JOHANNES KEPLER
UNIVERSITÄT LINZ
Netzwerk für Forschung, Lehre und Praxis

TNF

Technisch-Naturwissenschaftliche
Fakultät

Generalized Penalty Methods for Elliptic Neumann Boundary Control Problem with State and Control Constraints

MASTERARBERIT

zur Erlangung des akademischen Grades

DIPLOM-INGENIEUR

in der Studienrichtung

INDUSTRIAL MATHEMATICS

Angefertigt am *Institut für Numerische Mathematik*

Betreuung:

a. Univ. Prof. Dipl. Ing. Dr. Helmut Gfrerer

Eingereicht von:

Esubalewe Lakie Yedeg

Linz, July 2010

Dedicated to
Tigist Muluken and
Wubit Aragaw

Abstract

In this work we studied an optimal control problem constrained with boundary control. It is an infinite dimensional convex optimization problem that consist of minimizing a cost function subject to pointwise state and control constraints and governed by elliptic differential equation with a Neumann boundary condition. The control being distributed only on the boundary. A generalized penalty function approach is then used to reformulate the original constrained problem as an unconstrained problem. The convergence results and the error estimates of the penalty method are stated. To solve the resulting subproblems numerically we used Newton's method with line search in function spaces and its superlinear convergence is presented. The finite element method is used to discretize the subproblems and to transform into finite dimensional problems. Finally, numerical examples are given to illustrate the theoretical results.

Acknowledgement

First of all, I would like to thank my advisor Professor Helmut Gfrerer, for supervising my thesis, and the countless discussions, guidance and inspirations throughout this work. I owe my deepest gratitude to him for making the thesis possible.

I would like to take this moment to thank my program coordinates Dr. Ewald Lindner (Johannes Kepler University, Linz) and Dr. Martijn Anthonissen (Eindhoven University of Technology) for their coordination and encouragement in my study. I am very grateful to my Lecturers at both universities for providing the necessary knowledge during my study.

My dearest thanks go to Tigist Muluken and my family for their love, care and support. I would not forget my friends for their mental and physical support.

Last, but not least, I would like to acknowledge the European Union for financing my study and stay in Europe for two years through the Erasmus Mundus Scholarship.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Partial differential equation constrained optimization problem is an optimization of systems controlled by partial differential equations (PDEs) as constraints. PDE-constrained optimization problem is a recently emerging research area and it arises in the description of science and engineering applications including optimal control, design and parameter estimation [3]. Some examples of the application of PDE-constrained optimization problem arises in aerodynamics, mathematical finance, medicine and environmental engineering [5, 6, 7, 8]. In many situations, PDE-constrained optimization problems take the form of optimal control or optimal design problems. Solving optimization problems subject to partial differential equations with additional constraints is one of the challenging problems in scientific applications.

In this work we study an optimal control problem governed by Elliptic PDE with additional constraints on control and state variables. An optimal control problem of such a system involves finding an optimal control and optimal state which satisfy the Elliptic PDE and the additional constraints such that the given cost function is minimized. It is one of the infinite dimensional optimization problems.

The optimization problem we are considering has the following general form

$$
\begin{aligned}
(P) \quad & \min_{z \in Z} J(z) \\
& \text{subject to} \\
& \qquad Ez = 0, \\
& \qquad g_i(z) \leq \varphi_i, \ \ \mu_i - \text{a.e. in } \Omega_i, \ i = 1, \ldots, m
\end{aligned}
$$

where the cost function $J : Z \to \mathbb{R}$ is defined on a Hilbert space Z and the operator $E \in \mathcal{L}(Z, V)$ is a bounded linear operator from Z into another Banach space V. Further we have finitely many pointwise inequality constraints defined by the mappings $g_i : Z \to L^{r_i}(\Omega_i)$ and bounds $\varphi_i \in L^{r_i}(\Omega_i), \ i = 1, \ldots, m$, where for each i, we have $r_i \geq 1$ and Ω_i denotes a finite measure space $(\Omega_i, \mathcal{A}_i, \mu_i)$ given by the set Ω_i, σ-algebra $\mathcal{A}_i \subset \mathcal{P}(\Omega_i)$ of subsets of Ω_i and a finite measure μ_i on $\mathcal{A}_i$.

We assume the following basic assumptions for optimization problems:

Assumption:

(P1) There exist a feasible point z_f for the feasible set Z_F of (P), where $Z_F := \{z \in Z : Ez = 0, g_i(z) \leq \varphi_i, \ \mu_i - \text{ a.e. } \Omega_i, \ i = 1 : m\}$.

(P2) The cost function J is convex and lower semi-continuous (l.s.c.) on Z_E, where $Z_E := \{z \in Z : Ez = 0\}$ is the kernel of E.

(P3) There exists some positive real number α such that for all $z_1, z_2 \in Z_E$ we have

$$J(z_2) \geq J(z_1) + J'(z_1, z_2 - z_1) + \frac{\alpha}{2}\|z_2 - z_1\|^2.$$

(P4) For each $i = 1, \ldots, m$ and for each closed set $U \subset L^{r_i}(\Omega)$, the set $\{z \in Z_E : g_i(z) \in U\}$ is closed. Further, $\forall z_1, z_2 \in Z_E$ and all $t \in [0, 1]$ we have

$$t g_i(z_1)(x) + (1 - t) g_i(z_2)(x) \geq g_i(t z_1 + (1 - t) z_2)(x) \text{ for almost all } x \in \Omega_i.$$

Here, $J'(z_1, z_2 - z_1)$ is the directional derivative of J at z_1 in the direction of $z_2 - z_1$. Since we assume that J is convex and lower semi-continuous, its directional derivative is well defined.

Note that the assumptions (P1)–(P4) guarantee the existence of a unique solution $\bar{z}$ for the optimization problem (P).

The analysis of the abstract problem (P) is studied by Gfrerer in [1]. He used the generalized penalty method to solve the problem numerically and presented the numerical results for elliptic PDE with Dirichlet boundary constrained optimal control problem.

The main objective of this work is to solve the optimal control problem governed by elliptic Neumann boundary PDE with additional constraints on the control and the state variables. We use the results of Gfrerer [1] to solve this model problem. The generalized penalty method is applied to the model problem to find the optimal control and state which minimize the objective function by transforming the constrained optimal control problem into a sequence of unconstrained optimization problem.

The remaining of this work is organized as follows. In chapter 2 the nature and the existence of unique solution of the model problem are presented. The generalized penalty method for the model problem and its convergence are introduced in chapter 3. The error estimate of the method is presented in chapter 4. The dual problem of the model problem, the Lagrangian multiplier and its approximate by the penalty method are introduced in chapter 5. A Newton-type algorithm to solve the subproblems is introduced in chapter 6. In chapter 7 we used finite element method to discretize the problem. Finally, the numerical results are presented in chapter 8.

Chapter 2

Model Problem

This chapter deals with the nature of the model problem. The model problem is an optimal control problem subject to elliptic partial differential equation with Neumann boundary condition and additional constraints on the control and the state variables. Its formulation has the following terms:

- The definition of the control function u that represents the driving force which controls the environment of the system. It is defined only on the boundary.
- The elliptic partial differential equation with Neumann boundary condition modeling the controlled system, represented by the state function y.
- The cost function J which models the cost of the control and the state. It is a function of state and control variables.

In this work we assume the state variable y is defined on the open and bounded Lipschitz domain $\Omega = (0,1)^2 \subset \mathbb{R}^2$ and the control on the boundary $\partial\Omega$. The domain $\Omega = (0,1)^2$ and its boundary are shown shown on Figure 2.1.

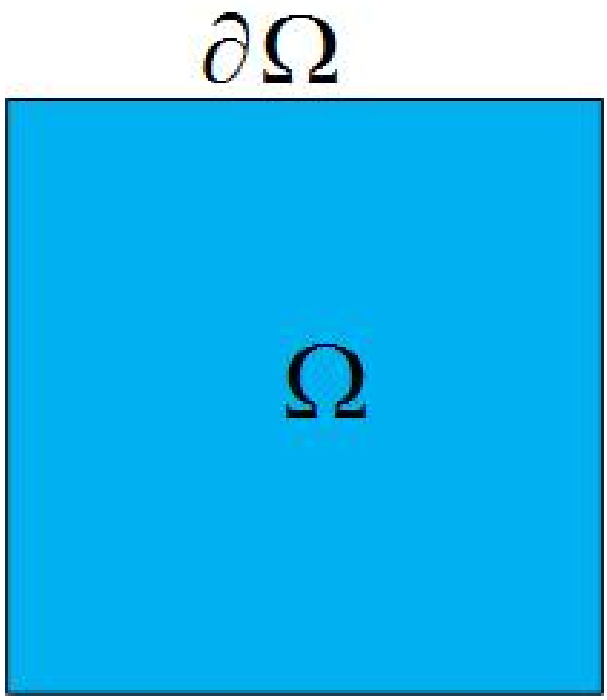

Figure 2.1: Model Domain Ω and its boundary $\partial\Omega$

The model problem we are considering has the following structure

$$(MP) \qquad \min_{y\in H^1(\Omega), u\in L^2(\partial\Omega)} J(z) := \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|u - u_d\|^2_{L^2(\partial\Omega)} \tag{2.1a}$$

subject to

$$-\Delta y + cy = r \text{ in } \Omega, \tag{2.1b}$$

$$\frac{\partial y}{\partial \nu} = u \text{ on } \partial\Omega, \tag{2.1c}$$

$$u \leq u_b \text{ on } \partial\Omega, \tag{2.1d}$$

$$y \leq y_b \text{ in } \Omega. \tag{2.1e}$$

where

- $\beta \in \mathbb{R},\ \beta > 0,$

- $c \in L^\infty(\Omega),\ c > 0$,
- $u_b \in L^2(\partial\Omega)$ is an upper bound on the control,
- $y_b \in L^2(\Omega)$ is an upper bound on the state,
- $r \in H^1(\Omega)^*$ is a given source term.

Next we present the existence of a unique weak solution to the the state equation for each control $u \in L^2(\partial\Omega)$. To show this we use the Lax-Milgram lemma.

Theorem 2.1 ***(Lax-Milgram).*** *Let Z be a Hilbert space and let*

$$a : Z \times Z \to \mathbb{R}$$

be a bilinear form. Suppose there exists some positive constants C_1 and C_2 such that

$$|a(x, y)| \leq C_1 \|x\|_Z \|y\|_Z, \ \text{for all } x \text{ and } y \in Z \tag{2.2}$$

and

$$a(x, x) \geq C_2 \|x\|_Z^2, \ \text{for all } x \in Z. \tag{2.3}$$

Then for every $F \in Z^*$ there exists a unique $y \in Z$ such that

$$a(x, y) = F(x) \text{ for all } x \in Z.$$

Furthermore, it satisfies an estimate

$$\|y\|_Z \leq \frac{1}{C_2} \|F\|_{Z^*}$$

for C_2 from (2.3).

Let us assume y satisfies the state equation (2.1b) including the boundary condition (2.1c). Multiplying both sides of (2.1b) by $v \in C^1(\bar{\Omega})$ and integrating over Ω yields

$$-\int_\Omega (\Delta y v + cyv) dx = \int_\Omega rv dx \ \ \forall v \in C^1(\bar{\Omega}). \tag{2.4}$$

By applying Green's formula on (2.4), we have

$$\int_{\Omega}(\nabla y \cdot \nabla v + cyv)dx - \int_{\partial\Omega}\frac{\partial y}{\partial \nu}vd\mu x = \int_{\Omega} rvdx \;\; \forall v \in C^1(\bar{\Omega}). \tag{2.5}$$

Then, substitution of the Neumann boundary condition (2.1c) in (2.5) leads to the following equation

$$\int_{\Omega}(\nabla y \cdot \nabla v + cyv)dx - \int_{\partial\Omega} uvd\mu x = \int_{\Omega} rvdx \;\; \forall v \in C^1(\bar{\Omega}). \tag{2.6}$$

Since $C^1(\bar{\Omega})$ is dense in $H^1(\Omega)$ the variational equation (2.6) makes sense in $H^1(\Omega)$. Thus the variational (weak) formulation of the state equation including boundary condition is

$$\int_{\Omega}(\nabla y \cdot \nabla v + cyv)dx - \int_{\partial\Omega} uvd\mu x = \int_{\Omega} rvdx \;\; \forall v \in H^1(\Omega). \tag{2.7}$$

Now let us introduce the following abstract notations

$$a(y,u) = \int_{\Omega}(\nabla y \cdot \nabla v + cyv)dx \;\; \forall y, v \in H^1(\Omega) \tag{2.8}$$

and

$$F(v) = \int_{\Omega} rvdx + \int_{\partial\Omega} uvd\mu x \;\; \forall v \in H^1(\Omega). \tag{2.9}$$

Then the variational equation (2.7) can be written as:

$$\text{Find } y \in H^1(\Omega) : \; a(y,v) = F(v) \;\; v \in H^1(\Omega)$$

where

$$a : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R} \text{ is a bilinear form and}$$

$$F : H^1(\Omega) \to \mathbb{R} \text{ is a linear functional.}$$

In operator form the variational equation (2.7) can be written as

$$Ay = r + Bu$$

where,

$$\begin{aligned} A \in \mathcal{L}(H^1(\Omega), H^1(\Omega)^*) \text{ defined by} & \\ \langle Ay, v\rangle_{H^1(\Omega)^*, H^1(\Omega)} &= \int_\Omega (\nabla y \cdot \nabla v + cyv)dx, \\ B \in \mathcal{L}(L^2(\partial\Omega), H^1(\Omega)^*) \text{ defined by} & \\ \langle Bu, v\rangle_{H^1(\Omega)^*, H^1(\Omega)} &= \int_{\partial\Omega} uvd_s(x). \end{aligned}$$

This shows that the optimal control problem (MP) can be written as a linear quadratic optimization problem of the form

$$\begin{aligned} \min_{y\in H^1(\Omega), u\in L^2(\partial\Omega)} \quad & J(z) := \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \tfrac{\beta}{2}\|u - u_d\|^2_{L^2(\partial\Omega)} \\ \text{subject to} & \\ & Ay - Bu = r \text{ in } \Omega, \\ & u \le u_b \text{ on } \partial\Omega \\ & y \le y_b \text{ in } \Omega. \end{aligned} \tag{2.10}$$

Hereafter, without lose of generality we assume that $r = 0$.

Now let us denote the spaces and the obstacles in problem (2.10) by

$$\begin{aligned} &Z = H^1(\Omega) \times L^2(\partial\Omega), \\ &V = H^1(\Omega)^* \\ &\varphi_1 = u_b \in L^2(\partial\Omega) \\ &\varphi_2 = y_b \in L^2(\Omega) \end{aligned}$$

and define the functions $g_1 : Z \to L^2(\partial\Omega)$, $g_2 : Z \to H^1(\Omega)$ and the linear operator $E : Z \to V$ by

$$\begin{aligned} g_1(y, u) &= u, \\ g_2(y, u) &= y \\ \langle E(y, u), v\rangle &= \langle Ay - Bu, v\rangle \\ &= \int_\Omega (\nabla y \cdot \nabla v + cyv)dx - \int_{\partial\Omega} uvd_s(x). \end{aligned} \tag{2.11}$$

It is known that the cartesian product of two Hilbert spaces is also a Hilbert space. Hence, $Z = H^1(\Omega) \times L^2(\partial\Omega)$ is a Hilbert space with inner product

$$\langle (y_1, u_1), (y_2, u_2)\rangle_{Z\times Z} = \langle y_1, y_2\rangle_{H^1(\Omega)\times H^1(\Omega)} + \langle u_1, u_2\rangle_{L^2(\partial\Omega)\times L^2(\partial\Omega)}$$

which induces the following norm in Z

$$\|(y, u)\|_Z = (\|y\|^2_{H^1(\Omega)} + \|u\|^2_{L^2(\partial\Omega)})^{1/2}.$$

Note that $V = H^1(\Omega)^*$ is a Hilbert space and the functions g_1 and g_2 are the projections of the space Z into the spaces $L^2(\partial\Omega)$ and $H^1(\Omega)$ respectively.

Theorem 2.2 *Let $c \in \mathcal{L}^\infty(\Omega)$, $c > 0$. Then the mapping $E : Z \to V$ defined by (2.12) is bounded.*

Proof For any $(y, u) \in Z$ and $v \in H^1(\Omega)$ we have

$$\begin{aligned}
|\langle E(y,u), v\rangle| &= |\int_\Omega \nabla y \cdot \nabla v + cy\, v dx - \int_{\partial\Omega} uv d\mu(x)| \\
&\le \int_\Omega |\nabla y \cdot \nabla v| dx + \int_\Omega |cy\, v| dx + \int_{\partial\Omega} |uv| d\mu(x)| \\
&\le \Big(\int_\Omega |\nabla y|^2 dx\Big)^{1/2}\Big(\int_\Omega |\nabla v|^2 dx\Big)^{1/2} + \|c\|_{L^\infty}\Big(\int_\Omega |y|^2 dx\Big)^{1/2}\Big(\int_\Omega |v|^2 dx\Big)^{1/2} \\
&\qquad + \Big(\int_{\partial\Omega} |u|^2 dx\Big)^{1/2}\Big(\int_{\partial\Omega} |v|^2 dx\Big)^{1/2} \\
&\qquad \text{(by Cauchy-Schwartz inequality)} \\
&\le \|\nabla y\|_{L^2(\Omega)}\|\nabla v\|_{L^2(\Omega)} + \|c\|_{L^\infty}\|y\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)} + \|u\|_{L^2(\partial\Omega)}\|v\|_{L^2(\partial\Omega)} \\
&\le \Big(\|y\|_{H^1(\Omega)} + \|c\|_{L^\infty}\|y\|_{H^1(\Omega)} + \|u\|_{L^2(\partial\Omega)}\Big)\|v\|_{H^1(\Omega)} \\
&\le (1 + \|c\|_{L^\infty})\|(y,u)\|_Z\|v\|_{H^1(\Omega)} \\
&< \infty.
\end{aligned}$$

Hence, $E : Z \to V$ is bounded.

Thus problem (2.10) becomes

$$\begin{aligned}
\min_{y\in H^1(\Omega), u\in L^2(\partial\Omega)} \quad & J(z) := \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \tfrac{\beta}{2}\|u - u_d\|^2_{L^2(\partial\Omega)} \\
\text{subject to} \quad & \\
& E(y,u) = 0 \text{ in } \Omega, \\
& u \le u_b \text{ on } \partial\Omega \\
& y \le y_b \text{ in } \Omega.
\end{aligned} \tag{2.12}$$

Denote the subspace Z_E of Z by $Z_E = \ker E$, and define the feasible set Z_F of problem (2.12) by

$$Z_F := \{z \in Z : Ez = 0, g_1(z)(x) \le \varphi_1(x), \text{ a.e. on } \partial\Omega, g_2(z)(x) \le \varphi_2(x), \text{ a.e. in } \Omega\}.$$

Note that, the set Z_E is closed and convex subset of Z and Z_F is a convex subset of Z.

Assumption 2.1

There exists a feasible point $z_f \in Z_E$, i.e. the model problem has non empty constraint set.

Theorem 2.3 *The cost function J is convex on Z_E.*

Proof Let $z_1 = (y_1, u_1)$, $z_2 = (y_2, u_2) \in Z_E$ and $\lambda \in [0, 1]$, we have

$$\begin{aligned}
J(\lambda z_1 + (1-\lambda)z_2) &= J(\lambda(y_1, u_1) + (1-\lambda)(y_2, u_2)) \\
&= J(\lambda y_1 + (1-\lambda)y_2, \lambda u_1 + (1-\lambda)u_2) \\
&= \frac{1}{2}\|\lambda y_1 + (1-\lambda)y_2 - y_d\|_{L^2(\Omega)} + \frac{\alpha}{2}\|\lambda u_1 + (1-\lambda)u_2 - u_d\|_{L^2(\partial\Omega)} \\
&= \frac{1}{2}\|\lambda(y_1 - y_d) + (1-\lambda)(y_2 - y_d)\|_{L^2(\Omega)} + \\
&\qquad \frac{\alpha}{2}\|\lambda(u_1 - u_d) + (1-\lambda)(u_2 - u_d)\|_{L^2(\partial\Omega)} \\
&\leq \lambda(\frac{1}{2}\|y_1 - y_d\|_{L^2(\Omega)} + \frac{\alpha}{2}\|(u_1 - u_d)\|_{L^2(\partial\Omega)}) + \\
&\qquad (1-\lambda)(\frac{1}{2}\|y_2 - y_d\|_{L^2(\Omega)} + \frac{\alpha}{2}\|u_2 - u_d\|_{L^2(\partial\Omega)}) \\
&= \lambda J(y_1, u_1) + (1-\lambda)J(y_2, u_2)) \\
&= \lambda J(z_1) + (1-\lambda)J(z_2)).
\end{aligned}$$

Hence, the cost function J is convex on Z_E.

As the norm function is continuous, the cost function $J : Z \to \mathbb{R}$ is continuous. This implies that J is convex and continuous on Z_E. Thus J is lower semicontinuous on Z_E, by Theorem 3.5 in ([11], chapter 3).

Lemma 2.4 *The bilinear form $a : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$ defined in (2.8) is bounded, i.e. there exist a constant $M_1 \geq 0$ such that*

$$|a(y, v)| \leq M_1 \|y\|_{H^1} \|v\|_{H^1}, \quad \forall y, v \in H^1(\Omega) \tag{2.13}$$

and the linear functional $F : H^1(\Omega) \to \mathbb{R}$ defined in (2.9) is also bounded, i.e. there exist a constant $M_2 \geq 0$ such that

$$|F(v)| \leq M_2 \|v\|_{H^1}, \quad \forall v \in H^1(\Omega). \tag{2.14}$$

Proof The boundedness of a :

$$\begin{aligned}
|a(y,v)| &\leq \int_\Omega |\nabla y \cdot \nabla v| dx + \int_\Omega |cyv| dx \\
&\leq \Big(\int_\Omega |\nabla y|^2 dx\Big)^{1/2} \Big(\int_\Omega |\nabla v|^2 dx\Big)^{1/2} + \|c\|_{L^\infty} \Big(\int_\Omega |y|^2 dx\Big)^{1/2} \Big(\int_\Omega |v|^2 dx\Big)^{1/2} \\
&\qquad \text{(by Cauchy-Schwartz inequality)} \\
&= \|\nabla y\|_{L^2} \|\nabla v\|_{L^2} + \|c\|_{L^\infty} \|y\|_{L^2} \|v\|_{L^2} \\
&\leq \Big(1 + \|c\|_{L^\infty}\Big) \|y\|_{H^1} \|v\|_{H^1}.
\end{aligned}$$

To show the second assertion: for $r \in H^1(\Omega)^*$ and $u \in L^2(\partial\Omega)$ we have

$$\begin{aligned}
|F(v)| &\leq \int_\Omega |rv| dx + \int_{\partial\Omega} |uv| d\mu x \\
&\leq \|r\|_{H^1(\Omega)^*} \|v\|_{H^1(\Omega)} + \|u\|_{L^2(\partial\Omega)} \|v\|_{H^1(\Omega)} \\
&= \Big(\|r\|_{H^1(\Omega)^*} + \|u\|_{L^2(\partial\Omega)}\Big) \|v\|_{H^1(\Omega)}.
\end{aligned}$$

From the definition of the bilinear form a and Lemma 2.4 we observe that $a(y,\cdot) \in H^1(\Omega)^*$ for all $y \in H^1(\Omega)$ and the mapping $y \in H^1(\Omega) \to a(y,\cdot) \in H^1(\Omega)^*$ is continuous and linear. Thus by Theorem 8.3.1 ([12], Chapter 8) there exists a bounded linear operator $A : H^1(\Omega) \to H^1(\Omega)^*$ with

$$a(y,v) = \langle Ay, v \rangle_{H^1(\Omega)^*, H^1(\Omega)} \ \ \forall y, v \in H^1(\Omega).$$

Similarly, we can write

$$F(v) = \langle r, v \rangle + \langle Bu, v \rangle \ \ \forall v \in H^1(\Omega)$$

where $B : L^2(\partial\Omega) \to H^1(\Omega)^*$ is a bounded linear operator defined by

$$\langle Bu, v \rangle = \int_{\partial\Omega} uv d\mu x \ \ \forall v \in H^1(\Omega).$$

Application of the Lax-Milgram Lemma yields the existence of a unique solution for the state equation for each control $u \in L^2(\partial\Omega)$. Thus we have the following theorem.

Theorem 2.5 *Let $\Omega \in \mathbb{R}^n$ be open and bounded. Then the bilinear form in (2.8) is bounded, coercive and the associated operator $A : H^1(\Omega) \to H^1(\Omega)^*$ has a bounded inverse. In particular, for a given $r \in H^1(\Omega)^*$ and for all $u \in L^2(\partial\Omega)$ the state equation (2.1b) with the boundary*

condition (2.1c) has a unique weak solution $y^* \in H^1(\Omega)$ *and there exist a constant* $C_\Omega > 0$ *such that*

$$\|y^*\|_{H^1(\Omega)} \leq C_\Omega(\|r\|_{H^1(\Omega)^*} + \|u\|_{L^2(\partial\Omega)}).$$

Proof The boundedness of a follows from (2.13). The coercivity of a follows from

$$\begin{aligned} a(y,y) &= \int_\Omega |\nabla y \cdot \nabla y| dx + \int_\Omega |cy\, y| dx \\ &= \int_\Omega |\nabla y|^2 dx + c\int_\Omega y^2 dx \\ &\geq \min\{1,c\} \int_\Omega (|\nabla y|^2 + y^2) dx \\ &= \min\{1,c\}\|y\|^2_{H^1(\Omega)} =: \frac{1}{C_0}\|y\|^2_{H^1(\Omega)}. \end{aligned}$$

The boundedness of F is given by (2.14). Hence, the result holds by the Lax-Milgram Lemma with $C_\Omega = C_0$.

From the theorem, we can conclude that there exist a continuous linear operator $S : L^2(\partial\Omega) \to H^1(\Omega)$ such that

$$a(y,v) := a(Su,v) = F(v), \quad \forall v \in H^1(\Omega).$$

Theorem 2.6 *There exists some real number* $\alpha > 0$ *such that for all* $z_1, z_2 \in Z_E$ *we have*

$$J(z_2) \geq J(z_1) + J'(z_1, z_2 - z_1) + \frac{\alpha}{2}\|z_2 - z_1\|^2.$$

Proof The cost function J is directionally differentiable in Z and the directional derivative at $(y,u) \in Z$ in the direction $(\triangle y, \triangle u) \in Z_E$ is given by

$$J'\Big((y,u),(\triangle y, \triangle u)\Big) = \langle y - y_d, \triangle y\rangle_{L^2(\Omega)} + \beta\langle u - u_d, \triangle u\rangle_{L^2(\partial\Omega)}.$$

Let $z_1 = (y_1, u_1)$, $z_2 = (y_2, u_2) \in Z_E$ and put $\triangle y = y_2 - y_1, \triangle u = u_2 - u_1$. Then we have

$$\begin{aligned}
&J(y_1 + \triangle y, u_1 + \triangle u) - J(y_1, u_1) - J'\Big((y_1, u_1), (\triangle y, \triangle u)\Big) \\
&\quad = \tfrac{1}{2}\|y + \triangle y - y_d\|^2_{L^2(\Omega)} + \tfrac{\beta}{2}\|u + \triangle u - u_d\|^2_{L^2(\partial\Omega)} - \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} - \tfrac{\beta}{2}\|u - u_d\|^2_{L^2(\partial\Omega)} \\
&\qquad -J'\Big((y, u), (\triangle y, \triangle u)\Big) \\
&\quad = \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \tfrac{1}{2}\|\triangle y\|^2_{L^2(\Omega)} + \langle y - y_d, \triangle y\rangle_{L^2(\Omega)} + \tfrac{\beta}{2}\|u - u_d\|^2_{L^2(\partial\Omega)} \\
&\qquad +\tfrac{\beta}{2}\|\triangle u\|^2_{L^2(\partial\Omega)} + \langle u - u_d, \triangle u\rangle_{L^2(\partial\Omega)} - \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} - \tfrac{\beta}{2}\|u - u_d\|^2_{L^2(\partial\Omega)} \\
&\qquad -\langle y - y_d, \triangle y\rangle_{L^2(\Omega)} - \langle u - u_d, \triangle u\rangle_{L^2(\partial\Omega)} \\
&\quad = \underbrace{\frac{1}{2}\|\triangle y\|^2_{L^2(\Omega)}}_{\geq 0} + \tfrac{\beta}{2}\|\triangle u\|^2_{L^2(\partial\Omega)} \\
&\quad \geq \tfrac{\beta}{4}\|\triangle u\|^2_{L^2(\partial\Omega)} + \tfrac{\beta}{4}\|\triangle u\|^2_{L^2(\partial\Omega)}.
\end{aligned}$$

Since $(\triangle y, \triangle u) \in Z_E$ we have $E(\triangle y, \triangle u) = 0$. Thus by Theorem 2.5 there exist a constant $C > 0$ such that

$$\|\triangle y\|^2_{H^1(\Omega)} \leq C\|\triangle u\|^2_{L^2(\partial\Omega)}.$$

This implies that

$$\begin{aligned}
J(y_1 + \triangle y, u_1 + \triangle u) - J(y_1, u_1) - J'\Big((y_1, u_1), (\triangle y, \triangle u)\Big) \\
\geq \tfrac{\beta}{4}(\tfrac{1}{C}\|\triangle y\|^2_{H^1(\Omega)} + \|\triangle u\|^2_{L^2(\partial\Omega)}) \\
\geq \tfrac{\beta}{4}\min\{1, \tfrac{1}{C}\}(\|\triangle y\|^2_{H^1(\Omega)} + \|\triangle u\|^2_{L^2(\partial\Omega)}).
\end{aligned}$$

It is equivalent to

$$\begin{aligned}
J(y_2, u_2) - J(y_1, u_1) - J'\Big((y_1, u_1), (y_2 - y_1, u_2 - u_1)\Big) \\
\geq \tfrac{\beta}{4}\min\{1, \tfrac{1}{C}\}(\|y_2 - y_1\|^2_{H^1(\Omega)} + \|u_2 - u_1\|^2_{L^2(\partial\Omega)}).
\end{aligned}$$

Therefore,

$$J(z_2) \geq J(z_1) + J'(z_1, z_2 - z_1) + \frac{\alpha}{2}\|z_2 - z_1\|^2, \ \forall z_1, z_2 \in Z_E$$

for $\alpha = \frac{\beta}{4}\min\{1, \frac{1}{C}\} > 0$.

Theorem 2.6 implies that the objective function J is strongly convex on Z_E.

Let us denote $\Omega_1 = \partial\Omega$ and $\Omega_2 = \Omega$. Since the functions $g_1 : Z \to L^2(\partial\Omega)$ and $g_2 : Z \to H^1(\Omega)$ are projections from Z to $L^2(\partial\Omega)$ and $H^1(\Omega)$ respectively they are linear and continuous.

Hence, for each $i = 1, 2$ and for each closed convex set $U \subset L^2(\Omega_i)$, the set $\{z \in Z_E : g_i(z) \in U\}$ is closed. Further since linear functions are convex, for all $z_1, z_2 \in Z_E$ and all $t \in [0, 1]$ we have

$$tg_i(z_1)(x) + (1-t)g_i(z_2)(x) \geq g_i(tz_1 + (1-t)z_2)(x) \text{ for almost all } x \in \Omega_i, i = 1, 2.$$

Thus, from Assumption 2.1, Theorem 2.3 and Theorem 2.6 we can see that all the assumptions of problem (P) are fulfilled by the model problem (MP). Therefore, the model problem has a unique optimal solution.

Similarly, we can see that the model problem (MP) and problem (P) have the same structure with $m = 2$. Hence, all the theoretical analysis of problem (P) holds true for (MP). In the next chapters we present all the additional assumptions and results of Gfrerer [1] on problem (P), and check wether they are fulfilled or not by the model problem.

Chapter 3

Generalized Penalty Method

The penalty method, which uses penalty functions, is a standard technique in constrained optimization. Its basic principle is to guarantee the fulfilment of the constraints in an asymptotic sense by including in the objective function an additional term (the penalty) that acts against the optimization goal if constraints are violated. In this way the given constrained programming problem is embedded in a family of variational problems that depend upon some parameters appearing in the penalty term and contain no restrictions, i.e. are unconstrained.

In this section penalty method is designed to solve problem (P) and hence the model problem by instead solving a sequence of specially constructed unconstrained optimization problems. In this method the feasible region of (P) is expanded from Z_F to the vector space Z_E by a large cost or "penalty" is added to the objective function for points that lie outside of the original feasible region Z_F.

Let the indicator function of the set of non-positive real numbers is defined by

$$\mathcal{I}_{(-\infty,0]}(t) := \begin{cases} 0 & \text{if } t \leq 0 \\ \infty & \text{if } t > 0. \end{cases}$$

Let $\bar{z} \in Z$ be a solution of problem (P), then $\bar{z}$ fulfills

i. $E\bar{z} = 0$

ii. $g_i(\bar{z})(x) \leq \varphi_i(x)$ a.e. in Ω_i and for all $i \in \{1, \ldots, m\}$, and

iii. $J(\bar{z}) \leq J(z), \ \forall z \in Z_f$.

Thus from ii. we have

$$\mathcal{I}_{(-\infty,0]}(g_i(\bar{z})(x) - \varphi_i(x)) = 0 \quad \text{and}$$

$$\mathcal{I}_{(-\infty,0]}(g_i(\bar{z})(x) - \varphi_i(x)) \leq \mathcal{I}_{(-\infty,0]}(g_i(z)(x) - \varphi_i(x)) \forall i \in \{1, \ldots, m\}, \ \forall z \in Z.$$

This implies that

$$\sum_{i=1}^{m} \int_{\Omega_i} \mathcal{I}_{(-\infty,0]}(g_i(\bar{z})(x) - \varphi_i(x)) \leq \sum_{i=1}^{m} \int_{\Omega_i} \mathcal{I}_{(-\infty,0]}(g_i(z)(x) - \varphi_i(x)) \ \forall z \in Z \tag{3.1}$$

Therefore, from (iii.) and (3.1) we can conclude that $\bar{z}$ is also the solution of

$$\min_z J(z) + \Psi(z) := J(z) + \sum_{i=1}^{m} \int_{\Omega_i} \mathcal{I}_{(-\infty,0]}(g_i(z)(x) - \varphi_i(x))dx,$$
$$\text{subject to}$$
$$Ez = 0.$$

We are going to use the notation of a generalized penalty method to choose a triple $(Q, \bar{q}, (\psi_{i,q})_{i=1,\ldots,m,q\in Q})$, where the parameter set Q is some topological space and $\bar{q} \in Q$ is a target parameter, which is assumed to have a countable basis of neighborhoods.

Here we assume that for each parameter $q \in \dot{Q} := Q \setminus \{\bar{q}\}$ and each $i \in \{1, \ldots, m\}$ there exist a family of function $\psi_{i,q} : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ such that $\psi_{i,q}$ approaches the indicator functions $\mathcal{I}_{(-\infty,0]}$ for the parameter q which converge to the target parameter $\bar{q}$.

The aim of the penalty method is to approximate the solution $\bar{z}$ of (P) by the solutions $\bar{z}_q$ of the following unconstrained problems

$$(\mathrm{P_q}) \qquad \min_z J_q(z) := J(z) + \Psi_q(z)$$
$$:= J(z) + \sum_{i=1}^{m} \int_{\Omega_i} \psi_{i,q_i}(g_i(z)(x) - \varphi_i(x))dx,$$
$$\text{s.t.}$$
$$Ez = 0.$$

Next we will see the existence of a unique solution $\bar{z}_q$ of the problem (P_q) and the convergence of the solution $\bar{z}_q$ to the solution $\bar{z}$ of problem (P) as $q \to \bar{q}$. We denote the system of neighborhoods of $\bar{q}$ by $\mathcal{Q}$ and the system of the weak neighborhoods of z in Z_E by $\mathcal{Z}(z)$.

Theorem 3.1 *Assume the assumption (P1)–(P4) fulfilled. Assume that for each $q \in \dot{Q}$ the function $\Psi_q : Z \to \mathbb{R} \cup \{\infty\}$ is well defined, convex, l.s.c. on Z_E and $\Psi_q(z_f) < \infty$ for $z_f \in Z_F$. Further assume that for each $i \in \{1, \ldots, m\}$ there is some linear functional $\eta_i^* \in Z_E^*$ and sequences of real numbers $(a_{i,q})_{q\in\dot{Q}}, (b_{i,q})_{q\in\dot{Q}}$ with $\lim_{q\to\bar{q}} a_{i,q} = \lim_{q\to\bar{q}} b_{i,q} = 0$ such that*

$$\Psi_q(z) \geq \sum_{i=1}^{m} (b_{i,q}\langle \eta_i^*, z\rangle + a_{i,q}) \quad \text{for all } z \in Z_E.$$

Then for each $q \in \dot{Q}$ the problem (P_q) has a unique solution denoted by $\bar{z}_q$.

The following theorem shows the convergence of the solutions $\bar{z}_q$ to the solution $\bar{z}$.

Theorem 3.2 *Assume all the assumptions of Theorem 3.1 hold. In addition assume that*

1. *for each $\hat{z} \in Z_E$ not feasible for (P) and for each real $R > 0$ there are some neighborhoods $U_{\hat{z}} \in \mathcal{Z}(\hat{z})$ and $U_{\bar{q}} \in \mathcal{Q}$ such that*

$$\inf_{q\in U_{\bar{q}}} \inf_{z\in U_{\hat{z}}} \Psi_q(z) \geq R$$

2. *for each $z' \in Z_F$ feasible for (P) there are some family $(z'_q)_{q\in\dot{Q}} \subset Z_E$ with $\lim_{q\to\bar{q}} z'_q = z'$ such that $\limsup_{q\to\bar{q}} \Psi_q(z'_q) \leq 0$,*

then

$$\lim_{q\to\bar{q}} \bar{z}_q = \bar{z}.$$

Theorem 3.1 and Theorem 3.2 are proved by Gfrerer in [1].

In order to have the function $\Psi_q : Z \to \mathbb{R} \cup \{\infty\}$ that satisfies the necessary conditions of Theorem 3.1 and Theorem 3.2 we choose some penalty functions $\psi_{i,q} : Z \to \mathbb{R} \cup \{\infty\}$ that satisfy the following assumptions. It is because the property of Ψ_q is determined by ψ_q.

Assumption:

(Q1) $\psi_{i,q}$ is convex, l.s.c. and increasing function with $(-\infty, 0] \subset \text{dom}\psi_{i,q}$.

(Q2) $\lim_{q\to\bar{q}} \psi_{i,q}(t) = \mathcal{I}_{(-\infty,0]}(t),\ \ \forall t \neq 0.$

The specific choice of the penalty functions is dependent on the nature of the problem. The following are some the well known possible examples of penalty functions for the case of $m = 1$.

1. Quadratic penalty function:

   In this case the parameter set is given by $Q = [0, \infty]$, the target parameter is $\bar{q} = \infty$ and the penalty function is given by

   $$\psi_k(t) = k \max\{0, t\}^2 \ \text{ for every } q \in \dot{Q}.$$

2. The Logarithmic barrier function:

   For the case of Logarithmic barrier function the parameter set is given by $Q = \mathbb{R}_+$, the target parameter is $\bar{q} = 0$ and the penalty function is given by

   $$\psi_k(t) = \begin{cases} -k \ln(-t) & \text{if } t < 0 \\ \infty & \text{if } t \geq 0 \end{cases}$$

3. The Combined logarithmic-quadratic penalty function:

   In this case the parameter set Q is given by $Q = \{(\kappa, \epsilon) \in \mathbb{R}^2_+ : \kappa \geq \alpha\epsilon^{3/2} > 0, \kappa^{1/2} \ln \epsilon > -\beta\}$ where α and β are some positive constants, the target parameter is $\bar{q} = (0, 0)$ and the penalty function is

   $$\psi_{\kappa,\epsilon}(t) = \begin{cases} -\kappa \ln(-t) & \text{if } t \leq -\epsilon \\ \kappa(-\ln(\epsilon) + \frac{t+\epsilon}{\epsilon} + \frac{(t+\epsilon)^2}{2\epsilon^2}) & \text{if } t > -\epsilon \end{cases} \tag{3.2}$$

   for $(\kappa, \epsilon) \in \dot{Q}$.

Note that the combined logarithmic-quadratic penalty function has the following important properties.

For each (κ, ϵ) $\psi_{\kappa,\epsilon}$ is twice continuously differentiable in $\mathbb{R}$ and the second derivative satisfies:

$$|D^2\psi_{\kappa,\epsilon}(t)| \leq \frac{\kappa}{\epsilon^2}, \forall t \in \mathbb{R} \text{ and}$$
$$|D^2\psi_{\kappa,\epsilon}(t_1) - D^2\psi_{\kappa,\epsilon}(t_2)| \leq 2\kappa\frac{t_1 - t_2}{\epsilon^3}, \forall t_1, t_2 \in \mathbb{R}.$$

Furthermore, we have $\psi_{\kappa,\epsilon}(0) = -\kappa \ln \epsilon < \beta\sqrt{\kappa} \to 0$ and $|D^2\psi_{\kappa,\epsilon}(t)| \leq \frac{\kappa}{\epsilon^2} \geq \frac{\alpha}{\sqrt{\epsilon}} \to \infty$ for $(\kappa, \epsilon) \to (0, 0)$.

The choice of a penalty function must also be related to some feasible points of the problem. Hence we make the following additional assumptions on the choice of penalty functions.

Assumption:

(PQ1) For each $i \in \{1, \ldots, m\}$ we have either

$$g_i(z_f) \leq \varphi_i - \delta \text{ a.e. in } \Omega_i \tag{3.3}$$

for some real number $\delta > 0$ or the condition

$$0 \in \text{dom}\psi_{i,q} \; \forall q \in \dot{Q}, \text{ and } \lim_{q \to \bar{q}} \psi_{i,q}(0) = 0 \tag{3.4}$$

are fulfilled.

Note that the assumption (3.3) is satisfied if (3.4) holds for all $i \in \{1, \ldots, m\}$, i.e. $\psi_{i,q}$ converges pointwise to $\mathcal{I}_{(-\infty,0]}$ on the whole real axis.

In the remaining of this chapter we will show that the assumptions of Theorem 3.1 and Theorem 3.2 are satisfied by the penalty functions $\psi_{i,q}$ which satisfy the assumptions (Q1), (Q2) and (PQ1). We start this discussion by showing the existence of some families of real

numbers $(a_{i,q})_{q\in\dot{Q}}$ and $(b_{i,q})_{q\in\dot{Q}}$ with $\lim_{q\to\bar{q}} a_{i,q} = \lim_{q\to\bar{q}} b_{i,q} = 0$ and the existence of some linear functionals $\eta_i^* \in Z_E^*$ for each $i \in \{1,\ldots,m\}$ such that the following inequality holds

$$\Psi_q(z) \geq \sum_{i=1}^{m}(b_{i,q}\langle\eta_i^*, z\rangle + a_{i,q}) \quad \text{for all } z \in Z_E.$$

To do this we use the following construction. From Assumption (Q1), we have the convexity of $\psi_{i,q}$ and hence the directional differentiability of $\psi_{i,q}$ at $-1 \in \mathrm{dom}\psi_{i,q}$ is well defined. For each $i \in \{1,\ldots,m\}$ and each $q \in \dot{Q}$ we set

$$\begin{aligned}&\tilde{a}_{i,q} = \psi_{i,q}(-1) + \psi'_{i,q}(-1,1) \in \mathbb{R} \text{ and}\\ &b_{i,q} = \psi'_{i,q}(-1,1) \in \mathbb{R}.\end{aligned}$$

From the monotonicity assumption of $\psi_{i,q}$ we have $b_{i,q} \geq 0$. From the assumption $\lim_{q\to\bar{q}} \psi_{i,q} = \mathcal{I}_{(-\infty,0]}$ and the convexity of $\psi_{i,q}$ we obtain

$$\begin{aligned}0 \leq \lim_{q\to\bar{q}} \tilde{a}_{i,q} &= \lim_{q\to\bar{q}} b_{i,q}\\ &= \lim_{q\to\bar{q}} \psi'_{i,q}(-1,1)\\ &\leq \lim_{q\to\bar{q}} 2(\psi_{i,q}(-1/2) - \psi_{i,q}(-1)) = 0.\end{aligned}$$

In addition it is easy to show

$$\alpha_{i,q} = \tilde{a}_{i,q} + b_{i,q}t = \psi_{i,q}(-1) + \psi'_{i,q}(-1,1)(t+1)$$

is an affine minorant of $\psi_{i,q}$.

By Lemma 2.3 in Gfrerer [1] the function $z \to \int_{\Omega_i}(g_i(z)(x) - \varphi_i(x))$ is convex and l.s.c. on Z_E. Hence by using Proposition 3.1 in [4], there exists some affine minorant $\gamma_i + \langle\eta_i^*, z\rangle$, $\gamma_i \in \mathbb{R}$, $\eta_i^* \in Z_E^*$ of this function. Then for every $z \in Z_E$ we obtain

$$\begin{aligned}\Psi_{i,q}(z) &= \int_{\Omega_i} \psi_{i,q}(g_i(z)(x) - \varphi_i(x))d\mu_i(x)\\ &\geq \int_{\Omega_i} (\tilde{a}_{i,q} + b_{i,q}(g_i(z)(x) - \varphi_i(x)))d\mu_i(x)\\ &= \tilde{a}_{i,q}\mu_i(\Omega_i) + b_{i,q}\int_{\Omega_i}(g_i(z)(x) - \varphi_i(x))d\mu_i(x)\\ &\geq \tilde{a}_{i,q}\mu_i(\Omega_i) + b_{i,q}(\gamma_i + \langle\eta_i^*, z\rangle)\\ &= (\tilde{a}_{i,q}\mu_i(\Omega_i) + b_{i,q}\gamma_i) + b_{i,q}\langle\eta_i^*, z\rangle\\ &\geq a_{i,q} + b_{i,q}\langle\eta_i^*, z\rangle\end{aligned}$$

where $a_{i,q} = -|\tilde{a}_{i,q}|\mu_i(\Omega_i) - |b_{i,q}\gamma_i|$.

Note that $\lim_{q\to\bar{q}} \tilde{a}_{i,q} = \lim_{q\to\bar{q}} b_{i,q} = 0$.

Remark 3.1. From the above discussion, we can see that all the assumptions of Theorem 3.1 are fulfilled.

From the following two lemmas we have, if the assumptions (P1)–(P4), (Q1),(Q2) and (PQ1) are fulfilled then the assumptions of Theorem 3.2 satisfied.

Lemma 3.3 *Assume that the assumptions (P1)–(P4), (Q1),(Q2) and (PQ1) are fulfilled. Then for each $i \in \{1, \ldots, m\}$ we have $z_f \in dom\Psi_{i,q}$. Further, for each $\hat{z} \in Z_E$ feasible for the problem (P) there is some family $(\hat{z}_q)_{q\in\dot{Q}}$ such that $\lim_{q\to\bar{q}} \hat{z}_q = \hat{z}$ and $\limsup_{q\to\bar{q}} \Psi_q(\hat{z}_{i,q}) \leq 0$, $\forall i$.*

Lemma 3.4 *Assume that the assumptions (P1)–(P4), (Q1),(Q2) and (PQ1) are fulfilled. Then for each $\hat{z} \in Z_E$ not feasible for the problem (P) and for each real $R > 0$ there are some neighborhoods $U_{\hat{z}} \in \mathcal{Z}(\hat{z})$ and $U_{\bar{q}} \in \mathcal{Q}$ such that*

$$\inf_{q\in U_{\bar{q}}} \inf_{z\in U_{\hat{z}}} \sum_{i=1}^{m} \Psi_{i,q}(z) \geq R.$$

Thus from Remark 3.1, Lemma 3.3 and Lemma 3.4 one can observe that all the assumption of Theorem 3.1 and Theorem 3.2 are fulfilled. Hence we have the following convergence result.

Corollary 3.5 *Assume that the assumptions (P1)–(P4), (Q1), (Q2) and (PQ1) are fulfilled. Then all the assumptions of Theorem 3.1 and Theorem 3.2 are fulfilled. In particular, for each $q \in \dot{Q}$ the problem (P_q) has a unique solution $\bar{z}_q$ and*

$$\lim_{q\to\bar{q}} \bar{z}_q = \bar{z}.$$

Chapter 4

Error Estimates

In this chapter, we are interested in estimates of the distance from $\bar{z}_q$ to $\bar{z}$ via some computable quantity, that is, in error estimate of the form

$$\|\bar{z}_q - \bar{z}\| = \mathcal{O}(\gamma(x_\delta, \delta))$$

where $\gamma : Z \times \mathbb{R}_+ \to \mathbb{R}_+$ is some (easily) computable function such that, at least, $\gamma(z_q, q) \to 0$ as $q \to \bar{q}$ and $\bar{z}_q \to \bar{z}$.

We start our analysis by considering the existence of the directional derivative of J_q under the necessary assumptions. Let $\hat{z}, z \in Z_E$ and $i \in \{1, 2, \ldots, m\}$. Then by assumption (P4) we have

$$g_i(\hat{z} + t(z - \hat{z})) - g_i(\hat{z}) \leq t(g_i(z) - g_i(\hat{z})). \tag{4.1}$$

This implies that, the function

$$\zeta(t) := \frac{g_i(\hat{z} + t(z - \hat{z})) - g_i(\hat{z})}{t}$$

is bounded above $\forall t \in [0, 1]$.

Note that ζ monotonically increasing.

Hence there exist measurable function $h : \Omega_i \to \mathbb{R}$ such that for every sequence (t_n) of positive

numbers converging to 0 we have

$$h(x) := \lim_n \frac{g_i(\hat{z} + t_n(z - \hat{z})) - g_i(\hat{z})}{t_n}, \ \mu_i - \text{ a.e. in } \Omega_i$$

The function h is the directional derivative of g_i at $\hat{z}$ in the direction of $z - \hat{z}$, we denote it by $g_i'(\hat{z}, z - \hat{z})$. Note that $g_i'(\hat{z}, z - \hat{z})$ is uniquely defined up to a set of measure zero and from (4.1) it satisfies

$$g_i(z)(x) \geq g_i(\hat{z})(x) + g_i'(\hat{z}, z - \hat{z})(x)$$

for almost all $x \in \Omega_i$.

Lemma 4.1 *Let $i \in \{1, 2, \ldots, m\}$, $q \in \dot{Q}$ be fixed and let $\hat{z} \in dom\Psi_{i,q} \cap Z_E$, $z \in Z_E$ be such that $\hat{z} + t_0(z - \hat{z}) \in dom\Psi_{i,q}$ for some positive real $t_0 > 0$. Then*

$$\begin{aligned} \lim_{t \downarrow 0} \frac{\Psi_{i,q}(\hat{z} + t(z - \hat{z})) - \Psi_{i,q}(\hat{z})}{t} &= \int_{\Omega_i} \psi_{i,q}^{\#}(g_i(\hat{z})(x) - \varphi_i(x), g_i'(\hat{z}, z - \hat{z})(x)) d\mu_i(x) \\ &\in \mathbb{R} \cup \{-\infty\} \end{aligned} \tag{4.2}$$

holds, where

$$\psi_{i,q}^{\#}(u, h) := \begin{cases} \psi_{i,q}'(u, h) & \text{if } u \in \text{ int dom } \psi_{i,q}, \\ \lim_{t \downarrow 0} \frac{\psi_{i,q}(u+th) - \psi_{i,q}(u)}{t} \in \mathbb{R} \cup \{-\infty\} & \text{if } u \in \text{ bd dom } \psi_{i,q}, \ h < 0 \\ 0 & \text{otherwise.} \end{cases}$$

From the definition of $\psi_{i,q}^{\#}(u, \cdot)$ and the convexity of $\psi_{i,q}$ it is easy to show that for each u the mapping $\psi_{i,q}^{\#}(u, \cdot)$ is positively homogeneous, i.e.,

$$\psi_{i,q}^{\#}(u, \lambda h) = \lambda \psi_{i,q}^{\#}(u, h), \ \forall \lambda \geq 0, \ h \in \mathbb{R}$$

and sublinear, i.e.

$$\psi_{i,q}^{\#}(u, h_1 + h_2) \leq \psi_{i,q}^{\#}(u, h_1) + \psi_{i,q}^{\#}(u, h_2), \ \forall h_1, h_2 \in \mathbb{R}.$$

Moreover from the monotonicity of $\psi_{i,q}$ we have $\psi_{i,q}^{\#}(u, h) \geq 0, \forall h \geq 0$ and $\psi_{i,q}^{\#}(u, h) \leq 0, \forall h \leq 0$. From this and the sublinearity of $\psi_{i,q}^{\#}(u, \cdot)$ we obtain the following monotonicity property

$$\psi_{i,q}^{\#}(u, h_1) \leq \psi_{i,q}^{\#}(u, h_2), \ \forall h_1 \leq h_2. \tag{4.3}$$

Here we have to notice that for $u \in$ bd dom $\psi_{i,q}$ we can have $\psi_{i,q}^{\#}(u,h) = -\infty$ and as a result the limit in (4.2) can be $-\infty$. However, it is not possible for the case $\hat{z} = \bar{z}_q$, it is because $\bar{z}_q$ is a minimizer J_q. Now we assume the following assumption to guarantee the applicability of Lemma 4.1 for the case of a feasible point z.

Assumption:

(Q3) For each $i \in \{1, \ldots, m\}$ and each $q \in \dot{Q}$ there are nonnegative constants $c_{i,q}\, d_{i,q}$ such that

$$\psi_{i,q}(t/2) \leq c_{i,q}\psi_{i,q}(t) + d_{i,q} \;\; \forall t \in [-1, 0).$$

Assumption (Q3) is fulfilled if $0 \in$ dom $\psi_{i,q}$ or $\psi_{i,q}$ grows like $-ln(-t)$ or $(-t)^{-p}$ for $t \to 0_-$ as in the case of the usual barrier functions.

Corollary 4.2 *Assume that the assumptions (P1)–(P4), (Q1)–(Q3) and (PQ1) are fulfilled. Let $q \in \dot{Q}$ be fixed and let $z \in Z_E$ be feasible for the problem (P). Then J_q is directionally differentiable at $\bar{z}_q$ in the direction $z - \bar{z}_q$ with directional derivative*

$$J_q'(\bar{z}_q, z - \bar{z}_q) = J'(\bar{z}_q, z - \bar{z}_q) + \sum_{i=1}^{m} \int_{\Omega_i} \psi_{i,q}^{\#}(g_i(\bar{z}_q)(x) - \varphi_i(x), g_i'(\bar{z}_q, z - \bar{z}_q)(x))d\mu_i(x)$$

Theorem 4.3 *Suppose that the assumptions (P1)–(P4), (Q1)–(Q3) and (PQ1) are fulfilled. Then there is a neighborhood $U_{\bar{q}} \subset Q$ and a constant L such that for all $q \in U_{\bar{q}}$ we have*

$$\|\bar{z}_q - \bar{z}\|_Z \leq \frac{2}{\alpha}\Big(L\, dis(\bar{z}_q, Z_F) + \sum_{i=1}^{m} \int_{\Omega_i} \psi_{i,q}^{\#}(g_i(\bar{z}_q)(x) - \varphi_i(x), -(g_i(\bar{z}_q)(x) - \varphi_i(x)))dx\Big).$$

Theorem 4.3 provides us an efficient tool for estimating the quality of our approximate $\bar{z}_q$ of $\bar{z}$, if we can estimate the distance of $\bar{z}_q$ to the feasible set Z_F. If the distance between $\bar{z}_q$ and the feasible set is known, then we have a computable error bound. This error bound is very important to provide a reliable stopping criteria for the related numerical algorithm.

For instance, in case of the combined logarithmic-quadratic penalty function given by (3.2) we have

$$\psi^{\#}_{i,(\kappa,\epsilon)}(u,h) := \begin{cases} \kappa & \text{if } t \leq -\epsilon \\ -\kappa(\frac{2t}{\epsilon} + \frac{t^2}{\epsilon^2}) < \kappa & \text{if } t > -\epsilon \end{cases}$$

and

$$\int_{\Omega_i} \psi^{\#}_{i,(\kappa,\epsilon)}(g_i(\bar{z}_{(\kappa,\epsilon)})(x) - \varphi_i(x), -(g_i(\bar{z}_{(\kappa,\epsilon)})(x) - \varphi_i(x)))d\mu_i x \leq \kappa\mu_i(\Omega_i)$$

Therefore, we have the following a-priori bound

$$\|\bar{z}_{(\kappa,\epsilon)} - \bar{z}\|_Z \leq \frac{2}{\alpha}\Big(L\,\mathrm{dis}(\bar{z}_{(\kappa,\epsilon)}, Z_F) + \sum_{i=1}^{m} \kappa\mu_i(\Omega_i)\Big).$$

If we want to approximate the exact solution $\bar{z}$ within some given precision this formula allows us to choose κ in advance and then to adjust ϵ in such a way that the $\mathrm{dis}(\bar{z}_q, Z_F)$ is sufficiently small.

Chapter 5

Duality

In this chapter we introduce the existence of multipliers for the model problem by assuming a constraint qualification condition. We use the generalized penalty method, discussed in Chapter 3, to approximate these multipliers. To approximate the multipliers we need the following additional assumption on the penalty functions.

Assumption:

(Q4) For each $i \in \{1, \ldots, m\}$ and each $q \in \dot{Q}$ assume that $\psi_{i,q}$ is differentiable in $\mathbb{R}$ and $\limsup_{t\to\infty} t^{-r_i}\psi_{i,q}(t) < \infty$. Furthermore, assume $\lim_{q\to\bar{q}} \psi_{i,q}(0) = 0$.

Note that if $\psi_{i,q}$ satisfies (Q4), it also satisfies the assumptions (PQ1) and (Q3). In addition, the assumptions (Q4) and (Q1) gives us the bound $|\psi_{i,q}(t)| \leq C_{i,q}(1 + |t|^{r_i})$ for all $t \in \mathbb{R}$, $i \in \{1, \ldots, 2\}$ and some nonnegative constante $C_{i,q}$. Consequently, for each $u \in L^{r_i}(\Omega_i)$ the function which maps x into $\psi_{i,q}(u(x))$ is μ_i-integrable.

It is known that for $r_i \geq 1$, $L^{r_i}(\Omega_i)$ is continuously embedded in $L^1(\Omega_i)$. Hence from Lemma 2.2 in [1] the mapping

$$\rho_{i,q}(u) := \int_{\Omega_i} \psi_{i,q}(u(x))d\mu_i x$$

is lower semicontinuous, convex and real valued on $L^{r_i}(\Omega_i)$. By using Corollaries 2.4 and 2.5 in ([4], Chapter 1) we have that $\rho_{i,q}$ is continuous and locally Lipschitz and hence everywhere directionally differentiable. By using similar arguments as in the proof of Lemma 3.1 in [1] we can show that

$$\rho'_{i,q}(u,h) = \int_{\Omega_i} \psi'_{i,q}(u(x),h(x))d\mu_i x.$$

From Lipschitz continuity of $\rho_{i,q}$ near u with some modulus L we have the following bound for its directional derivative,

$$\rho'_{i,q}(u,h) \leq L\|h\|_{L^{r_i}(\Omega_i)}, \forall h \in L^{r_i}(\Omega_i).$$

Thus from boundedness and differentiability of $\psi_{i,q}$ we deduce that the mapping $h \to \rho'_{i,q}(u,h)$ is bounded linear functional on $L^{r_i}(\Omega_i)$.

For each $i \in \{1,\ldots,m\}$ and each $q \in \dot{Q}$ we define linear functional $l^*_{i,q} \in L^{r_i}(\Omega_i)^*$ by

$$\begin{aligned}\langle l^*_{i,q}, h\rangle &:= \rho'_{i,q}(g_i(\bar{z}_q) - \varphi, h)\\ &= \int_{\Omega_i} \psi'_{i,q}(g_i(\bar{z}_q)(x) - \varphi(x), h(x))d\mu_i x.\end{aligned}$$

Assumption (Q1) gives the monotonicity of $\psi_{i,q}$, thus we have $\langle l^*_{i,q}, h\rangle \leq 0$ for every $h \in L^{r_i}(\Omega_i)$ satisfying $h(x) \leq 0$ a.e. in Ω_i and hence from assumption (P4) the mapping $z \to \langle l^*_{i,q}, h\rangle$ is convex on Z_E.

Lemma 5.1 *For each $q \in \dot{Q}$ the point $\bar{z}_q$ is the unique solution of the problem*

$$\min_{z\in Z_E} \tilde{J}_q(z) := J(z) + \sum_{i=1}^m \langle l^*_{i,q}, g_i(z)\rangle.$$

Moreover we have

$$\lim_{q\to\bar{q}} \sum_{i=1}^m \langle l^*_{i,q}, g_i(\bar{z}) - g_i(\bar{z}_q)\rangle = \lim_{q\to\bar{q}} \sum_{i=1}^m \langle l^*_{i,q}, \varphi_i - g_i(\bar{z}_q)\rangle = 0$$

and

$$\limsup_{q\to\bar{q}} \sum_{i=1}^{m} \langle l^*_{i,q}, u_i - g_i(\bar{z}_q)\rangle \le 0$$

for each $(u_1,\ldots,u_m) \in \prod_{i=1}^m L^{r_i}(\Omega_i)$ *such that* $u_i \le \varphi_i$ μ_i*-a.e. in* Ω_i.

Next we will show that the weak-*limit point of the family of linear functionals $l^*_{i,q}$ as $q \to \bar{q}$ is a solution of some dual problem. However, such weak-*limit point will not exist in $L^{r_i}(\Omega_i)$ generally but it exists in some other dual spaces. The existence of such limit points has a close relation with some constraint qualification conditions. In order to have all the necessary results we consider the following problem with bilateral constraints.

$$\begin{aligned}(\hat{P}) \qquad & \min_{z\in Z} J(z) \\ & \text{subject to} \\ & \qquad Ez = 0, \\ & \qquad \underline{\varphi}_i \le A_i(z) \le \overline{\varphi}_i \ \ \mu_i - \text{a.e. in } \hat{\Omega}_i,\ i = 1,\ldots,m', \\ & \qquad \hat{g}_i(z) \le \hat{\varphi}_i, \ \ \mu_i - \text{a.e. in } \hat{\Omega}_i,\ i = m'+1,\ldots,m'',\end{aligned}$$

where $A_i \in \mathcal{L}(Z, L^{p_i}(\hat{\Omega}_i)), i = 1,\ldots,m'$ are bounded linear operators, and $\hat{g}_i : Z \to L^{p_i}(\hat{\Omega}_i)$. We assume that with the setting $m = m' + m''$, $\Omega_{2i-1} = \Omega_{2i} = \hat{\Omega}_i, r_{2i-1} = r_{2i} = p_i, g_{2i-1} = -g_{2i} = A_j, \varphi_{2i-1} = \overline{\varphi}_i, \varphi_{2i} = -\underline{\varphi}_i, 1 \le i \le m'$ and $\Omega_{m'+i} = \hat{\Omega}_i, r_{m'+i} = p_{m'+i}, g_{m'+i} = \hat{g}_i, g_{m'+i} = \hat{g}_i, \varphi_{m'+i} = \hat{\varphi}_i, m'+1 \le i \le m''$ our problem (P) is an equivalent formulation of problem $(\hat{P})$ by splitting the bilateral constraints.

In addition to the above assumptions we assume that there is some Banach space $\hat{Z}$ continuously embedded in Z such that the solution $\bar{z}$ of (P) is in $\hat{Z}$ and some Banach space Y continuously embedded in $\prod_{i=1}^{m''} L^{p_i}(\hat{\Omega}_i)$ such that $\hat{G}(\hat{Z}) \subset Y$, where
$\hat{G}(z) = (A_1 z,\ldots,A_{m'}z, g_{m'+1}(z),\ldots,g_{m''}(z))$. We also defined the closed convex set $C \subset Y$ by

$$C := \{c = (c_1,\ldots,c_{m''}) \in Y : \begin{array}{l} \underline{\varphi}_i \le c_i \le \overline{\varphi}_i \ \ \mu_i - \text{a.e. in } \hat{\Omega}_i,\ i = 1,\ldots,m', \\ c_i \le \hat{\varphi}_i, \ \ \mu_i - \text{a.e. in } \hat{\Omega}_i,\ i = m'+1,\ldots,m'' \end{array} \}. \qquad (5.1)$$

Since E is a bounded operator from Z into V and $\hat{Z} \subset Z$, E is also a bounded operator from $\hat{Z}$ into V. Thus the kernel of E induces a subspace $\hat{Z}_E = Z_E \cap \hat{Z}$ of $\hat{Z}$. $\hat{Z}_E$ is indeed a Banach space by itself.

Note that the above assumptions imply that the solution $\bar{z}$ of (P) is also a solution for the problem

$$\begin{aligned} (\hat{P}') \qquad & \min_{z \in \hat{Z}} J(z) \\ & \text{subject to} \\ & \qquad Ez = 0, \\ & \qquad \hat{G}(z) \in C. \end{aligned}$$

Now for each $q \in \dot{Q}$, let us define new linear functionals $y_q^* \in (\prod_{i=1}^{m''} L^{p_i}(\hat{\Omega}_i))^* \subset Y^*$ by

$$\langle y_q^*, y \rangle = \sum_{i=1}^{m'} \langle l_{2i-1,q}^* - l_{2i,q}^*, y_i \rangle + \sum_{i=m'+1}^{m''} \langle l_{i,q}^*, y_i \rangle.$$

The following theorem shows that the weak-* limit point y^* of the subsequence of (y_q^*) belongs to the normal cone of C defined in (5.1) at $\hat{G}(\bar{z})$ and it is proved in [1].

Theorem 5.2 *Assume that the problem (P) is an equivalent formulation of the problem $(\hat{P})$ and suppose that the assumptions (P1)–(p4),(Q1),(Q2) and (Q4) are fulfilled. Let $(q_n) \subset \dot{Q}$ be a sequence converging to $\bar{q}$ and assume that $y^* \in Y^*$ is a weak-* limit point of the sequence (y_q^*). Then y^* belongs to the normal cone of C at $\hat{G}(\bar{z})$, denoted by $N_C(\hat{G}(\bar{z}))$, i.e. $\langle y^*, c - \hat{G}(\bar{z}) \rangle \leq 0$, $\forall c \in C$, and $\bar{z}$ is the unique solution for the convex problem*

$$\min_{z \in \hat{Z}_E} J(z) + \langle y^*, \hat{G}(z) \rangle.$$

Proposition 5.3 *Assume that the problem (P) is an equivalent formulation of the problem $(\hat{P})$ and suppose that the assumptions (P1)–(p4),(Q1),(Q2) and (Q4) are fulfilled. Further assume that the mapping $\hat{G}$ is continuous on $\hat{Z}_E$ and*

$$0_Y \in \mathit{int}(\hat{G}(\hat{Z}_E) - C). \tag{5.2}$$

Then

$$\limsup q \to \bar{q} \|y_q^*\|_{Y^*} < \infty.$$

By the Alaoglu-Bourbaki Theorem, the unit ball of Y^* is weakly-* compact. Hence, the assumptions of Proposition 5.3 guarantees, for every sequence q^n converging to $\bar{q}$ the sequence $y^*_{q^n}$ has at least one weak-* limit point.

Note that the condition (5.2) in Proposition 5.3 is one of the commonly used constraint qualification condition in convex programming. It is easy to observe that the condition (5.2) is satisfied if either there exist some $c \in C$ such that $c \in \hat{G}(\hat{Z}_E)$ or there exist $\hat{z} \in \hat{Z}_E$ such that $\hat{G}(\hat{z}) \in \text{int}_Y C$, i.e. a Slater condition is fulfilled.

Proposition 5.4 *Assume that the constraints can be partitioned into* $C = C_1 \times C_2 \subset Y_1 \times Y_2 = Y$, $\hat{G} = (\hat{G}_1, \hat{G}_2), \hat{G}_i : \hat{Z} \to Y_i, i = 1, 2$ *such that* $c_1 \in int\hat{G}_1(\hat{Z}_E)($ *or more generally,* $0 \in int(\hat{G}_1(\hat{Z}_E) - C_1))$ *and* $\hat{G}_1(\hat{z}) \in C_1, \hat{G}_2(\hat{z}) \in int_Y C_2$ *for some* $c_1 \in C_1, \hat{z} \in \hat{Z}_E$. *Then condition (5.2) is fulfilled.*

If we have continuously differentiability assumption of J and $\hat{G}$ on $\hat{Z}$, then by the assumptions of Theorem 5.2 we have that $DJ(\bar{z}) + D\hat{G}(\bar{z})^* y^* \in \hat{Z}_E^{\perp}$. Moreover, if E maps $\hat{Z}$ continuously onto some Banach space $\hat{V}$, which is continuously embedded in V we can apply the Closed Range Theorem to obtain the first order necessary condition of the form $DJ(\bar{z}) + D\hat{G}(\bar{z})^* y^* + E^* \hat{v}^*$, where $\hat{v}^* \in \hat{V}^*$.

Chapter 6

Numerical Method for Solving the Subproblems

In this section we present numerical methods that we use to solve the subproblems (P_q). Throughout this section let $q \in \dot{Q}$ be a fixed parameter. We begin by considering a Newton-type method to find the effective solution of the subproblem. Then we modify the method in such away that the solutions of the subproblems converges superlinearly to the solution of the main problem (P).

Before going to the discussion of the numerical method we assume the following additional smoothness assumptions on the problem functions.

Assumption:

(P5a) The objective function J is twice continuously differentiable on Z_E and the second derivative $D^2J(z)$ is bounded for all z belonging to bounded subsets of Z_E.

(P5b) For all $i \in \{1, \ldots, m\}$ the functions g_i have the representation

$$g_i(z)(x) = \eta_i\Big(A_i z(x)\Big)$$

for some continuous linear operator $A_i \in \mathcal{L}(Z, L^2(\Omega_i)^{d_i})$, $d_i \geq 1$ and some convex functions $\eta_i : \mathbb{R}^{d_i} \to \mathbb{R}$.

Note that the assumption (P5b) does not imply in general that g_i is twice continuous differentiable, even for smooth function η_i. The continuity depends on the operator A and the space.

In addition to the above smoothness assumptions on the problem functions, we assume for each $i \in \{1, \ldots, m\}$ the choice of the penalty functions $\psi_{i,q}$ satisfies the following hypothesis:

Assumption:

(PQ2a) The function $\pi : \mathbb{R}^{di} \times \mathbb{R} \to \mathbb{R}$ defined by $\pi(s,t) = \psi_{i,q}(\eta_i(s) - t)$ is twice continuously differentiable with respect to the first variable, s for all $t \in \mathbb{R}$ and there exist some constant L such that the second derivative of π with respect to s satisfies

$$|D^2_{ss}\pi_i(s,t)|_2 \leq L \text{ and } |D^2_{ss}\pi_i(s_1,t) - D^2_{ss}\pi_i(s_2,t)|_2 \leq L|s_1 - s_2|_2 \tag{6.1}$$

for all $s_1, s_2 \in \mathbb{R}^{d_i}$ and $t \in \mathbb{R}$ where the norm $|\cdot|_2$ is defined by

$$|D^2_{ss}\pi_i(s,t)|_2 := \sup\{\langle D^2_{ss}\pi_i(s,t)h, k\rangle : |h|_2 = 1, |k|_2 = 1\}.$$

(PQ2b) For each $i \in \{1, \ldots, m\}$ the functions $\Psi_{i,q}$ is continuous on Z.

Note that according to our new setting we have

$$\begin{aligned}
\Psi_{i,q}(z) &= \int_{\Omega_i} \psi_{i,q}(g_i(z)(x) - \varphi_i(x))d\mu_i(x) \\
&= \int_{\Omega_i} \psi_{i,q}(\eta_i(A_i z(x)) - \varphi_i(x))d\mu_i(x) \\
&= \int_{\Omega_i} \pi_i(A_i z(x), \varphi_i(x))d\mu_i(x).
\end{aligned}$$

For our model problem the assumption (P5) is definitely satisfied with the continuous linear operators defined by $A_1(y,u) := u$, $A_2(y,u) := y$ and convex functions $\eta_i : \mathbb{R} \to \mathbb{R}$ defined by

$\eta_i(x) := x$ for $i = 1, 2$.

Moreover, if we choose the combined logarithmic-quadratic penalty function given by (3.2) as the penalty function for the model problem, it is twice continuously differentiable, then the assumption (PQ2) fulfilled with $d_1 = d_2 = 1$, and

$$\pi_1(s,t) = \psi_{1,q}(s-t),$$
$$\pi_2(s,t) = \psi_{2,q}(s-t).$$

The following Lemma presents the differentiability of $\Psi_{i,q}$.

Lemma 6.1 *Assume the assumptions (P5b) and (PQ2) are fulfilled. Then for each $i \in \{1, \ldots, m\}$ the mapping $\Psi_{i,q}$ is twice Gâteaux-differentiable and $\Psi_{i,q} \in C^{1,1}(Z)$, where*

$$\langle D\Psi_{i,q}(z), h\rangle = \int_{\Omega_i} \langle D_s\pi_i(A_i z(x), \varphi_i(x)), A_i h(x)\rangle d\mu_i(x)$$

$$\langle D^2\Psi_{i,q}(z)h, k\rangle = \int_{\Omega_i} \langle D^2_{ss}\pi_i(A_i z(x), \varphi_i(x))A_i h(x), A_i k(x)\rangle d\mu_i(x)$$

Further, for all $z \in Z$ we have

$$\lim_{z' \to z} \sup_{\substack{h \in \mathcal{B} \\ \tilde{k} \in \tilde{\mathcal{K}}}} \int_{\Omega_i} |\langle (D^2_{ss}\pi_i(A_i z'(x), \varphi_i(x)) - D^2_{ss}\pi_i(A_i z'(x), \varphi_i(x)))h(x), \tilde{k}(x)\rangle| d\mu_i(x) = 0 \quad (6.2)$$

for every bounded subset $\mathcal{B} \subset L^2(\Omega_i)^{d_i}$ and whenever $\tilde{\mathcal{K}}$ is either bounded in $L^{r_i}(\Omega_i)^{d_i}$ for some $r_i > 0$ or the elements of $\tilde{\mathcal{K}}$ are of the form $\tilde{k}(x) = R(x)k(x)$ where k belongs to a compact subset $\tilde{\mathcal{K}} \subset L^2(\Omega_i)^{d_i}$ and R is from a bounded subset $\mathcal{R} \subset L^\infty(\Omega_i)^{d_i \times d_i}$.

For the detailed proof of the Lemma see Gfrerer [1]. The following Corollary gives the twice continuously differentiability of $\Psi_{i,q}$.

Corollary 6.2 *Assume the assumptions (P5b) and (PQ2) are fulfilled and either*

- *$A_i : Z_E \to L^2(\Omega_i)^{d_i}$ is a compact linear operator or*
- *$A_i \in \mathcal{L}(Z_E, L^{r_i}(\Omega_i)^{d_i})$ with $r_i > 2$.*

Then $\Psi_{i,q}$ is twice continuously differentiable on Z_E.

Note that due to the Assumption (P5) and the twice differentiability assumption on $\Psi_{i,q}$, the objective function J_q is at least twice Gâteaux-differentiable on the space Z_E and its second derivative $D^2J_q(z)$ is bounded for all z belonging to bounded subsets of Z_E.

From the convexity assumption of $\psi_{i,q}$ we have the convexity of $\Psi_{i,q}$ (see Corollary 2.4 in [1]) and hence for each $i \in \{1, \ldots, m\}$ we have

$$\langle D^2\Psi_{i,q}(z)h, h\rangle \geq 0 \ \ \forall z \in Z_E \tag{6.3}$$

Due to the assumption (P3) we have

$$\begin{aligned} \langle D^2J(z)h, h\rangle &= \lim_{t\to 0^+} \frac{J(z+th) - J(z) - t\langle DJ(z), h\rangle}{\frac{1}{2}t^2} \\ &\geq \lim_{t\to 0^+} \frac{\frac{\alpha}{2}t^2\|h\|_Z^2}{\frac{1}{2}t^2} \\ &= \alpha\|h\|_Z^2 \end{aligned} \tag{6.4}$$

Hence from equations (6.3) and (6.4) $\forall z, h \in Z_E$ we have

$$\begin{aligned} \langle D^2J_q(z)h, h\rangle &= \langle D^2J(z)h, h\rangle + \langle D^2\Psi_{i,q}(z)h, h\rangle \\ &\geq \langle D^2J(z)h, h\rangle \\ &= \alpha\|h\|_Z^2. \end{aligned}$$

Since $\bar{z}_q$ is a solution of (P_q) by the first order optimality condition we have that $DJ_q(\bar{z}_q) = 0$. Further, let U be any bounded and convex neighborhood of $\bar{z}_q$. Then for every $z \in U \cap Z_E$, there exist some $\lambda_1,\ \lambda_2 \in [0, 1]$ such that

$$\begin{aligned} J_q(z) - J_q(\bar{z}_q) &= \langle DJ(\bar{z}_q), z - \bar{z}_q\rangle + \tfrac{1}{2}\langle D^2J_q(\lambda_1\bar{z}_q + (1-\lambda_1)z)z - \bar{z}_q, z - \bar{z}_q\rangle \\ &= \tfrac{1}{2}\langle D^2J_q(\lambda_1\bar{z}_q + (1-\lambda_2)z)(z - \bar{z}_q), z - \bar{z}_q\rangle \end{aligned}$$

and

$$\begin{aligned} \langle DJ_q(z), z - \bar{z}_q\rangle &= \langle DJ_q(z) - DJ_q(\bar{z}_q), z - \bar{z}_q\rangle \\ &= \langle D^2J_q(\lambda_2\bar{z}_q + (1-\lambda_2)z)(z - \bar{z}_q), z - \bar{z}_q\rangle \end{aligned}.$$

Hence, from the convexity of J_q we have that

$$\begin{aligned} \tfrac{\alpha}{2}\|z-\bar{z}_q\|_Z^2 &\leq J_q(z)-J_q(\bar{z}_q) \\ &\leq \langle DJ_q(z), z-\bar{z}_q\rangle \\ &\leq C_U\|z-\bar{z}_q\|_Z^2 \end{aligned} \tag{6.5}$$

where $C_U := \sup\{\|D^2J_q(\bar{z}_q)\|_{\mathcal{L}(Z_E,Z_E^*)} : z \in U \cap Z_E\}$.

6.1 Newton's Method with Line Search

Here we present some basic algorithm for solving the subproblem (P_q). The algorithm we use is Newton's method with line search. In each iteration the iterates have the form $z_{n+1} = z_n + \sigma_n h_n$, where σ_n is the step length and h_n is the search direction. To determine the step length we use the Armijo Rule. It is stated as follows: Given a search direction h_n of J_q at z_n, choose the maximum $\sigma_n \in \{1, \frac{1}{2}, \frac{1}{4}, \dots\}$ for which

$$J_q(z_n+\sigma_n h_n) \leq J_q(z_n) + \gamma\sigma_n\langle DJ_q(z_n), h_n\rangle_{Z^*,Z}.$$

Here $\gamma \in (0,1)$ is a constant.

To determine the search direction, we compute the first-order optimal point h_n of the quadratic problem

$$\min_{h\in Z_E} \langle DJ_q(z_n), h\rangle_{Z^*,Z} + \frac{1}{2}\langle D^2J_q(z_n)h, h\rangle_{Z^*,Z}.$$

Since J_q is twice differentiable due to our assumptions we can use Newton's method. The Newton method with line search is stated as follows.

Algorithm (Newton Method with Line Search)

0. Choose $0 < \gamma < 1$, $z_0 \in Z_E$

   For $n = 0, 1, 2, \cdots$:

1. Compute $h_n \in Z_E$ as the solution of the quadratic problem

$$\min_{h\in Z_E} \langle DJ_q(z_n), h\rangle_{Z^*,Z} + \frac{1}{2}\langle D^2J_q(z_n)h, h\rangle_{Z^*,Z}. \tag{6.6}$$

2. Line search: Choose the step size as the maximum $\sigma_n \in \{1, \frac{1}{2}, \frac{1}{4}, \dots\}$ for which

$$J_q(z_n + \sigma_n h_n) \leq J_q(z_n) + \gamma \sigma_n \langle DJ_q(z_n), h_n \rangle_{Z^*,Z}.$$

3. Set $z_{n+1} = z_n + \sigma_n h_n$.

Note that the search direction $h_n \in Z_E$ in step 1 of the above algorithm can be solved by using the optimality system:
Find a stationary point $(h, p) \in Z \times V$ of

$$\langle DJ_q(z_n), h \rangle_{Z^*,Z} + \frac{1}{2} \langle D^2 J_q(z_n)h, h \rangle_{Z^*,Z} + \langle Eh, p \rangle.$$

Which leads to a KKT system

$$\begin{aligned} D^2 J_q(z_n)h + E^* p &= -DJ_q(z_n) \\ Eh &= 0. \end{aligned}$$

It is uniquely solvable due to (6.5) and the Lax-Milgram lemma, see e.g. [2]. The following theoretical result and its proof are given in Gfrerer [1].

Theorem 6.3 *Assume that the assumptions (P1)–(P5),(Q1),(Q2) and (PQ2) are fulfilled and let z_n be generated by Newton's method with line search. Then* $\limsup_{n\to\infty} z_n = \bar{z}_q$.

If the objective function J_q of the subproblem (P_q) is twice continuously differentiable on Z_E, then the rate of convergence of the above algorithm will be q-superlinear. That is, there exists an increasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\lim_{t\to 0_+} \omega(t) = 0$ such that $\|z_{n+1} - \bar{z}_q\|_{Z_E} \leq \omega(\|z_n - \bar{z}_q\|_{Z_E}) \|z_n - \bar{z}_q\|_{Z_E}$. More precisely,

$$\lim_{n\to\infty} \frac{\|z_{n+1} - \bar{z}_q\|_{Z_E}}{\|z_n - \bar{z}_q\|_{Z_E}} = 0.$$

If J_q is not twice continuously differentiable, the convergence will be only q-linear, that is

$$\lim_{n\to\infty} \frac{\|z_{n+1} - \bar{z}_q\|_{Z_E}}{\|z_n - \bar{z}_q\|_{Z_E}} = c$$

where $0 < c < 1$. For the model problem (MP) the objective function J_q is not twice continuously differentiable. Hence we can't expect q-superlinear convergence. To have a q-superlinear convergence we need the modification of the Newton's method with line search.

In the following section we consider a modification of the algorithm in such a way that we can achieve superlinear convergence in case of $\Psi_{i,q}$ is not twice continuously differentiable on Z_E for some $i \in \{1, \ldots, m\}$. We call this modification by a *Smoothed Newton step.*

6.2 Smoothed Newton Step

Before stating the smoothed Newton step we need the following additional assumptions on the structure of the problem.

Assumption:

(P6a) Suppose there exist some integers $m', \tilde{d}$, a finite measure space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mu})$, some real numbers γ_i, a linear operator $T \in \mathcal{L}(Z, L^2(\tilde{\Omega})^{\tilde{d}})$ and compact operators $C_i \in \mathcal{L}(Z, L^2(\tilde{\Omega})^{\tilde{d}})$ such that $d_i = \tilde{d}$, $(\Omega_i, \mathcal{A}_i, \mu_i) = (\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mu})$, $A_i = \gamma_i T + C_i$ for all $i = 1, \ldots, m'$. Further we assume $\Psi_{i,q}$ is twice continuously differentiable on Z_E for all $i = m' + 1, \ldots, m$.

(P6b) The mapping $\mathcal{H} : Z \to V \times L^2(\tilde{\Omega})^{\tilde{d}}$ defined by $\mathcal{H}(z) = (Ez, Tz)$ is surjective.

Denote the topological complement space of Ker T by $\mathcal{U} \subset Z$. For each $z \in Z_E$ we denote by $D\hat{J}(z) \in Z^*$ some continuous extension of the linear operator $DJ(z) \in Z_E^*$ and by $D^2\hat{J}(z) \in \mathcal{L}(Z, Z^*)$ some continuous extension of $D^2J(z) \in \mathcal{L}(Z_E, Z_E^*)$ such that $\langle D^2\hat{J}(z)h, k\rangle = \langle D^2\hat{J}(z)k, h\rangle$, $\forall h, k \in Z$ and $\langle D^2\hat{J}(z)h, h\rangle \geq \|h\|_Z^2, \forall h \in \mathcal{U}$.

For each $z \in Z_E$ we define the quadratic functions F_z and G_z on Z by

$$F_z(h) := \langle D\hat{J}(z), h\rangle + \frac{1}{2}\langle D^2\hat{J}(z)h, h\rangle + \sum_{i=m'+1}^{m} (\langle D\Psi_{i,q}(z), h\rangle + \frac{1}{2}\langle D^2\Psi_{i,q}(z)h, h\rangle),$$

and

$$G_z(h) := F_z(h) + \sum_{i=1}^{m'}(\langle D\Psi_{i,q}(z), h\rangle + \frac{1}{2}\langle D^2\Psi_{i,q}(z)h, h\rangle)$$

respectively. Note that the quadratic function G_z satisfies $\langle DG_z(0), h\rangle = \langle DJ_q(z), h\rangle$ and $\langle D^2G_z(0)h, k\rangle = \langle D^2J_q(z)h, k\rangle$. Let's now consider the smoothed Newton step, it consists of the following steps.

1. Given an iterate z_n, compute a solution $\zeta_1 \in Z$ of the quadratic problem

$$\begin{aligned} &\min_{\zeta} G_{z_n}(\zeta) \\ &s.t. \\ &\mathcal{H}\zeta = 0 \end{aligned}$$

with a multiplier $(v_1^*, v^*) \in V^* \times (L^2(\tilde{\Omega})^{\tilde{d}})^*$ such that

$$DG_{z_n}(\zeta_1) + E^*v_1^* + T^*v^* = 0.$$

2. Compute a solution $\zeta_2 \in \mathcal{U}$ of the nonlinear problem

$$\begin{aligned} \min_{\zeta\in\mathcal{U}} T_n(\zeta) := \langle E^*v_1^*, \zeta\rangle + F_{z_n}(\zeta_1 + \zeta) + \sum_{i=1}^{m'} \Big(\int_{\Omega_i} \pi_i((A_i\tilde{z}_n + \gamma_i T\zeta)(x), \varphi_i(x)) \\ + \langle D_s\pi_i(A_i\tilde{z}_n(x), \varphi_i(x)), C_i\zeta(x)\Big)\rangle d\tilde{\mu}(x), \end{aligned}$$

where $\tilde{z}_n = z_n + \zeta_1$.

3. Compute a solution $\zeta_3 \in Z$ of the quadratic problem

$$\begin{aligned} \min_{\zeta\in Z} S_n(\zeta) := F_{z_n}(\zeta_1 + \zeta_2 + \zeta) + \sum_{i=1}^{m'}(\langle D\Psi_{i,q}(\hat{z}_n), \zeta\rangle + \frac{1}{2}\langle D^2\Psi_{i,q}(\hat{z}_n)\zeta, \zeta\rangle) \\ s.t. \\ E(\zeta_2 + \zeta) = 0, \end{aligned}$$

where $\hat{z}_n = z_n + \zeta_1 + \zeta_2$.

4. Set $z_{n+1} = z_n + \zeta_1 + \zeta_2 + \zeta_3$.

Lemma 6.4
$\|\zeta_1\|_Z + \|\zeta_2\|_Z + \|\zeta_3\|_Z + \|\nu^*\|_{L^2(\Omega)^d} + \|v_3^* - v_1^*\|_{V^*} = \mathcal{O}(\|z_n - \bar{z}_q\|_{Z_E})$.

Theorem 6.5 *Assume that the assumptions (P1)–(P6),(Q1),(Q2) and (PQ2) are fulfilled and assume that there exists a bounded set $\tilde{\mathcal{K}} \subset L^2(\tilde{\Omega})^d$ which is either bounded in $L^{\hat{r}}(\tilde{\Omega})^{\tilde{d}}, \hat{r} > 2$ or the elements are of the form $\tilde{k}(x) = R(x)k(x)$ where k belongs to a compact subset $\mathcal{K} \subset L^2(\tilde{\Omega})^{\tilde{d}}$ and R is from a bounded subset $\mathcal{R} \subset L^\infty(\Omega)^{\tilde{d}\times\tilde{d}}$ such that for the smoothed Newton steps we have dist$(T\zeta_3, \|T\zeta_3\|\bar{\mathcal{K}}) = o(\|z_n - \bar{z}_q\|_Z)$ for all $z_n \in Z_E$ in some neighborhood of $\bar{z}_q$. Then there exists an increasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\lim_{t\to 0_+} \omega(t) = 0$ such that*

$$\|z_{n+1} - \bar{z}_q\|_Z \leq \omega(\|z_n - \bar{z}_q\|_Z)\|z_n - \bar{z}_q\|_Z$$

where z_{n+1} is the outcome of the smoothed Newton step.

6.3 Application to the model problem

In this section we present the application of smoothed Newton step to the model problem. In the model problem we have that $A_1(y,u) = u$, $A_2(y,u) = y$. Note that A_1 and A_2 are the projection of $Z = H^1(\Omega) \times L^2(\partial\Omega)$ into $L^2(\partial\Omega)$ and $H^1(\Omega)$ respectively. Since Ω is assumed to be bounded and Lipschitz domain by Theorem 7.15, ([9], Chapter 7) the Hilbert space $H^1(\Omega)$ is compactly embedded in $L^2(\Omega)$. Hence, the operator $A_2(y,u) = y$ is definitely compact. Consequently the mapping $\Psi_{2,q}$ is twice continuously differentiable on Z_E by Corollary 6.2. In other hand the operator $A_1(y,u) = u$ do not satisfy the assumptions of Corollary 6.2 and hence we have no twice continuously differentiability of $\Psi_{1,q}$ on Z_E. However, we can decompose it as in assumption (P6):

$$A_1 = \gamma_1 T + C_1$$

with $m' = 1, T(y,u) := A_1(y,u) = u, C_1(y,u) = 0, \gamma_1 = 1$ and $\mathcal{U} := \{0\} \times L^2(\partial\Omega)$.

For each $z \in Z_E$ define the quadratic functions F_z and G_z on Z for the model problem (MP) by

$$\begin{aligned} F_z(h) &:= \langle DJ(z), h\rangle + \tfrac{1}{2}\langle D^2J(z)h, h\rangle + \langle D\Psi_{2,q}(z), h\rangle + \tfrac{1}{2}\langle D^2\Psi_{i,q}(z)h, h\rangle) \\ &= \int_\Omega \Big((y-y_d)h_y + \frac{1}{2}h_y^2 + D_s\pi_2(y,\varphi_y)h_y + \frac{1}{2}D^2_{ss}\pi_2(y,\varphi_y)h_y^2\Big) + \\ &\quad \int_{\partial\Omega} \beta\Big((u-u_d)h_u + \frac{1}{2}h_u^2\Big) \end{aligned}$$

and

$$\begin{aligned} G_z(h) &:= F_z(h) + \langle D\Psi_{1,q}(z), h\rangle + \tfrac{1}{2}\langle D^2\Psi_{1,q}(z)h, h\rangle \\ &= \int_\Omega \Big((y-y_d)h_y + \frac{1}{2}h_y^2 + D_s\pi_2(y,\varphi_y)h_y + \frac{1}{2}D^2_{ss}\pi_2(y,\varphi_y)h_y^2\Big) + \\ &\quad \int_{\partial\Omega} \Big(\beta((u-u_d)h_u + \frac{1}{2}h_u^2) + D_s\pi_1(u,\varphi_u)h_u + \frac{1}{2}D^2_{ss}\pi_1(u,\varphi_u)h_u^2\Big) \end{aligned}$$

for $h = (h_y, h_u) \in Z$.

Now let us consider the smoothed Newton step at a given iterate $z_n = (y_n, u_n)$. In the first step we have to compute a solution $\zeta_1 = (\zeta_{1,y}, \zeta_{1,u}) \in Z$ of the quadratic problem

$$\begin{aligned} \min_{\zeta_y,\zeta_u} &\int_\Omega \Big((y_n-y_d)\zeta_y + \frac{1}{2}\zeta_y^2 + D_s\pi_2(y_n,\varphi_y)\zeta_y + \frac{1}{2}D^2_{ss}\pi_2(y_n,\varphi_y)\zeta_y^2\Big) + \\ &\int_{\partial\Omega} \Big(\beta((u_n-u_d)\zeta_u + \frac{1}{2}\zeta_u^2) + D_s\pi_1(u_n,\varphi_u)\zeta_u + \frac{1}{2}D^2_{ss}\pi_1(u_n,\varphi_u)\zeta_u^2\Big) \\ s.t.& \\ &-\Delta\zeta_y + c\zeta_y = 0 \text{ in } \Omega, \\ &\tfrac{\partial\zeta_y}{\partial\nu} = \zeta_u \text{ on } \partial\Omega, \\ &\zeta_u = 0 \text{ on } \partial\Omega. \end{aligned}$$

Obviously, the constraint set has only one feasible point $(0,0)$. Therefore, $\zeta_1 = (0,0)$ is the minimizer of the problem. The multiplier $v_1^* \in H^1(\Omega)$ corresponding to ζ_y is the unique solution of the variational problem

$$\int_\Omega \Big(\langle \nabla v_1^*, \nabla v\rangle + cv_1^*v + (y_n-y_d)v + D_s\pi_2(y_n,\varphi_y)v\Big) = 0 \ \forall v \in H^1(\Omega).$$

In the second step we have to find a minimizer $\zeta_2 = (\zeta_{2,y}, \zeta_{2,u}) \in \mathcal{U}$ (i.e. $\zeta_{2,y} = 0$ and $\zeta_{2,u} \in L^2(\partial\Omega)$) of the nonlinear problem

$$\min_{\zeta_u} \int_{\partial\Omega} \big(-v_1^*\zeta_u + \beta((u_n-u_d)\zeta_u + \frac{1}{2}\zeta_u^2) + \pi_1(u_n+\zeta_u,\varphi_u)\big) \ .$$

Which can be solved pointwise, that is, for each $x \in \partial\Omega$ the value $\zeta_{2,u}(x)$ is a minimizer of the following one-dimensional unconstrained convex programming problem

$$\min_{\zeta\in\mathbb{R}} -v_1^*(x)\zeta + \beta((u_n(x) - u_d(x))\zeta + \frac{1}{2}\zeta^2) + \pi_1(u_n(x) + \zeta, \varphi_u(x)). \tag{6.7}$$

In the third step we solve the solution $\zeta_3 = (\zeta_{3,y}, \zeta_{3,u}) \in Z$ of the following quadratic problem

$$\begin{aligned}
\min_{\zeta_y,\zeta_u} &\int_\Omega \Big((y_n - y_d)\zeta_y + \frac{1}{2}\zeta_y^2 + D_s\pi_2(y_n, \varphi_y)\zeta_y + \frac{1}{2}D_{ss}^2\pi_2(y_n, \varphi_y)\zeta_y^2\Big) \\
&+ \int_{\partial\Omega} \Big(\beta((u_n - u_d)(\zeta_{2,u} + \zeta_u) + \frac{1}{2}(\zeta_{2,u} + \zeta_u)^2) \\
&\quad + D_s\pi_1(u_n + \zeta_{2,u}, \varphi_u)\zeta_u + \tfrac{1}{2}D_{ss}^2\pi_1(u_n + \zeta_{2,u}, \varphi_u)\zeta_u^2\Big) \\
s.t.& \\
&-\Delta\zeta_y + c\zeta_y = 0 \text{ in } \Omega, \\
&\tfrac{\partial\zeta_y}{\partial\nu} = \zeta_{2,u} + \zeta_u \text{ on } \partial\Omega,
\end{aligned}$$

The corresponding multiplier $v_3^* \in H^1(\Omega)$ is a solution of the following equations:

$$\begin{aligned}
\int_\Omega \langle\nabla v_3^*, \nabla v\rangle + (cv_3^* + y_n - y_d + \zeta_{3,y} + D_s\pi_2(y_n, \varphi_y) + \\
D_{ss}^2\pi_2(y_n, \varphi_y)\zeta_{3,y})v = 0 \ \ \forall v \in H^1(\Omega),
\end{aligned} \tag{6.8}$$

and

$$\begin{aligned}
\beta(u_n - u_d + \zeta_{2,u} + \zeta_{3,u}) + D_s\pi_1(u_n + \zeta_{2,u}, \varphi_u) + \\
\tfrac{1}{2}D_{ss}^2\pi_1(u_n + \zeta_{2,u}, \varphi_u)\zeta_{3,u} - v_3^* = 0 \ \ \forall x \in \partial\Omega.
\end{aligned} \tag{6.9}$$

Finally, we update the iterate by setting $z_{n+1} = z_n + \zeta_1 + \zeta_2 + \zeta_3$.

From first order optimality condition of (6.7) we obtain

$$-v_1^*(x) + \beta(u_n(x) - u_d(x) + \zeta_{2,u}(x)) + D_s\pi_1(u_n(x) + \zeta_{2,u}(x), \varphi_u(x)) = 0 \ \ \forall x \in \partial\Omega \tag{6.10}$$

and hence from (6.9) we deduce

$$(\beta + D_{ss}^2\pi_1(u^n(x) + \zeta_{2,u}(x), \varphi_u(x)))\zeta_{3,u}(x) = v_3^*(x) - v_1^*(x), \ \ x \in \partial\Omega.$$

From the convexity assumption of $\pi_1(s,t)$ with respect to s we have that $D_{ss}^2\pi_1(u^n(x) + \zeta_{2,u}(x), \varphi_u(x)) \geq 0$ and hence

$$\zeta_{3,u}(x) = \frac{v_3^*(x) - v_1^*(x)}{(\beta + D_{ss}^2 \pi_1(u^n(x) + \zeta_{2,u}(x), \varphi_u(x)))}, \quad x \in \partial\Omega.$$

Since $H^1(\Omega)$ is compactly embedded in $L^2(\Omega)$, together with Lemma 6.4, the assumptions of Theorem 6.5 are satisfied. Therefore, the q-superlinear convergence of the smoothed Newton step is assured.

Chapter 7

Discretization

The Finite Element Method is a general discretization technique for the numerical solution of partial differential equation related problems. In this chapter we will see how this technique can be employed to discretize our optimal control problem.

The following are some of the basic futures of the Finite Element Methods (for the detailed description see ([10],Chapter 4)):

i. The physical region of the problem is subdivided into non overlapping subregions, called *finite elements*.

ii. The solution of the governing equations is over each element approximated by polynomial functions (constants, linear, quadratic, etc). The global solution is approximated by piecewise polynomial using thus local approximates and the coefficients of this piecewise polynomial becomes the unknown of the discretized problem.

iii. Substitution of the approximats into the governing equation gives a system of equations in the unknown parameters whose matrix, by construction, is sparse.

To simplify the discretization we assume that the domain $\Omega \subset \mathbb{R}^n$ is a polyhedron. For our

model problem $\Omega \subset \mathbb{R}^2$ is $\Omega = (0,1)^2$, a unit square in $\mathbb{R}^2$. We discretize Ω into small triangles, denoted by T_i for $i = 1, \ldots, n$. In this discretization the construction of the small triangles is in such away that the vertices on the boundary $\partial\Omega$ of Ω coincide with the vertices of some of the small triangles. The individual triangles are called the *triangular elements* of the FEM and the subdivision of the domain into small triangles is called *triangulation*.

Note that triangulation is not the only way to discretize the domain. For example quadrilaterals in 2D and tetrahedron and parallelepiped in 3D can be used for discretization of a domain.

Definition(Triangulation) Let $\Omega \subset \mathbb{R}^2$ be a polyhedral domain. A subdivision $\mathcal{T}_h = \{T_1, \ldots, T_n\}$ of $\bar{\Omega}$ into triangular elements is called *admissible* if the following properties hold:

1) $\bar{\Omega} = \cup_{T \in \mathcal{T}_h} T$.

2) For each $T_i, T_j \in \mathcal{T}_h$, $int T_i \cap T_j = \emptyset$.

3) Every face of any element $T_j \in \mathcal{T}_h$ is either a face of another element T_i or a part of the boundary $\partial\Omega$.

Where $h = \max_{1 \leq i \leq n} diam(T_i)$ denotes the maximum diameter of all $T \in \mathcal{T}_h$.
A set $\mathcal{T} = \{\mathcal{T}_h\}$ of triangulations of $\bar{\Omega}$ is said to be a family of triangulations if for any $\epsilon > 0$ there exists some $\mathcal{T}_h \in \mathcal{T}$ with $h < \epsilon$.

Having such a triangulation of the domain Ω, we can construct finite dimensional function spaces that approximates the infinite dimensional spaces that govern the control problem. Here we illustrate the simplest *finite element space* in arbitrary space dimensions, it is called

the *space of linear mappings.* The elements are called *linear elements.*

Let $x_i,\ i = 1, \ldots, N$ be the set of all vertices of the triangulation $\mathcal{T}_h$. Thus the space of all continuous piecewise linear functions V_h defined on $\mathcal{T}_h$ is a finite dimensional space with dimension N. The space V_h is called finite element space. A basis of V_h is given by the set $\{\phi_i\}_{i=1}^N$ where ϕ_i for $i = 1, \ldots, N$ satisfies

$$\phi_i(x_j) = \delta_{ij} \text{ for } \ i, j = 1, \ldots, N.$$

Note that ϕ_i vanishes identically in every element of the grid except the element having the vertex x_i. Clearly the space V_h is subset of $C(\bar{\Omega})$. Each element $v \in V_h$ has the form

$$v = \sum_{i=1}^{N} a_i \phi_i$$

and it is uniquely identified by the vector $a = (a_i) \in \mathbb{R}^N$.

In the model problem (MP) the space $H^1(\Omega)$ and $L_2(\partial\Omega)$ are both infinite dimensional. In the numerical solution of the problem we approximate both spaces by finite dimensional spaces using FEM.

Let $V_h \subset H^1(\Omega)$ be a conforming finite element space with basis of functions ϕ_i, $i = 1, \ldots, N$, i.e.

$$V_h = Span\{\phi_i\}_{i=1}^N.$$

Similarly, let $W_h \subset L^2(\partial\Omega)$ be a conforming finite element space with basis of functions φ_i, $i = 1, \ldots, M$, i.e.

$$W_h = Span\{\varphi_i\}_{i=1}^M.$$

Thus we have the following discretization of the model problem (MP)

$$\begin{aligned} \min\, & J(y_h, u_h) := \tfrac{1}{2}\|y_h - y_d\|^2 + \tfrac{\beta}{2}\|u_h - u_d\|^2 \\ \text{s.t.} \quad & \\ & y_h \in V_h,\ a(y_h, v_h) = (u_h, v_h) \text{ for all } v_h \in V_h, \\ & y_h \in Y_h \subset V_h, \\ & u_h \in U_h \subset W_h. \end{aligned}$$

Here Y_h and U_h are given by

$$Y_h := \{y_h \in V_h : y_h(x) \leq y_b(x) \text{ a.e in } \Omega\}$$

and

$$U_h := \{u_h \in W_h : u_h(x) \leq u_b(x) \text{ a.e in } \partial\Omega\}$$

respectively.

Chapter 8

Numerical results

In this chapter we present the numerical results of the model problem. We mainly focus on the practical performance of our method. Numerical results are given for two different examples. For both examples, the parameter space Q consists of 3 parameters $q = (\kappa, \epsilon_u, \epsilon_y)$. For each of the penalty functions $\psi_{1,q}$ and $\psi_{2,q}$ we use the combined logarithmic-quadratic function given by (3.2) with parameters (κ, ϵ_u) and (κ, ϵ_y) respectively.

With each point $z = (y, u)$ we associate the following values

$$\begin{aligned}
\tau_J(z) &:= 10^4 h^{-2} \frac{\sum_{i=1}^{2} \int_{\Omega_i} \psi_{i,q}^{\#}(g_i(z)(x)-\varphi_i(x), -(g_i(z)(x)-\varphi_i(x)))}{\sqrt{10^{-8}+\|y\|^2_{H^1(\Omega)}+\|u\|^2_{L^2(\partial\Omega)}}}, \\
\tau_y(z) &:= 100 h^{-2} \frac{\|(y-\varphi_y)^+\|_{L^\infty(\Omega)}}{10^{-4}+\|y\|_{H^1(\Omega)}}, \\
\tau_u(z) &:= 100 h^{-2} \frac{\|(u-\varphi_u)^+\|_{L^2(\partial\Omega)}}{10^{-4}+\|u\|_{L^2(\partial\Omega)}}, \\
\tau(z) &= \max\{\tau_J(z), \tau_y(z), \tau_u(z)\}
\end{aligned}$$

where $f^+(x) = \max\{f(x), 0\}$ for a function $f : \Omega \to \mathbb{R}$. We start the algorithm with parameters $\kappa_1 = 10h^2, \epsilon_{y,1} = \sqrt{\kappa}, \epsilon_{u,1} = \sqrt{\kappa/\beta}$ and the solution of the following elliptic PDE constrained control problem without additional constraints on the control and the state variables

$$\begin{aligned} \min_{y\in H^1(\Omega), u\in L^2(\partial\Omega)} \quad & J(z) := \tfrac{1}{2}\|y-y_d\|^2_{L^2(\Omega)} + \tfrac{\beta}{2}\|u-u_d\|^2_{L^2(\partial\Omega)} \\ \text{subject to} & \\ & -\Delta y + cy = 0 \text{ in } \Omega, \\ & \tfrac{\partial y}{\partial \nu} = u \text{ on } \partial\Omega. \end{aligned}$$

Let us denote the n-th iterate of Newton's method with line search for solving the subproblem (P_{q_k}) by $z_{k,n}$. Consider the quadratic problem (6.6) in the Newton steps, the expected decrease in the objective function for a step length $\sigma = 1$ is $\xi_{k,n} = \frac{1}{2}\langle DJ_q(z_{k,n}), h_{k,n}\rangle$. Thus we accept the iterate $z_{k,n+1}$ as approximate solution of the subproblem (P_{q_k}), if $|\xi_{k,n}| \leq (10^{-1}\min\{\frac{\tau(z_{k,n+1})-1}{10}, 1\}^3 + 10^{-14}(1-\min\{\frac{\tau(z_{k,n+1})-1}{10}, 1\}^3))(1+|J_{q_k}(z_{k,n+1})|)$.

We perform a smoothed Newton step to find the next iterate, $z_{k,n+1}$, of the Newton iterate $z_{k,n}$ only when the ratio between the L^∞ and the L^2 norm of the part of Newton's direction corresponding to the control u (i.e. h_u) exceeds 10. The approximate solution $z_{k,n+1}$ of the subproblem (P_{q_k}) is accepted as a solution of the model problem (MP), if $\tau(z_{k,n+1}) \leq 1$, i.e. the required accuracy has of order $\mathcal{O}(h^2)$.

However, if $\tau(z_{k,n+1}) \geq 1$, we solve the next subproblem $(P_{q_{k+1}})$ with parameter vector q_{k+1}. We compute the parameter vector q_{k+1} according to

$$\begin{aligned} \kappa_{k+1} &= \frac{\kappa_k}{\theta(\tau_J(z_{k,n+1}), m_k)}, \\ \epsilon_{u,k+1} &= \frac{\epsilon_{u,k}}{\theta(\tau_J(z_{k,n+1}), m_k)}, \\ \epsilon_{y,k+1} &= \frac{\epsilon_{y,k}}{\theta(\tau_J(z_{k,n+1}), m_k)}, \end{aligned}$$

where $\theta(\tau, m) = \max\{\min\{1.1\tau, m\}, 1.05\}$. The factor m_k is given by the formula

$$m_k = \max\Big\{1.2, \min\{2, 1 + \frac{m_{k-1}-1}{\min\{2, \max\{0.5, \delta_k\}\}}\}\Big\},$$

where

$$\delta_k = \max\{\frac{|J_{q_k}(z_{k,1}+h_{k,1}) - J_{q_k}(z_{k,1}) - \xi_{k,1}|}{0.1|\xi_{k,1}|}, \frac{|\xi_{k,1}|}{0.01(1+|J_{q_k}(z_{k,1})|)}\}.$$

Here, δ_k measures the quality of the first Newton step when solving the subproblem (P_{q_k}). We use the solution $z_{k-1,n+1}$ of the subproblem $(P_{q_{k-1}})$ as the starting point for the next subproblem (P_{q_k}). Using this strategy we usually need only 1 or 2 Newton steps to approximate the solution of each subproblem.

8.1 Example 1

In this example we solve the model problem (MP) with the following data:
The state bound y_b is given by

$$y_b(x_1, x_2) = \frac{1}{3}(x_1 - \frac{1}{2})^2 e^{4x_2} + x_2 + 1,$$

the control bound u_b is given by

$$u_b(x_1, x_2) = \frac{1}{4} + \frac{1}{4}\big(\cos(4\pi x_1) + \cos(4\pi x_2)\big),$$

and the desired state y_d is

$$y_d(x_1, x_2) = \frac{1}{3}(x_1 - \frac{1}{2})^2 e^{x_2} + \frac{1}{4}x_1 + 1.$$

Moreover, we have chosen the desired control $u_d(x_1, x_2) \equiv 0$ and $\beta = 0.01$. The following figures show the desired state y_d and the desired control u_d.

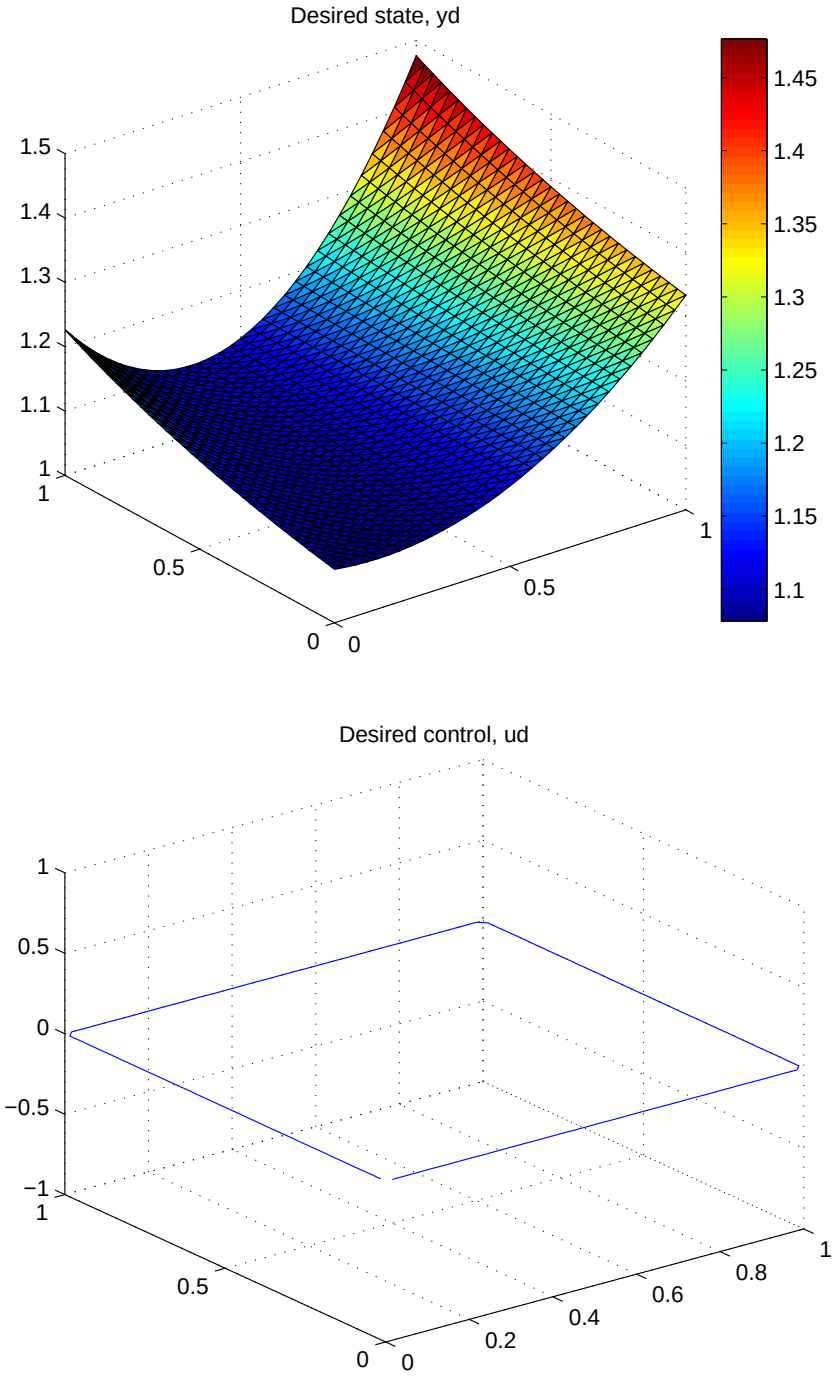

Figure 8.1: Desired state (above) and Desired control (below) for Example 1 with grid size $h = 1/32$.

The following figures show the state bound y_b and the control bound u_b.

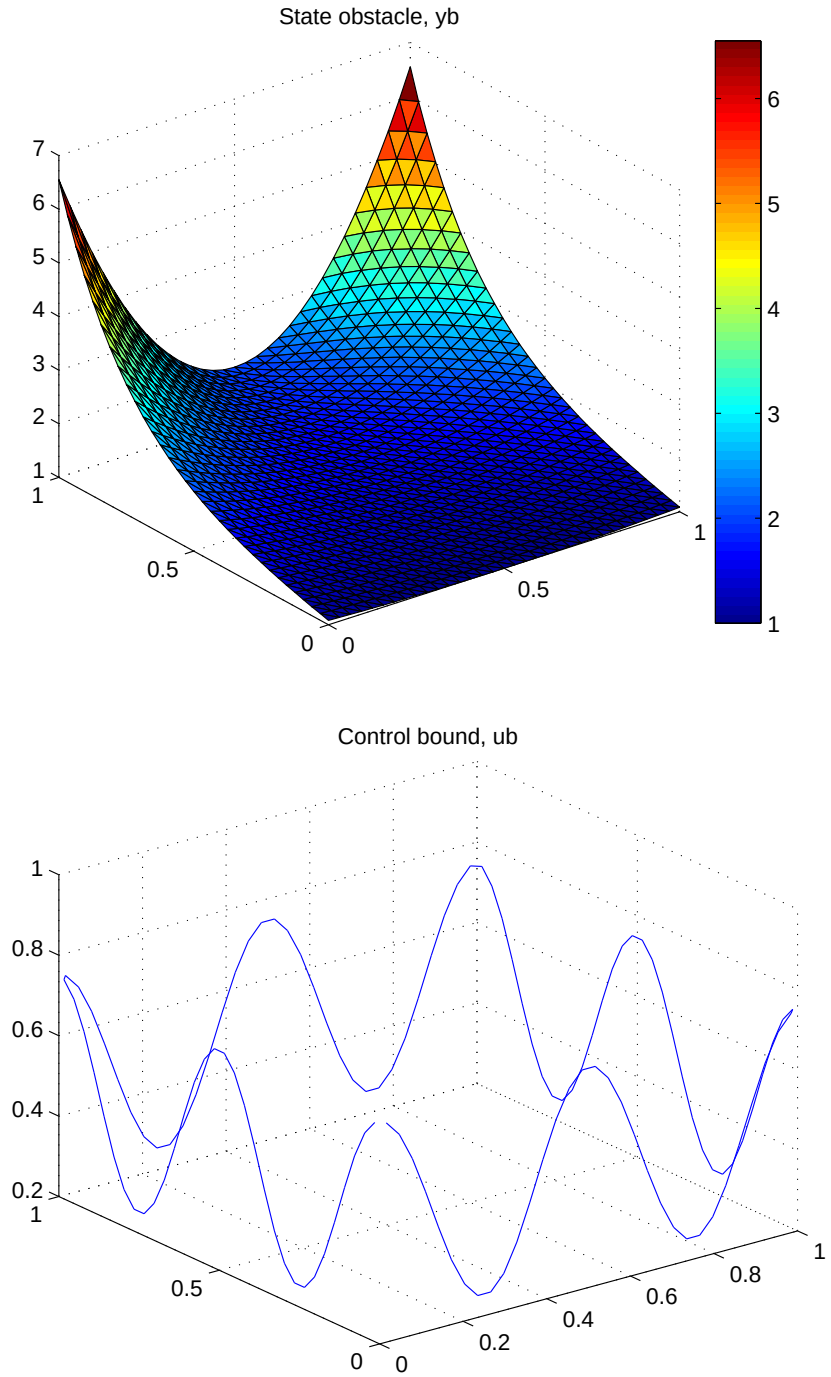

Figure 8.2: State bound (above) and Control bound (below) for Example 1 with grid size $h = 1/32$.

The following figures show the optimal state y_{opt} and the optimal control u_{opt}.

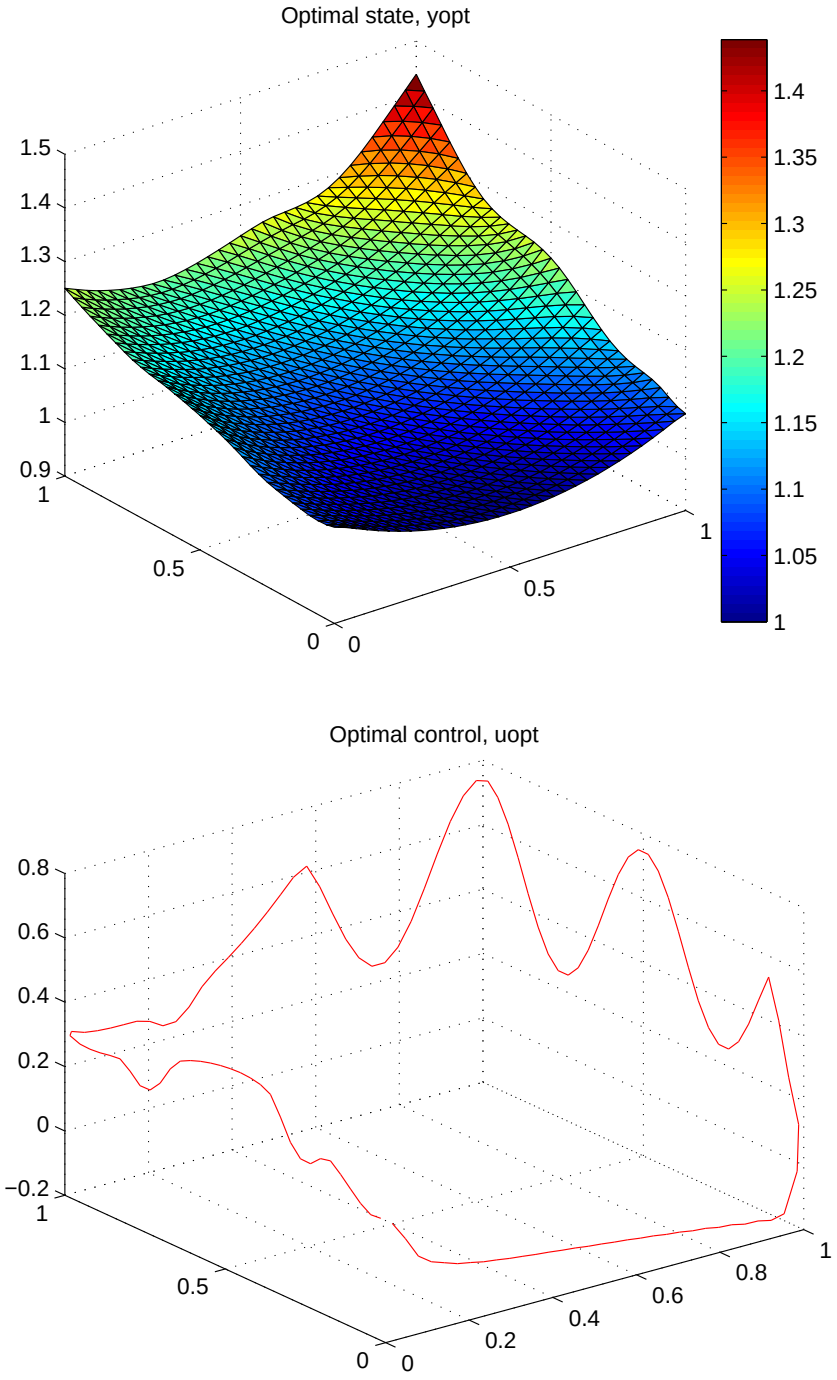

Figure 8.3: Optimal state (above) and Optimal control (below) for Example 1 with grid size $h = 1/32$.

The following figures show the differences $y_{opt} - y_b$ and $u_{opt} - u_b$. The difference is less than zero everywhere in the domain hence the solution satisfies the constraints.

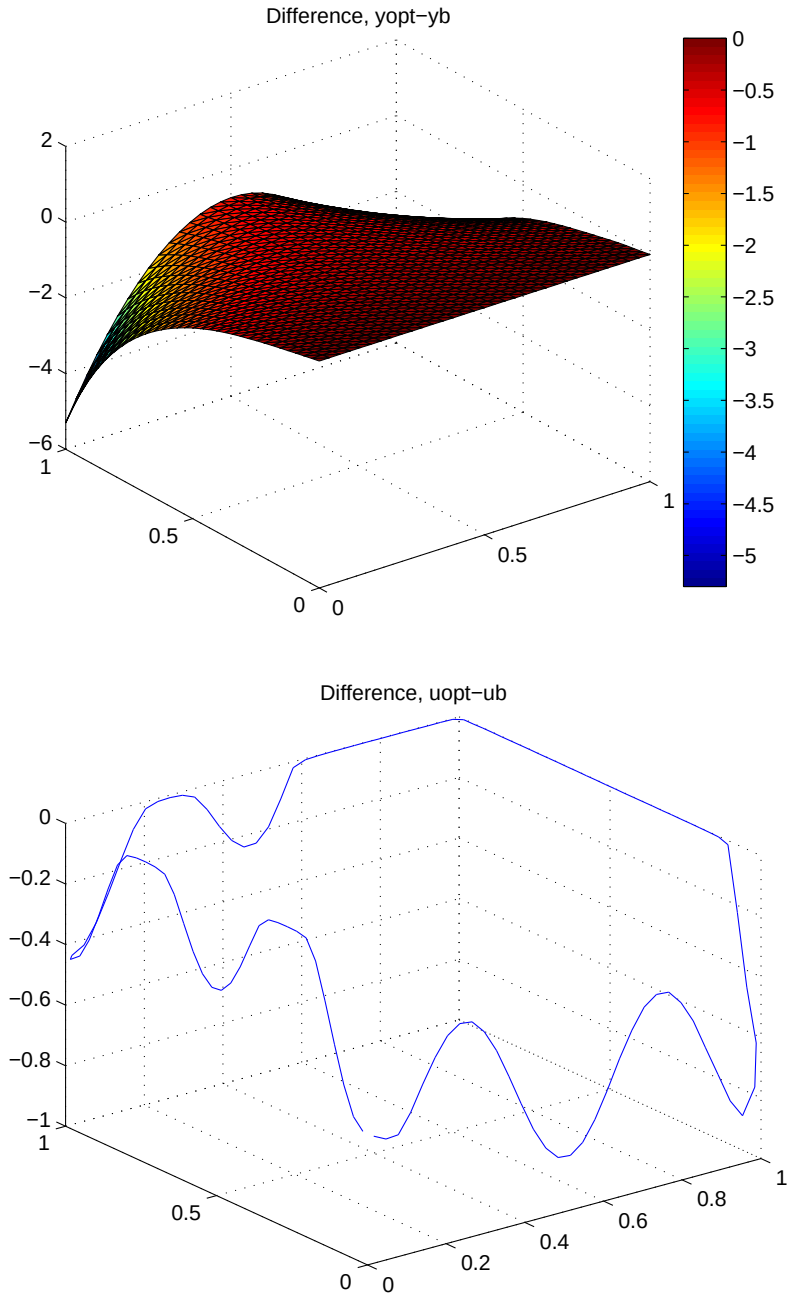

Figure 8.4: Differences $y_{opt} - y_b$ (above) and $u_{opt} - u_b$ (below) for Example 1 with grid size $h = 1/32$.

The following figures show the multipliers with respect to the state variable and the control variable respectively. It also shows that the constraints are active at the solution point.

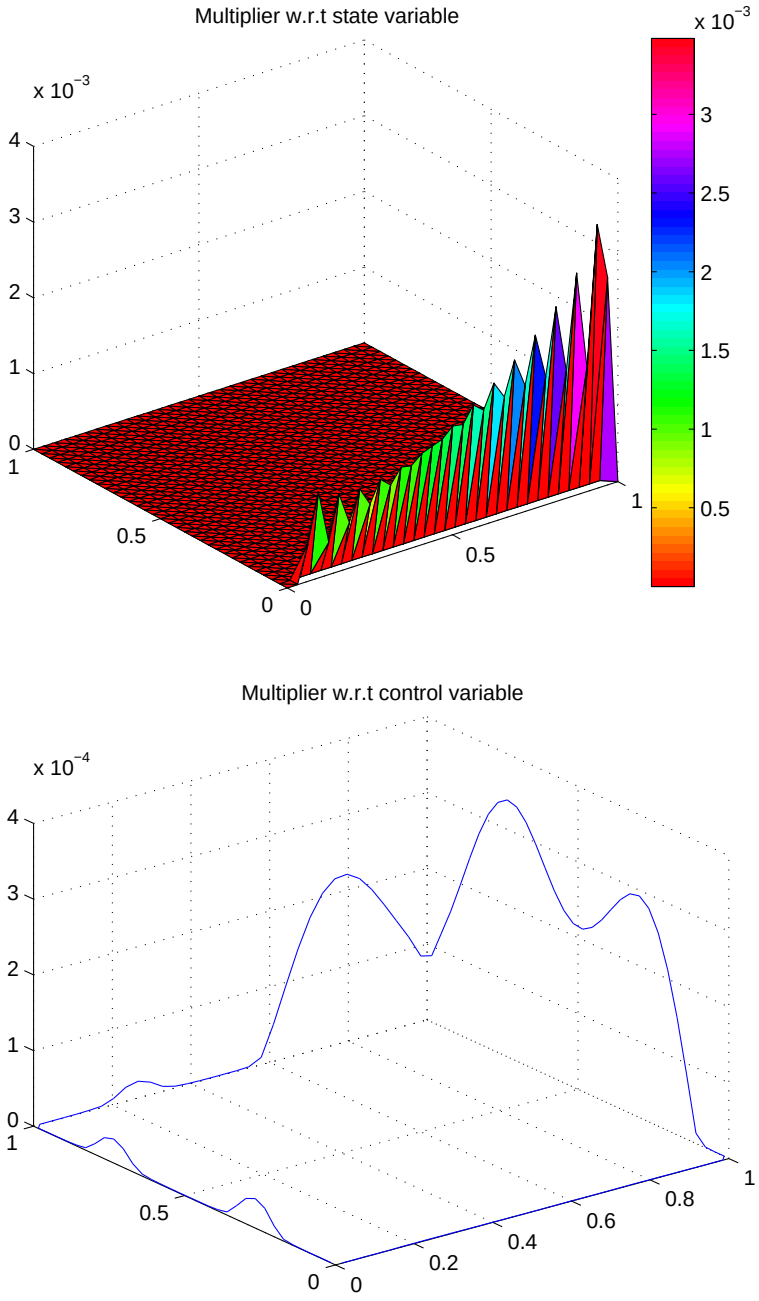

Figure 8.5: Multipliers with respect to state variable (above) and control variable (below) for Example 1 with grid size $h = 1/32$.

Grid size h	No. of Sub-problems	No. of Newton steps	No. of Smoothed Newton steps
$\frac{1}{8}$	7	12	0
$\frac{1}{16}$	10	13	0
$\frac{1}{32}$	12	13	0
$\frac{1}{64}$	13	15	1
$\frac{1}{128}$	15	16	1
$\frac{1}{256}$	17	20	1

Table 8.1: No. of Subproblems, Newton steps, Smoothed Newton steps needed to solve the (MP) for different grid sizes

8.2 Example 2

The data of this example are the same as Example 1, except the bounds on the state and the control as well as the desired state are changed as follows.

The state bound y_b is given by

$$y_b(x_1, x_2) = \frac{1}{2}(1 + \frac{1}{4}|\frac{1}{2} - x_1|),$$

the control bound u_b is given by

$$u_b(x_1, x_2) = \frac{1}{2}\cos(2\pi x_1) + \frac{1}{5}x_2,$$

and the desired state y_d is

$$y_d(x_1, x_2) = \frac{3}{2} - 2x_1x_2.$$

Moreover, we have chosen the desired control $u_d(x_1, x_2) \equiv 0$ and $\beta = 0.01$.

The following figures show the desired state y_d and the desired control u_d.

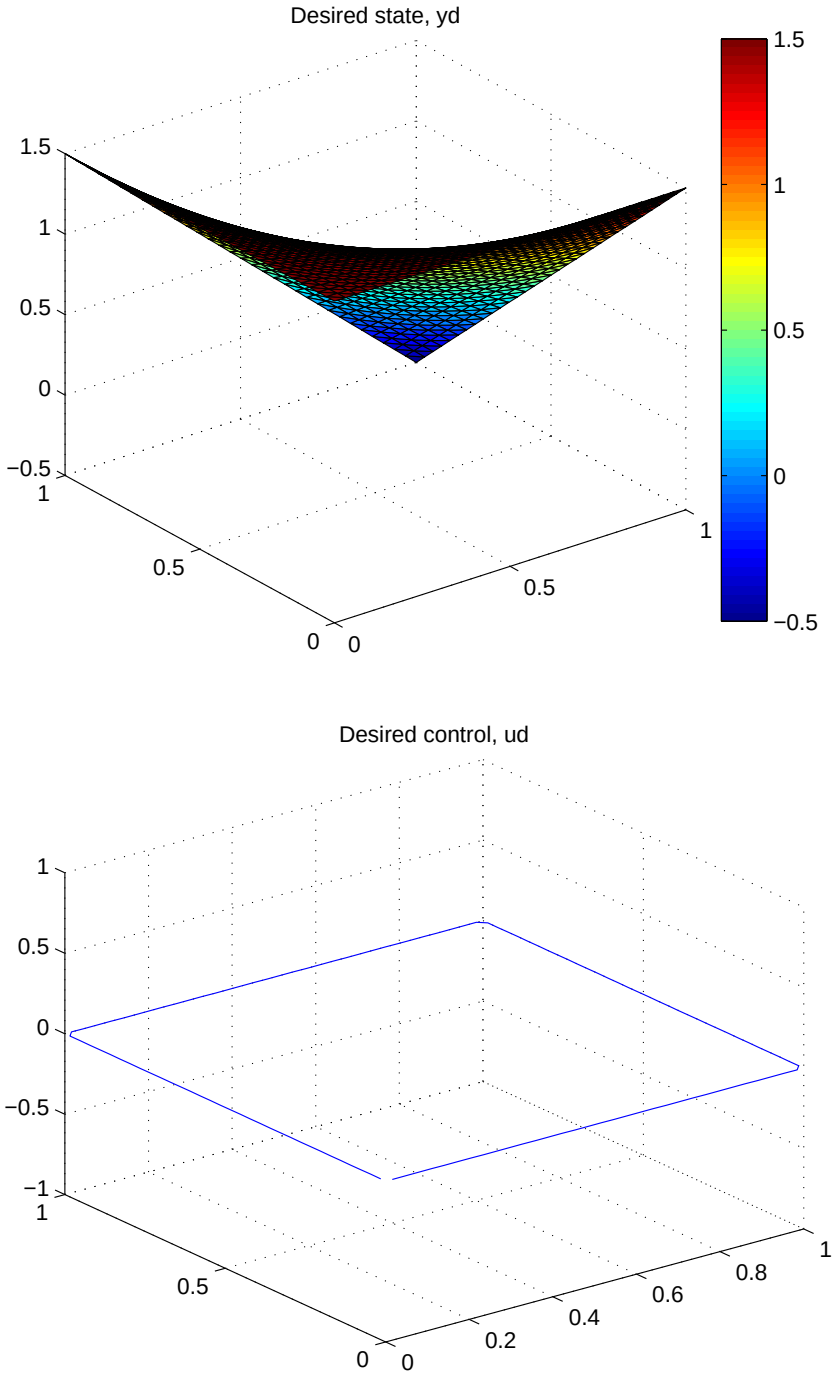

Figure 8.6: Desired state (above) and Desired control (below) for Example 2 with grid size $h = 1/32$.

The following figures show the state bound y_b and the control bound u_b.

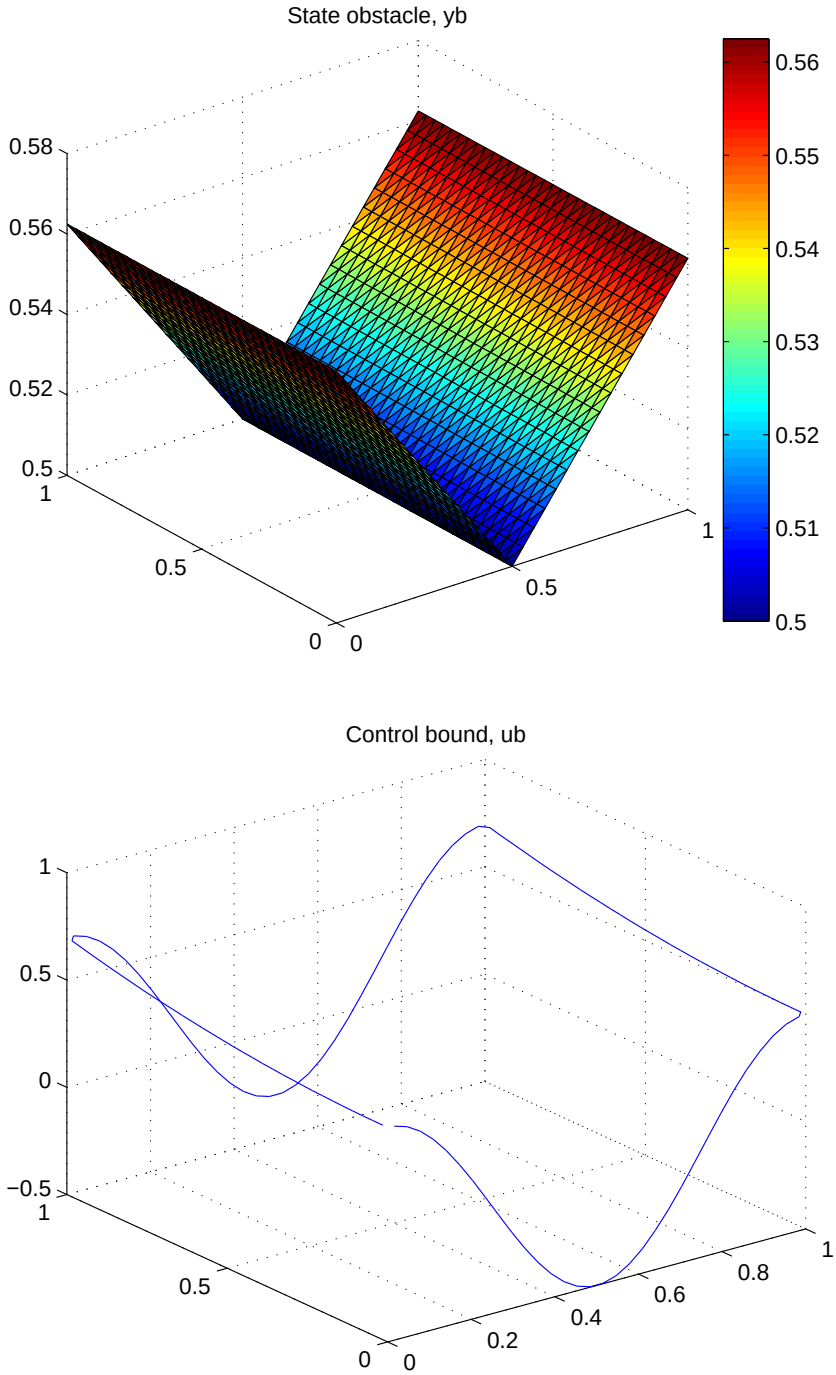

Figure 8.7: State bound (above) and Control bound (below) for Example 2 with grid size $h = 1/32$.

The following figures show the optimal state y_{opt} and the optimal control u_{opt}.

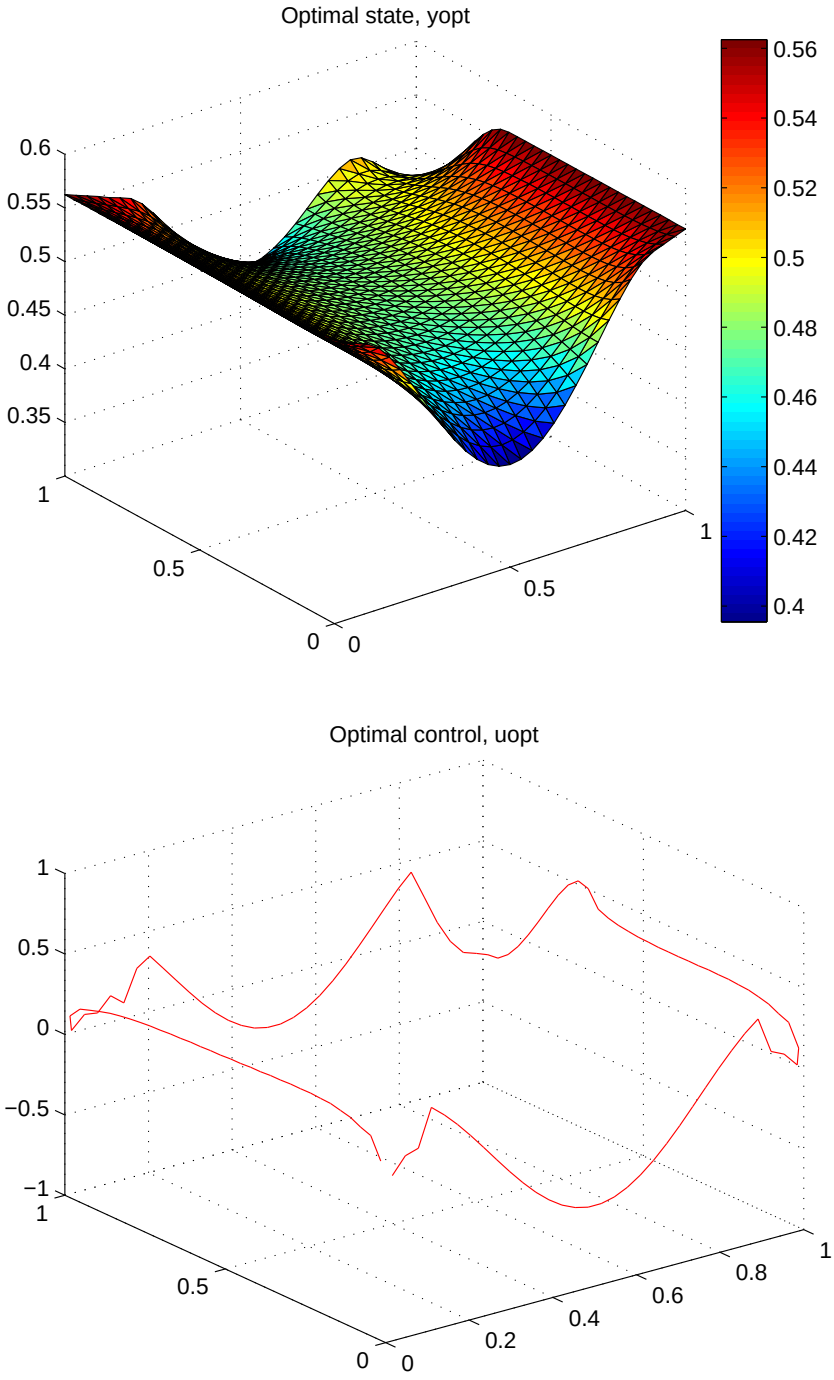

Figure 8.8: Optimal state (above) and Optimal control (below) for Example 2 with grid size $h = 1/32$.

The following figures show the differences $y_{opt} - y_b$ and $u_{opt} - u_b$. The difference is less than zero everywhere in the domain hence the solution satisfies the constraints.

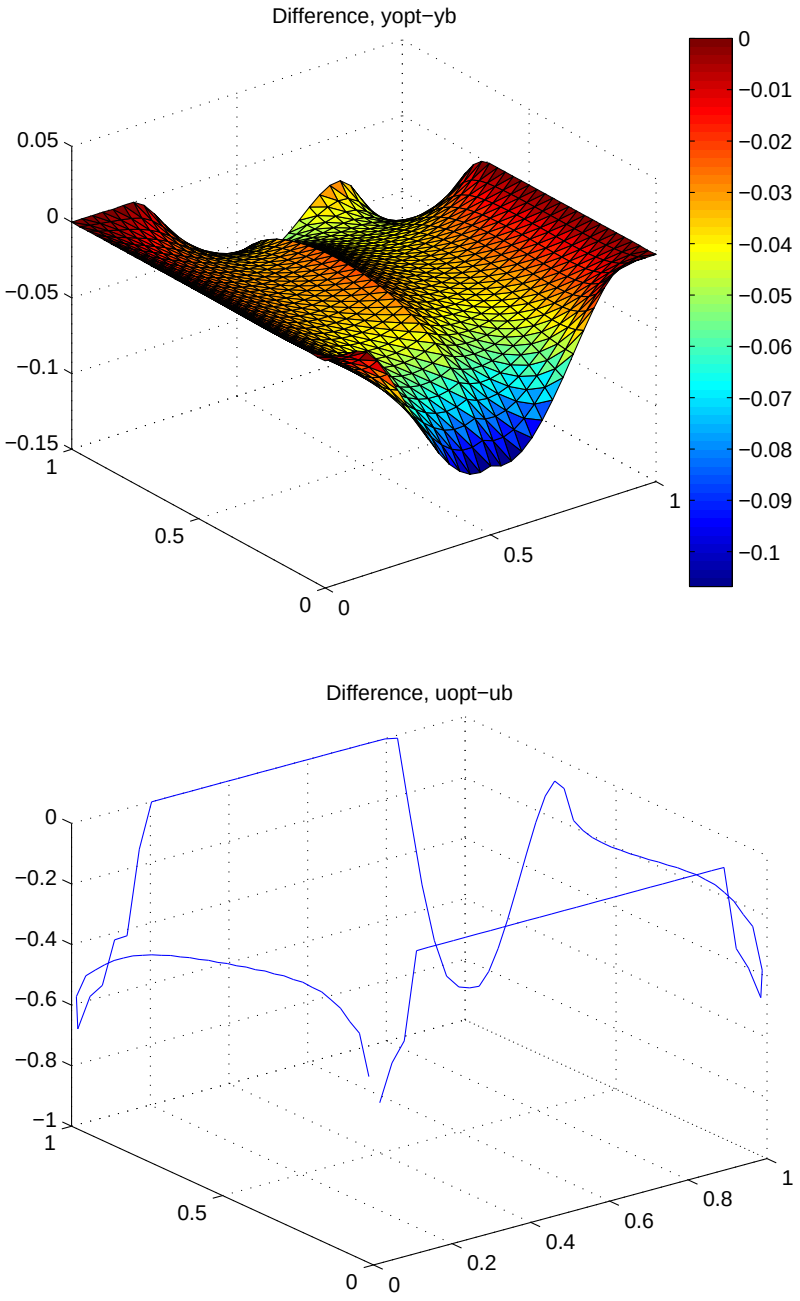

Figure 8.9: Differences $y_{opt} - y_b$ (above) and $u_{opt} - u_b$ (below) for Example 2 with grid size $h = 1/32$.

The following figures show the multipliers with respect to the state variable and the control variable respectively. It also shows that the constraints are active at the solution point.

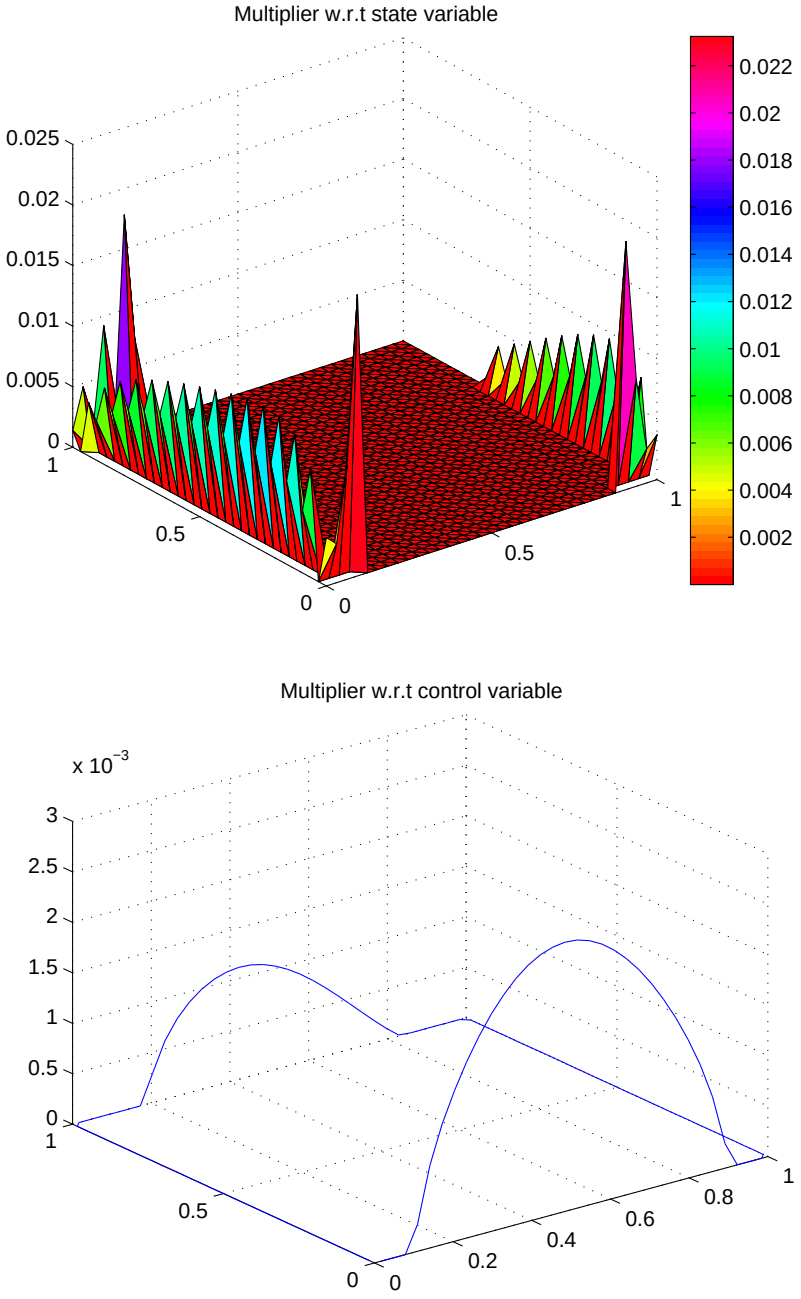

Figure 8.10: Multipliers with respect to state variable (above) and control variable (below) for Example 2 with grid size $h = 1/32$.

Grid size h	No. of Sub-problems	No. of Newton steps	No. of Smoothed Newton steps
$\frac{1}{8}$	15	20	0
$\frac{1}{16}$	17	21	0
$\frac{1}{32}$	21	23	1
$\frac{1}{64}$	25	29	1
$\frac{1}{128}$	29	32	1
$\frac{1}{256}$	35	41	2

Table 8.2: No. of Subproblems, Newton steps, Smoothed Newton steps needed to solve the (MP) for different grid sizes

In order to see the effect of the smoothed Newton step we include the results when solving the last subproblem with grid size $h = \frac{1}{256}$. Let h_y and h_u denotes the Newton direction corresponding to the state and the control respectively. The algorithm performs 6 iterations with normal Newton steps which seems to converge linearly. Then it performs one smoothed Newton step, which gives a very accurate approximation of the solution, and after performing one more smoothed Newton step the algorithm terminates. The following table shows that the effect of smoothed Newton step and shows the superlinear convergence.

Iteration	$\|h_y\|_{L^\infty(\Omega)}$	$\|h_u\|_{L^\infty(\partial\Omega)}$	$\|h_y\|_{L^2(\partial\Omega)}$
1	4.6E-4	2.1E-2	4.2E-3
2	7.7E-5	8.5E-3	6.4E-4
3	5.4E-5	3.6E-3	3.1E-4
4	1.0E-5	4.7E-3	1.5E-4
5	5.6E-6	5.9E-4	6.3E-5
6	7.9E-7	1.8E-4	1.3E-5
7^*	3.1E-7	3.8E-5	3.1E-6
8^*	7.1E-13	4.3E-11	7.3E-12

Table 8.3: Iterates for the last subproblem ($h = \frac{1}{256}$), $*$ shows smoothed Newton step.

Summery

In this work we considered pointwise state and control constrained optimal control problem governed by elliptic equationwith Neumann boundary condition. By introducing variational formation we transform the problem into infinite dimensional convex minimization problem with pointwise inequality constraints. The existence of a unique solution is presented by assuming the problem has nonempty feasible set.

Generalized penalty method is used to transform the constrained problem into a sequence of unconstrained optimization problems. The unconstrained problems are constructed by adding a term to the objective function that consists of a penalty parameter and a measure of a violation of the constraints. The convergence of the subproblems as well as the error estimates of the penalty method are discussed.

To solve the subproblems numerically we used Newton's method with line search. The convergence of this method is also included, in order to achieve a q-superlinear convergence we modified the method by introducing smoothed Newton steps. For numerical implementation of the method finite element discretization with picewise linear elements is introduced.

Bibliography

[1] H. Gfrerer, *Generalized Penalty Methods for a class of convex optimization problems with pointwise inequality constraints*, (2009), NuMa-report No.2009-06

[2] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints. Mathematical Modelling: Theory and Applications*, vol. 23, Springer Verlag, 2009.

[3] L. Biegler, O. Ghattas, M. Heinkenschloss, and B. Waanders, *Large-Scale PDEConstrained Optimization*, Lecture Notes in Computational Science and Engineering, Vol. 30, Springer-Verlag, New York, 2003.

[4] I. Ekeland, R. Temam, *Convex analysis and variational problems*, North-Holland-Elsevier, Amsterdam, 1976.

[5] C. Orozco and O. Ghattas, *Massively parallel aerodynamic shape optimization*, Comp. Syst. Eng., 1 (1992), pp. 311-320.

[6] I. Bouchouev and V. Isakov, *Uniqueness, stability and numerical methods for the inverse problem that arises in Lnancial markets, Inverse Problems*, 15 (1999), pp. R95-R116.

[7] S. R. Arridge, *Optical tomography in medical imaging*, Inverse Problems, 15 (1999), pp. R41-R93.

[8] V. Akcelik, G. Biros, and O. Ghattas, *Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation*, Proceedings of the IEEE/ACM Conference (2002), pp. 1-15.

[9] S. Salsa, *Partial differential equations in action: from modelling to theory*, Springer-Verlag, Milano, 2008

[10] P.Neittaanm, J. Sprekels, D. Tiba, *Optimization of Elliptic Systems. Theory and Applications*, Springer, Berlin, 2006

[11] Thomas S. Angell, Andreas Kirsch *Optimization methods in electromagnetic radiation.* Springer Verlag, New York, 2004.

[12] K. Atkinson and Weimin Han. *Theoretical Numerical Analysis: A Functional Analysis Framework*, 3rd edition, Springer-Verlag, New York, 2009.

Eidesstattliche Erklärung

Ich, Esubalewe Lakie Yedeg, erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Linz, July 2010

_____

Esubalewe Lakie Yedeg