Sufficient Optimality Conditions for Nonlinear Optimization Problems

Von der Fakultät für Mathematik der Universität Duisburg-Essen
(Standort Essen)
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Dissertation
von

Martin Naß

aus Kevelaer

Referent: Prof. Dr. Arnd Rösch
Koreferent: Prof. Dr. Daniel Wachsmuth

Datum der mündlichen Prüfung: 26. Februar 2014

Acknowledgements

At first, I like to thank my supervisor Prof. Dr. Arnd Rösch for the support and encouragement throughout this work. I want to thank him and the DFG-Priority Program 1253 for giving me the opportunity to work on this thesis. I also want to thank Prof. Dr. Daniel Wachsmuth for his willingness to act as a referee as well as for the fast work on his review.

Many thanks go to my present and former colleagues at the University of Duisburg-Essen. Particularly, I want to thank the members of my research group, Ute Aßmann and Hendrik Feldhordt, for the numerous fruitful discussions and the positive working atmosphere.

Finally, I want to thank my family for their support and help during the work on this thesis.

Contents

Chapter 1

Introduction

The topic of optimization, especially optimal control governed by partial differential equations, contains a large field of mathematical disciplines reaching from the foundations of functional analysis into the depths of numerical mathematics. It deals with the theoretical aspects of optimization, such as existence/uniqueness of solutions or necessary and sufficient optimality conditions, as well as the numerical implementation and the acompanying aspects of a priori/a posteriori error analysis, stability of solutions, conditions for and rate of convergence and many more. Due to this broad spectrum of aspects it is influenced by many different mathematical communities resulting in a huge variety of approaches and ideas.

Optimality conditions were always a point of interest and with the step from convex optimization problems to differentiable but non-convex settings the necessary conditions where no longer sufficient. It was necessary to consider sufficient conditions of higher order. Standard sufficient optimality conditions for finite dimension employ differentiability of $f(\bar{u})$ and that $f''(\bar{u})$ is positive definite at a local minimum $\bar{u}$. If one wants to adept these conditions to the infinite dimensional case one is often confronted with the following problem:

If one considers the functional f in an $L^2(\Omega)$-space it satisfies that the second derivative $f''(\bar{u})$ is positive definite, but it is not twice differentiable in $L^2(\Omega)$, which means $f''(\bar{u})$ does not belong to the correct functional space.

But if one interprets the same functional f as an $L^\infty(\Omega)$-functional, one can show that f satisfies the differentiability conditions while it is not positive definite in $\bar{u}$ with regard to $L^\infty(\Omega)$.

This phenomenon is called 2-norm discrepancy and it shows that the choice of suitable functional spaces for an optimization problem is very important. In the late 1970s A.D. Ioffe [25] and H. Maurer and I. Zowe [31] developed sufficent optimal condition for problems in Banach spaces and presented ways to deal with the 2-norm discrepancy.

Throughout the years there were many works dealing with the application of the abstract results on different classes of problems. We want to point out the works of H. Goldberg and F. Tröltzsch [19] for the control constrained case, E. Casas, J.-P. Raymond and F. Tröltzsch [11], [15], [39] for the state constrained case and A. Rösch and F. Tröltzsch [40] for the mixed constrained case.

In this work we study general nonlinear optimization problems in Banach and Hilbert spaces and discretizations of such problems. We use this approach, because we are in particular interested in nonlinear optimization problems with state constraints and this setting allows us to formulate such conditions in a mathematical way. Our goal is to derive sufficient optimality conditions that enable us to show optimality when an exact solution is unknown but a solution of the discretized problem is at hand. This is a quite common situation, if one, for example, has computed a numerical solution using a discrete model and wants to know if an exact solution exists in a neighborhood of this discrete solution. We assume that we have a numerical method, with certain properties, to solve these nonlinear problems. We develop sufficient optimality conditions based only on the numerical solution and other known quantities. Throughout this process we also deliver error estimates regarding the numerical solution.
We want to mention the results of D. Wachsmuth and S. Akindeinde, who worked on non-convex optimal control problems with finite dimensional control space [2], [3], and the work of I. Neitzel, J. Pfefferer and A. Rösch [37] regarding state-constrained elliptic optimal control problems with semilinear state equation and their finite element discretization.

1.1 Motivation

The usual approach to determine sufficient optimality conditions, which is for example utilizied by E.Casas and F.Tröltzsch in [10],[14] for elliptic problems with state constraints, is to formulate necessary optimality conditions and sufficient optimality conditions of second order for an optimal solution $\bar{u}$. Employing additional conditions one can prove further desirable properties of the solution. One can, for example, ensure stability of the solution of state constrained problems, if one requires uniqueness of the dual variables. The catch of this approach is that the optimality conditions as well as the additional conditions have to be checked for the optimal solution $\bar{u}$. Some properties depend on the discretization parameter h to be below a certain constant h_0. This can lead to uncertainties for some kinds of problems, where one has difficulties to obtain such an optimal solution as well as computing the actual value of h_0 . Of course there are cases in which it is possible to check these conditions, for example demonstrated by H. Goldberg and F. Tröltzsch in [19], but one cannot expect that for every problem. This encourages the idea to derive a different type of condition to prove optimality of a solution, based on known informations, for example on a numerical solution. As a simple approach to get such a condition one can think of using the discrete Hessian, i.e use its eigenvalues to check if the discrete Hessian is positive definite. But as A.Rösch and D. Wachsmuth showed in [42] this approach does not work in general, which is illustrated in a simple example.

Example 1.1. [42] Let U be a Hilbert-space. We look at the following problem

$$\min_{u \in U} f(u) = \frac{1}{2} \|u - u_1\|_U^2 \|u - u_2\|_U^2 \tag{1.1}$$

We see that $\tilde{u} = \frac{u_1+u_2}{2}$ is a saddle point for this example. If we choose $u_1 = x^{-1/2+\varepsilon}$ and $u_2 = -u_1$, we can compute "critical values" of the mesh size. The smallest eigenvalue of the discrete Hessian at $\tilde{u}$ is positive, if h is above the critical value h_0. This means that we get a false positive indication for an optimum, if we use this criteria with unsufficient mesh refinement. The problem arouses because the direction $u_1 - u_2$, which is the only direction with negative curvature, is approximated poorly and thus 'overlooked' until the refinement is fine enough. In Table 1.1 one can see that this false indication can occur even for rather small discretization parameters h. For detailed information on this example we refer to [42] Section 3.

ε	h_0	
0.05	1/18	= 0.056
0.04	1/106	= 0.0094
0.03	1/1917	$= 5.2 \cdot 10^{-4}$
0.02	1/619660	$= 1.6 \cdot 10^{-6}$

Table 1.1: Critical mesh sizes found in [42]

A second example illustrates another phenomenon. Lets take a look at

Example 1.2.

$$\min_{u \in L^2([0,1])} \|S(u) - y_d\|_{L^2([0,1])}$$

with

$$(Su)(t) = \int_0^t u(x)\ \mathrm{dx}$$

and

$$y_d(x) = \begin{cases} 0 \text{ for } x \in [0, 0.5] \\ 1 \text{ for } x \in (0.5, 1]. \end{cases}$$

It is possible to approximate y_d with differentiable functions for any given error margin, which means one can find $y = Su$ with an abitrary small L^2-error $\|S(u) - y_d\|_{L^2([0,1])}$. But on the other hand, since every $y = Su$ is differentiable, it is not possible to find a solution $\bar{u}$ with $\|S(\bar{u}) - y_d\|_{L^2([0,1])} = 0$. We see that the minimizing sequence of $y_n = S(u_n)$ exists but does not converge since the limit itself is not admissible and thus we see that this continuous problem does not have an solution. Now lets take a look at an arbitrary linear FE discretization of this problem. Let $0.5 \in [x_j, x_{j+1}]$ then we see that the state

$$\bar{y}_h = \begin{cases} 0 & \text{for } 0 \le x \le x_j \\ \dfrac{x - x_j}{x_{j+1} - x_j} & \text{for } x_j < x \le x_{j+1} \\ 1 & \text{for } x_{j+1} < x \le 1 \end{cases}$$

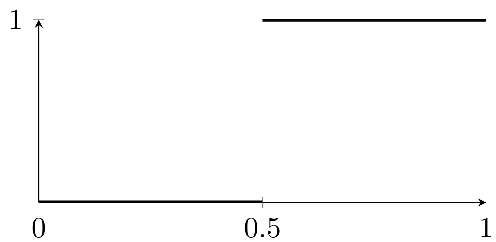
delivers an optimal functional value over all discrete states. Thus we get the corresponding optimal control

$$\bar{u}_h = \begin{cases} 0 & \text{for } 0 \le x \le x_j \\ \dfrac{1}{x_{j+1} - x_j} & \text{for } x_j < x \le x_{j+1} \\ 0 & \text{for } x_{j+1} < x \le 1 \end{cases}$$

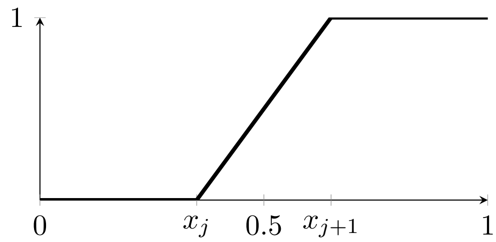
which means the discrete problem delivers a solution even if the continuous problem does not have one. Note that the $L^2([0,1])$-norm of $\bar{u}_h$, i.e.

$$\|\bar{u}_h\|_{L^2([0,1])} = \frac{1}{\sqrt{x_{j+1} - x_j}},$$

tends to infinity with finer discretizations of $[0,1]$, which means that the limit of $\bar{u}_h$ for $h \to 0$ does not belong to $L^2([0,1])$.



(a) y_d



(b) $\bar{y}_h$

Figure 1.2: Desired state y_d and optimal discrete state $\bar{y}_h$

These two examples illustrate, in rather simple settings, that it can be wrong to draw conclusions from computed solutions to the actual continuous solutions and that one has to make further efforts to reach solid results. We use the approach first presented in [41] by A.Rösch and D.Wachsmuth in 2008 and generalize their ideas to an abstract nonlinear optimization problem.

Chapter 2

Mathematical background

In this section we present the mathematical concepts, which enable us to formulate and discuss abstract optimization problems.

2.1 Banach and Hilbert spaces

Following Adams [1] we introduce the concepts of Banach and Hilbert spaces. We begin with Definition (1.7):

Definition 2.1. A **norm** on a vector space X is a real-valued function f on X satisfying the following conditions:

1. $f(x) \geq 0$ for all $x \in X$ and $f(x) = 0$ if and only if $x = 0$,
2. $f(cx) = |c|f(x)$ for every $x \in X$ and $c \in \mathbb{R}$,
3. $f(x + y) \leq f(x) + f(y)$ for every $x, y \in X$.

A vector space X provided with a norm is called **normed space**. We will denote the norm with $|| \cdot ||_X$. Now we can define convergent sequences and Cauchy sequences (see [1] (1.8),(1.9)) :

Definition 2.2. A sequence $\{x_n\}$ in a normed space X is **convergent** to the limit x_0 if and only if

$$\lim_{n \to \infty} \|x_n - x_0\|_X = 0.$$

Definition 2.3. A sequence $\{x_n\}$ in a normed space X is called **Cauchy sequence** if and only if for every $\varepsilon > 0$ there exists an integer N such that $\|x_m - x_n\|_X < \varepsilon$ holds whenever $m, n > N$.

Thus we can define

Definition 2.4. X is **complete** and a **Banach space** if every Cauchy sequence in X converges to a limit in X.

We proceed with the definition of Hilbert spaces. First we need the definition of an inner product (see [1] (1.10))

Definition 2.5. If X is a vector space, a functional $(\cdot,\cdot)_X$ defined on $X \times X$ is called **inner product** on X provided that for every $x, y \in X$ and $a, b \in \mathbb{R}$

1. $(x,y)_X = (y,x)_X$,
2. $(ax+by,z)_X = a(x,z)_X + b(y,z)_X$,
3. $(x,x)_X = 0$ if and only if $x = 0$.

Equipped with such a functional, X is called an **inner product space** and the functional

$$\|x\|_X = \sqrt{(x,x)_X}$$

is, in fact, a norm on X.

Definition 2.6. If X is complete (i.e. a Banach space) under the norm $\|x\|_X = \sqrt{(x,x)_X}$ it is called a **Hilbert space**.

We take a look at the normed dual of a normed space X (see [1] (1.11)):

Definition 2.7. A norm on the dual X^* of a normed space X can be defined by setting

$$\|x^*\|_{X^*} = \sup\{|x^*(x)| : \|x\|_X \leq 1\}$$

for each $x^* \in X^*$. Since $\mathbb{R}$ is complete, with the topology induced by this norm X^* is a Banach space (wether or not X is) and its called the **normed dual** of X.

We want to note several concepts involving dual spaces :

Definition 2.8. A sequence $\{x_n\} \subset X$ is called **weakly convergent to $\mathbf{x} \in \mathbf{X}$**, if

$$\langle x_n, f\rangle_{X,X^*} \to \langle x, f\rangle_{X,X^*} \; \forall f \in X^*$$

holds. An often used notation for this convergence is $x_n \rightharpoonup x$.

Definition 2.9. A map $F : X \to Y$ between two Banach spaces X and Y is called **weakly continuous**, if a weakly convergent sequence $\{x_n\}$ in X is mapped to a weakly convergent sequence $\{F(x_n)\}$ in Y, i.e.

$$x_n \rightharpoonup x \;\Rightarrow F(x_n) \rightharpoonup F(x),\; n \to \infty.$$

Definition 2.10. A functional $f : X \to \mathbb{R}$ is called **weakly lower semi-continuous**, if for $x_n \rightharpoonup x$ with $n \to \infty$
$$\liminf_{n\to\infty} f(x_n) \geq f(x)$$
holds.

Definition 2.11. A functional $f : X \to \mathbb{R}$ is called **radially unbounded**, if for a sequence $x_n \in U$ and $n \to \infty$
$$\|x_n\|_U \to \infty \Rightarrow f(x_n) \to \infty$$
holds.

Definition 2.12. A subset $U \subset X$ is called **weakly closed**, if for every weakly convergent sequence $\{u_n\}$ with limit $u \in X$ also $u \in U$ holds, i.e.
$$u_n \rightharpoonup u \in X,\ n \to \infty \Rightarrow u \in U.$$

Definition 2.13. An operator $G : X \to Y$ is called **weakly closed**, if $G(U) = \{G(u) : u \in U\}$ is a weakly closed subset of Y for every subset U of X.

Definition 2.14. The set $M \subset X$, X a Banach space, is called **weakly relatively compact**, if every sequence $\{x_n\} \subset M$ has a weakly convergent partial sequence. It is called **weakly compact** when it is additionally weakly closed.

The following two results, see [45] Theorem 2.10 and 2.11, will help us to ensure existence of solution for optimization problems:

Theorem 2.15. If X is a reflexive Banach space and $M \subset X$ is bounded then M is weakly relatively compact.

Theorem 2.16. If X is a Banach space and $M \subset X$ is convex and closed, then M is also weakly closed.
If X is reflexive and M convex, closed and bounded, then M is weakly compact.

To conclude this section we want to point out that, if X is a Hilbert space, it can be identified with its normed dual. This is showed by the following theorem (see [1] (1.12)).

Theorem 2.17. (Riesz representation) Let X be a Hilbert space. A linear functional x^* on X belongs to X^* if and only if there exists $x \in X$ such that for every $y \in X$ we have
$$x^*(y) = (y, x)_X,$$
and in this case $\|x^*\|_{X^*} = \|x\|_X$. Moreover, x is uniquely determined by $x^* \in X^*$.

2.2 Differentiability in Banach spaces

Let X, Y be Banach spaces and $G : X \to Y$ an operator from X to Y.

Definition 2.18. Let x and h be in X. If the limit

$$\lim_{t \searrow 0} \frac{1}{t}(G(x+th) - G(x)) =: \delta G(x,h),\ t \in \mathbb{R}$$

exists in Y, then it is called **directional derivative of G at x in direction of h**. If it exists for all $h \in X$ then the map $h \mapsto \delta G(x,h)$ is called **first variation of G at x**.

Definition 2.19. If the first variation $\delta G(x,h)$ exists as well as a linear and continuous operator $A : X \mapsto Y$ with

$$\delta G(x,h) = Ah,\ \forall h \in X$$

then A is called **Gâteaux-derivative of G at x $\in$ X**.

Definition 2.20. $G : X \to Y$ is called **Fréchet-differentiable at x $\in$ X** if there exist an operator $A \in \mathcal{L}(X,Y)$ and a map $r : X \times X \to Y$, such that

$$G(x+h) = G(x) + Ah + r(x,h)\ \forall h \in X$$

holds with

$$\frac{\|r(x,h)\|_Y}{\|h\|_x} \to 0 \text{ for } \|h\|_X \to 0.$$

A is called **Fréchet-derivative of G at x** and we use the notation $A = G'(x)$.

Definition 2.21. Is $G : X \to Y$ Fréchet-differentiable for all $x \in X$ then it is called **Fréchet-differentiable**. Let G be Fréchet-differentiable in a neighborhood of $x \in X$. If the map $x \mapsto G'(x)$ from X to $\mathcal{L}(X,Y)$ is continuous, then G is called **continuous Fréchet-differentiable at x**.

2.3 L^p and Sobolev spaces

Let Ω be a domain with Lipschitz boundary Γ. We denote by $L^p(\Omega)$, $1 \le p < \infty$, the space of real valued functions, which are defined on Ω and integrable to the p-th power with respect to the Lebesgue measure dx, i.e.

$$u \in L^p(\Omega) \Leftrightarrow \int_\Omega u^p \,\mathrm{dx} < \infty.$$

$L^p(\Omega)$ is a Banach space with the norm

$$\|u\|_{L^p(\Omega)} := \left(\int_\Omega |u(x)|^p \, \mathrm{dx}\right)^{1/p}.$$

With $p = 2$, $L^2(\Omega)$ is a Hilbert space with the scalar product

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

By $L^\infty(\Omega)$ we denote the space of all real valued functions, which are essentially bounded on Ω. The norm is given by

$$\|u\|_{L^\infty(\Omega)} = \operatorname*{ess\,sup}_{\Omega} |u(x)|.$$

Let m be an nonnegative integer and p a real number with $1 \leq p < \infty$.
$W^{m,p}(\Omega)$ denotes the Sobolev space of functions whose weak derivatives of order m lie in $L^p(\Omega)$. $W^{m,p}(\Omega)$ with the norm

$$\|u\|_{W^{m,p}(\Omega)} = \left(\sum_{|\alpha| \leq m} \|D^\alpha u\|_{L^p}^p\right)^{1/p}$$

is a Banach space.
For $p = 2$ we use the abbreviation

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

For $m = 1$ and $m = 2$, $H^m(\Omega)$ is a Hilbert space with the scalar product

$$(u, v)_{H^1(\Omega)} = \int_\Omega uv \, \mathrm{dx} + \int_\Omega \nabla u \nabla v \, \mathrm{dx},$$

$$(u, v)_{H^2(\Omega)} = \int_\Omega uv \, \mathrm{dx} + \int_\Omega \nabla u \nabla v \, \mathrm{dx} + \int_\Omega \nabla^2 u \nabla^2 v \, \mathrm{dx}$$

respectively.

Theorem 2.22 (Sobolev embedding theorem)**.** The following imbeddings are well defined and continuous for bounded $\Omega \in \mathbb{R}^n$ with Lipschitz boundary, $1 < p < \infty$ and a nonnegative integer m :

$$\begin{aligned}
&\text{For } mp < n \;:\; W^{m,p}(\Omega) \hookrightarrow L^q(\Omega), \text{ if } 1 \leq q \leq \frac{np}{n - mp};\\
&\text{For } mp = n \;:\; W^{m,p}(\Omega) \hookrightarrow L^q(\Omega), \text{ if } 1 \leq q < \infty;\\
&\text{For } mp > n \;:\; W^{m,p}(\Omega) \hookrightarrow C(\bar{\Omega}).
\end{aligned}$$

Remark 2.23. Every $L^p(\Omega)$ is a seperable space and $L^2(\Omega)$ is as a Hilbert space also reflexive.

2.4 Abstract optimization problem

Following [45] Chapter 6 we present the theory on optimization problems in Banach spaces. For the general setting we assume that U and Z are Banach spaces and $C \subset U$ is a nonempty convex subset of U.
To describe general optimization problems in Banach spaces we will utilize convex cones:

Definition 2.24. A convex set $K \subset Z$ is called **convex cone**, if $\lambda z \in K$ holds for every $z \in K$ and $\lambda > 0$.

Example 2.25. We want to give some examples for convex cones:

- For any Banach space Z :

$$K = \{0\} \text{ and } K = Z$$

- For $Z = L^2(\Omega)$ with a bounded domain $\Omega \subset \mathbb{R}^N$:

$$K = \{z \in L^2(\Omega) \;:\; z(x) \geq 0 \;\text{ f.a.a. } x \in \Omega\}$$

- For $Z = \mathbb{R}^3$:

$$K = \{z \in \mathbb{R}^3 \;:\; z_1 = 0,\; z_2 \leq 0,\; z_3 \geq 0\}$$

We use a convex cone to define a relation in Z with respect to this cone :

Definition 2.26. Let $K \subset Z$ be a convex cone and $z \in Z$. We set $z \geq_K 0$ if, and only if $z \in K$. Analogous we set $z \leq_K 0$, if $-z \in K$. Furthermore we set $z >_K 0$ and $z <_K 0$, if $z \in \text{int } K$ and $-z \in \text{int } K$, respectively.

Remark 2.27. For $z \geq_K 0$ one sees the elements of K as 'nonnegative'. The definition above can result in the fact that this nonnegativity does not comply with the natural sense of nonnegativity. If we take a look at the last example we see that the nonnegativity $z \geq_K 0$ only implies nonnegativity for z_3.

To define a relation in dual spaces and to introduce the Langrange multipliers we need to define the dual cone:

Definition 2.28. Let $K \subset Z$ be a convex cone. The dual cone belonging to K is defined as

$$K^+ = \{z^* \in Z^* \;: \langle z^*, z\rangle_{Z^*,Z} \geq 0 \;\forall z \in K\}$$

Example 2.29. We want to illustrate this definition by means of the first examples of convex cones.

- Let Z be a Banach space and $K = 0$, then $z \geq_K 0 \Leftrightarrow z = 0$ holds. This means $K^+ = Z^*$, because $\langle z^*, 0\rangle_{Z^*,Z} = 0 \geq 0$ is satisfied for every $z^* \in Z^*$.
- For a Banach space Z and $K = Z$ we get $K^+ = \{0\}$.
- With $Z = L^2(\Omega)$, with a bounded domain $\Omega \subset \mathbb{R}^N$, and

  $$K = \{z \in L^2(\Omega) \ : \ z(x) \geq 0 \ \text{ f.a.a. } x \in \Omega\}$$

  we see via the Riesz theorem that $Z = Z^*$ holds and we get $K^+ = K$.

With these definitions we can formulate the general problem for Fréchet-differentiable f and G, with $f : U \to \mathbb{R}$ and $G : U \to Z$, and a convex cone $K \subset Z$:

Problem 2.30.

$$\min_{u \in C} f(u) \tag{2.1}$$

s.t.:

$$G(u) \leq_K 0 \tag{2.2}$$

Definition 2.31. $\bar{u} \in C$ is called a local solution of Problem 2.30, if $\bar{u}$ is a feasible point and

$$f(\bar{u}) \leq f(u)$$

is fulfilled for all $u \in C$ with $G(u) \leq_K 0$ and $\|u - \bar{u}\|_U \leq \varepsilon$ with a suitable $\varepsilon > 0$.

We define the Langrange function $L(u, z^*)$:

Definition 2.32. The function

$$L(u, z^*) = f(u) + \langle z^*, G(u)\rangle_{Z^*,Z},$$

$L : U \times Z^* \to \mathbb{R}$, is called Lagrange function.

A Lagrange multiplier is defined as:

Definition 2.33. $z^* \in K^+$ is called a Lagrange multiplier for a local solution $\bar{u}$ of Problem 2.30 if the following conditions are fulfilled:

$$D_u L(\bar{u}, z^*)(u - \bar{u}) \geq 0 \quad \forall u \in C$$
$$\langle z^*, G(\bar{u})\rangle_{Z^*,Z} = 0$$

The existence of Lagrange multipliers can be ensured via regularity conditions, also called constraint qualifications, such as the regularity condition of Kurcyusz and Zowe [29]:

Definition 2.34. Let $\bar{u} \in C$ be with $G(\bar{u}) \leq_K 0$. The sets

$$C(\bar{u}) = \{\lambda(u - \bar{u}) \ : \ \lambda \geq 0,\ u \in C\}$$

and

$$K(\bar{z}) = \{\lambda(z - \bar{z}) \ : \ \lambda \geq 0,\ z \in K\}$$

are the conical hulls on C in $\bar{u}$ and K in $\bar{z}$. The regularity condition of Kurcyusz and Zowe can be formulated as

$$G'(\bar{u})C(\bar{u}) + K(-G(\bar{u})) = Z \tag{2.3}$$

which is equivalent to the fact that the equation

$$\alpha G'(\bar{u})(u - \bar{u}) + \beta(v + G(\bar{u})) = z$$

has a solution for every given $z \in Z$ with $u \in C$, $v \geq_K 0$ and nonnegative α and β.

Theorem 2.35. [[45] Theorem 6.3] Let $\bar{u}$ be a local solution of Problem 2.30 and f, G continuous Fréchet-differentiable in an open neighborhood of $\bar{u}$. Then there exists a Lagrange multiplier $z^* \in Z^*$ belonging to $\bar{u}$, if regularity condition (2.3) is fulfilled. The set of Lagrange multipliers belonging to $\bar{u}$ is bounded.

We will use an formulation, which is sufficient for (2.3), if K and C have a nonempty interior:

$$\exists \tilde{u} \in \text{int } C(\bar{u}) \ : \ G(\bar{u}) + G'(\bar{u})\tilde{u} <_K 0 \tag{2.4}$$

Condition (2.4) is called **M**angasarian **F**romovitz **C**onstraint **Q**ualification.

Remark 2.36. The regularity condition depends on the nonempty interior of K, which cones in $Z = L^p(\Omega)$, with $1 \leq p < \infty$, do not possess. We take for example the natural choice of the nonnegative cone in $L^2([0,1])$

$$K = \{z \in L^2([0,1]) \ : \ z(x) \geq 0 \text{ a.e. in } [0,1]\}.$$

One would think that a function such as $z(x) \equiv 1$ is an interior point of K. But if we look at the sequence

$$v_n(x) = \begin{cases} 1 \text{ in } [0, 1 - 1/n) \\ -1 \text{ in } [1 - 1/n, 1] \end{cases}$$

it does not belong to K, but it converges to z with respect to the L^2-norm. This effect occurs for every L^p space with $1 \leq p < \infty$, which makes it necessary to choose $Z \subset L^\infty(\Omega)$ if we want to employ formulation (2.4).

Chapter 3

Optimality conditions and main result

3.1 Optimality conditions

We consider the following general problem setting :

Assumption 3.1. (Setting of P)
Let U be a Hilbert space, Z a Banach space, $f : U \to \mathbb{R}$, $G : U \to Z$ and

$$\text{f a twice continuously Fréchet-differentiable functional} \tag{3.1}$$

$$\text{G a weakly closed operator} \tag{3.2}$$

$$\text{G a twice continuously Fréchet-differentiable nonlinear operator.} \tag{3.3}$$

With a weakly closed and non-empty subset U_{ad} of U we consider the problem

$$\min_{u \in U_{ad}} f(u) \tag{P}$$

and describe U_{ad} as

$$U_{ad} = \{u \in U \ : \ Gu \leq_K 0\}$$

while $K \subset Z$ is a convex cone. To use our approach we make several additional assumptions. We start with some properties of the functional f:

Assumption 3.2. (Properties of P)
Assume that $f : U \to \mathbb{R}$ has the following properties:

$$\text{f is bounded from below, i.e. } f(u) \geq b \text{ , for all } u \in U \text{ and one } b \in \mathbb{R}. \tag{3.4}$$

$$\text{f is weakly lower semicontinuous.} \tag{3.5}$$

$$\text{f is radially unbounded.} \tag{3.6}$$

$$|(f''(u) - f''(w))[v_1, v_2]| \leq M\|u - w\|_U \|v_1\|_U \|v_2\|_U, \tag{3.7}$$
$$\forall u, w \in U_{ad}, \; v_1, v_2 \in U$$

Lemma 3.3. (Existence of a solution for the continuous problem)
If Assumption 3.2 is fulfilled, then there exists at least one control $\bar{u} \in U_{ad}$ such that

$$f(\bar{u}) \leq f(u) \; \forall u \in U_{ad}$$

holds.

Proof. Since f is bounded from below there exists a $j \in \mathbb{R}$ with

$$j = \inf_{u \in U_{ad}} f(u)$$

We choose a minimizing sequence u_n such that

$$f(u_n) \to j \text{ for } n \to \infty.$$

Since f is radially unbounded we know that $\|u_n\|_U \leq C$, with a certain positive number C, holds for all $n \in \mathbb{N}$. Thus $\{u_n\}$ is a bounded set and consequently weakly relatively compact. (See Theorem 2.15) This means we can choose a subsequence $\{u_{n_k}\} \subset U_{ad}$ such that

$$u_{n_k} \rightharpoonup \bar{u} \text{ for } k \to \infty$$

for some $\bar{u} \in U_{ad}$. (See Note 2.23.) Note that $\bar{u}$ is in U_{ad}, because U_{ad} is weakly closed.
Since f is weakly lower semicontinuous we also see that

$$f(\bar{u}) \leq \liminf_{k \to \infty} \; f(u_{n_k})$$

holds. This leads to $f(\bar{u}) = j \leq f(u) \; \forall u \in U_{ad}$. □

From this point on we denote a local minimizer of (P) by $\bar{u}$.
To formulate the optimality conditions for the continuous problem we recall the Lagrange function:

Definition 3.4. The function $\mathcal{L} : U \times Z^* \to \mathbb{R}$

$$\mathcal{L}(u, z^*) = f(u) + \langle z^*, Gu \rangle_{Z^*, Z} \tag{3.8}$$

is called Lagrange function of (P).

To ensure the existence of Lagrange multipliers we assume a regularity condition

Assumption 3.5. (MFCQ-type)
There exists a $d \in U$ such that

$$G(\bar{u}) + G'(\bar{u})d <_K 0$$

holds.

Thus we get via Theorem 2.35:

Lemma 3.6. If Assumption 3.5 is fulfilled, then there exists a Lagrange multiplier $\mu \in K^* \subset Z^*$, such that the following properties are fulfilled :

$$D_u\mathcal{L}(\bar{u}, \mu)(u - \bar{u}) \geq 0 \ \forall u \in U_{ad} \tag{3.9}$$

$$\langle \mu, G(\bar{u}) \rangle_{Z^*,Z} = 0 \tag{3.10}$$

We denote μ as a Lagrange multiplier corresponding to $\bar{u}$.

At this point we want to introduce a second optimization problem (P_h) as a discrete counterpart to (P) and discuss it in a similar way.

Assumption 3.7. (Setting of P_h)
Let U be a Hilbert space, $Z_h \subset Z$ a Banach space, $f_h : U \to \mathbb{R}$, $G_h : U \to Z_h$ and

$$f_h \text{ a twice continuously Fréchet-differentiable functional} \tag{3.11}$$

$$G_h \text{ a weakly closed operator} \tag{3.12}$$

$$G_h \text{ a twice continuously Fréchet-differentiable nonlinear operator.} \tag{3.13}$$

We consider

$$\min_{u \in U_{ad}^h} f_h(u) \tag{P_h}$$

with

$$U_{ad}^h = \{u \in U \ : \ G_h u \leq_K 0\}$$

as a discrete problem.

Remark 3.8. At this point we want to emphasize two things :

- U has not been discretized, which means we use the approach of M.Hinze presented in [24].
- For linear finite element examples there is no difference between this so called Hinze discretization and a standard discretization of U.

We impose the same assumptions on the functional f_h as on f.

Assumption 3.9. (Properties of P_h)
Assume that $f_h : U \to \mathbb{R}$ has the following properties:

$$f_h \text{ is bounded from below} \tag{3.14}$$
$$f_h \text{ is weakly lower semicontinuous} \tag{3.15}$$
$$f_h \text{ is radially unbounded} \tag{3.16}$$

Lemma 3.10. (Existence of a solution for P_h)
If Assumption 3.9 is fulfilled, then there exists at least one control $\bar{u}_h$ such that

$$f_h(\bar{u}_h) \leq f_h(u) \ \forall u \in U_{ad}^h$$

holds.

This can be proven in the same way as Lemma 3.3.
By $\bar{u}_h$ we denote a local minimizer of (P_h). We assume a slightly different regularity condition for (P_h)

Assumption 3.11. (Regularity condition of P_h)
There exists an $d_h \in U$ such that

$$\begin{aligned} -G_h\bar{u}_h - sG_h'(\bar{u}_h)d_h - z \in K, \\ \forall z \in Z : \|z\|_Z \leq s\tau,\ s \in [0,1] \end{aligned} \tag{3.17}$$

and formulate the optimality condition for (P_h).

Lemma 3.12. If assumption 3.11 is fulfilled, then there exists a Lagrange multiplier $\mu_h \in K^* \subset Z_h^*$, such that the following properties are fulfilled:

$$D_u\mathcal{L}_h(\bar{u}_h, \mu_h)(u - \bar{u}_h) \geq 0,\ u \in U_{ad}^h \tag{3.18}$$
$$\langle \mu_h, G_h(\bar{u}) \rangle_{(Z_h)^*, Z_h} = 0 \tag{3.19}$$

The Lagrange function $\mathcal{L}_h$ is defined as

$$\mathcal{L}_h(u, z^*) = f_h(u) + \langle z^*, G_h u \rangle_{(Z_h)^*, Z_h}$$

and μ_h is a Langrange multiplier of $\bar{u}_h$.

We introduced the first order optimality conditions and showed the existence of a solution for the two problems. We interpret (P_h) as the discretized problem of the continuous problem (P) via the following assumptions:

Assumption 3.13. (Approximation properties of G_h)
There exist constants $c_G, c_{G'}$ and k such that

$$\|G(u_h) - G_h(u_h)\|_Z \le c_G h^k \|u_h\|_U,\ u_h \in U_{ad}^h \tag{3.20}$$
$$\|[G'(\bar{u}_h) - G'_h(\bar{u}_h)]u\|_Z \le c_{G'} h^k \|u\|_U,\ \forall u \in U \tag{3.21}$$

hold.

Assumption 3.14. (Approximation properties of f_h)
There exist constants $c_{f'}$ and k such that

$$|[f'(\bar{u}_h) - f'_h(\bar{u}_h)]u| \le c_{f'} h^k \|u\|_U,\ \ \forall u \in U \tag{3.22}$$

Throughout the estimation process we will impose the following properties on f,f_h,G and G_h

Assumption 3.15. (Coercivity, boundedness and Lipschitz-type conditions)
There exist constants L, M, N, R and $\alpha > 0$ such that

$$\mathcal{L}''(\bar{u}_h, \mu_h)v^2 = f''(\bar{u}_h)[v,v] + \langle \mu_h, G''(\bar{u}_h)[v,v]\rangle \ge \alpha \|v\|_U^2,\ \forall v \in U \tag{3.23}$$
$$\|[G'_h(u) - G'_h(\bar{u}_h)]v\| \le L\|u - \bar{u}_h\|_U \|v\|_U,\ \forall v \in U, \tag{3.24}$$
$$\forall u \in U : \|u - \bar{u}_h\|_U \le \|\tilde{u}_h - \bar{u}_h\|_U$$
$$|(f''(u) - f''(\bar{u}_h))[v_1, v_2]| \le M\|u - \bar{u}_h\|_U \|v_1\|_U \|v_2\|_U,\ \forall u \in U_{ad} \tag{3.25}$$
$$\|u - \bar{u}_h\|_U \le R,\ \ v_1, v_2 \in U$$
$$\|[G''(u) - G''(\bar{u}_h)][v_1, v_2]\|_Z \le N\|u - \bar{u}_h\|_U \|v_1\|_U \|v_2\|_U,\ \forall u \in U_{ad} \tag{3.26}$$
$$\|u - \bar{u}_h\|_U \le R,\ \ v_1, v_2 \in U$$

hold.

Remark 3.16. We want to point out

- that assumption (3.23) implies the coercivity of the second order derivative of f in $\bar{u}_h$
- that assumption (3.25) is a confinement of assumption (3.7).

To conclude this section we introduce a class of example, on which we will take a closer look in Chapter 5:

Example 3.17. We set $U = L^2(\Omega)$, $Z = L^\infty(\Omega)$, $\Omega \subset \mathbb{R}^n, n = 1,..,3$, and consider the problem

$$\begin{aligned} \min\ J(y,u) &:= \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\|u - u_d\|^2_{L^2(\Omega)} \\ -\Delta y + d(x,y) &= u \text{ in } \Omega \\ y(x) &= 0 \text{ on } \Gamma \\ y(x) = Gu(x) &\le c_c(x) \text{ a.e. in } \Omega \\ y(x) = Gu(x) &\ge -c_c(x) \text{ a.e. in } \Omega \end{aligned}$$

We set $\hat{G}$ as

$$\hat{G} = \begin{cases} Gu + c_c \\ -Gu + c_c \end{cases}$$

while G is the control-state operator belonging to the PDE above, and with

$$K = \{z \in Z : \ z(x) \ge 0, \text{ a.e. in } \Omega\}$$

we express the constraints of Example 3.17 as $\hat{G}u \le_K 0$.

3.2 Estimation strategy

Our goal is to formulate the final part of a second order sufficient condition for the optimization problem (P) and give an estimate of the discrepancy of the continuous optimal functional value $f(\bar{u})$ and the functional value $f(\bar{u}_h)$ as well as an estimate of the error $\|\bar{u} - \bar{u}_h\|_U$ between a continuous local minimizer $\bar{u}$ and a discrete local minimizer $\bar{u}_h$. In this section we want to present the main ideas we pursued to achieve the SSC and the estimate. The strategy can be divided in four major steps.
Step 1: Since we can not expect that $\bar{u}_h$ is a feasible control for the continuous problem we start with the construction of a feasible control u_δ. We use (3.18) and set

$$u_\delta := \bar{u}_h + s\delta d_h,$$

which is feasible for sufficient small h and a adequate choice of s and δ. (See Section 3.1)
Step 2: We derive an estimate of $|f(u_\delta) - f(\bar{u}_h)|$ depending on the discretization parameter h, the functional f and other known quantities and get:

$$|f(u_\delta) - f(\bar{u}_h)| \le c_f h^k \ .$$

Step 3: We consider all feasible $u \in U_{ad}$, which lie on the boundary of $B(\bar{u}_h, r)$, and get an inequality of the following structure:

$$f(u) - f(\bar{u}_h) \ge \alpha r^2 - \beta r - \gamma r^3 - \delta$$

with constants $\alpha, \beta, \gamma, \delta$, which depend on the quantities L,M,N,α,c_G,$c_{G'}$,$c_{f'}$,h, which we introduced in Section 3.1, and $\bar{u}_h$.

Step 4: We show that a continuous local minimizer $\bar{u}$ lies in the interior of a ball with radius r, $0 \le r \le R$, around $\bar{u}_h$.

This is equivalent to the fact that the solutions of the restricted problem

$$\min_{u \in U_{ad} \cap B(\bar{u}_h, r)} f(u) \tag{P_r}$$

are inner points of $B(\bar{u}_h, r) = \{u \in U : \|u - \bar{u}_h\|_U \le r\}$.

We also show that

$$\alpha r^2 - \beta r - \gamma r^3 - \delta \ge c_f h^k$$

holds for an adequate choice of r, $0 \le r \le R$.

We see that a u on the boundary of $B(\bar{u}_h, r)$ cannot be optimal for (P_r), which means it is no local solution of (P) . This leads to our desired estimates

$$|f(\bar{u}) - f(\bar{u}_h)| \le c_f h^k \tag{3.27}$$

$$\|\bar{u} - \bar{u}_h\|_U \le r \tag{3.28}$$

for an appropriate set of constants c_f, k and r.

Lemma 3.18. (Existence of a solution for P_r)

Let $\bar{u}_h$ be a solution of P_h and $r \in \mathbb{R}$ with $0 \le r \le R$. If Assumption 3.2 is fulfilled, then there exists at least one control $\bar{u} \in U_{ad} \cap B(\bar{u}_h, r)$ such that

$$f(\bar{u}) \le f(u) \ \forall u \in U_{ad} \cap B(\bar{u}_h, r)$$

holds.

Proof. Since $B(\bar{u}_h, r)$ is convex, closed and bounded we know because of Theorem 2.15, Theorem 2.16 and Remark 2.23 that $B(\bar{u}_h, r)$ is weakly compact. This means that $U_{ad} \cap B(\bar{u}_h, r)$ is weakly closed and that we can use the same techniques as in the proof of Lemma 3.3. $\square$

Chapter 4

Derivation of the main result

4.1 Construction of a feasible point

As stated in Chapter 3 $\bar{u}_h$ is a local minimizer of (P_h). Furthermore we choose a d_h, which fulfilles Assumption 3.11. Then we define u_δ as

$$u_\delta := \bar{u}_h + s\delta d_h$$

for $\delta \in [0, 1]$.
For every h, we denote m as the maximum of $\|\bar{u}_h\|_U$ and $\|\bar{u}_h + d_h\|_U$, i.e.

$$m = \max\{\|\bar{u}_h\|_U, \|\bar{u}_h + d_h\|_U\}.$$

We show that the control u_δ is feasible for an adequate choice of s and δ:

Theorem 4.1. For sufficient small h, $0 < s \leq 1$ and $u_\delta := \bar{u}_h + s\delta d_h$ the following implication holds:

$$\delta \geq \frac{2ch^k m}{s(\tau - sL\|d_h\|_U^2)} \Rightarrow u_\delta \in U_{ad} \tag{4.1}$$

Proof.

$$
\begin{aligned}
Gu_\delta &= Gu_\delta + G_h u_\delta - G_h u_\delta \\
&= Gu_\delta - G_h u_\delta + G_h \bar{u}_h \\
&+ \int_0^1 G_h'(\bar{u}_h + ts\delta d_h) s\delta d_h dt \\
&= Gu_\delta - G_h u_\delta + G_h \bar{u}_h \\
&+ \int_0^1 G_h'(\bar{u}_h + ts\delta d_h) s\delta d_h \\
&+ (G_h'(\bar{u}_h) - G_h'(\bar{u}_h)) s\delta d_h dt \\
&= Gu_\delta - G_h u_\delta + G_h \bar{u}_h + \delta s G_h'(\bar{u}_h) d_h \\
&+ \delta \int_0^1 [G_h'(\bar{u}_h + ts\delta d_h) - G_h'(\bar{u}_h)] s d_h dt
\end{aligned}
$$

That means we have to show that

$$
\begin{aligned}
&- G_h \bar{u}_h - \delta s G_h'(\bar{u}_h) d_h - Gu_\delta + G_h u_\delta \\
&- \delta \int_0^1 [G_h'(\bar{u}_h + ts\delta d_h) - G_h'(\bar{u}_h)] s d_h dt \in K
\end{aligned}
$$

holds. We know that $-G_h \bar{u}_h - \delta s G_h'(\bar{u}_h) d_h - z \in K$ holds for all $z \in Z : \|z\|_Z \leq s\delta\tau$ because of Assumption 3.11 and the convexity of K. That means if

$$
\| - Gu_\delta + G_h u_\delta - \delta \int_0^1 [G_h'(\bar{u}_h + ts\delta d_h) - G_h'(\bar{u}_h)] s d_h dt \|_Z \leq \delta s \tau
$$

holds, than $Gu_\delta \leq_K 0$ is fulfilled and u_δ is a feasible control. We derive a lower bound of δ:

$$
\begin{aligned}
&\| - Gu_\delta + G_h u_\delta - \delta \int_0^1 [G_h'(\bar{u}_h + ts\delta d_h) - G_h'(\bar{u}_h)] s d_h dt \|_Z \\
&\leq \| - Gu_\delta + G_h u_\delta \|_Z + \| - \delta \int_0^1 [G_h'(\bar{u}_h + ts\delta d_h) - G_h'(\bar{u}_h)] s d_h dt \|_Z \\
&\leq c_G h^k \|u_\delta\|_U + \delta \int_0^1 \| [G_h'(\bar{u}_h + ts\delta d_h) - G_h'(\bar{u}_h)] s d_h \|_Z dt \\
&\leq 2 c_G h^k m + \delta^2 L s^2 \|d_h\|_U^2 \leq 2 c_G h^k m + \delta L s^2 \|d_h\|_U^2
\end{aligned}
$$

Thus u_δ is feasible if

$$\begin{aligned} 2c_G h^k m + \delta L s^2 \|d_h\|_U^2 &\leq s\delta\tau \\ \Leftrightarrow 2c_G h^k m &\leq s\delta\tau - \delta L s^2 \|d_h\|_U^2 \\ \Leftrightarrow 2c_G h^k m &\leq \delta(s\tau - L s^2 \|d_h\|_U^2) \\ \Leftrightarrow \frac{2c_G h^k m}{s(\tau - Ls\|d_h\|_U^2)} &\leq \delta \end{aligned}$$

□

To get u_δ as close as possible to $\bar{u}_h$ we want the lower bound for δ to be as small as possible. To accomplish that we set

$$s = \min\{1, \frac{\tau}{2L\|d_h\|_U^2}\}.$$

This leads to

$$\delta = \frac{2c_G h^k m}{\frac{\tau}{2L\|d_h\|_U^2}(\tau - \frac{\tau L\|d_h\|_U^2}{2L\|d_h\|_U^2})} = \frac{8c_G h^k m L\|d_h\|_U^2}{\tau^2}$$

if we set δ on the lower bound and $s = \frac{\tau}{2L\|d_h\|_U^2} < 1$ holds. This means for sufficient small h we get a $\delta \leq 1$, for which u_δ is feasible. With this we have completed Step 1 of the estimation strategy.

4.2 Error of u_δ

Now we come to the second step presented in Section 3.2. We will estimate the difference of the functional values of u_δ and $\bar{u}_h$. We assume that $s = \frac{\tau}{2L\|d_h\|_U^2} < 1$ holds, because we get the same results for $s = 1$ only with a slightly different constant c_f.

Theorem 4.2. For sufficient small h, $m = \max\{\|\bar{u}_h\|_U, \|\bar{u}_h + d_h\|_U\}$, $s = \frac{\tau}{2L\|d_h\|_U^2}$ and $\delta = \frac{8c_G h^k m L\|d_h\|_U^2}{\tau^2}$ the following inequality holds:

$$|f(u_\delta) - f(\bar{u}_h)| \leq c_f h^k \tag{4.2}$$

Proof.

$$
\begin{aligned}
|f(u_\delta) - f(\bar{u}_h)| &= |\int_0^1 f'(\bar{u}_h + t(u_\delta - \bar{u}_h))(u_\delta - \bar{u}_h)\mathrm{dt}| \\
&\leq \int_0^1 |f'(\bar{u}_h + t(u_\delta - \bar{u}_h))(u_\delta - \bar{u}_h)|\mathrm{dt} \\
&\leq \|f'\|\|\bar{u}_h - u_\delta\|_U \\
&= \|f'\|\|\bar{u}_h - \bar{u}_h - s\delta d_h\|_U \\
&= \|f'\|s\delta\|d_h\|_U \\
&= \|f'\|\frac{\tau}{2L\|d_h\|_U^2}\frac{8c_G h^k mL\|d_h\|_U^2}{\tau^2}\|d_h\|_U \\
&= \|f'\|\frac{4c_G h^k m\|d_h\|_U}{\tau} \\
&\leq \|f'\|\frac{8c_G h^k m^2}{\tau} \\
&\leq c_f h^k
\end{aligned}
$$

□

This concludes Step 2.

4.3 Error on the boundary of $B(\bar{u}_h, r)$

As mentioned before we deal with the third step and derive a lower bound for the error on the boundary of $B(\bar{u}_h, r)$. We recall

$$\mathcal{L}(u, \mu) = f(u) + \langle \mu, Gu \rangle_{Z^*,Z}, \ \forall u \in U_{ad}$$

as the Lagrange-function of the problem (P) and

$$\mathcal{L}_h(u, \mu) = f_h(u) + \langle \mu, G_h u \rangle_{Z^*,Z}, \ \forall u \in U_{ad}^h$$

as the Lagrange-function of the discrete problem (P_h). Note that $\bar{u}_h$ satisfies a first order condition, i.e.

$$\langle f_h'(\bar{u}_h) + G_h'(\bar{u}_h)^* \mu_h, u - \bar{u}_h \rangle_{U^*,U} \geq 0, \ \forall u \in U$$

which leads to

$$f_h'(\bar{u}_h) + G_h'(\bar{u}_h)^* \mu_h = 0 \tag{4.3}$$

Now we consider the $u \in U_{ad}$ on the boundary of $B(\bar{u}_h, r)$.

Theorem 4.3. Let $R \geq r > 0$, $u \in U_{ad}$ be on the boundary of $B(\bar{u}_h, r)$, i.e. $\|u - \bar{u}_h\|_U = r$, $\bar{u}_h$ a solution of the discrete problem (P_h) and $\mu_h \in K^*$ a Lagrange multiplier of $\mathcal{L}_h$ with respect to $\bar{u}_h$. Then the following inequality holds:

$$f(u) - f(\bar{u}_h) \geq \frac{\alpha}{2} r^2 - h^k r(c_{f'} + c_{G'} \|\mu_h\|_{Z^*}) - \frac{r^3}{6}(M + N\|\mu_h\|_{Z^*}) - c_G h^k \|\mu_h\|_{Z^*} \|\bar{u}_h\|_U$$

Proof.

$$\begin{aligned} f(u) - f(\bar{u}_h) &\geq f(u) - f(\bar{u}_h) + \langle \mu_h, Gu \rangle_{Z^*,Z} \\ &- \langle \mu_h, G_h \bar{u}_h \rangle_{Z^*,Z} \\ &= f(u) - f(\bar{u}_h) + \langle \mu_h, Gu \rangle_{Z^*,Z} \\ &- \langle \mu_h, G\bar{u}_h \rangle_{Z^*,Z} \\ &+ \langle \mu_h, G\bar{u}_h \rangle_{Z^*,Z} - \langle \mu_h, G_h \bar{u}_h \rangle_{Z^*,Z} \\ &= \mathcal{L}(u, \mu_h) - \mathcal{L}(\bar{u}_h, \mu_h) + \langle \mu_h, (G - G_h)\bar{u}_h \rangle_{Z^*,Z} \end{aligned}$$

This leads to:

$$\begin{aligned} f(u) - f(\bar{u}_h) &\geq \mathcal{L}(u, \mu_h) - \mathcal{L}(\bar{u}_h, \mu_h) - \langle \mu_h, (G - G_h)\bar{u}_h \rangle_{Z^*,Z} \\ &\geq \underbrace{\mathcal{L}'(\bar{u}_h, \mu_h)(u - \bar{u}_h)}_{i)} + \frac{1}{2} \underbrace{\mathcal{L}''(\bar{u}_h, \mu_h)(u - \bar{u}_h)^2}_{ii)} \\ &+ \underbrace{\int_0^1 \int_0^s [\mathcal{L}''(\bar{u}_h + t(\bar{u}_h - u), \mu_h) - \mathcal{L}''(\bar{u}_h, \mu_h)](u - \bar{u}_h)^2 \,\mathrm{dtds}}_{iii)} \\ &+ \langle \mu_h, (G - G_h)\bar{u}_h \rangle_{Z^*,Z} \end{aligned}$$

where $\mathcal{L}'(u, \mu)$ is the partial derivative in direction of u and

$$\mathcal{L}''(\bar{u}_h, \mu_h)(u - \bar{u}_h)^2 := \mathcal{L}''(\bar{u}_h, \mu_h)[u - \bar{u}_h, u - \bar{u}_h]$$

Ad i):

$$\begin{aligned} \mathcal{L}'(\bar{u}_h, \mu_h)(u - \bar{u}_h) &= f'(\bar{u}_h)(u - \bar{u}_h) + G'(\bar{u}_h)^* \mu_h (u - \bar{u}_h) \\ &= [f'(\bar{u}_h) - f'_h(\bar{u}_h)](u - \bar{u}_h) + \underbrace{[f'_h(\bar{u}_h) + G'_h(\bar{u}_h)^* \mu_h](u - \bar{u}_h)}_{=0 \text{ by } (4.3)} \\ &+ [G'(\bar{u}_h)^* - G'_h(\bar{u}_h)^*]\mu_h (u - \bar{u}_h) \\ &\geq -c_{f'} h^k \|u - \bar{u}_h\|_U - c_{G'} h^k \|\mu_h\|_{Z^*} \|u - \bar{u}_h\|_U \end{aligned}$$

Ad ii):

$$\mathcal{L}''(\bar{u}_h, \mu_h)(u - \bar{u}_h)^2 \geq \alpha \|u - \bar{u}_h\|_U^2 \text{ by } (3.23)$$

Ad iii):

$$\int_0^1 \int_0^s [\mathcal{L}''(\bar{u}_h + t(\bar{u}_h - u), \mu_h) - \mathcal{L}''(\bar{u}_h, \mu_h)](u - \bar{u}_h)^2 \text{ dtds}$$
$$= \int_0^1 \int_0^s [f''(\bar{u}_h + t(u - \bar{u}_h)) - f''(\bar{u}_h)](u - \bar{u}_h)^2$$
$$+\langle \mu_h, [G''(\bar{u}_h + t(u - \bar{u}_h)) - G''(\bar{u}_h)](u - \bar{u}_h)^2 \rangle_{Z^*,Z} \text{dtds}$$
$$\geq -(\frac{M}{6} \|u - \bar{u}_h\|_U^3 + \frac{N}{6} \|\mu_h\|_{Z^*} \|u - \bar{u}_h\|_U^3)$$

For $u \in U_{ad}$ with $\|u - \bar{u}_h\|_U = r$, these three inequalities imply:

$$\begin{aligned} f(u) - f(\bar{u}_h) \geq & \frac{\alpha}{2} \|u - \bar{u}_h\|_U^2 - ch^k \|u - \bar{u}_h\|_U \\ & - ch^k \|\mu_h\|_{Z^*} \|u - \bar{u}_h\|_U - \frac{M}{6} \|u - \bar{u}_h\|_U^3 \\ & - \frac{N}{6} \|u - \bar{u}_h\|_U^3 - \langle \mu_h, (G - G_h)\bar{u}_h \rangle_{Z^*,Z} \\ \geq & \frac{\alpha}{2} r^2 - h^k r(c_{f'} + c_{G'} \|\mu_h\|_{Z^*}) - \frac{1}{6} r^3 (M + N \|\mu_h\|_{Z^*}) \\ & - c_G h^k \|\mu_h\|_{Z^*} \|\bar{u}_h\|_U \end{aligned}$$

□

Now that Step 3 is completed we combine the previous theorems to prove the existence of a local minimizer $\bar{u}$ of (P) in a neighborhood of $\bar{u}_h$:

4.4 Main result

We prove that a local minimizer $\bar{u}$ of (P) lies in a ball around $\bar{u}_h$ such that the discrepancy of $f(\bar{u})$ and $f(\bar{u}_h)$ is bounded by $c_f h^k$.

Theorem 4.4. Let $s = \frac{\tau}{L\|\tilde{u}_h - \bar{u}_h\|_U^2}$ and $\delta = \frac{8 c_G h^k m L \|\tilde{u}_h - \bar{u}_h\|_U^2}{\tau^2}$.
If there is a radius r, $0 \leq r \leq R$, for which

$$\frac{\alpha}{2} r^2 - h^k r(c_{f'} + c_{G'} \|\mu_h\|_{Z^*}) - \frac{1}{6} r^3 (M + N \|\mu_h\|_{Z^*}) - c_G h^k \|\mu_h\|_{Z^*} \|\bar{u}_h\|_U - c_f h^k > 0$$

is fulfilled, then

$$\|\bar{u} - \bar{u}_h\|_U < r \tag{4.4}$$

holds, which means that a local solution $\bar{u}$ of (P) lies within a ball of radius r around $\bar{u}_h$.

Proof. It results from Theorem 4.3 that for every $u \in U_{ad}$ with $\|u - \bar{u}_h\|_U = r$ the following inequality applies :

$$f(u)-f(\bar{u}_h) \geq \frac{\alpha r^2}{2} - h^k r(c_{f'} + c_{G'}\|\mu_h\|_{Z^*}) - \frac{r^3}{6}(M + N\|\mu_h\|_{Z^*}) - c_G h^k \|\mu_h\|_{Z^*}\|\bar{u}_h\|_U$$

On the other hand we get from Theorem 4.2 that

$$|f(u_\delta) - f(\bar{u}_h)| < c_f h^k$$

holds. For sufficient small h we get that $\|u_\delta - \bar{u}_h\|_U < r$ holds, which means that the optimal solution of the restricted problem

$$\min_{u \in U_{ad} \cap B(\bar{u}_h, r)} f(u) \tag{P_r}$$

is not on the boundary of $B(\bar{u}_h, r)$, which means that the solution of (P_r) is a local solution of (P). □

With that Step 4 is completed and our strategy was succesful. We derived a sufficient optimality condition in Theorem 4.4 as well as the error estimate

$$|f(u_\delta) - f(\bar{u}_h)| \leq c_f h^k$$

in Theorem 4.2.

Remark 4.5. After determining a discrete solution $\bar{u}_h$ and a Lagrange multiplier μ_h we reduce our problem by means of Theorem 4.4. It remains to show existence of a root of a third order polynomial in the interval $[0, R]$.

Chapter 5

Example

We consider a semilinear elliptic problem:

Example 5.1.

$$
\begin{aligned}
\min\ J(y,u) := \frac{1}{2}\|y-y_d\|^2_{L^2([0,1])} - \frac{\lambda}{2}\|u-u_d\|^2_{L^2([0,1])} \\
-\Delta y + d(x,y) = u \text{ in } [0,1] \\
y(0) = y(1) = 0 \\
y(x) = Gu(x) \leq c_c(x) \text{ a.e. in } [0,1] \\
y(x) = Gu(x) \geq -c_c(x) \text{ a.e. in } [0,1]
\end{aligned}
$$

with the control to state operator $G : L^2([0,1]) \mapsto H^1_0([0,1])$ belonging to the ODE above.
With

$$
\hat{G} = \begin{cases} Gu + c_c \\ -Gu + c_c \end{cases}
$$

and

$$
K = \{z \in Z : \ z(x) \geq 0, \ \text{a.e. in } [0,1]\}
$$

we can express these condition via the cone relation :

$$
\hat{G}u \leq_K 0
$$

We set d as :

$$
d(x,y) := y(x) + y^3(x)
$$

That leads to :

$$
\begin{aligned}
d'(x,y)h = h(x) + 3y^2(x)h(x) \\
d''(x,y)(h_1(x), h_2(x)) = 6y(x)h_1(x)h_2(x)
\end{aligned}
$$

where d' is the partial derivative of d in direction of y.
The first order Frechét-derivative of G is determined as :

$$G'(\hat{u})u = y \tag{5.1}$$

where y is the weak solution of

$$\begin{aligned} -\Delta y + y + 3\hat{y}^2 y &= u \text{ in } \Omega \\ y &= 0 \text{ on } \Gamma \end{aligned}$$

with $G(\hat{u}) = \hat{y}$.
The second order Frechét-derivative is determined as

$$G''(\hat{u})[u_1, u_2] = \hat{z} \tag{5.2}$$

where $\hat{z}$ is the weak solution of

$$\begin{aligned} -\Delta \hat{z} + \hat{z} + 3\hat{y}^2 \hat{z} &= -6\hat{y}y_1 y_2 \text{ in } \Omega \\ \hat{z} &= 0 \text{ on } \Gamma \end{aligned}$$

with $G(\hat{u}) = \hat{y}$ and $G'(\hat{u})u_i = y_i$ for $i = 1, 2$.

5.1 Verifying the assumptions

We compute the constants of Chapter 4 for this class of examples. These computations are quite technical even for this rather simple case as we will see. At the end of this chapter we will give an overview of the results for all involved constants. We start with several underlying constants, which we utilize for the desired estimates.

5.1.1 Lagrange operator and imbedding constants

We look at the imbedding constants I_p of the imbeddings $H_0^1([0,1]) \hookrightarrow L^p([0,1])$, i.e.

$$\|y\|_{L^p([0,1])} \leq I_p \|y\|_{H^1([0,1])} \; \forall y \in H_0^1([0,1]).$$

Computation of I_p

We start with the derivation of I_2, which we will then use to compute the imbedding constants for $L^p([0,1])$ with $p > 2$.
Note that

$$\|\sin(n\pi x)\|_{L^2([0,1])}^2 = \int_0^1 \sin^2(n\pi x)dx = \frac{1}{2}$$

and

$$\|\sin(n\pi x)\|^2_{H^1([0,1])} = \int_0^1 \sin^2(n\pi x) + n^2\pi^2\cos^2(n\pi x)dx = \frac{1}{2} + n^2\pi^2\frac{1}{2} = \frac{1+n^2\pi^2}{2}$$

hold. This leads to

$$\|\sin(n\pi x)\|_{L^2([0,1])} = \frac{1}{\sqrt{1+n^2\pi^2}}\|\sin(n\pi x)\|_{H^1([0,1])} \leq \frac{1}{\sqrt{1+\pi^2}}\|\sin(n\pi x)\|_{H^1([0,1])}$$

We set $y = \sum_{n=0}^{\infty} y_n \sin(n\pi x)$ and get :

$$\begin{aligned}
\|y\|^2_{L^2([0,1])} &= \int_0^1 y^2 dx = \int_0^1 \sum_{n=0}^{\infty}(y_n \sin(n\pi x))^2 dx \\
&= \sum_{n=0}^{\infty} y_n^2 \int_0^1 \sin^2(n\pi x)\ \mathrm{dx} \\
&= \sum_{n=0}^{\infty} y_n^2 \|\sin(n\pi x)\|^2_{L^2([0,1])} \\
&= \sum_{n=0}^{\infty} y_n^2 \frac{1}{1+n^2\pi^2}\|\sin(n\pi x)\|^2_{H^1([0,1])} \\
&\leq \frac{1}{1+\pi^2}\sum_{n=0}^{\infty} y_n^2 \|\sin(n\pi x)\|^2_{H^1([0,1])} \\
&= \frac{1}{1+\pi^2}\int_0^1 \sum_{n=0}^{\infty} y_n^2 \sin^2(n\pi x) + y_n^2 n^2\pi^2\cos^2(n\pi x)\ \mathrm{dx} \\
&= \frac{1}{1+\pi^2}\int_0^1 y^2 + (y')^2 dx \\
&= \frac{1}{1+\pi^2}\|y\|^2_{H^1([0,1])}
\end{aligned}$$

Thus we see that

$$\|y\|_{L^2([0,1])} \leq \frac{1}{\sqrt{1+\pi^2}}\|y\|_{H^1([0,1])}$$

holds. This leads to the following imbedding constants:

$$\begin{aligned}
I_2 &= \frac{1}{\sqrt{\pi^2+1}} \approx 0.3033 \\
I_4 &= (\frac{1}{2}I_2^2)^{\frac{1}{4}} = 2^{-\frac{1}{4}} I_2^{\frac{1}{2}} \approx 0.4631 \\
I_6 &= (\frac{9}{8}I_4^4)^{\frac{1}{6}} = (\frac{9}{16})^{\frac{1}{6}} I_2^{\frac{1}{3}} \approx 0.6105
\end{aligned}$$

Computation of c_∞

In the next step we look at the imbedding of $H^1([0,1]) \hookrightarrow L^\infty([0,1])$, on which we will rely for several of the other constants :

$$\|z - \bar{z}_h\|_{L^\infty([0,1])} \leq c_\infty \|z - \bar{z}_h\|_{H^1([0,1])} \tag{5.3}$$

Let y be an arbitrary function in $H_0^1([0,1])$, then there exists a $x_0 \in [0,1]$ such that $\|y\|_{L^\infty([0,1])} = |y(x_0)|$. We split the estimation process in two cases:
Case 1 : $x_0 \in [0, \frac{1}{2}]$

$$\begin{aligned}
\|y\|_{L^\infty([0,1])} &= |y(x_0)| = |y(x_0)| - y(0) \\
&= \int_0^{x_0} |y'(t)| \text{ dt} \leq \sqrt{\int_0^{x_0} 1 \text{ dt}} \sqrt{\int_0^{x_0} y'(t)^2 \text{ dt}} \\
&\leq \sqrt{\int_0^{\frac{1}{2}} 1 \text{ dt}} \sqrt{\int_0^1 y'(t)^2 \text{ dt}} \leq \frac{1}{\sqrt{2}} \|y\|_{H^1([0,1])}
\end{aligned}$$

Case 2 : $x_0 \in (\frac{1}{2}, 1]$

$$\begin{aligned}
\|y\|_{L^\infty([0,1])} &= |y(x_0)| = |y(x_0)| - y(1) \\
&\leq \int_{x_0}^1 |y'(t)| \text{ dt} \leq \sqrt{\int_{x_0}^1 1 \text{ dt}} \sqrt{\int_{x_0}^1 y'(t)^2 \text{ dt}} \\
&\leq \sqrt{\int_{\frac{1}{2}}^1 1 \text{ dt}} \sqrt{\int_0^1 y'(t)^2 \text{ dt}} \leq \frac{1}{\sqrt{2}} \|y\|_{H^1([0,1])}
\end{aligned}$$

Hence we get

$$c_\infty = \frac{1}{\sqrt{2}}$$

Properties of the Lagrange operator

We end this section with some properties of the Lagrange operator.
For

$$Ay = -\Delta y = -y_{xx}$$

there exist constants δ_0 and δ_1, such that

$$\begin{aligned}
\delta_0 \|y\|^2_{H^1([0,1])} &\leq \langle Ay, y \rangle \\
\langle Ay_1, y_2 \rangle &\leq \delta_1 \|y_1\|_{H^1([0,1])} \|y_2\|_{H^1([0,1])} \ \forall y \in H_0^1([0,1])
\end{aligned}$$

hold. We see that

$$\langle Ay_1, y_2 \rangle = \langle -\Delta y_1, y_2 \rangle = \langle \nabla y_1, \nabla y_2 \rangle \leq \|y_1\|_{H^1([0,1])} \|y_2\|_{H^1([0,1])}$$

holds, which means $\delta_1 = 1$. Furthermore we see that

$$\langle Ay, y\rangle = \|\nabla y\|^2_{L^2([0,1])} = \|y\|^2_{H^1([0,1])} - \|y\|^2_{L^2([0,1])} \geq (1 - I_2^2)\|y\|^2_{H^1([0,1])}$$

holds. Thus we get $\delta_0 = 1 - I_2^2$.
For δ_0 we get :

$$\delta_0 = 1 - I_2^2 \approx 0.9080$$

5.1.2 Computation of c_S

We look at the constant c_S from [41] Theorem 4.1:

$$\|y\|_{H^2([0,1])} \leq c_S \|u\|_{L^2([0,1])}$$

We recall the ODE:

$$-y_{xx} + y + y^3 = u \;\Rightarrow\; y_{xx} = y + y^3 - u$$

This leads to

$$\begin{aligned} y_{xx}^2 &= y^2 + y^3 - uy + y^4 + y^6 - y^3u - uy - uy^3 + u^2 \\ &= y^6 + 2y^4 + y^2 - 2y^3u - 2uy + u^2 \\ &= (y^4 + 2y^2 + 1)y^2 - 2uy(y^2 + 1) + u^2. \end{aligned}$$

Furthermore

$$\int_0^1 y^2 + y_x^2 + y^4 \text{ dx} = \int_0^1 uy \text{ dx}$$

holds. Thus we get

$$\begin{aligned} \|y\|^2_{H^2([0,1])} &= \int_0^1 y^2 + y_x^2 + y_{xx}^2 \text{ dx} = \int_0^1 y^2 + y_x^2 + y^4 + y_{xx}^2 - y^4 \text{ dx} \\ &= \int_0^1 uy + (y^4 + 2y^2 + 1)y^2 - 2uy(y^2 + 1) + u^2 - y^4 \text{ dx} \\ &= \int_0^1 -(2y^2 + 1)uy + (y^4 + y^2 + 1)y^2 + u^2 \text{ dx} \\ &= \int_0^1 -(2y^2 + 1)(y^2 + y_x^2 + y^4) + (y^4 + y^2 + 1)y^2 + u^2 \text{ dx} \\ &= \int_0^1 -2y^4 - 2y^2y_x^2 - 2y^6 - y^2 - y_x^2 - y^4 + y^6 + y^4 + y^2 + u^2 \text{ dx} \\ &= \int_0^1 u^2 \underbrace{-y^6 - 2y^4 - 2y^2y_x^2 - y_x^2}_{\leq 0} \text{ dx} \leq \int_0^1 u^2 \text{ dx} \\ &\leq \|u\|^2_{L^2([0,1])} \end{aligned}$$

$\Rightarrow$

$$c_S \leq 1$$

5.1.3 Computation of c_{inv}

We want to compute the inverse estimate constant c_{inv}:

$$\|v_h\|_{L^\infty([0,1])} \le \frac{c_{inv}}{\sqrt{h}} \|v_h\|_{L^2([0,1])} \quad \forall v_h \in V_h$$

In order to do this we will utilize two known results. Regarding symmetric matrices we want to recall the Rayleigh quotient and its properties:

Definition 5.2. For a given matrix $A \in \mathbb{R}^{n\times n}$ and a nonzero vector $x \in \mathbb{R}^n$ the Rayleigh quotient $R(A,x)$ is defined as

$$R(A,x) = \frac{x^T A x}{x^T x}.$$

Theorem 5.3. The Raleigh quotient fulfills

$$\lambda_{min} \le R(A,x) \le \lambda_{max}, x \in \mathbb{R}^n \setminus \{0\}$$

for a symmetric matrix $A \in \mathbb{R}^{n\times n}$, its smallest eigenvalue λ_{min} and its largest eigenvalue λ_{max}.

And regarding estimates of eigenvalues we quote Gershgorins circle theorem:

Theorem 5.4. For every diagonal entry a_{ii} of $A \in \mathbb{R}^{n\times n}$ the Gershgorin circle is defined as:

$$\bar{B}_i := \bar{B}(a_{ii}, \sum_{j=1, j\neq i}^{n} |a_{ij}|)$$

and the spectrum of A lies in $\cup_{i=1}^n \bar{B}_i$

Now we derive c_{inv}:
Let $v_h \in V_h$, then v_h can be written as $v_h = \sum_{i=1}^n v_i\phi_i$ and the L^2-norm can be derived as

$$\|v_h\|^2_{L^2([0,1])} = \underline{v}^T M \underline{v}$$

with $\underline{v} = (v_i)_{i=1..n}$ and $M = (m_{ij})_{i,j=1..n}$ with $m_{ij} = \langle \phi_i, \phi_j \rangle$. We define an L^2-type norm as

$$\|v_h\|^2_{equ} = h\underline{v}^T \underline{v}$$

and prove that this norm is equivalent to the L^2-norm :
We set

$$A = \frac{1}{6}\begin{pmatrix} 4 & 1 & & & \\ 1 & 4 & 1 & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & & 1 & 4 \end{pmatrix}$$

and use the Rayleigh quotient to get

$$\lambda_{min}\underline{v}^T\underline{v} \leq \underline{v}^T A\underline{v} \leq \lambda_{max}\underline{v}^T\underline{v}.$$

Thus we get

$$h\lambda_{min}\underline{v}^T\underline{v} \leq \underline{v}^T M\underline{v} \leq h\lambda_{max}\underline{v}^T\underline{v}.$$

Using Gershgorins circle theorem we get the bounds

$$\lambda_{min} \geq \frac{1}{3} \text{ and } \lambda_{max} \leq 1.$$

This leads to

$$\begin{aligned} &\frac{1}{3}\|v_h\|^2_{equ} \leq \|v_h\|^2_{L^2([0,1])} \leq \|v_h\|^2_{equ} \\ \Rightarrow &\|v_h\|_{equ} \leq \sqrt{3}\|v_h\|_{L^2([0,1])} \leq \sqrt{3}\|v_h\|_{equ}. \end{aligned}$$

Now we look at the minimization problem

$$\begin{aligned} \min \sqrt{h\sum_{i=1}^n v_i^2} &= \|v_h\|_{equ} \\ \text{s.t. } \|v_h\|_{L^\infty([0,1])} &= \max_i v_i = m. \end{aligned}$$

It is clear that $\bar{v}$ with $\bar{v}_i = m$ for one certain $i \in 1,..,n$ and $\bar{v}_j = 0,\ \forall j \neq i$, is a solution of this problem, which leads us to

$$\|\bar{v}\|_{equ} = \sqrt{h}m \leq \|v_h\|_{equ} \quad \forall v_h \in V_h \text{ with } \|v_h\|_{L^\infty([0,1])} = m.$$

Thus we get:

$$\|v_h\|_{L^\infty([0,1])} = m = \frac{\sqrt{h}m}{\sqrt{h}} \leq \frac{\|v_h\|_{equ}}{\sqrt{h}} \leq \frac{\sqrt{3}}{\sqrt{h}}\|v_h\|_{L^2([0,1])}$$

which means

$$c_{inv} = \sqrt{3}.$$

5.1.4 Computation of N

We consider Assumption (3.26)

$$\begin{aligned} \|[G''(u) - G''(\bar{u}_h)][v_1, v_2]\|_{H^1([0,1])} \leq N\|u - \bar{u}_h\|_{L^2([0,1])}\|v_1\|_U\|v_2\|_{L^2[(0,1])}, \\ \forall u \in U_{ad},\ \|u - \bar{u}_h\|_{L^2([0,1])} \leq R,\ v_1, v_2 \in L^2([0,1]) \end{aligned}$$

and want to compute N. From (5.2) we get

$$(G''(u) - G''(\bar{u}_h))[u_1, u_2] = z - \bar{z}_h$$

where $z - \bar{z}_h$ is the weak solution of

$$-\Delta(z-\bar{z}_h) + (z-\bar{z}_h) + 3(y^2 z - \bar{y}_h^2 \bar{z}_h) = -6yy_1y_2 + 6\bar{y}_h y_{h1} y_{h2} \text{ in } [0,1]$$
$$z(0) - \bar{z}_h(0) = z(1) - \bar{z}_h(1) = 0$$

with $G(u) = y$, $G(\bar{u}_h) = \bar{y}_h$, $G'(u)u_i = y_i$ and $G'(\bar{u}_h)u_2 = y_{hi}$ for $i = 1, 2$.
This leads to the weak formulation

$$\int_0^1 \nabla(z-\bar{z}_h)\nabla v + (z-\bar{z}_h)v + 3(y^2 z - \bar{y}_h^2\bar{z}_h)v \text{ dx} =$$
$$\int_0^1 -6yy_1y_2v + 6\bar{y}_h y_{h1}y_{h2}v \text{ dx}, \ \forall v \in H^1([0,1]).$$

With $v = z - \bar{z}_h$ we get

$$\int_0^1 \nabla(z-\bar{z}_h)^2 + (z-\bar{z}_h)^2 + 3(y^2 z - \bar{y}_h^2\bar{z}_h)(z-\bar{z}_h) \text{ dx} =$$
$$\int_0^1 -6(z-\bar{z}_h)(yy_1y_2 - \bar{y}_h y_{h1}y_{h2}) \text{ dx}.$$

This is equivalent to:

$$\int_0^1 \nabla(z-\bar{z}_h)^2 \text{ dx} + \langle d'(x,y)z - d'(x,\bar{y}_h)\bar{z}_h, z-\bar{z}_h\rangle =$$
$$\int_0^1 -6(z-\bar{z}_h)(yy_1y_2 - \bar{y}_h y_{h1}y_{h2}) \text{ dx}$$

We see that the following equation applies:

$$\begin{aligned}\langle d'(x,y)z - d'(x,\bar{y}_h)\bar{z}_h, z-\bar{z}_h\rangle &= \langle d'(x,\bar{y}_h)(z-\bar{z}_h), z-\bar{z}_h\rangle \\ &\quad + \langle d'(x,y) - d'(x,\bar{y}_h)z, z-\bar{z}_h\rangle \\ &= \int_0^1 (z-\bar{z}_h)^2 + 3\bar{y}_h^2(z-\bar{z}_h)^2 \\ &\quad + 3y^2 z(z-\bar{z}_h) - 3\bar{y}_h^2 z(z-\bar{z}_h) \text{ dx}\end{aligned}$$

This results in

$$\begin{aligned}\|z-\bar{z}_h\|^2_{H^1([0,1])} + \int_0^1 3\bar{y}_h^2(z-\bar{z}_h)^2 \text{ dx} = -\int_0^1 3(z-\bar{z}_h)[(y^2 z - \bar{y}_h^2 z) \\ + 2yy_1y_2 - 2\bar{y}_h y_{h1}y_{h2}] \text{ dx}\end{aligned}$$

and we gain

$$\|z-\bar{z}_h\|^2_{H^1([0,1])} \leq |3\int_0^1 (z-\bar{z}_h)[y^2 z - \bar{y}_h^2 z + 2yy_1y_2 - 2\bar{y}_h y_{h1}y_{h2}] \text{ dx}|$$

from the fact that $\int_0^1 3\bar{y}_h^2(z-\bar{z}_h)^2 \text{ dx} \geq 0$. We get the estimate

$$\|z-\bar{z}_h\|^2_{H^1([0,1])} \leq 3\|z-\bar{z}_h\|_{L^\infty([0,1])}(\int_0^1 |y^2z-\bar{y}_h^2 z+2yy_1y_2-2\bar{y}_h y_{h1}y_{h2}| \text{ dx})$$
$$\leq 3c_\infty\|z-\bar{z}_h\|_{H^1([0,1])}[|\langle y^2-\bar{y}_h^2, z\rangle|+2\int_0^1 |yy_1y_2-\bar{y}_h y_{h1}y_{h2}| \text{ dx}]$$

which leads to

$$\|z-\bar{z}_h\|_{H^1([0,1])} \leq 3c_\infty|\langle y^2-\bar{y}_h^2, z\rangle|+6c_\infty\int_0^1 |yy_1y_2-\bar{y}_h y_{h1}y_{h2}| \text{ dx}|$$
$$= 3c_\infty|\langle y^2-\bar{y}_h^2, z\rangle|+6c_\infty\int_0^1 |y(y_1y_2-y_{h1}y_2)$$
$$+ yy_{h1}y_2-\bar{y}_h(y_{h1}y_{h2}-y_{h1}y_2)-\bar{y}_h y_{h1}y_2| \text{ dx}|$$
$$= 3c_\infty|\langle y^2-\bar{y}_h^2, z\rangle|+6c_\infty\int_0^1 |y([G'(u)-G'(\bar{u}_h)]u_1G'(u)u_2)$$
$$- \bar{y}_h([G'(\bar{u}_h)-G'(u)]u_2G'(\bar{u}_h)u_1)$$
$$+ (y-\bar{y}_h)(G'(\bar{u}_h)u_1G'(u)u_2)| \text{ dx}|$$
$$\leq 3c_\infty\Big[|\langle y^2-\bar{y}_h^2, z\rangle|$$
$$+ 2\big[\|y\|_{L^\infty([0,1])}\|[G'(u)-G'(\bar{u}_h)]u_1\|_{L^\infty([0,1])}\|G'(u)\|\|u_2\|_{L^2([0,1])}$$
$$+ \|\bar{y}_h\|_{L^\infty([0,1])}\|[G'(\bar{u}_h)-G'(u)]u_2\|_{L^\infty([0,1])}\|G'(\bar{u}_h)\|\|u_1\|_{L^2([0,1])}$$
$$+ \|y-\bar{y}_h\|_{L^\infty([0,1])}\|G'(\bar{u}_h)\|_{L^2([0,1])\to L^\infty([0,1])}\|u_1\|_{L^2([0,1])}$$
$$\|G'(u)\|_{L^2([0,1])\to L^\infty([0,1])}\|u_2\|_{L^2([0,1])}\big]\Big].$$

We assume

$$\|G'(u)\|_{L^2([0,1])\to L^\infty([0,1])} \leq K' \tag{5.4}$$
$$\|G''(u)\|_{L^2([0,1])\to \mathcal{L}(L^2([0,1]),L^\infty([0,1]))} \leq K'' \tag{5.5}$$
$$\|Gu-G\bar{u}_h\|_{L^\infty([0,1])} \leq L_G\|u-\bar{u}_h\|_{L^2([0,1])} \tag{5.6}$$
$$\|y^2-\bar{y}_h^2\|_{L^2([0,1])} = \|G(u)^2-G(\bar{u}_h)^2\|_{L^2([0,1])} \leq L_{G^2}\|u-\bar{u}_h\|_{L^2([0,1])} \tag{5.7}$$
$$\|[G'(\bar{u}_h)-G'(u)]u_i\|_{L^\infty([0,1])} \leq L_{G'}\|u-\bar{u}_h\|_{L^2([0,1])}\|u_i\|_{L^2([0,1])} \tag{5.8}$$

with $\|u-\bar{u}_h\|_{L^2([0,1])} \leq R$ for (5.3)-(5.8) and $u \in U_{ad}$ for (5.6)-(5.8) and get

$$\|z-\bar{z}_h\|_{H^1([0,1])} \leq 3c_\infty(\|y^2-\bar{y}_h^2\|_{L^2([0,1])}\|G''(u)\|\|u_1\|_{L^2([0,1])}\|u_2\|_{L^2([0,1])}$$
$$+ 2(2\|c_c\|_{L^\infty([0,1])}L_{G'}K'\|u-\bar{u}_h\|_{L^2([0,1])}\|u_1\|_{L^2([0,1])}\|u_2\|_{L^2([0,1])}$$
$$+ L_G(K')^2\|u-\bar{u}_h\|_{L^2([0,1])}\|u_1\|_{L^2([0,1])}\|u_2\|_{L^2([0,1])}))$$
$$\leq 3c_\infty\|u-\bar{u}_h\|_{L^2([0,1])}\|u_1\|_{L^2([0,1])}\|u_2\|_{L^2([0,1])}$$
$$(L_{G^2}K''+4\|c_c\|_{L^\infty([0,1])}L_{G'}K'+2L_G(K')^2). \tag{5.9}$$

Computation of K'

To compute K' we consider the weak formulation of (5.1)

$$(\nabla y, \nabla v) + (y, v) + (3\hat{y}^2 y, v) = (u, v)$$

and set $v = y$ which leads to

$$(\nabla y, \nabla y) + (y, y) + \underbrace{(3\hat{y}^2 y, y)}_{\geq 0} = (u, y)$$

$\Rightarrow$

$$\|y\|^2_{H^1([0,1])} \leq (u, y) \leq \|u\|_{L^2([0,1])}\|y\|_{L^2([0,1])} \leq \|u\|_{L^2([0,1])}\|y\|_{H^1([0,1])}$$

$\Rightarrow$

$$\|y\|_{H^1([0,1])} \leq \|u\|_{L^2([0,1])} \Rightarrow \ \|G'(\hat{u})\| \leq K' = 1$$

Computation of K''

To compute K'' we use the same strategy:
We consider the weak formulation belonging to (5.2)

$$(\nabla \hat{z}, \nabla v) + (\hat{z}, v) + (3\hat{y}^2 \hat{z}, v) = (-6\hat{y}y_1y_2, v)$$

and set $v = \hat{z} \Rightarrow$

$$(\nabla \hat{z}, \nabla \hat{z}) + (\hat{z}, \hat{z}) + \underbrace{(3\hat{y}^2 \hat{z}, \hat{z})}_{\geq 0} = (-6\hat{y}y_1y_2, \hat{z})$$

$\Rightarrow$

$$\begin{aligned}
\|\hat{z}\|^2_{H^1([0,1])} &\leq |(-6\hat{y}y_1y_2, \hat{z})| \\
&\leq \|\hat{z}\|_{L^\infty([0,1])} \int_0^1 |6G\hat{u}G'(\hat{u})u_1G'(\hat{u})u_2| \ \mathrm{dx} \\
&\leq 6c_\infty \|\hat{z}\|_{H^1([0,1])}\|G\hat{u}\|_{L^\infty([0,1])}K'\|u_1\|_{L^2([0,1])}K'\|u_2\|_{L^2([0,1])}
\end{aligned}$$

$\Rightarrow$

$$\begin{aligned}
\|\hat{z}\|_{H^1([0,1])} &\leq 6c_\infty\|G\hat{u}\|_{L^\infty([0,1])}\|u_1\|_{L^2([0,1])}\|u_2\|_{L^2([0,1])} \\
&\leq 6c_\infty c_{SL^\infty}\|\hat{u}\|_{L^2([0,1])}\|u_1\|_{L^2([0,1])}\|u_2\|_{L^2([0,1])}
\end{aligned}$$

$\Rightarrow$

$$\|G''(\hat{u})\| \leq K'' = 6c_\infty c_{SL^\infty}$$

with

$$c_{SL^\infty} = 4\frac{I_6^2}{\delta_0}|\Omega|^{\frac{1}{6}}.$$

(See [41] Corollary 8.1)
With

$$c_{SL^\infty} \approx 1.6419$$

we get

$$K'' \approx 6.9660.$$

Computation of L_G

For (5.6) we get from [41] Lemma 4.2 that

$$L_G = \frac{I_2}{\delta_0} \approx 0.3340 \tag{5.10}$$

holds.

Computation of L_{G^2}

For (5.7) we see that

$$\begin{aligned}\|y^2 - \bar{y}_h^2\|_{L^2([0,1])} &\leq \|y - \bar{y}_h\|_{L^4([0,1])}\|y + \bar{y}_h\|_{L^4([0,1])} \\ &\leq \|y - \bar{y}_h\|_{L^\infty([0,1])}\|y + \bar{y}_h\|_{L^\infty([0,1])} \\ &\leq L_G\|u - \bar{u}_h\|_{L^2([0,1])}\|y + \bar{y}_h\|_{L^\infty([0,1])}\end{aligned}$$

holds. Thus we need to estimate $\|y + \bar{y}_h\|_{L^\infty([0,1])}$:

$$\|y + \bar{y}_h\|_{L^\infty([0,1])} \leq \|y\|_{L^\infty([0,1])} + \|\bar{y}_h\|_{L^\infty([0,1])}$$

This means that we have to find an estimate of $\|y\|_{L^\infty([0,1])}$ for all y with $G(u) = y$ and $\|u - \bar{u}_h\|_{L^2([0,1])} \leq R$. We know that

$$\|y\|_{L^\infty([0,1])} \leq \frac{1}{\sqrt{2}}\|y\|_{H^1([0,1])}$$

holds. Hence we get

$$\|y + \bar{y}_h\|_{L^\infty([0,1])} \leq \frac{1}{\sqrt{2}}(\|y\|_{H^1([0,1])} + \|\bar{y}_h\|_{H^1([0,1])})$$

Now we want to derive a upper bound for $\|y\|_{H^1([0,1])}$ and $\|\bar{y}_h\|_{H^1([0,1])}$ respectively: We consider the weak formulation of the underlying problem:

$$(\nabla y, \nabla v) + (y, v) + (y^3, v) = (u, v)$$

With $v = y$ we get:

$$\begin{aligned}
\|y\|^2_{H^1([0,1])} + \underbrace{(y^3, y)}_{\geq 0} &= (u, y) \\
\Rightarrow \ \|y\|^2_{H^1([0,1])} &\leq \|u\|_{L^2([0,1])}\|y\|_{H^1([0,1])} \\
\Rightarrow \ \|y\|_{H^1([0,1])} &\leq \|u\|_{L^2([0,1])}
\end{aligned}$$

Thus we need to find an estimate for $\|u\|_{L^2([0,1])}$. Obviously

$$\|u\|_{L^2([0,1])} \leq \|\bar{u}_h\|_{L^2([0,1])} + R$$

holds $\forall u \in B_R(\bar{u}_h)$. Hence,

$$\|y\|_{H^1([0,1])} \leq \|\bar{u}_h\|_{L^2([0,1])} + R$$

holds. This leads to

$$\|y + \bar{y}_h\|_{L^\infty([0,1])} \leq \sqrt{2}(\|\bar{u}_h\|_{L^2([0,1])} + R).$$

Finally we see that

$$\|y^2 - \bar{y}_h^2\|_{L^2([0,1])} \leq \sqrt{2}L_G\|u - \bar{u}_h\|_{L^2([0,1])}(\|\bar{u}_h\|_{L^2([0,1])} + R)$$

holds and thus we get

$$L_{G^2} = \sqrt{2}L_G(\|\bar{u}_h\|_{L^2([0,1])} + R).$$

Computation of $L_{G'}$

$L_{G'}$ from (5.8) can be derived as follows:
We denote

$$G'(u)u_i = \hat{y}$$

and

$$G'(\bar{u}_h)u_i = \tilde{y}$$

as the weak solutions of

$$-\Delta\hat{y} + \hat{y} + 3y^2\hat{y} = u_i$$

respectively

$$-\Delta\tilde{y} + \tilde{y} + 3\bar{y}_h^2\tilde{y} = u_i.$$

This leads to

$$-\Delta(\hat{y} - \tilde{y}) + \hat{y} - \tilde{y} + 3(y^2\hat{y} - \bar{y}_h^2\tilde{y}) = 0$$

and the weak formulation reads as

$$(\nabla(\hat{y}-\tilde{y}),\nabla v)+\underbrace{(\hat{y}-\tilde{y},v)+(3(y^2\hat{y}-\bar{y}_h^2\tilde{y},v)}_{=\langle d'(x,y)\hat{y}-d'(x,\bar{y}_h)\tilde{y},v\rangle}=0,\ \forall v\in H_0^1([0,1]).$$

We set $v=\hat{y}-\tilde{y}$ and get

$$\|\nabla(\hat{y}-\tilde{y})\|^2+\langle d'(x,y)\hat{y}-d'(x,\bar{y}_h)\tilde{y},\hat{y}-\tilde{y}\rangle=0$$

We consider the second term of this equation:

$$\begin{aligned}\langle d'(x,y)\hat{y}-d'(x,\bar{y}_h)\tilde{y},\hat{y}-\tilde{y}\rangle &= \langle d'(x,\bar{y}_h)(\hat{y}-\tilde{y}),\hat{y}-\tilde{y}\rangle\\ &+\langle d'(x,y)-d'(x,\bar{y}_h)\hat{y},\hat{y}-\tilde{y}\rangle\\ &=\int_0^1(\hat{y}-\tilde{y})^2+3\bar{y}_h^2(\hat{y}-\tilde{y})^2+3(y^2-\bar{y}_h^2)\hat{y}(\hat{y}-\tilde{y})\mathrm{dx}\end{aligned}$$

This results in:

$$\|\nabla(\hat{y}-\tilde{y})\|^2+\|\hat{y}-\tilde{y}\|^2+\underbrace{\int_0^1 3\bar{y}_h^2(\hat{y}-\tilde{y})^2}_{\geq 0}=-3\int_0^1(y^2-\bar{y}_h^2)\hat{y}(\hat{y}-\tilde{y})\mathrm{dx}$$

$$\begin{aligned}\Rightarrow\|\hat{y}-\tilde{y}\|^2_{H^1([0,1]}&\leq|3\int_0^1(y^2-\bar{y}_h^2)\hat{y}(\hat{y}-\tilde{y})\mathrm{dx}|\\ &\leq 3\|y^2-\bar{y}_h^2\|_{L^2([0,1])}\|\hat{y}\|_{L^2([0,1])}\|\hat{y}-\tilde{y}\|_{L^\infty([0,1])}\\ &\leq 3\|y^2-\bar{y}_h^2\|_{L^2([0,1])}\|G'(u)\|_{\mathcal{L}}\|u_i\|_{L^2([0,1])}c_\infty\|\hat{y}-\tilde{y}\|_{H^1([0,1])}\\ \Rightarrow\|\hat{y}-\tilde{y}\|_{H^1([0,1])}&\leq 3L_{G^2}c_\infty\|u-\bar{u}_h\|K'\|u_i\|_{L^2([0,1])}\\ \Rightarrow L_{G'}=3c_\infty K'L_{G^2}&\end{aligned}$$

Final result

If we combine all the results above we get :

$$\begin{aligned}N&=3c_\infty(L_{G^2}K''+4\|c_c\|_{L^\infty([0,1])}L_{G'}+2L_G(K')^2)\\ &=\frac{3}{\sqrt{2}}\Bigg[24(\frac{9}{16})^{\frac{1}{3}}\frac{(\pi^2+1)(\pi^2+1)^{\frac{1}{6}}}{\pi^4}(\|\bar{u}_h\|_{L^2([0,1])}+R)\\ &+12\|c_c\|_{L^\infty([0,1])}\frac{\sqrt{\pi^2+1}}{\pi^2}(\|\bar{u}_h\|_{L^2([0,1])}+R)+2\frac{\sqrt{\pi^2+1}}{\pi^2}\Bigg]\end{aligned}$$

5.1.5 Computation of L

We consider Assumption (3.24)

$$\begin{aligned}\|[G_h'(u) - G_h'(\bar{u}_h)]v\|_{H^1([0,1)]} \leq L\|u - \bar{u}_h\|_{L^2([0,1])}\|v\|_{L^2([0,1])}, \, \forall v \in L^2([0,1]),\\ \forall u \in L^2([0,1]) : \|u - \bar{u}_h\|_{L^2([0,1])} \leq \|\tilde{u}_h - \bar{u}_h\|_{L^2([0,1])}\end{aligned}$$

and compute L in a analogous way to N. We set

$$G_h'(u)w = z_u \;:\; -\Delta z_u + z_u + 3\hat{y}_u^2 z_u = w$$

with $G_h(u) = \hat{y}_u$ and get

$$[G_h'(u) - G_h'(v)]w = z_u - z_v \;:\; -\Delta(z_u - z_v) + (z_u - z_v) + 3\hat{y}_u^2 z_u - 3\hat{y}_v^2 z_v = 0$$

$\forall u, v \in U$. We consider the weak formulation

$$(\nabla(z_u - z_v), t) + (z_u - z_v, t) + 3(\hat{y}_u^2 z_u - \hat{y}_v^2 z_v, t) = 0, \; \forall t \in Z_h$$

and with $t = z_u - z_v$ we get

$$(\nabla(z_u - z_v), \nabla(z_u - z_v)) + (z_u - z_v, z_u - z_v) + 3(\hat{y}_u^2 z_u - \hat{y}_v^2 z_v, z_u - z_v) = 0$$

$\Rightarrow$

$$\int_0^1 \nabla(z_u - z_v)^2 + (z_u - z_v)^2 + 3(\hat{y}_u^2 z_u - \hat{y}_v^2 z_v)(z_u - z_v) \text{ dx} = 0$$

$\Rightarrow$

$$\int_0^1 \nabla(z_u - z_v)^2 \text{ dx} + \langle d'(x, \hat{y}_u)z_u - d'(x, \hat{y}_v)z_v, z_u - z_v\rangle = 0$$

For $\langle d'(x, \hat{y}_u)z_u - d'(x, \hat{y}_v)z_v, z_u - z_v\rangle$ we get the following equation:

$$\begin{aligned}\langle d'(x, \hat{y}_u)z_u - d'(x, \hat{y}_v)z_v, z_u - z_v\rangle &= \langle d'(x, \hat{y}_v)(z_u - z_v), z_u - z_v\rangle\\ &\quad + \langle [d'(x, \hat{y}_u) - d'(x, \hat{y}_v)]z_u, z_u - z_v\rangle\\ &= \int_0^1 (z_u - z_v)^2 + 3\hat{y}_v^2(z_u - z_v)^2\\ &\quad + 3(\hat{y}_u^2 - \hat{y}_v^2)z_u(z_u - z_v) \text{ dx}\end{aligned}$$

$\Rightarrow$

$$\int_0^1 \nabla(z_u - z_v)^2 + (z_u z_v)^2 + 3\hat{y}_v^2(z_u - z_v)^2 + 3(\hat{y}_u^2 - \hat{y}_v^2)z_u(z_u - z_v) \text{ dx} = 0$$

which is equivalent to

$$\|z_u - z_v\|_{H^1([0,1])}^2 + \underbrace{\int_0^1 3\hat{y}_v^2(z_u - z_v)^2 \text{ dx}}_{\geq 0} + \int_0^1 3(\hat{y}_u^2 - \hat{y}_v^2)z_u(z_u - z_v) \text{ dx} = 0$$

This leads to

$$\begin{aligned}\|z_u - z_v\|^2_{H^1([0,1])} &\leq |-3\int_0^1 (\hat{y}_u^2 - \hat{y}_v^2) z_u (z_u - z_v)\ \mathrm{dx}| \\ &\leq 3\|z_u - z_v\|_{L^\infty([0,1])} \int_{[0,1]} |(\hat{y}_u^2 - \hat{y}_v^2) z_u|\ \mathrm{dx} \\ &\leq 3c_\infty \|z_u - z_v\|_{H^1([0,1])} \|\hat{y}_u^2 - \hat{y}_v^2\|_{L^2(\Omega)} \|z_u\|_{L^2([0,1])}\end{aligned}$$

Thus we get

$$\begin{aligned}\|z_u - z_v\|_{H^1([0,1])} &\leq \frac{3}{\sqrt{2}} \|G_h(u)^2 - G_h(v)^2\|_{L^2([0,1])} \|G_h'(u)w\|_{L^2([0,1])} \\ &\leq \frac{3}{\sqrt{2}} L_{G_h^2} \|u - v\|_{L^2([0,1])} K_h' \|w\|_{L^2([0,1])}\end{aligned}$$

under the assumptions

$$\|G_h(u)^2 - G_h(v)^2\|_{L^2([0,1])} \leq L_{G_h^2} \|u - v\|_{L^2([0,1])} \tag{5.11}$$

$$\|G_h'(u)\| \leq K_h' \tag{5.12}$$

Computation of K_h'

We consider the weak formulation

$$(\nabla z_u, \nabla t) + (z_u, t) + (3\hat{y}_u^2 z_u, t) = (w, t)$$

to verify (5.12), set $t = z_u$ and get

$$(\nabla z_u, \nabla z_u) + (z_u, z_u) + \underbrace{(3\hat{y}_u^2 z_u, z_u)}_{\geq 0} = (w, z_u).$$

This leads to

$$\|z_u\|^2_{H^1([0,1])} \leq (w, z_u) \leq \|w\|_{L^2([0,1])} \|z_u\|_{L^2([0,1])} \leq \|w\|_{L^2([0,1])} \|z_u\|_{H^1([0,1])}$$

$\Rightarrow$

$$\|z_u\|_{H^1([0,1])} \leq \|w\|_{L^2([0,1])}$$

Thus we see that

$$\|G_h'(u)\| \leq 1$$

holds and get

$$K_h' = 1.$$

Computation of L_{G^2}

To proof (5.11) we fix v at $\bar{u}_h$ and consider only $u \in U$ which fulfill $\|u - \bar{u}_h\|_{L^2([0,1])} \leq \|\tilde{u}_h - \bar{u}_h\|_{L^2([0,1])}$. We get :

$$\begin{aligned}
\|G_h(u)^2 - G_h(v)^2\|_{L^2([0,1])} &\leq \|y_u^2 - y_{\bar{u}_h}^2\|_{L^2([0,1])} \\
&\leq \|y_u - y_{\bar{u}_h}\|_{L^2([0,1])} \|y_u + y_{\bar{u}_h}\|_{L^2([0,1])} \\
&\leq \|y_u - y_{\bar{u}_h}\|_{L^\infty([0,1])} \|y_u + y_{\bar{u}_h}\|_{L^\infty([0,1])} \\
&\leq L_{G_h} \|u - \bar{u}_h\|_{L^2([0,1])} \|y_u + y_{\bar{u}_h}\|_{L^\infty([0,1])}
\end{aligned}$$

with

$$\begin{aligned}
\|y_u - y_{\bar{u}_h}\|_{L^\infty([0,1])} &\leq L_{G_h} \|u - \bar{u}_h\|_{L^2([0,1])} \\
&\forall u \in U : \|u - \bar{u}_h\|_{L^2([0,1])} \leq \|\tilde{u}_h - \bar{u}_h\|_{L^2([0,1])}
\end{aligned} \tag{5.13}$$

Since L_G only depends on the imbedding constant I_2 and the Laplace operator Δ, $L_G = L_{G_h}$ holds.
Now we consider $\|y_u + y_{\bar{u}_h}\|_{L^\infty([0,1])}$. We see that

$$\begin{aligned}
\|y_u + y_{\bar{u}_h}\|_{L^\infty([0,1])} &\leq \|y_u\|_{L^\infty([0,1])} + \|y_{\bar{u}_h}\|_{L^\infty([0,1])} \\
&\leq \frac{1}{\sqrt{2}} (\|y_u\|_{H^1([0,1])} + \|y_{\bar{u}_h}\|_{H^1([0,1])})
\end{aligned}$$

holds. The weak formulation reads as

$$(\nabla y_u, \nabla t) + (y_u, t) + (y_u^3, t) = (u, t).$$

With $t = y_u$ this leads to

$$(\nabla y_u, \nabla y_u) + (y_u, y_u) + (y_u^3, y_u) = (u, y_u).$$

Thus we get

$$\begin{aligned}
\|y_u\|_{H^1([0,1])}^2 &\leq (\|u\|_{L^2([0,1])}) \|y_u\|_{L^2([0,1])} \\
&\leq (\|u\|_{L^2([0,1])}) \|y_u\|_{H^1([0,1])} \\
\Rightarrow \|y_u\|_{H^1([0,1])} &\leq \|u\|_{L^2([0,1])}.
\end{aligned}$$

Since the considered u fulfill $\|u - \bar{u}_h\|_{L^2([0,1])} \leq \|\tilde{u}_h - \bar{u}_h\|_{L^2([0,1])}$ it is obvious that

$$\|u\|_{L^2([0,1])} \leq \|\bar{u}_h\|_{L^2([0,1])} + \|\tilde{u}_h - \bar{u}_h\|_{L^2([0,1])}$$

holds. Thus

$$\|y_u\|_{H^1([0,1])} \leq \|\bar{u}_h\|_{L^2([0,1])} + \|\tilde{u}_h - \bar{u}_h\|_{L^2([0,1])}$$

holds, which leads to

$$\begin{aligned}\|y_u^2 - y_v^2\|_{L^2([0,1])} &\leq L_{G_h}\|u - \bar{u}_h\|_{L^2([0,1])}\frac{2}{\sqrt{2}}(\|\bar{u}_h\|_{L^2([0,1])} + \|\tilde{u}_h - \bar{u}_h\|_{L^2([0,1])}) \\ &= \sqrt{2}L_{G_h}(\|\bar{u}_h\|_{L^2([0,1])} + \|\tilde{u}_h - \bar{u}_h\|_{L^2([0,1])})\|u - \bar{u}_h\|_{L^2([0,1])} \\ &\forall u \in U : \|u - \bar{u}_h\|_{L^2([0,1])} \leq \|\tilde{u}_h - \bar{u}_h\|_{L^2([0,1])}.\end{aligned}$$

Hence we see

$$L_{G_h^2} = \sqrt{2}L_{G_h}(\|\bar{u}_h\|_{L^2([0,1])} + \|\tilde{u}_h - \bar{u}_h\|_{L^2([0,1])}).$$

Final result

$$\begin{aligned}L &= 3c_\infty L_{G_h^2} K_h' \\ &= 3\frac{\sqrt{\pi^2+1}}{\pi^2}(\|\bar{u}_h\|_{L^2([0,1])} + R)\end{aligned}$$

5.1.6 Computation of M

First we prove the following inequaility

$$\|Gu - G\bar{u}_h\|_{L^\infty([0,1])} \leq \frac{3}{\sqrt{2}}\|u - \bar{u}_h\|_{L^2([0,1])}\ ,\ \forall u \in U\ : \|u - \bar{u}_h\|_{L^2([0,1])} \leq R$$

Since G is Fréchet-differentiable we know that

$$Gu = G\bar{u}_h + G'(\bar{u}_h)(u - \bar{u}_h) + r(\bar{u}_h, u - \bar{u}_h)$$

holds, which we reformulate to

$$Gu - G\bar{u}_h = G'(\bar{u}_h)(u - \bar{u}_h) + r(\bar{u}_h, u - \bar{u}_h).$$

This leads to

$$\begin{aligned}
\|Gu - G\bar{u}_h\|_{L^\infty([0,1])} &\leq \|G'(\bar{u}_h)(u-\bar{u}_h)\|_{L^\infty([0,1])} + \|r(\bar{u}_h, u-\bar{u}_h)\|_{L^\infty([0,1])} \\
&= \|G'(\bar{u}_h)(u-\bar{u}_h)\|_{L^\infty([0,1])} + \|Gu - G\bar{u}_h - G'(\bar{u}_h)(u-\bar{u}_h)\|_{L^\infty([0,1])} \\
&\leq \frac{1}{\sqrt{2}}(\|G'(\bar{u}_h)(u-\bar{u}_h)\|_{H^1([0,1])} \\
&+ \|Gu - G\bar{u}_h - G'(\bar{u}_h)(u-\bar{u}_h)\|_{H^1([0,1])}) \\
&\leq \frac{1}{\sqrt{2}}(\underbrace{\|G'(\bar{u}_h)\|_{L^2([0,1])\to H^1([0,1])}}_{\leq 1}\|u-\bar{u}_h\|_{L^2([0,1])} \\
&+ \sup_{\tau\in[0,1]}\|G'(\bar{u}_h+\tau(u-\bar{u}_h)) - G'(\bar{u}_h)\|_{L^2([0,1])\to H^1([0,1])}\|u-\bar{u}_h\|_{L^2([0,1])}) \\
&\leq \frac{1}{\sqrt{2}}(\|u-\bar{u}_h\|_{L^2([0,1])} \\
&+ \sup_{\tau\in[0,1]}\underbrace{\|G'(\bar{u}_h+\tau(u-\bar{u}_h))\|_{L^2([0,1])\to H^1([0,1])}}_{\leq 1}\|u-\bar{u}_h\|_{L^2([0,1])} \\
&+ \underbrace{\|G'(\bar{u}_h)\|_{L^2([0,1])\to H^1([0,1])}}_{\leq 1}\|u-\bar{u}_h\|_{L^2}) \\
&\leq \frac{3}{\sqrt{2}}\|u-\bar{u}_h\|_{L^2}.
\end{aligned}$$

Now we consider

$$\begin{aligned}
|(f''(u) - f''(\bar{u}_h))[v_1, v_2]| \leq M\|u-\bar{u}_h\|_{L^2([0,1])}\|v_1\|_{L^2([0,1])}\|v_2\|_{L^2([0,1])} \\
\forall u \in B_R(\bar{u}_h) \text{ and } v_1, v_2 \in L^2([0,1])
\end{aligned}$$

with

$$f(u) = \frac{1}{2}\|Gu - y_d\|_{L^2([0,1])} + \frac{\lambda}{2}\|u\|_{L^2([0,1])}$$

and compute the derivatives of first and second order:

$$\begin{aligned}
f'(u)v_1 &= (G'(u)v_1, Gu - y_d)_{L^2([0,1])} + \lambda(v_1, u)_{L^2([0,1])} \\
f''(u)[v_1, v_2] &= (G''(u)[v_1, v_2], Gu - y_d)_{L^2([0,1])} \\
&\quad + (G'(u)v_2, G'(u)v_1)_{L^2([0,1])} + \lambda(v_1, v_2)_{L^2([0,1])}
\end{aligned}$$

Thus we get

$$
\begin{aligned}
[f''(u) - f''(\bar{u}_h)][v_1, v_2] &= (G''(u)[v_1, v_2], Gu - y_d)_{L^2([0,1])} + (G'(u)v_2, G'(u)v_1)_{L^2([0,1])} \\
&- (G''(\bar{u}_h)[v_1, v_2], G\bar{u}_h - y_d)_{L^2([0,1])} - (G'(\bar{u}_h)v_2, G'(\bar{u}_h)v_1)_{L^2([0,1])} \\
&= (G''(u)[v_1, v_2], Gu - G\bar{u}_h + G\bar{u}_h - y_d)_{L^2([0,1])} \\
&- (G''(\bar{u}_h)[v_1, v_2], G\bar{u}_h - y_d)_{L^2([0,1])} \\
&+ (G'(u)v_2, G'(u)v_1 - G'(\bar{u}_h)v_1 + G'(\bar{u}_h)v_1)_{L^2([0,1])} \\
&- (G'(\bar{u}_h)v_2, G'(\bar{u}_h)v_1)_{L^2([0,1])} \\
&= (G''(u)[v_1, v_2], Gu - G\bar{u}_h)_{L^2([0,1])} \\
&+ (G''(u)[v_1, v_2] - G''(\bar{u}_h)[v_1, v_2], G\bar{u}_h - y_d)_{L^2([0,1])} \\
&+ (G'(u)v_2, G'(u)v_1 - G'(\bar{u}_h)v_1)_{L^2([0,1])} \\
&+ (G'(u)v_2 - G'(\bar{u}_h)v_2, G'(\bar{u}_h)v_1)_{L^2([0,1])} \\
&\leq \|G''(u)[v_1, v_2\|_{L^2([0,1])}\|Gu - G\bar{u}_h\|_{L^2([0,1])} \\
&+ \|[G''(u) - G''(\bar{u}_h)][v_1, v_2]\|_{L^2([0,1])}\|G\bar{u}_h - y_d\|_{L^2([0,1])} \\
&+ \|G'(u)v_2\|_{L^2([0,1])}\|G'(u)v_1 - G'(\bar{u}_h)v_1\|_{L^2([0,1])} \\
&+ \|G'(u)v_2 - G'(\bar{u}_h)v_2\|_{L^2([0,1])}\|G'(\bar{u}_h)v_1\|_{L^2([0,1])} \\
&\leq K''\|v_1\|_{L^2([0,1])}\|v_2\|_{L^2([0,1])}\|Gu - G\bar{u}_h\|_{L^\infty([0,1])} \\
&+ N\|u - \bar{u}_h\|_{L^2([0,1])}\|v_1\|_{L^2([0,1])}\|v_2\|_{L^2([0,1])}\|G\bar{u}_h - y_d\|_{L^2([0,1])} \\
&+ \|v_2\|_{L^2([0,1])}L\|u - \bar{u}_h\|_{L^2([0,1])}\|v_1\|_{L^2([0,1])} \\
&+ L\|u - \bar{u}_h\|_{L^2([0,1])}\|v_2\|_{L^2([0,1])}\|v_1\|_{L^2([0,1])} \\
&\leq (\frac{3}{\sqrt{2}}K'' + N\|G\bar{u}_h - y_d\|_{L^2([0,1])} + 2L) \\
&\|u - \bar{u}_h\|_{L^2([0,1])}\|v_1\|_{L^2([0,1])}\|v_2\|_{L^2([0,1])}.
\end{aligned}
$$

5.1.7 Computation of c_G

We consider

$$\|G(u_h) - G_h(u_h)\|_{L^\infty([0,1])} \leq c_G h\|u_h\|_{L^2([0,1])} \ , \ \forall u_h \in U_{ad}^h$$

and set

$$
\begin{aligned}
G(u_h) = z \ &: \ -\Delta z + z + z^3 = u_h \\
&\Leftrightarrow (\nabla z, \nabla v) + (z, v) + (z^3, v) = (u_h, v) \ , \ \forall v \in V \\
G_h(u_h) = z_h \ &: \ -\Delta z_h + z_h + z_h^3 = u_h \\
&\Leftrightarrow (\nabla z_h, \nabla v_h) + (z_h, v_h) + (z_h^3, v_h) = (u_h, v_h) \ , \ \forall v_h \in V_h.
\end{aligned}
$$

We get

$$
\begin{aligned}
\|z-z_h\|_{L^\infty([0,1])} &\le \|z - I_h z\|_{L^\infty([0,1])} + \|I_h z - z_h\|_{L^\infty([0,1])} \\
&\le \frac{1}{\sqrt{2}}\|z - I_h z\|_{H^1([0,1])} + \frac{\sqrt{3}}{\sqrt{h}}\|I_h z - z_h\|_{L^2([0,1])} \\
&\le \frac{2}{\sqrt{2}} h\|z\|_{H^2([0,1])} + \frac{\sqrt{3}}{\sqrt{h}}(\|I_h z - z\|_{L^2([0,1])} + \|z - z_h\|_{L^2([0,1])}) \\
&\le \sqrt{2} h c_S \|u_h\|_{L^2([0,1])} + \frac{\sqrt{3}}{\sqrt{h}}(h^2\|z\|_{H^2([0,1])} + c_{L^2} c_S h^2 \|u_h\|_{L^2([0,1])}) \\
&\le \sqrt{2} h \|u_h\|_{L^2([0,1])} + \frac{\sqrt{3}}{\sqrt{h}}(h^2\|u_h\|_{L^2([0,1])} + h^2 c_{L^2([0,1])} \|u_h\|_{L^2([0,1])}) \\
&= (\sqrt{2} + \sqrt{3}\sqrt{h} + \sqrt{3}\sqrt{h} c_{L^2([0,1])}) \|u_h\|_{L^2([0,1])} h \\
&= \left(\sqrt{2} + \sqrt{3}\sqrt{h} + \sqrt{3}\sqrt{h}\frac{(\delta_1 c_2 + c_f c_1)^2}{\delta_0}\sqrt{c_L^2 + (1 + (1 + 3c_c^2)c_L)^2}\right) \\
&\qquad h\|u_h\|_{L^2([0,1])} \\
&= \left(\sqrt{2} + \sqrt{3}\sqrt{h} + \sqrt{3}\sqrt{h}(3 + 3c_c^2)^2 \frac{1+\pi^2}{\pi^2}\sqrt{\frac{I_2^2}{\delta_0^2} + (1 + (1 + 3c_c^2)\frac{I_2}{\delta_0})^2}\right) \\
&\qquad h\|u_h\|_{L^2([0,1])}.
\end{aligned}
$$

5.1.8 Computation of $c_{G'}$

We look at the second approximation property

$$
\|[G'(\bar{u}_h) - G_h'(\bar{u}_h)]u\|_{L^\infty([0,1])} \le c_{G'} h^k \|u\|_{L^2([0,1])}, \ \forall u \in L^2([0,1]).
$$

We set

$$
G'(\bar{u}_h)u = z
$$

and

$$
G_h'(\bar{u}_h)u = z_h
$$

with z solution of

$$
\begin{aligned}
-z_{xx} + z + 3\bar{y}_h^2 z &= u \text{ in } [0,1] \\
z(0) = z(1) &= 0
\end{aligned}
$$

and z_h solution of

$$
\begin{aligned}
-(z_h)_{xx} + z_h + 3\bar{y}_h^2 z_h &= u \text{ in } [0,1] \\
z_h(0) = z_h(1) &= 0.
\end{aligned}
$$

We write this equations as

$$(Az)(x) + f(y(x)) = u(x) \text{ in } [0,1]$$

with

$$Az = -\Delta z = -z_{xx} \text{ and } f(z) = z + 3\bar{y}_h^2 z.$$

Now we use Lemma 5.1 in [41] to derive the following estimate:

$$\begin{aligned}
\|z - z_h\|_{L^\infty([0,1])} &\leq \frac{1}{\sqrt{2}}\|z - z_h\|_{H^1([0,1])} \\
&\leq \frac{1}{\sqrt{2}}\frac{\delta_1 c_2 + c_3 c_1}{\delta_0} h\|G'(\bar{u}_h)\|_{L^2 \to H^2}\|u\|_{L^2([0,1])}
\end{aligned}$$

with

$$\begin{aligned}
\delta_0 &= 1 - I_2^2 = \frac{\pi^2}{1+\pi^2} \\
\delta_1 &= 1 \\
c_1 &= 1 \\
c_2 &= 1 + \sqrt{\frac{7}{3}} \\
c_3 &= 1 + 3c_c^2.
\end{aligned}$$

c_1 and c_2 fulfill the interpolation properties of assumption (A3) in [41],i.e.

$$\begin{aligned}
\|y - I_h y\|_{L^2([0,1])} &\leq c_1 h^2 \|y\|_{H^2([0,1])} \\
\|y - I_h y\|_{H^1([0,1])} &\leq c_2 h \|y\|_{H^2([0,1])}
\end{aligned}$$

which is proven in [8] Chapter 4.5.
c_3 has to fulfill assumption (A2) of [41], i.e. for a function $f = f(y) \ : \mathbb{R} \to \mathbb{R}$ of class C^2 with $f(0) = 0$, there exists a constant c_3 such that

$$|f(y_1) - f(y_2)| \leq c_3 |y_1 - y_2|$$

holds for all $y_1, y_2 \in \mathbb{R}$. We want to use this result for the first derivative of G, which means we have to show that this assumption holds for $f(y) = y + 3y_h^2 y$, $y \in \mathbb{R}$. We get for $y_1, y_2 \in \mathbb{R}$:

$$\begin{aligned}
|f(y_1) - f(y_2)| &= |y_1 + 3y_h^2 y_1 - y_2 - 3y_h^2 y_2| \\
&= |(1 + 3y_h^2)y_1 - (1 + 3y_h^2)y_2| = (1 + 3y_h^2)|y_1 - y_2| \\
&\leq (1 + 3c_c^2)|y_1 - y_2|
\end{aligned}$$

We utilized that y_h is a solution of the optimal control problem and thus fullfills the pointwise state constraint $|y(x)| \leq c_c$.

At last we have to compute the operator norm $\|G'(\bar{u}_h)\|_{L^2\to H^2}$:
We set

$$G'(\bar{u}_h)u = z$$

with

$$\begin{aligned} -z_{xx} + z + 3\bar{y}_h^2 z &= u \text{ in } [0,1] \\ z(0) = z(1) &= 0. \end{aligned}$$

This is equivalent to

$$z_{xx} = z + 3\bar{y}_h^2 z - u$$

wich leads to

$$\begin{aligned} z_{xx}^2 &= z + 3\bar{y}_h^2 z^2 - uz + 3\bar{y}_h^2 + 9\bar{y}_h^4 z^2 - 3\bar{y}_h^2 zu - uz - 3\bar{y}_h^2 zu + u^2 \\ &= z^2 + 6\bar{y}_h^2 z^2 + 9\bar{y}_h^2 - 6\bar{y}_h^2 zu - 2zu + u^2 \\ &= (1 + 6\bar{y}_h^2 + 9\bar{y}_h^4)z^2 - (6\bar{y}_h^2 + 2)uz + u^2. \end{aligned}$$

Thus we get

$$\begin{aligned} \|G'(\bar{u}_h)u\|_{H^2([0,1])}^2 &= \int_0^1 z^2 + z_x^2 + z_{xx}^2 \text{ dx} \\ &= \int_0^1 z^2 + z_x^2 + 3\bar{y}_h^2 z^2 + z_{xx}^2 - 3\bar{y}_h^2 z^2 \text{ dx} \\ &= \int_0^1 uz + (1 + 6\bar{y}_h^2 + 9\bar{y}_h^4)z^2 - (6\bar{y}_h^2 + 2)uz + u^2 - 3\bar{y}_h^2 z^2 \text{ dx} \\ &= \int_0^1 (1 + 6\bar{y}_h^2 + 9\bar{y}_h^4)z^2 - (6\bar{y}_h^2 + 1)uz + u^2 - 3\bar{y}_h^2 z^2 \text{ dx} \\ &= \int_0^1 (1 + 6\bar{y}_h^2 + 9\bar{y}_h^4)z^2 - (6\bar{y}_h^2 + 1)(z^2 + z_x^2 + 3\bar{y}_h^2 z^2) + u^2 - 3\bar{y}_h^2 z^2 \text{ dx} \\ &= \int_0^1 (1 + 6\bar{y}_h^2 + 9\bar{y}_h^4)z^2 - (6\bar{y}_h^2 + 1)z^2 - (6\bar{y}_h^2 + 1)z_x^2 \\ &\quad - 18\bar{y}_h^4 z^2 - 3\bar{y}_h^2 z^2 + u^2 - 3\bar{y}_h^2 z^2 \text{ dx} \\ &= \int_0^1 u^2 \underbrace{-9\bar{y}_h^2 z^2 - (6\bar{y}_h^2 + 1)z_x^2 - 6\bar{y}_h^2 z^2}_{\leq 0} \text{ dx} \\ &\leq \int_0^1 u^2 \text{ dx} = \|u\|_{L^2([0,1])}^2 \end{aligned}$$

which means

$$\begin{aligned} \|G'(\bar{u}_h)u\|_{H^2([0,1])} &\leq \|u\|_{L^2([0,1])} \\ \Rightarrow \|G'(\bar{u}_h)\|_{L^2([0,1])\to H^2([0,1])} &\leq 1. \end{aligned}$$

Finally we get

$$\begin{aligned}\|z - z_h\|_{L^\infty([0,1])} &\le \frac{1}{\sqrt{2}} \frac{1 + \sqrt{\frac{7}{3}} + 1 + 3c_c^2}{\frac{\pi^2}{1+\pi^2}} h\|u\|_{L^2([0,1])} \\ &= \frac{(1+\pi^2)(2 + \sqrt{\frac{7}{3}} + 3c_c^2)}{\pi^2\sqrt{2}} h\|u\|_{L^2([0,1])}.\end{aligned}$$

5.1.9 Computation of c_f

Because of Theorem 4.2 we know that c_f can be derived as

$$c_f = \|f'\| \frac{8c_G m^2}{\tau}$$

with

$$\|f'\| \ge \max_{t\in[0,1]} \|f'(\bar{u}_h + td_h)\|_{\mathcal{L}(L^2([0,1]),\mathbb{R})}.$$

We estimate $\|f'(u_h)\|_{\mathcal{L}(L^2([0,1]),\mathbb{R})}$ with $u_h = \bar{u}_h + td_h,\ t \in [0,1]$. With

$$\begin{aligned}f'(\bar{u}_h)u &= \int_0^1 G'(\bar{u}_h)u(G\bar{u}_h - y_d) + \lambda u(\bar{u}_h - u_d)\ \mathrm{dx} \\ &\le |\int_0^1 G'(\bar{u}_h)u(G\bar{u}_h - y_d)\ \mathrm{dx}| + \lambda|\int_0^1 u(\bar{u}_h - u_d)\ \mathrm{dx}| \\ &\le \|G'(\bar{u}_h)u\|_{L^2([0,1])}\|G\bar{u}_h - y_d\|_{L^2([0,1])} + \lambda\|u\|_{L^2([0,1])}\|\bar{u}_h - u_d\|_{L^2([0,1])} \\ &\le \|u\|_{L^2([0,1])}(\|G\bar{u}_h - y_d\|_{L^2([0,1])} + \lambda\|\bar{u}_h - u_d\|_{L^2([0,1])})\end{aligned}$$

we see that

$$\|f'(u_h)\|_{\mathcal{L}(L^2([0,1]),\mathbb{R})} \le \|Gu_h - y_d\|_{L^2([0,1])} + \lambda\|u_h - u_d\|_{L^2([0,1])}$$

holds.

5.1.10 Computation of $c_{f'}$

We derive $c_{f'}$ of

$$|[f'(\bar{u}_h) - f_h'(\bar{u}_h)]u| \le c_{f'} h^k \|u\|_{L^2([0,1])}, \quad \forall u \in L^2([0,1])$$

with

$$f(u) = \frac{1}{2}\|Gu - y_d\|^2_{L^2([0,1])} + \frac{\lambda}{2}\|u\|^2_{L^2([0,1])}$$

and

$$f_h(u) = \frac{1}{2}\|G_h u - y_d\|^2_{L^2([0,1])} + \frac{\lambda}{2}\|u\|^2_{L^2([0,1])}.$$

We compute the derivatives

$$f'(\bar{u}_h)u = (G'(\bar{u}_h)u, G\bar{u}_h - y_d)_{L^2([0,1])} + \lambda(u, \bar{u}_h)_{L^2([0,1])}$$

and

$$f'_h(\bar{u}_h)u = (G'_h(\bar{u}_h)u, G_h\bar{u}_h - y_d)_{L^2([0,1])} + \lambda(u, \bar{u}_h)_{L^2([0,1])}.$$

Thus we get :

$$\begin{aligned}
f'(\bar{u}_h)u - f'_h(\bar{u}_h)u &= (G'(\bar{u}_h)u, G\bar{u}_h - y_d)_{L^2([0,1])} - (G'_h(\bar{u}_h)u, G_h\bar{u}_h - y_d)_{L^2([0,1])} \\
&= (G'(\bar{u}_h)u, G\bar{u}_h - y_d)_{L^2([0,1])} - (G'_h(\bar{u}_h)u, G_h\bar{u}_h - y_d)_{L^2([0,1])} \\
&+ (G'(\bar{u}_h)u, G_h\bar{u}_h - y_d)_{L^2([0,1])} - (G'(\bar{u}_h)u, G_h\bar{u}_h - y_d)_{L^2([0,1])} \\
&= (G'(\bar{u}_h)u, G\bar{u}_h - G_h\bar{u}_h)_{L^2([0,1])} + (G'(\bar{u}_h)u - G'_h(\bar{u}_h)u, G_h\bar{u}_h - y_d)_{L^2([0,1])} \\
&\leq \|G'(\bar{u}_h)u\|_{L^2([0,1])}\|G\bar{u}_h - G_h\bar{u}_h\|_{L^2([0,1])} \\
&+ \|G'(\bar{u}_h)u - G'_h(\bar{u}_h)u\|_{L^2([0,1])}\|G_h\bar{u}_h - y_d\|_{L^2([0,1])} \\
&\leq \|G'(\bar{u}_h)u\|_{H^1([0,1])}\|G\bar{u}_h - G_h\bar{u}_h\|_{L^\infty([0,1])} \\
&+ \|G'(\bar{u}_h)u - G'_h(\bar{u}_h)u\|_{L^\infty([0,1])}\|G_h\bar{u}_h - y_d\|_{L^\infty([0,1])} \\
&\leq \|u\|_{L^2([0,1])}c_G h\|\bar{u}_h\|_{L^2([0,1])} + c_{G'}h\|u\|_{L^2([0,1])}\|G_h\bar{u}_h - y_d\|_{L^\infty([0,1])} \\
&= (c_G\|\bar{u}_h\|_{L^2([0,1])} + c_{G'}\|G_h\bar{u}_h - y_d\|_{L^\infty})h\|u\|_{L^2([0,1])} \\
\Rightarrow |[f'(\bar{u}_h) - f'_h(\bar{u}_h)]u| &\leq (c_G\|\bar{u}_h\|_{L^2([0,1])} + c_{G'}\|G_h\bar{u}_h - y_d\|_{L^\infty([0,1])})h\|u\|_{L^2([0,1])}
\end{aligned}$$

5.1.11 Coercivity

We want to derive an α, which fulfills

$$\mathcal{L}''(\bar{u}_h, \mu_h)v^2 = f''(\bar{u}_h)[v,v] + \langle \mu_h, G''(\bar{u}_h)[v,v]\rangle \geq \alpha\|v\|_U^2, \ \forall v \in U.$$

First of all we have to compute the derivatives of $\mathcal{L}$:

$$\begin{aligned}
D_u\mathcal{L}(u,y,p,\mu_a,\mu_b)v &= (\lambda(u-u_d), v) \\
D_y\mathcal{L}(u,y,p,\mu_a,\mu_b)y_v &= (y-y_d, y_v) - (\nabla y_v, \nabla p) - (y_v, p) \\
&\quad - (3y^2 y_v, p) + (\mu_a, y_v) - (\mu_b, y_v)
\end{aligned}$$

$$\begin{aligned}
D_{u^2}\mathcal{L}(u,y,p,\mu_a,\mu_b)[v,v] &= \lambda(v,v) = \int_\Omega \lambda v^2 \text{ dx} \\
D_{y^2}\mathcal{L}(u,y,p,\mu_a,\mu_b)[y_v,y_v] &= (y_v,y_v) - (6yy_vy_v,p) = \int_\Omega y_vy_v - 6yy_vy_vp \text{ dx} \\
&= \int_\Omega (1-6yp)y_v^2 \text{ dx}
\end{aligned}$$

Thus

$$D_{u^2,y^2}\mathcal{L}(\bar{u}_h,\bar{y}_h,\bar{p}_h,\mu_a^h,\mu_b^h)[(v,y_v),(v,y_v)] = \int_\Omega (1-6\bar{y}_h\bar{p}_h)y_v^2dx + \int_\Omega \lambda v^2 dx \geq \alpha\|v\|^2_{L^2(\Omega)} \ \forall v \in L^2(\Omega)$$

with $G'(\bar{u}_h)v = y_v$:

$$-\Delta y_v + y_v + 3\bar{y}_h^2 y_v = \bar{u}_h \text{ in } [0,1]$$
$$y_v(0) = y_v(1) = 0$$

holds with $\alpha = \lambda$, if the pointwise condition

$$1 - 6\bar{y}_h(x)\bar{p}_h(x) \geq 0 \text{ a.e in } [0,1]$$

is fulfilled.

5.2 Summary

To conclude this section we summarize our results. These results enable us to compute the assumed constants once we derived a numerical solution $\bar{u}_h$.
We start the summary with the Lipschitz-type constants $L,\ M$ and N.

5.2.1 Lipschitz-type constants $L,\ M$ and N

The assumption (3.24)

$$\|[G_h'(u) - G_h'(\bar{u}_h)]v\|_{H^1([0,1])} \leq L\|u-\bar{u}_h\|_{L^2([0,1])}\|v\|_{L^2([0,1])},\ \forall v \in L^2([0,1]),$$
$$\forall u \in L^2([0,1]) : \|u-\bar{u}_h\|_{L^2([0,1])} \leq \|\tilde{u}_h - \bar{u}_h\|_{L^2([0,1])}$$

holds for

$$L = 3\frac{\sqrt{\pi^2+1}}{\pi^2}(\|\bar{u}_h\|_{L^2([0,1])} + R).$$

For the constant M of (3.25) with

$$|(f''(u) - f''(\bar{u}_h))[v_1,v_2]| \leq M\|u-\bar{u}_h\|_{L^2([0,1])}\|v_1\|_{L^2([0,1])}\|v_2\|_{L^2([0,1])}$$
$$\forall u \in B_R(\bar{u}_h) \text{ and } v_1, v_2 \in L^2([0,1])$$

we concluded

$$M = 36(\frac{9}{16})^{\frac{1}{3}}\frac{(\pi^2+1)^{\frac{2}{3}}}{\pi^2} + N\|G\bar{u}_h - y_d\|_{L^2[0,1]} + 2L$$

For assumption (3.26)

$$\|[G''(u) - G''(\bar{u}_h)][v_1, v_2]\|_Z \le N\|u - \bar{u}_h\|_U \|v_1\|_U \|v_2\|_U, \\ \forall u \in U_{ad},\ \|u - \bar{u}_h\|_U \le R,\ v_1, v_2 \in U$$

we derived the estimation

$$N = \frac{3}{\sqrt{2}}(24(\frac{9}{16})^{\frac{1}{3}} \frac{(\pi^2+1)(\pi^2+1)^{\frac{1}{6}}}{\pi^4}(\|\bar{u}_h\|_{L^2([0,1])} + R) \\ + 12\|c_c\|_{L^\infty([0,1])} \frac{\sqrt{\pi^2+1}}{\pi^2}(\|\bar{u}_h\|_{L^2([0,1])} + R) + 2\frac{\sqrt{\pi^2+1}}{\pi^2}).$$

5.2.2 Approximation properties

We derived the approximation constants $c_G, c_{G'}$ and $c_{f'}$ as follows:
We proved for assumption (3.20), i.e.

$$\|G(u_h) - G_h(u_h)\|_{L^\infty([0,1])} \le c_G h \|u_h\|_{L^2([0,1])} \ ,\ \forall u_h \in U_{ad}^h$$

that

$$c_G = \sqrt{2} + \sqrt{3}\sqrt{h} + \sqrt{3}\sqrt{h}(3 + 3c_c^2)^2 \frac{1+\pi^2}{\pi^2}\sqrt{\frac{\pi^2+1}{\pi^4} + (1 + (1+3c_c^2)\frac{\sqrt{\pi^2+1}}{\pi^2})^2}$$

fulfills it.
For $c_{G'}$ of assumption (3.21),

$$\|[G'(\bar{u}_h) - G'_h(\bar{u}_h)]u\|_{L^\infty([0,1])} \le c_{G'} h^k \|u\|_{L^2([0,1])},\ \forall u \in L^2([0,1]),$$

we derived

$$c_{G'} = \frac{(1+\pi^2)(2 + \sqrt{\frac{7}{3} + 3c_c^2})}{\pi^2\sqrt{2}}.$$

We optained for $c_{f'}$ of assumption (3.22),

$$|[f'(\bar{u}_h) - f'_h(\bar{u}_h)]u| \le c_{f'} h^k \|u\|_{L^2([0,1])},\quad \forall u \in L^2([0,1]),$$

that

$$c_{f'} = (c_G\|\bar{u}_h\|_{L^2([0,1])} + c_{G'}\|G_h\bar{u}_h - y_d\|_{L^\infty([0,1])})$$

holds.

5.2.3 Coercivity condition

We considered the coercivity of $\mathcal{L}''$ in $(\bar{u}_h, \mu_h)$,

$$\mathcal{L}''(\bar{u}_h, \mu_h)v^2 = f''(\bar{u}_h)[v,v] + \langle \mu_h, G''(\bar{u}_h)[v,v]\rangle \ge \alpha\|v\|_U^2,\ \forall v \in U,$$

and showed that it is coercive with $\alpha = \lambda$, if the following pointwise condition is fulfilled:

$$1 - 6\bar{y}_h(x)\bar{p}_h(x) \ge 0 \text{ a.e. in } [0,1]$$

5.2.4 Error bound c_f for u_δ

We recall Theorem 4.2 :

Theorem 5.5. For sufficient small h, $m = \max\{\|\bar{u}_h\|_U, \|\bar{u}_h + d_h\|_U\}$, $s = \frac{\tau}{2L\|d_h\|_U^2}$ and $\delta = \frac{8c_G h^k mL\|d_h\|_U^2}{\tau^2}$ the following inequality holds :

$$|f(u_\delta) - f(\bar{u}_h)| < c_f h^k \tag{5.14}$$

We showed that

$$c_f = (\|Gu_h - y_d\|_{L^2([0,1])} + \lambda\|u_h - u_d\|_{L^2([0,1])})\frac{8c_G m^2}{\tau}$$

holds for the right choice of $t \in [0,1]$ in $u_h = \bar{u}_h + td_h$.

5.2.5 Auxiliary constant K''

As a last point we want to recall the auxiliary constant K'', which is used in more than one estimate. K'' is an upper bound of the operator norm of G"(u). It fulfills (5.5), i.e.

$$\|G''(u)\|_{L^2([0,1])\to\mathcal{L}(L^2([0,1]),L^\infty([0,1]))} \leq K''.$$

We showed that

$$K'' \approx 6.9660$$

holds.

Now we are able to compute all of the assumed constants and check the sufficient optimality condition, once we derived a numerical solution $\bar{u}_h$. Chapter 6 is dedicated to all these numerical aspects.

Chapter 6

Numerical experiments

In the first part of this chapter we introduce the numerical methods we put to use throughout the computation of a numerical solution for the example. In the second part we will present the results of our computations and the conclusions for the optimality conditions and the predicted error estimates.

6.1 FEM

We present a short look into the 1-dimensional **F**inite **E**lement **M**ethod. We follow [27] Chapter 3 and adapt it for an example.
We look at the following problem :
Find a $y \in C^2(0,1) \cap C^1(0,1] \cap C[0,1]$, such that

$$\begin{aligned} -y''(x) + y(x) &= u(x) \ \forall x \in \Omega = [0,1] \\ y(0) &= 0 \\ y(1) &= 0 \end{aligned}$$

hold for a given function $u \in H^2[0,1]$. This can be converted into the following variational formulation:
For a given $u \in H^2[0,1]$ find $y \in V = \{y \in H^2[0,1] : y(0) = y(1) = 0\}$, such that

$$a(y,v) = \langle F, v \rangle$$

holds for all $v \in V = v \in H_0^2[0,1]$ with

$$\begin{aligned} a(y,v) &= \int_0^1 y'(x)v'(x) + y(x)v(x) \ \mathrm{dx} \\ \langle F, v \rangle &= \int_0^1 u(x)v(x) \ \mathrm{dx}. \end{aligned}$$

We want to derive an approximate solution for this problem. To discretize the continuous problem we divide the intervall $[0, 1]$ into $n + 1$ equal parts and get the points $x_0, ..., x_{n+1}$ with

$$x_j = x_0 + jh \text{ and } h = \frac{1}{n+1}.$$

For every x_j we define an ansatzfunction ϕ_j, $j = 1, .., n$, as follows:

$$\phi_j = \begin{cases} 0 & \text{for } 0 \leq x \leq x_{j-1} \\ \dfrac{x - x_{j-1}}{h} & \text{for } x_{j-1} < x \leq x_j \\ \dfrac{x_{j+1} - x}{h} & \text{for } x_j < x \leq x_{j+1} \\ 0 & \text{for } x_{j+1} < x \leq 1 \end{cases}$$

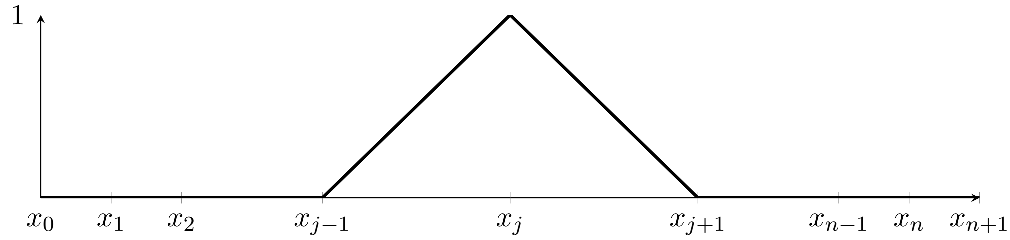


Figure 6.1: ϕ_j

Additionally ϕ_0 and ϕ_{n+1} are defined as

$$\phi_0 = \begin{cases} \dfrac{x_1 - x}{h} & \text{for } 0 < x \leq x_1 \\ 0 & \text{for } x_1 < x \leq 1 \end{cases}$$

and

$$\phi_{n+1} = \begin{cases} 0 & \text{for } 0 \leq x \leq x_n \\ \dfrac{x - x_n}{h} & \text{for } x_n < x \leq 1 \end{cases}$$

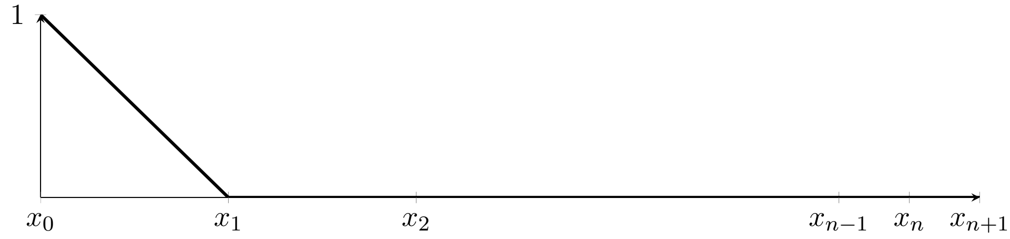


Figure 6.2: ϕ_0

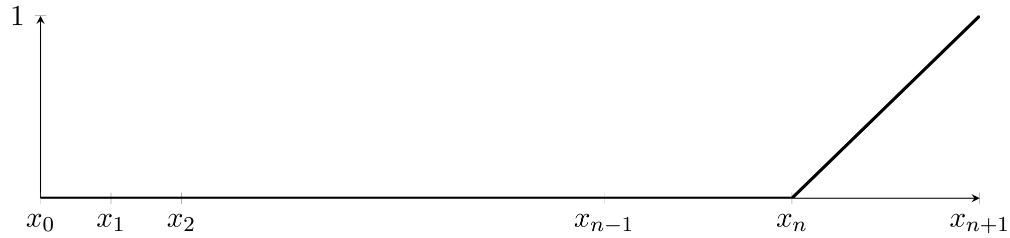


Figure 6.3: ϕ_{n+1}

Using these definitions we can define the general ansatzspace

$$V_h = \{v_h(x) : v_h(x) = \sum_{i=0}^{n+1} v_i \phi_i(x),\ v_h(0) = v_h(1) = 0\}.$$

Now we can formulate the discrete problem :
For a given $u \in H^2[0,1]$ find $y_h \in V_h$, such that

$$\begin{aligned} a(y_h, v_h) &= \langle F, v_h \rangle \\ \text{with } a(y_h, v_h) &= \int_0^1 y_h'(x) v_h'(x) + y_h(x) v_h(x) \text{ dx} \\ \text{and } \langle F, v_h \rangle &= \int_0^1 u(x) v_h(x) \end{aligned}$$

hold for all $v_h \in V_h$.
We express this problem via matrices :

$$K_h \underline{y}_h = \underline{u}_h$$

with

$$\begin{aligned} K_h &= [K_{ij}]_{i,j=1}^n \\ &= [\int_0^1 \phi_i' \phi_j' + \phi_i \phi_j \text{ dx}]_{i,j=1}^n \end{aligned}$$

and

$$\underline{u}_h = [u_i]_{i=1}^n = [\int_0^1 u(x)\phi_i(x)\ \mathrm{dx}]_{i=1}^n$$

Note that the first and last entry of the coefficient vector for every $v_h \in V_h$ is equal to zero due to the dirichlet boundary condition. Thus we only have to consider the entries belonging to the inner knots $x_1, ..., x_n$.

We write K_h as $K_h = K_{h,1} + K_{h,2}$ and see in [27] Section 3.3 that

$$K_{h,1} = \frac{1}{h}\begin{pmatrix} 2 & -1 & 0 & . & . & . & 0 \\ -1 & 2 & -1 & 0 & . & . & 0 \\ 0 & -1 & 2 & -1 & 0 & . & 0 \\ . & . & . & . & . & . & . \\ 0 & . & . & 0 & -1 & 2 & -1 \\ 0 & . & . & . & 0 & -1 & 2 \end{pmatrix}$$

and

$$K_{h,2} = \frac{h}{6}\begin{pmatrix} 4 & 1 & 0 & . & . & . & 0 \\ 1 & 4 & 1 & 0 & . & . & 0 \\ 0 & 1 & 4 & 1 & 0 & . & 0 \\ . & . & . & . & . & . & . \\ 0 & . & . & 0 & 1 & 4 & 1 \\ 0 & . & . & . & 0 & 1 & 4 \end{pmatrix}$$

hold. The matrices $K_{h,1}$ and $K_{h,2}$ are called stiffness matrix and mass matrix.

$\underline{u}_h = (u_1, ..., u_n)^T$ can be derived as

$$u_i = \int_{x_{i-1}}^{x_{i+1}} u(x)\phi_i(x)\ \mathrm{dx}.$$

Thus we converted the underlying continuous problem into a system of equations which can be solved numerically.

6.2 SQP method

To compute numerical results we use one **S**equentially **Q**uadratic **P**rogramming (SQP) method, the Lagrange-Newton SQP.

The general theory is taken from [45] Chapter 4.11, with minor modifications to address the dirichlet boundary conditions of our example.

We consider a problem with distributed control:

$$\begin{aligned} \min J(y,u) := \int_\Omega \phi(x, y(x))\ \mathrm{dx} + \int_\Omega \psi(x, u(x))\ \mathrm{dx} \\ \text{s.t.} -\Delta y + d(x,y) = u \text{ in } \Omega \\ y = 0 \text{ on } \Gamma \end{aligned} \tag{P}$$

An adjoint state is defined via

$$\begin{aligned} -\Delta p + d_y(x,y)p &= \phi_y(x,y) \text{ in } \Omega \\ p &= 0 \text{ on } \Gamma. \end{aligned}$$

The main principles of the SQP algorithm can be described in three steps.
The first step is to linearize the nonlinear problem (P) at a feasible point (y_k, u_k) and a corresponding adjoint state p_k. This leeds to a quadratic problem (QP_k):

$$\min \left\{ J'(y_k,u_k)(y-y_k,u-u_k) + \frac{1}{2}\mathcal{L}''(y_k,u_k,p_k)(y-y_k,u-u_k)^2 \right\} \qquad (QP_k)$$

subject to

$$\begin{aligned} -\Delta y + d(x,y_k) + d_y(x,y_k)(y-y_k) &= u \text{ in } \Omega \\ y &= 0 \text{ on } \Gamma \end{aligned}$$

In the second step we solve the quadratic problem (QP_k) and derive the control u_{k+1} and state y_{k+1}. Then p_{k+1} can be computed via

$$\begin{aligned} -\Delta p + pd_y(x,y_k) + p_k d_{yy}(x,y_k)(y_{k+1}-y_k) &= \phi_y(x,y_k) + \phi_{yy}(x,y_k)(y_{k+1}-y_k) \text{ in } \Omega \\ p &= 0 \text{ on } \Gamma \end{aligned}$$

The last step is to linearize (P) at $(y_{k+1}, u_{k+1}, p_{k+1})$ and start anew.
The algorithm is to be terminated, if the solution of (QP_k) is equal to the solution of the previous iteration.

6.3 Numerical implementation

As mentioned in Chapter 5 we want to illustrate our results on the following example.

Example 6.1.

$$\begin{aligned} \min J(y,u) &= \frac{1}{2}\|y-y_d\|^2_{L^2([0,1])} + \frac{\lambda}{2}\|u-u_d\|^2_{L^2([0,1])} \\ -\Delta y + y + y^3 &= u \text{ in } [0,1] \\ y(0) &= y(1) = 0 \\ |y(x)| &\le c_c \text{ a.e. in } [0,1] \end{aligned}$$

with $c_c \in \mathbb{R}_+$.

To compute a numerical solution we want to employ the Lagrange-Newton-SQP method and derive the following quadratic problem for a feasible point (u_k, y_k, p_k):

$$\min\{\int_0^1 (y_k - y_d)(y - y_k) + \lambda(u_k - u_d)(u - u_k)\ \mathrm{dx}$$
$$-\frac{1}{2}\int_0^1 p_k 6y_k^2(y - y_d)^2\ \mathrm{dx} + \frac{1}{2}\int_0^1 (y - y_k)^2 + \lambda(u - u_k)^2\ \mathrm{dx}\}$$

subject to

$$-\Delta y + y_k + y_k^3 + y - y_k + 3y_k^2(y - y_k) = u \text{ in } \Omega$$
$$y = 0 \text{ on } \Gamma,$$

which can be reformulated to

$$\min\{\frac{1}{2}\int_0^1 y^2(1 - 6p_k y_k^2) + \lambda u^2\ \mathrm{dx} + \int_0^1 y(6y_k^3 p_k - y_d) - \lambda u u_d\ \mathrm{dx}$$
$$+\int_0^1 y_k y_d - \frac{1}{2}y_k^2 - 3p_k y_k^4 + \lambda u_k u_d - \frac{\lambda}{2}u_k^2\ \mathrm{dx}\}$$

subject to

$$-\Delta y + y + 3y_k^2 y - 2y_k^3 = u \text{ in } \Omega$$
$$y = 0 \text{ on } \Gamma.$$

We skip the constant term and get the equivalent problem

$$\min\{\frac{1}{2}\int_0^1 y^2(1 - 6p_k y_k^2) + \lambda u^2\ \mathrm{dx} + \int_0^1 y(6y_k^3 p_k - y_d) - \lambda u u_d\ \mathrm{dx}\}$$

subject to

$$-\Delta y + y + 3y_k^2 y - 2y_k^3 = u \text{ in } \Omega$$
$$y = 0 \text{ on } \Gamma.$$

For optimal y_{k+1} and u_{k+1} we can derive the corresponding adjoint state p_{k+1} via :

$$-\Delta p + p + 3y_k^2 p = 2y_k + 6y_k^2 p_k - y_d - 6y_k p_k y \text{ in } \Omega$$
$$p = 0 \text{ on } \Gamma$$

We translate this quadratic problem into its FE formulation:
Let y, u, y_k and p_k be functions in V_h, i.e. $y = \sum_{i=1}^n y^i \phi_i$, $u = \sum_{i=1}^n u^i \phi_i$, $y_k = \sum_{i=1}^n y_k^i \phi_i$ and $p_k = \sum_{i=1}^n p_k^i \phi_i$, with corresponding coefficient vectors $\underline{y} = (y^1, ..., y^n)^T$, $\underline{u} = (u^1, ..., u^n)^T$, $\underline{y}_k = (y_k^1, ..., y_k^n)^T$ and $\underline{p}_k = (p_k^1, ..., p_k^n)^T$. Then we can formulate the FE problem setting:

Problem 6.2.

$$\min_{(\underline{y},\underline{u})\in\mathbb{R}^{2n}} J(\underline{y},\underline{u}) = \frac{1}{2}(\underline{y},\underline{u})^T H(\underline{y},\underline{u}) + f^T(\underline{y},\underline{u})$$

subject to

$$EQ(\underline{y},\underline{u}) = b$$
$$lb \le (\underline{y},\underline{u}) \le ub$$

with

$$H = \begin{pmatrix} K_{h,2} - 6A_{y_k}^{p_k} & 0 \\ 0 & \lambda K_{h,2} \end{pmatrix}$$

where $A_{y_k}^{p_k}$ is defined as

$$A_{y_k}^{p_k} = \frac{1}{5(n+1)} \begin{pmatrix} a & b & 0 & . & . & . & 0 \\ *_1 & *_2 & *_3 & 0 & . & . & 0 \\ 0 & *_1 & *_2 & *_3 & 0 & . & 0 \\ . & . & . & . & . & . & . \\ 0 & . & . & 0 & *_1 & *_2 & *_3 \\ 0 & . & . & . & 0 & f & g \end{pmatrix}$$

with

$$\begin{aligned} a &= 2y_k^1 p_k^1 + \frac{1}{4}y_k^1 p_k^2 + \frac{1}{4}y_k^2 p_k^1 + \frac{1}{6}y_k^2 p_k^2 \\ b &= \frac{1}{4}y_k^1 p_k^1 + \frac{1}{6}y_k^1 p_k^2 + \frac{1}{6}y_k^2 p_k^1 + \frac{1}{4}y_k^2 p_k^2 \\ f &= \frac{1}{4}y_k^{n-1} p_k^{n-1} + \frac{1}{6}y_k^{n-1} p_k^n + \frac{1}{6}y_k^n p_k^{n-1} + \frac{1}{4}y_k^n p_k^n \\ g &= \frac{1}{6}y_k^{n-1} p_k^{n-1} + \frac{1}{4}y_k^{n-1} p_k^n + \frac{1}{4}y_k^n p_k^{n-1} + 2y_k^n p_k^n \end{aligned}$$

as well as

$$\begin{aligned} *_1 &= \frac{1}{4}y_k^{l-1} p_k^{l-1} + \frac{1}{6}y_k^{l-1} p_k^l + \frac{1}{6}y_k^l p_k^{l-1} + \frac{1}{4}y_k^l p_k^l \\ *_2 &= \frac{1}{4}y_k^{l-1} p_k^{l-1} + \frac{1}{4}y_k^{l-1} p_k^l + \frac{1}{4}y_k^l p_k^{l-1} + 2y_k^l p_k^l + \frac{1}{4}y_k^l p_k^{l+1} \\ *_3 &= \frac{1}{4}y_k^l p_k^l + \frac{1}{6}y_k^l p_k^{l+1} + \frac{1}{6}y_k^{l+1} p_k^l + \frac{1}{4}y_k^{l+1} p_k^{l+1} \end{aligned}$$

for $l \in \{2...n-1\}$. For the computation of $A_{y_k}^{p_k}$ we refer to Section 7.1. f is determined via

$$f = \begin{pmatrix} K_{h,2}(\underline{v} - \underline{y}_d) \\ -\lambda K_{h,2}\underline{u}_d \end{pmatrix}$$

with $\underline{v} = (v_1, ..., v_n)^T$ and $v_i = 6(y_k^i)^2 p_k^i$ for $i = 1, .., n$.
The equation matrix is defined as

$$EQ = \begin{pmatrix} NB & -K \end{pmatrix}$$

with $NB = K + B$. While

$$B = \eta \begin{pmatrix} 6 * (y_k^1)^2 & (y_k^1)^2 + (y_k^2)^2 & 0 & . & . & . & 0 \\ *_1 & *_2 & *_3 & 0 & . & . & 0 \\ 0 & *_1 & *_2 & *_3 & 0 & . & 0 \\ . & . & . & . & . & . & . \\ 0 & . & . & 0 & *_1 & *_2 & *_3 \\ 0 & . & . & . & 0 & (y_k^{n-1})^2 + (y_k^n)^2 & (y_k^{n-1})^2 + 6(y_k^n)^2 \end{pmatrix}$$

with

$$\eta = \frac{1}{4(n+1)}$$
$$*_1 = (y_k^{l-1})^2 + (y_k^l)^2$$
$$*_2 = (y_k^{l-1})^2 + 6(y_k^l)^2 + (y_k^{l+1})^2$$
$$*_3 = (y_k^l)^2 + (y_k^{l+1})^2$$

for $l \in \{2...n-1\}$. For the computation of B we refer to Section 7.2.
The lower and upper bound lb and ub can be written as :

$$lb = -(\underbrace{c_c, ...c_c}_{n}, \underbrace{c, .., c}_{n})^T$$
$$ub = (\underbrace{c_c, ...c_c}_{n}, \underbrace{c, .., c}_{n})^T$$

with a sufficiently large $c \in \mathbb{N}$, such that these partial boundaries are never active.
To start the computation we have to derive a feasible starting point and to choose y_d and u_d.
For the desired state y_d we choose

$$y_d = c_c \sin(\pi x)$$

and set u_d as the corresponding control

$$u_d = c_c(\pi^2 + 1)\sin(\pi x) + c_c^3 \sin(\pi x)^3.$$

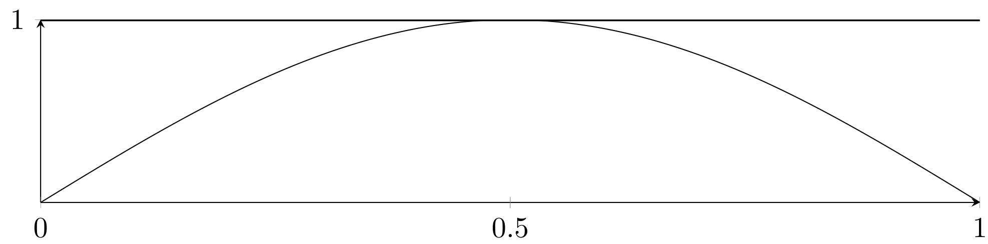


Figure 6.4: y_d

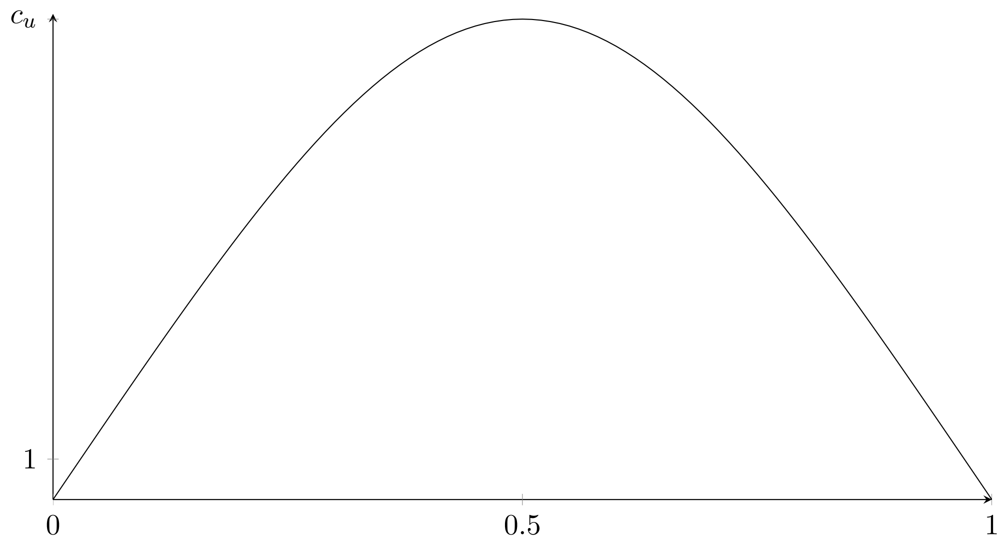


Figure 6.5: u_d with $c_u = \pi^2 + 2$

We illustrate these choices in Figure 6.4 and Figure 6.5 for $c_c = 1$.

Finally we have to determine a starting point for the SQP method.
It is clear that $u_k(x) = y_k(x) = 0,$ a.e. in $[0, 1]$ are feasible, which leaves us with the computation of their adjoint state p_k.
p_k has to fulfill the PDE :

$$-p'' + p = -c_c \sin(\pi x) \tag{6.1}$$

With $\tilde{p}_k = a \sin(\pi x)$ we get

$$\begin{aligned}\tilde{p}_k &= a \sin(\pi x)\\ \tilde{p}_k' &= \pi a \sin(\pi x)\\ \tilde{p}_k'' &= -\pi^2 a \sin(\pi x).\end{aligned}$$

We insert $\tilde{p}_k$ into (6.1) and get

$$(\pi^2 + 1)a\sin(\pi x) = -c_c \sin(\pi x) \Rightarrow a = \frac{-c_c}{\pi^2 + 1}$$

which means

$$p_k(x) = -\frac{c_c}{\pi^2 + 1}\sin(\pi x).$$

6.4 Numerical results

We used MATLAB to implement the Finite Element discretization and the Lagrange-Newton-SQP and solved the quadratic problems with the integrated solver QUADPROG. For the actual computation we choose $c_c = 0.01$ which is rather low but enables us to observe the fulfillment of the sufficient condition at a relatively coarse refinement.
We derive the polynom for several choices of n starting with $n = 63$ to illustrate the progress throughout the refinement of the grid. From step to step we choose n such that $h = \frac{1}{n+1}$ is cut in half. Table 6.6 shows the observed errors regarding the functional value, the control and the state as well as the first nonnegative root of the sufficient condition polynom. The state and the control error are cut in half with each step, which means we see convergence of order h. For r_1 we observe a slower decrease but it still serves as an upper bound for the control error. The slower rate of convergence is likely caused by overestimation of some expression throughout the estimation of the involved constants.

Note that Table 6.6 only shows the error between the numerical optimal control $\bar{u}_h$ and the projection $I_h\bar{u}$ of the continuous optimal control $\bar{u}$. The same holds for the states $\bar{y}_h$ and $\bar{y}$. In order to derive the actual errors $\|\bar{u}_h - \bar{u}\|_{L^2([0,1])}$ and $\|y_h - \bar{y}\|_{L^2([0,1])}$ we proceed as follows :
For the control error we see that

$$\|\bar{u}_h - \bar{u}\|_{L^2([0,1])} \leq \|\bar{u}_h - I_h\bar{u}\|_{L^2([0,1])} + \|I_h\bar{u} - \bar{u}\|_{L^2([0,1])}$$

n	$\lvert f_h(\bar{u}_h, \bar{y}_h) - f(\bar{u}, \bar{y})\rvert$	$\lVert\bar{y}_h - I_h\bar{y}\rVert_{L^2}$	$\lVert\bar{u}_h - I_h\bar{u}\rVert_{L^2}$	r_1
63	$3.9904e-7$	$1.8848e-6$	$8.9335e-4$	–
127	$1.0079e-7$	$9.3983e-7$	$4.4897e-4$	0.064571
255	$2.5250e-8$	$4.7021e-7$	$2.2472e-4$	0.036586
511	$6.3267e-9$	$2.3497e-7$	$1.1249e-4$	0.022508
1023	$1.5834e-9$	$1.1745e-7$	$5.6275e-5$	0.014403
2047	$3.9608e-10$	$5.8716e-8$	$2.8147e-5$	0.009470

Table 6.6: Numerical errors and the control error estimate r_1

holds as well as

$$\|\bar{y}_h - \bar{y}\|_{L^2([0,1])} \leq \|\bar{y}_h - I_h\bar{y}\|_{L^2([0,1])} + \|I_h\bar{y} - \bar{y}\|_{L^2([0,1])}$$

holds for the state error. Due to the regularity of $\bar{u}$ and therefore $\bar{y}$ we can use the interpolation estimates and their constants (see [8]). We get

$$\|I_h\bar{u} - \bar{u}\|_{L^2([0,1])} \leq h^2\|\bar{u}''\|_{L^2([0,1])}$$

and

$$\|I_h\bar{y} - \bar{y}\|_{L^2([0,1])} \leq h^2\|\bar{y}''\|_{L^2([0,1])}.$$

with

$$\begin{aligned}\|\bar{u}''\|_{L^2([0,1])} &= \sqrt{\frac{c_c^2(\pi^4+\pi^2)^2}{2} + \frac{3c_c^4\pi^2(\pi^4+\pi^2)}{4} + \frac{45c_c^6\pi^4}{16}} \\ &\approx 0.5755\end{aligned}$$

and

$$\|y_d''\|_{L^2([0,1])} = \frac{c_c\pi^2}{\sqrt{2}} \approx 0.0698$$

(see Section 7.3). This leads to

$$\|\bar{u}_h - \bar{u}\|_{L^2([0,1])} \leq \|\bar{u}_h - I_h\bar{u}\|_{L^2([0,1])} + 0.7586h^2$$

and to

$$\|\bar{y}_h - \bar{y}\|_{L^2([0,1])} \leq \|\bar{y}_h - I_h\bar{y}\|_{L^2([0,1])} + 0.0698h^2$$

which means that both errors approximately behave in the same way as the computed numerical errors.
We present the computation of the necessary constants for $n = 2047$ inner knots, $R = 0.1,\ k = 1,\ \tau = 1,$ and $m = 0.18$. Before we start with the actual computation we state the necessary informations which gained through the numerics :

$$\begin{aligned}\|\bar{u}_h\|_{L^2([0,1])} &= 7.6800e-2 \\ \|\bar{y}_h\|_{L^2([0,1])} &= 7.1000e-3 \\ \|I_h y_d\|_{L^2([0,1])} &= 7.1000e-3 \\ \|\bar{y}_h^2\|_{L^\infty([0,1])} &= 1.0000e-4 \\ \|\mu_h\|_{L^\infty([0,1])} &= 2.1440e-4 \\ \|G_h\bar{u}_h - I_h y_d\|_{L^2([0,1])} &= 5.8716e-8 \\ \|G_h\bar{u}_h - I_h y_d\|_{L^\infty([0,1])} &= 8.7558e-8\end{aligned}$$

With these informations we are able to compute the assumed constants with exception of M and $c_{f'}$ were we used the errors $\|G\bar{u}_h - y_d\|_{L^2([0,1])}$ and $\|G_h\bar{u}_h - y_d\|_{L^\infty([0,1])}$

during the estimation process, which are not direct results of the numerical computations. For these two errors we use the estimates

$$\|G\bar{u}_h - y_d\|_{L^2([0,1])} \leq c_G h\|\bar{u}_h\|_{L^2([0,1])} + \|G_h\bar{u}_h - I_h y_d\|_{L^2([0,1])} + \frac{c_c\pi^2}{\sqrt{2}}h^2$$

and

$$\|G\bar{u}_h - y_d\|_{L^\infty([0,1])} \leq \|G_h\bar{u}_h - I_h y_d\|_{L^\infty([0,1])} + \frac{c_c\pi^2}{2}(h^2 + h).$$

We show the derivation of these estimates in Section 7.4. Now we can compute all constants and the computation yields :

$$\begin{aligned} \alpha &= 1, & c_G &= 1.9743..., \\ c_{G'} &= 2.7472..., & c_{f'} &= 0.1516..., \\ c_f &= 0.0859..., & L &= 0.1771..., \\ M &= 15.1290..., & N &= 2.6662... \end{aligned}$$

Note that the computation of c_f involves the estimation of $\|f'\|$. We showed that

$$\|f'\| \leq \|G\bar{u}_h - y_d\|_{L^2([0,1])} + \lambda\|\bar{u}_h - u_d\|_{L^2([0,1])}$$

holds (see 5.1.9), which can be estimated as

$$\|f'\| \leq 2\|y_d\|_{L^2([0,1])} + 2\lambda\|u_d\|_{L^2([0,1])} = 0.1684...$$

due to the choice of $R = 0.1$.
Now we can compute the corresponding sufficient condition polynom

$$\begin{aligned} P(r) = &\frac{\alpha}{2}r^2 - h^k r(c_{f'} + c_{G'}\|\mu_h\|_{(L^\infty([0,1]))^*}) \\ &- \frac{1}{6}r^3(M + N\|\mu_h\|_{(L^\infty([0,1]))^*}) - c_G h^k\|\mu_h\|_{(L^\infty([0,1]))^*}\|\bar{u}_h\|_{L^2([0,1])} - c_f h^k \\ \approx &\frac{1}{2}r^2 - 0.000075r - 2.5220r^3 - 0.00005 \end{aligned}$$

which is illustrated in Figure 6.7.

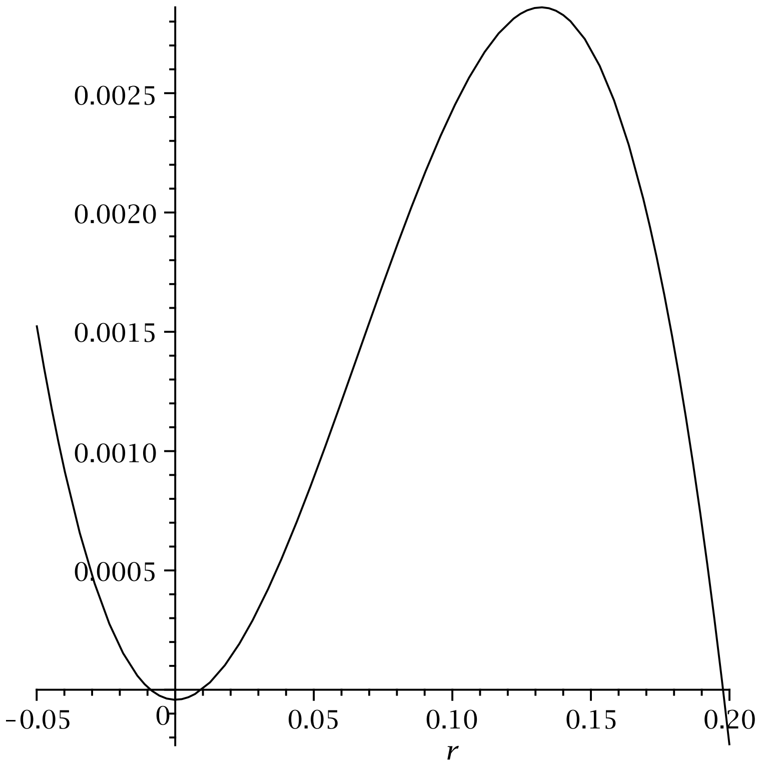


Figure 6.7: Polynom of the SSC for $n = 2047$

The second and third root of the polynom are :

$$r_1 = 0.0094...$$
$$r_2 = 0.1976...$$

We already showed that

$$\|\bar{u}_h - \bar{u}\|_{L^2([0,1])} \leq \|\bar{u}_h - I_h\bar{u}\|_{L^2([0,1])} + 0.7586h^2$$

holds. Thus we see for $h = \frac{1}{2048}$ that

$$\begin{aligned}\|\bar{u}_h - \bar{u}\|_{L^2([0,1])} &\leq \|\bar{u}_h - I_h\bar{u}\|_{L^2([0,1])} + 0.7586h^2 \\ &= 2.8147e-5 + \frac{0.7586}{2048^2} \\ &\approx 2.8327e-5 < 9.466e-3 = r_1\end{aligned}$$

holds, which means that $\|\bar{u} - \bar{u}_h\|_{L^2([0,1])} < r$ holds for all positive r with $P(r) > 0$. Thus we see that the second order conditions works and that the error estimate also delivers an upper bound of the L^2-control error.

Chapter 7

Further computations

To implement the Lagrange-Newton-SQP method into a finite element framework we have to discretize the linearized PDE in an adequate form.

7.1 Conversion of $\int y_k p_k y^2$ dx (Computation of $A_{y_k}^{p_k}$)

We assume that y_k,p_k and y have the corresponding finite element representations $\sum_{i=1}^n y_k^i \phi_i$, $\sum_{i=1}^n p_k^i \phi_i$ and $\sum_{i=1}^n y^i \phi_i$. We want to derive a matrix $A_{y_k}^{p_k}$ such that $\int_0^1 y_k p_k y^2 \text{ dx} = y^T A_{y_k}^{p_k} y$ holds.

$$
\begin{aligned}
&\int_0^1 y_k p_k y^2 \text{ dx} = \\
&\quad \int_0^1 \sum_{l=1}^n y^l \phi_l \sum_{m=1}^n y^m \phi_m \sum_{i=1}^n y_k^i \phi_i \sum_{j=1}^n p_k^j \phi_j \text{ dx} \\
&\quad = \int_0^1 \sum_{l=1}^n y^l \phi_l \sum_{m=1}^n y^m \phi_m \sum_{i=1}^n y_k^i \phi_i (p_k^{i-1} \phi_{i-1} + p_k^i \phi_i + p_k^{i+1} \phi_{i+1}) \text{ dx} \\
&\quad = \int_0^1 \sum_{l=1}^n y^l \phi_l \sum_{m=1}^n y^m \phi_m \sum_{i=1}^n (y_k^i p_k^{i-1} \phi_i \phi_{i-1} + y_k^i p_k^i \phi_i^2 + y_k^i p_k^{i+1} \phi_i \phi_{i+1}) \text{ dx} \\
&\quad = \int_0^1 \sum_{l=1}^n y^l \phi_l \sum_{m=1}^n y^m (y_k^{m-1} p_k^{m-1} \phi_{m-1}^2 \phi_m + y_k^{m-1} p_k^m \phi_{m-1} \phi_m^2 \\
&\quad + y_k^m p_k^{m-1} \phi_m^2 \phi_{m-1} + y_k^m p_k^m \phi_m^3 + y_k^m p_k^{m+1} \phi_m^2 \phi_{m+1} \\
&\quad + y_k^{m+11} p_k^m \phi_{m+1} \phi_m^2 + y_k^{m+1} p_k^{m+1} \phi_{m+1}^2 \phi_m) \text{ dx}
\end{aligned}
$$

Using the properties of ϕ_i we get the expanded expression

$$
\begin{aligned}
&\int_0^1 y_k p_k y^2 \text{ dx} \\
&= \int_0^1 \sum_{l=2}^{n-1} \{ y^l y^{l-1} (y_k^{l-1} p_k^{l-1} \phi_{l-1}\phi_l + y_k^{l-1} p_k^l \phi_{l-1}^2 \phi_l^2 + y_k^l p_k^{l-1} \phi_l^2 \phi_{l-1}^2 + y_k^l p_k^l \phi_l^3 \phi_{l-1}) \\
&+ y^l y^l (y_k^{l-1} p_k^{l-1} \phi_{l-1}^2 \phi_l^2 + y_k^{l-1} p_k^l \phi_{l-1} \phi_l^3 + y_k^l p_k^{l-1} \phi_{l-1} \phi_l^3 + y_k^l p_k^l \phi_l^4 + y_k^l p_k^{l+1} \phi_l^3 \phi_{l+1} \\
&+ y_k^{l+1} p_k^l \phi_{l+1} \phi_l^3 + y k^{l+1} p_k l + 1 \phi_{l+1}^2 \phi_l^2) \\
&+ y^l y^{l+1} (y_k^l p_k^l \phi_l^3 \phi_{l+1} + y_k^l p_k^{l+1} \phi_l^2 \phi_{l+1}^2 + y_k^{l+1} p_k^l \phi_{l+1}^2 \phi_l^2 + y_k^{l+1} p_k^{l+1} \phi_{l+1}^3 \phi_l)\} \\
&+ y^1 y^1 (y_k^1 p_k^1 \phi_1^4 + y_k^1 p_k^2 \phi_1^3 \phi_2 + y_k^2 p_k^1 \phi_1^3 \phi_2 + y_k^2 p_k^2 \phi_2^2 \phi_1^2) \\
&+ y^1 y^2 (y_k^1 p_k^1 \phi_1^3 \phi_2 + y_k^1 p_k^2 \phi_1^2 \phi_2^2 + y_k^2 p_k^1 \phi_2^2 \phi_1^2 + y_k^2 p_k^2 \phi_2^3 \phi_1) \\
&+ y^n y^{n-1} (y_k^{n-1} p_k^{n-1} \phi_{n-1}^3 \phi_n + y_k^{n-1} p_k^n \phi_{n-1}^2 \phi_n^2 + y_k^n p_k^{n-1} \phi_n^2 \phi_{n-1}^2 + y_k^n p_k^n \phi_n^3 \phi_{n-1}) \\
&+ y^n y^n (y_k^{n-1} p_k^{n-1} \phi_{n-1}^2 \phi_n^2 + y_k^{n-1} p_k^n \phi_{n-1} \phi_n^3 + y_k^n p_k^{n-1} \phi_{n-1} \phi_n^3 + y_k^n p_k^n \phi_n^4) \text{ dx.}
\end{aligned}
$$

We need to compute the integrals over the occuring products and potencies of the ansatzfunctions. For suitable $i \in \{1...n\}$ we get:

$$
\begin{aligned}
\int_0^1 \phi_i^2 \phi_{i+1}^2 \text{ dx} = \int_0^1 \phi_{i-1}^2 \phi_i^2 \text{ dx} &= \frac{1}{30(n+1)} \\
\int_0^1 \phi_i^3 \phi_{i+1} \text{ dx} = \int_0^1 \phi_{i-1} \phi_i^3 \text{ dx} &= \frac{1}{20(n+1)} \\
\int_0^1 \phi_i^4 \text{ dx} &= \frac{2}{5(n+1)}
\end{aligned}
$$

This leads us to

$$\begin{aligned}
\int_0^1 y_k p_k y^2 \text{ dx} = \frac{1}{(n+1)} \sum_{l=2}^{n-1} \{ & y^l y^{l-1} (\frac{1}{20} y_k^{l-1} p_k^{l-1} + \frac{1}{30} y_k^{l-1} p_k^l + \frac{1}{30} y_k^l p_k^{l-1} + \frac{1}{20} y_k^l p_k^l) \\
& + y^l y^l (\frac{1}{20} y_k^{l-1} p_k^{l-1} + \frac{1}{20} y_k^{l-1} p_k^l + \frac{1}{20} y_k^l p_k^{l-1} + \frac{2}{5} y_k^l p_k^l + \frac{1}{20} y_k^l p_k^{l+1}) \\
& + \frac{1}{20} y_k^{l+1} p_k^l + \frac{1}{30} yk^{l+1} p_k^{l+1}) \\
& + y^l y^{l+1} (\frac{1}{20} y_k^l p_k^l + \frac{1}{30} y_k^l p_k^{l+1} + \frac{1}{30} y_k^{l+1} p_k^l + \frac{1}{20} y_k^{l+1} p_k^{l+1}) \} \\
& + \frac{2}{5} y^1 y^1 (y_k^1 p_k^1 + \frac{1}{20} y_k^1 p_k^2 + \frac{1}{20} y_k^2 p_k^1 + \frac{1}{30} y_k^2 p_k^2) \\
& + y^1 y^2 (\frac{1}{20} y_k^1 p_k^1 + \frac{1}{30} y_k^1 p_k^2 + \frac{1}{30} y_k^2 p_k^1 + \frac{1}{20} y_k^2 p_k^2) \\
& + y^n y^{n-1} (\frac{1}{20} y_k^{n-1} p_k^{n-1} + \frac{1}{30} y_k^{n-1} p_k^n + \frac{1}{30} y_k^n p_k^{n-1} + \frac{1}{20} y_k^n p_k^n) \\
& + y^n y^n (\frac{1}{30} y_k^{n-1} p_k^{n-1} + \frac{1}{20} y_k^{n-1} p_k^n + \frac{1}{20} y_k^n p_k^{n-1} + \frac{2}{5} y_k^n p_k^n)
\end{aligned}$$

which means $\int_0^1 y_k p_k y^2 \text{ dx} = y^T A y$ with

$$A_{y_k}^{p_k} = \frac{1}{5(n+1)} \begin{pmatrix} a & b & 0 & . & . & . & 0 \\ *_1 & *_2 & *_3 & 0 & . & . & 0 \\ 0 & *_1 & *_2 & *_3 & 0 & . & 0 \\ . & . & . & . & . & . & . \\ 0 & . & . & 0 & *_1 & *_2 & *_3 \\ 0 & . & . & . & 0 & f & g \end{pmatrix}$$

and

$$\begin{aligned}
a &= 2 y_k^1 p_k^1 + \frac{1}{4} y_k^1 p_k^2 + \frac{1}{4} y_k^2 p_k^1 + \frac{1}{6} y_k^2 p_k^2 \\
b &= \frac{1}{4} y_k^1 p_k^1 + \frac{1}{6} y_k^1 p_k^2 + \frac{1}{6} y_k^2 p_k^1 + \frac{1}{4} y_k^2 p_k^2 \\
f &= \frac{1}{4} y_k^{n-1} p_k^{n-1} + \frac{1}{6} y_k^{n-1} p_k^n + \frac{1}{6} y_k^n p_k^{n-1} + \frac{1}{4} y_k^n p_k^n \\
g &= \frac{1}{6} y_k^{n-1} p_k^{n-1} + \frac{1}{4} y_k^{n-1} p_k^n + \frac{1}{4} y_k^n p_k^{n-1} + 2 y_k^n p_k^n
\end{aligned}$$

as well as

$$\begin{aligned}
*_1 &= \frac{1}{4} y_k^{l-1} p_k^{l-1} + \frac{1}{6} y_k^{l-1} p_k^l + \frac{1}{6} y_k^l p_k^{l-1} + \frac{1}{4} y_k^l p_k^l \\
*_2 &= \frac{1}{4} y_k^{l-1} p_k^{l-1} + \frac{1}{4} y_k^{l-1} p_k^l + \frac{1}{4} y_k^l p_k^{l-1} + 2 y_k^l p_k^l + \frac{1}{4} y_k^l p_k^{l+1} \\
*_3 &= \frac{1}{4} y_k^l p_k^l + \frac{1}{6} y_k^l p_k^{l+1} + \frac{1}{6} y_k^{l+1} p_k^l + \frac{1}{4} y_k^{l+1} p_k^{l+1}
\end{aligned}$$

for $l \in \{2...n-1\}$.

7.2 Computation of B

We want to derive the FEM matrix for the third therm of the linearized PDE

$$-\Delta y + y + 3y_k^2 y = u.$$

Let $y_k = \sum_{i=1}^n y_k^i \phi_i$ and $y = \sum_{i=1}^n y^i \phi_i$ be the FEM representations of y_k and y. We look at the third term of the weak formulation :

$$3\int_0^1 y_k^2 y v_h = 3\int_0^1 (\sum_{i=1}^n (y_k^i)^2 \phi_i)(\sum_{i=1}^n y_k \phi_i) v_h \text{ dx}$$

To derive the entries of the j-th column of B we set $v_h = \phi_j$. Due to the disjoint support of non-neighboring ansatzfunctions the term above can be reduced to:

$$\begin{aligned}
&3\int_0^1 ((y_k^{j-1})^2\phi_{j-1} + (y_k^j)^2\phi_j + (y_k^{j+1})^2\phi_{j+1})(y^{j-1}\phi_{j-1} + y^j\phi_j + y^{j+1}\phi_{j+1})\phi_j \text{ dx} \\
= &3\int_0^1 (y_k^{j-1})^2\phi_{j-1}^2\phi_j y^{j-1} + (y_k^{j-1})^2\phi_{j-1}\phi_j^2 y^j \\
&+ (y_k^j)^2\phi_j^2\phi_{j-1}y^{j-1} + (y_k^j)^2\phi_j^3 y^j + (y_k^j)^2\phi_j^2\phi_{j+1}y^{j+1} \\
&+ (y_k^{j+1})^2\phi_j^2\phi_{j+1}y^j + (y_k^{j+1})^2\phi_j\phi_{j+1}^2 y^{j+1} \text{ dx} \\
= &3[(y_k^{j-1})^2\int_0^1 \phi_{j-1}^2\phi_j \text{ dx } y^{j-1} + (y_k^{j-1})^2\int_0^1 \phi_{j-1}\phi_j^2 \text{ dx } y^j \\
&+ (y_k^j)^2\int_0^1 \phi_j^2\phi_{j-1} \text{ dx } y^{j-1} + (y_k^j)^2\int_0^1 \phi_j^3 \text{ dx } y^j + (y_k^j)^2\int_0^1 \phi_j^2\phi_{j+1} \text{ dx } y^{j+1} \\
&+ (y_k^{j+1})^2\int_0^1 \phi_j^2\phi_{j+1} \text{ dx } y^j + (y_k^{j+1})^2\int_0^1 \phi_j\phi_{j+1}^2 \text{ dx } y^{j+1}]
\end{aligned}$$

We compute the involved integrals and get for suitable $j \in \{1, ..., n\}$:

$$\begin{aligned}
\int_0^1 \phi_{j-1}^2\phi_j \text{ dx} = \int_0^1 \phi_{j-1}\phi_j^2 \text{ dx} = \int_0^1 \phi_j^2\phi_{j+1} \text{ dx} = \int_0^1 \phi_j\phi_{j+1}^2 \text{ dx} = \frac{1}{12(n+1)} \\
\int_0^1 \phi_j^3 \text{ dx} = \frac{1}{2(n+1)}
\end{aligned}$$

This leads to

$$
\begin{aligned}
&3[(y_k^{j-1})^2 \int_0^1 \phi_{j-1}^2\phi_j \text{ dx } y^{j-1} + (y_k^{j-1})^2 \int_0^1 \phi_{j-1}\phi_j^2 \text{ dx } y^j \\
&\quad + (y_k^j)^2 \int_0^1 \phi_j^2\phi_{j-1} \text{ dx } y^{j-1} + (y_k^j)^2 \int_0^1 \phi_j^3 \text{ dx } y^j + (y_k^j)^2 \int_0^1 \phi_j^2\phi_{j+1} \text{ dx } y^{j+1} \\
&\quad + (y_k^{j+1})^2 \int_0^1 \phi_j^2\phi_{j+1} \text{ dx } y^j + (y_k^{j+1})^2 \int_0^1 \phi_j\phi_{j+1}^2 \text{ dx } y^{j+1} \text{ dx }] \\
&= 3[(y_k^{j-1})^2 y^{j-1}\frac{1}{12(n+1)} + (y_k^{j-1})^2 y^j \frac{1}{12(n+1)} \\
&\quad + (y_k^j)^2 y^{j-1}\frac{1}{12(n+1)} + (y_k^j)^2 y^j \frac{1}{2(n+1)} + (y_k^j)^2 y^{j+1}\frac{1}{12(n+1)} \\
&\quad + (y_k^{j+1})^2 y^j \frac{1}{12(n+1)} + (y_k^{j+1})^2 y^{j+1}\frac{1}{12(n+1)}] \\
&= 3[y^{j-1}\frac{(y_k^{j-1})^2 + (y_k^j)^2}{12(n+1)} + y^j(\frac{(y_k^{j-1})^2}{12(n+1)} + \frac{(y_k^j)^2}{2(n+1)} \\
&\quad + \frac{(y_k^{j+1})^2}{12(n+1)}) + y^{j+1}\frac{(y_k^j)^2 + (y_k^{j+1})^2}{12(n+1)}] \\
&\quad = \frac{1}{4(n+1)}[y^{j-1}((y_k^{j-1})^2 + (y_k^j)^2) + y^j((y_k^{j-1})^2 + 6(y_k^j)^2 + (y_k^{j+1})^2) \\
&\quad + y^{j+1}((y_k^j)^2 + (y_k^{j+1})^2)].
\end{aligned}
$$

Thus we see

$$
B = \eta \begin{pmatrix}
6 * (y_k^1)^2 & (y_k^1)^2 + (y_k^2)^2 & 0 & . & . & . & 0 \\
*_1 & *_2 & *_3 & 0 & . & . & 0 \\
0 & *_1 & *_2 & *_3 & 0 & . & 0 \\
. & . & . & . & . & . & . \\
0 & . & . & 0 & *_1 & *_2 & *_3 \\
0 & . & . & . & 0 & (y_k^{n-1})^2 + (y_k^n)^2 & (y_k^{n-1})^2 + 6(y_k^n)^2
\end{pmatrix}
$$

with

$$
\begin{aligned}
\eta &= \frac{1}{4(n+1)} \\
*_1 &= (y_k^{l-1})^2 + (y_k^l)^2 \\
*_2 &= (y_k^{l-1})^2 + 6(y_k^l)^2 + (y_k^{l+1})^2 \\
*_3 &= (y_k^l)^2 + (y_k^{l+1})^2
\end{aligned}
$$

for $l \in \{2...n-1\}$.

7.3 L^2-norms of the second derivatives of u_d and y_d

To evaluate the numerical results we used estimations involving the second order derivatives of u_d and y_d. We start with the optimal/desired control u_d. We have

$$
\begin{aligned}
u_d &= c_c(\pi^2+1)\sin(\pi x) + c_c^3 \sin^3(\pi x) \\
u_d' &= c_c(\pi^2+1)\pi\cos(\pi x) + 3c_c^3 \sin^2(\pi x)\cos(\pi x)\pi \\
u_d'' &= c_c(\pi^2+1)\pi^2(-\sin(\pi x) + 6c_c^3\pi^2 \sin(\pi x)\cos^2(\pi x) - 3c_c^3\pi^2\sin^3(\pi x)
\end{aligned}
$$

and

$$
\begin{aligned}
(u_d'')^2 = {} & c_c^2(\pi^4+\pi^2)^2\sin^2(\pi x) - 12c_c^4\pi^2(\pi^4+\pi^2)\sin^2(\pi x)\cos^2(\pi x) \\
& + 6c_c^4\pi^2(\pi^4+\pi^2)\sin^4(\pi x) - 36c_c^6\pi^4\sin^4(\pi x)\cos^2(\pi x) \\
& + 36c_c^6\pi^4\sin^2(\pi x)\cos^4(\pi x) + 9c_c^6\pi^4\sin^6(\pi x).
\end{aligned}
$$

Thus we get:

$$
\begin{aligned}
\|u_d''\|_{L^2([0,1])} = {} & \Big[c_c^2(\pi^4+\pi^2)^2\int_0^1 \sin^2(\pi x)\ \mathrm{dx} \\
& - 12c_c^4\pi^2(\pi^4+\pi^2)\int_0^1 \sin^2(\pi x)\cos^2(\pi x)\ \mathrm{dx} \\
& + 6c_c^4\pi^2(\pi^4+\pi^2)\int_0^1 \sin^4(\pi x)\ \mathrm{dx} - 36c_c^6\pi^4\int_0^1 \sin^4(\pi x)\cos^2(\pi x)\ \mathrm{dx} \\
& + 36c_c^6\pi^4\int_0^1 \sin^2(\pi x)\cos^4(\pi x)\ \mathrm{dx}\ + 9c_c^6\pi^4\int_0^1 \sin^6(\pi x)\ \mathrm{dx}\ \Big]^{1/2} \\
= {} & \Big[\frac{c_c^2(\pi^4+\pi^2)^2}{2} - \frac{12c_c^4\pi^2(\pi^4+\pi^2)}{8} \\
& + \frac{18c_c^4\pi^2(\pi^4+\pi^2)}{8} - \frac{36c_c^6\pi^4}{16} + \frac{36c_c^6\pi^4}{16} + \frac{45c_c^6\pi^4}{16}\Big]^{1/2} \\
= {} & \Big[\frac{c_c^2(\pi^4+\pi^2)^2}{2} + \frac{3c_c^4\pi^2(\pi^4+\pi^2)}{4} + \frac{45c_c^6\pi^4}{16}\Big]^{1/2}
\end{aligned}
$$

This leaves us with the computation of $\|y_d''\|_{L^2([0,1])}$:

$$
y_d = c_c\sin(\pi x),\ \ y_d' = c_c\pi\cos(\pi x),\ \ y_d'' = -c_c\pi^2\sin(\pi x)
$$

This results in:

$$
\begin{aligned}
\|y_d''\|^2_{L^2([0,1])} &= \int_0^1 (-c_c\pi^2\sin(\pi x)^2\ \mathrm{dx} \\
&= \int_0^1 c_c^2\pi^4\sin^2(\pi x)\ \mathrm{dx} = c_c^2\pi^4\int_0^1 \sin^2(\pi x)\ \mathrm{dx} \\
&= \frac{c_c^2\pi^4}{2} \\
\Rightarrow\ \ \|y_d''\|_{L^2([0,1])} &= \frac{c_c\pi^2}{\sqrt{2}}
\end{aligned}
$$

7.4 $G\bar{u}_h - y_d$ error

Throughout the estimation process we used the errors $\|G\bar{u}_h - y_d\|_{L^2[0,1]}$ and $\|G_h\bar{u}_h - y_d\|_{L^\infty[0,1]}$. But our numerical calculations provide only the errors for the discrete control-state operator and the discrete representation of y_d, i.e. $\|G_h\bar{u}_h - I_h yd\|_{L^2[0,1]}$ and $\|G_h\bar{u}_h - I_h yd\|_{L^\infty[0,1]}$. Thus we have to estimate the former errors using the latter. Using the informations at hand we estimate the L^2-error and L^∞-error:

$$\begin{aligned}
\|G\bar{u}_h - y_d\|_{L^2([0,1])} &= \|G\bar{u}_h - G_h\bar{u}_h + G_h\bar{u}_h - I_h y_d + I_h y_d - y_d\|_{L^2([0,1])} \\
&\le \|G\bar{u}_h - G_h\bar{u}_h\|_{L^2([0,1])} + \|G_h\bar{u}_h - I_h y_d\|_{L^2([0,1])} + \|I_h y_d - y_d\|_{L^2([0,1])} \\
&\le c_G h\|\bar{u}_h\|_{L^2([0,1])} + \|G_h\bar{u}_h - I_h y_d\|_{L^2([0,1])} + h^2\|y_d''\|_{L^2([0,1])}
\end{aligned}$$

$$\begin{aligned}
\|G_h\bar{u}_h - y_d\|_{L^\infty([0,1])} &= \|G_h\bar{u}_h - I_h y_d + I_h y_d - y_d\|_{L^\infty([0,1])} \\
&\le \|G_h\bar{u}_h - I_h y_d\|_{L^\infty} + \|I_h y_d - y_d\|_{L^\infty([0,1])} \\
&\le \|G_h\bar{u}_h - I_h y_d\|_{L^\infty} + \frac{1}{\sqrt{2}}\|I_h y_d - y_d\|_{H^1([0,1])} \\
&\le \|G_h\bar{u}_h - I_h y_d\|_{L^\infty([0,1])} + \frac{1}{\sqrt{2}}(h^2 + h)\|y_d''\|_{L^2([0,1])}
\end{aligned}$$

Using the result of Section 7.3 we get:

$$\begin{aligned}
\|G\bar{u}_h - y_d\|_{L^2([0,1])} &\le c_G h\|\bar{u}_h\|_{L^2([0,1])} + \|G_h\bar{u}_h - I_h y_d\|_{L^2([0,1])} + \frac{c_c\pi^2}{\sqrt{2}}h^2 \\
\|G\bar{u}_h - y_d\|_{L^\infty([0,1])} &\le \|G_h\bar{u}_h - I_h y_d\|_{L^\infty([0,1])} + \frac{c_c\pi^2}{2}(h^2 + h)
\end{aligned}$$

Chapter 8

Conclusion and perspectives

In this thesis we studied abstract nonlinear optimization problems in Banach and Hilbert spaces.
In the first part we derived the sufficient optimality condition and the error estimate. We assumed the existence of a discretized and thus numerical solvable version of such a problem. Depending on the discrete solution $\bar{u}_h$ and properties of both involved problems, the continuous as well as the discretized, we developed a set of sufficient optimality conditions, which ensure existence of a solution in a neighborhood of $\bar{u}_h$ and also delivered an error estimate for this solution. The presented method has the benefit that it only depends on computable quantities and that the conditions can therefore be checked when there is a numerical solution at hand.

In the second part we applied the theory on an one-dimensional example. We derived the estimates for all involved constants and developed the techniques, which were essential to conduct the estimation process.

The last part was dedicated to the numerical methods. We introduced the FE method, which we used for the discretization of inifinite dimensional spaces. To deal with nonlinear optimization problems we introduced the Lagrange-Newton SQP. We conducted the necessary computations and transformed the example into the numerical problem and applied the numerical methods. We presented the results and deduced the actual errors from the observed numerical quantities. We interpreted the data and saw that the sufficient optimality conditions were satisfied and that the error estimate holds, although we observed weaker convergence for the error estimate. But even with optimized estimates one can show, using the results of [37], that the maximal achieveable estimate is of order h. This result is based on the satisfied SSC. If one uses only a priori arguments one can only expect an order of $h^{3/4}$ using the presented technique. This is mainly caused by the fact that we have to deal with an L^2 environment.

To conclude this work let us comment on some further aspects:

We presented our computations by means of an one dimensional example. However, many of them can be conducted for higher dimensions with the presented techniques. The crucial estimates, which cannot be transferred to higher dimension, are those involving the imbedding $H^1([0,1]) \hookrightarrow L^\infty([0,1])$. Essentially these are the estimates L, M, N and $c_{G'}$.
For $\Omega \subset \mathbb{R}^d$, with $d = 2, 3$, we know that $H^2(\Omega) \hookrightarrow C(\bar{\Omega})$ holds, which could be used as substitute for the higher dimensional estimates. The inverse estimate can be formulated as

$$\|\bar{v}_h\|_{L^\infty(\Omega)} \leq c(\Omega) h^{-d/2} \|\bar{v}_h\|_{L^2(\Omega)},$$

which could be used to derive L^∞-estimates for FE-errors. In higher dimensions one has to include the geometry of Ω into the estimation process, as one can for example see in the inverse estimate above. Especially the involved imbedding constants depend on Ω and their computation leads to several eigenvalue problems. This adds another difficulty to the technical aspects, which one has to keep in mind.

Another question, which can be interesting in the future, is if and when the SSC of the discrete solution $\bar{u}_h$ entails an SSC for the continuous solution $\bar{u}$.
While [2], [3], [41] and [42] investigated this question thoroughly for control constrained problems, state constrained problems pose different kind of challenges due to the low regularity of the Lagrange multipliers.
It is desirable to overcome these difficulties and find a positive answer to this question, because it would enable us to employ a-priori-theory, which would lead to better FE-error estimates, especially for higher dimensions.
On the other hand it would also effect regularity approaches, as presented for example in [28]. At this point it does not seem possible to achieve this goal for the general abstract problem. However, the ideas presented in [20] may be a key to reach results for a special class of problems.

List of Figures

List of Tables

References

[1] R.A. Adams. *Sobolev Spaces.* Academic Press, 1978.

[2] S. Akindeinde. *Numerical verification of optimality and optimality conditions for optimal control problems.* dissertation, Universität Würzburg, 2013.

[3] S. Akindeinde and D. Wachsmuth. A posteriori verification of optimality conditions for control problems with finite-dimensional control space. *Numerical Functional Analysis and Optimization*, 33:473–523, 2012.

[4] W. Alt. The Lagrange-Newton method for infinite-dimensional optimization problems. *Numer. Funct. Anal. Optim.*, 11:201–224, 1990.

[5] W. Alt. *Nichtlineare Optimierung.* Vieweg, 2002.

[6] W. Alt, R. Griesse, N. Metla, and A. Rösch. Lipschitz stability for elliptic optimal control problems with mixed control-state constraints. *Optimization*, 59:833–849, 2010.

[7] W. Alt, R. Sontag, and F. Tröltzsch. An SQP method for optimal control of weakly singular Hammerstein integral equations. *Appl. Math. Optim.*, 33 (3):227–252, 1996.

[8] M. Braack. Finite Elemente,Vorlesungsscript. 2012.

[9] E. Casas. Control of an elliptic problem with pointwise state constraints. *SIAM Journal on Control and Optimization*, 24 (6):1309–1318, 1986.

[10] E. Casas. Error estimates for the numerical approximation of semilinear elliptic control problems with finitely many state constraints. *ESAIM, Control Optim. Calc. Var.*, 8:345–374, 2002.

[11] E. Casas, J.C. De los Reyes, and F. Tröltzsch. Sufficient second-order optimality conditions for semilinear control problems with pointwise state constraints. *SIAM Journal on Optimization*, 19 (2):616–643, 2008.

[12] E. Casas, M. Mateos, and F. Tröltzsch. Error estimates for the numerical approximation of boundary semilinear elliptic equations. *Computational Optimization and Applications*, 31:193–220, 2005.

[13] E. Casas and F. Tröltzsch. Second-order necessary optimality conditions for some state-constrained control problems of semilinear elliptic equations. *Appl. Math. Optim.*, 39 (2):211–227, 1999.

[14] E. Casas and F. Tröltzsch. Recent advances in the analysis of pointwise state-constrainted elliptic optimal control problems. *ESAIM*, 16:581–600, 2010.

[15] E. Casas, F. Tröltzsch, and A. Unger. Second order sufficient optimality conditions for some state-constrained control problems of semilinear elliptic equations. *SIAM Journal on Control and Optimization*, 38 (5):1369–1391, 2000.

[16] C.Meyer and A.Rösch. L^∞-estimates for approximated optimal control problems. *SIAM Journal on Control and Optimization*, 44:1636–1649, 2005.

[17] K. Deckelnick and M. Hinze. Convergence of a finite element approximation to a state-constrained elliptic control problem. *SIAM J. Numer. Anal.*, 45:1937–1953, 2007.

[18] K. Deckelnick and M. Hinze. Numerical analysis of a control and state constrained elliptic control problem with piecewise constant control approximations. *Numerical Mathematics and Advanced Applications, Proceedings of ENUMATH 2007*, pages 597–604, 2008.

[19] H. Goldberg and F. Tröltzsch. Second order optimality conditions for a class of control problems governed by nonlinear integral equations with application to a parabolic boundary control. *Optimization*, 20 (5):687–698, 1989.

[20] R. Griesse. Lipschitz stability of solutions to some state-constrained elliptic optimal control problems. *Journal for Analysis and its Applications*, 25:435–455, 2006.

[21] R. Griesse, N. Metla, and A. Rösch. Local quadratic convergence of SQP for elliptic optimal control problems with mixed control-state constraints. *Control Cybernet.*, 39 (3):717–738, 2010.

[22] M. Heinkenschloss and F. Tröltzsch. Analysis of the Lagrange-SQP-Newton method for the control of a phase field equation. *Control Cybernet.*, 28 (2):177–211, 1999.

[23] M. Hintermüller and K. Kunisch. Path-following methods for a class of constrained minimization problems in function space. *SIAM J. Optimization*, 17:159–187, 2006.

[24] M. Hinze. A variational discretization concept in control constrained optimization: the linear quadratic case. *J. Computational Optimization and Applications*, 30:45–63, 2005.

[25] A.D. Ioffe. Necessary and sufficient conditions for a local minimum. 3: Second order conditions and augemented duality. *SIAM Journal on Control and Optimization*, 17:266–288, 1979.

[26] J. Douglas Jr., T. Dupont, and L. Wahlbin. The stability in L^q of the L^2-projection into finite element function spaces. *Numerical Mathematics*, 23:193–197, 1975.

[27] M. Jung and U. Langer. *Methode der finiten Elemente für Ingenieure*. Teubner, 2001.

[28] K. Krumbiegel, I. Neitzel, and A. Rösch. Sufficient optimality conditions for the Moreau-Yosida-type regularization concept applied to semilinear elliptic optimal control problems with pointwise state constraints. *Ann. Acad. Rom. Sci. Ser. Math. Appl.*, 2:222–246, 2010.

[29] S. Kurcyusz and J. Zowe. Regularity and stability for the mathematical programming problem in Banach spaces. *Appl. Math. Optimization*, 5:49–62, 1979.

[30] J.L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer-Verlag, 1971.

[31] H. Maurer and J. Zowe. First- and second-order conditions in infinite-dimensional programming problems. *Math. Programming*, 16:98–110, 1979.

[32] C. Meyer. Error estimates for the finite-element approximation of an elliptic control problem with pointwise state and control constraints. *Control Cybern.*, 37:51–83, 2008.

[33] C. Meyer, A. Rösch, and F. Tröltzsch. Optimal control of PDEs with regularized pointwise state constraints. *Computational Optimization and Applications*, 33:209–228, 2006.

[34] C. Meyer and F. Tröltzsch. On an elliptic optimal control problem with pointwise mixed control-state constraints. *Recent advances in optimization, Lecture notes in Econom. and Math. Systems*, 563:187–204, 2006.

[35] H.D. Mittelmann and F.Tröltzsch. Sufficient optimality in a parabolic control problem. *Trends in industrial and applied mathematics, Appl. Optim*, 72:305–316, 2002.

[36] N.Arada, J.-P. Raymond, and F. Tröltzsch. On an augmented Lagrangian SQP method for a class of optimal control problems in Banach spaces. *Computational Optimization and Applications*, 22:369–398, 2002.

[37] I. Neitzel, J. Pfefferer, and A. Rösch. Finite element discretization of state-constrained elliptic optimal control problems with semilinear state equation. page in preparation.

[38] R. Rannacher. Numerische Mathematik 2, Vorlesungsscript. 2004.

[39] J.-P. Raymond and F. Tröltzsch. Second order sufficient optimality conditions for nonlinear parabolic control problems with state constraints. *Discrete Contin. Dynam. Systems.*, 6 (2):431–450, 2000.

[40] A. Rösch and F. Tröltzsch. On regularity of solutions and lagrange multipliers of optimal control problems for semilinear equations with mixed pointwise control-state constraints. *SIAM Journal on Control and Optimization*, 46:1098–1115, 2007 (3).

[41] A. Rösch and D. Wachsmuth. Numerical verification of optimality conditions. *SIAM Journal on Control and Optimization*, 47:2557–2581, 2008.

[42] A. Rösch and D. Wachsmuth. How to check numerically the suffient optimality conditions for infinite-dimensional optimization problems. *Internat. Ser. Numer. Math*, 158:297–317, 2009.

[43] A. Schiela. Barrier methods for optimal control problems with state constraints. *SIAM J. Optimization*, 20:1002–1031, 2009.

[44] D. Tiba and F. Tröltzsch. Error estimates for the discretization of state constrained convex control problems. *Num. Funct. Anal. Optim.*, 17:1005–1028, 1996.

[45] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen.* Vieweg, 2005.

[46] F. Tröltzsch. Lipschitz stability of solutions of linear-quadratic parabolic control problems with respect to perturbations. *Dynam. Contin. Discrete Impuls System*, 7 (2):289–306, 2009.

[47] F. Tröltzsch and S. Volkwein. The SQP method for control constrained optimal control of the Burgers equation. *ESAIM Control Optim. Calc. Var.*, 6:649–674, 2001.