Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

Anton Schiela, Martin Weiser

Superlinear convergence of the Control Reduced Interior Point Method for PDE Constrained Optimization[1]

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

ZIB-Report 05-15 (February 2005)

Superlinear convergence of the Control Reduced Interior Point Method for PDE Constrained Optimization†

Anton Schiela, Martin Weiser

February 18, 2005

Abstract

A thorough convergence analysis of the Control Reduced Interior Point Method in function space is performed. This recently proposed method is a primal interior point pathfollowing scheme with the special feature, that the control variable is eliminated from the optimality system. Apart from global linear convergence we show, that this method converges locally almost quadratically, if the optimal solution satisfies a certain non-degeneracy condition. In numerical experiments we observe, that a prototype implementation of our method behaves as predicted by our theoretical results.

AMS MSC 2000: 90C51, 65N15, 49M15

Keywords: interior point methods in function space, optimal control

1 Introduction

In the last years, interior point methods enjoyed an increasing popularity in optimal control with stationary elliptic PDEs, most often applied to a finite dimensional discretized problem in a *discretize then optimize* fashion [6, 7]. More recently, they have also been applied to the infinite-dimensional continuous control problem, forming the outermost loop of an *optimize then discretize* approach [12, 16].

For the convergence analysis of interior point methods, the regularity of the involved variables is of crucial importance. Therefore, the choice of appropriate function spaces and norms is essential — a problem that is also faced by the discretize then optimize approach if the discretization is sufficiently fine to capture the structure of the continuous problem. Typically the least regular variables, which are the Lagrange multipliers for state constraints, or the control and the multipliers for control constraints in the absence of state constraints, cause the most troubles for the analysis as well as for numerical approximation.

Little is known about state constrained problems, even though numerical experience indicates that interior point methods perform reasonably well on practical

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

problems. We just mention [9] for regularized state constraints and [7] for discretized problems.

In contrast primal-dual interior point methods have been shown to converge linearly for control constrained problems [14, 16]. This behavior is also observed in numerical implementations.

A different type of interior point methods in function space was analysed by Ulbrich and Ulbrich in [12]. Here an affine scaling method was augmented by a smoothing step in order to achieve superlinear local convergence. These methods are related to semi-smooth Newton methods [11]. In recent talks Stefan Ulbrich reports on a primal-dual interior point method with smoothing step that converges superlinearly [13].

For a faithful approximation of the control, a massive mesh refinement around the boundaries of the active sets is necessary, due to reduced regularity of the control at the boundaries of the active sets which are usually not aligned with the coarse grid.

In order to obtain higher accuracy on coarse meshes, Hinze [4] proposed to eliminate the control pointwise from the optimality system, such that only the more regular variables, the state and the adjoint state, need to be discretized. This approach leads to a semismooth equation that characterizes the solution. The very same idea has been translated to primal interior point methods by Weiser, Gänzler, and Schiela [15], where the semismooth equation is substituted by a homotopy of smooth equations. Linear convergence of a short step pathfollowing method has been established, and a priori spatial error bounds for finite element discretizations have been derived.

Distantly related work includes Rösch [8], who obtains similar accuracy on coarse meshes by a postprocessing step, and Hintermüller and Kunisch [3], who perform a regularization homotopy in the context of primal-dual active set methods.

The pointwise elimination of the control does not only alleviate the need for mesh refinement, but does also close the remaining gap to finite dimensional interior point methods. In the present paper, we concentrate on the convergence of an abstract pathfollowing method in function space. Under reasonable assumptions about the solution's structure we show *r-convergence of order 2*. This means r-superlinear convergence that approaches quadratic convergence asymptotically. This result is comparable to the convergence speed of finite dimensional interior point methods as presented in [17]. The analysis mainly depends on two related quantities: the slope of the central path, and an affine invariant Lipschitz constant that governs the convergence of Newton's method.

2 Basic ideas

We will now give a short summary of the ideas of the control reduced interior point method. Consider the model problem

$$\min_{y\in H_0^1, u\in L_2} \frac{1}{2}\|y-y_d\|_{L_2}^2 + \frac{\alpha}{2}\|u\|_{L_2}^2 \quad \text{subject to} \quad Ly = u, \quad -1 \le u \le 1 \tag{1}$$

on some convex polyhedric domain $\Omega \subset \mathbb{R}^d$ $(1 \le d \le 3)$. $y_d \in L_2$ is the desired state, and $\alpha > 0$ is a fixed regularization parameter. $Ly = -\operatorname{div}(a(x)\nabla y) + b(x)y$ with symmetric $a(x) \in \mathbb{R}^{d\times d}$ uniformly positive definite and $b(x) \in \mathbb{R}$ non-negative is a second order elliptic differential operator. Note, that the state equation is an H^2-regular problem.

As already described in [5] for problem (1) the first order necessary conditions state the existence of a Lagrange multiplier $\lambda \in H_0^1$, such that

$$\begin{aligned} y - y_d + L\lambda &= 0 \\ Ly - u &= 0 \\ u - \operatorname{Proj}_{[-1,1]}\left(\alpha^{-1}\lambda\right) &= 0. \end{aligned} \tag{2}$$

Primal interior point methods substitute (2) by the regularized equation

$$g(\lambda, u; \mu) := \alpha u - \lambda - \frac{\mu}{u+1} + \frac{\mu}{1-u} = 0 \tag{3}$$

for $\mu > 0$ and thus define the central path $\mu \mapsto (y, u, \lambda)$ via the system of equations

$$\begin{aligned} y - y_d + L\lambda &= 0 \\ Ly - u &= 0 \\ \alpha u - \lambda - \frac{\mu}{u+1} + \frac{\mu}{1-u} &= 0 \\ -1 \le u &\le 1. \end{aligned} \tag{4}$$

These are just the first order necessary conditions for the logarithmic barrier reformulation of (1),

$$\min \frac{1}{2}\|y-y_d\|_{L_2}^2 + \frac{\alpha}{2}\|u\|_{L_2}^2 + \mu\int (\ln(u+1) + \ln(1-u))\, dx \quad \text{subject to} \quad Ly = u.$$

2.1 Elimination of the control

We can use (3) and (4) in order to eliminate $u = u(\lambda;\mu)$. We recapitulate and extend the analysis of the properties of the function $u(\lambda;\mu)$.

Lemma 2.1. *For each $\lambda \in \mathbb{R}$, $\mu > 0$ there is exactly one $u(\lambda;\mu) \in]-1,1[$ that satisfies (3). Moreover, $u(\lambda;\mu)$ is continuously differentiable with respect to μ and twice continuously differentiable with respect to λ. Defining*

$$h_1(u;\mu) := (1+u)(1-u)g_u = \alpha(1-u^2) + \mu\left(\frac{1-u}{1+u} + \frac{1+u}{1-u}\right)$$

we can formulate

$$u_\lambda = \frac{1-u^2}{h_1} > 0, \quad u_{\lambda\lambda} = -\frac{4\mu}{h_1^3}u(3+u^2), \quad u_\mu = -\frac{2u}{h_1}, \quad u_{\lambda\mu} = -\frac{2(1+u^2)}{h_1^2}.$$

Moreover $0 \le u_\lambda \le (\alpha + 2\mu)^{-1}$.

Proof. The uniqueness of $u(\lambda;\mu)$ follows immediately from the monotonicity of g with respect to u and from $\lim_{u\to\pm 1} g(\lambda, u;\mu) = \pm\infty$.

Differentiating $g(\lambda, u(\lambda;\mu);\mu) = 0$ with respect to λ yields:

$$\begin{aligned} g_u u_\lambda - 1 = 0 \quad &\Rightarrow \quad u_\lambda = \frac{1}{g_u} \\ g_{uu}u_\lambda^2 + g_u u_{\lambda\lambda} = 0 \quad &\Rightarrow \quad u_{\lambda\lambda} = -\frac{1}{g_u}(g_{uu}u_\lambda^2) = -\frac{g_{uu}}{g_u^3}. \end{aligned}$$

Clearly $g_u \ge \alpha + 2\mu$ yields positivity and boundedness of u_λ and:

$$\begin{aligned} u_{\lambda\lambda} &= -\frac{1}{h_1^3} g_{uu}(1-u)^3(u+1)^3 = -\frac{1}{h_1^3}\left(\frac{-2\mu}{(u+1)^3} + \frac{2\mu}{(1-u)^3}\right)(1-u)^3(u+1)^3 \\ &= -\frac{2\mu}{h_1^3}(-(1-u)^3 + (u+1)^3) = -\frac{4\mu}{h_1^3}u(3+u^2). \end{aligned}$$

Similarly, differentiating $g(\lambda, u(\lambda;\mu);\mu) = 0$ with respect to μ gives us

$$0 = g_u u_\mu + g_\mu \quad \Rightarrow \quad u_\mu = -\frac{g_\mu}{g_u} = -\frac{g_\mu(1-u)(1+u)}{h_1} = -\frac{2u}{h_1}.$$

The statement for $u_{\lambda\mu}$ follows analogously from differentiating u_λ with respect to μ. □

Defining

$$V := H_0^1 \times H_0^1, \qquad v(\mu) := \begin{pmatrix} y(\mu) \\ \lambda(\mu) \end{pmatrix}$$

Lemma 2.1 asserts, that the homotopy

$$F(\,\cdot\,;\mu) : V \to V^*, \quad F(v;\mu) = \begin{bmatrix} y - y_d + L\lambda \\ Ly - u(\lambda;\mu) \end{bmatrix} = 0 \tag{5}$$

is well defined and in turn defines the central path. Its existence and differentiability with respect to μ in V has been discussed in [9] and [15]. In Section 3 we will refine those results substantially.

Lemma 2.2. *Let* h_1 *be defined as in Lemma 2.1 and*

$$h_2(u;\mu) := h_1(u;\mu)(1-u^2) = \alpha(1-u^2)^2 + 2\mu\left(1+u^2\right).$$

Then

$$\begin{aligned} h_1(u;\mu) &\ge \max\{2\sqrt{2\alpha\mu}, 2\mu\} \\ h_2(u;\mu) &\ge 2\mu(1+u^2). \end{aligned}$$

Proof. First we note: $h_1 \geq h_2 \geq 2\mu(1+u^2)$. By the relation $a, b \geq 0 \Rightarrow a+b \geq 2\sqrt{ab}$ we infer:

$$h_1(u;\mu) \geq 2\sqrt{\alpha(1-u^2)\mu\left(\frac{1-u}{1+u}+\frac{1+u}{1-u}\right)} = 2\sqrt{\alpha\mu 2\,(1+u^2)} \geq 2\sqrt{2\alpha\mu}.$$

□

2.2 Newton linearization and local norms

Now we discuss the solvability of the linearization of (5). This is important for the analysis of our Newton pathfollowing algorithm. The linearized system reads:

$$\begin{pmatrix} I & L \\ L & -u_\lambda(\lambda;\mu) \end{pmatrix}\begin{pmatrix} \delta y \\ \delta\lambda \end{pmatrix} = \begin{pmatrix} r \\ s \end{pmatrix}. \tag{6}$$

Since $u_\lambda > 0$ we may define a special family of norms for $\delta v := (\delta y, \delta\lambda)$ by

$$\|\delta v\|_{v,\mu} := \|\delta y\|_{L_2} + \|\delta\lambda\|_{L_\infty} + \|\sqrt{u_\lambda(v;\mu)}\delta\lambda\|_{L_2} \tag{7}$$

depending on a given $v \in V$ and μ. Thus we call it a family of *local norms.*

Theorem 2.3. *The linearized optimality system* (6) *posesses a unique solution $\delta v \in V$. If $r, s \in L_2$ then additionally $\delta v \in H^2 \times H^2$ and the following estimates hold:*

$$\|\delta v\|_{v,\mu} + \|\delta\lambda\|_{H^2} \leq c\left(\|r\|_{L_2} + \min\left\{\|s\|_{L_1}, \left\|\frac{s}{\sqrt{u_\lambda}}\right\|_{L_2}\right\}\right), \tag{8}$$

$$\|\delta y\|_{H^2} \leq \frac{c}{\sqrt{\alpha+2\mu}}\left(\|r\|_{L_2} + \min\left\{\|s\|_{L_1}, \left\|\frac{s}{\sqrt{u_\lambda}}\right\|_{L_2}\right\}\right) + \|s\|_{L_2}. \tag{9}$$

Proof. Unique solvability of (6) has been established in [15] together with continuity of

$$F_v^{-1} : ((H_0^1)^* \times (H_0^1)^*) \to (H_0^1 \times H_0^1).$$

However, its special structure allows us to refine this result if $r, s \in L_2$.

By the first row of (6) and $r \in L_2$ we have $L\delta\lambda \in L_2$, which implies $\delta\lambda \in H^2 \cap H_0^1$. By the second row we similarly obtain $L\delta y \in L_2$. Thus we can take the scalar product of the second row and $-\delta\lambda$:

$$-\langle\delta\lambda, L\delta y\rangle + \langle\delta\lambda, u_\lambda\delta\lambda\rangle = -\langle\delta\lambda, s\rangle \tag{10}$$

and apply partial integration twice to obtain $\langle\delta\lambda, L\delta y\rangle = \langle L\delta\lambda, \delta y\rangle$, using $\delta\lambda, \delta y \in H_0^1$. Then we eliminate δy inserting the first row of (6) and finally arrive at the identity

$$\langle L\delta\lambda, L\delta\lambda\rangle + \langle\delta\lambda, u_\lambda\delta\lambda\rangle = -\langle\delta\lambda, s\rangle + \langle L\delta\lambda, r\rangle.$$

The Cauchy Schwartz inequality and Hölder's inequality yield the following estimate in the energy norm

$$\begin{aligned}\|L\delta\lambda\|_{L_2}^2 + \|\sqrt{u_\lambda}\delta\lambda\|_{L_2}^2 &\le |\langle\delta\lambda, s\rangle| + \|r\|_{L_2}\|L\delta\lambda\|_{L_2} \\ &\le \min\left\{\|s/\sqrt{u_\lambda}\|_{L_2}\,\|\sqrt{u_\lambda}\delta\lambda\|_{L_2}, \|s\|_{L_1}\|\delta\lambda\|_{L_\infty}\right\} + \|r\|_{L_2}\|L\delta\lambda\|_{L_2}.\end{aligned}$$

By the Sobolev imbedding $H^2 \cap H_0^1 \hookrightarrow L_\infty$ and standard regularity theory it follows $\|\delta\lambda\|_{L_\infty} \le c\|\delta\lambda\|_{H^2} \le c\|L\delta\lambda\|_{L_2}$ and thus

$$\sqrt{\|L\delta\lambda\|_{L_2}^2 + \|\sqrt{u_\lambda}\delta\lambda\|_{L_2}^2} \le c\left(\|r\|_{L_2} + \min\{\|s/\sqrt{u_\lambda}\|_{L_2}, \|s\|_{L_1}\}\right).$$

The estimates for δy are now obtained easily. The first row of (6) yields $\|\delta y\|_{L_2} \le \|L\delta\lambda\|_{L_2} + \|r\|_{L_2}$ and thus (8). From the second row of (6) we conclude that $\|\delta y\|_{H^2} \le \|\sqrt{u_\lambda}\|_{L_\infty}\|\delta\lambda\|_{v,\mu} + \|s\|_{L_2}$ and thus (9). $\square$

Remark 2.4. Compared to the results in [16] the inverse Jacobian F_v^{-1} has the remarkable property to map from a rough space into a smooth space. This property is a consequence of the elimination of u via the optimality condition. Moreover, it is essential to show superlinear convergence.

Here we see a relation between control reduction and the application of a smoothing step to u as performed for example in [12]. There a control iterate u is smoothed by inserting it into the state equation, which yields a smoothed u_s via the optimality condition. This way the missing smoothing property of the inverse Jacobian is compensated. Thus, both methods use the optimality condition to overcome the lack of smoothness in the unreduced system.

Next we recapitulate an affine covariant Newton convergence theorem with local norms. It is a slight extension of the Newton Mysovskii Theorem presented in [2].

Theorem 2.5. *Suppose V, Z are Banach spaces, $D \subset V$ is open and $F : D \to Z$ is Gâteaux-differentiable. Additionally, assume $F_v(v)$ is invertible for all $v \in D$. Let $\|\cdot\|_v : Z \to \mathbb{R}$ form a γ-continuous family of norms. Let ω be a constant, such that:*

$$\|F_v^{-1}(v)\left(F_v(v + t\Delta v) - F_v(v)\right)\Delta v\|_v \le t\omega\|\Delta v\|_v\|\Delta v\|_{v+t\Delta v} \tag{11}$$

holds for all $v, v + \Delta v \in D$.

Define $e := v - v_*$ and $h := \omega\|e\|$. Let $v_0 \in D$ be given, v_* a solution of $F(v) = 0$ and assume

$$D \supset S_* := \overline{S}(v_*, \|\cdot\|_{v_*}, (1 + \gamma\|e_0\|_{v_0})\|e_0\|_{v_0}).$$

If for the starting error

$$(1 + \gamma\|e_0\|_{v_0})^3\, h_0 < 2 \tag{12}$$

holds, then the Newton sequence $v_{k+1} = v_k - F_v(v_k)^{-1}F(v_k)$ converges quadratically towards v_* in the sense that

$$\|e_{k+1}\|_{v_{k+1}} < (1 + \gamma\|e_k\|_{v_k})^3\,\omega/2\|e_k\|_{v_k}^2. \tag{13}$$

Roughly speaking, γ and ω capture the nonlinearity of the problem and determine the radius of convergence. For the analysis of a Newton pathfollowing method with local norms we have to derive bounds for both parameters. We obtain bounds for γ in this section and for ω in Section 3. For this purpose we first analyse some additional continuity properties of the scaling function u_λ.

Lemma 2.6. *The following continuity properties hold for $u_\lambda(\lambda;\mu)$:*

$$|\sqrt{u_\lambda(\hat\lambda;\mu)} - \sqrt{u_\lambda(\bar\lambda;\mu)}| \;\le\; \frac{2}{\sqrt{\mu}}\sqrt{u_\lambda(\hat\lambda;\mu)}\sqrt{u_\lambda(\bar\lambda;\mu)}|\hat\lambda-\bar\lambda| \tag{14}$$

$$|\sqrt{u_\lambda(\lambda;\hat\mu)} - \sqrt{u_\lambda(\lambda;\bar\mu)}| \;\le\; \frac{1}{\sqrt{\alpha\hat\mu}}|\hat\mu-\bar\mu|,\ \text{for } \hat\mu\le\bar\mu \tag{15}$$

Proof. To show (14) we suppress the argument μ in u_λ. The mean-value theorem yields

$$|u_\lambda(\bar\lambda)^{-1/2} - u_\lambda(\hat\lambda)^{-1/2}| \le \left|\left(u_\lambda(\tilde\lambda)^{-1/2}\right)'(\hat\lambda-\bar\lambda)\right| = \left|\frac{1}{2}u_{\lambda\lambda}(\tilde\lambda)u_\lambda(\tilde\lambda)^{-3/2}\right||\hat\lambda-\bar\lambda|$$

for some $\tilde\lambda\in[\bar\lambda,\hat\lambda]$. Defining $\tilde u = u(\tilde\lambda)$ and using Lemma 2.1 and Lemma 2.2 we continue with

$$\left|\frac{1}{2}u_{\lambda\lambda}(\tilde\lambda)u_\lambda(\tilde\lambda)^{-3/2}\right| \le \frac{2(3+u^2)\mu}{(h_1(1-\tilde u^2))^{3/2}} = \frac{(4+2(1+u^2))\mu}{h_2^{3/2}} \le \frac{2}{\sqrt{\mu}}.$$

Multiplication with $\sqrt{u_\lambda(\hat\lambda)}\sqrt{u_\lambda(\bar\lambda)}$ yields (14). For (15) we similarly deduce:

$$\begin{aligned}\left|\sqrt{u_\lambda(\hat\mu)} - \sqrt{u_\lambda(\bar\mu)}\right| &\le \frac{|u_{\lambda\mu}(\tilde\mu)|}{2\sqrt{u_\lambda(\tilde\mu)}}|\hat\mu-\bar\mu| \le \frac{2}{h_1(\tilde\mu)\sqrt{h_2(\tilde\mu)}}|\hat\mu-\bar\mu| \\ &\le \frac{1}{\sqrt{\alpha\tilde\mu}}|\hat\mu-\bar\mu| \le \frac{1}{\sqrt{\alpha\hat\mu}}|\hat\mu-\bar\mu|.\end{aligned}$$

□

An immediate consequence is the γ-continuity of our norm.

Corollary 2.7. *The norm $\|\cdot\|_{v,\mu}$ is γ-continuous with $\gamma = 2(\alpha\mu)^{-1/2}$, or more precisely:*

$$\left|\|z\|_{\hat v,\mu} - \|z\|_{\bar v,\mu}\right| \;\le\; 2(\alpha\mu)^{-1/2}\|\hat\lambda-\bar\lambda\|_{L_\infty}\|z\|_{\hat v,\mu}. \tag{16}$$

Moreover, the norm $\|\cdot\|_{v,\mu}$ is Lipschitz continuous with respect to μ:

$$\left|\|z\|_{v,\hat\mu} - \|z\|_{v,\bar\mu}\right| \le \frac{1}{\sqrt{\alpha\hat\mu}}|\hat\mu-\bar\mu|\|z\|_{v,\hat\mu} \quad \text{for } \hat\mu\le\bar\mu. \tag{17}$$

Proof. By definition

$$\begin{aligned}\left|\|z\|_{\hat{v},\hat{\mu}} - \|z\|_{\bar{v},\bar{\mu}}\right| &= \left|\left\|\sqrt{u_\lambda(\hat{\lambda};\hat{\mu})}\lambda_z\right\|_{L_2} - \left\|\sqrt{u_\lambda(\bar{\lambda};\bar{\mu})}\lambda_z\right\|_{L_2}\right| \\ &\leq \left\|\left(\sqrt{u_\lambda(\hat{\lambda};\hat{\mu})} - \sqrt{u_\lambda(\bar{\lambda};\bar{\mu})}\right)\lambda_z\right\|_{L_2}.\end{aligned}$$

To show (16) we set $\bar{\mu} = \hat{\mu} = \mu$ and use Lemma 2.6 and the Hölder inequality to continue:

$$\begin{aligned}\left|\|z\|_{\hat{v},\mu} - \|z\|_{\bar{v},\mu}\right| &\leq 2\mu^{-1/2}\left\|\sqrt{u_\lambda(\hat{\lambda})}\sqrt{u_\lambda(\bar{\lambda})}|\hat{\lambda} - \bar{\lambda}|\lambda_z\right\|_{L_2} \\ &\leq 2\mu^{-1/2}\left\|\sqrt{u_\lambda(\bar{\lambda})}(\hat{\lambda} - \bar{\lambda})\right\|_{L_\infty}\left\|\sqrt{u_\lambda(\hat{\lambda})}\lambda_z\right\|_{L_2} \\ &\leq 2(\alpha\mu)^{-1/2}\|\hat{\lambda} - \bar{\lambda}\|_{L_\infty}\|z\|_{\hat{v},\mu}.\end{aligned}$$

To show (17) we set $\bar{\lambda} = \hat{\lambda} = \lambda$ and continue similarly using (15)

$$\left|\|z\|_{v,\hat{\mu}} - \|z\|_{v,\bar{\mu}}\right| \leq \left\|\hat{\mu}^{-1}|\hat{\mu} - \bar{\mu}|\sqrt{u_\lambda(\hat{\mu})}\lambda_z\right\|_{L_2} \leq \hat{\mu}^{-1}\alpha^{-1/2}|\hat{\mu} - \bar{\mu}|\|z\|_{v,\hat{\mu}}.$$

□

3 The length of the central path and affine invariant Lipschitz constants.

Our mathematical framework for Newton pathfollowing methods relies on two quantities. On the one hand we need an estimate about the slope of the central path to assess how the solution of (5) changes when the homotopy parameter μ is reduced. On the other hand we have to estimate the convergence radius of Newton's method. Theorem 2.5 states that this quantity mainly depends on the affine invariant Lipschitz constant ω.

3.1 Estimates for global linear convergence

First we consider the slope of the central path, which is defined by the norm of

$$v_\mu(\mu) = -F_v(v(\mu);\mu)^{-1}F_\mu(v(\mu);\mu). \tag{18}$$

We will derive a bound for the slope that is independent of the choice of α and the structure of the optimal solution.

Theorem 3.1. *There is a constant $\tilde{c} < \infty$ independent of α and μ, such that the slope η of the central path is bounded by*

$$\eta := \|v_\mu(\mu)\|_{v(\mu),\mu} + \|\lambda_\mu(\mu)\|_{H^2} \leq \tilde{c}\min\{\mu^{-1/2}, \|u_\mu\|_{L_1}\}, \tag{19}$$

and simultaneously for $0 \leq \sigma \leq 1$ the estimates

$$\|y(\sigma\mu) - y(\mu)\|_{L_2} + \|\lambda(\sigma\mu) - \lambda(\mu)\|_{H^2} \leq \tilde{c}(1 - \sqrt{\sigma})\sqrt{\mu} \tag{20}$$

$$\|\lambda(\sigma\mu) - \lambda(\mu)\|_{L_\infty} \leq \tilde{c}(1 - \sqrt{\sigma})\sqrt{\mu} \tag{21}$$

$$\|y(\sigma\mu) - y(\mu)\|_{H^2} \leq \tilde{c}(1 - \sqrt{\sigma})\sqrt{\mu/\alpha} \tag{22}$$

hold.

Proof. Regarding the system (5) we observe that its first row is independent of μ. Furthermore, since

$$\frac{d}{d\mu}(Ly - u(\lambda;\mu)) = -u_\mu(\lambda;\mu),$$

we have $-F_\mu = (0, u_\mu)^T$. To evaluate $\|v_\mu(\mu)\|$ via (18) we apply Theorem 2.3 to F_v and the right hand side $(r,s)^T = -F_\mu = (0, u_\mu)^T$ to obtain

$$\|v_\mu\|_{v(\mu),\mu} + \|\lambda_\mu\|_{H^2} \leq c \min\left\{ \left\| \frac{u_\mu}{\sqrt{u_\lambda}} \right\|_{L_2}, \|u_\mu\|_{L_1} \right\}.$$

Now we use the representations of u_μ and u_λ from Lemma 2.1 and the estimate for h_2 in Lemma 2.2 to show

$$\left\| \frac{u_\mu}{\sqrt{u_\lambda}} \right\|_{L_2} = \left\| \frac{-2u}{h_1} \sqrt{\frac{h_1}{1-u^2}} \right\|_{L_2} = \left\| \frac{-2u}{\sqrt{h_2}} \right\|_{L_2} \leq \frac{2}{\sqrt{\min h_2}} \leq c\mu^{-1/2}, \tag{23}$$

which yields (19) and especially

$$\|y_\mu(\mu)\|_{L_2} + \|\lambda_\mu(\mu)\|_{L_\infty} + \|\lambda_\mu(\mu)\|_{H^2} \leq c\mu^{-1/2}.$$

We use this estimate to show (20) by integration of the slope:

$$\begin{aligned} \|y(\sigma\mu) - y(\mu)\|_{L_2} + \|\lambda(\sigma\mu) - \lambda(\mu)\|_{H^2} &\leq \int_{\sigma\mu}^{\mu} \|y_\mu(m)\|_{L_2} + \|\lambda_\mu(m)\|_{H^2} dm \\ &\leq c(\sqrt{\mu} - \sqrt{\sigma\mu}) \\ &= c(1 - \sqrt{\sigma})\sqrt{\mu} \end{aligned}$$

as well as (21). The estimate (22) follows by the same argumentation using (9) and

$$\|u_\mu(\mu)\|_{L_2} \leq c(\min h_1)^{-1} \leq c(\alpha\mu)^{-1/2}.$$

The constant $\tilde{c}$ is defined as the maximum of all constants c used in this proof. □

Next we derive a general estimate for the affine invariant Lipschitz constant ω. Here we make use of our local norm to obtain sharp results with respect to α. Since μ is a constant parameter in one Newton step, we suppress it as an argument.

Theorem 3.2. *The following estimates for the affine invariant Lipschitz constant ω hold:*

$$\begin{aligned} \|F_v^{-1}(v)\,(F_v(v+t\Delta v)-F_v(v))\,\Delta v\|_{v,\mu} &\leq t\omega_g \|\Delta\lambda\|_{v,\mu}\|\Delta\lambda\|_{v+t\Delta v,\mu} \\ \omega_g &= \frac{c}{\sqrt{\mu(\alpha+2\mu)}}. \end{aligned}$$

Proof. We define $\delta v := F_v^{-1}(v)\,(F_v(v+t\Delta v)-F_v(v))\,\Delta v$ and $\bar\lambda := \lambda + t\Delta\lambda$ to compute

$$(F_v(v+t\Delta v)-F_v(v))\,\Delta v = \begin{pmatrix} 0 \\ (u_\lambda(\bar\lambda)-u_\lambda(\lambda))\Delta\lambda \end{pmatrix} =: \begin{pmatrix} r \\ s \end{pmatrix}.$$

Now we can use Theorem 2.3 and (14) to estimate

$$\begin{aligned} \|\delta v\|_{v,\mu} \leq c\|s\|_{L_1} &= c\left\|\left(\sqrt{u_\lambda(\bar\lambda)}+\sqrt{u_\lambda(\lambda)}\right)\left(\sqrt{u_\lambda(\bar\lambda)}-\sqrt{u_\lambda(\lambda)}\right)\Delta\lambda\right\|_{L_1} \\ &\leq 2(\alpha+2\mu)^{-1/2}\frac{c}{\sqrt{\mu}}\left\|\sqrt{u_\lambda(\bar\lambda)u_\lambda(\lambda)}t|\Delta\lambda|^2\right\|_{L_1} \\ &\leq \frac{c}{\sqrt{\mu(\alpha+2\mu)}}t\|\Delta\lambda\|_{\bar v,\mu}\|\Delta\lambda\|_{v,\mu}. \end{aligned}$$

□

3.2 Estimates for local superlinear convergence

If the adjoint state λ_* corresponding to the optimal solution u_*, y_* of (1) satisfies additional assumptions, we can refine our estimates substantially.

Assumption 3.3. Assume there are $e_0 > 0, \Gamma > 0$, such that the adjoint state λ_* associated with the optimal solution of (1) fulfills the following requirement:

$$\Big|\{t \in \Omega : \lambda_*(t) \in [\alpha-2e, \alpha+2e] \cup [-\alpha-2e, -\alpha+2e]\}\Big| < \Gamma e$$

for all $e < e_0$.

Assumption 3.3 can be viewed as a function space analogue of a strict complementarity or non-degeneracy condition. In finite dimensional optimization such a condition is necessary to show superlinear convergence of interior point methods, as noted for example in [17]. In function space we cannot impose strict complementarity. Instead we can restrict the size of the regions, where strict complementarity is almost violated. This idea has already been used in [12] and [11] in order to show a certain rate of superlinear convergence of an affine scaling interior point method and of semismooth Newton methods.

Lemma 3.4. *Let Assumption 3.3 hold. Then for $e_0 > e > \|\lambda-\lambda_*\|_{L_\infty}$ and $u = u(\lambda;\mu)$ the auxilliary function h_1 defined in Lemma 2.1 satisfies*

$$|\{t \in \Omega : h_1(t) < e\}| \leq \Gamma e. \tag{24}$$

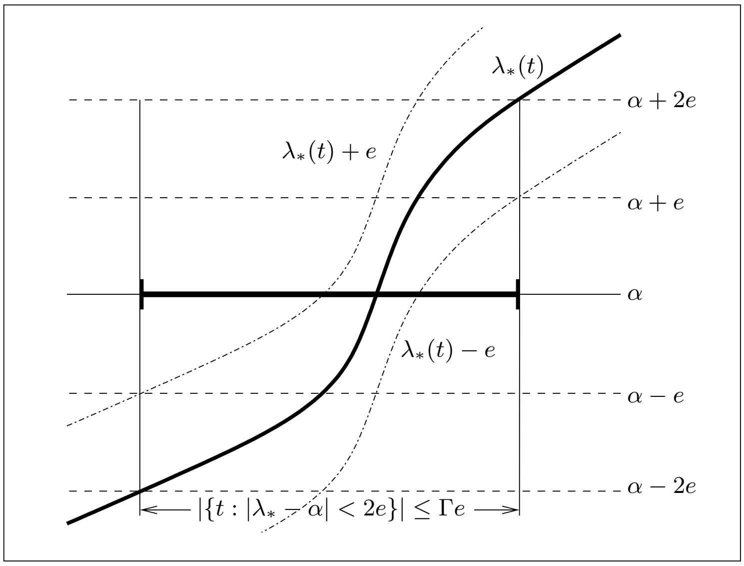

Figure 1: Illustration of Assumption 3.3

Proof. Since $\|\lambda - \lambda_*\|_{L_\infty} < e$ we conclude

$$|\lambda_* \pm \alpha| \ge 2e \Rightarrow |\lambda \pm \alpha| \ge |\lambda_* \pm \alpha| - |\lambda_* - \lambda| \ge 2e - e = e.$$

Thus by Assumption 3.3 we have

$$|\{t \in \Omega : |\lambda(t) \pm \alpha| < e\}| \le \Gamma e. \tag{25}$$

Now we assume $|\alpha - \lambda| \ge e$ and $|\alpha + \lambda| \ge e$ and solve (3) for λ to obtain

$$\begin{aligned}
\left|\alpha(1-u) + \frac{\mu}{1+u} - \frac{\mu}{1-u}\right|(1+u) &\ge e(1+u), \\
\left|\alpha(1+u) - \frac{\mu}{1+u} + \frac{\mu}{1-u}\right|(1-u) &\ge e(1-u).
\end{aligned}$$

Taking the maximum of both inequalities yields

$$\begin{aligned}
e &\le e \max\{1+u, 1-u\} \\
&\le \max\left\{\left|\alpha(1-u)(1+u) + \mu - \mu\frac{1+u}{1-u}\right|, \left|\alpha(1+u)(1-u) + \mu - \mu\frac{1-u}{1+u}\right|\right\} \\
&\le \max\left\{\left|\alpha(1-u^2) + \mu\right|, \left|\mu\frac{1+u}{1-u}\right|, \left|\mu\frac{1-u}{1+u}\right|\right\} \\
&\le \alpha(1-u^2) + \mu\frac{1+u}{1-u} + \mu\frac{1-u}{1+u} = h_1.
\end{aligned}$$

Thus $\{t \in \Omega : h_1(t) < e\} \subset \{t \in \Omega : |\lambda(t) \pm \alpha| < e\}$ and the assertion of the lemma follows. $\square$

The following lemma is cited from [10]. Its proof can be found there.

Lemma 3.5. *Let $f \geq 0 \in L_\infty(\Omega), \psi \geq 0 \in C^0[\underline{e}, \overline{e}]$ monotone increasing, $\varphi \geq 0 \in C^1[\underline{e}, \overline{e}]$ monotone decreasing and*

$$|\{t \in \Omega : f(t) > \varphi(e)\}| \leq \psi(e) \quad \forall\, \underline{e} \leq e \leq \overline{e}$$

Then

$$\int_\Omega f(t)\, dt \leq \psi(\underline{e})\|f\|_{L_\infty} + \int_{\underline{e}}^{\overline{e}} |\varphi'(e)\psi(e)|\, de + |\Omega|\varphi(\overline{e}).$$

Now we have gathered all auxiliary results necessary to show sharpened bounds on the length of the central path and the affine invariant Lipschitz constant.

Theorem 3.6. *Let Assumption 3.3 hold. Then there is $\mu_s > 0$, such that for all $\mu < \mu_s$ the slope η of the central path is bounded by*

$$\eta := \|v_\mu(\mu)\|_{v,\mu} + \|\lambda_\mu(\mu)\|_{H^2} \leq c\Gamma(1 - \ln \alpha\mu) + ce_0^{-1}, \tag{26}$$

and the length of the central path can be bounded by

$$\|y_* - y(\mu)\|_{L_2} + \|\lambda_* - \lambda(\mu)\|_{H^2} \leq c\mu(\Gamma(1 - \ln \alpha\mu) + e_0^{-1}) \leq \sqrt{\alpha\mu}. \tag{27}$$

Proof. We recall, that Theorem 3.1 bounds the slope of the central path in terms of $\|u_\mu\|_{L_1}$ and also provides the a-priori bound

$$\|\lambda_* - \lambda(\mu)\|_{L_\infty} \leq \tilde{c}\sqrt{\mu}. \tag{28}$$

By Lemma 2.1 $|u_\mu| = \frac{2u}{h_1} \leq \frac{2}{h_1}$ and thus Lemma 3.4 yields

$$\left|\left\{t \in \Omega : |u_\mu(v;\mu)(t)| > \frac{2}{e}\right\}\right| \leq \Gamma e$$

for $\underline{e} \leq e \leq e_0$. By (28) we can choose $\underline{e} \geq \tilde{c}\sqrt{\mu} \geq \|\lambda_* - \lambda(\mu)\|_{L_\infty}$. Now we use Lemma 3.5 with $\varphi(e) = 2e^{-1}$, $\psi(e) = \Gamma e$, and $\overline{e} = e_0$ to obtain

$$\begin{aligned}
\|u_\mu(v;\mu)(t)\|_{L_1} &\leq \psi(\underline{e})\|u_\mu(v;\mu)(t)\|_{L_\infty} + \int_{\underline{e}}^{\overline{e}} |\varphi'(e)\psi(e)|\, de + |\Omega|\varphi(\overline{e}) \\
&\leq \Gamma\underline{e}\|u_\mu\|_{L_\infty} + \int_{\underline{e}}^{e_0} \frac{2}{e^2}\Gamma e\, de + \frac{2|\Omega|}{e_0} \\
&\leq c\Gamma\left(\underline{e}\|h_1^{-1}\|_{L_\infty} + e_0^{-1}\right) + 2\Gamma \ln e\Big|_{\underline{e}}^{e_0} \\
&\leq c\Gamma(\underline{e}(\alpha\mu)^{-1/2} + e_0^{-1}) + 2\Gamma \ln\, e_0 - 2\Gamma \ln \underline{e} \\
&\leq c\Gamma(1 + \underline{e}(\alpha\mu)^{-1/2} - \ln \underline{e}^2) + ce_0^{-1} \\
&\leq c\Gamma(1 + \tilde{c}\sqrt{\mu}(\alpha\mu)^{-1/2} - \ln \tilde{c}\mu) + ce_0^{-1}
\end{aligned} \tag{29}$$

Thus Theorem 3.1 gives us

$$\|\lambda_\mu(\mu)\|_{L_\infty} \le \|v_\mu(\mu)\|_{v(\mu),\mu} \le c\|u_\mu\|_{L_1} \le c\Gamma(1+\alpha^{-1/2}-\ln\mu)+ce_0^{-1}.$$

Integration of this slope estimate yields:

$$\begin{aligned}\|\lambda_* - \lambda(\mu)\|_{L_\infty} &\le \int_0^\mu c\Gamma(1+\alpha^{-1/2}-\ln m)+ce_0^{-1}\,dm\\ &= c\Gamma m(1+\alpha^{-1/2}-\ln m)+mce_0^{-1}\Big|_0^\mu\\ &= c\mu(\Gamma(1+\alpha^{-1/2}-\ln\mu)+e_0^{-1}).\end{aligned}$$

Since the error decreases almost linearly in μ there exists a μ_s, such that for all $\mu \le \mu_s$ it follows

$$\|\lambda_* - \lambda(\mu)\|_{L_\infty} \le \sqrt{\alpha\mu}.$$

For such μ we can insert $\underline{e} = \sqrt{\alpha\mu}$ into (29) and obtain by the same argumentation as above (26) and especially

$$\|y_\mu(\mu)\|_{L_2} + \|\lambda_\mu(\mu)\|_{H^2} \le c\Gamma(1-\ln\alpha\mu)+ce_0^{-1}.$$

Integration of this slope estimate finally yields (27). □

We can also improve our Lipschitz estimates for small μ. The use of local norms is not necessary here.

Theorem 3.7. *There is a constant $c<\infty$, such that for $\mu \le \mu_s$ and*

$$\omega_s := c\left(\Gamma(1+\alpha^{-1})+\mu e_0^{-3}\right)$$

the Lipschitz condition

$$\|F_v^{-1}(v)\left(F_v(v+t\Delta v)-F_v(v)\right)\Delta v\|_{v,\mu} \le t\omega_s\|\Delta\lambda\|_{L_\infty}^2 \tag{30}$$

holds for all v, Δv with λ, $\lambda+\Delta\lambda$ in $S(\lambda(\mu),\|\cdot\|_{L_\infty},\sqrt{\alpha\mu})$.

Proof. The proof is similar to the proof of Theorem 3.2. However, now we apply a different estimate on

$$s = (u_\lambda(\lambda)-u_\lambda(\lambda+t\Delta\lambda))\Delta\lambda = \int_0^1 u_{\lambda\lambda}(\lambda+\Theta t\Delta\lambda)t\Delta\lambda\,d\Theta\,\Delta\lambda$$

and proceed using Theorem 2.3 to obtain for $\delta v := F_v^{-1}(v)\left(F_v(v+t\Delta v)-F_v(v)\right)\Delta v$

$$\begin{aligned}\|\delta v\|_{v,\mu} \le c\|s\|_{L_1} &\le c\int_0^1 \|u_{\lambda\lambda}(\lambda+\Theta t\Delta\lambda)\|_{L_1}d\Theta\, t\|\Delta\lambda\|_{L_\infty}^2\\ &\le c\sup_{\Theta\in[0,1]}\|u_{\lambda\lambda}(\lambda+\Theta t\Delta\lambda)\|_{L_1}t\|\Delta\lambda\|_{L_\infty}^2. \end{aligned} \tag{31}$$

Now we recall that $|u_{\lambda\lambda}| \le \frac{c\mu}{h_1^3}$, and by Theorem 3.6 that

$$\|\lambda_* - \lambda\|_{L_\infty} \le \|\lambda_* - \lambda(\mu)\|_{L_\infty} + \|\lambda(\mu) - \lambda\|_{L_\infty} \le 2\sqrt{\alpha\mu}$$

for $\mu < \mu_s$. Thus in Lemma 3.4 the choice of $\underline{e} := 2\sqrt{\alpha\mu}$ is possible and we obtain

$$\left|\left\{t \in \Omega : |u_{\lambda\lambda}(v;\mu)(t)| > \frac{c\mu}{e^3}\right\}\right| \le \Gamma e$$

for $2\sqrt{\alpha\mu} \le e \le e_0$. Using Lemma 3.5 with $\varphi(e) = c\mu e^{-3}$, $\psi(e) = \Gamma e$, $\underline{e} = 2\sqrt{\alpha\mu}$, $\overline{e} = e_0$ we conclude

$$\begin{aligned}
\|u_{\lambda\lambda}(\lambda;\mu)\|_{L_1} &\le \psi(\underline{e})\|u_{\lambda\lambda}(\lambda;\mu)\|_{L_\infty} + \int_{\underline{e}}^{\overline{e}} |\varphi'(e)\psi(e)|\,de + |\Omega|\varphi(\overline{e}) \\
&\le \Gamma\underline{e}c\mu\|h_1^{-1}\|_{L_\infty}^3 + \int_{\underline{e}}^{e_0} \frac{c\mu}{e^4}\Gamma e\,de + \frac{2\mu|\Omega|}{e_0^3} \\
&\le c\Gamma\underline{e}\mu(\alpha\mu)^{-3/2} - c\Gamma\frac{\mu}{e^2}\Big|_{\underline{e}}^{e_0} + \frac{2\mu|\Omega|}{e_0^3} \\
&\le c\left(\Gamma\sqrt{2\mu\alpha}\mu(\mu\alpha)^{-3/2} + \Gamma\left(\frac{\mu}{(2\sqrt{\mu\alpha})^2} - \frac{\mu}{e_0^2}\right) + \mu e_0^{-3}\right) \\
&\le c\Gamma(\alpha^{-1} + 1) + c\mu e_0^{-3}.
\end{aligned}$$

$\square$

Remark 3.8. The assertion in Theorem 3.7 can be formulated differently. With δv defined as in the proof we even have by Theorem 2.3

$$\|\delta y\|_{L_2} + \|\delta\lambda\|_{H^2} \le \omega_s\|\Delta\lambda\|_{L_\infty}^2$$

and Newton's method will converge also in the norm used on the left hand side of this estimate.

3.3 Estimates for a rapid starting phase

After our local analysis for small μ, we shortly discuss bounds for η and ω that are an improvement over Theorem 3.1 and Theorem 3.2 for large μ. These bounds reflect the fact, that for $\mu \to \infty$ the central path converges towards the analytical center $u \equiv 0$ of the admissible set.

Corollary 3.9. *The slope η of the central path can be bounded by*

$$\eta = \|v_\mu(\mu)\|_{v,\mu} + \|\lambda_\mu(\mu)\|_{H^2} \le c\frac{\|\lambda(\mu)\|_{L_1} + \alpha}{\mu^2}. \tag{32}$$

Moreover, the estimate (30) *in Theorem 3.7 also holds for*

$$\omega_s := c\frac{r + \alpha}{\mu^3} \tag{33}$$

for all v, $v + \Delta v$, such that $\lambda, \lambda + \Delta\lambda \in S(0, \|\cdot\|_{L_1}, r)$.

Proof. From Lemma 2.1 we know that $|u_\mu(t)| \le \frac{2u(t)}{2\mu}$. Furthermore (3) yields pointwise for feasible $u \in]-1;1[$:

$$|u| = \left|\frac{\lambda - \alpha u}{2\mu}(1-u^2)\right| \le \frac{|\lambda| + \alpha}{2\mu}$$

and thus

$$\|u_\mu\|_1 \le \frac{\|\lambda(\mu)\|_{L_1} + \alpha}{4\mu^2}$$

which yields (32) by Theorem 3.1. Considering (33) we use Lemma 2.1 again to obtain

$$|u_{\lambda\lambda}(t)| \le \frac{|u(t)|6\mu}{(2\mu)^3} \quad \Rightarrow \quad \|u_{\lambda\lambda}\|_1 \le \frac{\mu(\|\lambda\|_{L_1} + \alpha)}{2\mu(2\mu)^3}$$

inserting this bound into the estimate (31) in the proof of Theorem 3.7 yields the assertion. □

The consequence of our Corollary 3.9 is, that a properly controlled algorithm can reduce μ rapidly, if the initial value for μ was chosen larger than necessary.

4 A theoretical pathfollowing method

This section is devoted to the analysis of a pathfollowing method for solving (5). We assume for simplicity, that the Newton step in function space

$$\partial_v F(v_k; \mu_k)\delta v_k = -F(v_k; \mu) \tag{34}$$

can be computed exactly.

Algorithm 4.1.
select $\mu_0 > 0$, $\mu_s > 0$, $0 < \rho$, $0 < \sigma < 1$, $\tau > 0$,
    and v_0 with $\|v_0 - v(\mu_0)\|_{v_0,\mu_0} \le \rho\sqrt{\alpha\mu_0}$
For $k = 0, \ldots$
  solve (34) for δv_k
  $v_{k+1} = v_k + \delta v_k$
  if $\mu_k \ge \mu_s$
    $\mu_{k+1} = \sigma\mu_k$
  else
    $\mu_{k+1} = \tau(\Gamma(1 - \ln\alpha\mu_k) + e_0^{-1})^2\mu_k^2$

Remark 4.2. The algorithm sketched here is of course not designed for implementation in practice. Apart from the fact, that we cannot solve the Newton step without discretization error, several parameters cannot be determined a-priori for a particular problem. An efficient and robust algorithm will have to choose its stepsizes adaptively during the course of the homotopy and take into account the errors introduced by solving (34) inexactly. Our aim here is much more to explore the potential of such an algorithm by presenting a possible fixed choice of stepsizes that yields superlinear convergence.

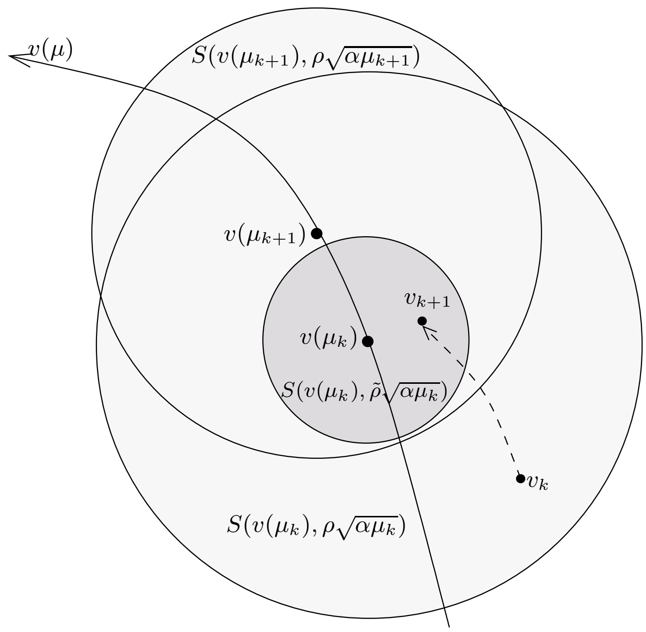

Figure 2: Neighbourhoods of the central path during Algorithm 4.1.

Our algorithm is constructed as follows: for $0 < \rho$ we choose $0 < \sigma < 1$, $\tau > 0$, and $\tilde{\rho} := \sqrt{\sigma}\rho/4$, such that one Newton step maps

$$v_k \in S(v(\mu_k), \rho\sqrt{\alpha\mu_k}) \quad \longrightarrow \quad v_{k+1} \in S(v(\mu_k), \tilde{\rho}\sqrt{\alpha\mu_k})$$

and after a reduction of μ the inclusion

$$v_{k+1} \in S(v(\mu_{k+1}), \rho\sqrt{\alpha\mu_{k+1}}) \supset S(v(\mu_k), \tilde{\rho}\sqrt{\alpha\mu_k})$$

holds. This mechanism is depicted graphically in Figure 2.

In the remainder of the section we show that there is a suitable choice of parameters, such that Algorithm 4.1 is well defined and computes iterates that converge to the solution $v(0)$. For this purpose we will use the estimates derived in Section 3, that mainly depend on α, μ, e_0, and Γ.

Since we are interested in small α and our estimates are not uniform for very large values of this parameter we bound α a-priori by $\alpha_{\max}$. The generic constants we use may depend on $\alpha_{\max}$.

First, we analyse the effect of a μ-reduction by a constant factor σ, which is technically involved due to the necessary use of local norms.

Lemma 4.3. *There is a constant $\bar{c}$, such that for $0 < \rho \leq 1$ and*

$$\sigma \geq \max\left\{\left(1 + \frac{\rho\sqrt{\alpha}}{8\bar{c}}\right)^{-2}, \frac{1}{4}\right\}. \tag{35}$$

we can conclude

$$\|v - v(\mu)\|_{v,\mu} \le \tilde{\rho}\sqrt{\alpha\mu} \quad \Longrightarrow \quad \|v - v(\sigma\mu)\|_{v,\sigma\mu} \le \rho\sqrt{\alpha\sigma\mu}$$

with $\tilde{\rho} = \sqrt{\sigma}\rho/4$ *defined as above.*

Proof. By the triangle inequality and (17) we derive

$$\begin{aligned}
\|v - v(\sigma\mu)\|_{v,\sigma\mu} &\le \|v - v(\mu)\|_{v,\sigma\mu} + \|v(\mu) - v(\sigma\mu)\|_{v,\sigma\mu} \\
&\le \|v - v(\mu)\|_{v,\mu}\left(1 + \frac{1}{\sqrt{\alpha\sigma\mu}}|\mu - \sigma\mu|\right) + \left\|\int_{\sigma\mu}^{\mu} v_\mu(m)dm\right\|_{v,\sigma\mu} \\
&\le \tilde{\rho}\sqrt{\alpha\mu}\left(1 + \frac{\sigma^{-1}-1}{\sqrt{\alpha}}\right) + \int_{\sigma\mu}^{\mu} \|v_\mu(m)\|_{v,\sigma\mu}\, dm. \qquad (36)
\end{aligned}$$

Using (21) we conclude for $\mu \ge m \ge \sigma\mu$ and $\tilde{c}$ as in Theorem 3.1 that

$$\|\lambda - \lambda(m)\|_{L_\infty} \le \|\lambda - \lambda(\mu)\|_{L_\infty} + \|\lambda(\mu) - \lambda(m)\|_{L_\infty} \le \tilde{\rho}\sqrt{\alpha\mu} + \tilde{c}(1 - \sqrt{\sigma})\sqrt{\mu}.$$

Thus the integrand can be estimated via (16) and (17) from Corollary 2.7, and (19) from Theorem 3.1:

$$\begin{aligned}
\|v_\mu(m)\|_{v,\sigma\mu} &\le \left(1 + \frac{2}{\sqrt{\alpha m}}\|\lambda - \lambda(m)\|_{L_\infty}\right)\|v_\mu(m)\|_{v(m),\sigma\mu} \\
&\le \left(1 + \frac{2\sqrt{\mu}}{\sqrt{m}}\left(\tilde{\rho} + \tilde{c}\frac{1-\sqrt{\sigma}}{\sqrt{\alpha}}\right)\right)\left(1 + \frac{1}{\sqrt{\alpha m}}|\mu - m|\right)\|v_\mu(m)\|_{v(m),m} \\
&\le \left(1 + \frac{2\tilde{\rho}}{\sqrt{\sigma}} + 2\tilde{c}\frac{\sigma^{-1/2}-1}{\sqrt{\alpha}}\right)\left(1 + \frac{\sigma^{-1}-1}{\sqrt{\alpha}}\right)\|v_\mu(m)\|_{v(m),m}. \qquad (37)
\end{aligned}$$

Now we use Assumption (35) to bound the factors in parentheses that appear in (36) and (37). Setting $\bar{c} := \max\{\tilde{c}, 1\}$ we derive

$$\begin{aligned}
\tilde{c}\alpha^{-1/2}(\sigma^{-1/2} - 1) &\le \rho/8 \le 1/8, \qquad (38) \\
\sigma^{-1/2} + 1 &\le (1/4)^{-1/2} + 1 \le 3, \\
\alpha^{-1/2}(\sigma^{-1} - 1) &= \alpha^{-1/2}(\sigma^{-1/2} - 1)(\sigma^{-1/2} + 1) \le \frac{\rho}{8\bar{c}} \cdot 3 \le \frac{3}{8}, \qquad (39)
\end{aligned}$$

and thus inserting these inequalities and the definition of $\tilde{\rho}$

$$\begin{aligned}
\int_{\sigma\mu}^{\mu} \|v_\mu(m)\|_{v,\sigma\mu}\, dm &\le \int_{\sigma\mu}^{\mu}\left(1 + \frac{1}{2} + \frac{1}{4}\right)\left(1 + \frac{3}{8}\right)\|v_\mu(m)\|_{v(m),m}\, dm \\
&\le \frac{5}{2}\int_{\sigma\mu}^{\mu}\frac{\tilde{c}}{\sqrt{m}}dm \le 5\,\tilde{c}(1 - \sqrt{\sigma})\sqrt{\mu} \\
&\le 5\,\tilde{c}\frac{\sigma^{-1/2}-1}{\sqrt{\alpha}}\sqrt{\sigma\alpha\mu}.
\end{aligned}$$

Inserting this estimate and (39) into (36) we obtain

$$\|v - v(\sigma\mu)\|_{v,\sigma\mu} \leq \frac{11}{8}\tilde{\rho}\sqrt{\alpha\mu} + \frac{5}{8}\rho\sqrt{\sigma\alpha\mu} \leq \left(\frac{11}{32} + \frac{5}{8}\right)\rho\sqrt{\sigma\alpha\mu} \leq \rho\sqrt{\alpha\sigma\mu}.$$

□

Theorem 4.4. *There exists* $0 < \rho < 1$ *independent of* μ, α, e_0, *and* Γ, *such that if* σ *satisfies* (35) *and*

$$\|v_0 - v(\mu_0)\|_{v_0,\mu_0} \leq \rho\sqrt{\alpha\mu_0}$$

Algorithm 4.1 is well defined and converges linearly:

$$\|\lambda_k - \lambda_*\|_{H^2} + \|y_k - y_*\|_{L_2} \leq c\sqrt{\mu_k}. \tag{40}$$

If Assumption 3.3 holds, then there are $\mu_s > 0$ *and* τ *independent of* μ, *such that Algorithm 4.1 is of "r-order 2" for* $\mu < \mu_s$*:*

$$\lim_{k\to\infty} \frac{\ln \mu_{k+1}}{\ln \mu_k} = 2 \tag{41}$$

and

$$\|\lambda_{k+1} - \lambda_*\|_{H^2} + \|y_{k+1} - y_*\|_{L_2} \leq c\mu_k(\Gamma(1 - \ln \alpha\mu_k) + e_0^{-1}) \tag{42}$$

Proof. Assume for the purpose of induction that

$$\|v_k - v(\mu_k)\|_{v_k,\mu_k} \leq \rho\sqrt{\alpha\mu_k}.$$

Then by Theorem 2.5 using the estimates for γ in Lemma 2.7 and for ω in Theorem 3.2 we have

$$\begin{aligned}
\|v_{k+1} - v(\mu_k)\|_{v_{k+1},\mu_k} &\leq (1 + \gamma\|v_k - v(\mu_k)\|_{v_k,\mu_k})^3\, \omega_g/2\|v_k - v(\mu_k)\|^2_{v_k,\mu_k} \\
&\leq \left(1 + 2(\alpha\mu_k)^{-1/2}\rho\sqrt{\alpha\mu_k}\right)^3 \frac{c}{2\sqrt{\alpha\mu_k}}\rho^2\alpha\mu_k \\
&\leq c\,(1 + 2\rho)^3\, \frac{1}{2}\rho^2\sqrt{\alpha\mu_k}.
\end{aligned}$$

Since ρ appears quadratically for sufficiently small ρ independent of α and μ it holds

$$\|v_{k+1} - v(\mu_k)\|_{v_{k+1},\mu_k} \leq \tilde{\rho}\sqrt{\alpha\mu_k}.$$

Now Lemma 4.3 completes the induction, showing that after a reduction of μ

$$\|v_{k+1} - v(\mu_{k+1})\|_{v_{k+1},\mu_{k+1}} \leq \rho\sqrt{\alpha\mu_{k+1}}.$$

Equation (40) follows by the triangle inequality and (20) (with $\sigma\mu = 0$) in Theorem 3.1.

If Assumption 3.3 holds, Theorem 3.6 states the existence of a constant μ_s, such that for $\mu < \mu_s$ refined estimates hold. We assume $\mu_s \leq e_0^2\alpha^{-1}$ and consider $\mu < \mu_s$.

From Phase 1 of the algorithm we know that $\|\lambda_k - \lambda(\mu_k)\|_{L_\infty} \le \rho\sqrt{\alpha\mu_k}$. Since our estimate for ω_s in Theorem 3.7 makes no use of local norms we can set $\gamma = 0$ in Theorem 2.5 and conclude

$$\begin{aligned}\|\lambda_{k+1} - \lambda(\mu_k)\|_{L_\infty} &\le \|v_{k+1} - v(\mu_k)\|_{v,\mu} \le \frac{\omega_s}{2}\|\lambda_k - \lambda(\mu_k)\|_{L_\infty}^2 \\ &\le c\frac{\Gamma(\alpha^{-1}+1) + \mu_k e_0^{-3}}{2}\rho^2\alpha\mu_k \le c(\Gamma(1+\alpha) + \alpha\mu_k e_0^{-3})\,\rho^2\mu_k \\ &\le c(\Gamma(1+\alpha) + e_0^{-1})\,\rho^2\mu_k.\end{aligned}$$

Consequently for sufficiently small ρ we have

$$\|\lambda_{k+1} - \lambda(\mu_k)\|_{L_\infty} \le c(\Gamma + e_0^{-1})\rho\mu_k.$$

Moreover, considering Remark 3.8 a closer look on Theorems 2.3, 3.2 and 2.5 yields that Newton's method also converges in the H^2-norm for λ and thus

$$\|\lambda_{k+1} - \lambda(\mu_k)\|_{H^2} + \|y_{k+1} - y(\mu_k)\|_{L_2} \le c(\Gamma + e_0^{-1})\mu_k. \tag{43}$$

Estimating the error after a μ-reduction we compute

$$\begin{aligned}\|\lambda_{k+1} - \lambda(\mu_{k+1})\|_{L_\infty} &\le \|\lambda_{k+1} - \lambda(\mu_k)\|_{L_\infty} + \|\lambda(\mu_k) - \lambda(\mu_{k+1})\|_{L_\infty} \\ &\le c\mu_k(\Gamma + e_0^{-1})\rho + c\mu_k(\Gamma(1 - \ln\alpha\mu_k) + ce_0^{-1}) \\ &\le \rho\sqrt{\alpha}\left((c\alpha^{-1/2}\rho^{-1})\mu_k\left(\Gamma(1-\ln\alpha\mu_k) + e_0^{-1}\right)\right).\end{aligned}$$

Now we notice, that there is a constant c, such that for $\tau := c^2\alpha^{-1}\rho^{-2}$ we can continue with

$$\begin{aligned}\|\lambda_{k+1} - \lambda(\mu_{k+1})\|_{L_\infty} &\le \rho\sqrt{\alpha\tau\mu_k^2(\Gamma(1-\ln\alpha\mu_k) + e_0^{-1})^2} \\ &\le \rho\sqrt{\alpha\mu_{k+1}},\end{aligned}$$

which completes the induction. The estimate (42) follows from (43) together with (27) in Theorem 3.6. Finally, relation (41) holds for our update rule $\mu_k \to \mu_{k+1}$, since

$$\lim_{x\to 0}\frac{\ln x^2(c-\ln x)^2}{\ln x} = 2\frac{\ln x + \ln(c - \ln x)}{\ln x} = 2.$$

□

Remark 4.5. Since ρ can be chosen independently of α and μ, assumption (35) requires $1 - \sigma = \mathcal{O}(\sqrt{\alpha})$ for small α whereas $\tau = \mathcal{O}(\alpha^{-1})$. These are worst case bounds for the efficiency of Algorithm 4.1.

5 Numerical experiments

We present some numerical results to illustrate the theory developed in the previous sections. We implemented a prototype version of a control reduced interior point algorithm in MATLAB in order to obtain computational estiamtes of the quanities relevant for the progress of the homotopy. We have implemented a simple adaptive step size strategy, which already shows superlinear convergence behavior.

We considered the numerical solution of (1) on $\Omega =]0,1[\times]0,1[$, with $L = -\Delta$ and piecewise constant $y_d \in L_2(\Omega)$. Defining χ_M as the characteristic function of $M \subset \Omega$, we used

$$y_d := a\left(4\chi_{[0,3/4]\times[0,1/2]} - 10\chi_{[0,3/4]\times[1/2,1]} - 2\chi_{[3/4,1]\times[0,1/2]} + 50\chi_{[3/4,1]\times[1/2,1]}\right)$$

for different values of a. This unsymmetric choice prevents solutions from getting a too simple structure. A general observation was, that problems with both small α and small residual $\|y - y_d\|$ were more expensive to compute than problems where one of these quantities is large.

We present estimates for the slope η of the central path and for the Lipschitz constant ω. The slope was approximated by the finite differences

$$\eta(\mu_k; \mu_{k+1}) \approx \frac{\|v(\mu_k) - v(\mu_{k+1})\|_{v(\mu_{k+1}),\mu_{k+1}}}{\mu_k - \mu_{k+1}}.$$

The Lipschitz constant ω was estimated by monitoring the convergence behaviour of Newtons method similar to the methodology presented in [1]. Both estimates have to be used with care in the presence of round-off errors. In order to obtain smooth graphs in our diagrams we restrict the step size σ to $\sigma \geq 0.1$.

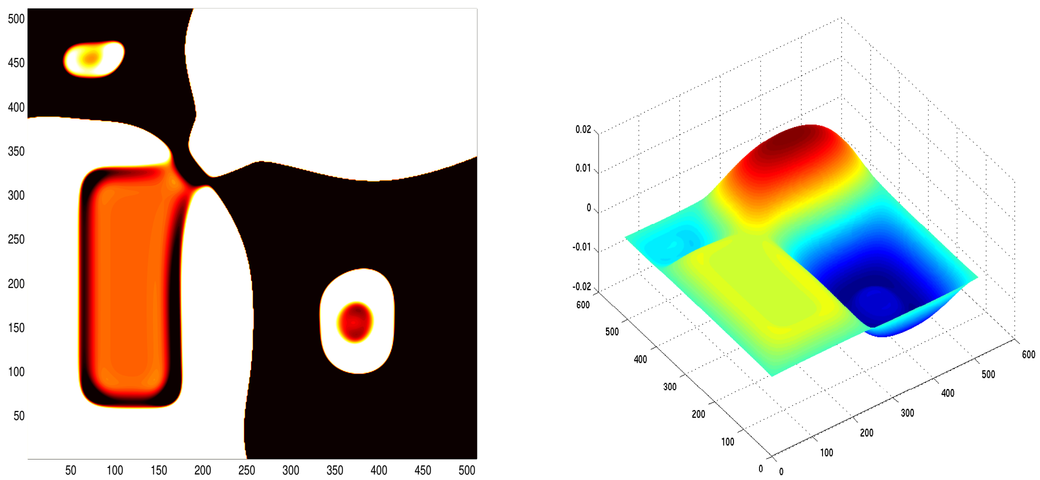

Figure 3: Control u and state y for $\alpha = 10^{-8}$ and $a = 0.001$.

In our first test set we consider $a = 0.001$ chosen, such that e_0 is relatively small since $|\lambda|$ is small on a large portion of Ω, furthermore we set $\alpha = 10^{-8}$. Optimal control and state are depicted in Figure 3. We see, that there are large active sets but also large areas where the control lies in the interior of the feasible region.

Considering Figure 4 we observe that our estimates for η and ω are largely independent of the mesh sizes $h = 2^{-k}$ with $k = 6 \dots 9$ we have used. Moreover, we can identify three phases. For large μ the quantities $\eta \sim \mu^{-2}$ and $\omega \sim \mu^{-3}$ are very small and increase rapidly. This is the "rapid starting phase". Then a phase of moderate increase starts: the "global phase". Here $\omega, \eta \sim \mu^{-1/2}$ and constant stepsizes σ will have to be taken. In difficult examples this will be the phase, where most of the steps are taken. Finally, the increase in η and ω stagnates almost entirely. We have reached the "local phase of superlinear convergence". Here the estimated values of η and ω for different h diverge slightly, but clearly behave very similarly. If the restriction on σ is dropped, this phase usually consists of not more that three or four steps.

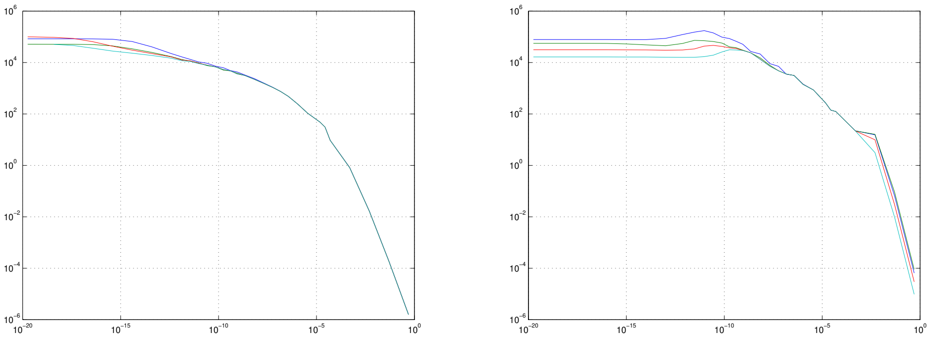

Figure 4: Logarithmic plot of estimates for η (left) and ω (right) with respect to μ for several mesh sizes.

Next we perform test runs for $a = 0.001$ and several choices of α, namely $\alpha = 10^{-2k}$ with $k = 1 \dots 6$ at constant mesh-size $h = 1/128$. In Figure 5 we see, that the phase of superlinear convergence starts later with larger final values of η and ω as α becomes smaller. This may be a consequence of e_0 becoming smaller for smaller α.

In a second example we set $a = 0.1$ which leads to a large e_0. Now we observe, that the neither the slope η, nor the starting point μ_s depend significantly on α. Moreover, we see that the starting phase immediately blends into the phase of local superlinear convergence. The Lipschitz constant ω seems to be very small in this example, and we cannot give any reliable estimates for numerical reasons as indicated above. If the restriction on σ is dropped, the algorithm takes 5 μ-reduction steps to reduce μ from 10 to 10^{-15}.

Conclusion

We have shown robust global and fast local convergence of the control reduced primal interior point method for PDE constrained optimization. Its properties make

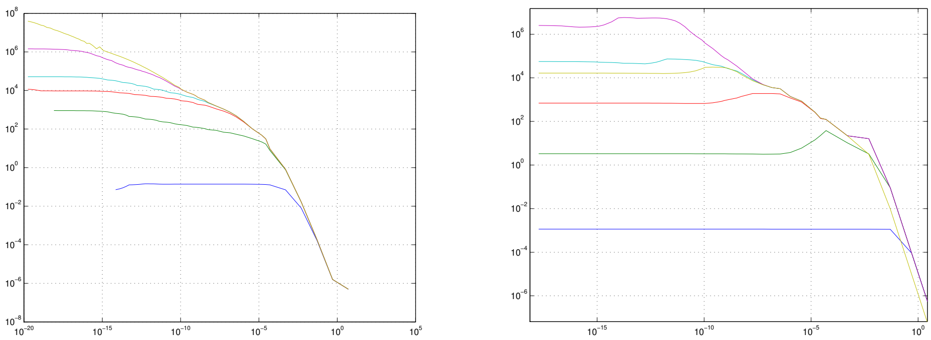

Figure 5: Logarithmic plot of estimates for η (left) and ω (right) with respect to μ for several choices of α and small a. The smaller α is chosen, the larger the final value of the estimates.

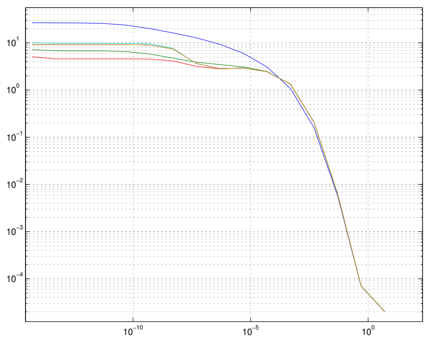

Figure 6: Logarithmic plot of estimate for η for several choices of α and large a

it an excellent basis for an efficient algorithm in optimal control. Clearly our algorithm in function space has to be complemented by both an efficient and adaptive discretization scheme and stepsize control. The development of these features is subject of ongoing work.

References

[1] P. Deuflhard. *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms.* Springer, 2003.

[2] P. Deuflhard and F.A. Potra. Asymptotic mesh independence of Newton-Galerkin methods via a refined Mysovskii theorem. *SIAM J. Numer. Anal.*, 29(5):1395–1412, 1992.

[3] M. Hintermüller and K. Kunisch. Path-following methods for a class of constrained minimization problems in function space. Technical report, Karl-Franzens-Universität Graz, Juli 2004.

[4] M. Hinze. A variational discretization concept in control constrained optimization: the linear-quadratic case. *COMOA*, 30:45–63, 2005.

[5] J. L. Lions. *Optimal control of systems governed by partial differential equations*, volume 170 of *Grundlehren der methematischen Wissenschaften*. Springer-Verlag, 1973.

[6] H. Maurer and H. D. Mittelmann. Optimization techniques for solving elliptic control problems with control and state constraints. i: Boundary control. *Comp. Optim. Applic.*, 16:29–55, 2000.

[7] H. Maurer and H. D. Mittelmann. Optimization techniques for solving elliptic control problems with control and state constraints. part 2: Distributed control. *Comp. Optim. Applic.*, 18:141–160, 2001.

[8] C. Meyer and A. Rösch. Superconvergence properties of optimal control problems. *SIAM J. Control Optim.*, to appear.

[9] 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, ZIB, 2004.

[10] A. Schiela and M. Weiser. Interior point methods in function space for bang-bang control. in preparation, Zuse Institute Berlin, 2005.

[11] M. Ulbrich. Semismooth Newton methods for operator equations in function spaces. *SIAM J. Optim.*, 13:805–842, 2003.

[12] M. Ulbrich and S. Ulbrich. Superlinear convergence of affine-scaling interior-point Newton methods for infinite-dimensional nonlinear problems with pointwise bounds. *SIAM J. Control Optim.*, 38(6):1938–1984, 2000.

[13] S Ulbrich. Multilevel primal-dual interior point methods for PDE-constrained optimization. Talk at the Workshop: Control of PDEs at TU Berlin, December 2003.

[14] M. Weiser. Interior point methods in function space. ZIB Report 03-35, Zuse Institute Berlin, 2003.

[15] M. Weiser, T. Gänzler, and A. Schiela. A control reduced primal interior point method for pde constrained optimization. ZIB Report 04-38, Zuse Institute Berlin, 2004.

[16] M. Weiser and A. Schiela. Function space interior point methods for PDE constrained optimization. ZIB Report 04-27, Zuse Institute Berlin, 2004.

[17] S.J. Wright. *Primal-Dual Interior-Point Methods*. SIAM, Philadelphia, 1997.