the number of the interior vertices $V_{i}^{i}$, $n_i$ is the $xy$-coordinates of $i$-th interior vertex, and $N_m$ is the number of the interior midpoints of the triangulation $T_h$. Here $Nd = N_{iv} + N_m$. Note that $\mathbf{v}(1:N_{iv})$ indicates the values of $\mathbf{v}$ at the interior node points and $\mathbf{v}(N_{iv} + 1:N_d)$ indicates the values of $\mathbf{v}$ at the interior midpoints of the triangulation $T_h$. Also note that the number of the interior midpoints of the triangulation $T_h$, $N_m$, is the same as the number of interior edges in $E_{h}^i$.

The corresponding matrix-vector form of the variational inequality with pointwise state constraints is

$$y^T A(x - y) - f^T (x - y) \geq 0 \quad \forall \, x \in \mathbb{R}^{Nd},$$  \hspace{1cm} (6.6)$$

$$\Psi_\pm \leq \mathbf{S} y \leq \Psi_+,$$  \hspace{1cm} (6.7)$$

where the vectors $\Psi_\pm \in \mathbb{R}^{N_{iv}}$ such that

$$\Psi_-(i) = \psi_{\pm}(n_i), \quad \text{and} \quad \Psi_+(i) = \psi_{+}(n_i) \quad \text{for } i = 1, \ldots, N_{iv},$$  \hspace{1cm} (6.8)$$

and the matrix $\mathbf{S} \in \mathbb{R}^{N_{iv} \times Nd}$ such that

$$\mathbf{S} = \left[ \begin{array}{cc} I_{N_{iv} \times N_{iv}} & O_{N_{iv} \times N_m} \end{array} \right].$$  \hspace{1cm} (6.9)$$

### 6.1.2. Matrix-Vector Form of Discrete KKT Conditions

By introducing the dual unknown $\lambda$, we have the discrete problem which is equivalent to (6.4)-(6.5):

$$\begin{cases} \mathbf{A} y - \mathbf{f} + \mathbf{S}^T \lambda = 0 \\ \lambda = \max\{0, \lambda + C(\mathbf{S} y - \Psi_+)\} + \min\{0, \lambda + C(\mathbf{S} y - \Psi_-)\}. \end{cases}$$  \hspace{1cm} (6.10)$$
The KKT conditions give us that

\[
\begin{cases}
Ay + S^T \lambda = f \\
\Psi_- \leq Sy \leq \Psi_+ \\
\lambda \geq 0 & \text{if } Sy = \Psi_+ \\
\lambda \leq 0 & \text{if } Sy = \Psi_- \\
\lambda = 0 & \text{if } \Psi_- < Sy < \Psi_+,
\end{cases}
\]

(6.11)

where \( \lambda \in \mathbb{R}^{N_{iv}} \).

### 6.1.3. Primal Dual Active Set Algorithm

1. Initial unknown \( y_0 \) is given. Set initial dual unknown \( \lambda_0 = 0 \) and initial step \( k = 0 \).

2. Define the active sets

\[
\mathcal{A}_-^k = \{ i : \lambda^k(i) + C(Sy^k(i) - \Psi_-(i)) < 0 \},
\]

\[
\mathcal{A}_+^k = \{ i : \lambda^k(i) + C(Sy^k(i) - \Psi_+(i)) > 0 \},
\]

and the inactive set

\[
\mathcal{J}^k = \{ 1, 2, \ldots, N_{iv} \} \setminus (\mathcal{A}_-^k \cup \mathcal{A}_+^k).
\]

3. Solve

\[
Ay^{k+1} + S^T \lambda^{k+1} = f,
\]

\[
Sy^{k+1} = \Psi_- \quad \text{on } \mathcal{A}_-^k,
\]

\[
Sy^{k+1} = \Psi_+ \quad \text{on } \mathcal{A}_+^k,
\]

\[
\lambda^{k+1} = 0 \quad \text{on } \mathcal{J}^k.
\]
(4) Update

$$\mathcal{A}^{k+1} = \{ i : \lambda^{k+1}(i) + C(Sy^{k+1}(i) - \Psi_{\pm}(i)) < 0 \},$$

$$\mathcal{A}^{k+1}_{+} = \{ i : \lambda^{k+1}(i) + C(Sy^{k+1}(i) - \Psi_{+}(i)) > 0 \},$$

(5) Terminate the iteration when $$\mathcal{A}^{k+1} = \mathcal{A}^{k}$$ and $$\mathcal{A}^{k+1}_{+} = \mathcal{A}^{k}_{+}$$.

We can solve the system in step (3) simultaneously by computing the following block matrix:

$$\begin{bmatrix}
A(N_d, N_d) & S^T(N_d, \mathcal{A}^{k}_{-}) & S^T(N_d, \mathcal{A}^{k}_{+}) \\
S(\mathcal{A}^{k}, N_d) & 0(\mathcal{A}^{k}, \mathcal{A}^{k}_{-}) & 0(\mathcal{A}^{k}, \mathcal{A}^{k}_{+}) \\
S(\mathcal{A}^{k}_{+}, N_d) & 0(\mathcal{A}^{k}_{+}, \mathcal{A}^{k}_{-}) & 0(\mathcal{A}^{k}_{+}, \mathcal{A}^{k}_{+})
\end{bmatrix}
\begin{bmatrix}
y^{k+1} \\
\lambda^{k+1}(\mathcal{A}^{k}) \\
\lambda^{k+1}_{+}(\mathcal{A}^{k}_{+})
\end{bmatrix}
= 
\begin{bmatrix}
f \\
\Psi_{-}(\mathcal{A}^{k}) \\
\Psi_{+}(\mathcal{A}^{k}_{+})
\end{bmatrix}.$$

**Theorem 6.1.** (cf. [48, Theorem 3.1]) Let $$y_{\ast}$$ and $$\lambda_{\ast}$$ be the unique solution to (6.10). The primal dual active set algorithm converges superlinearly provided that the initial state $$y_{0}$$ and the initial dual $$\lambda_{0}$$ are sufficiently close to $$y_{\ast}$$ and $$\lambda_{\ast}$$.

### 6.2. Numerical Experiments

In this section we report the results for various numerical experiments on square domains where the computations were carried out on uniform meshes. Consequently $$\tau$$ equals $$1 - \epsilon$$ for any $$\epsilon > 0$$ in Theorem 5.3, Corollary 5.5 and Corollary 5.6.

Consider the reduced optimal control problem with one-sided pointwise state constraint and general tracking:

$$\min_{y \in K} \frac{1}{2} \left[ \beta \|\Delta y\|_{L^2(\Omega)}^2 + \sum_{j=1}^{J} w_0(p_j) [y(p_j) - y_0(p_j)]^2 \\
+ \sum_{i=1}^{L} \int_{\mathcal{E}_i} w_1(y - y_1)^2 \, ds + \int_{\Omega} w_2(y - y_2)^2 \, dx \right],$$

(6.12)
where \( p_1, \ldots, p_J \in \Omega \), and

\[
K = \{ y \in H^2(\Omega) \cap H^1_0(\Omega) : y \leq \psi_+ \text{ on } \Omega \}.
\]

Let \( V_h \) be the \( P_2 \) Lagrange finite element space. The discrete problem of (6.12) is given by

\[
\min_{y_h \in K_h} \frac{1}{2} \left[ \beta a_h(y_h, y_h) + \sum_{j=1}^{J} w_0(p_j)[y_h(p_j) - y_0(p_j)]^2 \right. \\
\left. + \sum_{l=1}^{L} \int_{E_l} w_1(y_h - y_1)^2 \, ds + \int_{\Omega} w_2(y_h - y_2)^2 \, dx \right], \tag{6.13}
\]

where

\[
K_h = \{ y_h \in V_h : I_h y_h \leq I_h \psi_+ \text{ on } \Omega \},
\]

\( a_h(\cdot, \cdot) \) is the bilinear form corresponding to \( \| \Delta \cdot \|_{L^2(\Omega)}^2 \), and \( I_h \) is the nodal Lagrange \( P_1 \) interpolation operator.

**Example 6.2.1.** The example only involves point tracking and region tracking, and the exact solution is constructed.

Consider the optimal control problem (6.12) with \( w_1 = 0 \) and \( w_2 = 1 \). Then the problem becomes to find

\[
\bar{y} = \arg\min_{y \in K} \frac{1}{2} \left[ \beta \| \Delta y \|_{L^2(\Omega)}^2 + \| y - y_2 \|_{L^2(\Omega)}^2 + \sum_{j=1}^{J} w_0(p_j)[y(p_j) - y_0(p_j)]^2 \right], \tag{6.14}
\]

where \( p_1, \ldots, p_J \in \Omega \) and

\[
K = \{ y \in H^2(\Omega) \cap H^1_0(\Omega) : y \leq \psi_+ \text{ on } \Omega \}. \tag{6.15}
\]

**Exact Solution.** We will construct an exact solution to the problem. Suppose a function \( v \in H^3(\Omega) \) is piecewise smooth with respect to \( \mathcal{A} \) and satisfies (cf. [14, Example 7.1])
(i) $v = \psi_+$ on $\mathcal{A}$ and $v \leq \psi_+$ on $\Omega \setminus \mathcal{A}$,

(ii) $v = \Delta v = 0$ on $\partial \Omega$, and

(iii) $\left[ \frac{\partial (\Delta v)}{\partial n} \right] \geq 0$ on $\partial \mathcal{A}$.

Choose the points $p_1, \ldots, p_J \in \Omega \setminus \mathcal{A}$ and define

$$
\phi_j(x) = \frac{1}{8\pi} |x - p_j|^2 \ln |x - p_j|.
$$

Note that

$$
\phi_j(p_j) = 0, \quad \Delta \phi_j = \frac{1}{2\pi} \left( \ln |x - p_j| + 1 \right), \quad \text{and} \quad \Delta^2 \phi_j = \delta_{p_j},
$$

where $\delta_{p_j}$ is the Dirac delta function associated with $p_j$.

We introduce a function $\Phi \in C^3_c(\Omega)$ such that $\Phi = 1$ near $p_1, \ldots, p_J$ and $\Phi = 0$ in a neighborhood of $\mathcal{A}$. The exact solution $\bar{y}$ is given by

$$
\bar{y} = v + \epsilon \sum_{j=1}^{J} \phi_j \Phi,
$$

where $\epsilon$ is sufficiently small so that $\bar{y} = \psi_+$ on $\mathcal{A}$ and $\bar{y} < \psi_+$ on $\Omega \setminus \mathcal{A}$.

Now we set positive weights $w_0(p_j)$ and the target functions $y_0(p_j)$ by

$$
y_0(p_j) = \bar{y}(p_j) + \frac{\epsilon \beta}{w_0(p_j)} = v(p_j) + \epsilon \left( \sum_{j=1}^{J} \phi_i(p_j) + \frac{\beta}{w_0(p_j)} \right),
$$

for $j = 1, \ldots, J$.

We finally set the target function $y_2 \in L_2(\Omega)$ by

$$
y_2 = \begin{cases} 
Lv + \epsilon \left( \beta \sum_{j=1}^{J} R(\phi_j, \Phi) + \sum_{j=1}^{J} \phi_j \Phi \right) & \text{on} \Omega \setminus \mathcal{A}, \\
Lv + \gamma & \text{on} \mathcal{A},
\end{cases}
$$

where $L$ is the differential operator $\beta \Delta^2 + 1$, $R(\cdot, \cdot)$ is the bilinear form defined by

$$
R(f, g) = 4\nabla(\Delta f) \cdot \nabla g + 2(\Delta f)(\Delta g) + 4D^2f : D^2g + 4\nabla f \cdot \nabla(\Delta g) + f(\Delta^2 g),
$$

66
and $\gamma$ is a nonnegative constant.

Note that by the settings of $\phi_j$, $\Phi$, and integration by parts,

$$\int_{\Omega} \Delta(\phi_j \Phi) \Delta z \, dx = z(p_j) + \int_{\Omega} R(\phi_j, \Phi) z \, dx.$$

One can show that

$$\beta \int_{\Omega} (\Delta \bar{y})(\Delta z) \, dx + \int_{\Omega} (\bar{y} - y_2)z \, dx + \sum_{j=1}^{J} w_0(p_j) [\bar{y}(p_j) - y_0(p_j)] z(p_j)$$

$$= -\gamma \int_{\partial \mathcal{A}} z \, d\mathcal{A} - \beta \int_{\partial \mathcal{A}} \left[ \partial(\Delta v)/\partial n \right] z \, ds, \quad \forall \, z \in H^2(\Omega) \cap H^1_0(\Omega).$$

Let $\mu$ be a regular Borel measure such that

$$\int_{\Omega} z \, d\mu = -\gamma \int_{\partial \mathcal{A}} z \, ds - \beta \int_{\partial \mathcal{A}} \left[ \partial(\Delta v)/\partial n \right] z \, ds,$$

for any $z \in H^2(\Omega) \cap H^1_0(\Omega)$. Since $\gamma$ and $\beta$ are positive constants, and $\left[ \partial(\Delta v)/\partial n \right] \geq 0$ on $\partial \mathcal{A}$, it implies that $\mu$ is a nonpositive measure. If we set $z = \bar{y} - \psi_+$, then

$$\int_{\Omega} (\bar{y} - \psi_+) \, d\mu = -\gamma \int_{\partial \mathcal{A}} (\bar{y} - \psi_+) \, ds - \beta \int_{\partial \mathcal{A}} \left[ \partial(\Delta v)/\partial n \right] (\bar{y} - \psi_+) \, ds = 0,$$

because $\bar{y} = \psi_+$ on $\mathcal{A}$. Thus $\bar{y}$ satisfies the KKT conditions (3.28), (3.30), (3.31), and (3.50), i.e., $\bar{y}$ is the exact solution to (6.14)-(6.15).

**Numerical Experiments.** Set $\Omega = [-4, 4] \times [-4, 4]$ and $\beta = 1$. The state constraint $\psi_+$ is given by

$$\psi_+(x) = |x|^2 - 1,$$

and the active set

$$\mathcal{A} = \{ x : |x| \leq 1 \}.$$
We track the state at points $p_j, j = 1, 2, 3, 4$, which do not belong to $V_h$, are given by

$$
p_1 = (-2.49, -2.51), \quad p_2 = (2.51, -2.15), \quad p_3 = (2.49, 2.51), \quad p_4 = (-2.51, 2.49),
$$

with the weights $w_0(p_j) = 100, j = 1, 2, 3, 4$. The function $\Phi \in C^3_c(\Omega)$ is given by

$$
\Phi(x) = \zeta(x_1)\zeta(x_2),
$$

where

$$
\zeta(t) = \begin{cases}
0 & \text{if } r_4 \leq t \\
\eta\left(\frac{t-r_3}{r_4-r_3}\right) & \text{if } r_3 \leq t \leq r_4 \\
1 & \text{if } r_2 \leq t \leq r_3 \\
\eta\left(\frac{r_2-t}{r_2-r_1}\right) & \text{if } r_1 \leq t \leq r_2 \\
0 & \text{if } l_1 \leq t \leq r_1 \\
\eta\left(\frac{t-l_4}{l_1-l_2}\right) & \text{if } l_2 \leq t \leq l_1 \\
1 & \text{if } l_3 \leq t \leq l_2 \\
\eta\left(\frac{l_3-t}{l_3-l_4}\right) & \text{if } l_4 \leq t \leq l_3 \\
0 & \text{if } t \leq l_4,
\end{cases}
$$

with $\eta(t) = (1 + 4t + 10t^2 + 20t^3)(1 - t)^4$ and

$$
l_4 = -7/2, \quad l_3 = -3, \quad l_2 = -2, \quad l_1 = -3/2, \\
r_4 = 7/2, \quad r_3 = 3, \quad r_2 = 2, \quad r_1 = 3/2.
$$

We set $\epsilon = 0.1$ in the construction of the exact solution $\bar{y}$ defined by (6.16).

We use the $P_2$ Lagrange finite element on uniform meshes to discretize the problem and the penalty parameter $\sigma = 100$. The numerical results are presented in Tables 5.1-5.2,
where the relative errors of the state are defined by

\[ e_\infty = \max_{p \in \mathcal{V}_h} |\bar{y}(p) - \bar{y}_h(p)|/\|\bar{y}\|_{L_\infty(\Omega)}, \quad e_0 = \|\bar{y} - \bar{y}_h\|_{L_2(\Omega)}/\|\bar{y}\|_{L_2(\Omega)}, \]

\[ e_1 = \|\bar{y} - \bar{y}_h\|_{H^1(\Omega)}/\|\bar{y}\|_{H^1(\Omega)}, \quad e_h = \|\bar{y} - \bar{y}_h\|_h/\|\bar{y}\|_{H^2(\Omega)}. \]

The convergence for the state in the $\|\cdot\|_h$ norm (resp., the convergence of the control in the $L_2$ norm) agrees with Theorem 5.3 (resp., Corollary 5.5) where $\tau = 1 - \epsilon$. The convergence of the state in the other norms are better than the convergence predicted by Corollary 5.6. The graphs for the optimal state, optimal control and active set are displayed in Figure 5.1 and Figure 5.2.

Table 6.1. Relative error of the state versus mesh size $h$ and orders of convergence for Example 6.2.1.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e_{\infty}^r$</th>
<th>order</th>
<th>$e_0^r$</th>
<th>order</th>
<th>$e_1^r$</th>
<th>order</th>
<th>$e_h^r$</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^{-1}$</td>
<td>3.5132e-1</td>
<td>2.0660e-1</td>
<td>6.6432e-1</td>
<td>1.9206e0</td>
<td>1.2239e0</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$2^{-2}$</td>
<td>1.4843e-1</td>
<td>1.24</td>
<td>8.5745e-2</td>
<td>1.27</td>
<td>3.1322e-1</td>
<td>1.08</td>
<td>5.3464e-1</td>
<td>0.93</td>
</tr>
<tr>
<td>$2^{-3}$</td>
<td>1.4030e-1</td>
<td>0.05</td>
<td>4.6489e-2</td>
<td>0.88</td>
<td>2.0671e-1</td>
<td>0.60</td>
<td>1.0183e0</td>
<td>0.27</td>
</tr>
<tr>
<td>$2^{-4}$</td>
<td>5.7321e-2</td>
<td>1.32</td>
<td>1.7819e-2</td>
<td>1.39</td>
<td>8.9037e-2</td>
<td>1.22</td>
<td>5.3464e-1</td>
<td>0.93</td>
</tr>
<tr>
<td>$2^{-5}$</td>
<td>2.1617e-2</td>
<td>1.41</td>
<td>5.8463e-3</td>
<td>1.62</td>
<td>3.5021e-2</td>
<td>1.35</td>
<td>2.6932e-1</td>
<td>0.99</td>
</tr>
<tr>
<td>$2^{-6}$</td>
<td>6.7055e-3</td>
<td>1.69</td>
<td>1.6982e-3</td>
<td>1.77</td>
<td>1.1445e-2</td>
<td>1.62</td>
<td>1.2753e-1</td>
<td>1.08</td>
</tr>
</tbody>
</table>

Table 6.2. Relative error of the control versus mesh size $h$ and orders of convergence for Example 6.2.1.

<table>
<thead>
<tr>
<th>$h$</th>
<th>2$^{-1}$</th>
<th>2$^{-2}$</th>
<th>2$^{-3}$</th>
<th>2$^{-4}$</th>
<th>2$^{-5}$</th>
<th>2$^{-6}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|\bar{u} - \bar{u}<em>h|</em>{L_2(\Omega)}$</td>
<td>9.8501e-1</td>
<td>6.2440e-1</td>
<td>5.2082e-1</td>
<td>2.7114e-1</td>
<td>1.3591e-1</td>
<td>6.3854e-2</td>
</tr>
<tr>
<td>$|\bar{u}|_{L_2(\Omega)}$</td>
<td>9.8501e-1</td>
<td>6.2440e-1</td>
<td>5.2082e-1</td>
<td>2.7114e-1</td>
<td>1.3591e-1</td>
<td>6.3854e-2</td>
</tr>
<tr>
<td>order</td>
<td>0.66</td>
<td>0.26</td>
<td>0.94</td>
<td>1.00</td>
<td>1.09</td>
<td></td>
</tr>
</tbody>
</table>

**Example 6.2.2.** Set $\Omega = (0, 1)^2$, $\beta = 1$, and $\psi = 5 - ((x_1 - 0.5)^2 + (x_2 - 0.5)^2)$. The weights are $w_0 = 10^3$, $w_1 = 10^4$, and $w_2 = 10^5$.

We use the $P_2$ Lagrange finite element on uniform meshes to discretize the problem and the penalty parameter $\sigma = 100$. 

69
Figure 6.1. Graphs of optimal state and optimal control with $h = 2^{-5}$ for Example 6.2.1.

Figure 6.2. Active set with $h = 2^{-5}$ for Example 6.2.1.

The errors are computed by

$$e_\infty = \max_{p \in \mathcal{V}_h} |\bar{y}_{h/2}(p) - \bar{y}_h(p)|, \quad e_0 = \|\bar{y}_{h/2} - \bar{y}_h\|_{L^2(\Omega)},$$

$$e_1 = \|\bar{y}_{h/2} - \bar{y}_h\|_{H^1(\Omega)}, \quad e_h = \|\bar{y}_{h/2} - \bar{y}_h\|_h.$$

The figures of each trackings are shown in Figure 5.3.

- **Example 6.2.2.a** The target function is $y_2 = 4.5$ and the weight is $w_2 = 10^5$.

The convergence of the state in the given norms gets better than Corollary 5.6 as the mesh size gets smaller. The graphs for the optimal state, optimal control and active set are displayed in Figure 5.4 and Figure 5.5. The numerical results are reported in Ta-
Table 6.3. Error of the state versus mesh size $h$ and orders of convergence for Example 6.2.2.a.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e_\infty$</th>
<th>order</th>
<th>$e_0$</th>
<th>order</th>
<th>$e_1$</th>
<th>order</th>
<th>$e_h$</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^{-2}$</td>
<td>7.7360e-1</td>
<td></td>
<td>2.4626e-1</td>
<td></td>
<td>3.1891e0</td>
<td></td>
<td>7.2011e1</td>
<td></td>
</tr>
<tr>
<td>$2^{-3}$</td>
<td>7.4386e-1</td>
<td>0.06</td>
<td>1.8492e-1</td>
<td>0.41</td>
<td>2.7570e0</td>
<td>0.21</td>
<td>6.6700e1</td>
<td>0.11</td>
</tr>
<tr>
<td>$2^{-4}$</td>
<td>3.8736e-1</td>
<td>0.94</td>
<td>1.3651e-1</td>
<td>0.44</td>
<td>2.1843e0</td>
<td>0.34</td>
<td>5.2354e1</td>
<td>0.35</td>
</tr>
<tr>
<td>$2^{-5}$</td>
<td>2.3739e-1</td>
<td>0.71</td>
<td>8.0143e-2</td>
<td>0.77</td>
<td>1.3179e0</td>
<td>0.73</td>
<td>3.2764e1</td>
<td>0.68</td>
</tr>
<tr>
<td>$2^{-6}$</td>
<td>9.3902e-2</td>
<td>1.34</td>
<td>2.8588e-2</td>
<td>1.49</td>
<td>5.0711e-1</td>
<td>1.38</td>
<td>1.3834e1</td>
<td>1.24</td>
</tr>
<tr>
<td>$2^{-7}$</td>
<td>2.8186e-2</td>
<td>1.74</td>
<td>8.1316e-3</td>
<td>1.81</td>
<td>1.5362e-1</td>
<td>1.72</td>
<td>4.6234e0</td>
<td>1.58</td>
</tr>
</tbody>
</table>

**Example 6.2.2.b** The target function is $y_2 = 4.875$ and the weight is $w_2 = 10^5$.

The convergence of the state in the given norms is better than the convergence predicted by Corollary 5.6. The graphs for the optimal state, optimal control and active set are displayed in Figure 5.6 and Figure 5.7. The numerical results are reported in Table 5.4.
Figure 6.4. Graphs of optimal state and optimal control for Example 6.2.2.a.

Figure 6.5. Active set for Example 6.2.2.a.

Table 6.4. Error of the state versus mesh size $h$ and orders of convergence for Example 6.2.2.b.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e_\infty$</th>
<th>order</th>
<th>$e_0$</th>
<th>order</th>
<th>$e_1$</th>
<th>order</th>
<th>$e_h$</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^{-2}$</td>
<td>8.0969e-1</td>
<td></td>
<td>3.8160e-1</td>
<td></td>
<td>2.6850e0</td>
<td></td>
<td>4.1215e1</td>
<td></td>
</tr>
<tr>
<td>$2^{-3}$</td>
<td>1.7108e-1</td>
<td>2.24</td>
<td>1.0167e-1</td>
<td>1.91</td>
<td>8.8094e-1</td>
<td>1.61</td>
<td>1.6113e1</td>
<td>1.35</td>
</tr>
<tr>
<td>$2^{-4}$</td>
<td>8.0752e-2</td>
<td>1.08</td>
<td>3.5409e-2</td>
<td>1.52</td>
<td>4.5163e-1</td>
<td>0.96</td>
<td>8.0301e0</td>
<td>1.00</td>
</tr>
<tr>
<td>$2^{-5}$</td>
<td>3.3320e-2</td>
<td>1.28</td>
<td>1.5962e-2</td>
<td>1.15</td>
<td>2.2442e-1</td>
<td>1.01</td>
<td>3.8588e0</td>
<td>1.06</td>
</tr>
<tr>
<td>$2^{-6}$</td>
<td>1.3262e-2</td>
<td>1.33</td>
<td>5.8786e-3</td>
<td>1.44</td>
<td>8.4251e-2</td>
<td>1.41</td>
<td>1.5188e0</td>
<td>1.35</td>
</tr>
<tr>
<td>$2^{-7}$</td>
<td>3.9506e-3</td>
<td>1.75</td>
<td>1.6467e-3</td>
<td>1.84</td>
<td>2.4213e-2</td>
<td>1.80</td>
<td>4.7759e-1</td>
<td>1.67</td>
</tr>
</tbody>
</table>
Figure 6.6. Graphs of optimal state and optimal control for Example 6.2.2.b.

Figure 6.7. Active set for Example 6.2.2.b.
• **Example 6.2.2.c** The target functions are $y_0 = 4.875$, $y_2 = 4.875$ and the weights are $w_0 = 10^3$, $w_2 = 10^5$.

The convergence of the state in the given norms is better than the convergence predicted by Corollary 5.6. The graphs for the optimal state, optimal control and active set are displayed in Figure 5.8 and Figure 5.9. The numerical results are reported in Table 5.5.

Table 6.5. Error of the state versus mesh size $h$ and orders of convergence for Example 6.2.2.c.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e_\infty$</th>
<th>order</th>
<th>$e_0$</th>
<th>order</th>
<th>$e_1$</th>
<th>order</th>
<th>$e_h$</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^{-3}$</td>
<td>3.2945e-1</td>
<td>9.7720e-2</td>
<td>1.0086e0</td>
<td></td>
<td>2.2398e1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$2^{-4}$</td>
<td>1.7619e-1</td>
<td>0.90</td>
<td>4.4548e-2</td>
<td>1.13</td>
<td>5.5639e-1</td>
<td>0.86</td>
<td>1.1866e1</td>
<td>0.92</td>
</tr>
<tr>
<td>$2^{-5}$</td>
<td>8.5321e-2</td>
<td>1.05</td>
<td>2.2446e-2</td>
<td>0.99</td>
<td>3.1602e-1</td>
<td>0.82</td>
<td>6.7512e0</td>
<td>0.81</td>
</tr>
<tr>
<td>$2^{-6}$</td>
<td>3.4077e-2</td>
<td>1.32</td>
<td>8.8179e-3</td>
<td>1.35</td>
<td>1.4102e-1</td>
<td>1.61</td>
<td>3.5388e0</td>
<td>0.93</td>
</tr>
<tr>
<td>$2^{-7}$</td>
<td>1.1995e-2</td>
<td>1.51</td>
<td>2.7765e-3</td>
<td>1.67</td>
<td>4.8778e-2</td>
<td>1.53</td>
<td>1.6191e0</td>
<td>1.13</td>
</tr>
</tbody>
</table>

Figure 6.8. Graphs of optimal state and optimal control for Example 6.2.2.c.

• **Example 6.2.2.d** The target function is $y_1 = 4.875$ and the weight is $w_1 = 10^4$.

The convergence for the state in the $\| \cdot \|_\infty$ norm agrees with Theorem 5.3 where $\tau = 1 - \epsilon$. The convergence of the state in the other norms is better than the convergence predicted by Corollary 5.6. The graphs for the optimal state, optimal control and active
set are displayed in Figure 5.10 and Figure 5.11. The numerical results are reported in Table 5.6.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e_{\infty}$</th>
<th>order</th>
<th>$e_0$</th>
<th>order</th>
<th>$e_1$</th>
<th>order</th>
<th>$e_h$</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^{-2}$</td>
<td>1.4081e0</td>
<td>5.7257e-1</td>
<td>4.6188e0</td>
<td>4.9850e1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$2^{-3}$</td>
<td>2.7697e-1</td>
<td>1.1406e-1</td>
<td>1.0613e0</td>
<td>2.12</td>
<td>1.8972e1</td>
<td>1.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$2^{-4}$</td>
<td>1.7717e-1</td>
<td>5.0881e-2</td>
<td>5.5257e-1</td>
<td>0.94</td>
<td>8.7797e0</td>
<td>1.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$2^{-5}$</td>
<td>9.4905e-2</td>
<td>2.3143e-2</td>
<td>2.7791e-1</td>
<td>0.99</td>
<td>4.7698e0</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$2^{-6}$</td>
<td>5.2915e-2</td>
<td>1.0922e-2</td>
<td>1.2736e-1</td>
<td>1.13</td>
<td>2.3439e0</td>
<td>1.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$2^{-7}$</td>
<td>2.8603e-2</td>
<td>5.5587e-3</td>
<td>5.8891e-2</td>
<td>1.11</td>
<td>1.0540e0</td>
<td>1.15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example 6.2.3.** This example involves all three types of tracking functions, where the exact solution is not available.

Set $\Omega = (0,1)^2$, $\beta = 1$, and $\psi = 5 - ((x_1 - 0.5)^2 + (x_2 - 0.5)^2)$. The weights are $w_0(p_1) = 2 \times 10^3$, $w_0(p_2) = 5 \times 10^2$, $w_1 = 7 \times 10^2$, and $w_2 = 10^3$.

Tracking points are $p_1 = (0.375, 0.625)$, and $p_2 = (0.625, 0.625)$, tracking lines are connecting $(0.125, 0.125)$, $(0.875,0.125)$, $(0.875, 0.875)$, and $(0.125, 0.875)$, and tracking region is a rectangle with the vertices $(0.375, 0.25)$, $(0.625, 0.25)$, $(0.625, 0.375)$, and $(0.375, 0.375)$. The target functions are

$$ y_0(p_1) = 4.8, \quad y_0(p_2) = 4.6, \quad y_1 = 4.5, \quad \text{and} \quad y_2 = 4.7. $$

Note that for this example the pointwise state constraints are satisfied by the target functions $y_0$, $y_1$ and $y_2$. Therefore it can be interpreted as a least-squares data fitting problem.

We use the $P_2$ Lagrange finite element on uniform meshes to discretize the problem and the penalty parameter $\sigma = 100$. The numerical results are reported in Table 5.7 and
Figure 6.9. Active set for Example 6.2.2.c.

Figure 6.10. Graphs of optimal state and optimal control for Example 6.2.2.d.

Figure 6.11. Active set for Example 6.2.2.d.
Table 5.8, where the relative error of the state are defined by

\[ e_{\infty}^r = \max_{p \in \mathcal{V}_h} |\bar{y}_{h/2}(p) - \bar{y}_h(p)| / \max_{p \in \mathcal{V}_h} |y_{2^{-9}}(p)|, \quad e_0^r = \|\bar{y}_{h/2} - \bar{y}_h\|_{L_2(\Omega)} / \|y_{2^{-9}}\|_{L_2(\Omega)}, \]

\[ e_1^r = \|\bar{y}_{h/2} - \bar{y}_h\|_{H^1(\Omega)} / \|y_{2^{-9}}\|_{H^1(\Omega)}, \quad e_h^r = \|\bar{y}_{h/2} - \bar{y}_h\|_{h} / \|y_{2^{-9}}\|_{h}. \]

The convergence for the state in the \| \cdot \|_h norm (resp., the convergence of the control in the \( L_2 \) norm) agrees with Theorem 5.3 (resp., Corollary 5.5) where \( \tau = 1 - \epsilon \). The convergence of the state in the other norms is better than the convergence predicted by Corollary 5.6. The graphs for the optimal state, optimal control and active set at \( h = 2^{-7} \) are displayed in Figure 5.9 and Figure 5.10.

![Figure 6.12. Tracking for Example 6.2.3.](image)

Table 6.7. Relative error of the state versus mesh size \( h \) and orders of convergence for Example 6.2.3.

<table>
<thead>
<tr>
<th>( h )</th>
<th>( e_{\infty}^r )</th>
<th>order</th>
<th>( e_0^r )</th>
<th>order</th>
<th>( e_1^r )</th>
<th>order</th>
<th>( e_h^r )</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2^{-4} )</td>
<td>3.4311e-2</td>
<td>2.5090e-2</td>
<td>4.2391e-2</td>
<td>4.8085e-2</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 2^{-5} )</td>
<td>1.1588e-2</td>
<td>7.7981e-3</td>
<td>1.4364e-2</td>
<td>2.3055e-2</td>
<td>1.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 2^{-6} )</td>
<td>3.6465e-3</td>
<td>2.1994e-3</td>
<td>4.6165e-3</td>
<td>1.1124e-2</td>
<td>1.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 2^{-7} )</td>
<td>3.6465e-3</td>
<td>2.1994e-3</td>
<td>4.6165e-3</td>
<td>1.1124e-2</td>
<td>1.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 2^{-8} )</td>
<td>7.0256e-4</td>
<td>1.5988e-4</td>
<td>3.6717e-4</td>
<td>2.6061e-3</td>
<td>1.03</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6.8. Relative error of the control versus mesh size $h$ and orders of convergence for Example 6.2.3.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$2^{-3}$</th>
<th>$2^{-4}$</th>
<th>$2^{-5}$</th>
<th>$2^{-6}$</th>
<th>$2^{-7}$</th>
<th>$2^{-8}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|u_h/2 - u_h|_{L^2(\Omega)}$</td>
<td>2.6978e-1</td>
<td>1.4143e-1</td>
<td>7.0276e-2</td>
<td>3.4365e-2</td>
<td>1.6486e-2</td>
<td>8.0367e-3</td>
</tr>
<tr>
<td>$|u_{2^{-9}}|_{L^2(\Omega)}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>order</td>
<td>-</td>
<td>0.93</td>
<td>1.01</td>
<td>1.03</td>
<td>1.06</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Figure 6.13. Graphs of optimal state and optimal control with $h = 2^{-7}$ for Example 6.2.3.

Figure 6.14. Active set with $h = 2^{-7}$ for Example 6.2.3.
Chapter 7. Elliptic Optimal Control Problems with State and Control Constraints

We conduct numerical simulations for the elliptic optimal control problem with general tracking and both state and control constraints (cf. [22]).

7.1. Continuous Problem

We consider the following optimal control problem:

\[
\begin{aligned}
\text{Find } \quad (\bar{y}, \bar{u}) = \arg\min_{(y,u) \in \mathbb{K}} \left[ \frac{1}{2} \| y - y_d \|_{L^2(\Omega;\nu)}^2 + \frac{\beta}{2} \| u \|_{L^2(\Omega)}^2 \right],
\end{aligned}
\]

(7.1)

where \( \mathbb{K} \subset H^1_0(\Omega) \times L^2(\Omega) \) and \((y,u) \in H^1_0(\Omega) \times L^2(\Omega) \) belongs to \( \mathbb{K} \) if and only if

\[
\int_{\Omega} \nabla y \cdot \nabla z \, dx = \int_{\Omega} u z \, dx \quad \forall z \in H^1_0(\Omega),
\]

(7.2)

\[
\psi_- \leq y \leq \psi_+ \quad \text{a.e. in } \Omega,
\]

(7.3)

\[
\phi_- \leq u \leq \phi_+ \quad \text{a.e. in } \Omega.
\]

(7.4)

We assume that the functions \( \psi_- \), \( \psi_+ \), \( \phi_- \) and \( \phi_+ \) satisfy (i) \( \psi_- , \psi_+ \in W^{2,\infty}(\Omega) \cap H^3(\Omega) \), (ii) \( \psi_- < \psi_+ \) on \( \bar{\Omega} \), (iii) \( \psi_- < 0 < \psi_+ \) on \( \partial \Omega \), (iv) \( \phi_- , \phi_+ \in W^{1,\infty}(\Omega) \) and (v) \( \phi_- < \phi_+ \) on \( \bar{\Omega} \).

Since \( \Omega \) is convex, it follows from elliptic regularity and (7.2) that \( y \in H^2(\Omega) \cap H^1_0(\Omega) \) and hence the optimal control problem defined by (7.1)-(7.4) is equivalent to find

\[
\bar{y} = \arg\min_{y \in K} \left[ \frac{1}{2} \| y - y_d \|_{L^2(\Omega;\nu)}^2 + \frac{\beta}{2} \| \Delta y \|_{L^2(\Omega)}^2 \right],
\]

(7.5)

where

\[
K = \{ y \in H^2(\Omega) \cap H^1_0(\Omega) : \psi_- \leq y \leq \psi_+ \text{ and } \phi_- \leq -\Delta y \leq \phi_+ \text{ a.e. in } \Omega \}.
\]

(7.6)
We assume the following Slater condition:

There exists \( y \in H^2(\Omega) \cap H^1_0(\Omega) \) such that (i) \( \psi_\ast < y < \psi_\ast \) in \( \Omega \), and

\[
\text{(ii) } u = -\Delta y \text{ satisfies the constraint (7.4).}
\]

Then \( K \) is a nonempty, closed and convex subset of \( H^2(\Omega) \cap H^1_0(\Omega) \). According to Theorem 3.7, the reduced problem defined by (7.5)-(7.6) has a unique solution \( \bar{y} \) in \( K \). With the same argument in Section 3.2.2, the unique solution \( \bar{y} \) is characterized by the fourth order variational inequality

\[
\beta \int_{\Omega} (\Delta \bar{y})(\Delta y - \Delta \bar{y}) \, dx + \int_{\Omega} (\bar{y} - y_d)(y - \bar{y}) \, d\nu \geq 0 \quad \forall y \in K. \tag{7.8}
\]

7.2. Discrete Problem

Let \( V_h \subset H^2(\Omega) \cap H^1_0(\Omega) \) be the cubic Hermite finite element space associated with quasi-uniform triangulation \( T_h \) of \( \Omega \) (cf. \([39, 19]\)). Note that the cubic Hermite finite element space consists of piecewise cubic polynomial functions that are continuous up to first order derivatives at the vertices of \( T_h \). The degrees of freedom for the cubic Hermite element \( T \) (see Figure 7.1) consists of

(i) three vertices \( p_i \) of \( T \),
(ii) first order derivatives at three vertices of $T$,

(iii) the center $c$ of $T$.

To handle the control constraints (7.4), we introduce a modified Hermite finite element. Let $\varphi_i$ be the nodal basis function associated with $p_i$ of $T$ and $\varphi_T = \varphi_1 \varphi_2 \varphi_3$ be the cubic bubble function on $T$. According to Lemma 3.1. of [22], a cubic polynomial $v$ is uniquely determined by the following degrees of freedom:

(i) the values of $v$ at the three vertices $p_i$ of $T$,

(ii) the value of first order derivatives of $v$ at three vertices of $T$,

(iii) $\int_T (1 + \gamma \varphi_T)(\Delta v) \, dx$, where $\gamma$ is any nonnegative number.

We use $C^0$ interior penalty method to discrete the continuous problem (7.5)-(7.6).

By using (4.16), the discrete problem is to find

$$
\bar{y}_h = \arg\min_{y_h \in K_h} \frac{1}{2} \left[ \beta a_h(y_h, y_h) + \|y_h - y_d\|_{L^2(\Omega; \nu)}^2 \right],
$$

where

$$
K_h = \{ y_h \in V_h : I_h \psi_- \leq I_h y_h \leq I_h \psi_+ \text{ and } Q_h \varphi_- \leq Q_h (-\Delta_h y_h) \leq Q_h \varphi_+ \} \quad (7.10)
$$

with $\Delta_h$ is the piecewise defined Laplacian operator. Here, $Q_h$ is the $L_2$ projection operator onto the space of piecewise constant functions associated with $T_h$ defined by

$$(Q_h v)|_T := \frac{1}{|T|} \int_T v \, dx \quad \forall v \in L_2(\Omega), \ T \in T_h,$$

where $|T|$ is the area of the triangle $T$.

**7.3. Numerical Simulations**

Consider the following discrete problem:

$$
\min_{y_h \in K_h} \frac{1}{2} \left[ \sum_{j=1}^J w_0(p_j)[y_h(p_j) - y_0(p_j)]^2 + \sum_{l=1}^L \int_{\phi_l} w_1(y_h - y_1)^2 \, ds + \int_\Omega w_2(y_h - y_2)^2 \, dx + \beta a_h(y_h, y_h) \right],
$$
where
\[ K_h = \{ y_h \in V_h : I_h y_h \leq I_h \psi_+ \text{ and } Q_h ( - \Delta_h y_h ) \leq Q_h \phi_+ \text{ in } \Omega \}. \]

We set \( \Omega = (0, 1)^2 \), \( \beta = 1 \), \( \psi_+ = 10 \), and \( \phi_+ = 300 \). We use the modified Hermite finite element on uniform meshes to discretize the problem and the penalty parameter \( \sigma = 10^6 \). The discrete problem is solved by a primal-dual active set algorithm.

The relative errors of the state are computed by
\[
\begin{align*}
e_{\infty}^r &= \max_{p \in V_h} | \bar{y}_{h/2} (p) - \bar{y}_h (p) | / \max_{p \in V_h} | y_{2^{-7}} (p) |, \quad e_0^r = \| \bar{y}_{h/2} - \bar{y}_h \|_{L_2(\Omega)} / \| y_{2^{-7}} \|_{L_2(\Omega)}, \\
e_1^r &= \| \bar{y}_{h/2} - \bar{y}_h \|_{H^1(\Omega)} / \| y_{2^{-7}} \|_{H^1(\Omega)}, \quad e_h^r = \| \bar{y}_{h/2} - \bar{y}_h \|_h / \| y_{2^{-7}} \|_h.
\end{align*}
\]

The figures of each trackings are shown in Figure 6.2.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.2}
\caption{Tracking for Example 6.3.1 and Example 6.3.2}
\end{figure}

- **Example 6.3.1.** Tracking points are \( p_1 = (0.25, 0.75) \), and \( p_2 = (0.75, 0.75) \). The weights are \( w_0 (p_1) = w_0 (p_2) = 800 \), and the desired states are \( y_0 (p_1) = y_0 (p_2) = 10 \).

The graphs for the optimal state, optimal control and active set at \( h = 2^{-5} \) are displayed in Figure 6.3 and Figure 6.4. According to the results in Table 6.1 and Table 6.2, the convergence of state and control behaves similar to the case where only state constraints exist.

- **Example 6.3.2.** Tracking points are \( p_1 = (0.25, 0.75) \), and \( p_2 = (0.75, 0.75) \),
Table 7.1. Relative error of the state versus mesh size $h$ and orders of convergence for Example 6.3.1.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e_{\infty}^r$</th>
<th>order</th>
<th>$e_0^r$</th>
<th>order</th>
<th>$e_1^r$</th>
<th>order</th>
<th>$e_h^r$</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^{-2}$</td>
<td>9.1115e-2</td>
<td>2.00</td>
<td>8.0266e-2</td>
<td>2.00</td>
<td>1.9256e-1</td>
<td>2.00</td>
<td>6.5876e-1</td>
<td>2.00</td>
</tr>
<tr>
<td>$2^{-3}$</td>
<td>4.4082e-2</td>
<td>1.05</td>
<td>2.4919e-2</td>
<td>1.05</td>
<td>7.3026e-2</td>
<td>1.05</td>
<td>4.1113e-1</td>
<td>1.05</td>
</tr>
<tr>
<td>$2^{-4}$</td>
<td>6.6635e-3</td>
<td>2.73</td>
<td>3.2081e-3</td>
<td>2.73</td>
<td>1.5849e-2</td>
<td>2.73</td>
<td>1.8685e-1</td>
<td>2.73</td>
</tr>
<tr>
<td>$2^{-5}$</td>
<td>2.1942e-3</td>
<td>1.60</td>
<td>9.9091e-4</td>
<td>1.60</td>
<td>3.7818e-3</td>
<td>1.60</td>
<td>8.9186e-2</td>
<td>1.60</td>
</tr>
<tr>
<td>$2^{-6}$</td>
<td>1.0608e-3</td>
<td>1.05</td>
<td>8.4763e-4</td>
<td>1.05</td>
<td>1.3493e-3</td>
<td>1.05</td>
<td>4.2645e-2</td>
<td>1.05</td>
</tr>
</tbody>
</table>

Table 7.2. Relative error of the control versus mesh size $h$ and orders of convergence for Example 6.3.1.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$|u_{h/2} - u_h|_{L^2(\Omega)}$</th>
<th>$|u_{2^{-7}}|_{L^2(\Omega)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^{-2}$</td>
<td>6.6874e-1</td>
<td>2.00</td>
</tr>
<tr>
<td>$2^{-3}$</td>
<td>4.1736e-1</td>
<td>1.14</td>
</tr>
<tr>
<td>$2^{-4}$</td>
<td>1.8968e-1</td>
<td>1.07</td>
</tr>
<tr>
<td>$2^{-5}$</td>
<td>9.0539e-2</td>
<td>1.07</td>
</tr>
<tr>
<td>$2^{-6}$</td>
<td>4.3292e-2</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Figure 7.3. Graphs of optimal state and optimal control with $h = 2^{-5}$ for Example 6.3.1.

Figure 7.4. Active set with $h = 2^{-5}$ for Example 6.3.1.
tracking lines are connecting (0.5, 0.5) and (0.5, 0.75), and tracking region is a rectangle
with the vertices (0.25, 0.25), (0.75, 0.25), (0.75, 0.5), and (0.25, 0.5). The weights are
\( w_0(p_1) = w_0(p_2) = 500 \), \( w_1 = 200 \), and \( w_2 = 100 \), and the desired states are \( y_0(p_1) = y_0(p_2) = 10 \), \( y_1 = 10 \), and \( y_2 = 10 \).

The graphs for the optimal state, optimal control and active set at \( h = 2^{-5} \) are
displayed in Figure 6.5 and Figure 6.6. According to the results in Table 6.3 and Table
6.4, the convergence of state and control behaves similar to the case where only state con-
straints exist.

Table 7.3. Relative error of the state versus mesh size \( h \) and orders of convergence for
Example 6.3.2

<table>
<thead>
<tr>
<th>( h )</th>
<th>( e_\infty )</th>
<th>order</th>
<th>( e_0^r )</th>
<th>order</th>
<th>( e_1^r )</th>
<th>order</th>
<th>( e_2^r )</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2^{-2} )</td>
<td>7.0966e-2</td>
<td></td>
<td>6.0871e-2</td>
<td></td>
<td>9.3068e-2</td>
<td></td>
<td>3.4885e-1</td>
<td></td>
</tr>
<tr>
<td>( 2^{-3} )</td>
<td>1.1470e-2</td>
<td>2.63</td>
<td>7.1798e-3</td>
<td>3.08</td>
<td>2.5356e-2</td>
<td>1.88</td>
<td>1.6966e-1</td>
<td>1.04</td>
</tr>
<tr>
<td>( 2^{-4} )</td>
<td>2.6846e-3</td>
<td>2.10</td>
<td>1.5911e-3</td>
<td>2.17</td>
<td>5.5579e-3</td>
<td>2.19</td>
<td>8.3186e-2</td>
<td>1.03</td>
</tr>
<tr>
<td>( 2^{-5} )</td>
<td>6.4116e-3</td>
<td>2.07</td>
<td>3.3309e-4</td>
<td>2.26</td>
<td>1.2713e-3</td>
<td>2.13</td>
<td>3.9616e-2</td>
<td>1.07</td>
</tr>
<tr>
<td>( 2^{-6} )</td>
<td>2.3468e-4</td>
<td>1.45</td>
<td>1.9565e-4</td>
<td>0.77</td>
<td>3.7347e-4</td>
<td>1.77</td>
<td>1.9169e-2</td>
<td>1.05</td>
</tr>
</tbody>
</table>

Table 7.4. Relative error of the control versus mesh size \( h \) and orders of convergence for
Example 6.3.2

<table>
<thead>
<tr>
<th>( h )</th>
<th>( | u_{h/2} - u_h |_{L^2(\Omega)} )</th>
<th>( | u_{2^{-7}} |_{L^2(\Omega)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2^{-2} )</td>
<td>3.5586e-1</td>
<td>1.7307e-1</td>
</tr>
<tr>
<td>( 2^{-3} )</td>
<td>8.4856e-2</td>
<td>4.0414e-2</td>
</tr>
<tr>
<td>( 2^{-4} )</td>
<td>1.9555e-2</td>
<td>1.05</td>
</tr>
<tr>
<td>( 2^{-5} )</td>
<td>1.04</td>
<td>1.03</td>
</tr>
<tr>
<td>( 2^{-6} )</td>
<td>1.07</td>
<td>1.05</td>
</tr>
</tbody>
</table>

84
Figure 7.5. Graphs of optimal state and optimal control with $h = 2^{-5}$ for Example 6.3.2

Figure 7.6. Active set with $h = 2^{-5}$ for Example 6.3.2
Chapter 8. Conclusion and Future Work

In this dissertation, elliptic optimal control problems with pointwise constraints are considered. The cost functional tracks points, curves and regions of the domain in two spatial dimensions. The optimal control problem is to find

$$(\bar{y}, \bar{u}) = \arg\min_{(y,u) \in K} \frac{1}{2} \left[ G(y) + \beta \|u\|_{L_2(\Omega)}^2 \right],$$

where

$$G(y) = \sum_{j=1}^{J} (y(p_j) - y_0(p_j))^2 w_0(p_j) + \sum_{l=1}^{L} \int_{\phi_l} (y - y_1)^2 w_1 ds + \int_{\Omega} (y - y_2)^2 w_2 dx,$$

$K \subset H^1_0(\Omega) \times L_2(\Omega)$, and $(y, u) \in H^1_0(\Omega) \times L_2(\Omega)$ belongs to $K$ if and only if

$$\int_{\Omega} \nabla y \cdot \nabla z dx = \int_{\Omega} uz dx \quad \forall z \in H^1_0(\Omega),$$

$$\psi_- \leq y \leq \psi_+ \quad \text{a.e. in } \Omega.$$ 

The optimal control problem has been reduced into the minimization problem which involves only the state variable $y$:

Find $$\bar{y} = \arg\min_{y \in K} \frac{1}{2} \left[ \beta \|\Delta y\|_{L_2(\Omega)}^2 + \|y - y_d\|_{L_2(\Omega;\nu)}^2 \right]$$,

where

$$K = \{ y \in H^2(\Omega) \cap H^1_0(\Omega) : \psi_- \leq y \leq \psi_+ \text{ in } \Omega \}.$$ 

The reduced minimization problem has a unique solution $\bar{y} \in K$ characterized by the fourth order variational inequality

$$\beta \int_{\Omega} (\Delta \bar{y})(\Delta y - \Delta \bar{y}) dx + \int_{\Omega} (\bar{y} - y_d)(y - \bar{y}) d\nu \geq 0 \quad \forall y \in K,$$
which is equivalent to the generalized Karush-Kuhn-Tucker (KKT) conditions:

$$
\beta \int_\Omega (\Delta \bar{y})(\Delta z) \, dx + \int_\Omega (\bar{y} - y_d) z \, d\nu = \int_\Omega z \, d\mu \quad \forall \ z \in H^2(\Omega) \cap H^1_0(\Omega),
$$

where $\mu$ is a regular Borel measure, such that

$$
\mu \geq 0 \quad \text{if} \quad \bar{y} = \psi_-, \\
\mu \leq 0 \quad \text{if} \quad \bar{y} = \psi_+, \\
\mu = 0 \quad \text{otherwise}.
$$

We have the global regularity for the state

$$\bar{y} \in H^{2+\alpha}(\Omega),$$

for some $\alpha \in (0, 1)$, where the index of elliptic regularity $\alpha$ is determined by $\Omega$. For a rectangular domain,

$$\alpha = 1 - \epsilon, \quad \epsilon > 0.$$

The quadratic $C^0$ interior penalty method has been applied to discretize the optimal control problem.

The discrete problem is to find

$$\bar{y}_h = \arg\min_{y_h \in K_h} \frac{1}{2} \left[ \beta a_h(y_h, y_h) + \|y_h - y_d\|_{L^2(\Omega, \nu)}^2 \right],$$

where

$$K_h = \{ y_h \in V_h : I_h \psi_- \leq I_h y_h \leq I_h \psi_+ \},$$

$$a_h(y_h, z_h) = \sum_{T \in T_h} \int_T D^2 y_h : D^2 z_h \, dx + \sum_{e \in E^i_h} \sigma h_e^{-1} \int_e \begin{bmatrix} \partial y_h / \partial n \end{bmatrix} \begin{bmatrix} \partial z_h / \partial n \end{bmatrix} \, ds$$

$$+ \sum_{e \in E^i_h} \int_e \begin{bmatrix} \{ \partial^2 y_h / \partial n^2 \} \{ \partial^2 z_h / \partial n^2 \} + \{ \partial^2 z_h / \partial n^2 \} \{ \partial y_h / \partial n \} \end{bmatrix} \, ds,$$

87
$h_e$ is the diameter of the edge $e$, and $\sigma > 0$ is a penalty parameter.

We have the following estimate for the optimal state:

$$\| \tilde{y} - \bar{y}_h \|_h \leq C h^\tau,$$

where $\tau$ is given by

$$\tau = \begin{cases} 
\alpha & \text{if } \mathcal{T}_h \text{ is quasi-uniform} \\
1 - \epsilon & \text{if } \mathcal{T}_h \text{ is graded around the corners of } \Omega \\
& \text{where the interior angles are } > (\pi/2)
\end{cases}$$

Here $\alpha$ is the index of elliptic regularity and $\epsilon$ can be any positive number.

We have performed numerical experiments on rectangular domains with uniform meshes which is sufficient to guarantee $O(h^{1-\epsilon})$ estimates for the errors. The $O(h^{1-\epsilon})$ estimates are also valid for a general convex domain provided we use local mesh refinements around the corners where the interior angles are strictly larger than $\pi/2$.

One can consider an analogous optimal control problem on three dimensional domains, where the cost function also includes tracking on surfaces inside the domain in addition to tracking on points, curves and regions. In this case the singularity caused by point-tracking is more severe than the two dimensional case. In this case the regularity estimate (3.41) for the adjoint state $\bar{p}$ becomes (cf. [33, Theorem 1])

$$\bar{p} \in W_0^{1,s}(\Omega) \quad \forall \ s < \frac{3}{2},$$

and hence the regularity estimate (3.44) for the optimal state $\tilde{y}$ becomes

$$\tilde{y} \in W^{3,s}_{loc}(\Omega) \quad \forall \ s < \frac{3}{2}.$$

88
Consequently the error estimates on uniform meshes are $O(h^{1/2-\epsilon})$ even for a rectangular parallelepiped. However, since the deterioration of the elliptic regularity is due to the existence of point tracking at known positions, one can include these points in $\mathcal{V}_h$ and use local mesh refinement around them (and around $\partial \Omega$ for a general convex $\Omega$) to recover $O(h^{1-\epsilon})$ error estimates.

We can extend the optimal control problem to the ones where both state and control constraints exist [22]. In this case, we would apply the cubic bubble function to the local basis function. One can show the discretization error due to the control constraints and the state constraints and establish an error estimate. The key is the estimates for the Dirac measure.
Bibliography


[64] C. Meyer, A. Rösch, and F. Tröltzsch. Optimal control of PDEs with regularized

94


Vita

SeongHee Jeong was born in 1989, in Bucheon, Gyeonggi-do, South Korea. She received the degree of Bachelor of Science in Mathematics, with minor in Economics at Sookmyung Women’s University in February 2013. She earned the degree of Master of Science in Computational Mathematics from Yonsei University in February 2016. In August 2017 she came to Louisiana State University to pursue graduate studies in Mathematics. She is currently a candidate for the degree of Doctor of Philosophy in Mathematics, which will be awarded in May 2023.