Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

ANTON SCHIELA

An Interior Point Method in Function Space for the Efficient Solution of State Constrained Optimal Control Problems [1]

[1] Supported by the DFG Research Center MATHEON "Mathematics for key technologies"

ZIB-Report 07-44 (March 2008)

An Interior Point Method in Function Space for the Efficient Solution of State Constrained Optimal Control Problems †

Anton Schiela

March 14, 2008

Abstract

We propose and analyse an interior point path-following method in function space for state constrained optimal control. Our emphasis is on proving convergence in function space and on constructing a practical path-following algorithm. In particular, the introduction of a pointwise damping step leads to a very efficient method, as verified by numerical experiments.
AMS MSC 2000: 90C51, 49M05

Keywords: interior point methods in function space, optimal control, state constraints

1 Introduction

The construction and analysis of efficient algorithms for state constrained optimal control problems is still a considerable challenge. Presently, most popular methods that admit a (partial) analysis in function space are path-following methods, such as exterior penalty methods [7], Lavrentiev regularization [9, 10] and interior point methods [13, 14, 15]. Except for [13] and partially [10] (for a fixed Lavrentiev parameter) the available results are restricted to properties of the *homotopy path*, such as its existence, convergence and continuity. Except for these two works, not much is known about convergence of the associated *path-following algorithms*. This includes the important question if it is at all possible to follow the homotopy path by a practical algorithm, or if the sequence of iterates may stagnate far away from the desired solution. Closely connected and even more relevant from a practical point of view is the question how to choose homotopy parameters to obtain a fast and robust algorithm. These questions can certainly not be answered by an analysis of the path alone.

The aim of this paper is to propose and analyse an interior point method in function space that is capable of solving state constrained optimal control problems efficiently. The corresponding homotopy path has been analysed in [14, 15], so

†Supported by the DFG Research Center Matheon "Mathematics for key technologies"

our emphasis here is on the Newton path-following method and on giving positive answers to the above questions. We establish qualitative convergence results in the following sense. Under suitable conditions there is a sequence of homotopy parameters μ_k that converges to 0 and a sequence of corresponding iterates x_k produced by a Newton corrector scheme that converges to the solution of the original problem x_*. The quantities used in the analysis, which yields convergence of the scheme as an a-priori result, can be modelled and estimated inside a numerical algorithm to yield a criterion for controlling the path-following algorithm efficiently. This is done in the spirit of [4, Chapter 5], but modified in a way that fits into our particular setting in function space.

To establish a rigorous analysis we essentially need estimates for two quantities. The first one, which reflects the most basic analytic properties of the homotopy path, is its local Lipschitz constant $\eta(\mu)$. The second captures the nonlinearity of the equations that define the homotopy path. This quantity, which governs the local convergence behaviour of Newton's method and in particular its radius of convergence, is an affine covariant Lipschitz constant for the Jacobian, denoted by $\omega(\mu)$. Since good a-posteriori estimates are available for η and ω, their role is not a purely analytic one, but they establish a close connection between a-priori theory and algorithmic implementation. In some sense, the algorithm is driven by an a-posteriori counterpart of the convergence theory established in this work.

Compared to [13] we introduce, as an algorithmic modification, a pointwise damping step, which prevents Newton's method from leaving the feasible domain and enhances the efficiency of the path-following scheme significantly. It is motivated by the idea to exploit the pointwise structure of the problem and has several useful interpretations. In our numerical experiments we observe that this modification allows the solution of state constrained optimal control problems in a few Newton steps.

Acknowledgement. The author wishes to thank Dr. Martin Weiser for helpful discussions and the close cooperation during the development of the computational framework.

2 A Class of State Constrained Optimal Control Problems

Let Ω be an open and smoothly bounded domain in $\mathbb{R}^d$, $d = 1 \dots 3$ and $\overline{\Omega}$ its closure. Let Y denote the space of states and U the space of controls. Define $Z := Y \times U$ with $z := (y, u)$ and consider the following convex minimization problem, the details of which are fixed in the remaining section.

$$\begin{aligned} \min_{z \in Z} J(z) \quad \text{s.t. } & Ay - Bu = 0 \\ & \underline{y} \le y. \end{aligned} \tag{1}$$

We set $Y = C(\overline{\Omega})$, and $U = L_2(Q)$ for a measurable set Q equipped with an appropriate norm. This setting includes optimal control problems subject to linear elliptic partial differential equations with distributed control ($Q = \Omega$), boundary control ($Q = \partial\Omega$) and finite dimensional control ($Q = \{1, \ldots, n\}$, equipped with the counting measure).

We will now specify our abstract theoretical framework, which holds throughout this work and collect a couple of basic results about this class of problems. Our framework is placed in the context of convex analysis, whose fundamentals can e.g. be looked up in [5].

Convex Functionals. For simplicity, let J be a quadratic tracking type functional with Tychonov regularization term:

$$J(z) = \frac{1}{2} \|y - y_d\|^2_{L_2(\Omega)} + \frac{\alpha}{2} \|u\|^2_{L_2(\Omega)}$$

Obviously, this functional is strictly convex and continuous in Z, and hence subdifferentiable. Its subdifferential is single valued and given by

$$Z^* \ni \partial J(z) = \begin{pmatrix} y - y_d \\ \alpha u \end{pmatrix}.$$

Equality Constraints. The equality constraint $Ay - Bu = 0$ is introduced to model a partial differential equation.

Let R be a reflexive Banach space and $B : U \to R$ be continuous. We assume that $A : Y \supset \operatorname{dom} A \to R$ is a linear operator, which is *densely defined, closed* that maps $\operatorname{dom} A$ to R *bijectively.*

In the context of optimal control R is often the dual of a Sobolev space and the operator B is usually defined as the adjoint of an embedding or a trace operator (cf. e.g. the discussion in [8] or [15]).

We consider A as a model of a differential operator, which may be unbounded. This depends of course on the choice of topology in Y. Closed, densely defined operators between Banach spaces are a classical concept of functional analysis. They generalize the concept of continuous operators and retain much of their structure. In particular, there is an open-mapping theorem, a closed range theorem, and adjoint operators are well defined. In this work and in [14, 15] only these basic properties of A are needed for a successful analysis. A classical introduction to unbounded operators is [6], but most elementary facts can also be found in standard textbooks on functional analysis.

There is a simple correspondence between a bijective closed operator and its inverse.

Lemma 2.1. *For Banach spaces Y and R let $A : Y \supset \operatorname{dom} A \to R$ be a linear operator. A is closed and bijective if and only if A possesses a continuous inverse $A^{-1} : R \to \operatorname{dom} A \subset Y$ in the sense that $A^{-1}A = id_{\operatorname{dom} A}$ and $AA^{-1} = id_R$.*

Proof. Assume first that a continuous inverse A^{-1} exists. Then in particular A is bijective. Let $y_k \to y$ and $r_k = Ay_k \to r$. By surjectivity of A there is $\tilde{y} \in \operatorname{dom} A$: $A\tilde{y} = r$, hence $Ay_k \to A\tilde{y}$. We have to show $y = \tilde{y}$. Because A^{-1} is continuous, we conclude $y_k = A^{-1}Ay_k \to A^{-1}A\tilde{y} = \tilde{y}$, hence $y = \tilde{y}$.

If in converse A is closed and bijective, then existence of a continuous inverse follows from the open mapping theorem (cf. e.g. [17, Satz IV.4.4]), which not only holds for continuous, but also for closed operators. □

Hence, if the partial differential equation defined by $Ay = f$ is uniquely solvable and admits an a-priori estimate, then A is closed and bijective. Hence, our assumption of closedness of A holds, if the solution operator A^{-1} maps R into $\operatorname{dom} A \subset C(\overline{\Omega})$ continuously.

We define E as the subspace of $Y \times U$ of all pairs (y, u) that satisfy $Ay - Bu = 0$. It is closed, because it is the kernel of the closed operator $(A, -B)$. Because A^{-1} is continuous we can eliminate $y = A^{-1}Bu$. Then if follows from reflexivity of U that E is weakly sequentially compact.

We exploit density of $\operatorname{dom} A$ in Y to define an adjoint operator A^* by the following standard construction. Here and in the following we denote by $\langle \cdot, \cdot \rangle$ the dual pairing. For every $l \in R^*$ the mapping $y \to \langle l, Ay \rangle$ is a linear functional on $\operatorname{dom} A$. We define $\operatorname{dom} A^*$ as the subspace of all $l \in R^*$ for which $y \to \langle l, Ay \rangle$ is continuous on $\operatorname{dom} A$ and can thus by density be extended uniquely to a continuous functional on Y. Hence, for each $l \in \operatorname{dom} A^*$ there is a unique linear functional $A^*l \in Y^*$ for which

$$\langle l, Ay \rangle = \langle A^*l, y \rangle \qquad \forall\, y \in \operatorname{dom} A. \tag{2}$$

This yields the definition of $A^* : R^* \supset \operatorname{dom} A^* \to Y^*$.

Because R is reflexive, $\operatorname{dom} A^*$ is dense in R^*. This is due to [6, Theorem II.2.14].

Inequality Constraints. We assume that $\underline{y} \in C^{1,1}(\overline{\Omega})$, which means that its spacial derivatives are Lipschitz continuous. The inequality constraints in (1) are interpreted to hold pointwise almost everywhere and they define a closed set $G \subset Y$. We assume that there is a *strictly feasible point* $\breve{z} = (\breve{y}, \breve{u})$ that satisfies $A\breve{y} - B\breve{u} = 0$ and

$$0 < d_{\min} := \operatorname*{ess\,inf}_{t \in \overline{\Omega}} \{\breve{y}(t) - \underline{y}(t)\}. \tag{3}$$

We call such a condition a (uniform pointwise) *Slater condition* and $\breve{z}$ a Slater point. This condition together with the topology of Y defined by $\|\cdot\|_\infty$ is used in the analysis of dual variables and subdifferentials and in the derivation of first order optimality conditions.

Combined Functionals. It is a popular strategy in convex analysis to combine the functional J and the constraints $z \in E$ and $z \in G$ to a single functional. This is done via indicator functions. The *indicator function* $\chi_C(z)$ of a set $C \subset Z$ is defined by

$$\chi_C(z) := \begin{cases} 0 & : \quad z \in C \\ \infty & : \quad \text{otherwise.} \end{cases}$$

If C is non-empty, convex, and closed, then χ_C is a proper ($\chi_C \not\equiv +\infty$), convex, and lower semi-continuous function. In particular, χ_E and χ_G enjoy these properties.

With the help of indicator functions we can rewrite (1) as an unconstrained minimization problem defined by the following functional:

$$\begin{aligned} &F : Z \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\} \\ &F := J + \chi_E + \chi_G. \end{aligned} \tag{4}$$

By our assumptions F is a proper, lower semi-continuous, strictly convex, and coercive functional and does thus admit a unique minimizer by weak compactness of E (cf. e.g. [5][Proposition II.1.2])

3 Barrier Regularizations for State Constraints

Let us recapitulate known results about the regularization of state constrained optimal control problems with barrier functions. The proofs for the following results can be found in [14, 15].

Definition 3.1. For all $q \geq 1$ and $\mu > 0$ the functions $l(z; \mu; q) : \mathbb{R}_+ \to \overline{\mathbb{R}}$ defined by

$$l(z; \mu; q) := \begin{cases} -\mu \ln(z) & : \quad q = 1 \\ \dfrac{\mu^q}{(q-1)z^{q-1}} & : \quad q > 1 \end{cases}$$

are called *barrier functions of order* q. We extend their domain of definition to $\mathbb{R}$ by setting $l(z; \mu; q) = \infty$ for $z \leq 0$. We include finite sums of these barrier functions, and define their order to be the maximum order of the summands. We denote their derivatives, which are defined for $z > 0$ by l' and l'' .

Usually we do not have to consider special values of q or μ. In these cases we may abbreviate the notation $l(z; \mu; q)$ by $l(z; \mu)$ or even $l(z)$.

Using these barrier *functions* $l(z; \mu; q)$ we construct barrier *functionals* $b(y; \mu; q)$ to implement constraints of the form $y \geq 0$ on a compact set $B \subset \overline{\Omega}$ by computing the integral over l:

$$\begin{aligned} b(\cdot; \mu; q) &: C(B) \to \overline{\mathbb{R}} \\ y &\mapsto \int_B l(y(t); \mu; q)\, dt. \end{aligned}$$

It is easy to see that b is a well defined, extended real valued functional on $C(B)$. With these definitions we may regularize F in (4) by replacing χ_G with $b(y; \mu)$. Hence,

$$F_\mu(z) = J(z) + \chi_E(z) + b(y; \mu). \tag{5}$$

We denote by b' and b'' the *formal* derivatives of b. Here,

$$\langle b'(z;\mu;q), \delta z\rangle = \int_B l'(z(t);\mu;q)\delta z(t)\,dt, \tag{6}$$

if the right hand side it is well defined. An analogous definition holds for b''. We call these quantities *formal derivatives*, because in general they may not have the properties of a derivative, and it is not even clear, a-priori if (6) is well defined, because for given z, δz $l'(z;\mu;q)\delta z$ may not be an integrable function on B. However, we have the following result:

Theorem 3.2. *For all $\mu \geq 0$, problem* (5) *admits a unique solution* $(y(\mu), u(\mu))$. *Moreover, the system* (7)-(8)

$$0 = y(\mu) - y_d + A^*p + m \tag{7}$$
$$0 = \alpha \cdot u(\mu) - B^*p. \tag{8}$$

admits a unique solution $(p(\mu), m(\mu)) \in R^* \times M(\overline{\Omega})$ satisfying the following conditions.

The non-positive measure $m(\mu)$ can be represented as the sum of a formal derivative of a barrier functional (which is in particular well defined) and a non-positive measure $\tilde{m}(\mu)$:

$$\langle m(\mu), v\rangle = \int_\Omega l'(y;\mu)v\,dt + \int_{y=0} v\,d\tilde{m}(\mu). \tag{9}$$

If $\mu = 0$, then the first term vanishes. The second term vanishes for $y > \underline{y}$. Hence,

$$\int_\Omega y(\mu)\,d\tilde{m}(\mu) = 0. \tag{10}$$

Moreover, the set of all $m(\mu)$ is uniformly bounded in $M(\overline{\Omega})$ on every fixed interval $\mu \in [0;\mu_0]$.

Proof. This is a special case of [15, Theorems 2.6, 2.7]. □

Our next assertion captures the analytic properties of our of solutions. It holds for general $m(\mu)$, i.e., also if $\tilde{m}(\mu) \neq 0$.

Theorem 3.3. *The set of solutions of* (5) *forms a path that converges to the solution of* (1) *with the error estimates*

$$J(y(\mu)) - J(y_*) \leq C\mu \tag{11}$$
$$\|y(\mu) - y_*\|_Y + \|u(\mu) - u_*\|_{L_2} \leq c\sqrt{\mu}. \tag{12}$$

This path is locally Lipschitz continuous for each $\mu > 0$, and satisfies

$$\|y(\mu) - y(\nu)\|_Y + \|u(\mu) - u(\nu)\|_{L_2} \leq c\mu^{-1/2}|\mu - \nu|. \tag{13}$$

If $\mu \geq \nu \geq \mu/2$, then

$$\left\|\sqrt{b''(y(\mu))}(y(\mu) - y(\nu))\right\|_{L_2} \leq c\mu^{-1/2}|\mu - \nu|. \tag{14}$$

Proof. Equations (12) and (13) follow from [14, Theorem 5.3, 5.5], and [15, Theorem 2.8]. A close look at the proof of [14, Theorem 5.5], in particular equation (37) there shows (14). □

As usual in interior point methods we will call this homotopy path of solutions the *central path.*

Proposition 3.4. *If $y(\mu)$ is strictly feasible, then points $(y(\mu), p(\mu)) \in Y \times R^*$ on the central path are characterized by being solutions of the following system of equations in $Y^* \times R$*

$$\begin{aligned} y - y_d + b'(y;\mu) + A^*p &= 0 \\ Ay - BB^*\alpha^{-1}p &= 0. \end{aligned} \tag{15}$$

Proof. We may use (8) to compute $u = \alpha^{-1}B^*p$, and insert this into the state equation, which yields the second row of (15). The first row of (15) follows from (7) and the assumed feasibility of $y(\mu)$, which implies $m = b'(y;\mu)$. □

The system of equations (15) will be in the center of our considerations. Our path-following method is based on solving this system approximately by Newton's method. For an analysis in function space it is therefore necessary to guarantee strict feasibility of $y(\mu)$.

Strict feasibility. Because the barrier subgradients $m(\mu)$ are uniformly bounded in $M(\overline{\Omega})$, we conclude uniform boundedness of $p(\mu)$ in R^* via (7) and thus also of $u(\mu)$ via (8). The space R^* depends on the state equation considered in the application and is often a Sobolev space $W^{1,t}(\Omega)$. An example for an elliptic PDE is analysed in [3]. This increased regularity of $u(\mu)$ yields in turn better regularity of $y(\mu)$, say in some Sobolev space $W^{s,p}$, together with a uniform norm-bound. In view of the Sobolev-embedding theorems this motivates our following assumption:

Assumption 3.5. Assume that the set $y(\mu)$ is uniformly bounded on some positive interval $(0;\mu_0]$ in $C^\beta(\overline{\Omega})$ for some $0 < \beta \leq 2$. Here $C^\beta(\overline{\Omega})$ denotes the spaces of Hölder continuous functions for $\beta < 1$, and of differentiable functions with Hölder-continuous derivatives for $1 < \beta < 2$.

Proposition 3.6. *If Assumption 3.5 holds, then there is a positive integer q, and a function $\psi(\mu)$, depending on q and β, which is strictly positive on every compact positive interval $[\underline{\mu};\mu_0]$ and monotonically decreasing such that the following assertion holds:*

If $y(\mu)$ is a point on the central path, induced by a barrier function of order q, then

$$\inf_{t\in\overline{\Omega}} y(\mu)(t) - \underline{y} \geq \psi(\mu). \tag{16}$$

Proof. This follows from [14, Lemma 6.1] and the following discussion there. □

Hence, by an appropriate choice of q we can force the central path solutions to be strictly feasible, approaching the bounds for $\mu \to 0$ in a controlled fashion.

4 A Simple Newton Path-Following Method

Our aim in this section is to prove Theorem 4.9, a qualitative convergence result for a class of Newton path-following methods in function space, applied to our state constrained optimal control problem.

Our analysis is based on two quantities, which describe the behaviour of our Newton path-following scheme. As for the structure of the central path, we use its local Lipschitz constant $\eta(\mu)$. Equation (13) provides us with the fairly good estimate $\eta(\mu) = O(\mu^{-1/2})$. To be able to capture the behaviour of the Newton corrector, we give estimates for the Newton contraction $\Theta(x;\mu)$, defined below. It is hard to obtain sharp bounds for Θ, and we will content ourselves with a rough quantitative estimate. Refinements are conceivable, but highly technical. Note in this context that for state constrained problems even qualitative results in function space are very sparse in the literature. In compensation we describe in Section 6 a method to estimate the quantities η and Θ locally in order to drive an adaptive path-following algorithm.

Our prototype is Algorithm 4.1. We will show in this section that a choice μ_k is possible, such that Algorithm 4.1 is well defined and converges to the optimal solution of the problem. In Section 6 we describe how to choose the sequence μ_k in practice, based on a-posteriori quantities.

Algorithm 4.1.
select $\mu_0 > 0$, and x_0 with y_0 sufficiently close to $y(\mu_0)$
for $k = 0, \dots$
  $x_{k+1} := x_k - F'(x_k, \mu_k)^{-1} F(x_k, \mu_k)$
  select μ_{k+1}

Let in the following $X := Y \times R^*$. Reflexivity of R yields $X^* = Y^* \times R$. $F(x;\mu)$ is given by (15). Throughout this section we assume that Assumption 3.5 holds, and choose q sufficiently large, such that (16) holds.

Let us define a domain of definition for F. There are essentially two requirements on $x = (y, p)$. First, $Ay \in R$ and $A^*p \in Y^*$ have to be well defined. Thus we require $y \in \operatorname{dom} A$, $p \in \operatorname{dom} A^*$. Second, we need some feasibility condition for y, which we address by the following restriction, using Assumption 3.5. We define a (strictly feasible) neighbourhood Y_μ of the central path by

$$Y_\mu = \rho\psi(\mu) \cdot B_{L_\infty}(y(\mu)), \tag{17}$$

for some fixed $0 < \rho < 1$.

Now set the domain of definition of F to $D_\mu := (\operatorname{dom} A \cap Y_\mu) \times \operatorname{dom} A^* \subset X$. Then the following mapping is well defined:

$$F(x;\mu) : X \supset D_\mu \to X^*.$$

For our analysis we will choose the sequence μ_k such that all iterates remain in Y_μ. In Y_μ the relation $c(y - \underline{y})(t) \le (y(\mu) - \underline{y})(t) \le C(y - \underline{y})(t)$ holds, which we will use often in the following. This helps us to derive *a-priori* estimates for $\Theta(x;\mu)$. In a practical algorithm, where *a-posteriori* estimates are available, this neighbourhood can be dropped.

4.1 Analysis of the Newton corrector

The formal linearization of this system at a point $x \in D_\mu$ reads

$$F'(x;\mu) := \begin{pmatrix} I + b''(y;\mu) & A^* \\ A & -\alpha^{-1}BB^* \end{pmatrix}.$$

This is a formal linearization, because we do not specify in which sense $F'(x;\mu)$ is a derivative of $F(x;\mu)$. However, we will show that Newton's method is locally quadratically convergent, if this formal linearization is used in the role of the Jacobian matrix.

Observe that in contrast to $F(x;\mu)$, $F'(x;\mu)$ is defined for all $x \in Y_\mu \times U$, not only in D_μ and does not depend on p. For fixed x we have

$$F'(x;\mu)(\cdot) : X \supset \operatorname{dom} A \times \operatorname{dom} A^* \to X^*.$$

Moreover, because A and A^* are *linear*:

$$\begin{aligned} \left\|(F'(x;\mu) - F'(\tilde{x};\mu))\delta x\right\|_{Y^* \times R} &= \left\|(b''(y) - b''(\tilde{y}))\delta y\right\|_{Y^*} \\ &\leq \left\|(b''(y) - b''(\tilde{y}))\right\|_{Y^*} \|\delta y\|_Y . \end{aligned} \tag{18}$$

Because $l''(y;\mu)$ is uniformly continuous in Y_μ we conclude that $F'(x;\mu)$ depends continuously on x (w.r.t. the operator norm) in this region.

We introduce the following local scaled norm for corrections δy:

$$\|\delta y\|_{x,\mu} := \left\|\alpha^{1/2}\delta y\right\|_Y + \left\|\sqrt{1 + l''(y;\mu)}\delta y\right\|_{L_2(\Omega)} . \tag{19}$$

Let us first establish the solvability of the linear system $F'(x;\mu)\delta x = r$, which reads in detail:

$$\begin{pmatrix} I + b''(y;\mu) & A^* \\ A & -\alpha^{-1}BB^* \end{pmatrix} \begin{pmatrix} \delta y \\ \delta p \end{pmatrix} = \begin{pmatrix} r_a \\ r_s \end{pmatrix}. \tag{20}$$

For our analysis it will be sufficient to consider the case $r_s = 0$.

Theorem 4.2. *For $y \in Y_\mu$ and $r_a \in Y^*$, $r_s = 0$ the system* (20) *admits a unique solution $(\delta y, \delta p) \in X$, with $\delta y \in \operatorname{dom} A$ and $\delta p \in \operatorname{dom} A^*$. The following estimate holds:*

$$\|\delta y\|_{x,\mu} + \alpha^{-1/2} \|B^*\delta p\|_{L_2(\Omega)} \leq C \sup_{v \in Y} \frac{\langle r_a, v\rangle}{\|v\|_{x,\mu}}. \tag{21}$$

Here C is independent of x and μ.

Proof. Consider the quadratic minimization problem:

$$\min_{(\tilde{y},\tilde{u})\in Y\times U} q(\tilde{y},\tilde{u}) := \frac{1}{2}\left\|\sqrt{1+l''}\tilde{y}\right\|_{L_2}^2 + \frac{\alpha}{2}\|\tilde{u}\|_U^2 - \langle r_a, \tilde{y}\rangle \quad \text{s.t. } A\tilde{y} - B\tilde{u} = 0,$$

which can be written as $\min(q+\chi_E)(\tilde{y},\tilde{u})$. By our assumptions this problem has a unique solution $(\delta y, \delta u) \in Y \times U$, and because $q(0) = 0$, we have $q(\delta y, \delta u) \le 0$. Hence, we conclude

$$\left\|\sqrt{1+l''}\delta y\right\|_{L_2}^2 + \alpha\|\delta u\|_U^2 \le 2|\langle r_a, \delta y\rangle|$$

and thus, dividing by the square-root of the left hand side, using $\|\delta y\|_Y \le C\|\delta u\|_U$ (which holds, because $A\delta y - B\delta u = 0$) we obtain

$$\|\delta y\|_{x,\mu} + \alpha^{1/2}\|\delta u\|_U \le C\frac{\langle r_a, \delta y\rangle}{\|\delta y\|_{x,\mu}} \le C\sup_{v\in Y}\frac{\langle r_a, v\rangle}{\|v\|_{x,\mu}}. \tag{22}$$

It remains to connect this estimate to (20). Since q is continuous on $Y \times U$, we can apply the sum-rule of convex analysis [5, Thm. II.5.6] to derive optimality conditions for our minimization problem. We obtain $0 \in \partial q + \partial\chi_E$. Since q is Gâteaux differentiable, $\partial q(\delta z) = \{((I + b'')\delta y - r_a, \alpha u)\}$. Using our assumptions on A and B, [15, Proposition 2.5] yields $\partial\chi_E = \mathrm{ran}(A, -B)^*$ and thus existence of $\delta p \in \mathrm{dom}\, A^*$, which satisfies the equations

$$\begin{aligned}(I + b'')\delta y + A^*\delta p &= r_a\\ \alpha\delta u - B^*\delta p &= 0.\end{aligned}$$

The first row is identical to the first row of (20). The second row yields, $\delta u = \alpha^{-1}B^*\delta p$. Inserting this into the equation $A\delta y - B\delta u = 0$ yields the second row of (20), inserting it into (22) yields (21). □

Remark 4.3. Observe that the inverse Jacobian possesses a strong smoothing property. In particular, $\|\delta y\|_Y \le \|r_a\|_{Y^*}$. This is possible, because the corresponding system of equations is a system of *partial differential equations* only. Such a smoothing property is important for the robustness of function space oriented methods. The drawback of this primal formulation is that the nonlinearity of the barrier terms is high.

The most popular interior point methods in finite dimensions are primal-dual methods, which introduce additional *algebraic* equations. Then the resulting system is only relatively mildly nonlinear. However, the presence of purely algebraic equations spoils the smoothing property of the inverse Jacobian. In Section 5 we propose an algorithmic variant that retains the smoothing property of the Jacobian, but similarly to primal-dual methods alleviates the nonlinearity of the barrier terms.

Since in Y_μ all scaled norms $\|\delta y\|_{x,\mu}$ are equivalent up to a constant, we introduce the following scaled norm for simplicity:

$$\|\delta x\|_\mu := \|\delta y\|_{x(\mu),\mu} + \left\|\alpha^{-1/2} B^* \delta p\right\|_{L_2(\Omega)} .$$

Its scaling is fixed for fixed μ. If μ is decreased by a factor $0 < \sigma < 1$, and $y(\sigma\mu) \in Y_\mu$, then $\|\cdot\|_\mu$ and $\|\cdot\|_{\sigma\mu}$ are equivalent up to a constant $C_N(\sigma)$, which tends to 1, as $\sigma \to 1$.

Next we capture the local behaviour of Newton's method. For our analysis it is sufficient to consider one single step of it.

Theorem 4.4. *Let $x(\mu)$ be the solution of the nonlinear equation $F(x;\mu) = 0$. The Newton mapping*

$$\begin{aligned} N : X \supset D_\mu &\to X \\ x &\mapsto x_+ := x - F'(x;\mu)^{-1} F(x;\mu) \end{aligned}$$

defined by a Newton step yields $x_+ \in \operatorname{dom} A \times \operatorname{dom} A^*$.

Moreover, N extends uniquely and continuously to $Y_\mu \times R^*$. For this extended mapping still $x_+ = N(x) \in \operatorname{dom} A^* \times \operatorname{dom} A$ holds.

Define

$$\Theta(x;\mu) := \frac{\left\|F'(x;\mu)^{-1}\big(F'(x;\mu)(x - x(\mu)) - (F(x;\mu) - F(x(\mu);\mu))\big)\right\|_\mu}{\|x - x(\mu)\|_\mu}. \tag{23}$$

Then the following contraction estimate holds

$$\|x_+ - x(\mu)\|_\mu = \Theta(x;\mu)\,\|x - x(\mu)\|_\mu . \tag{24}$$

Proof. Because $F(x;\mu)$ is well defined on D_μ and by Theorem 4.2, the Newton correction $\delta x := F'(x;\mu)^{-1}F(x;\mu)$ is well defined, $\delta x \in \operatorname{dom} A^* \times \operatorname{dom} A$, and hence $x_+ = x - \delta x$, too.

Using $F(x(\mu);\mu) = 0$ and (23) we have

$$\begin{aligned} \|x_+ - x(\mu)\|_\mu &= \left\|x - x(\mu) - F'(x;\mu)^{-1}F(x;\mu)\right\|_\mu \\ &= \left\|F'(x;\mu)^{-1}\big(F'(x;\mu)(x - x(\mu)) - (F(x;\mu) - F(x(\mu);\mu))\big)\right\|_\mu \\ &= \Theta(x;\mu)\,\|x - x(\mu)\|_\mu , \end{aligned}$$

which yields (24) for all $x \in D_\mu$.

Let us now extend our results from D_μ to $Y_\mu \times R^*$. We know that for $x \in D_\mu$

$$\begin{aligned} x_+ - x(\mu) &= F'(x;\mu)^{-1}\big(F'(x;\mu)(x - x(\mu)) - (F(x;\mu) - F(x(\mu);\mu))\big) \\ &= F'(x;\mu)^{-1}\begin{pmatrix} b''(y;\mu)(y - y(\mu)) - (b'(y;\mu) - b'(y(\mu);\mu)) \\ 0 \end{pmatrix} \\ &= F'(x;\mu)^{-1}\begin{pmatrix} r_a(y) \\ 0 \end{pmatrix}. \end{aligned}$$

The last expression is not only well defined for $x \in D_\mu$, but for all $x \in Y_\mu \times R^*$. Moreover, $r_a(y) \in L_\infty \subset Y^*$. Hence,

$$\tilde{N}(x) := F'(x;\mu)^{-1} \begin{pmatrix} r_a(y) \\ 0 \end{pmatrix} + x(\mu) \in \operatorname{dom} A^* \times \operatorname{dom} A$$

is well defined for all $x \in Y_\mu \times R^*$ and coincides with $N(x)$ on D_μ which is dense in $Y_\mu \times R^*$, because $\operatorname{dom} A$ is dense in Y and $\operatorname{dom} A^*$ is dense in R^*. It remains to show that $\tilde{N}$ depends continuously on x, which implies that it is the unique continuous extension of N.

This is not hard, because $r_a(y)$ depends continuously on y, and $F'(x;\mu)^{-1}$ depends continuously on x by the following argument (which is known as an operator perturbation lemma).

First of all the identity

$$F'(\tilde{x};\mu)^{-1} = (I - F'(x;\mu)^{-1}(F'(x;\mu) - F'(\tilde{x};\mu)))^{-1} F'(x;\mu)^{-1}$$

holds. By continuity of $F'(x;\mu)$ with respect to x due to (18) and by invertibility of $F'(x;\mu)$ we have $T(x;\tilde{x})(\cdot) := F'(x;\mu)^{-1}(F'(x;\mu) - F'(\tilde{x};\mu))(\cdot) \in L(X)$ and $\|T(x;\tilde{x})\| \to 0$ for $\tilde{x} \to x$. This implies via construction of the Neumann series that $(I - T(x;\tilde{x}))^{-1} \to I$ for $\tilde{x} \to x$, and hence continuity of $F'(x;\mu)^{-1}$ at x. Now

$$\begin{aligned} \left\| \tilde{N}(x) - \tilde{N}(\tilde{x}) \right\|_\mu \le & \left\| F'(x;\mu)^{-1} \right\| \|r_a(y) - r_a(\tilde{y})\|_{Y^*} \\ & + \left\| F'(x;\mu)^{-1} - F'(\tilde{x};\mu)^{-1} \right\| \|r_a(\tilde{y})\|_{Y^*} \end{aligned}$$

and continuity follows by (18). □

Remark 4.5. Theorem 4.4 states that Newton steps can be canonically defined on all of $Y_\mu \times R^*$ by unique continuous extension. This result is useful in Section 5, where a pointwise modification is introduced, which is an L_∞ perturbation. Theorem 4.4 asserts that this perturbation does not interfere with the well definedness of Newton steps.

Equation (24) gives us the interpretation of

$$\Theta(x;\mu) = \frac{\|x_+ - x(\mu)\|_\mu}{\|x - x(\mu)\|_\mu}$$

as a local Newton contraction. If $\Theta(x;\mu) \le k < 1$ in a neighbourhood of $x(\mu)$, then Newton's method converges. We will show now $\Theta(x;\mu) = O(\|x - x(\mu)\|_\mu)$, which implies local quadratic convergence of Newton's method for each $\mu > 0$. Proving local quadratic convergence alone, however, would not be sufficient in the context of path-following. In addition we need a more quantitative result of the form (26) that relates $\Theta(x;\mu)$ to μ and that yields bounds from below on the *radius of convergence* to conclude convergence of the overall path-following method.

Lemma 4.6. *Let $y_1, y_2 > \underline{y}$, and $\tilde{y} := \min\{y_1; y_2\}$. For the barrier function $l(y) = l(y;\mu;q)$ the following pointwise estimate holds:*

$$|l''(y_1)(y_1 - y_2) - (l'(y_1) - l'(y_2))| \leq \frac{c}{\tilde{y} - \underline{y}} |l''(\tilde{y})(y_1 - y_2)^2|. \tag{25}$$

The constant c depends only on q.

Proof. Since l' is a sum of functions of the form $\mu^q y^{-q}$, it is twice differentiable for positive y and all derivatives are monotonically decreasing in absolute value. Hence, application of the fundamental theorem of calculus twice yields (25), taking into account the rules of differentiation. □

Proposition 4.7. *For each $\mu > 0$ Newton's method converges locally quadratically to the solution $x(\mu)$. More precisely, there is a positive function $\omega(\mu)$, which is bounded on every compact positive interval, such that for $x \in Y_\mu \times R^*$*

$$\Theta(x;\mu) \leq \frac{1}{2}\omega(\mu)\|x - x(\mu)\|_\mu, \tag{26}$$

together with the bound $\omega(\mu) \leq c\psi^{-1}(\mu)$.

Proof. By definition of Θ we have, just as in the proof of Theorem 4.4

$$\Theta(x;\mu) \leq \frac{\|\delta x\|_\mu}{\|x - x(\mu)\|_\mu} := \frac{\left\| F'(x;\mu)^{-1} \begin{pmatrix} r_a(y) \\ 0 \end{pmatrix} \right\|_\mu}{\|x - x(\mu)\|_\mu},$$

with $r_a(y) = b''(y)(y - y(\mu)) - (b'(y) - b'(y(\mu)))$. Because $y \in Y_\mu$, and thus

$$C(y(\mu) - \underline{y}) \leq y - \underline{y} \leq c(y(\mu) - \underline{y}),$$

Lemma 4.6 gives us the pointwise estimate (dropping the argument $t \in \Omega$):

$$|r_a(y)| \leq \frac{c}{\tilde{y} - \underline{y}} \left| l''(\tilde{y})(y - y(\mu))^2 \right| \leq \frac{c}{y(\mu) - \underline{y}} \left| \sqrt{l''(y(\mu))}(y - y(\mu)) \right|^2.$$

Let $\|v\|_Y = \|v\|_\infty = 1$ be arbitrary. Then by the Hölder inequality

$$\begin{aligned}
|\langle r_a, v\rangle| &\leq \int_\Omega \frac{c}{y(\mu) - \underline{y}} |\sqrt{l''(y(\mu))}(y - y(\mu))|^2 |v|\, dt \\
&\leq \left\|c(y(\mu) - \underline{y})^{-1}\right\|_\infty \|y - y(\mu)\|_{x,\mu}^2 \|v\|_{L_\infty} \\
&\leq c\psi(\mu)^{-1} \|x - x(\mu)\|_\mu^2.
\end{aligned}$$

Hence, Theorem 4.2 yields

$$\|\delta x\|_\mu \leq c\psi(\mu)^{-1} \|x - x(\mu)\|_\mu^2,$$

which implies (26). □

By the fundamental theorem of calculus the quantity $\omega(\mu)$ can be interpreted as an *affine invariant Lipschitz constant* for $F'(x;\mu)$. It is a measure for the nonlinearity of the problem at hand. Its prominent role for the analysis and control of algorithms based on Newton's method has been pointed out in [4]. The crucial fact is that good computational a-posteriori estimates are available for Θ and thus for ω, as will be elaborated in Section 6.

4.2 Convergence of the Path-Following Method

In the following lemma we connect the continuity properties of the central path and the convergence properties of the Newton corrector to obtain a convergent path-following method. It turns out that all we need is an estimate for the Lipschitz constant $\eta(\mu)$ of the central path, and an estimate for the quantity $\omega(\mu)$, defined in (26), which governs the radius of convergence of Newton's method. With these two quantities we can show, using mostly algebraic arguments, that there is a sequence μ_k such that Algorithm 4.1 produces a sequence of iterates that remains in a prescribed neighbourhood of the central path and converges to the solution of the original problem. With this method it is principally possible but rather technical to compute a rate of convergence. In the context of control constraints this has been done in [12].

Lemma 4.8. *Let the functions $\omega(\mu), \eta(\mu)$ be majorants for the quantities introduced above. Assume further that on each interval $[\underline{\mu}, \mu_0] \subset]0, \mu_0]$, $\eta(\mu)$ and $\omega(\mu)$ are bounded from above, and $r(\mu)$ is a positive function, bounded from below.*

If the initial value x_0 satisfies the inequality

$$\|x_0 - x(\mu_0)\|_{\mu_k} \le \min\left\{\omega(\mu_0)^{-1}, r(\mu_0)\right\}, \tag{27}$$

then we can choose a sequence σ_k with $0 < \sigma_k = \sigma(\mu_k) < 1$ depending only on μ_k and the functions η, ω, r such that Algorithm 4.1 produces iterates that remain inside $r(\mu_k)B(x(\mu_k))$ for each k and

$$\|x_{k+1} - x(\mu_k)\|_{\mu_k} \le \frac{1}{2}\|x_k - x(\mu_k)\|_{\mu_k}, \tag{28}$$

$$\|x_k - x(\mu_k)\|_{\mu_k} \le \min\left\{\omega(\mu_k)^{-1}, r(\mu_k)\right\}. \tag{29}$$

Moreover,

$$\lim_{k\to\infty} \mu_k = 0, \qquad \lim_{k\to\infty} \|x_k - x_*\|_{\mu_k} = 0.$$

Proof. Assume w.l.o.g. that ω, η, r are continuous and thus uniformly continuous in each interval $[\underline{\mu}, \mu_0]$, and that $r(\mu)$ tends to 0 for $\mu \to 0$. Otherwise, we can easily construct majorants of ω, η and a positive minorant for r on $]0, \mu_0]$ having these properties.

We perform a proof by induction in k. Assume that $x_k \in r(\mu_k)B(x(\mu_k))$ and

$$\|x_k - x(\mu_k)\|_{\mu_k} \le \omega(\mu_k)^{-1},$$

which holds by (27) for $k = 0$. Then by Theorem 4.4 one Newton step yields

$$\|x_{k+1} - x(\mu_k)\|_{\mu_k} \leq \frac{1}{2}\omega(\mu_k)^{-1}.$$

Reduction of μ_k via a factor $\sigma = \sigma(\mu_k)$, setting $\mu_{k+1} := \sigma\mu_k$ gives us (recall that $C_N(\sigma)$ describes the equivalence between the norms $\|\cdot\|_\mu$ and $\|\cdot\|_{\sigma\mu}$):

$$\begin{aligned}
\|x_{k+1} - x(\mu_{k+1})\|_{\mu_{k+1}} &\leq C_N(\sigma)\,\|x_{k+1} - x(\mu_{k+1})\|_{\mu_k} && (30)\\
&\leq C_N(\sigma)(\|x_{k+1} - x(\mu_k)\|_{\mu_k} + \|x(\mu_k) - x(\mu_{k+1})\|_{\mu_k}) && (31)\\
&\leq C_N(\sigma)\left(\frac{1}{2}\cdot\omega(\mu_k)^{-1} + \eta(\mu_k)(\mu_k - \mu_{k+1})\right).
\end{aligned}$$

To complete the induction we have to achieve

$$\|x_{k+1} - x(\mu_{k+1})\|_{\mu_{k+1}} \leq \min\{\omega(\mu_{k+1})^{-1}, r(\mu_{k+1})\}.$$

Thus, we have to choose $\sigma < 1$ such that simultaneously

$$C_N(\sigma)\eta(\mu_k)(\mu_k - \mu_{k+1}) \leq r(\mu_{k+1}) - C_N(\sigma)\frac{1}{2}r(\mu_k) \tag{32}$$

$$C_N(\sigma)\eta(\mu_k)(\mu_k - \mu_{k+1}) \leq \omega(\mu_{k+1})^{-1} - C_N(\sigma)\frac{1}{2}\omega(\mu_k)^{-1}. \tag{33}$$

By our boundedness and continuity assumptions on η, ω, and r, and because $C_N(\sigma) \to 1$ for $\sigma \to 1$ this is obviously possible since $\eta(\mu_k) < \infty$, $\lim_{\sigma\to 1} r(\sigma\mu_k) = r(\mu_k)$, and $\lim_{\sigma\to 1}\omega(\sigma\mu_k)^{-1} = \omega(\mu_k)^{-1}$. Moreover, it is easy to verify that by our boundedness and uniform continuity assumptions, for each $\underline{\mu}$ there is a $\sigma_{\min}(\underline{\mu}) < 1$ such that $\sigma_k < \sigma_{\min}(\underline{\mu})$ for all $\mu_k \geq \underline{\mu}$. Thus, $\mu_k \leq \underline{\mu}$ after finitely many steps, which implies $\mu_k \to 0$.

The convergence result $x_k \to x_*$ follows now by (29) and the assumed convergence of $x(\mu) \to x_*$. □

Theorem 4.9. *If Assumption 3.5 holds, then for sufficiently high order q there is a sequence $\mu_k \to 0$, and a sequence $x_k \to x_*$ in X, such that Algorithm 4.1 is well defined and $y_k \in Y_{\mu_k}$. Moreover, the following estimate holds:*

$$\|y_k - y_*\|_Y + \|u_k - u_*\|_U \leq C\sqrt{\mu_k}.$$

Proof. We have to verify the assumptions of Lemma 4.8. First, we choose

$$r(\mu) = \min\{\rho\psi(\mu); \sqrt{\mu}\}$$

as in (17), which guarantees $y_k \in Y_{\mu_k}$. Boundedness of $\eta(\mu)$ follows from (13) and (14) and boundedness of $\omega(\mu)$ was shown in Proposition 4.7. This yields existence of a sequence $\mu_k \to 0$. The error estimate for the iterates then follows from (12) and our choice for $r(\mu)$. □

Finite dimensional interior point methods can invoke equivalence of norms in $\mathbb{R}^n$ to prove linear convergence at a rate that quickly degenerates with increasing dimension. These are the so called complexity estimates.

5 A Pointwise Modification

Barrier methods rely on iterates that are feasible with respect to the inequality constraints. Since in the barrier context Newton's method approximates a rational function by a linear one, Newton steps tend to be too large in direction towards the constraints. So it is likely that iterates become infeasible. This issue should be addressed algorithmically. Otherwise, this may restrict the speed of convergence of practical algorithms.

In the following we propose a modification of Newton's method, which may be considered as a pointwise damping strategy. The idea exploits the pointwise structure of the problem and guarantees feasibility of the iterates.

Consider the first row of the Newton equation $F(x;\mu) + F'(x;\mu)\delta x = 0$. It is posed in Y^* and reads

$$y - y_d + b'(y) + A^*p + (I + b''(y))(y_+ - y) + A^*(p_+ - p) = 0. \tag{34}$$

Our principle idea is to construct a modified feasible iterate y_C that satisfies

$$y_C - y_d + b'(y_C) + A^*p_+ = 0. \tag{35}$$

It is not obvious at first sight that this idea is sensible, because A^*p_+ is not necessarily a function. In particular, in the context of finite elements and weak formulations (35) cannot be interpreted as a pointwise equation.

However, subtraction of (34) and (35) yields a pointwise equation for y_C that depends on y and y_+:

$$y_C - y + l'(y_C) - l'(y) = (1 + l''(y))(y_+ - y) \quad \text{almost everywhere in } \Omega. \tag{36}$$

If p_+ is sufficiently smooth, then (35) and (36) are equivalent. Hence, (36) extends (35) for general p_+. The idea is now to solve this equation pointwise, but to use only those y_C, for which $|y_C - y| \leq |y_+ - y|$. We obtain a pointwise damping step.

In the case of a logarithmic barrier function (36) is a quadratic equation in y_C and can be solved explicitely as such. For rational barrier functions we may use iterative techniques, of which bisection is the simplest. Because this computation is a pointwise operation to be performed at each node of the discretization, its contribution to the overall computational effort is marginal.

5.1 Interpretation as a Pointwise Damped Primal Correction

In the following lemma we gather the basic properties of our pointwise modification.

Lemma 5.1. *Let $y \in \mathbb{R}$ be strictly feasible. Then for every $y_+ \in \mathbb{R}$* (36) *admits a unique solution y_C, which is strictly feasible. If $y_+ \leq y$, then $y_+ \leq y_C \leq y$. Otherwise $y \leq y_+ \leq y_C$. Moreover,*

$$|y_C - y_+| \leq c(\mu)|y - y_C|^2. \tag{37}$$

Hence, for $y_+ \leq y$

$$|y_C - y_+| \leq c(\mu)|y - y_+|^2. \tag{38}$$

Proof. On $(\underline{y}, \infty)$ the function $f(y_C) := y_C + l'(y_C)$ is well defined, monotonically increasing, continuous, $\lim_{y_C \to \underline{y}} = -\infty$, and $\lim_{y_C \to \infty} = +\infty$. By the mean value theorem this implies unique solvability of the equation $f(y_C) = r$ for any $r \in \mathbb{R}$ with $y_C \in (\underline{y}, \infty)$. By (36) and the fundamental theorem of calculus:

$$(1 + l''(y))(y_+ - y) = f(y_C) - f(y) = \int_y^{y_c} (1 + l''(\eta))d\eta \, (y_C - y). \qquad (39)$$

Since l'' is monotonically decreasing in y, this relation yields $y_+ \le y_C \le y$ for $y_+ \le y$ and $y \le y_+ \le y_C$ otherwise. We may rewrite (39) as

$$y_C - y_+ = l''(y)(y_+ - y) - (l'(y_C) - l'(y)),$$

and hence

$$(y_C - y_+)(1 + l''(y)) = l''(y)(y_C - y) - (l'(y_C) - l'(y)),$$

which, via Lemma 4.6 implies (37). □

In order to obtain a *damping* strategy we use (36) only for the case $y_+ \le y$. In this case Lemma 5.1 asserts that $y_+ \to y_C$ is indeed a pointwise damping. Moreover, Theorem 4.9, which asserts local convergence of the undamped Newton corrector in L_∞ and (38) assert that the damped Newton corrector converges with the same properties as the undamped variant. In practice, however, the pointwise damped variant is far more efficient.

Theorem 5.2 (Convergence Theorem for Damping)**.** *The conclusions of Theorem 4.9 remain valid, if a damping step* (36) *is used.*

Proof. By Theorem 4.4 Newton steps are well defined for all $y \in Y_\mu$. Because of (38) the damping step is only a small perturbation of the undamped Newton step, and, after a possible reduction of step size the results of Theorem 4.9 carry over. □

5.2 Interpretation as a Blended Primal-Dual Correction.

The pointwise modification (35) has another useful interpretation in terms of a dual method. Let us introduce the variable v, defined by

$$v(t) = y(t) + l'(y(t); \mu).$$

We can solve this equation for y to obtain a nonlinear function $y(v)$ with derivative

$$y_v(v) = \left(I + b''(y(v); \mu)\right)^{-1},$$

and a nonlinear system of equations in the variables p and v:

$$\begin{aligned} v - y_d + A^*p &= 0 \\ Ay(v) - \alpha^{-1}BB^*p &= 0, \end{aligned}$$

which we will abbreviate by $\tilde{F}(\tilde{x};\mu)$, setting $\tilde{x} := (v,p)$. In contrast to F, which is only defined for $\underline{y} \le y$, $\tilde{F}$ is well defined for all sufficiently smooth $\tilde{x}$. Its (again formal) linearization is given by

$$
\begin{aligned}
\tilde{F}'(\tilde{x};\mu) &:= \begin{pmatrix} I & A^* \\ Ay_v(v) & -\alpha^{-1}BB^* \end{pmatrix} \\
&= \begin{pmatrix} I+b'' & A^* \\ A & -\alpha^{-1}BB^* \end{pmatrix} \begin{pmatrix} (I+b'')^{-1} & 0 \\ 0 & I \end{pmatrix}.
\end{aligned}
\tag{40}
$$

Although this formulation has the advantage of guaranteed feasibility, it suffers from the poor regularity properties of this formulation. The presence of the nonlinearity $Ay(v)$ in $\tilde{F}$ makes a pure dual method rather unstable.

If we consider the factorization of $\tilde{F}'(\tilde{x};\mu)^{-1}$ in (40), we see that performing one Newton step for $F(x;\mu)$ and one correction of the form (36) is equivalent to performing one Newton step for $\tilde{F}(\tilde{x};\mu)$ and computing $y(v)$.

Hence, by our damping strategy we implicitly compute both the primal and the dual Newton step and take the pointwise minimum in absolute value to obtain a blended correction that has the favourable properties of both methods and avoids their problems.

6 An Adaptive Path-Following Scheme

In this section we consider the construction of a practical path-following algorithm that adaptively chooses the sequence μ_k. Our starting point is again the system (15). Hence, our algorithm uses the state y and the adjoint state p as iteration variables.

For efficient path-following several extensions of Algorithm 4.1 are useful. Most importantly, we have to provide a practical criterion for the choice of the sequence μ_k. If this choice is made, then performing one single Newton step is a too rigid concept in practice. Rather, one should aim for some (loose) convergence criterion. If the choice of μ_k was too aggressive, it may be useful to reject the path-following step and select μ_k more conservatively, based on more accurate information. This leads to Algorithm 6.1.

Our considerations are based on the ideas of [4, Chapter 5]. The main idea is to introduce parameterized models for the quantities $\eta(\mu)$ and $\Theta(x;\mu)$ (and closely related $\omega(\mu)$) used in the a-priori analysis and supply computational estimates for the parameters. This strategy guarantees a close connection between a-priori results and the algorithmic realization and helps to take into account the special structure of the function space problem. In the following we will introduce these models. For additional details we refer to [12, Section 8.2].

Algorithm 6.1.
select $\mu_0 > 0$, and x_0 with $\|x_0 - x(\mu_0)\| \le r(\mu_0)$, $k := 0$
do *(Homotopy Method)*
    $\widetilde{x}_0 = x_k$, $j = 0$
    do *(Newton Corrector)*
        $\delta\tilde{x}_j \leftarrow$ compute pointwise damped Newton step
        (failure,converged) $\leftarrow$ estimate Newton contraction
        if(not failure) $\tilde{x}_{j+1} := \tilde{x}_j + \delta\tilde{x}_j$
        j := j+1
    while not(converged or failure)
    if(converged)
        $x_{k+1} := \tilde{x}_{j+1}$
        $\mu_{k+1} \leftarrow$ predict new step size
        k := k+1
    if(failure)
        $\mu_k \leftarrow$ reduce step size
while(termination criterion not reached)

For the evaluation of our algorithmic quantities we have to choose a norm. Our convergence theory and numerical experience suggest to use a scaled local norm, similar to $\|\cdot\|_{x,\mu}$ defined in (19). Experience shows that it is favourable in practice to drop the L_∞ part and use the following L_2-type norm:

$$\|\delta x\|^2 := \|\delta x\|^2_{x,\mu,2} = \left\|\sqrt{1 + b''(x;\mu)}\delta y\right\|^2_{L_2} + \alpha^{-1/2}\,\|B^*\delta p\|^2_U$$

It is of practical importance that this norm can be evaluated easily and accurately. In particular, we do not rely on the evaluation of norms of residuals. Corresponding residual norms would be dual norms, which are cumbersome and expensive to evaluate.

Following the ideas of [16] it is possible and useful to take into account algorithmically that our scaled norm depends on μ and x. We will, however, neglect this issue here for simplicity.

6.1 A Model of the Central Path

A direct way to estimate the Lipschitz constant η would be using finite differences:

$$[\eta]_{Exact}(\mu_k) := \frac{\|x(\mu_k) - x(\mu_{k+1})\|}{|\mu_k - \mu_{k+1}|}.$$

However, because the exact solutions of the central path are not available, we replace them by finite differences of our computational values:

$$[\eta](\mu_k) := \frac{\|x_k - x_{k+1}\|}{|\mu_k - \mu_{k+1}|}. \tag{41}$$

The closer x_k and $x(\mu_k)$, the more accurate the estimate $[\eta]$. With this estimate we may model $\eta(\mu)$ by

$$\eta_M(\mu) := [\eta](\mu_k)\left(\frac{\mu}{\mu_k}\right)^{-1/2}. \tag{42}$$

Once $[\eta](\mu)$ is computed we may also use it to estimate the remaining length of the central path via integration of $\eta_M(\mu)$:

$$\|x_{\mu_k} - x_*\| \approx 2[\eta](\mu_k)\cdot\mu_k. \tag{43}$$

6.2 Estimating the Newton Contraction

An important feature of Algorithm 6.1 is the evaluation of a Newton step that defines the corrector: $\tilde{x}_{j+1} \leftarrow \tilde{x}_j$. To capture the behaviour of Newton's method we derive a model for the contraction $\Theta(x;\mu)$. For this purpose we consider (23), which is of course computationally unavailable, because $x(\mu)$ is unknown. However, after one Newton step $x \to x_+$ we may replace $x(\mu)$ by x_+ in (23) and obtain

$$\begin{aligned}
[\Theta]^N(x;\mu) :&= \frac{\left\|F'(x;\mu)^{-1}\big(F'(x;\mu)(x-x_+) - (F(x;\mu)-F(x_+;\mu))\big)\right\|}{\|x-x_+\|}\\
&= \frac{\left\|(x - F'(x;\mu)^{-1}F(x;\mu)) - (x_+ - F'(x;\mu)^{-1}F(x_+;\mu))\right\|}{\|x-x_+\|}\\
&= \frac{\|x_+ - \overline{x}_+\|}{\|x-x_+\|},
\end{aligned}$$

where $\overline{x}_+$ is the result of a simplified Newton step performed at x_+.

If we use our pointwise damping strategy we have to design a modification of $[\Theta](x;\mu)$ in terms of x_C, because x_+ may be infeasible. The most obvious modification is to replace the result of a Newton step $x \to x_+$ by a pointwise damped Newton step $x \to x_C$, and the simplified Newton step at $x_+ \to \overline{x}_+$ by a pointwise damped simplified Newton step $x_C \to \overline{x}_C$:

$$[\Theta]^{PD}(x;\mu) := \frac{\|x_C - \overline{x}_C\|}{\|x - x_C\|}.$$

By (38) the pointwise damped Newton steps merge into ordinary Newton steps close to the solution, and the same holds for the simplified versions. So $[\Theta]^{PD}(x;\mu)$ is asymptotically equivalent to $[\Theta]^N(x;\mu)$ close to the solution.

If $[\Theta]$ is small enough, then the simplified Newton step needed for this evaluation can be used to improve the quality of the solution. If direct sparse solvers are used for the linear equations, then the additional computational effort needed is rather small. We only have to assemble another right hand side and perform one forward-backward substitution. In our implementation we even perform a second simplified Newton step in case of very small $[\Theta]$, because this improves the efficiency slightly.

If iterative solvers are used, then the relative additional effort for a simplified Newton step depends on the relation between assembly of the stiffness matrix, construction of a preconditioner and iterative solution. Also in this case the obtained simplified Newton step can be used to improve the solution, or as a starting value for the next iterative solution process.

Let us now concentrate on $[\Theta]^{PD}$. Since Newton's method is a *locally* convergent method it may happen that for bad initial values the iteration diverges. This happens, if $\Theta(x;\mu) > 1$. So it is natural to terminate the Newton corrector with failure, if $[\Theta]^{PD}(x;\mu) > 1$ in order to start a correction step with a more conservative choice of μ_k.

Next we derive a convergence criterion for the Newton corrector. Motivated by the triangle inequality $\|x - x_*\| \le \|x - x_C\| + \|x_C - x_*\|$ we may estimate the distance $r(x;\mu) := \|x - x(\mu)\|$ by setting

$$[r](x;\mu) := (1 + [\Theta]^{PD})\,\|x - x_C\|\,, \quad [r](x_C;\mu) := [\Theta]^{PD}(1 + [\Theta]^{PD})\,\|x - x_C\|\,.$$

If $[r](x_C;\mu)$ is sufficiently small, then the Newton corrector is terminated successfully. A useful convergence criterion for a Newton correction method is to require

$$[r](\tilde{x}_{j+1};\mu) \le \rho\,\|\tilde{x}_0 - \tilde{x}_{j+1}\|\,, \tag{44}$$

which means that the corrector has reduced the error about a factor ρ. Useful choices are in a range of 0.01 to 0.5.

6.3 Step Size Selection

Proposition 4.7 suggests to model the Newton contraction $\Theta(x;\mu)$ by

$$\Theta_M(x;\mu) := \omega_M(\mu)\|x - x(\mu)\|, \tag{45}$$

where we – similarly to η_M – define

$$\omega_M(\mu) := [\omega](\mu_k)\left(\frac{\mu}{\mu_k}\right)^{-1/2} \tag{46}$$

and, again replacing $x(\mu)$ by x_C:

$$[\omega](\mu_k) := \frac{[\Theta]^{PD}(x;\mu)}{\|x - x_C\|}.$$

Assume first that the corrector for μ_k has terminated successfully. By the triangle inequality we have

$$\|x_{k+1} - x(\mu_{k+1})\| \le \|x_{k+1} - x(\mu_k)\| + \|x(\mu_{k+1}) - x(\mu_k)\|$$

Recalling that $x_{k+1} = \tilde{x}_{j+1}$ the first summand is estimated by

$$\|x_{k+1} - x(\mu_k)\| \approx [r](x_{k+1};\mu_k),$$

while the second summand is estimated via (42) as

$$\|x(\mu_{k+1}) - x(\mu_k)\| \approx \eta_M(\mu_{k+1})|\mu_k - \mu_{k+1}|.$$

To compute μ_{k+1} it is sensible to aim for a certain contraction Θ_d, supplied by the user, which should be achieved by the next Newton correction step. The requirement $\Theta_M(x_{k+1};\mu_{k+1}) = \Theta_d$ yields

$$\|x_{k+1} - x(\mu_{k+1})\| = \frac{\Theta_d}{\omega_M(\mu_{k+1})}.$$

Inserting our estimates into these equations we obtain, setting $\sigma := \mu_{k+1}/\mu_k$:

$$[r](x_{k+1};\mu_k) + [\eta](\mu_k)\mu_k(1-\sigma)\sigma^{-1/2} = \frac{\Theta_d}{[\omega](\mu_k)\sigma^{-1/2}}.$$

Since $[r](x_{k+1};\mu_k), [\eta](\mu_k), [\omega](\mu_k)$ are computationally available and Θ_d is given we may compute σ, and thus $\mu_{k+1} = \sigma\mu_k$ from this equation.

If the corrector has terminated with a failure, we still have $[r](x_k;\mu_{k-1}), [\eta](\mu_{k-1})$, and $[\omega](\mu_k)$ at hand. Note that $[\omega](\mu_k)$ is the result of the evaluation of the failed Newton step and thus gives rise to a step size reduction. Hence we can compute σ analogously to the successful case with $\mu_k < \sigma\mu_{k-1} < \mu_{k-1}$, via

$$[r](x_k;\mu_{k-1}) + [\eta](\mu_{k-1})\mu_{k-1}(1-\sigma)\sigma^{-1/2} = \frac{\Theta_d}{[\omega](\mu_k)\sigma^{-1/2}}.$$

which serves as a step size reduction.

6.4 Termination Criteria

Depending on the application there are several termination criteria conceivable. For example we may use (43) or a modification to obtain a stopping criterion in terms of a norm of interest.

As an alternative we may stop if the estimated error in the functional is below a certain bound. Theorem 3.3 provides us with a linear convergence result via (11). By evaluation of the function values during the iteration we may estimate the missing constant and arrive at a good estimate for the error in the functional. Because the function values converge linearly in μ the difference $J(x(\mu)) - J(x_*)$ becomes small very quickly.

It is worth pointing out that interior point methods terminate at a *feasible* suboptimal solution. This feature, not shared by exterior penalty methods, reliefs users from the difficulty to decide how much infeasibility they are willing to accept. Rather, users can balance between optimality and computational effort.

If a quantitative discretization error estimate is available, then the error bounds can be matched with the these estimates. For a-priori error estimates for interior point methods we refer to [8], while quantitative a-posteriori error estimates together with an adaptive grid refinement strategy are subject of current research.

7 Numerical Examples

In our numerical examples we will investigate the performance of the proposed variants of interior point methods for some model problems. In particular, we are interested in the qualitative convergence behaviour of the path-following scheme and a-posteriori estimates for the quantities which govern the path-following method, namely $[\eta]$ and $[\omega]$. Further we are interested in the convergence behaviour of the function values at the iterates. Finally we are interested in the efficiency of the proposed method.

For a simple numerical example we consider an elliptic distributed control problem on the unit square $\Omega =]0;1[\times]0;1[$. For some $\infty > p > 2$ and $1/p + 1/p' = 1$ we define

$$\begin{aligned} A : W^{1,p}(\Omega) &\to W^{1,p'}(\Omega)^* \\ y \mapsto Ay : \langle Ay, v\rangle &:= \int_\Omega \langle \nabla y, \nabla v\rangle + yv\, dt. \end{aligned}$$

It follows from regularity theory [1, Thm. 9.2] that A is an isomorphism. Moreover, by $p > 2$ there exists a continuous Sobolev embedding $W^{1,p}(\Omega) \hookrightarrow C(\overline{\Omega})$, which is dense. These two results allow us to define A as a closed, densely defined, bijective operator via Lemma 2.1:

$$A : C(\overline{\Omega}) \supset W^{1,p}(\Omega) \to W^{1,p'}(\Omega)^*.$$

The operator B is defined by

$$\begin{aligned} B : L_2(\Omega) &\to W^{1,p'}(\Omega)^* \\ u \mapsto Bu : \langle Bu, v\rangle &:= \int_\Omega uv\, dt. \end{aligned}$$

B is continuous by the Sobolev embedding theorems, if $W^{1,p'}(\Omega) \hookrightarrow L_2(\Omega)$ is continuous, i.e. $p' > 1$.

As for state constraints we choose $\overline{y} = 0.5$ as an upper bound, for the definition of the functional J we choose $y_d = 2 \cdot x_1 \cdot x_2$, $\alpha = 10^{-3}$. As boundary conditions we choose homogenous Neumann conditions. The optimal state has a relatively large active set, and the Lagrange multiplier apparently consists of a regular part and a line measure, concentrated at the boundary of the active set. The control reveals an edge at the boundary of the active set.

As a second numerical example we change the boundary conditions to homogenous Dirichlet conditions and choose $\overline{y} = 0.55$. Inspection of the numerical solution yields that the active constraint set seems to be concentrated on an single point with a point measure as Lagrange multiplier. The adjoint state (and thus the control) has a sharp peak at the active point.

The discretization of y and p is performed by linear finite elements as described in [8] on a uniform triangular grid. The implementation is based on the DUNE library [2]. For the evaluation of the barrier integrals we use the trapezoidal rule, as analyzed in [8]. The resulting linear systems of equations are solved by the direct sparse solver PARDISO [11].

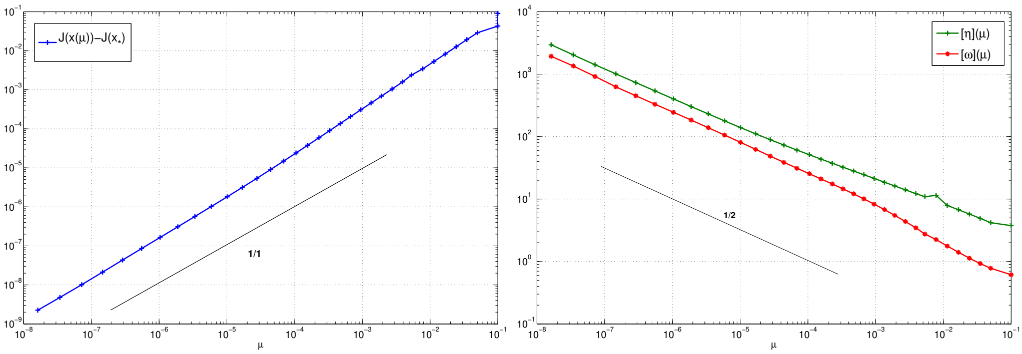

Figure 1: First problem. Left: Error in functional values. Right: Algorithmic quantities.

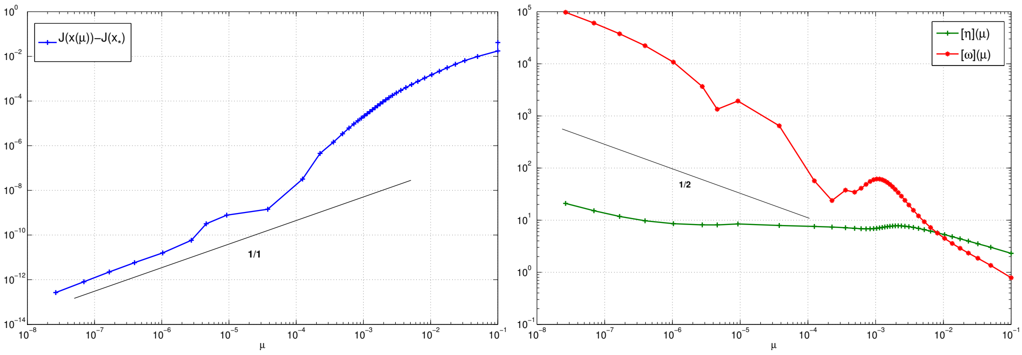

Figure 2: Second problem. Left: Error in functional values. Right: Algorithmic quantities.

Let us first have a look at the algorithmic quantities $[\eta]$ and $[\omega]$. It will turn out below that our method is able to perform very large reductions of μ per step. To obtain smooth plots we deliberately set the algorithmic parameters to very conservative values (in particular $\Theta_d = 0.05$) in the following. We also choose the stopping value for $\mu_{end} = 10^{-8}$ much smaller than appropriate for a practical application.

Comparing Figure 1 to our theoretic results we conclude that the theoretic predictions $J(x(\mu)) - J(x_*) = O(\mu)$ and $\eta(\mu) = O(\mu^{-1/2})$ made in Theorem 3.3 correspond rather well to the computational estimates. A very close look at the graphs suggests that the convergence is slightly faster in this particular problem. In contrast the a-posteriori estimate $[\omega](\mu) = O(\mu^{-1/2})$ is much better than the predicted bound from Proposition 4.7 for this particular problem.

$j \setminus i$	0	1	2	3
0	13	-	-	-
1	12	19	-	-
2	17	21	21	-
3	17	22	23	23

$j \setminus i$	0	1	2	3
0	14	-	-	-
1	12	12	-	-
2	13	15	12	-
3	15	15	13	13

Figure 3: Number of Newton steps used by the various barrier functions $l_{i,j}(y;\mu)$. Left: First problem. Right: Second problem.

Figure 2, which corresponds to the case with a point functional shows a quite different behaviour. While $J(x(\mu)) - J(x_*) = O(\mu)$ seems to hold asymptotically, $\eta(\mu)$ grows much more slowly than in our first example, while $\omega(\mu)$ grows faster and less regularly with local maximum near $\mu = 10^{-3}$. Observe how our algorithm reduces the stepsize in this difficult region to be able to comply to our (very restrictive) contraction demands. Summarizing, the second problem seems to be more nonlinear than the first, while having a shorter central path. This underlines the necessity of modeling the Newton nonlinearity as well as the properties of the central path.

Let us turn to the efficiency of our algorithm. The results of [8] suggest that for mesh sizes up to $h = 2^{-8}$ the discretization error (at least for the first problem) is above $2.5 \cdot 10^{-3}$. Hence, the choice of 10^{-3} as an accuracy requirement seems appropriate. For the first problem our algorithm detects this accuracy around $\mu \approx 5 \cdot 10^{-7}$, for the second problem this criterion is reached around $\mu \approx 10^{-5}$. The error in the functional is around 10^{-7} in the first problem and around 10^{-9} in the second problem.

For the desired contraction Θ_d we now choose a more aggressive value $\Theta_d = 0.8$ and a relative accuracy $\rho = 0.5$ (cf. (44)) for the corrector. To assess the influence of the order parameter q on the computational performance we compute the solutions of our problem with the help barrier functions of the form

$$l_{i,j}(y;\mu) = \sum_{k=i}^{j} l(y;\mu;q = 1 + k/2)$$

with $0 \le i \le j \le 3$. The table indicates that for the first problem and for $h = 2^{-8}$ low order barrier functions seem to be the more efficient than their higher order counterparts. In particular, efficiency degrades, if low order terms are dropped. For the second problem there seems to be no clear advantage for any type of barrier function.

To inspect mesh dependence of our algorithm (using the pure logarithmic barrier function from now on) test runs were performed for $h = 2^{-k}$ for $k = 4 \ldots 9$. While the number of Newton steps used for the first problem appears to be constant, for the second problem iteration counts increase slightly for finer discretizations. This reflects the fact that the structure of the solution of the first problem is already well resolved on coarse grids, while for the second problem the peak in the control is resolved only gradually when the grid is refined.

Problem $\setminus k$	4	5	6	7	8	9
#1	13	14	12	12	13	13
#2	9	11	11	12	14	13

Figure 4: Number of Newton steps depending on the mesh size $h = 2^{-k}$

Finally, for comparison we solved the first problem by a short step Newton path-following method without pointwise modification, but with a save-guard damping to prevent iterates from becoming infeasible. The total number of Newton steps for $h = 2^{-6}$ was 92, which is of course not competitive at all. The main problem is that even close to the solution the homotopy steps are small, because the pointwise nonlinearity introduced by the barrier functions is high.

References

[1] H. Amann. Nonhomogeneous linear and quasilinear elliptic and parabolic boundary value problems. In H.J. Schmeisser and H. Triebel, editors, *Function Spaces, Differential Operators and Nonlinear Analysis.*, pages 9–126. Teubner, Stuttgart, Leipzig, 1993.

[2] P. Bastian, M. Blatt, C. Engwer, A. Dedner, Klöfkorn, Kuttanikkad R., M. S., Ohlberger, and O. Sander. The Distributed and Unified Numerics Environment (DUNE). In *Proc. of the 19th Symposium on Simulation Technique in Hannover, September 12-14*, 2006.

[3] E. Casas. Control of an elliptic problem with pointwise state constraints. *SIAM J. Control Optim.*, 24(6):1309–1318, 1986.

[4] P. Deuflhard. *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms*, volume 35 of *Series Computational Mathematics*. Springer, 2004.

[5] I. Ekeland and R. Témam. *Convex Analysis and Variational Problems.* Number 28 in Classics in Applied Mathematics. SIAM, 1999.

[6] S. Goldberg. *Unbounded Linear Operators.* Dover Publications, Inc., 1966.

[7] M. Hintermüller and K. Kunisch. Feasible and non-interior path-following in constrained minimization with low multiplier regularity. *SIAM J. Control Optim.*, 45(4):1198–1221, 2006.

[8] M. Hinze and A. Schiela. Discretization of interior point methods for state constrained elliptic optimal control problems: Optimal error estimates and

parameter adjustment. Technical Report SPP1253-08-03, Priority Program 1253, German Research Foundation, 2007.

[9] C. Meyer, F. Tröltzsch, and A. Rösch. Optimal control problems of PDEs with regularized pointwise state constraints. *Computational Optimization and Applications*, 33:206–228, 2006.

[10] U. Prüfert, F. Tröltzsch, and M. Weiser. The convergence of an interior point method for an elliptic control problem with mixed control-state constraints. ZIB Report 04-47, Zuse Institute Berlin, 2004. to appear at Computational Optimization and Applications.

[11] O. Schenk and K. Gärtner. On fast factorization pivoting methods for sparse symmetric indefinite systems. *Elec. Trans. Numer. Anal*, 23:158–179, 2006.

[12] A. Schiela. *The Control Reduced Interior Point Method - A Function Space Oriented Algorithmic Approach.* PhD thesis, Free University of Berlin, Dept. Math. and Comp. Sci., 2006.

[13] A. Schiela. Convergence of the control reduced interior point method for PDE constrained optimal control with state constraints. ZIB Report 06-16, Zuse Institute Berlin, 2006.

[14] A. Schiela. Barrier methods for optimal control problems with state constraints. ZIB Report 07-07, Zuse Institute Berlin, 2007.

[15] A. Schiela. An extended mathematical framework for barrier methods in function space. ZIB Report 08-07, Zuse Institute Berlin, 2008.

[16] M. Weiser. *Function Space Complementarity Methods for Optimal Control Problems.* PhD thesis, Free University of Berlin, Dept. Math. and Comp. Sci., 2001.

[17] D. Werner. *Funktionalanalysis.* Springer, 3^{rd} edition, 2000.