JOHANNES KEPLER
UNIVERSITÄT LINZ
Netzwerk für Forschung, Lehre und Praxis

Exact and Inexact Semismooth Newton Methods for Elliptic Optimal Control Problems

BAKKALAUREATSARBEIT

zur Erlangung des akademischen Grades

BAKKALAUREUS DER TECHNISCHEN WISSENSCHAFTEN

in der Studienrichtung

TECHNISCHE MATHEMATIK

Angefertigt am *Institut für Numerische Mathematik*

Betreuung:

O. Univ. Prof. Dipl. Ing. Dr. Ulrich Langer

Eingereicht von:

Peter Gangl

Linz, Oktober 2010

Contents

Chapter 1

Introduction

1.1 Motivation

A wide range of processes in practical applications can be described by mathematical models which are based on partial differential equations (PDEs). Typically, the objective is to optimize those processes, e.g. to minimize the costs or to maximize the profit, so we are facing problems of PDE-constrained optimization. Such problems can be found in a wide variety of applications, e.g. in fluid mechanics or structural mechanics, but also more and more medical applications make use of this mathematical approach (see, e.g., [1]).

In this thesis, we will concentrate on one important class of PDE-constrained optimization problems, namely on **optimal control problems**. For simplicity we will only consider linear elliptic PDEs as constraints. Our optimal control problems will be of the following form:

$$\min_{y\in H_0^1(\Omega),u\in L^2(\Omega_c)} J(y,u) := \frac{1}{2}\|y-y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega_c)} \tag{1.1}$$
$$\text{s.t. } Ay = r + Bu \text{ and } \beta_l \le u \le \beta_r \text{ almost everywhere in } \Omega_c,$$

Here, $y \in H_0^1(\Omega)$ is the state, which is defined on the open bounded domain $\Omega \subset \mathbb{R}^n$. The control variable $u \in L^2(\Omega_c)$ is defined on the open bounded domain $\Omega_c \subset \mathbb{R}^m$ and regulates the state via some given PDE-constraint. The operator $A : H_0^1(\Omega) \to H^{-1}(\Omega)$ is a linear elliptic partial differential operator, e.g. $A = -\Delta$, $r \in H^{-1}(\Omega)$ is given and $B : L^{p'}(\Omega_c) \to H^{-1}(\Omega)$ is continuous and linear, with $p' \in [1, 2)$.
The control is subject to pointwise bounds β_l, $\beta_r \in \mathbb{R}$ with $\beta_l < \beta_r$. The objective is to drive the state y as close as possible to the desired state $y_d \in L^2(\Omega)$. The second part of the objective function penalizes excessive control costs, the regularization parameter $\alpha > 0$ is typically small.

The probably most typical example of optimal control problems are heat conduction problems where the control variable u represents a heat source which controls the temperature distribution y via some heat conduction PDE. Depending on whether this source is placed only on the boundary or on the whole domain (i.e. $\Omega_c = \partial\Omega$ or $\Omega_c = \Omega$), we are talking about *boundary control problems* or *distributed control problems*. In this thesis we will (for simplicity) only consider the case of distributed control problems, i.e. $\Omega_c = \Omega$. Due to limited heating and cooling capacities, we also have to take into account bounds for the heat source u.

In Section 1.3, we will see that we can reformulate such problems in terms of an operator equation of the form

$$G(x) = 0 \tag{1.2}$$

where $G : X \to Y$ is a not necessarily smooth operator from a Banach space X to a Banach space Y. This thesis mainly deals with the **semismooth Newton method**, a generalization of Newton's method that is capable of solving this type of operator equations and additionally maintains the superlinear convergence rate. After having derived the theory, we will illustrate our proceedings in terms of a simple heat conduction optimal control problem, which can be interpreted as a one-dimensional analogue of the hyperthermia treatment of prostate cancer (cf. e.g. [2] and the references therein).

1.2 Notation

The Frèchet-derivative (F-derivative) of an operator $G : X \to Y$ between Banach spaces is denoted by $G' : X \to \mathcal{L}(X, Y)$, where $\mathcal{L}(X, Y)$ is the set of all bounded linear operators from X to Y. For a Banach space U, we denote its dual space $\mathcal{L}(U, \mathbb{R})$ by U^*. In particular, the derivative of a real-valued function $f : U \to \mathbb{R}$ is denoted by $f' : U \to U^*$. In case of a Hilbert space U, the gradient $\nabla f : U \to U$ is the Riesz repsresentation of f', i.e.

$$(\nabla f(u), v)_U = \langle f'(u), v \rangle_{U^* \times U} \quad \forall v \in U.$$

Here $\langle f'(u), v \rangle_{U^* \times U}$ denotes the dual pairing between the dual space $U^* = \mathcal{L}(U, \mathbb{R})$ and U, and $(\cdot, \cdot)_U$ is the inner product. Note that in a Hilbert space we can do the identification $U^* = U$ via Riesz' isomorphism.
Furthermore, we will use the quantor $\dot{\forall}$ meaning "for almost all", i.e. for all up to a set of measure zero.

1.3 From the Optimal Control Problem to the Operator Equation

As already mentioned, we need to reformulate our problem as an operator equation of the form (1.2). We perform this reformulation in several steps (cf. [3]):
At first, assuming the solvability of the PDE constraint, we can eliminate the state variable y via the state equation, i.e. we formally replace y by $y(u) = A^{-1}(r + Bu)$. By defining the set of all admissible functions, $U_{ad} := \{u \in L^2(\Omega) : \beta_l \leq u \leq \beta_r \text{ a.e. on } \Omega\}$, we obtain the reduced problem

$$\min_{u \in U_{ad}} \hat{J}(u) := J(y(u), u) := \frac{1}{2}\|y(u) - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} \tag{1.3}$$

which is a quadratic optimization problem in the Hilbert space $L^2(\Omega)$.

Now, we reformulate this minimization problem in terms of a variational inequality:

Lemma 1.1. *A function $\bar{u} \in U_{ad}$ solves the quadratic minimization problem (1.3) if and only if it satisfies the following variational inequality:*

$$(\nabla \hat{J}(\bar{u}), u - \bar{u})_{L^2(\Omega)} \geq 0 \quad \forall u \in U_{ad}. \tag{1.4}$$

Proof
"$\Longrightarrow$": Let $\bar{u}$ be the solution of problem (1.3) and choose $u \in U_{ad}$ arbitrarily. Consider the convex combination

$$u(t) = \bar{u} + t(u - \bar{u})$$

for an arbitrary $t \in (0, 1]$. The convexity of U_{ad}, which can easily be seen, ensures $u(t) \in U_{ad}$. Exploiting the optimality of $\bar{u}$, we obtain $\hat{J}(u(t)) \geq \hat{J}(\bar{u})$ and therefore also

$$\frac{1}{t}\left(\hat{J}(\bar{u} + t(u - \bar{u})) - \hat{J}(\bar{u})\right) \geq 0.$$

If we let $t \downarrow 0$, we obtain

$$\hat{J}'(\bar{u})(u - \bar{u}) \geq 0 \quad \forall u \in U_{ad}$$

which is the same as the variational inequality (1.4).
"$\Longleftarrow$": Again, choose an arbitrary $u \in U_{ad}$. Due to the convexity of our functional $\hat{J}$ we can make use of the well-known estimate

$$\hat{J}(u) - \hat{J}(\bar{u}) \geq \hat{J}'(\bar{u})(u - \bar{u}).$$

By (1.4), the right hand side of the inequality is non-negative and we get $\hat{J}(u) \geq \hat{J}(\bar{u})$, which implies the optimality of $\bar{u}$.

□

As the next step, we will have a look at the point-wise optimality conditions. We state the following

Lemma 1.2. *A function $\bar{u} \in U_{ad}$ satisfies the variational inequality (1.4) if and only if the following relations hold for almost all $x \in \Omega$:*

$$\bar{u}(x) = \begin{cases} \beta_l & \text{for } \nabla \hat{J}(\bar{u})(x) > 0, \\ \in [\beta_l, \beta_r] & \text{for } \nabla \hat{J}(\bar{u})(x) = 0, \\ \beta_r & \text{for } \nabla \hat{J}(\bar{u})(x) < 0. \end{cases} \tag{1.5}$$

Proof
"$\Longrightarrow$": Assume that (1.5) is not satisfied. We define the measurable sets

$$\begin{aligned} A_+(\bar{u}) &= \{x \in \Omega : \nabla \hat{J}(\bar{u})(x) > 0\}, \\ A_-(\bar{u}) &= \{x \in \Omega : \nabla \hat{J}(\bar{u})(x) < 0\}. \end{aligned}$$

Since $\bar{u} \in U_{ad}$, by our assumption there exists either a set $E_+ \subset A_+(\bar{u})$ of positive measure with $\bar{u}(x) > \beta_l$ for all $x \in E_+$ or a set $E_- \subset A_-(\bar{u})$ of positive measure with $\bar{u}(x) < \beta_r$ for all $x \in E_-$. For the first case we define

$$u(x) = \begin{cases} \beta_l & \text{for } x \in E_+ \\ \bar{u}(x) & \text{for } x \in \Omega \setminus E_+ \end{cases}$$

We obtain

$$\int_\Omega \left(\nabla \hat{J}(\bar{u})(x)\right)(u(x)-\bar{u}(x))\,dx = \int_{E_+} \underbrace{\left(\nabla \hat{J}(\bar{u})(x)\right)}_{>0}\underbrace{(\beta_l - \bar{u}(x))}_{<0} dx < 0,$$

which is a contradiction to the variational inequality (1.4). For the second case, we proceed analogously by setting $u(x) = \beta_r$ on E_- and $u(x) = \bar{u}(x)$ in the remaining points.
"$\Longleftarrow$": For arbitrary $x \in A_+$ and $u \in U_{ad}$, we have $u(x) - \bar{u}(x) = u(x) - \beta_l \geq 0$. Since $\nabla \hat{J}(\bar{u})(x) > 0$ on $A_+(\bar{u})$, we obtain

$$\left(\nabla \hat{J}(\bar{u})(x)\right)(u(x) - \bar{u}(x)) \geq 0 \tag{1.6}$$

almost everywhere on $A_+(\bar{u})$. Analogously, (1.6) holds almost evereywhere on $A_-(\bar{u})$. In those points where $\nabla \hat{J}(\bar{u})$ vanishes, it is satisfied trivially, so it is valid almost everywhere in Ω. By integrating (1.6) over Ω we obtain (1.4).

□

Now we are able to reformulate our optimal control problem (1.1) as an equivalent operator equation of the form (1.2):

Theorem 1.3. *A function $\bar{u}$ is an optimal control for problem (1.1) if and only if it solves the operator equation*

$$\Phi(u) := u - P_{U_{ad}}(u - \theta \nabla \hat{J}(u)) = 0 \tag{1.7}$$

for almost all $x \in \Omega$, where $\theta > 0$ is arbitrary, but fixed. Here, $\hat{J}$ and U_{ad} are defined as in (1.3) and $P_{U_{ad}}$ is the L^2-projection onto U_{ad}, which is given by

$$P_{U_{ad}}(v)(x) = P_{[\beta_l,\beta_r]}(v(x)) = \max(\beta_l, \min(v(x), \beta_r)), \quad x \in \Omega.$$

Proof
By Lemma 1.1 and Lemma 1.2 as well as the reduction of the problem in (1.3) we know that a function $\bar{u} \in U_{ad}$ solves our optimal control problem (1.1) if and only if it satisfies (1.5). As one can see, $\bar{u}$ satisfying (1.5) is equivalent to

$$\bar{u}(x) = \begin{cases} \beta_l & \text{for } \bar{u}(x) - \theta\nabla\hat{J}(\bar{u})(x) < \beta_l, \\ \bar{u}(x) - \theta\nabla\hat{J}(\bar{u})(x) & \text{for } \bar{u}(x) - \theta\nabla\hat{J}(\bar{u})(x) \in [\beta_l, \beta_r], \\ \beta_r & \text{for } \bar{u}(x) - \theta\nabla\hat{J}(\bar{u})(x) > \beta_r, \end{cases} \tag{1.8}$$

which is just the definition of the L^2-projection of $\bar{u}(x) - \theta\nabla\hat{J}(\bar{u})(x)$ onto U_{ad}.

□

Remark 1.4.
- Note: Due to the projection operator $P_{U_{ad}}$ our operator equation (1.7) is non-smooth!

- The reformulation procedure illustrated above is applicable to all optimization problems of the form
$$\min_{u\in L^2(\Omega)} f(u) \quad \text{s.t.} \quad a \leq u \leq b \ \text{ a.e. on } \Omega$$
with convex and real-valued functional f and bounds a, $b \in L^\infty(\Omega)$. Those bounds can be transformed to constant bounds $\beta_l = 0$ and $\beta_r = 1$ via $u \mapsto \frac{u-a}{b-a}$.

- For the gradient of our functional $\hat{J}$, we obtain
$$\nabla \hat{J}(u) = \alpha u + B^*(A^{-1})^*(A^{-1}(r + Bu) - y_d) = \alpha u + H(u) \tag{1.9}$$
where $H(u) = B^*(A^{-1})^*(A^{-1}(r + Bu) - y_d)$. We will make use of this structure later.

Chapter 2

Generalized Newton's Method

In the following, we will investigate the general operator equation

$$G(x) = 0 \tag{2.1}$$

with the not necessarily differentiable operator $G : X \to Y$, where X and Y are Banach spaces.

If G is continuously Frèchet-(F-)differentiable and G' continuously invertible at the solution $\bar{x}$, we can apply the classical Newton method which converges superlinearly for an initial value x^0 sufficiently close to the solution $\bar{x}$.

Fortunately, we will see that we can generalize Newton's method also to not necessarily smooth operators and - under additional conditions - preserve the superlinear convergence behaviour:

Algorithm 1. (Generalized Newton's method)

0. Choose $x^0 \in X$ (sufficiently close to the solution $\bar{x}$)

For k=0, 1, 2, ...:

1. Choose an invertible operator $M_k \in \mathcal{L}(X, Y)$.
2. Obtain s^k by solving
$$M_k s = -G(x^k) \tag{2.2}$$
and set $x^{k+1} = x^k + s^k$.

We will now investigate the generated sequence (x^k) in a neighborhood of a solution $\bar{x}$ of (2.1), i.e. $G(\bar{x}) = 0$. For the distance $d^k := x^k - \bar{x}$ to the solution we have

$$\begin{aligned} M_k d^{k+1} &= M_k(x^{k+1} - \bar{x}) = M_k(x^k + s^k - \bar{x}) = M_k d^k - G(x^k) \\ &= G(\bar{x}) + M_k d^k - G(x^k). \end{aligned} \tag{2.3}$$

Using this observation we obtain the following convergence results:

Theorem 2.1. *Consider the operator equation (2.1) with $G : X \to Y$, where X and Y are Banach spaces. Let (x^k) be the sequence generated by the generalized Newton method (Algorithm 1). Then:*

1. *If x^0 is sufficiently close to $\bar{x}$ and*

$$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X \le \gamma \|d^k\|_X \quad \forall\, k \text{ with } \|d^k\|_X \text{ suff. small}, \tag{2.4}$$

then $x^k \to \bar{x}$ q-linearly with rate $\gamma \in (0,1)$.

2. *If x^0 is sufficiently close to $\bar{x}$ and*

$$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X = o(\|d^k\|_X) \text{ for } \|d^k\|_X \to 0, \tag{2.5}$$

then $x^k \to \bar{x}$ q-superlinearly.

3. *If x^0 is sufficiently close to $\bar{x}$ and for some $\alpha > 0$*

$$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X = \mathcal{O}(\|d^k\|_X^{1+\alpha}) \text{ for } \|d^k\|_X \to 0, \tag{2.6}$$

then $x^k \to \bar{x}$ q-superlinearly with order $1+\alpha$.

Proof

1. Let $\delta > 0$ be so small that (2.4) holds for all x^k with $\|d^k\|_X < \delta$. Then, for x^0 satisfying $\|x^0 - \bar{x}\|_X < \delta$, by (2.3) we have

$$\begin{aligned} \|x^1 - \bar{x}\|_X &= \|d^1\|_X = \|M_0^{-1}(G(\bar{x}+d^0) - G(\bar{x}) - M_0 d^0)\|_X \\ &\le \gamma\|d^0\|_X = \gamma\|x^0 - \bar{x}\|_X < \delta. \end{aligned}$$

Inductively, let $\|x^k - \bar{x}\|_X < \delta$. Then

$$\begin{aligned} \|x^{k+1} - \bar{x}\|_X &= \|d^{k+1}\|_X = \|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X \\ &\le \gamma\|d^k\|_X = \gamma\|x^k - \bar{x}\|_X < \delta. \end{aligned}$$

Hence, we have

$$\|x^{k+1} - \bar{x}\|_X \le \gamma \|x^k - \bar{x}\|_X \quad \forall k \ge 0.$$

2. Fix $\gamma \in (0,1)$ and let $\delta > 0$ be so small that (2.4) holds for all x^k with $\|d^k\|_X < \delta$. Then, for x^0 satisfying $\|x^0 - \bar{x}\|_X < \delta$, we can apply 1. to conclude $x^k \to \bar{x}$ with rate γ. Now, (2.5) immediately yields (again using (2.3))

$$\begin{aligned} \|x^{k+1} - \bar{x}\|_X &= \|d^{k+1}\|_X = \|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X = o(\|d^k\|_X) \\ &= o(\|x^k - \bar{x}\|_X) \qquad (\text{for } k \to \infty). \end{aligned}$$

3. As in 2, we can apply part 1 of this theorem and conclude the q-linear convergence of the sequence (x^k) to $\bar{x}$ with rate γ. Now, (2.6) yields

$$\begin{aligned} \|x^{k+1} - \bar{x}\|_X &= \|d^{k+1}\|_X = \|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X = \mathcal{O}(\|d^k\|_X^{1+\alpha}) \\ &= \mathcal{O}(\|x^k - \bar{x}\|_X^{1+\alpha}) \qquad (\text{for } k \to \infty). \end{aligned}$$

□

Remark 2.2. Conditions (2.4),(2.5) and (2.6) are meant uniformly in k.
For (2.4), that means that there exists $\delta_\gamma > 0$ such that

$$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X \le \gamma \|d^k\|_X \quad \forall\ k \text{ with } \|d^k\|_X < \delta_\gamma.$$

For (2.5), that means that for all $\eta \in (0,1)$, there exists $\delta_\eta > 0$ such that

$$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X \le \eta \|d^k\|_X \quad \forall\ k \text{ with } \|d^k\|_X < \delta_\eta.$$

The condition in (2.6) and is meant similarly.

Remark 2.3. For convenience we split the smallness assumption imposed on

$$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X$$

into two parts:

1. *Regularity condition:*
$$\forall k \ge 0 \exists C \in \mathbb{R} : \|M_k^{-1}\|_{Y\to X} \le C \tag{2.7}$$

2. *Approximation condition:*
$$\|G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k\|_Y = o(\|d^k\|_X) \text{ for } \|d^k\|_X \to 0. \tag{2.8}$$
or
$$\|G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k\|_Y = \mathcal{O}(\|d^k\|_X^{1+\alpha}) \text{ for } \|d^k\|_X \to 0. \tag{2.9}$$
From the definition of the F-derivative one can see immediately that for a F-differentiable operator G the F-derivative G' satisfies the superlinear approximation condition (2.8).

Since conditions (2.7) and (2.8) together are sufficient for (2.5), those two conditions also yield the superlinear convergence statement of Theorem 2.1. In practice it is more convenient to check (2.7) and (2.8) rather than (2.5).
Analogously, (2.7) and (2.9) yield the superlinear convergence of order $1+\alpha$.

Remark 2.4 (Classical Newton's Method)**.** If, for a continuously F-differentiable operator G with G' satisfying the regularity condition $\|G'(x^k)^{-1}\|_{Y\to X} \le C$ for all $k \ge 0$, we choose $M_k = G'(x^k)$ for each k in step 1 of Algorithm 1 we obtain the classical Newton's method, which converges locally q-superlinearly. Moreover, if G' is α-order Hölder continuous near $\bar{x}$, the order of convergence is $1+\alpha$.

Chapter 3

Generalized Differential and Semismoothness

Since the operators appearing in our applications are often non-smooth, the question arises if we can find suitable substitutes M_k for $G'(x^k)$. Therefore, we will consider set-valued mappings $\partial G : X \rightrightarrows \mathcal{L}(X, Y)$ as generalized differentials. Then we will choose M_k point-based in each step, i.e.

$$M_k \in \partial G(x^k). \tag{3.1}$$

In order to preserve the convergence statements of Theorem 2.1, the two conditions (2.7) and (2.8) (or (2.9)) have to be fulfilled. For the superlinear approximation condition (2.8) to be satisfied for every such choice M_k, we require

$$\sup_{M \in \partial G(x+d)} \|G(x+d) - G(x) - Md\|_Y = o(\|d\|_X) \text{ for } \|d\|_X \to 0. \tag{3.2}$$

This, together with the same consideration for the α-order approximation condition (2.9), leads us to the definition of semismoothness:

Definition 3.1. *(Semismoothness) Let X, Y be Banach spaces, $G : X \to Y$ a continuous operator and let be given the set-valued mapping $\partial G : X \rightrightarrows \mathcal{L}(X, Y)$ with non-empty images. Then*

1. *G is called ∂G-semismooth at $x \in X$ if*

$$\sup_{M \in \partial G(x+d)} \|G(x+d) - G(x) - Md\|_Y = o(\|d\|_X) \text{ for } \|d\|_X \to 0. \tag{3.3}$$

2. *G is called ∂G-semismooth of order $\alpha > 0$ at $x \in X$ if*

$$\sup_{M \in \partial G(x+d)} \|G(x+d) - G(x) - Md\|_Y = O(\|d\|_X^{1+\alpha}) \text{ for } \|d\|_X \to 0. \tag{3.4}$$

So, if we have an operator equation of type (2.1) with G being ∂G-semismooth for some set-valued mapping ∂G, we can - under some additional regularity condition - automatically apply Algorithm 1 which for adequate initial value now converges q-superlinearly (or q-superlinearly of order $1 + \alpha$) to the solution $\bar{x}$. However, the hard part will be to construct such a set-valued mapping ∂G such that the given operator G is ∂G-semismooth.

As a first result, which should be intuitively clear, we will show that a smooth (i.e. continuously F-differentiable) operator is also semismooth with respect to its F-derivative:

Lemma 3.2. *Let $G : X \to Y$ be continuously F-differentiable near $x \in X$ and $\{G'\}$ the set-valued operator $\{G'\} : X \rightrightarrows \mathcal{L}(X,Y)$ mapping x to the one element set $\{G'(x)\}$.*

1. *Then G is $\{G'\}$-semismooth at x.*
2. *Furthermore, if G' is α-order Hölder continuous near x, then G is $\{G'\}$- semismooth at x of order α.*

Proof

1. Here we use the definition of F-differentiability and the continuity of G':

$$\begin{aligned}
&\sup_{M\in\{G'\}(x+d)} \|G(x+d)-G(x)-Md\|_Y = \|G(x+d)-G(x)-G'(x+d)d\|_Y \\
&\le \|G(x+d)-G(x)-G'(x)\|_Y + \|G'(x)-G'(x+d)d\|_Y \\
&\le o(\|d\|_X) + \|G'(x)-G'(x+d)\|_{X\to Y}\|d\|_X = o(\|d\|_X).
\end{aligned}$$

2. Here we use the Hölder-continuity of G' of order α with Hölder constant L:

$$\begin{aligned}
&\sup_{M\in\{G'\}(x+d)} \|G(x+d)-G(x)-Md\|_Y = \|G(x+d)-G(x)-G'(x+d)d\|_Y \\
&= \left\| \int_0^1 (G'(x+td)-G'(x+d))d\mathrm{d}t \right\|_Y \le \int_0^1 \|G'(x+td)-G'(x+d)\|_{X\to Y}\mathrm{d}t\|d\|_X \\
&\le \int_0^1 L(1-t)^\alpha \|d\|_X^\alpha \mathrm{d}t\|d\|_X = \frac{L}{1+\alpha}\|d\|_X^{1+\alpha} = \mathcal{O}(\|d\|_X^{1+\alpha}).
\end{aligned}$$

□

To give the reader a notion of this idea of semismoothness, we will give a simple finite-dimensional example. First, we need a set-valued mapping with respect to which a function between finite-dimensional spaces is semismooth:

Definition 3.3. *(Clarke's generalized Jacobian) Let the function $G : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz-continuous.*
Then the set-valued mapping

$$\begin{aligned}
\partial^{cl} G : \mathbb{R}^n &\rightrightarrows \mathbb{R}^{m\times n} \\
\partial^{cl} G(x) &= \mathrm{conv}\{M \in \mathbb{R}^{m\times n} : \exists (x^k) \to x, G \text{ differentiable at } x^k : G'(x^k) \to M\} \qquad (3.5)
\end{aligned}$$

is called Clarke's generalized Jacobian.

Here, the convergence in $\mathbb{R}^{m\times n}$ is meant with respect to the norm associated to the vector norms in $\mathbb{R}^m$ and $\mathbb{R}^n$, i.e.

$$\|M\|_{\mathbb{R}^{m\times n}} = \underset{y\in\mathbb{R}^n\setminus\{0\}}{sup} \frac{\|My\|_{\mathbb{R}^m}}{\|y\|_{\mathbb{R}^n}}$$

Remark 3.4. Since - by Rademacher's Theorem (cf. [4]) - G' exists almost everywhere on $\mathbb{R}^n$ for Lipschitz-continuous G, $\partial^{cl}G(x)$ is non-empty for all $x \in \mathbb{R}^n$ and the definition is justified.
For locally Lipschitz-continuous functions $G : \mathbb{R}^n \to \mathbb{R}^m$, the standard choice for ∂G is $\partial^{cl}G$. The classical definition of semismoothness for functions $G : \mathbb{R}^n \to \mathbb{R}^m$ as in [5] is equivalent to $\partial^{cl}G$-semismoothness in connection with directional differentiability of G.

Now we can give a concrete example:

Example 1. For $a, b \in \mathbb{R}$ let $P_{[a,b]} : \mathbb{R} \to \mathbb{R}, P_{[a,b]}(x) = \max(a, \min(x, b))$ be the projection onto the interval $[a, b]$.
Consider $\psi : \mathbb{R} \to \mathbb{R}$, $\psi(x) = P_{[a,b]}(x)$, $a < b$, then Clarke's generalized derivative is

$$\partial^{cl}\psi(x) = \begin{cases} \{0\} & x < a \text{ or } x > b, \\ \{1\} & a < x < b, \\ \text{conv}\{0,1\} = [0,1] & x = a \text{ or } x = b \end{cases}$$

and ψ is $\partial^{cl}\psi$-semismooth, which we see as follows:

- For $x \notin \{a, b\}$ we have that ψ is continuously differentiable in a neighborhood of x and $\partial^{cl}\psi \equiv \{\psi'\}$. Hence, by Lemma 3.2, ψ is $\partial^{cl}\psi$-semismooth at x.
- For $x = a$, we estimate explicitly: For small $d > 0$, we have $\partial^{cl}\psi(x+d) = \{\psi'(a+d)\} = \{1\}$ and thus

  $$\sup_{M\in\partial^{cl}\psi(x+d)} |\psi(x + d) - \psi(x) - Md| = a + d - a - 1 \cdot d = 0.$$

  For small $d < 0$, we have $\partial^{cl}\psi(x + d) = \{\psi'(a + d)\} = \{0\}$ and thus

  $$\sup_{M\in\partial^{cl}\psi(x+d)} |\psi(x + d) - \psi(x) - Md| = a - a - 0 \cdot d = 0.$$

Hence, the $\partial^{cl}\psi$-semismoothness of ψ at $x = a$ is proved.
For $x = b$ we can can do exactly the same.

We proceed by establishing semismoothness results for the direct product, the sum and the composition of semismooth operators.

Theorem 3.5. *(Calculus) Let X, Y, Z, X_1, X_2, Y_1, Y_2 be Banach spaces.*

(a) If the operators $G_i : X \to Y_i$ are ∂G_i-semismooth at x, $i = 1, 2$, then the direct product

$$G := \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} : X \to Y_1 \times Y_2 =: Y$$

is ∂G-semismooth at x with

$$\begin{aligned}\partial G := \begin{pmatrix}\partial G_1\\ \partial G_2\end{pmatrix} : X &\rightrightarrows \mathcal{L}(X,Y_1)\times\mathcal{L}(X,Y_2)\\ \partial G(x) &= \{M\in\mathcal{L}(X,Y) : Mx = \begin{pmatrix}M_1x\\ M_2x\end{pmatrix} \text{ with } M_i\in\partial G_i(x), i=1,2\}\end{aligned}$$

(b) If $G_i : X\to Y$, $i=1,2$, are ∂G_i-semismooth at x then the sum

$$G := G_1 + G_2 : X \to Y$$

is ∂G-semismooth at x with

$$\begin{aligned}\partial G := \partial G_1 + \partial G_2 : X &\rightrightarrows \mathcal{L}(X,Y)\\ \partial G(x) &= \{M_1+M_2 : M_i\in\partial G_i(x),\ i=1,2\}\end{aligned}$$

(c) Let $G_1 : Y\to Z$ and $G_2 : X\to Y$ be ∂G_i-semismooth at $G_2(x)$ and x, respectively. Assume that ∂G_1 is bounded near $y=G_2(x)$ and that G_2 is Lipschitz continuous near x. Then $G=G_1\circ G_2$ is ∂G-semismooth with

$$\partial G(x) = \{M_1M_2 : M_1\in\partial G_1(G_2(x)),\ M_2\in\partial G_2(x)\}.$$

Proof

(a) By definition of ∂G we know that for every $M\in\partial G(x+d)$ there exist $M_1\in\partial G_1(x+d)$ and $M_2\in\partial G_2(x+d)$ with $Mx=\begin{pmatrix}M_1x\\ M_2x\end{pmatrix}$. Hence, using the norm $\|y\|_Y := \|y_1\|_{Y_1} + \|y_2\|_{Y_2}$ and exploiting the ∂G_i-semismoothness of G_i we obtain

$$\begin{aligned}&\sup_{M\in\partial G(x+d)}\|G(x+d)-G(x)-Md\|_Y = \sum_{i=1}^{2}\sup_{M_i\in\partial G_i(x+d)}\|G_i(x+d)-G_i(x)-M_id\|_y\\ &= o(\|d\|_X)+o(\|d\|_X) = o(\|d\|_X) \text{ as } \|d\|_X\to 0.\end{aligned}$$

(b) By the ∂G_i-semismoothness of G_i, $i=1,2$, and the triangular inequality we have

$$\begin{aligned}&\sup_{M\in\partial G(x+d)}\|G(x+d)-G(x)-Md\|_Y \le \sum_{i=1}^{2}\sup_{M_i\in\partial G_i(x+d)}\|G_i(x+d)-G_i(x)-M_id\|_Y\\ &= o(\|d\|_X) \text{ as } \|d\|_X\to 0.\end{aligned}$$

(c) Let $y=G_2(x)$ and consider $d\in X$. Let $h(d)=G_2(x+d)-y$. Then for $\|d\|_X$ sufficiently small, by the Lipschitz-continuity of G_2 we have

$$\|h(d)\|_Y = \|G_2(x+d)-G_2(x)\|_Y \le L^2|d\|_X,$$

so $\|h(d)\|_Y = \mathcal{O}(\|d\|_X)$. Furthermore, for $M_1\in\partial G_1(G_2(x+d))$ and $M_2\in\partial G_2(x+d)$, we obtain

$$\begin{aligned}&\|G_1(G_2(x+d))-G_1(G_2(x))-M_1M_2d\|_Z\\ &= \|G1(y+h(d))-G_1(y)-M_1h(d)+\underbrace{M_1(G_2(x+d)-G_2(x))}_{M_1h(d)}-M_1M_2d\|_Z\\ &= \|G1(y+h(d))-G_1(y)-M_1h(d)+M_1(G_2(x+d)-G_2(x)-M_2d)\|_Z\\ &\le \|G1(y+h(d))-G_1(y)-M_1h(d)\|_Z + \|M_1\|_{Y\to Z}\|G_2(x+d)-G_2(x)-M_2d\|_Y.\end{aligned}$$

By assumption, there exists a constant C with $\|M_1\|_{Y\to Z} \leq C$ if $\|d\|_X$ is sufficiently small. Taking the supremum with respect to M and using the ∂G_i-semismoothness of G_i, $i = 1, 2$, gives

$$\begin{aligned}
&\sup_{M\in\partial G(x+d)} \|G(x+d) - G(x) - Md\|_Z \\
&= \sup_{M\in\partial G(x+d)} \|G_1(G_2(x+d)) - G_1(G_2(x)) - M_1M_2d\|_Z \\
&\leq \sup_{M_1\in\partial G_1(y+h(d))} \|G_1(y+h(d)) - G_1(y) - M_1h(d)\|_Z \\
&+C \sup_{M_2\in\partial G_2(x+d)} \|G_2(x+d) - G_2(x) - M_2d\|_Y \\
&= o(\underbrace{\|h(d)\|_Y}_{\mathcal{O}(\|d\|_X)} + C\, o(\|d\|_X) = o(\|d\|_X).
\end{aligned}$$

□

Remark 3.6. Without any difficulty, the results of Theorem 3.5 can be extended to the case of α-order semismoothness.

As we have seen in Section 1.3, optimization problems of the form

$$\min_{u\in L^2(\Omega)} f(u) \qquad \text{s.t.} \quad a \leq u \leq b \quad \text{a.e. on } \Omega \tag{3.6}$$

where $\Omega \subset \mathbb{R}^n$ is a bounded measurable domain, $f : L^2(\Omega) \to \mathbb{R}$ is twice continuously F-differentiable and a, $b \in L^\infty(\Omega)$ with $a \leq b$ a. e. in Ω, can be reformulated in terms of a non-smooth operator equation of the form

$$\Phi(u) := u - P_S(u - \theta\nabla f(u)) = 0 \tag{3.7}$$

where $S = \{v \in L_2(\Omega) : a \leq v \leq b \textit{ a.e. in } \Omega\}$ is the set of admissible solutions, $\theta > 0$ is arbitrary, but fixed, and P_S denotes the L_2-projection onto S, which is given by

$$P_S(v)(x) = P_{[\beta_l,\beta_r]}(v(x)), \quad x \in \Omega.$$

with $P_{[\beta_l,\beta_r]}$ as in Example 1.
Note that, since P_S coincides with the pointwise projection onto $[\beta_l, \beta_r]$, we have

$$\Phi(u)(x) = u(x) - P_{[\beta_l,\beta_r]}(u(x) - \theta\nabla f(u)(x)).$$

Our aim now is to define a generalized differential $\partial\Phi$ for Φ in such a way that Φ is $\partial\Phi$-semismooth.
By the chain rule and sum rule we developed in Theorem 3.5, this reduces to the question how a suitable differential for the superposition $P_{[\beta_l,\beta_r]}(v(\cdot))$ can be defined.

For that purpose, we can use a result proved by M. Ulbrich stating the $\partial\Psi_G$-semismoothness of superposition operators Ψ_G of the form $\psi(G(\cdot))$ where G is a continuously F-differentiable operator, see [5].
First, we need to define this generalized differential $\partial\Psi_G$:

Definition 3.7. *Let $\psi : \mathbb{R}^m \to \mathbb{R}$ be Lipschitz continuous and ($\partial^{cl}\psi$)-semismooth. Furthermore, let $1 \leq q \leq p \leq \infty$ be given, consider*

$$\Psi_G : U \to L^q(\Omega), \quad \Psi_G(u)(x) = \psi(G(u)(x)),$$

where $G : U \to L^p(\Omega)^m$ is continuously F-differentiable and U is a Banach space. We define the differential

$$\begin{array}{rcl} \partial\Psi_G : U & \rightrightarrows & \mathcal{L}(U, L^q(\Omega)), \\ \partial\Psi_G(u) & = & \{M : Mv = g^T(G'(u)v), g \in L^\infty(\Omega)^m, \\ & & g(x) \in \partial^{cl}\psi(G(u)(x)) \ \dot{\forall} x \in \Omega\}. \end{array} \tag{3.8}$$

Remark 3.8. This is just the differential that we would obtain by the construction in part (c) of Theorem 3.5.

Now we can state the following semismoothness result:

Theorem 3.9. *Let $\Omega \subset \mathbb{R}^n$ be measurable with $0 < |\Omega| < \infty$. Furthermore, let $\psi : \mathbb{R}^m \to \mathbb{R}$ be Lipschitz continuous and $\partial^{cl}\psi$-semismooth. Let U be a Banach space, $1 \leq q < p \leq \infty$, and assume that the operator $G : U \to L^q(\Omega)^m$ is continuously F-differentiable and that G maps U locally Lipschitz continuously to $L^p(\Omega)^m$. Then, the operator*

$$\Psi_G : U \to L^q(\Omega), \quad \Psi_G(u)(x) = \psi(G(u)(x)),$$

is $\partial\Psi_G$-semismooth, where $\partial\Psi_G$ is defined in (3.8).

A proof can be found in [5].

In order to be able to apply this result to the second summand of our problem (3.7),

$$\begin{array}{rcl} \Psi_G : U = L^2(\Omega) & \to & L^2(\Omega), \\ \Psi_G(u)(x) & = & \psi(G(u)(x)) = P_{[\beta_l, \beta_r]}((u - \theta\nabla f(u))(x)), \end{array}$$

we have to make some assumptions on the structure of ∇f (which are fulfilled by many optimal control problems):
There exist $\alpha > 0$ and $p > 2$ sucht that

- $\nabla f(u) = \alpha u + H(u)$,
- $H : L^2(\Omega) \to L^2(\Omega)$ continuously F-differentiable,
- $H : L^2(\Omega) \to L^p(\Omega)$ locally Lipschitz continuous.

Under these assumptions and by setting the arbitrary parameter θ to $\frac{1}{\alpha}$, Ψ_G reduces to

$$\Psi_G(u)(x) = \psi(G(u)(x)) = P_{[\beta_l, \beta_r]}(-\frac{1}{\alpha}H(u)(x)),$$

so setting $q = 2$, $\psi = P_{[\beta_l, \beta_r]}$ and $G = -\frac{1}{\alpha}H$, we can apply *Theorem 3.9* and obtain that the operator Ψ_G is $\partial\Psi_G$-semismooth with $\partial\Psi_G$ defined in (3.8).

Therefore, by *Theorem 3.5* (a), the operator $\Phi = I - \Psi_G$ in our problem (3.7) is semismooth w.r.t $\partial\Phi = I - \partial\Psi_G$ defined by

$$\begin{aligned} \partial\Phi : L^2(\Omega) &\rightrightarrows \mathcal{L}(L^2(\Omega), L^2(\Omega)), \\ \partial\Phi(u) &= \{M : M = I + \frac{g}{\alpha} \cdot H'(u), g \in L^\infty(\Omega), \\ &\quad g(x) \in \partial^{cl} P_{[\beta_l,\beta_r]}(-(1/\alpha)H(u)(x)) \ \dot\forall x \in \Omega\}. \end{aligned} \tag{3.9}$$

Remark 3.10. As already mentioned in Remark 1.4, the gradient of the reduced optimal control problem we obtained in 1.3 is of the form $\alpha u + H(u)$ with

$$H(u) = B^*(A^{-1})^*(A^{-1}(r + Bu) - y_d).$$

Now it becomes clear why we required $B \in \mathcal{L}(L^{p'}, H^{-1}(\Omega))$ with $p' < 2$: To ensure that H is a linear mapping from $L^2(\Omega)$ to $(L^{p'}(\Omega))^* = L^p(\Omega)$ with $p = \frac{p'}{p'-1} > 2$.

In the next chapter, we will make use of the results we have established so far to provide an algorithm capable of solving optimization problems of the form (3.6).

Chapter 4

Exact Semismooth Newton Methods

The semismoothness concept ensures the approximation property required for the convergence of generalized Newton methods. In addition, we need the regularity condition (2.7) to be satisfied in each iteration of Algorithm 1. Therefore, we require:
There exist constants $C > 0$ and $\delta > 0$ such that

$$\|M^{-1}\|_{Y \to X} \leq C \quad \forall M \in \partial G(x) \quad \forall x \in X, \ \|x - \bar{x}\|_X < \delta. \tag{4.1}$$

Under these two assumptions, the following generalized Newton method for semismooth operator equations is q-superlinearly convergent:

Algorithm 2. (Exact Semismooth Newton method)

0. Choose $x^0 \in X$ (sufficiently close to the solution $\bar{x}$)

For k=0, 1, 2, ...:

1. Choose $M_k \in \partial G(x^k)$.
2. Obtain s^k by solving
$$M_k s = -G(x^k), \tag{4.2}$$
and set $x^{k+1} = x^k + s^k$.

The local convergence result is a simple corollary of Theorem 2.1:

Theorem 4.1. *Let $G : X \to Y$ be continuous and ∂G-semismooth at a solution $\bar{x}$ of (2.1). Furthermore, assume that the regularity condition (4.1) holds. Then there exists $\delta > 0$ such that for all $x^0 \in X$, $\|x^0 - \bar{x}\|_X < \delta$, the semismooth Newton method (Algorithm 2) converges q-superlinearly to $\bar{x}$.*
If G is ∂G-semismooth of order $\alpha > 0$ at $\bar{x}$, then the convergence is of order $1 + \alpha$.

Proof The regularity condition (4.1) implies (2.7) as long as x^k is close enough to $\bar{x}$. Furthermore, the semismoothness of G at $\bar{x}$ ensures the q-superlinear approximation condition (2.8).
In the case of α-order semismoothness, the approximation condition (2.9) with order $1 + \alpha$ holds.
Therefore, Theorem 2.1 yields the assertions. □

Since our problem (3.6) is equivalent to (3.7) and the operator Φ is $\partial\Phi$-semismooth with $\partial\Phi$ defined in (3.9), requiring the regularity condition

$$\exists C > 0 \; \exists \delta > 0 : \|M^{-1}\|_{L^2(\Omega)\to L^2(\Omega)} \leq C \quad \forall M \in \partial\Phi(u) \quad \forall u \in L^2(\Omega), \|u - \bar{u}\|_{L^2(\Omega)} < \delta, \tag{4.3}$$

and provided that we are given an initial value u^0 sufficiently close to the solution $\bar{u}$, we can apply the superlinearly convergent semismooth Newton method (*Algorithm 2)* and obtain an approximate solution u.

Chapter 5

Inexact Semismooth Newton Methods

So far, we have established an iterative algorithm where in each iteration step, we have to solve an operator equation of the form $M_k s^k = -G(x^k)$. In practical applications this operator M_k can be very complex and - even after discretization - computing an exact solution can be very expensive and may not be justified when the current iterate x^k is far from a solution. This difficulty motivates the notion of the inexact semismooth Newton method. The main idea behind this method is that an inexact solution of the Newton system (2.2) can be interpreted as a solution of the same system, but with M_k replaced by a perturbed operator $\tilde{M}_k$. Since the condition (2.5) (or the conditions (2.7) and (2.8)) remain valid if M_k is replaced by a perturbed operator $\tilde{M}_k$ and the perturbation is sufficiently small, we see that the fast convergence of the generalized Newton method is not affected if the system is solved inexactly and the accuracy of the solution is controlled suitably.
So, the idea of the inexact Newton method is not to solve the equation

$$M_k s^k = -G(x^k)$$

but to find some $\eta_k \in [0, \eta_{max})$ and $s^k \in X$ satisfying

$$\|M_k s^k + G(x^k)\|_Y \leq \eta_k \|G(x^k)\|_Y.$$

In practical applications, after discretization the Newton system (2.2) is usually a system of linear equations. Since its dimension depends on the complexity of the continuous problem and of course on the fineness of the discretization, the system which has to be solved in each iteration can become very large, so that it is recommendable to use iterative solvers to compute an approximate solution for the Newton correction s^k in each step. Therefore, the inexact semismooth Newton method is practically more relevant than its exact counterpart. The algorithm is still q-superlinearly convergent and looks as follows:

Algorithm 3. (Inexact Semismooth Newton method)

0. Choose $x^0 \in X$ (sufficiently close to the solution $\bar{x}$), select $\eta_{max} \in [0, 1)$

For k=0, 1, 2, ...:

1. Choose $M_k \in \partial G(x^k)$.

2. Find some $\eta_k \in [0, \eta_{max})$ and $s^k \in X$ that satisfy

$$\|M_k s^k + G(x^k)\|_Y \leq \eta_k \|G(x^k)\|_Y., \tag{5.1}$$

   and set $x^{k+1} = x^k + s^k$.

Since finding a proper initial value x^0 is a non-trivial matter, these algorithms are often equipped with some globalization techniques such as the general inexact-Newton trust region methods or some line search methods like the Armijo-Goldstein rule. For more information on those techniques as well as for a detailed convergence analysis of Algorithm 3 we refer to [6] and references therein.

Chapter 6

Application to a Model Optimal Control Problem

Now we are able to illustrate the method we have developed in theory in a concrete one-dimensional example. Consider the following elliptic optimal control model problem with a heat conduction problem as PDE-constraint:

$$\begin{aligned} \underset{y\in V_0, u\in L^2(0,1)}{min} \quad & J(y,u) := \tfrac{1}{2}\|y-y_d\|^2_{L^2(0,1)} + \tfrac{\alpha}{2}\|u\|^2_{L^2(0,1)} \\ & \text{s.t.} \; -(\lambda(x)y'(x))' = u(x) \\ & y(0) = y(1) = 0 \\ & \text{and } \beta_l \le u \le \beta_r \text{ almost everywhere in } (0,1), \end{aligned} \tag{6.1}$$

where $V_0 = H_0^1(0,1) = \{y \in H^1(0,1) : y(0) = y(1) = 0\}$, and $y_d \in L^2(0,1)$ and λ are piecewise constant functions,

$$\lambda(x) = \begin{cases} \lambda_1 & \text{for } 0 < x < a \vee b < x < 1 \\ \lambda_2 & \text{for } a < x < b \end{cases}$$

and

$$y_d(x) = \begin{cases} 0 & \text{for } 0 < x < a \vee b < x < 1 \\ 5 & \text{for } a < x < b, \end{cases}$$

where $0 < a < b < 1$. For the choice $\Omega_c = \Omega = (0,1)$, $A \in \mathcal{L}(V_0, V_0^*)$ given by the bilinear form

$$\langle Ay, z\rangle := a(y,z) := \int_0^1 \lambda(x)\nabla y(x)\nabla z(x)dx = \int_0^1 \lambda(x)y'(x)z'(x)dx,$$

$r = 0 \in V_0^*$ and B the natural embedding operator I of $L^2(0,1)$ into V_0^*, this problem is of the form (1.1).

We interpret this problem as a one-dimensional analogue of the hyperthermia treatment of prostate cancer in the following way:
The domain $\Omega = (0,1)$ represents the prostate of the patient, the subinterval (a,b) corresponds to the cancerous area. In this medical therapy, so-called “thermoseeds” made of

ferromagnetic material are placed into the prostate and an externally applied electromagnetic field induces heat. The control variable u corresponds to the heat source, the state variable y stands for the temperature distribution. The control is defined on the whole interval $(0,1)$, so we are facing a distributed control problem. The ultimate goal is to destroy the cancerous cells and at the same time not to harm the healthy tissue outside the interval (a,b), so the aim is to drive the temperature distribution y as close as possible to the prescribed desired distribution y_d, which has a jump of 5 centigrades at the transition to the cancerous area. The function λ represents the heat conduction coefficient, for which we assume two different values λ_1 and λ_2 outside and inside the interval (a,b). Owing to limited heating and cooling capacities, we also have to bound the heat source u form below and from above by β_l and β_r. Furthermore, for simplicity we assume homogeneous Dirichlet boundary conditions for y on $\Gamma = \Gamma_1 = \{0,1\}$.
Summarizing, the optimization problem consists in determining the distribution of the heat u that has to be induced in the domain $(0,1)$ such that the temperature distribution is optimized, so that the thermoseeds can be placed correspondingly.

In Section 1.3 we have seen that determining u from an optimal control problem like (6.1) is equivalent to solving the non-smooth operator equation in (1.7), i.e.

$$\Phi(u) := u - P_{U_{ad}}(u - \theta\nabla\hat{J}(u)) = 0$$

where

$$U_{ad} = \{u \in L^2(0,1) : \beta_l \le u(x) \le \beta_r \;\dot{\forall} x \in (0,1)\}$$

and

$$\hat{J}(u) := J(y(u),u) := \frac{1}{2}\|y(u) - y_d\|^2_{L^2(0,1)} + \frac{\alpha}{2}\|u\|^2_{L^2(0,1)}.$$

Here, $y(u)$ denotes the solution of the PDE-constraint for fixed u, formally

$$y(u) = A^{-1}(r + Bu) = A^{-1}Bu.$$

For the F-derivative $\nabla\hat{J}$ we obtain

$$\begin{aligned}\nabla\hat{J}(u) &= \alpha u + y'(u)^*(y(u) - y_d) \\ &= \alpha u + B^*(A^{-1})^*(A^{-1}(r + Bu) - y_d) =: \alpha u + H(u).\end{aligned} \tag{6.2}$$

Since $B \in \mathcal{L}(L^{p'}(\Omega), H^{-1}(\Omega))$, we have $B^* \in \mathcal{L}(H^1_0(\Omega), L^p(\Omega))$ with $p = p'/(p'-1) > 2$. Hence the affine linear operator $H(u)$ is a continuous affine linear mapping $L^2(\Omega) \to L^p(\Omega)$ and we can apply *Theorem 3.9* like in Chapter 3 and obtain that Φ is $\partial\Phi$-semismooth with $\partial\Phi$ defined in (3.9). Since the regularity condition (2.7) is satisfied, we can use the (exact) semismooth Newton method (Algorithm 2) to compute a solution of (6.1). Assuming that we have a proper initial value u^0, the Newton system in each step of the algorithm reads

$$M_k s^k = -\Phi(u^k) \tag{6.3}$$

where

$$M_k := I + \frac{1}{\alpha}g^k \cdot H'(u^k) = I + \frac{1}{\alpha}g^k \cdot B^*(A^{-1})^*A^{-1}B$$

and $g \cdot H'(u)$ stands for $v \mapsto g \cdot (H'(u)v)$ and $g^k \in L^\infty(\Omega)$ is chosen such that

$$g^k(x) = \begin{cases} = 0 & -(1/\alpha)H(u^k)(x) \notin [\beta_l, \beta_r], \\ = 1 & -(1/\alpha)H(u^k)(x) \in (\beta_l, \beta_r), \\ \in [0,1] & -(1/\alpha)H(u^k)(x) \in \{\beta_l, \beta_r\}. \end{cases}$$

For solving (6.3), we note that s^k solves (6.3) if and only if $s^k = d_u^k$ and $(d_y^k, d_u^k, d_\mu^k)^T$ solves

$$\begin{pmatrix} I & 0 & A^* \\ 0 & I & -\frac{1}{\alpha} g^k \cdot B^* \\ A & -B & 0 \end{pmatrix} \begin{pmatrix} d_y^k \\ d_u^k \\ d_\mu^k \end{pmatrix} = \begin{pmatrix} 0 \\ -\Phi(u^k) \\ 0 \end{pmatrix}$$

which can be easily seen.

For solving this system in each iteration step we use the Finite Element Method (FEM). We assume that all involved functions (i.e. d_y^k, d_u^k, d_μ^k and u^k) are in our working space V_0 and define the finite dimensional subspace

$$V_{0h} := \operatorname{span}\{\phi_i\}_{i=1}^{n_h - 1} = \left\{ v_h = \sum_{i=1}^{n_h-1} v_i \phi_i \,\middle|\, v_i \in \mathbb{R} \right\}$$

where the i-th nodal basis function ϕ_i is a piecewise linear function defined by

$$\phi_i(x_j) = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \quad \forall\, i, j = 0, \ldots, n_h$$

with $x_0, \ldots, x_{n_h}$ being the nodes of our mesh.
For the Finite Elment discretization we multiply the whole system by a test function $v \in V_0$, integrate over the whole domain $(0,1)$ and perform integration by parts in the principal terms. We restrict our V_0-functions to the finite dimensional subspace V_{0h} (i.e. $d_y^k \to d_{yh}^k$, $d_u^k \to d_{uh}^k$, $d_\mu^k \to d_{\mu h}^k$, $u^k \to {u^k}_h$ and $v \to v_h$) and end up with the variational system

$$\begin{aligned} &(d_{yh}^k, v_h)_{L^2(0,1)} & & +a(d_{\mu h}^k, v_h) &&= 0 \\ & & (d_{uh}^k, v_h)_{L^2(0,1)} & -\frac{1}{\alpha}(g^k \cdot d_{\mu h}^k, v_h)_{L^2(0,1)} &&= \langle \Phi({u^k}_h), v_h \rangle \\ &a(d_{yh}^k, v_h) & -(d_{uh}^k, v_h)_{L^2(0,1)} & &&= 0 \end{aligned}$$

with the bilinear forms

$$a(u,v) = \int_0^1 \lambda(x) u'(x) v'(x) \mathrm{d}x \qquad \text{and} \qquad (u,v)_{L^2(0,1)} = \int_0^1 u(x) v(x) \mathrm{d}x.$$

and the linear form

$$\langle F, v \rangle = \int_0^1 F(x) v(x) \mathrm{d}x.$$

Expressing each of the V_{0h}-functions as a linear combination of the basis functions ϕ_i, $i = 1, \dots, n_h - 1$ and exploiting the bilinearity of $a(\cdot,\cdot)$ and $(\cdot,\cdot)_{L^2(0,1)}$, we arrive at the following system of linear equations:

$$\begin{pmatrix} M & 0 & K^T \\ 0 & M & gM^k \\ K & -M & 0 \end{pmatrix} \begin{pmatrix} \underline{y^k} \\ \underline{u^k} \\ \underline{\mu^k} \end{pmatrix} = \begin{pmatrix} 0 \\ -\underline{\Phi^k} \\ 0 \end{pmatrix} \tag{6.4}$$

Here, $K = (K_{ij})_{i,j=1,\dots n_h-1}$ with $K_{ij} = a(\phi_j, \phi_i)$ denotes the stiffness matrix, $M = (M_{ij})_{i,j=1,\dots n_h-1}$ with $M_{ij} = (\phi_j, \phi_i)_{L^2(0,1)}$ denotes the mass matrix, $gM^k = (gM^k_{ij})_{i,j=1,\dots n_h-1}$ is defined by $gM^k_{ij} = \int_0^1 -\frac{1}{\alpha} g^k(x)\phi_j(x)\phi_i(x)\mathrm{d}x$ and the vector $(0, -\underline{\Phi^k}, 0)^T$ denotes the load vector with $\underline{\Phi^k} = \big(\langle \Phi(u^k), \phi_i \rangle\big)_{i=1,\dots,n_h-1}$. As we can see, for setting up the system only the matrix gM^k and the load vector $\underline{\Phi^k}$ have to be re-calculated in each step of the iteration.
The vectors $\underline{y^k} := \underline{d^k_{yh}}$, $\underline{u^k} := \underline{d^k_{uh}}$ and $\underline{\mu^k} := \underline{d^k_{\mu h}}$ are the coefficient vectors in the basis representation of d^k_{yh}, d^k_{uh} and $d^k_{\mu h}$, respectively. Due to the special choice of the basis, all the occurring matrices have a tridiagonal shape.

Note: We identify a finite element function $w_h(x) = \sum_{i=0}^{n_h} w_i \phi_i(x)$ by the vector of coefficients $\underline{w_h} = (w_i)_{i=0}^{n_h}$.

For solving this system of linear equations we perform a block-Gaussian elimination and finally use the primal Schur complement:
By subtracting KM^{-1} times the first block-row from the third one, expressing $\underline{\mu^k}$ from the new third block-row and plugging in into the second block-row, we arrive at the system

$$\Big(M - gM^k\, K^{-1}\, M\, K^{-1}\, M\Big)\, \underline{u^k} = -\underline{\Phi^k}. \tag{6.5}$$

Since we are only interested in the correction d^k_u of the control u, we do not have to make any further computations.

Remark 6.1. For solving this linear system it is more convenient to use an iterative method, such as the Conjugate Gradient (CG) method (cf. [7]). The advantage is that we do not need to solve nor to put up system (6.5) explicitly, but we only have to provide a matrix-vector multiplication for which we can exploit the tridiagonal shape of the stiffness matrix K.
Note: Since the linear system is solved inexactly in each Newton iteration, we are actually facing the inexact semismooth Newton method.

The last difficulty is the non-trivial question how to find a proper starting value u^0 for our Newton iteration. To avoid this problem, we start with an arbitrary initial value and perform several steps of the *dashed Newton method* (cf. [8]), i.e. we compute the Newton correction $s^k = d^k_u$ as described above, but set $u^{k+1} = u^k + \tau_k s^k$ where we choose $\tau = \tau_k \in (0, 1]$ such that the inequality

$$\|\Phi(u^{k+1})\|^2 = \|\Phi(u^k + \tau s^k)\|^2 < \|\Phi(u^k)\|^2$$

is satisfied as sharp as possible. As soon as the residual $\|\Phi(u^k)\|^2$ is small enough, we set the dashing parameter τ_k to 1 and return to the semismooth Newton method which - as we have seen in Chapters 4 and 5 - converges q-superlinearly to the solution $\bar{u}$. So, by this procedure, we have achieved globalization of the semismooth Newton method.

Bibliography

[1] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE Constraints*, chapter 2. Springer, 2009.

[2] P. Deuflhard. Hyperthermia simulation and treatment planning. `http://www.zib.de/visual/projects/hyperthermia/hyperthermia`.

[3] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen: Theorie, Verfahren und Anwendungen.* Vieweg, Wiesbaden, 2005.

[4] J. Heinonen. Lectures on lipschitz analysis. `http://www.math.jyu.fi/research/reports/rep100.pdf`.

[5] M. Ulbrich. Nonsmooth newton-like methods for variational inequalities and constrained optimization problems in function spaces. `http://www-m1.ma.tum.de/foswiki/pub/Lehrstuhl/PublikationenUlbrich/Habil.pdf`, 2001. Habilitation Thesis, TU München.

[6] J. Chen and L. Qi. Globally and superlinearly convergent inexact newton-krylov algorithms for solving nonsmooth equations. *NLAA*, 17:155–174, 2010.

[7] W. Zulehner. *Numerische Mathematik*, chapter 8.3. 2008.

[8] U. Langer and M. Jung. *Methode der finiten Elemente für Ingenieure*, chapter 6. Teubner, 2. edition, 2010.