Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

MARTIN WEISER PETER DEUFLHARD

The Central Path towards the Numerical Solution of Optimal Control Problems

ZIB-Report 01-12 (July 2001)

The Central Path towards the Numerical Solution of Optimal Control Problems

Martin Weiser   Peter Deuflhard

19th July 2001

Abstract

A new approach to the numerical solution of optimal control problems including control and state constraints is presented. Like hybrid methods, the approach aims at combining the advantages of direct and indirect methods. Unlike hybrid methods, however, our method is directly based on interior-point concepts in function space — realized via an adaptive multilevel scheme applied to the complementarity formulation and numerical continuation along the central path. Existence of the central path and its continuation towards the solution point is analyzed in some theoretical detail. An adaptive stepsize control with respect to the duality gap parameter is worked out in the framework of affine invariant inexact Newton methods. Finally, the performance of a first version of our new type of algorithm is documented by the successful treatment of the well-known intricate windshear problem.

Keywords: numerical optimal control, interior point methods in function space, affine invariance

MSC (2000): 65K10, 49M15, 90C48, 90C51

Contents

1 Introduction

In the last decade, the numerical solution of optimal control problems has reached a high level of sophistication. Present methods are able to treat important classes of large scale real life problems in science and engineering. Two types of methods are in common use: (a) *direct methods*, mostly based on some *robust collocation* including an ad hoc parametrization of the controls (see Bock and Plitt [7]), and (b) *indirect methods*, typically based on either *multiple shooting* techniques (see Bulirsch [9], Stoer and Bulirsch [27], Deuflhard [13, 14], Bock [6]) or *adaptive collocation methods* (see Ascher, Christiansen, and Russell [3], Ascher and Bader [2], Ascher, Mattheij, and Russell [4]). Whenever the necessary Euler-Lagrange conditions give a sufficient description of the problem, then indirect methods lead to a provably optimal solution [9]. However, they require rather detailed a-priori knowledge about the sequence of optimal subtrajectories. In contrast to that, direct methods may dispense of this severe constraint, but have a tendency to lead to non optimal solutions now and then. For this reason, *hybrid methods* seem to be the state of the art (see von Stryk and Bulirsch [29], Bulirsch, Nerz, Pesch, and von Stryk [12]): in a first step, some direct method, wherein the control variables are parameterized ad hoc, supplies a rough idea about the optimal subarc sequence; then, in the second step, an indirect method is employed to finally solve the problem to high accuracy.

The present paper advocates a *unified function space approach* realizing ideas of both direct and indirect methods in infinite dimension rather than in finite dimension — as opposed to the above hybrid methods. There are several possibilities of such an extension. In a recent monograph Pytlak [25] proposed an approach that he claims to be a function space approach. However, in our above wording, that approach is of the type robust collocation, leaving a considerable gap between the presented theory and the rather heuristic algorithmic realization. A genuine function space SQP method has been suggested by Alt and Malanowski [1]. However, implementing a finite dimensional interior point QP solver within their proposed method would involve the solution of a sequence of discrete problems with duality gap parameter $\mu \to 0$ on each of the successively finer grids. Such a procedure would suffer from severe difficulties in the presence of C^0-discontinuities of the control variables.

The present paper advocates a function space approach realized as a *nested* reduction of mesh size and duality gap parameter. The extension of interior point type methods from finite to infinite dimension is not at all straightforward: after all, the concept of logarithmic barrier functions is no longer useful in the infinite dimensional setting (cf. Jarre [20]). Fortunately, the *complementarity* version of IP methods, including the *central path* concept, carries over naturally to infinite dimension. However, some careful theoretical consideration is needed: the well-known two-norm discrepancy (see Maurer [22] and Malanowski [21]) strongly advises the use of different norms for handling differentiability on one hand and convexity on the other hand.

The paper is organized as follows. In Section 2, we turn to the central path in function space as the mathematical concept and derive the main theoretical framework. In Section 3, important details of an algorithm on this theoretical basis are worked out. Affine invariant norms are used to control the iteration process towards the numerical solution (see VOLKWEIN and WEISER [28], POTRA [24], or more generally, DEUFLHARD [15]). Our computational approach actually *exploits function space* via an *adaptive multilevel refinement* of all variables *including the controls*. Its present realization is done within the setting of collocation methods. In this context it seems worth mentioning, that our herein suggested method differs clearly from comparable finite dimensional multigrid techniques by SCHULZ [26], wherein the adaptive refinement of the *control variables* is done in the *outer loop*, as opposed to our infinite dimensional technique, wherein the refinement is performed in the *innermost loop*. An adaptive stepsize control along the central path is worked out as the infinite dimensional extension of finite dimensional suggestions due to DEUFLHARD [14]. Our present version of the algorithm is working satisfactorily, but far from optimized concerning the discretization and solution of linear subproblems. Nevertheless, in Section 4, we are able to document the successful solution of a well-known intricate optimal control problem, the abort landing in the presence of windshear (cf. MIELE et al. [23], BULIRSCH, MONTRONE, and PESCH [10, 11]). Even though our approach does not need any cumbersome analytic preparation which is required for any indirect method, our numerical results are in full agreement with those obtained by multiple shooting [11].

2 Central Path Continuation in Function Space

We consider the abstract optimization problem

$$\min J(x) \quad \text{subject to} \quad \begin{aligned} c(x) &= 0 \\ g(x) &\geq 0 \end{aligned} \tag{2.1}$$

where J is a functional defined on a real Banach space X and c, g are mappings from X into real Banach spaces Z_c and Z_g, respectively. The partial ordering on Z_g is assumed to be induced by a closed convex cone K. In order to fix the theoretical frame, we first collect the basic necessary and sufficient conditions for an optimal solution of such a problem — see [22]. Let $\langle \cdot, \cdot \rangle$ denote the dual pairing associated with Z_g and $Z_g^\star$. Throughout the paper we tacitly assume that the given problem has a locally unique solution. If not so, the algorithms to be designed should be able to detect any local non-uniqueness.

Theorem 2.1 (Necessary conditions). *Let a solution x of problem* (2.1) *be regular, i.e.* $0 \in \operatorname{int}(g(x) + g'(x)X - K)$, *and define the Lagrangian*

$$L(x, \lambda, \eta) = J(x) - \langle \lambda, c(x) \rangle - \langle \eta, g(x) \rangle .$$

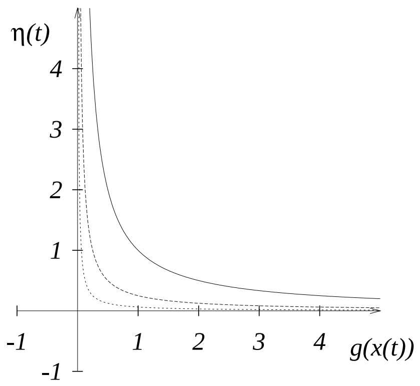

Figure 2.1: Zero level sets of the smoothed complementarity function (2.9) for $\mu \in \{1, 1/4, 1/16\}$.

Then there are $\lambda \in Z_c^\star$ and $\eta \in Z_g^\star$, $\eta \geq 0$, such that

$$\partial_x L(x, \lambda, \eta) = 0 \tag{2.2}$$

$$c(x) = 0 \tag{2.3}$$

$$\langle \eta, g(x) \rangle = 0\,. \tag{2.4}$$

In many cases of real life applications, these necessary conditions already define a unique optimal solution. However, there exist well-known counterexamples that require more than just the necessary conditions.

Theorem 2.2 (Sufficient conditions). *Suppose x is a regular point for the abstract problem* (2.1)*. Assume that J, c, and g are defined and twice continuously Fréchet differentiable on some larger space $X_p \supset X$. Assume x, λ, and η satisfy* (2.2–2.4)*. Suppose that there are $\delta > 0$ and $\beta > 0$ such that*

$$\langle \partial_x^2 L(x, \lambda, \eta) h, h \rangle \geq \delta \|h\|_p^2$$

for all $h \in \ker c'(x)$ with $g'(x)h \in K + \operatorname{span}\{g(x)\}$ and $\langle \eta, g'(x)h \rangle \leq \beta \|h\|_p$. Then there exists a neighborhood U of x in X_p such that x is the unique local solution of (2.1) *in U.*

For optimal control problems as studied here, equation (2.4) is equivalent to *pointwise complementarity* in the sense that

$$\eta(t) g(x)(t) = 0, \quad \eta(t) \geq 0, \quad g(x) \geq 0 \quad \text{for almost all } t. \tag{2.5}$$

The idea of *interior point methods* is to replace the unwieldy complementarity condition (2.5) by a relaxed substitute condition of the type

$$\eta(t) g(x)(t) = \mu \quad \text{for } \mu > 0\,, \tag{2.6}$$

wherein μ is the *duality gap* parameter. The connection between (2.5) and (2.6) is via a homotopy with respect to μ, the so-called *central path.* For the central path to be well-defined, the additional *feasibility condition*

$$\eta(t) \geq 0, \quad g(x)(t) \geq 0 \tag{2.7}$$

has to be satisfied. For $\mu \to 0$ we arrive at the condition (2.5) — see Figure 2.1.

So-called *complementarity methods* permit *infeasible* iterates as well replacing condition (2.5) by a condition of the type

$$\psi(g(x), \eta; \mu) = 0\,, \tag{2.8}$$

where the feasibility of the central path is guaranteed by the construction of ψ. Thus, infeasible intermediate iterates can be accepted — a feature, which increases the overall robustness of the method. Throughout the paper we specify ψ to be the so-called FISCHER-BURMEISTER function [18]

$$\psi(a, b; \mu) = a + b - \sqrt{a^2 + b^2 + 2\mu}\,, \tag{2.9}$$

the zero level set of which is characterized by the interior point conditions (2.6) and (2.7). Upon introducing slack variables w, we arrive at the canonical formulation

$$F(x, \lambda, \eta, w; \mu) = \begin{bmatrix} \partial_x L(x, \lambda, \eta) \\ -c(x) \\ w - g(x) \\ \psi(w, \eta; \mu) \end{bmatrix} = 0\,, \tag{2.10}$$

which will be the formulation to be actually solved numerically. Assuming sufficient differentiability, the associated derivative has the form

$$F'(x, \lambda, \eta, w; \mu) = \begin{bmatrix} \partial_x^2 L(x, \lambda, \eta) & -c'(x)^\star & -g'(x)^\star & \\ -c'(x) & & & \\ -g'(x) & & & I \\ & & \partial_\eta \psi(w, \eta; \mu) & \partial_w \psi(w, \eta; \mu) \end{bmatrix},$$

which clearly reveals the saddle point structure of the abstract optimization problem (2.1).

For ease of writing, we introduce the extended variables $v = (x, \lambda, \eta, w)$, with a typical splitting of the unknowns $x = (u, y)$ into control variables u and state variables y. In this notation, let now J, c, and g be specified as

$$J(x) = \int_0^1 f(x(t))\, dt\,,$$

$$c(x) = \begin{bmatrix} \bar{c}(x) \\ \mathfrak{c}_r(y(0), y(1)) \end{bmatrix}, \quad g(x)(t) = \begin{bmatrix} \mathfrak{g}_u(u(t)) \\ \mathfrak{g}_y(y(t)) \end{bmatrix}.$$

Variable	Dimension	Space
u	n_u	$L_p^{n_u}$
y	n_y	$(W_p^1)^{n_y}$
x	$n_u + n_y$	$X_p = L_p^{n_u} \times (W_p^1)^{n_y}$
λ	$n_\lambda + n_r$	$\Lambda_p = L_p^{n_\lambda} \times \mathbb{R}^{n_r}$
η_u	$\bar{n}_u$	$L_p^{\bar{n}_u}$
η_y	$\bar{n}_y$	$(W_p^1)^{\bar{n}_y}$
η	$\bar{n}_u + \bar{n}_y$	$E_p = L_p^{\bar{n}_u} \times (W_p^1)^{\bar{n}_y}$
v	n_v	V_p
		$\hat{X}_p^\star = L_p^{n_u} \times (W_p^{-1})$
		$Z_p = \hat{X}_p^\star \times \Lambda_p^\star \times E_p \times E_p$
$\mathfrak{c}$	$\mathbb{R}^{n_u+2n_y} \to \mathbb{R}^{n_\lambda}$	
$\mathfrak{c}_r$	$\mathbb{R}^{2n_y} \to \mathbb{R}^{n_r}$	

Table 2.1: Variables, their dimension, and the associated function spaces.

wherein $\bar{c}(x)(t) = \mathfrak{c}(u(t), y(t), \dot{y}(t))$ contains the ordinary differential equations as equality constraints. For the purpose of survey, a list of variables, dimensions, and associated spaces is given in Table 2.1.

The actual existence of a central path $v(\mu)$ for the above general mapping $F(v(\mu); \mu) = 0$ is harder to prove than in the finite dimensional setting. The proof is rather technical and therefore omitted here; it can be found in Weiser [30]. For the purpose of the present paper we merely follow the basic lines of that derivation and collect the essential results. The derivation proceeds in two steps: first, local continuation into an open set, which includes $\mu > 0$; second, continuation within a closed set up to $\mu = 0$.

The pointwise complementarity condition (2.5) enforces $\eta, w \in L_\infty$, hence we have to choose $p = \infty$, which implies $X = X_\infty$ and $V = V_\infty$.

Theorem 2.3. *Let f, $\mathfrak{c}$, $\mathfrak{c}_r$, and $\mathfrak{g}$ be twice Lipschitz-continuously differentiable with respect to their arguments, and $\mu > 0$. Then the mapping $F : V_\infty \times \mathbb{R}_+ \to Z$ is a continuosly differentiable mapping. Its derivative F' satisfies the Lipschitz condition*

$$\|F'(v; \mu) - F'(\bar{v}; \mu)\|_{V_\infty \to Z_\infty} \leq \mathrm{const}\, \mu^{-1} \|v - \bar{v}\|_{V_\infty} .$$

The above result clearly indicates numerical difficulties to be expected in the continuation process as $\mu \to 0$. In order for any Newton-type continuation method to work, we need to verify that the derivative is not only Lipschitz continuous, but also has a bounded inverse.

Theorem 2.4. *Assumptions of Theorem 2.3. Let $D \subset V$ be an open bounded set and let $\mu_0 > 0$ be a sufficiently small duality gap parameter. Assume that for $v \in D$ and $0 < \mu \leq \mu_0$ the following conditions hold uniformly:*

1. *The linearized state equation is solvable: There exists a constant $\beta > 0$ such that for every constraints variation $\delta \in L_p^{n_\lambda} \times \mathbb{R}^{n_r}$ there is a state variation $\delta y \in (W_p^1)^{n_y}$ with*
$$c_y(x)\delta y = \delta \quad \text{and} \quad \|\delta y\|_{(W_p^1)^{n_y}} \leq \beta \|\delta\|_{L_p^{n_\lambda} \times \mathbb{R}^{n_r}}$$
for $p = 1, 2, \infty$.

2. *The state equation $\mathfrak{c}(u(t), y(t), \dot{y}(t)) = 0$ is linear in $\dot{y}(t)$.*

3. *The strengthened Legendre-Clebsch condition holds: for*
$$\begin{aligned} H(t) := {} & f''(x(t)) - \mathfrak{c}''(u(t), y(t), \dot{y}(t))^T \lambda(t) - \mathfrak{g}''(x(t))^T \eta(t) \\ & + \mathfrak{g}'(x(t))^T \partial_\eta \psi(w(t), \eta(t); \mu)^{-1} \partial_w \psi(w(t), \eta(t); \mu) \mathfrak{g}'(u(t)), \end{aligned}$$
there exists a constant $\gamma > 0$ such that
$$\xi(t)^T H(t) \xi(t) \geq \gamma |\xi(t)|^2$$
for almost all $t \in [0, 1]$ and all $x \in \ker c'$.

Under these assumptions $F'(v; \mu)$ has an inverse, which on every closed set $D \times [\mu_-, \mu_0]$ with $\mu_- > 0$, is uniformly bounded.

It is interesting to note that the existence of the central path is directly connected with the sufficient conditions for an optimal solution — see Theorem 2.2.

Theorem 2.5. *Assumptions and notation as in Theorem 2.4. Assume there exists some starting point $v_0 \in D$ and $\mu_0 > 0$ such that $F(v_0; \mu_0) = 0$. Then there exists a path $v(\mu) \in D$ that can be continued up to the boundary of $D \times (0, \mu_0]$.*

Once the existence of the central path has been established, we are ready to proceed to the second step of our derivation, which involves the continuation of the central path within closed sets up to $\mu = 0$. Unfortunately, due to the smoothing effect of the complementarity regularization, continuation in V_∞ cannot be performed up to $\mu = 0$. Instead, we will resort to the coarser Hilbert space setting of V_2. In the presence of *state* constraints, the Lagrange multipliers usually contain Dirac distributions at boundary points of state constrained subarcs — in which case every reasonable approximation will leave any bounded set D even in V_2. In the presence of *control* constraints, the dual variables are known to be continuous apart from a singular set, which can be excluded by constraint qualifications. In view of this feature, we restrict our treatment to control constrained problems for the remaining part of this section — making sure that we exclude the singular set (see Assumption 2 in Theorem 2.8 below).

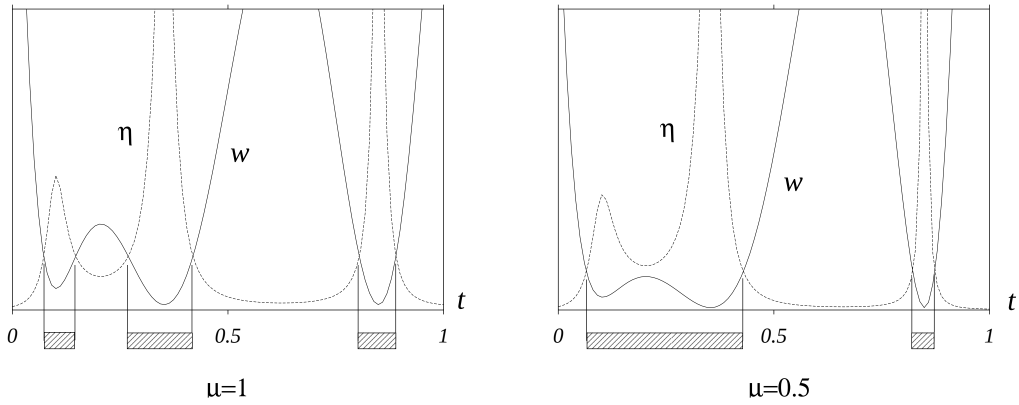

Figure 2.2: Nearly active sets computed from the slack variables w and the Lagrange multipliers η for $\rho = 1$. The reduction in the duality gap parameter μ from left to right leads to a better approximation of the solution's active set.

Note that in fact only the first part — continuation within open sets $\mu > 0$ — can be carried out *numerically.* The theoretical restriction to the Hilbert space V_2 is therefore not too severe. It may, however, influence the approximation order with respect to the continuation parameter μ. For an associated experimental result see Figure 4.8. Even our key example in Section 4 falls out of the present analytic setting, since it involves state constraints as well. However, in this case, the *adaptive multilevel algorithm* to be worked out in Section 3 produces successively sharper local peaks on successively finer meshes — thus realizing a multilevel approximation of the Dirac distribution. For an illustration of this effect see Figure 4.9 below.

In what follows we will prove uniform boundedness of $F'(v(\mu);\mu)^{-1}$ for $\mu \to 0$. This requires more subtle analytic techniques than those used to prove Theorem 2.4 above (cf. [30]). For this purpose, we exploit the well-known splitting of inequality constraints into active and inactive ones: active constraints are subsumed into the equality constraints c, whereas inactive constraints are simply dropped. Since the exact active/inactive splitting at $\mu = 0$ is typically not known for $\mu > 0$, we weaken the concept introducing an approximate splitting into *nearly* active and inactive inequality constraints — for an illustration see Figure 2.2.

For ease of presentation, we will consider $\bar{n}_u = 1$. The extension to more components is straightforward. To be precise, we define a *ρ-nearly active set* Ω_ρ by

$$\Omega_\rho(v) = \{t \in [0,1] : w(t;\mu) \le \rho\eta(t;\mu)\} \quad \text{and its complement} \quad \Omega_\rho^c = [0,1]\backslash\Omega_\rho$$

for $\rho > 0$. This splitting induces a splitting of variables

$$w = (w_\rho, w_\rho^c) \quad \text{and} \quad \eta = (\eta_\rho, \eta_\rho^c)\,,$$

of spaces

$$W_\rho = \{w|_{\Omega_\rho} : w \in L_2\} \quad \text{and} \quad W_\rho^c = \{w|_{\Omega_\rho^c} : w \in L_2\}\,, \tag{2.11}$$

and of inequalities and complementarity function

$$g = (g_\rho, g_\rho^c) \quad \text{and} \quad \psi = (\psi_\rho, \psi_\rho^c)\,.$$

Lemma 2.6. *The splitting* (2.11) *leads to the diagonal operator splittings*

$$\partial_w \psi(w,\eta;\mu) = \begin{bmatrix} \partial_w \psi_\rho & \\ & \partial_w \psi_\rho^c \end{bmatrix} \quad \textit{and} \quad \partial_\eta \psi(w,\eta;\mu) = \begin{bmatrix} \partial_\eta \psi_\rho & \\ & \partial_\eta \psi_\rho^c \end{bmatrix},$$

where $\|(\partial_w \psi_\rho)^{-1}\|_{W_\rho \to W_\rho}$ *and* $\|(\partial_\eta \psi_\rho^c)^{-1}\|_{W_\rho \to W_\rho}$ *are bounded independently of* μ.

Proof. For the nearly active set we infer

$$\begin{aligned}\partial_w \psi_\rho(w,\eta;\mu)^{-1} &= \left(1 - \frac{w}{\sqrt{w^2+\eta^2+2\mu}}\right)^{-1} \\ &= \frac{\sqrt{w^2+\eta^2+2\mu}}{\sqrt{w^2+\eta^2+2\mu}-w} \le \frac{\sqrt{2(w^2+\eta^2)}}{\sqrt{w^2+\eta^2}-w}\,.\end{aligned}$$

From $sw = \eta$ with $s \ge \rho^{-1}$ we conclude

$$\partial_w \psi_\rho(w,\eta;\mu)^{-1} \le \frac{\sqrt{2(1+s^2)w^2}}{\sqrt{(1+s^2)w^2}-w} = \frac{\sqrt{2(1+s^2)}}{\sqrt{1+s^2}-1} \le \frac{\sqrt{2(1+\rho^{-2})}}{\sqrt{1+\rho^{-2}}-1}\,.$$

Straightforward computation yields $\|(\partial_w \psi_\rho)^{-1}\|_{W_\rho \to W_\rho} \le \text{const}$, where the constant depends only on ρ. The analogous proof for $(\partial_\eta \psi_\rho^c)^{-1}$ is omitted. □

For the proof of the main theorem we will use the following Lemma by Braess and Blömer [8][1] on saddle point operators with penalty term.

Lemma 2.7. *Let X and Z be Hilbert spaces. Assume the following conditions hold:*

1. *The continuous linear operator $C : X \to Z$ satisfies the inf-sup-condition: There exists a constant $\beta > 0$ such that*

$$\inf_{x \in X} \sup_{\zeta \in Z} \frac{\langle \zeta, Cx \rangle}{\|x\|_X \|\zeta\|_Z} \ge \beta\,.$$

2. *The continuous linear operator $A : X \to X^\star$ is symmetric positive definite on the nullspace of C and positive semidefinite on the whole space X: There exists a constant $\alpha > 0$ such that*

$$\langle x, Ax \rangle \ge \alpha \|x\|_X^2 \quad \textit{for all } x \in \ker C$$

and

$$\langle x, Ax \rangle \ge 0 \quad \textit{for all } x \in X\,.$$

[1] Note that there is a misprint in Lemma B.1 in the article. The assumption that A is positive semidefinite on the whole space is used in the proof without being stated.

3. *The continuous linear operator $D : Z^\star \to Z$ is symmetric positive semidefinite.*

Then, the operator

$$\begin{bmatrix} A & C^\star \\ C & -D \end{bmatrix}$$

is invertible. The inverse is bounded by a constant depending only on α, β, and the norms of A, C, and D.

Theorem 2.8. *Let $\bar{n}_y = 0$. Suppose that the following conditions are satisfied uniformly for all $v = (x, \lambda, \eta, w)$ on the central path.*

1. *The control u occurs linearly in J, c, and g.*

2. *For the nearly active constraints $C := (c'(x), g'_\rho(x))^T$ the inf-sup-condition*

$$\inf_{\chi \in \Lambda_2, \zeta \in W^\star_\rho} \sup_{\xi \in X_2} \frac{\langle C\xi, (\chi, \zeta)^T \rangle}{\|\xi\|_{X_2} (\|\chi\|_{\Lambda_2} + \|\zeta\|_{W^\star_\rho})} \geq \beta$$

   holds for some $\beta > 0$.

3. *The modified Hessian of the Lagrangian is positive definite on the nullspace of the nearly active constraints and positive semidefinite on the whole space: For*

$$H := \partial^2_x L(x, \lambda, \eta) + g^{c\prime}_\rho(x)^\star \partial_\eta \psi^c_\rho(w, \eta; \mu)^{-1} \partial_w \psi^c_\rho(w, \eta; \mu) g^{c\prime}_\rho(x) \tag{2.12}$$

   there exists a constant $\alpha > 0$ such that

$$\langle Hx, x \rangle \geq \alpha \|x\|^2_{X_2} \quad \text{for all } x \in \ker C$$
$$\langle Hx, x \rangle \geq 0 \quad \text{for all } x \in X_2 \,.$$

If the central path $v(\mu)$ is defined on the half-open interval $(0, \mu]$, it can be continuously extended in V_2 to include a limit point $v(0)$ satisfying the KKT conditions (Theorem 2.1).

Remark 2.9. In general, whenever controls u arise nonlinearly, they can be expressed analytically in terms of state and dual variables via differentiating the Hamilton function — which might be the only possibly necessary analytic preprocessing in our approach. Therefore, the limitation to linearly occurring controls is no severe restriction.

Proof. The proof will be merely outlined. Details can be found in [30].

In a first step, we eliminate Δw and $\Delta\eta_\rho^c$ from the system $F'(v;\mu)\Delta v = \zeta = (z,s,r,q)^T$ to obtain

$$\begin{bmatrix} H & -(c')^\star & -(g_\rho')^\star \\ -c' & & \\ -g_\rho' & & -(\partial_w\psi_\rho)^{-1}\partial_\eta\psi_\rho \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta\eta_\rho \end{bmatrix} = \begin{bmatrix} z + (g_\rho^{c\prime})^\star(\partial_\eta\psi_\rho^c)^{-1}(q_\rho^c - \partial_w\psi_\rho^c r_\rho^c) \\ s \\ -(\partial_w\psi_\rho)^{-1}q_\rho + r_\rho\,. \end{bmatrix} \tag{2.13}$$

In a second step, we show that the inverse of the operator in (2.13) is bounded independently of μ. Note that $(\partial_w\psi_\rho)^{-1}$, $\partial_\eta\psi_\rho$, $(\partial_\eta\psi_\rho^c)^{-1}$, and $\partial_w\psi_\rho^c$ are all bounded uniformly for $\mu > 0$ by Lemma 2.6. Then, using Assumptions 2 and 3, we can apply Lemma 2.7 which guarantees the existence of a unique solution $\Delta x \in X_2$, $\Delta\lambda \in \Lambda_2$, and $\Delta\eta_\rho \in W_\rho$ with

$$\begin{aligned} \|\Delta x\|_{X_2} &\le \text{const}\,\|z + (g_\rho^{c\prime})^\star(\partial_\eta\psi_\rho^c)^{-1}(q_\rho^c - \partial_w\psi_\rho^c r_\rho^c)\|_{X_2^\star} \\ &\le \text{const}\,(\|z\|_{X_2^\star} + \|(g_\rho^{c\prime})^\star\|\|(\partial_\eta\psi_\rho^c)^{-1}\|(\|q_\rho^c\|_{W_2} + \|\partial_w\psi_\rho^c\|\|r_\rho^c\|_{W_2})) \\ &\le \text{const}\,(\|z\|_{X_2^\star} + \text{const}\,(\|q_\rho^c\|_{W_2} + \|r_\rho^c\|_{W_2})) \\ &\le \text{const}\,\|\zeta\|_{Z_2} \end{aligned}$$

and with similar calculations,

$$\|\Delta\lambda\|_{\Lambda_2} + \|\Delta\eta_\rho\|_{W_2} \le \text{const}\,\|\zeta\|_{Z_2}\,,$$

where the constants are independent of μ. Tracing the eliminations of the first step back yields

$$\|\Delta\eta_\rho^c\|_{W_2} \le \text{const}\,\|\zeta\|_{Z_2} \quad \text{and} \quad \|\Delta w\|_{W_2} \le \text{const}\,\|\zeta\|_{Z_2}\,.$$

Hence, $F'(v(\mu);\mu)^{-1}$ is bounded independently of μ. Finally, from

$$\partial_\mu\psi(w,\eta;\mu) = \frac{1}{\sqrt{w^2+\eta^2+2\mu}} \le \mu^{-\frac{1}{2}}$$

and

$$\partial_\mu F(v(\mu)) = \begin{bmatrix} 0 & 0 & 0 & -\partial_\mu\psi \end{bmatrix}^T$$

we infer for the derivative of the central path $v(\mu)$ that

$$\begin{aligned} \|v'(\mu)\|_{V_2} &= \|F'(v(\mu);\mu)^{-1}\partial_\mu F(v(\mu))\|_{V_2} \\ &\le \|F'(v(\mu);\mu)^{-1}\|_{Z_2\to V_2}\|\partial_\mu F(v(\mu))\|_{Z_2} \le \text{const}\,\mu^{-\frac{1}{2}}\,. \end{aligned}$$

Therefore, the path is uniformly continuous and thus can be continuously extended to include a limit point $v(0)$.

If u occurs linearly in J, c, and g, both the path $v(\mu)$ and F are continuous in V_2, such that $F(v(0),0) = 0$ and $v(0)$ satisfies the first order optimality conditions of Theorem 2.1. □

Remark 2.10. In general, Assumption 2 imposes an upper bound on the choice of ρ due to the monotonicity

$$\rho_1 \le \rho_2 \Rightarrow W_{\rho_1} \subset W_{\rho_2} .$$

If W_ρ gets too large, C may become non-injective and thus no longer satisfy the inf-sup-condition.

3 Numerical Algorithm

For the numerical computation of the solution point $v(0)$ we employ a Newton type continuation method following the central path $v(\mu)$ defined by (2.10). When applied to $F(v) = 0$ and $AF(Bv) = 0$, where A and B are invertible linear transformations, Newton's method generates equivalent sequences of iterates. This *invariance* property should be inherited by numerical algorithms and accompanying convergence theory. Unfortunately, full invariance is impossible due to the necessity of measuring convergence in some appropriate norm. Fixing $B = I$ one obtains affine covariant (error oriented) methods [16], whereas setting $A = I$ yields affine contravariant (residual oriented) methods [19]. Coupling $A = B^\star$ results in affine conjugate (energy oriented) methods for convex unconstrained optimization problems [17]. For an in-depth treatment of affine invariance we refer to the upcoming research monograph[15].

3.1 Affine Invariant Norms

Neither of the above-mentioned invariance classes reflects the structure of the optimization problem (2.1). A new class of affine invariance and a corresponding invariant norm for equality constrained problems ($\bar{n}_u = \bar{n}_y = 0$) that has been worked out in Volkwein and Weiser [28] needs to be extended to include inequality constraints as well.

First we recollect the norm construction for equality constrained problems. We utilize the positive definiteness of $\partial_x^2 L$ on the nullspace of the active constraints by constructing an affine conjugate seminorm. This seminorm is complemented by a residual-oriented approach to the constraints in order to define a norm.

Lemma 3.1. *The adjoint equation operator $T(v) : \ker c'(x) \times \Lambda_2 \to X_2^\star$ defined by*

$$T(v) := \begin{bmatrix} \partial_x^2 L(v) & -c'(x)^\star \end{bmatrix}$$

is an isomorphism.

A proof is given in [30].
Let $R(v) : \ker c'(x) \times \Lambda_2 \to X_2^\star \times \Lambda_2^\star$ be defined by

$$\begin{bmatrix} \partial_x^2 L(v) & \\ & I_R \end{bmatrix} ,$$

where I_R denotes the Riesz-isomorphism. If $\partial_x^2 L(v)$ is positive definite on $\ker c'(x)$, we can define a local norm on $Z_2 = X_2^\star \times \Lambda_2^\star$ by

$$|||(a,l)^T|||_v^2 := \langle R(v)T(v)^{-1}a, T(v)^{-1}a\rangle + \|l\|_{\Lambda_2^\star}^2 \,. \tag{3.14}$$

Since $T(v)$ is an isomorphism and $R(v)$ positive definite, this norm is equivalent to the canonical norm on Z_2. Furthermore, it is easy to verify that it is invariant under linear transformations of the domain space X_2: with

$$B = \begin{bmatrix} \bar{B} & \\ & I \end{bmatrix}$$

we have

$$\hat{R}(Bv) = \begin{bmatrix} \bar{B}^\star \partial_x^2 L(Bv)\bar{B} & \\ & I \end{bmatrix} = B^\star R(Bv) B$$

and

$$\hat{T}(Bv) = \begin{bmatrix} \bar{B}^\star \partial_x^2 L(Bv)\bar{B} & -\bar{B}^\star c'(Bv)^\star \end{bmatrix} = \bar{B}^\star T(Bv) B \,,$$

such that

$$\begin{aligned} |||(\bar{B}a,l)^T|||_{Bv}^2 &= \langle \hat{R}(Bv)\hat{T}(Bv)^{-1}\bar{B}a, \hat{T}(Bv)^{-1}\bar{B}a\rangle + \|l\|_{\Lambda_2^\star}^2 \\ &= \langle B^\star R(v)BB^{-1}T(v)^{-1}\bar{B}^{-1}\bar{B}a, B^{-1}T(v)^{-1}\bar{B}^{-1}\bar{B}a\rangle + \|l\|_{\Lambda_2^\star}^2 \\ &= \langle R(v)T(v)^{-1}a, T(v)^{-1}a\rangle + \|l\|_{\Lambda_2^\star}^2 \\ &= |||(a,l)^T|||_{Bv}^2 \end{aligned}$$

Note that albeit its nontrivial definition, this norm is comparatively cheap to evaluate: the computation of $T(v)^{-1}a$ can be implemented by one more system solve with the same operator $F'(v)$ and a different right hand side.

The norm (3.14) can be used for *inequality* constrained problems, too, if the slack variables w and the Lagrange multipliers η for the inequality constraints are eliminated beforehand. Then the term $\partial_x^2 L(v)$ above has to be replaced by

$$H(v) := \partial_x^2 L(v) + g'(x)^\star \partial_\eta \psi(w,\eta;\mu)^{-1}\partial_w \psi(w,\eta;\mu) g'(x) \,,$$

which must again be positive definite on $\ker c'(x)$. However, this norm may suffer from the ill-conditioning of $\partial_\eta \psi(w,\eta;\mu)^{-1}\partial_w\psi(w,\eta;\mu)$ for $\mu \to 0$.

Alternatively, we can resort to the splitting of the inequality constraints into nearly active and inactive sets as introduced in Section 2. Elimination of the "nearly" inactive slack variables w_ρ^c and Lagrange multipliers η_ρ^c then leads to

$$H_\rho(v) := \partial_x^2 L(v) + (g'(x)_\rho^c)^\star \partial_\eta \psi_\rho^c(w,\eta;\mu)^{-1}\partial_w \psi_\rho^c(w,\eta;\mu) g'(x)_\rho^c \,, \tag{3.15}$$

which must be positive definite on $N_\rho := \ker c'(x) \times \ker g'(x)_\rho$. Since

$$\partial_\eta \psi_\rho^c(w,\eta;\mu)^{-1}\partial_w \psi_\rho^c(w,\eta;\mu)$$

is bounded for $\mu \to 0$, the ill-conditioning is avoided. Unfortunately, proceeding like that suffers from two drawbacks: first, the local norm varies nonsmoothly with the splitting into nearly active and inactive constraints and hence with the evaluation point v it is attached to, and second, it is expensive to evaluate, since $x \in N_\rho$ is difficult to obtain. In general, however, both drawbacks can be avoided in practice. Typically, not only the modified Hessian (3.15), but $\partial_x^2 L(v)$ itself is positive definite on the X-component of $F'(v)^{-1}(X_2^\star \times \{0_{\Lambda_2^\star}\} \times \{0_{W_2}\} \times \{0_{W_2}\})$, which is almost N_ρ at least for small μ. Thus we can define

$$|||(a,l,s,q)^T|||_{(v,\mu)}^2 := \langle \partial_x^2 L(v)\xi, \xi\rangle + \|\nu\|_{\Lambda_2}^2 + \|\sigma\|_{W_2}^2 + \|\vartheta\|_{W_2}^2 + \|l\|_{\Lambda_2^\star}^2 + \|s\|_{W_2}^2 + \|q\|_{W_2}^2$$

where

$$F'(v;\mu)(\xi,\nu,\sigma,\vartheta)^T = (a,0,0,0)^T \,. \tag{3.16}$$

It is immediately clear that this norm is equivalent to the canonical norm of Z_2. Furthermore, it is not affected by a change of splitting into nearly active and inactive inequality constraints. The positive definiteness of $\partial_x^2 L(v)$, however, must be checked in the algorithm.

So far, we have defined affine invariant norms for the coarser Hilbert space Z_2. In order to obtain an affine invariant norm on Z_∞, we assume that

$$\xi(t)^T \partial_x^2 L(v)(t)\xi(t) \ge \nu|\xi(t)|^2 \tag{3.17}$$

for all ξ obtained from (3.16) and some $\nu > 0$. This is related to the assumptions of Theorem 2.8. Then we define

$$\|(a,l,s,q)^T\|_{(v,\mu)}^2 := \|\phi\|_\infty + \|\nu\|_{\Lambda_\infty}^2 + \|\sigma\|_{W_\infty}^2 + \|\vartheta\|_{W_\infty}^2 + \|l\|_{\Lambda_\infty}^2 + \|s\|_{W_\infty}^2 + \|q\|_{W_\infty}^2 \tag{3.18}$$

with ξ, ν, σ, ϑ as in (3.16). Because of the boundedness of $F'(v;\mu)^{-1}$, this norm is equivalent to the canonical norm of Z_∞. Unfortunately, it restricts the invariance class to (block) diagonal transformations $\bar{B}$. A fully satisfactory affine invariant norm for Z_∞ has not been found yet.

Lemma 3.2. *If (3.17) holds, then there exists an affine invariant Lipschitz constant $\omega \ge 0$, such that*

$$\|F'(v_1;\mu) - F'(v_2;\mu)\|_{(v_1,\mu)} \le \omega \|F'(v_1;\mu)(v_1 - v_2)\|_{(v_1,\mu)} \tag{3.19}$$

for all v_1, v_2 such that $\mathrm{co}\{v_1, v_2\} \subset D$.

Proof. The existence of ω is a direct consequence of Theorem 2.3 and the equivalence of the norms $\|\cdot\|_{v_1}$ and $\|\cdot\|_{Z_\infty}$. □

Formulating inexact Newton continuation algorithms in terms of local norms, we need to have a Lipschitz-continuous dependence of the norm with respect to the evaluation point to which it is attached, i.e. the existence of affine invariant constants $\gamma_v \geq 0$ and $\gamma_\mu \geq 0$ such that

$$\left| \|r\|_{(v_1,\mu)} - \|r\|_{(v_2,\mu)} \right| \leq \gamma_v \|F'(v_1;\mu)(v_1 - v_2)\|_{(v_1,\mu)} \|r\|_{(v_1,\mu)} \tag{3.20}$$

and

$$\left| \|r\|_{(v,\mu_1)} - \|r\|_{(v,\mu_2)} \right| \leq \gamma_\mu |\mu_1 - \mu_2| \|r\|_{(v,\mu_1)} \tag{3.21}$$

for all r. This can be easily derived from the Lipschitz continuity of F' and the equivalence of the local norms to the canonical norm. In contrast, obtaining quantitative estimates analytically is a difficult task. Based on an assumption similar to (3.19), a rather crude estimate of $\gamma_v \leq \frac{3}{2}\omega$ has been derived in [28]. In the affine contravariant setting, $\gamma_v = 0$ is trivially obtained, $\gamma_v \leq \omega$ holds in the affine covariant case, and $\gamma_v \leq \frac{\omega}{2}$ in the affine conjugate case. For the norm (3.18), however, no quantitative estimate is known up to now.

Remark 3.3. Recently, Potra [24] employed an affine invariant norm for proving $\mathcal{O}(\sqrt{n}L)$-iteration complexity of an interior point algorithm applied to horizontal linear complementarity problems, which include linear and quadratic complementarity problems. In the notation of the current paper, that author uses a diagonal scaling like $D := \sqrt{\partial_w \psi^{-1} \partial_\eta \psi}$. The applicability of such a scaled norm to the function space optimal control setting is still under consideration.

3.2 Adaptive Central Path Following

Once the central path homotopy is theoretically established, a numerical continuation scheme for following the path towards the solution $v(0)$ must be developed. For *numerical* pathfollowing, an adaptive tangential predictor / Newton type corrector algorithm is worked out. The method is applied directly to the infinite dimensional function space formulation, involving only in the innermost loop when solving linear subproblems. Since a reduction of the discretization error is expensive, we substitute both the tangential predictor and the Newton corrector by their inexact counterparts and aim for linear convergence only — in the spirit of complexity estimates of [17].

As can be seen from the increasing Lipschitz constant of $F'(\cdot;\mu)$ as derived in Theorem 2.3, the local convergence domain of the Newton corrector can be expected to collapse for $\mu \to 0$ (see Figure 3.3). Nevertheless, Theorem 2.8 provides a qualitative upper bound on the error incurred by a premature termination of the numerical continuation along the central path. Experience shows that feasible and acceptably suboptimal solutions can indeed be obtained by following the central path up to some $\mu_* > 0$ — see Section 4.

Inexact Newton corrector. The corrector operates with constant duality gap parameter; thus we drop μ in order to simplify notation. Due to the inhibitive cost of reducing the discretization error, we cannot strive for highly accurate Newton

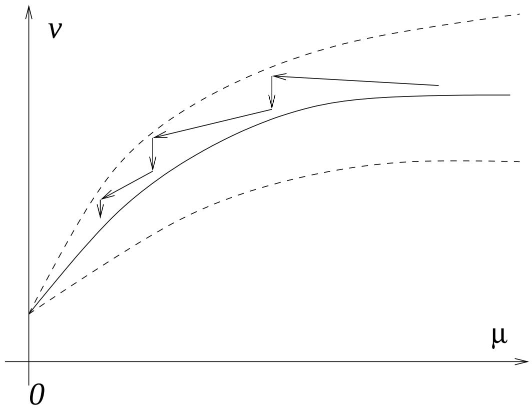

Figure 3.3: Inexact tangential continuation along the central path.

corrections. Instead, we will employ an *inexact Newton method*, where an *inner residual* remains:

$$F'(v^k)\delta v^k = -F(v^k) + r^k$$
$$v^{k+1} = v^k + \delta v^k \,.$$

The relative accuracy δ_k of the *inexact Newton correction* δv^k, given by

$$\delta_k := \frac{\|r^k\|_{v^k}}{\|F(v^k)\|_{v^k}}\,, \tag{3.22}$$

will play a crucial role. In actual computation, the inexact simplified Newton correction $\overline{\delta v^{k+1}}$ defined by

$$F'(v^k)\overline{\delta v^{k+1}} = -F(v^{k+1}) + \overline{r^{k+1}}$$

will also be used.

Theorem 3.4. *Assumptions and notation of Theorem 2.4. Let γ_v and ω be constants such that the local norm $\|\cdot\|_v$ satisfies (3.20) and the affine invariant Lipschitz condition*

$$\|(F'(\xi) - F'(v))(\xi - v)\|_\zeta \le \omega \|F_v(v)(\xi - v)\|_v^2 \tag{3.23}$$

holds for all collinear $v, \xi, \zeta \in V$ such that $\mathrm{co}\{v, \xi, \zeta\} \subset D$. Let $\Theta < 1$ and

$$\mathcal{L}(v) := \Big\{\xi \in D : \|F(\xi)\|_\xi \le \Big(1 + \frac{\gamma_v \Theta}{2\omega}\Big)\|F(v)\|_v\Big\}\,.$$

Assume that $v^0 \in D$ and that the level set $\mathcal{L}(v^0)$ is closed. If $\omega\|F(v^0)\|_{v_0} < 2\Theta$ and the inner iteration is controlled such that

$$\frac{1+\delta_k}{2}\omega\|F'(v^k)\delta v^k\|_{v^k} + (1 + \gamma_v\|F'(v^k)\delta v^k\|_{v^k})\delta_k \le \Theta\,, \tag{3.24}$$

then the iterates are well defined for all $k \in \mathbb{N}$, stay in $\mathcal{L}(v^0)$, and the residuals converge to zero at a rate of

$$\|F(v^{k+1})\|_{v^{k+1}} \leq \Theta \|F(v^k)\|_{v^k} .$$

Furthermore, if the inexact simplified Newton correction is computed with relative accuracy $\overline{\delta_{k+1}}$,

$$\|F'(v^k)\overline{\delta v^{k+1}}\|_{v^k} \leq (1+\overline{\delta_{k+1}})\Big(\frac{\delta_k}{1-\delta_k} + \frac{\omega}{2}\|F'(v^k)\delta v^k\|_{v^k}\Big)\|F'(v^k)\delta v^k\|_{v^k} \tag{3.25}$$

holds.

Proof. By induction, let $\mathcal{L}(v^k)$ be closed and $\omega\|F(v^k)\|_{v^k} < 2$. Then

$$\begin{aligned} F(v^k + s\delta v^k) &= F(v^k) + \int_0^s F_v(v^k + t\delta v^k)\delta v^k \, dt &(3.26)\\ &= (1-s)F(v^k) + sr^k + \int_0^s (F_v(v^k+t\delta v^k) - F_v(v^k))\delta v^k \, dt &(3.27)\end{aligned}$$

for all $s \in [0,1]$ with $\mathrm{co}\{v^k, v^k + s\delta v^k\} \subset D$.

Using the Lipschitz continuity (3.23) of F' and the norm continuity (3.20), for $\sigma \in [0,s]$ we have

$$\begin{aligned} &\|F(v^k + s\delta v^k)\|_{v^k+\sigma\delta v^k} \\ &\quad \leq (1-s)\|F(v^k)\|_{v^k+\sigma\delta v^k} + s\|r^k\|_{v^k+\sigma\delta v^k} \\ &\qquad + \int_0^s \|(F'(v^k+t\delta v^k) - F'(v^k))\delta v^k\|_{v^k+\sigma\delta v^k} \, dt \\ &\quad \leq (1-s)(1+\sigma\gamma_v\|F'(v^k)\delta v^k\|_{v^k})\|F(v^k)\|_{v^k} \\ &\qquad + s(1+\sigma\gamma_v\|F'(v^k)\delta v^k\|_{v^k})\|r^k\|_{v^k} + \int_0^s t\omega\|F'(v^k)\delta v^k\|^2_{v^k} \, dt \\ &\quad = (1+\sigma\gamma_v\|F'(v^k)\delta v^k\|_{v^k})((1-s)\|F(v^k)\|_{v^k} + s\delta_k\|F(v^k)\|_{v^k}) \\ &\qquad + \frac{s^2}{2}\omega\|F'(v^k)\delta v^k\|^2_{v^k} \end{aligned} \tag{3.28}$$

From (3.22) we have

$$(1-\delta_k)\|F(v^k)\|_{v^k} \leq \|F_v(v^k)\delta v^k\|_{v^k} = \|F(v^k) - r^k\|_{v^k} \leq (1+\delta_k)\|F(v^k)\|_{v^k} . \tag{3.29}$$

Thus we arrive at

$$\begin{aligned} &\frac{\|F(v^k + s\delta v^k)\|_{v^k+s\delta v^k}}{\|F(v^k)\|_{v^k}} \\ &\quad \leq (1+s\gamma_v\|F'(v^k)\delta v^k\|_{v^k})(1-s+s\delta_k) + \frac{1+\delta_k}{2}s^2\omega\|F'(v^k)\delta v^k\|_{v^k} . \end{aligned}$$

The ultimate goal is to establish a contraction property for the undamped Newton method ($s = 1$). Thus we have to require

$$(1 + \gamma_v \|F'(v^k)\delta v^k\|_{v^k})\delta_k + \frac{1+\delta_k}{2}\omega\|F'(v^k)\delta v^k\|_{v^k} \le \Theta < 1\,,$$

which is the accuracy condition (3.24). Defining $\chi := \gamma_v\|F'(v^k)\delta v^k\|_{v^k}$ and using $s \le 1$ and (3.24), we have

$$\begin{aligned}
\frac{\|F(v^k + s\delta v^k)\|_{v^k+s\delta v^k}}{\|F(v^k)\|_{v^k}} &\le (1+s\chi)(1-s+s\delta_k) + \frac{1+\delta_k}{2}s^2\omega\|F'(v^k)\delta v^k\|_{v^k} \\
&= (1+s\chi)(1-s) + (1+s\chi)s\delta_k + \frac{1+\delta_k}{2}s^2\omega\|F'(v^k)\delta v^k\|_{v^k} \\
&\le (1+s\chi)(1-s) + s\Theta \qquad (3.30) \\
&= 1 - s + s\Theta + s(1-s)\chi \\
&\le 1 + \frac{\chi}{4}\,.
\end{aligned}$$

From (3.24) we infer $\|F'(v^k)\delta v^k\|_{v^k} \le 2\Theta/\omega$, and hence

$$\|F(v^k + s\delta v^k)\|_{v^k+s\delta v^k} \le \left(1 + \frac{\gamma_v\Theta}{2\omega}\right)\|F(v^k)\|_{v^k}$$

holds. If $\mathrm{co}\{v^k, v^k + \delta v^k\} \not\subset D$, then there is some $s^* \in [0,1)$ with $\mathrm{co}\{v^k, v^k + s^*\delta v^k\} \subset D$ but $v^k + s^*\delta v^k \notin \mathcal{L}(v^k)$, i.e.

$$\|F(v^k + s^*\delta v^k)\|_{v^k+s^*\delta v^k} > (1 + \frac{\gamma_v\Theta}{2\omega})\|F(v^k)\|_{v^k}\,,$$

which is a contradiction. Thus, $v^{k+1} \in D$. Furthermore, setting $s = 1$ in (3.30) yields

$$\|F(v^{k+1})\|_{v^{k+1}} \le \Theta\|F(v^k)\|_{v^k} \qquad (3.31)$$

and therefore $\mathcal{L}(v^{k+1}) \subset \mathcal{L}(v^k)$. Since $\mathcal{L}(v^k)$ is closed, every Cauchy sequence in $\mathcal{L}(v^{k+1})$ converges to a limit point in $\mathcal{L}(v^k)$, which is, by continuity of the norm, also contained in $\mathcal{L}(v^{k+1})$. Hence, $\mathcal{L}(v^{k+1})$ is closed.

Inserting $\sigma = 0$, $s = 1$ into (3.28) yields

$$\begin{aligned}
\|F'(v^k)\overline{\delta v^{k+1}}\|_{v^k} &\le (1 + \overline{\delta_{k+1}})\|F(v^{k+1})\|_{v^k} \\
&\le (1 + \overline{\delta_{k+1}})\left(\delta_k\|F(v^k)\|_{v^k} + \frac{\omega}{2}\|F'(v^k)\delta v^k\|^2_{v^k}\right) \\
&\le (1 + \overline{\delta_{k+1}})\left(\frac{\delta_k}{1-\delta_k} + \frac{\omega}{2}\|F'(v^k)\delta v^k\|_{v^k}\right)\|F'(v^k)\delta v^k\|_{v^k}\,,
\end{aligned}$$

which completes the proof. □

Inexact prediction step. From numerical experience, we expect more or less constant reduction factors for μ, translating into constantly decreasing continuation step sizes. In order to avoid this biased step size behavior, the predictor is formulated in terms of $\tau = -\log \mu$. The inexact tangential predictor $\hat{v}(\tau)$ is defined by

$$F'(v_0;\tau_0)\phi = -\partial_\tau F(v_0;\tau_0) + r\,, \quad \hat{v}(\tau) = v_0 + (\tau - \tau_0)\phi\,, \tag{3.32}$$

where again a residual r remains.

Lemma 3.5. *Assumptions of Theorem 3.4. Let γ_τ and β be nonnegative constants such that the local norm $\|\cdot\|_{(v,\tau)}$ satisfies*

$$\big|\|\rho\|_{(v_2,\tau_2)} - \|\rho\|_{(v_1,\tau_1)}\big| \le \gamma_\tau(\tau_2 - \tau_1)\|\rho\|_{(v_1,\tau_1)} \tag{3.33}$$

and

$$\|F(v_2;\tau_2)\|_{(v_1,\tau_1)} \le \|F(v_1;\tau_1)\|_{(v_1,\tau)} + \|r\|_{(v_1,\tau)}(\tau_2 - \tau_1) + \beta(\tau_2 - \tau_1)^2 \tag{3.34}$$

for all $\rho \in Z_\infty$, v_1, v_2 such that $F'(v_1;\tau_1)(v_2 - v_1) = -(\tau_2 - \tau_1)(\partial_\mu F(v_1;\tau_1) + r)$, and $\mathrm{co}\{v_1, v_2\} \subset D$*. Then the inexact Newton corrector with starting point v_2 converges to the central path $v(\tau)$ for all stepsizes $\Delta\tau = \tau_2 - \tau_1$ satisfying*

$$(1 + \gamma_\tau\Delta\tau)(\|F(v_1;\tau_1)\|_{(v_1,\tau_1)} + \|r\|_{(v_1,\tau_1)}\Delta\tau + \beta\Delta\tau^2) < \frac{2}{\omega}\,. \tag{3.35}$$

Proof. Combining the convergence condition $\omega\|F(v;\tau)\|_{(v,\tau)} < 2$ from Theorem 3.4 with assumptions (3.33) and (3.34) yields the result. □

Note that, since (3.35) represents a monotone convex function of $\Delta\tau$, the maximum permitted stepsize can be easily computed by an ordinary Newton method starting from $\sqrt{2/(\omega\beta)}$.

Remark 3.6. Again, the constant γ_τ is needed because of the inexactness of the tangential predictor. In exact Newton continuation algorithms (see Deuflhard [14]), the change of local norms can be subsumed under the second order term β.

Computable Lipschitz Estimates. For actual computation we need easily computable estimates of the theoretical quantities ω, γ_v, γ_τ, and β, to be inserted into conditions (3.24) and (3.35). From (3.25) and (3.20), (3.29), respectively, we derive the computable estimates

$$[\omega]_k = \frac{2}{\|F'(v^k)\delta v^k\|_{v^k}}\left(\frac{\|F'(v^k)\overline{\delta v^{k+1}}\|_{v^k}}{(1+\overline{\delta_{k+1}})\|F'(v^k)\delta v^k\|_{v^k}} - \frac{\delta_k}{1-\delta_k}\right) \le \omega \tag{3.36}$$

and

$$[\gamma_v]_k = \frac{d\Big(\Phi\big(\delta_{k+1}, \|F'(v^{k+1})\delta v^{k+1}\|_{v^{k+1}}\big), \Phi\big(\overline{\delta_{k+1}}, \|F'(v^k)\overline{\delta v^{k+1}}\|_{v^k}\big)\Big)}{(1+\delta_{k+1})\|F'(v^{k+1})\delta v^{k+1}\|_{v^{k+1}}\|F'(v^k)\delta v^k\|_{v^k}} \le \gamma_v\,.$$

where $\Phi(a,b) = [\frac{b}{1+a}, \frac{b}{1-a}]$ denotes the inaccuracy interval and

$$d(A,B) = \inf_{a \in A, b \in B} |a-b|$$

is the usual set distance. Furthermore, computable estimates for γ_τ and β can be derived from (3.33) and (3.34) as

$$[\gamma_\tau] = \frac{d\Big(\Phi\big(\delta_2, \|F'(v_2;\tau_2)\delta v_2\|_{(v_2;\tau_2)}\big), \Phi\big(\overline{\delta_2}, \|F'(v_1;\tau_1)\overline{\delta v_2}\|_{(v_1;\tau_1)}\big)\Big)}{(1+\delta_2)\|F'(v_2;\tau_2)\delta v_2\|_{(v_2;\tau_2)}(\tau_2-\tau_1)} \le \gamma_\tau$$

and

$$[\beta] = \max\{0, \tilde{\beta}\} \le \beta$$

where

$$\tilde{\beta} = \frac{d\Big(\Phi\big(\overline{\delta_2}, \|F'(v_1;\tau_1)\overline{\delta v_2}\|_{(v_1,\tau_1)}\big), \Phi\big(\delta_1, \|F'(v_1;\tau_1)\delta v_1\|_{(v_1,\tau_1)}\big)\Big)}{(\tau_2-\tau_1)^2} - \frac{\|r\|_{(v_1,\tau_1)}}{\tau_2-\tau_1},$$

respectively. Of course, *reliable* estimates are obtained only if δ_k is sufficiently small, which imposes additional accuracy requirements on the computation of the predictor and corrector.

Since the computable estimates are based on *local sampling* only, they are necessarily too small. Therefore, the computed continuation stepsize $\Delta\tau$ is larger than intended and may even be too large for the corrector to converge. In this case, a *stepsize reduction* has to be performed on the basis of updated estimates. An iterative stepsize reduction scheme and its termination has been studied in [30].

Solution of Linear Subproblems. Applying Newton continuation methods in function space requires solving a *sequence of linear complementarity problems* of the same structure as the nonlinear complementarity problem (2.10). In principle, any standard linear BVP solver can be employed. If there is a stable direction for the initial value problem, a multiple shooting discretization of the linear problem is certainly apropriate. For the numerical example in Section 4 and further ones in [30], a collocation method with adaptive mesh refinement has been used. The successive grid adaptation is based on a not very sophisticated ad hoc error estimator. These and related algorithmic issues will be worked out to a higher level of sophistication in the near future.

4 Abort Landing in the Presence of Windshear

In this section we will consider a well-known intricate optimal control problem, the abort landing in the presence of windshear. Our numerical results are based on the precise model given in [5] — for the convenience of the reader, this problem formulation is arranged in the Appendix. The problem is of Chebyshev type, maximizing the minimal altitude. There is only one control variable u entering *linearly* — thus satisfying Assumption 1 of Theorem 2.8. The optimal solution consists of control and state constrained subarcs as well as touch points and singular subarcs, which makes the problem difficult to tackle by means of the maximum principle. It contains a third-order state constraint and a nondifferentiable wind model based on spline representation — and is therefore not fully covered by our theoretical presentation in Section 2. Nevertheless, as will be reported now, already the first version of our herein developed algorithm worked satisfactorily.

Originally, the problem has been modeled by MIELE et al. [23]; as for the numerical solution, these authors seem to have applied a *robust collocation* method based on a finite dimensional parametrization of the control and combined with a gradient restoration technique to find the corresponding optimal finite dimensional solution. Their paper does not present any numerical results for the control, which is the most critical variable.

As a preparation for the application of *multiple shooting*, BULIRSCH et al. [10] required 11 pages to present a brief outline of the analytic derivation of the necessary conditions. In contrast to that, our herein proposed method did not require any analytical preprocessing — thus saving considerable human effort. In a second paper [11], the application of the multiple shooting method has been described along with the homotopy necessary to obtain the correct switching structure. In 1995, BERKMANN and PESCH [5] solved the same problem even more accurately and claimed that "a competing direct method is unlikely to be able to produce solutions with such high resolution". In fact, our direct function space method did require a substantial computational effort to reach a comparably high accuracy. A comparison of computing times, however, would be too early, since our first focus was on developing a working algorithm within the rather new conceptual frame. There is enough space left for perfectioning our algorithm, which will be filled in the near future.

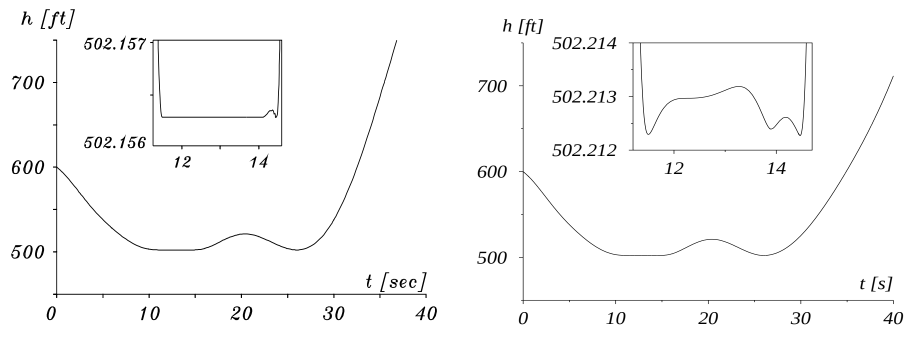

Figure 4.4: Altitude h for windshear problem. Left: multiple shooting result from [5]. Right: central path result at $\mu = 2.1 \cdot 10^{-4}$ (this paper).

Figure 4.4 shows a comparison of our altitude results with those obtained in [5].

The agreement is perfect within the interval up to the last touch point $t^* = 25.997s$. Beyond that point, there exists a continuum of optimal solutions. This can be understood from the fact that for $t > t^*$ all relevant Lagrange multipliers vanish (Figure 4.5) and none of the inequality constraints is active. Only the multiplier corresponding to the Chebyshev reformulation of the minimization problem does not vanish (see $\dot{\zeta} = 0$ in the Appendix).

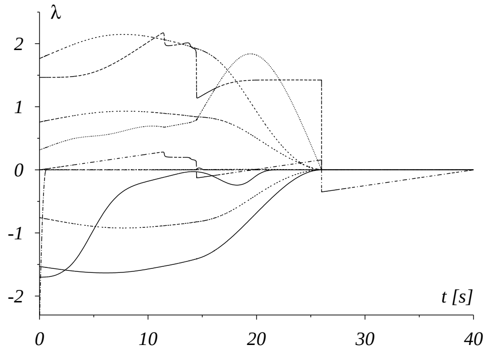

Figure 4.5: Lagrange multipliers λ_i corresponding to equality constraints (scaled).

As already mentioned above, the most critical variable is the control u, the angle of attack rate. That is why we present its rather complex behavior in Figure 4.6. As can be seen, our results once again are in perfect agreement with the multiple shooting results from [5] at least in the relevant interval $[0, t^*]$. Slight differences do occur as local Gibbs phenomena around two discontinuities, among which one arises at $t = t^*$. The latter occurrence does not deserve too much attention — we could have suppressed it in the canonical way by adding some term ϵu^2 to the functional for sufficiently small ϵ.

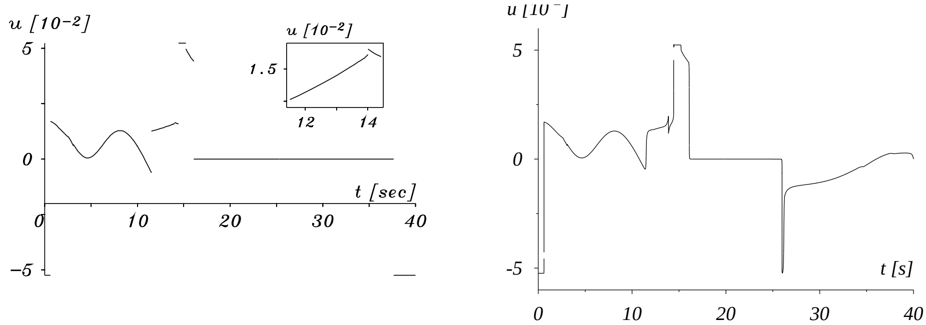

Figure 4.6: Control u for windshear problem. Left: multiple shooting result from [5]. Right: central path result at $\mu = 2.1 \cdot 10^{-4}$ (this paper).

In passing we note that PYTLAK [25] has also attacked this problem by his method and documented his results, but obtained a control behavior quite different from the one given in Figure 4.6.

As for the obtained functional value (minimum altitude $h_{\min}$), BULIRSCH et al. [11] report an improvement of 10ft over MIELE et al. [23]. On top of that, BERKMANN and PESCH [5] achieved a further improvement of $2.7 \cdot 10^{-6}$ ft to a value of $h_{\min} = 502.1562810$ ft. Our method led to an even better minimal altitude of $h_{\min} = 502.210661$ ft. In order to assess this value, we solved the initial value problem in both forward and backward direction using the computed control from our algorithm. As a numerical integrator we selected the Matlab implementation of Dormand-Prince RK45. In forward direction we obtained $h_{\min} = 502.210433$ ft, in backward direction $h_{\min} = 502.210438$ ft. This clearly confirms that our result is an improvement even over [5].

In order to throw some light into the performance of our new algorithm, we next give some details about the continuation process with respect to the duality gap parameter μ and the adaptive multilevel scheme.

The computations were started on a uniform initial grid with mesh size $h_0 = 1/25$. On this grid, the nonlinear KKT equations $F(v; 1) = 0$ with dimension 2748 have been solved using a Newton method with damping. The corresponding initial trajectory is depicted in Figure 4.7.

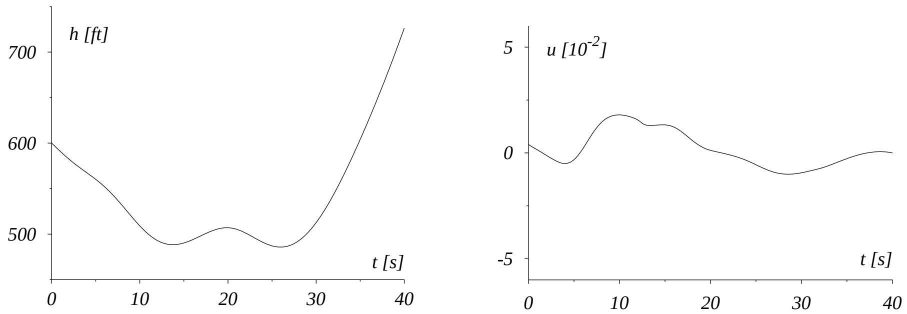

Figure 4.7: Initial trajectory for windshear problem. Left: altitude h. Right: control u.

An illustration of the adaptive continuation process along the central path is given in Figure 4.8. The log-log scale indicates that

$$J(v(\mu)) - J(v(0)) \sim \mu^\alpha \text{ with } \alpha \approx 1.44 .$$

Finally, the adaptive mesh refinement structure for this problem is presented in Figure 4.9. Successive refinement led to mesh sizes

$$h_j = 2^{-j} h_0 .$$

Obviously, the highly dynamic structure of the solution is captured reasonably well by the adaptive refinement procedure.

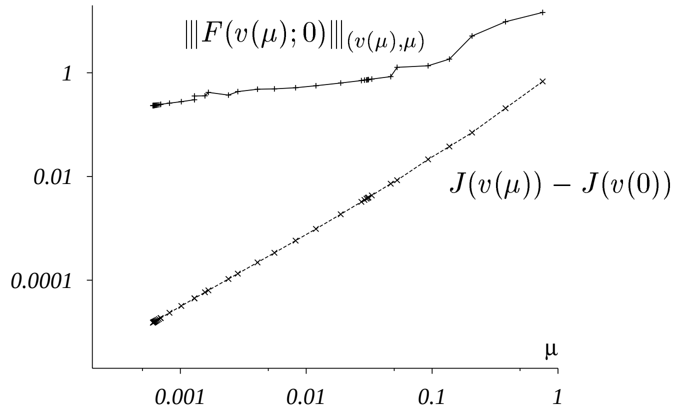

Figure 4.8: Central path continuation for windshear problem.

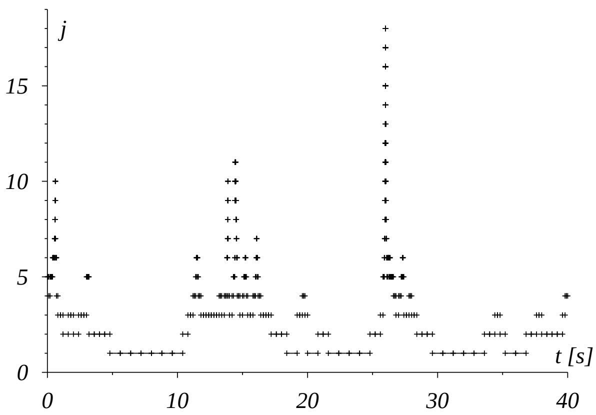

Figure 4.9: Adaptive mesh refinement in windshear problem.

Conclusion

In this paper we present a direct function space method for optimal control problems based on the complementarity formulation of interior point methods. The new method essentially dispenses of any analytical preprocessing — thus saving considerable human effort. In its algorithmic realization, function space is exploited via an adaptive multilevel method in combination with an adaptive central path following algorithm. A theoretical justification of the algorithm has only been achieved for control constrained problems where the controls appear linearly. However, numerical results for a well-known intricate optimal control problem with both control and state constraints seem to indicate that a much wider class of problems should be tractable by our algorithm. Even though a lot remains to be done both in theoretical justification and in algorithmic realization, we are confident to have opened a promising alternative path towards the numerical solution of complex optimal control problems from science and engineering.

References

[1] W. Alt and K. Malanowski. The Lagrange-Newton method for state constrained optimal control problems. *Comput. Optim. Appl.*, 4(3):217–239, 1995.

[2] U. Ascher and G. Bader. A new basis implementation for a mixed order boundary value ode solver. *SIAM J. Sci. Stat. Comput.*, 8:483–500, 1987.

[3] U. Ascher, J. Christiansen, and R. D. Russell. Collocation software for boundary-value odes. *ACM Trans. Math. Softw.*, 7:209–222, 1981.

[4] U. Ascher, R. M. M. Mattheij, and R. D. Russell. *Numerical solution of boundary value problems for ordinary differential equations.* Prentice Hall, 1988.

[5] P. Berkmann and H. J. Pesch. Abort landing in windshear: An optimal control problem with third-order state constraint and varied switching structure. *J. Optim. Theory Appl.*, 85(1):21–57, 1995.

[6] H. G. Bock. Numerische Behandlung von zustandsbeschränkten und Chebychef-Steuerungs-Problemen. Technical report, Carl-Cranz-Gesellschaft, Oberpfaffenhofen, 1981.

[7] H. G. Bock and K.-J. Plitt. A multiple shooting algorithm for direct solution of optimal control problems. In *Proceedings of the 9th IFAC World Congress, Budapest.* Pergamon Press, 1984.

[8] D. Braess and C. Blömer. A multigrid method for a parameter dependent problem in solid mechanics. *Numerische Mathematik*, 57, 1990.

[9] R. Bulirsch. Die Mehrzielmethode zur numerischen Lösung von nichtlinearen Randwertproblemen und Aufgaben der optimalen Steuerung. Technical report, Carl-Cranz-Gesellschaft, Oberpfaffenhofen, 1971.

[10] R. Bulirsch, F. Montrone, and H. J. Pesch. Abort landing in the presence of windshear as a minimax optimal control problem. I: Necessary conditions. *J. Optim. Theory Appl.*, 70(1):1–23, 1991.

[11] R. Bulirsch, F. Montrone, and H. J. Pesch. Abort landing in the presence of windshear as a minimax optimal control problem. II: Multiple shooting and homotopy. *J. Optim. Theory Appl.*, 70(2):223–254, 1991.

[12] R. Bulirsch, E. Nerz, H. J. Pesch, and O. von Stryk. Combining direct and indirect methods in optimal control: Range maximization of a hang glider. In R. Bulirsch, A. Miele, J. Stoer, and K. H. Well, editors, *Optimal Control*, volume 111 of *International Series of Numerical Mathematics*. Birkhäuser, 1993.

[13] P. Deuflhard. A modified Newton method for the solution of ill-conditioned systems of nonlinear equations with application to multiple shooting. *Numer. Math.*, 22:289–315, 1974.

[14] P. Deuflhard. A stepsize control for continuation methods and its special application to multiple shooting techniques. *Numerische Mathematik*, 33:115–146, 1979.

[15] P. Deuflhard. *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms*. Springer, to be published (2001).

[16] P. Deuflhard and G. Heindl. Affine invariant convergence theorems for Newton's method and extensions to related methods. *SIAM J. Numer. Anal.*, 16:1–10, 1979.

[17] P. Deuflhard and M. Weiser. Local inexact Newton multilevel FEM for nonlinear elliptic problems. In M.-O. Bristeau, G. Etgen, W. Fitzigibbon, J.-L. Lions, J. Periaux, and M. Wheeler, editors, *Computational science for the 21st century*, pages 129–138. Wiley, 1997.

[18] A. Fischer. A special Newton-type optimization method. *Optimization*, 24:269–284, 1992.

[19] A. Hohmann. *Inexact Gauss Newton Methods for Parameter Dependent Nonlinear Problems*. PhD thesis, Free University Berlin, 1994.

[20] F. Jarre. Comparing two interior-point approaches for semi-infinite programs. Technical report, Universität Trier, April 1999.

[21] K. Malanowski. Sufficient optimality conditions for optimal control subject to state constraints. *SIAM J. Control Optimization*, 35(1):205–227, 1997.

[22] H. Maurer. First and second order sufficient optimality conditions in mathematical programming and optimal control. *Math. Programming Study*, 14:163–177, 1981.

[23] A. Miele, T. Wang, C. Y. Tzeng, and W. W. Melvin. Optimal abort landing trajectories in the presence of a windshear. *J. Optim. Theory Appl.*, 55(1):165–202, 1987.

[24] F. Potra. A path-following method for linear complementarity problems based on the affine invariant Kantorovich theorem. ZIB-Report 00-30, Konrad-Zuse-Zentrum für Informationstechnik, 2000.

[25] R. Pytlak. *Numerical Methods for Optimal Control Problems with State Constraints*. Lecture Notes in Mathematics 1707. Springer, 1999.

[26] V. Schulz. Solving discretized optimization problems by partially reduced SQP methods. *Computing and Visualization in Science*, 1:83–96, 1998.

[27] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Springer, 1993.

[28] S. Volkwein and M. Weiser. Affine invariant convergence analysis for inexact augmented Lagrangian-SQP methods. ZIB-Report 00-56, Konrad-Zuse-Zentrum für Informationstechnik, 2000.

[29] O. von Stryk and R. Bulirsch. Direct and indirect methods for trajectory optimization. *Ann. Oper. Res.*, 37:357–373, 1992.

[30] M. Weiser. *Function Space Complementarity Methods for Optimal Control Problems*. PhD thesis, Free University Berlin, 2001.

Appendix: Mathematical Model of Windshear Problem

In what follows we present the full mathematical model as given in [5], which corrected a misprint in the parameters given in [10].

In order to convert the Chebyshev-type optimization problem to the Lagrange formulation that has been assumed throughout the work, we introduce a *lower bound* ζ of the *altitude* h together with the state constraint

$$h \geq \zeta \tag{A.1}$$

and the auxiliary equation

$$\dot{\zeta} = 0\,,$$

such that the cost functional can be written as

$$J = -\int_0^T \zeta\, dt\,.$$

Horizontal and *vertical wind velocity* is given by

$$\begin{aligned} W^x &= A(x) \\ W^h &= \frac{h}{h_\star} B(x) \end{aligned}$$

with

$$A(x) = \begin{cases} -p + ax^3 + bx^4 + qx^5, & 0 \leq x \leq x_1 \\ r(x - \frac{x_3}{2}), & x_1 \leq x \leq x_2 \\ p - a(x_3 - x)^3 - b(x_3 - x)^4 - q(x_3 - x)^5, & x_2 \leq x \leq x_3 \\ p, & x_3 \leq x \end{cases}$$

$$B(x) = \begin{cases} dx^3 + ex^4 + sx^5, & 0 \leq x \leq x_1 \\ -51 \exp(-c(x - \frac{x_3}{2})^4), & x_1 \leq x \leq x_2 \\ d(x_3 - x)^3 + e(x_3 - x)^4 + s(x_3 - x)^5, & x_2 \leq x \leq x_3 \\ 0, & x_3 \leq x \end{cases}.$$

The parameter sets for the wind model from [5] are given in Table A.1.

Note that the wind model has discontinuous second derivatives at the junction points x_1, x_2, and x_3.

Under the assumption of the airplane to be a particle of constant mass moving in a vertical plane and the wind field to be steady, the following equations of motion

$x_1=$	$5.000\cdot 10^2$	ft	$d=$	$-8.02881\cdot 10^{-8}$	$\text{s}^{-1}\text{ft}^{-2}$
$x_2=$	$4.100\cdot 10^3$	ft	$e=$	$6.28083\cdot 10^{-11}$	$\text{s}^{-1}\text{ft}^{-3}$
$x_3=$	$4.600\cdot 10^3$	ft	$p=$	$5.00000\cdot 10^1$	s^{-1}ft
$a=$	$6.000\cdot 10^{-8}$	$\text{s}^{-1}\text{ft}^{-2}$	$q=$	$0.0 \quad \cdot 10^0$	
$b=$	$-4.000\cdot 10^{-11}$	$\text{s}^{-1}\text{ft}^{-3}$	$r=$	$2.50000\cdot 10^{-2}$	s^{-1}
$c=$	$-10^{-12}\ln(\frac{25}{30.6})$	ft^{-4}	$s=$	$0.0 \quad \cdot 10^0$	

Table A.1: Parameter values for the wind model.

$A_0=$	$4.4560\cdot 10^4$	lb	$B_0=$	$1.5520\cdot 10^{-1}$	
$A_1=$	$-2.3980\cdot 10^1$	lb s ft^{-1}	$B_1=$	$1.2369\cdot 10^{-1}$	rad^{-1}
$A_2=$	$1.4420\cdot 10^{-2}$	lb s^2ft^{-2}	$B_2=$	$2.4203\cdot 10$	rad^{-2}
$C_0=$	$7.2150\cdot 10^{-1}$		$\rho=$	$2.2030\cdot 10^{-3}$	lb s^2ft^{-4}
$C_1=$	$6.0877\cdot 10^0$	rad^{-1}	$S=$	$1.5600\cdot 10^3$	ft^2
$C_2=$	$-9.0277\cdot 10^0$	rad^{-2} for $p_L=2$	$\alpha_\star=$	$2.0944\cdot 10^{-1}$	rad
$mg=$	$1.5000\cdot 10^5$	lb	$\delta=$	$3.4906\cdot 10^{-2}$	rad
$g=$	$3.2172\cdot 10^1$	s^{-2}ft			

Table A.2: Parameter values for the aerodynamic forces.

can be derived:

$$\begin{aligned}
\dot{x} &= V\cos\gamma + W^x \\
\dot{h} &= V\sin\gamma + W^h \\
\dot{V} &= \frac{T}{m}\cos(\alpha+\delta) - \frac{D}{m} - g\sin\gamma - (\dot{W}^x\cos\gamma + \dot{W}^h\sin\gamma) \\
V\dot{\gamma} &= \frac{T}{m}\sin(\alpha+\delta) + \frac{L}{m} - g\cos\gamma + (\dot{W}^x\sin\gamma - \dot{W}^h\cos\gamma) \\
\dot{\alpha} &= u\,.
\end{aligned}$$

Here, x denotes the *horizontal position*, h the *altitude*, V the *relative velocity*, γ the *relative path inclination*, and α the *relative angle of attack*. The derivative of the angle of attack u is chosen as control variable. Note that due to the occurrence of the derivatives $\dot{W}^x$ and $\dot{W}^h$ in the equations of motion, the problem is C^2 as assumed by the theory only if the wind model functions A and B are C^3.

Approximations for the *aerodynamic forces* thrust T, drag D, and lift L acting

$\beta_0=$	$3.825\cdot 10^0$	
$\dot{\beta}_0=$	$2.000\cdot 10^{-1}$	s^{-1}

Table A.3: Model parameter data for the power setting.

$u_{\max}=5.236 \cdot 10^{-2}$	rad s^{-1}
$\alpha_{\max}=3.002 \cdot 10^{-1}$	rad

Table A.4: Model parameter data for the inequality constraints.

initial conditions		terminal conditions	
$x_0=0.0 \quad \cdot 10$	ft	$t_f=4.000 \cdot 10^{1}$	s
$\gamma_0=3.925 \cdot 10^{-2}$	rad	$\gamma_f=1.297 \cdot 10^{-1}$	rad
$\alpha_0=1.283 \cdot 10^{-1}$	rad		
$h_0=6.000 \cdot 10^{2}$	ft		
$V_0=2.397 \cdot 10^{2}$	ft s^{-1}		

Table A.5: Model parameter data for boundary conditions.

on the aircraft are given by

$$T = \beta(A_0 + A_1 V + A_2 V^2) \tag{A.2}$$

$$D = \frac{1}{2}(B_0 + B_1\alpha + B_2\alpha^2)\rho S V^2 \tag{A.3}$$

$$L = \frac{1}{2}(C_0 + C_1\alpha + C_2 \max(0, \alpha - \alpha_\star)^{p_L})\rho S V^2 \tag{A.4}$$

with the *power setting*

$$\beta = \min(1, \beta_0 + \dot{\beta}_0 t)$$

resulting from the additional hypothesis that, upon sensing the aircraft to be in a windshear, the pilot increases the power setting at a constant time rate until the maximum power setting is reached. Note that the lift approximation from [10] with $p_L = 2$ has a discontinuous second derivative. The model parameter data given in Table A.2 refer to a Boeing B727 aircraft powered by three JT8D-17 turbofan engines.

Simple *bounds* are imposed on the angle of attack and its derivative:

$$|u| \le u_{\max}$$
$$\alpha \le \alpha_{\max}$$

Boundary conditions are given for the initial quasi-steady flight prior to the windshear onset and for terminal steepest climb in quasi-steady flight:

$$x(0) = x_0 \qquad V(0) = V_0$$
$$h(0) = h_0 \qquad \gamma(0) = \gamma_0$$
$$\gamma(t_f) = \gamma_f \,.$$