



SpezialForschungsBereich F 32



Karl-Franzens Universität Graz  
Technische Universität Graz  
Medizinische Universität Graz



# OPTIMAL CONTROL ON AN ELLIPTIC SYSTEM WITH CONVEX POLYGONAL CONTROL CONSTRAINTS

K. Kunisch      X. Lu

SFB-Report No. 2011-003

February 2011

A-8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the  
Austrian Science Fund (FWF)



SFB sponsors:

- **Austrian Science Fund (FWF)**
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**



# OPTIMAL CONTROL ON AN ELLIPTIC SYSTEM WITH CONVEX POLYGONAL CONTROL CONSTRAINTS

KUNISCH KARL\* AND XILIANG LU\*\*

**ABSTRACT.** The semi-smooth Newton method for optimal control problems for systems of partial differential equations with polygonal constraints on the controls is developed. The Newton derivatives are characterized for the case of systems of dimension two, superlinear convergence is verified, and a simple proof-of-concept numerical example is provided.

**Key words.** Optimal control problem, control constraint, semi-smooth Newton method, superlinear convergence.

**AMS subject classification.** 49M15, 65J15.

## 1. INTRODUCTION

In recent years optimal control problems governed by partial differential equations with pointwise constraints on the controls has received a considerable amount of attention. The challenge consists in finding efficient numerical methods in spite of the non-smoothness that is introduced due to the inequality constraints and, from the point of view of numerical analysis, to obtain rate of convergence results for finite dimensional approximations.

The numerical realization of optimal control problems with constraints on the control can advantageously be performed with semi-smooth Newton methods or, equivalently by the primal dual active set method. We refer to [4, 6, 9, 10], and the references cited there. An alternative approach is based on interior point methods. It was analyzed, for example in [15, 19]. Convergence analysis for finite dimensional approximations was carried out in [8, 13, 14, 17], for example. For diffusion type problems, the semi-smooth Newton methods differently from interior point methods relies on a penalty or barrier parameter. The contributions mentioned above focused, for the most part, on the case of scalar valued controls with unilateral or bilateral constraints. In case of vector valued controls which appear for the controls of the Navier Stokes equations, for example, the constraints where of separable type.

There are only few papers which deal with the case of mixed control constraints. In [11], the semi-smooth Newton methods was investigated for a pointwise Euclidean norm constraint and in [5], a class of affine control constraints, where the number of equations characterizing the constraints is no large than the number of controls, was studied. Second order sufficient optimality conditions for general convex control constraints were obtained in [18].

---

\*Supported in part by the Fonds zur Förderung der wissenschaftlichen Forschung under SFB 32, Mathematical Optimization and Applications in the Biomedical Sciences.

\*\*Supported by the Startup Research Funds 201-275-132 by Wuhan University in China.

In this paper, we analyze the semi-smooth Newton method for constraints of polygonal type, a topic that has not been treated before. We suppose that  $\Omega$  is a bounded domain in  $\mathbb{R}^d$  and that state and control vectors map from  $\Omega$  to  $\mathbb{R}^l$ . We assume  $l = 2$  in this paper and we expect that a generalization of our approach the case of  $l$ -dimensional controls is possible.

The control variables  $\vec{u}$  are supposed to satisfy pointwise polygonal constraints

$$\vec{u}(x) \in K, \text{ for a.e. } x \in \tilde{\Omega} \subset \Omega, \text{ where } K \subset \mathbb{R}^2 \text{ is a convex, closed polygon.}$$

At the end of this section the constraints will be expressed by means of inequalities. The cost-functional is chosen to be quadratic:

$$J(\vec{y}, \vec{u}) = \frac{1}{2} \|\vec{y} - \vec{y}_d\|^2 + \frac{\alpha}{2} \|\vec{u}\|_{L^2(\tilde{\Omega})}^2,$$

where  $\vec{y}_d \in L^2(\Omega, \mathbb{R}^2)$  is given and  $\tilde{\Omega}$  is a subdomain of  $\Omega$  where the controls are localized. The equation constraint is chosen as elliptic system in  $\Omega$  with zero Dirichlet boundary conditions,

$$(1.1) \quad \Lambda \vec{y} = \mathcal{B} \vec{u} \quad \text{in } \Omega, \quad \vec{y}|_{\partial\Omega} = \vec{0},$$

where  $\Lambda$  is a linear elliptic operator and  $\mathcal{B}$  is the extension-by-zero operator from the subdomain  $\tilde{\Omega} \subset \Omega$  to  $\Omega$ . Hence  $\mathcal{B}$  is a bounded linear operator from  $L^2(\tilde{\Omega}, \mathbb{R}^2)$  to  $L^2(\Omega, \mathbb{R}^2)$ .

The optimal control problem under consideration is then given by

**Problem 1.1.**

$$\min J(\vec{y}, \vec{u}), \quad \text{subject to (1.1) and } \vec{u}(x) \in K, \text{ a.e. } x \in \tilde{\Omega}.$$

The main aim of this work is to verify that it can be solved by a semi-smooth Newton method. In passing let us also recall that linear-quadratic problems also arise as the auxiliary problems in the sequential-quadratic programming approach to genuinely nonlinear optimal control problems.

For optimal control problem with unilateral or bilateral control constraints, semi-smooth Newton method have been proved to be an efficiently superlinearly convergent technique, see [7, 10]. We briefly recall the notion of differentiability which will be used in this paper. Let  $X$  and  $Y$  be Banach spaces, with  $D$  an open set in  $X$ . The mapping  $F : D \mapsto Y$  is called Newton (slantly) differentiable at  $x \in D$ , if there exists an open neighborhood  $N_x \subset D$ , and mappings  $G : N_x \mapsto \mathcal{L}(X, Y)$  such that

$$(1.2) \quad \lim_{\|h\|_X \rightarrow 0} \frac{\|F(x+h) - F(x) - G(x+h)(h)\|_Y}{\|h\|_X} = 0.$$

With the help of Newton derivative, we may apply Newton type methods for this kind of nonlinear equation and obtain the local superlinear convergence, see [3, 7, 9, 16]. The key step to utilize the semismooth Newton method is to reformulate the optimality condition for (1.1) as a nonlinear Newton differentiable operator equation in an appropriate function space setting.

The paper is organized as follows: in the remaining part of this section we will give the notation for the function spaces and the description of constraints. Section 2 contains existence and uniqueness of the optimal solution, the first order necessary condition and equivalent formulations. Moreover Newton differentiability for the nonlinear equation representing the optimality system is proved. A semi-smooth Newton algorithm, together with its convergence analysis are given in section 3. In

last section 4, a few numerical examples are presented to depict the super-linear convergence property of the algorithm.

Throughout the paper standard notation for Sobolev spaces is used.  $L^p(\Omega, \mathbb{R}^2)$  represents the space of vector valued  $L^p$  integrable functions defined in  $\Omega$ . The norm in  $L^2$  is expressed as  $\|\cdot\|$ . Further  $|\cdot|$  expresses the Euclidean vector norm in  $\mathbb{R}^2$  or  $\mathbb{R}^m$ . The inner product  $(\cdot, \cdot)$  is taken as the integration of two  $L^2$  functions over  $\Omega$ , and  $(\cdot, \cdot)_{\tilde{\Omega}}$  stands for integration over the subdomain  $\tilde{\Omega} \subset \Omega$ .

It is assumed that the underlying system (1.1) is strongly elliptic, i.e., there exists  $\eta > 0$ , such that  $\langle \Lambda \vec{y}, \vec{y} \rangle \geq \eta \|\vec{y}\|_{H^1}$  for every  $\vec{y} \in H_0^1(\Omega, \mathbb{R}^2)$ . By the Lax-Milgram lemma, for every  $\vec{u} \in L^2(\tilde{\Omega}, \mathbb{R}^2)$ , the underlying system (1.1) admits a unique variational solution  $y \in H_0^1(\Omega, \mathbb{R}^2)$ . It is further assumed that  $\Omega$  is sufficiently regular such that  $\vec{y} \in H^2 \cap H_0^1(\Omega, \mathbb{R}^2)$  and

$$(1.3) \quad \|\vec{y}\|_{H^2(\Omega)} \leq C \|\vec{u}\|_{L^2(\tilde{\Omega})},$$

for a constant  $C$  independent of  $\vec{u} \in L^2(\tilde{\Omega}, \mathbb{R}^2)$ .

We turn to the description of the polygonal set  $K$ . It is taken as the intersection of  $m$ -half spaces. Each half space is represented by the affine inequality  $\vec{n}_i \cdot \vec{u} \leq \psi_i$ , where  $\vec{n}_i$  is the unit outer normal vector to the halfspace. We denote  $\vec{n}_i = (\cos \theta_i, \sin \theta_i)'$ , with  $\theta_i \in [0, 2\pi)$ .

We can check that the angles must satisfy  $\theta_i \neq \theta_j$ , for all  $i \neq j$ . In fact, if  $\theta_i = \theta_j$  and  $\psi_i \leq \psi_j$ , then  $\vec{n}_i \cdot \vec{p} \leq \psi_i \Rightarrow \vec{n}_j \cdot \vec{p} \leq \psi_j$ , and hence the constraint  $\vec{n}_j \cdot \vec{p} \leq \psi_j$  is redundant. Without loss of generality, we utilize the ordering  $2\pi > \theta_1 > \theta_2 > \dots > \theta_m \geq 0$ . To guarantee that  $K$  is convex we assume that  $\theta_i - \theta_{i+1} < \pi$  for  $i = 1, \dots, m$ , where  $\theta_{m+1} = \theta_1 - 2\pi$ . We denote the sides of polygonal by  $l_i$  (the two end points are not included), and the vertexes by  $a_{i,i+1}$ , (at times we use  $a_{m,1} = a_{m,m+1} = a_{0,1}$ ). Thus the sides and vertices are order in a clockwise manner, see Figure 1.1.

The polygonal constraints can be put into matrix form by introducing

$$M : \mathbb{R}^2 \rightarrow \mathbb{R}^m$$

and the vector  $\psi \in \mathbb{R}^m$  by

$$(1.4) \quad M = \begin{pmatrix} \cos \theta_1 & \sin \theta_1 \\ \vdots & \vdots \\ \cos \theta_m & \sin \theta_m \end{pmatrix}, \quad \vec{\psi} = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_m \end{pmatrix}.$$

Then the convex polygonal domain  $K$  can be expressed as

$$K = \{\vec{u} \in \mathbb{R}^2 : M\vec{u} \leq \vec{\psi}\}.$$

## 2. OPTIMALITY SYSTEM

We define the subset of admissible controls  $\vec{u} \in L^2(\tilde{\Omega}, \mathbb{R}^2)$  as

$$(2.1) \quad D = \{\vec{u} \in L^2(\tilde{\Omega}, \mathbb{R}^2) : \vec{u}(x) \in K, \text{ a.e. } x \in \tilde{\Omega}\}.$$

Clearly  $D$  is a closed convex set in  $L^2(\tilde{\Omega}, \mathbb{R}^2)$ . Denote the control-to-state-mapping  $\vec{u} \rightarrow \vec{y}(\vec{u})$  by  $T$ , where  $T : L^2(\tilde{\Omega}, \mathbb{R}^2) \rightarrow L^2(\Omega, \mathbb{R}^2)$  is given by  $T = \Lambda^{-1}\mathcal{B}$ . Hence Problem 1.1 can equivalently be expressed as

$$(2.2) \quad \inf_{\vec{u} \in D} \hat{J}(\vec{u}),$$

where  $\hat{J}(\vec{u}) = J(T\vec{u}, \vec{u})$  denotes the reduced cost-functional.

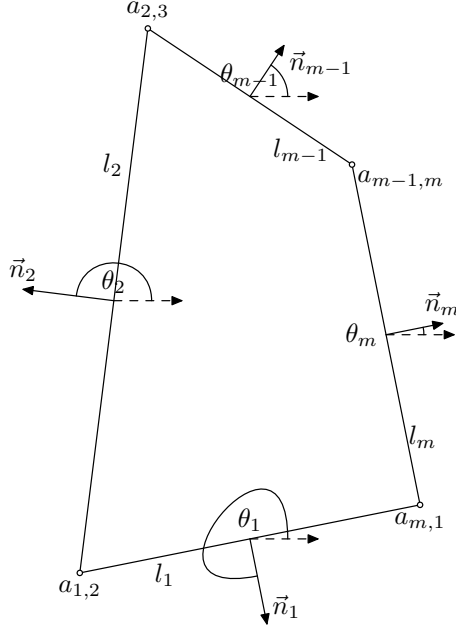


FIGURE 1.1. Notation for the constraints

**Theorem 2.1.** *There exists a unique solution to Problem 1.1.*

*Proof.* Let  $\{\vec{u}^n\} \subset D$  denote a minimizing sequence. Due to the control-cost term this sequence is bounded in  $L^2(\tilde{\Omega}, \mathbb{R}^2)$ . Passage to a subsequence, still denoted by  $\{\vec{u}^n\}$ , we have

$$\vec{u}^n \rightarrow \vec{u} \text{ weakly in } L^2(\tilde{\Omega}, \mathbb{R}^2),$$

for some  $\vec{u} \in L^2(\tilde{\Omega}, \mathbb{R}^2)$ . By closedness and convexity of  $D$  we have  $\vec{u} \in D$ . Next we notice that the control to state map  $T$  maps weakly convergent sequences in  $L^2(\tilde{\Omega}, \mathbb{R}^2)$  to weakly convergent sequences in  $L^2(\Omega, \mathbb{R}^2)$ . Together with the weakly lower semi-continuous of the norm functionals, we obtain that the weak limit function  $\vec{u}$  is indeed an optimal solution. Uniqueness is a consequence of strict convexity of the cost function  $\hat{J}(\vec{u})$  and convexity of  $D$ .  $\square$

Next we present the optimality system for Problem 1.1 for the minimizer  $(\vec{y}^*, \vec{u}^*)$ . Since  $\hat{J}$  is differentiable and  $D$  is a closed convex set, we have

$$(\hat{J}'(\vec{u}^*), \vec{u}^* - \vec{v}) \leq 0, \quad \text{for all } \vec{v} \in D.$$

From the definition,

$$\hat{J}'(\vec{u}^*) = T^*(T\vec{u}^* - \vec{y}_d) + \alpha\vec{u}^* = \mathcal{B}^*\Lambda^{-*}(\vec{y}^* - \vec{y}_d) + \alpha\vec{u}^*.$$

We define the dual variable  $\vec{p}^* = -\Lambda^{-*}(\vec{y}^* - \vec{y}_d)$ . Then we find the optimality system in the form:

$$\begin{aligned}
 (2.3) \quad & \text{Primal equation} & \Lambda \vec{y}^* &= \mathcal{B} \vec{u}^*, & \vec{y}^*|_{\partial\Omega} &= \vec{0}, \\
 & \text{Adjoint equation} & \Lambda^* \vec{p}^* &= \vec{y}_d - \vec{y}^*, & \vec{p}^*|_{\partial\Omega} &= \vec{0}, \\
 & \text{Optimality condition} & (\mathcal{B}^* \vec{p}^* - \alpha \vec{u}^*, \vec{u}^* - \vec{v}) &\geq 0, & \text{for all } \vec{v} \in D,
 \end{aligned}$$

where  $\mathcal{B}^* : L^2(\Omega, \mathbb{R}^2) \rightarrow L^2(\tilde{\Omega}, \mathbb{R}^2)$  denotes the adjoint operator of  $\mathcal{B}$ , which is the restriction operator from  $\Omega$  to  $\tilde{\Omega}$ .

**Lemma 2.2.** *The optimality condition in (2.3) is equivalent to*

$$(2.4) \quad \vec{u}^* = \text{Proj}_D\left(\frac{1}{\alpha} \mathcal{B}^* \vec{p}^*\right),$$

which is equivalent to the pointwise projection

$$(2.5) \quad \vec{u}^*(x) = \text{Proj}_K\left(\frac{1}{\alpha} \vec{p}^*(x)\right), \text{ for a.e. } x \in \tilde{\Omega}.$$

Here  $\text{Proj}_D$  and  $\text{Proj}_K$  denote the projections to  $D$  in  $L^2(\tilde{\Omega}, \mathbb{R}^2)$ , and to  $K$  in  $\mathbb{R}^2$ , respectively.

*Proof.* The first equivalence stated above follows from the projection to a general closed convex subset in a Hilbert space. Equivalence between the pointwise and function space projection follows, for instance, by observing that for  $x \in \tilde{\Omega}$

$$\text{Proj}_K\left(\frac{1}{\alpha} \vec{p}^*(x)\right) = \text{argmin}_{\vec{v} \in K} \left| \frac{1}{\alpha} \vec{p}^*(x) - \vec{v} \right|^2,$$

and uniqueness of projection onto convex sets.  $\square$

By introducing Lagrange multipliers we shall obtain a complementarity system that is equivalent to the pointwise projection.

**Lemma 2.3.** *For every vector  $\vec{p} \in \mathbb{R}^2$ , the projection  $\vec{u} = \text{Proj}_K(\frac{1}{\alpha} \vec{p})$  is equivalent to the existence of a unique vector  $\vec{\lambda} \in \mathbb{R}^m$ , such that*

$$(2.6) \quad \vec{p} = \alpha \vec{u} + M^T \vec{\lambda}, \quad \vec{\lambda} \geq \vec{0}, \quad M \vec{u} - \vec{\psi} \leq \vec{0}, \quad \vec{\lambda} \cdot (M \vec{u} - \vec{\psi}) = 0.$$

*Proof.* Note at first that

$$(2.7) \quad \vec{u} = \text{argmin}_{\vec{v} \in K} \frac{1}{2} \left| \frac{1}{\alpha} \vec{p} - \vec{v} \right|^2.$$

Due to the assumption  $\theta_i - \theta_{i+1} \in (0, \pi)$  the solution  $\vec{u}$  satisfies the regular point condition for the constraint  $M \vec{v} \leq \vec{\psi}$ , for each  $\vec{p} \in \mathbb{R}^2$ . Hence by standard Lagrange multiplier theory for inequality constraints there exists  $\vec{\lambda} \in \mathbb{R}^m$  such that (2.6) holds, [9, 12]. For proof of semi-smoothness in Lemma 2.6 below, we use an explicit expression for  $\vec{\lambda}$  in terms of  $\vec{p}$  which we derive next.

For this purpose the  $\vec{p}$ -space  $\mathbb{R}^2$  is decomposed into disjoint subsets,  $\mathcal{I}$ ,  $\mathcal{A}_i$ ,  $\mathcal{A}_{i,i+1}$  for  $i = 1, \dots, m$  ( $\mathcal{A}_{m,m+1} = \mathcal{A}_{m,1}$ ), see Figure 2.2. The choice of the

Lagrange multiplier  $\vec{\lambda}$  depends on which subdomain the vector  $\vec{p}$  belongs to. Three types of subdomains are defined as follows:

$$(2.8) \quad \begin{cases} \mathcal{I} = \{\vec{h} : Proj_K(\vec{h}) \in \text{interior of } K\}, \\ \mathcal{A}_i = \{\vec{h} : Proj_K(\vec{h}) \in l_i\}, \\ \mathcal{A}_{i,i+1} = \{\vec{h} : Proj_K(\vec{h}) = a_{i,i+1}\}. \end{cases}$$

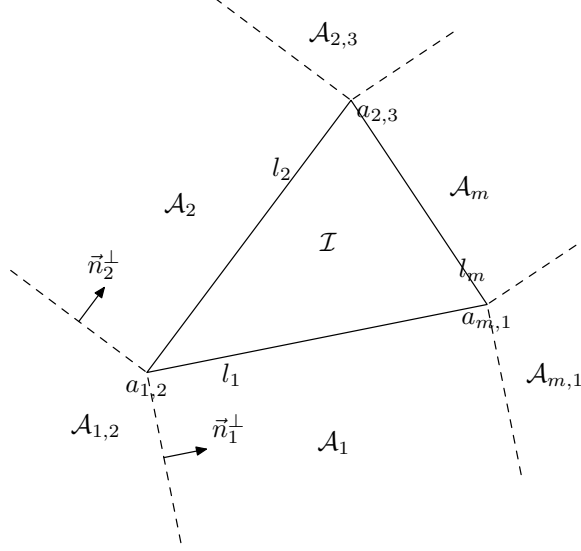


FIGURE 2.2. The decomposition of  $\vec{p}$  plane

The dotted lines in Figure 2.2 are perpendicular to adjacent side of polygon.

Case 1: If  $\frac{1}{\alpha}\vec{p} \in \mathcal{I}$ , then all the constraints are inactive, and hence  $\vec{u} = \frac{1}{\alpha}\vec{p}$  and  $\vec{\lambda} = \vec{0}$ .

Case 2: If  $\frac{1}{\alpha}\vec{p} \in \mathcal{A}_i$ , then the constraints  $\frac{1}{\alpha}\vec{n}_j \cdot \vec{p} \leq \psi_j$ , are not active for all  $j \neq i$ , and hence  $\lambda_j = 0$  for all  $j \neq i$ . By definition of projection operator we have for some  $\lambda_i \geq 0$

$$\frac{1}{\alpha}\vec{p} - \vec{u} = \frac{1}{\alpha}\lambda_i\vec{n}_i,$$

$$\vec{u} \cdot \vec{n}_i = \psi_i.$$

Solving this linear system, we obtain

$$\lambda_i = \vec{p} \cdot \vec{n}_i - \alpha\psi_i, \lambda_j = 0 \text{ for } j \neq i, \quad \vec{u} = \frac{1}{\alpha}\vec{p} - \frac{1}{\alpha}(\vec{p} \cdot \vec{n}_i)\vec{n}_i + \psi_i\vec{n}_i.$$

This provides the unique representation of  $\vec{\lambda}$  and  $\vec{u}$  on  $\cup_{i=1}^m \mathcal{A}_i$ .

Case 3: If  $\frac{1}{\alpha}\vec{p} \in \mathcal{A}_{i,i+1}$ , the constraints  $\frac{1}{\alpha}\vec{n}_j \cdot \vec{p} \leq \psi_i$ , for  $j \neq i, j \neq i+1$  are not active, and hence necessarily  $\lambda_j = 0$ , for all  $j = 1, \dots, m$  with  $j \neq i, j \neq i+1$ . The vertex  $a_{i,i+1}$  is presented by the intersection of two lines and we have  $\vec{u} = M_{i,i+1}^{-1}\vec{\psi}_{i,i+1}$ ,



where  $M_{i,i+1}$  is the square matrix arising from  $i$ -th and  $i+1$ -st rows of  $M$ . Here and below, for  $i = m$ , the index  $m+1$  is set equal to the index 1. Further  $\vec{\lambda}_{i,i+1}$  (or  $\vec{\psi}_{i,i+1}$ ) denotes the  $i$ -th and  $i+1$ -st coordinates of  $\vec{\lambda}$  (or  $\vec{\psi}$ ). Therefore

$$\vec{\lambda}_{i,i+1} = \alpha M_{i,i+1}^{-T} \left( \frac{1}{\alpha} \vec{p} - \vec{u} \right) = M_{i,i+1}^{-T} \vec{p} - \alpha (M_{i,i+1} M_{i,i+1}^T)^{-1} \vec{\psi}_{i,i+1},$$

where

$$M_{i,i+1}^{-T} = -\frac{1}{\sin(\theta_i - \theta_{i+1})} \begin{pmatrix} \sin \theta_{i+1} & -\cos \theta_{i+1} \\ -\sin \theta_i & \cos \theta_i \end{pmatrix},$$

$$(M_{i,i+1} M_{i,i+1}^T)^{-1} = \frac{1}{\sin^2(\theta_i - \theta_{i+1})} \begin{pmatrix} 1 & -\cos(\theta_i - \theta_{i+1}) \\ -\cos(\theta_i - \theta_{i+1}) & 1 \end{pmatrix}.$$

Then for  $\frac{1}{\alpha} \vec{p} \in \mathcal{A}_{i,i+1}$  we find  $\lambda_j = 0$  for  $j \neq i, j \neq i+1$ , and

$$(2.9) \quad \begin{aligned} \lambda_i &= \frac{1}{\sin^2(\theta_i - \theta_{i+1})} [-\sin(\theta_i - \theta_{i+1})(p_i \sin \theta_{i+1} - p_{i+1} \cos \theta_{i+1}) - \alpha \psi_i + \alpha \cos(\theta_i - \theta_{i+1}) \psi_{i+1}], \\ \lambda_{i+1} &= \frac{1}{\sin^2(\theta_i - \theta_{i+1})} [-\sin(\theta_i - \theta_{i+1})(p_{i+1} \cos \theta_i - p_i \sin \theta_i) - \alpha \psi_{i+1} + \alpha \cos(\theta_i - \theta_{i+1}) \psi_i]. \end{aligned}$$

We have thus obtained the unique representation of  $\vec{\lambda}$  as a function of  $\vec{p}$ .  $\square$

It will be convenient to summarize the representation of  $\vec{\lambda}(\vec{p})$  that was obtained in the proof of the previous lemma as:

$$(2.10) \quad \lambda_i(\vec{p}) = \begin{cases} \vec{p} \cdot n_i - \alpha \psi_i & \text{if } \vec{p} \in \mathcal{A}_i \\ \frac{1}{\sin^2(\theta_i - \theta_{i+1})} \begin{pmatrix} -\sin(\theta_i - \theta_{i+1})(p_i \sin \theta_{i+1} - p_{i+1} \cos \theta_{i+1}) \\ -\alpha \psi_i + \alpha \cos(\theta_i - \theta_{i+1}) \psi_{i+1} \end{pmatrix} & \text{if } \vec{p} \in \mathcal{A}_{i,i+1}, \\ \frac{1}{\sin^2(\theta_i - \theta_{i-1})} \begin{pmatrix} \sin(\theta_i - \theta_{i-1})(p_i \cos \theta_{i-1} - p_{i-1} \sin \theta_{i-1}) \\ -\alpha \psi_i + \alpha \cos(\theta_i - \theta_{i-1}) \psi_{i-1} \end{pmatrix} & \text{if } \vec{p} \in \mathcal{A}_{i-1,i}, \\ 0 & \text{if } \vec{p} \in \mathcal{I}, \end{cases}$$

where  $\theta_0 = \theta_m, p_0 = p_m, \mathcal{A}_{0,1} = \mathcal{A}_{m,1}, \theta_{m+1} = \theta_1, p_{m+1} = p_1, \mathcal{A}_{m,m+1} = \mathcal{A}_{m,1}$ .

**Lemma 2.4.** *The mapping  $H : \vec{p} \rightarrow \vec{\lambda}(\vec{p})$  from  $\mathbb{R}^2 \mapsto \mathbb{R}^m$  is piecewise affine and locally Lipschitz continuous.*

*Proof.* The fact that  $H$  is piecewise affine follows directly from (2.10). Locally Lipschitz continuity follows from an abstract result on the sensitivity of solutions and Lagrange multipliers with respect to problem data, in abstract optimization problems see e.g. [9] page 46, and the references given there. It is applied to (2.7), with  $\vec{p}$  denoting the perturbation parameter.  $\square$

By collecting the point-wise information obtained in the two previous lemmas, we obtain a Lagrange multiplier associated to  $\vec{u} = \text{Proj}_D(\frac{1}{\alpha} \mathcal{B}^* \vec{p})$  in function space.

**Lemma 2.5.** *For any  $\vec{p} \in L^2(\Omega, \mathbb{R}^2)$ , the projection  $\vec{u} = \text{Proj}_D(\frac{1}{\alpha} \mathcal{B}^* \vec{p})$  is equivalent to the existence of a unique  $\vec{\lambda} \in L^2(\hat{\Omega}, \mathbb{R}^m)$ , such that*

$$(2.11) \quad \mathcal{B}^* \vec{p} = \alpha \vec{u} + M^T \vec{\lambda}, \quad \vec{\lambda} \geq \vec{0}, \quad M \vec{u} - \vec{\psi} \leq \vec{0}, \quad (\vec{\lambda}, M \vec{u} - \vec{\psi})_{\hat{\Omega}} = 0.$$

*Proof.* For  $x \in \tilde{\Omega}$  we define pointwise almost everywhere  $\vec{\lambda}(x)$  according to Lemma 2.3. From the explicit representation (2.10) we follow that  $\vec{\lambda} \in L^2(\tilde{\Omega}, \mathbb{R}^m)$  for  $\vec{p} \in L^2(\Omega, \mathbb{R}^2)$ . The pointwise information (2.6) implies that (2.11) holds.  $\square$

**Lemma 2.6.**  $\vec{\lambda} = H(\mathcal{B}^* \vec{p})$  is Newton differentiable form  $L^q(\Omega, \mathbb{R}^2)$  to  $L^p(\tilde{\Omega}, \mathbb{R}^m)$ , for any  $1 \leq p < q \leq \infty$ .

*Proof.* It clearly suffices to consider the case when  $\tilde{\Omega} = \Omega$ , i.e. when  $\mathcal{B}^*$  equals the identity. In the first part of the proof we verify Newton differentiability for a special piecewise linear function. In the second step we reduce the general case to the special one. We shall use the fact that by the Hölder's inequality, for any  $1 \leq p < q \leq \infty$ ,

$$\|w\|_{L^p(\tilde{\Omega})} \leq |\tilde{\Omega}|^r \|w\|_{L^q(\tilde{\Omega})}, \quad \text{where } r = \begin{cases} \frac{q-p}{pq} & \text{for } q < \infty, \\ \frac{1}{p} & \text{for } q = \infty. \end{cases}$$

Subsequently we only consider the case  $q < \infty$ , and leave modifications for the case  $q = \infty$  to the reader.

Step 1: Let  $\mathbb{R}^2$  be decomposed to three mutually disjoint subsets according to

$$I_1 = \{\vec{p} : p_1 > 0, p_2 > 0\}, \quad I_2 = \{\vec{p} : p_1 + p_2 > 0, p_2 \leq 0\}, \quad I_3 = \mathbb{R}^2 - I_1 - I_2.$$

Accordingly we define the piecewise linear function  $f(\vec{p}) : \mathbb{R}^2 \rightarrow \mathbb{R}$  as

$$f(\vec{p}) = \begin{cases} p_1 & \vec{p} \in I_1 \\ p_1 + p_2 & \vec{p} \in I_2 \\ 0 & \vec{p} \in I_3 \end{cases},$$

see Figure 2.3.

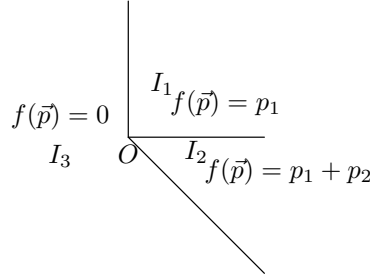


FIGURE 2.3. Piecewise linear function

Moreover we define the vector function  $G$ :

$$G(\vec{p}) = \begin{cases} (1, 0)' & \vec{p} \in I_1 \\ (1, 1)' & \vec{p} \in I_2 \\ (0, 0)' & \vec{p} \in I_3 \end{cases}.$$

We will proof the function  $G$  is one of Newton derivative of  $f$ . Next we consider the substitution operator generated by  $f$  and denoted be the same symbol, as mapping

from  $L^q(\Omega, \mathbb{R}^2)$  to  $L^p(\Omega, \mathbb{R})$ . We shall show that  $f$  is Newton-differentiable with Newton derivative given by the substitution operator generated by  $G$ .

For this purpose we partition  $\Omega$  as follows:

$$\Omega_0 = \{x \in \Omega : \vec{p}(x) \text{ and } (\vec{p} + \vec{h})(x) \text{ lie in same subset } I_i, \text{ for } i = 1, 2, 3\},$$

$$\Omega_{1,2} = \{x \in \Omega : \vec{p}(x) \in I_1, (\vec{p} + \vec{h})(x) \in I_2\},$$

and  $\Omega_{2,1}, \Omega_{1,3}, \Omega_{3,1}, \Omega_{2,3}, \Omega_{3,2}$  are defined analogously. Let the index set  $I = \{0, (1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2)\}$ . It is can be checked that  $\Omega_i, i \in I$  are disjoint subsets of  $\Omega$  and that their union is  $\Omega$ . For any increment  $\vec{h} \in L^q(\Omega, \mathbb{R}^2)$  we define the remainder term

$$R = f(\vec{p} + \vec{h}) - f(\vec{p}) - G(\vec{p} + \vec{h}) \cdot \vec{h}.$$

By definition of Newton derivative, it is sufficient to show that

$$\lim_{\|\vec{h}\|_{L^q(\tilde{\Omega})} \rightarrow 0} \frac{\|R\|_{L^p(\tilde{\Omega})}}{\|\vec{h}\|_{L^q(\tilde{\Omega})}} = 0.$$

A computation shows that for  $x \in \Omega_i, i \in I$ , we have

$$R = \begin{cases} 0 & i = 0 \\ p_2 & i = (1, 2) \\ -p_2 & i = (2, 1) \\ -p_1 & i = (1, 3) \\ p_1 & i = (3, 1) \\ -p_1 - p_2 & i = (2, 3) \\ p_1 + p_2 & i = (3, 2). \end{cases}$$

Since  $\Omega_i$  forms a disjoint partition of  $\Omega$ , we have

$$\|R\|_{L^p(\Omega)} \leq \sum_{i \in I} \|R\|_{L^p(\Omega_i)}.$$

By definition, we need to show that for any small  $\epsilon > 0$ , there exists a  $\delta$ , such that for any  $\|\vec{h}\|_{L^q(\tilde{\Omega})} \leq \delta$ , we have

$$\frac{\sum_{i \in I} \|R\|_{L^p(\Omega_i)}}{\|\vec{h}\|_{L^q(\tilde{\Omega})}} \leq \epsilon.$$

It is sufficient to check that this is true for each  $i \in I$  separately. Consider at first the domain  $\Omega_{1,2}$ . For any  $x \in \Omega_{1,2}$ , by definition we have

$$\vec{p}(x) \in I_1, \quad (\vec{p} + \vec{h})(x) \in I_2,$$

which implies that  $p_2(x) > 0$ ,  $(p_2 + h_2)(x) \leq 0$ , and in particular that

$$(2.12) \quad |p_2(x)| \leq |h_2(x)| \text{ for } x \in \Omega_{1,2}.$$

For the subdomain  $\Omega_\eta$  given by

$$\Omega_\eta = \{x \in \Omega : 0 < p_2(x) < \eta\}.$$

we have  $\lim_{\eta \rightarrow 0^+} |\Omega_\eta| = 0$ , where  $|\cdot|$  denotes the measure of the given domain. Hence there exists an  $\eta$  independent of  $h$ , such that

$$|\Omega_\eta| \leq \left(\frac{\epsilon}{2}\right)^{\frac{pq}{q-p}}.$$

For any  $x \in \Omega_{1,2} \setminus \Omega_\eta$ , we have  $|h_2(x)| \geq |p_2(x)| \geq \eta$ , which implies that

$$\|\vec{h}\|_{L^q(\Omega)} \geq \|h_2\|_{L^q(\Omega)} \geq \|h_2\|_{L^q(\Omega_{1,2} \setminus \Omega_\eta)} \geq \eta |\Omega_{1,2} \setminus \Omega_\eta|.$$

The choice  $\delta = \eta \left(\frac{\epsilon}{2}\right)^{\frac{pq}{q-p}}$  guarantees that for every  $\|\vec{h}\|_{L^q(\Omega)} \leq \delta$ , we have

$$\begin{aligned} \frac{\|R\|_{L^p(\Omega_{1,2})}}{\|\vec{h}\|_{L^q(\Omega)}} &\leq \frac{\|R\|_{L^p(\Omega_{1,2} \cap \Omega_\eta)}}{\|\vec{h}\|_{L^q(\Omega)}} + \frac{\|R\|_{L^p(\Omega_{1,2} \setminus \Omega_\eta)}}{\|\vec{h}\|_{L^q(\Omega)}} \\ &\leq |\Omega_{1,2} \cap \Omega_\eta|^{\frac{q-p}{pq}} \frac{\|p_2\|_{L^q(\Omega_{1,2} \cap \Omega_\eta)}}{\|h_2\|_{L^q(\Omega)}} + |\Omega_{1,2} \setminus \Omega_\eta|^{\frac{q-p}{pq}} \frac{\|p_2\|_{L^q(\Omega_{1,2} \setminus \Omega_\eta)}}{\|h_2\|_{L^q(\Omega)}} \\ &\leq |\Omega_\eta|^{\frac{q-p}{pq}} + |\Omega_{1,2} \setminus \Omega_\eta|^{\frac{q-p}{pq}} \leq \frac{\epsilon}{2} + \left(\frac{\|\vec{h}\|_{L^q(\Omega)}}{\eta}\right)^{\frac{q-p}{pq}} \leq \epsilon. \end{aligned}$$

The remaining subsets  $\Omega_i$  can be treated analogously. We notice that the observation (2.12) plays essential role in the proof. For the other subdomain  $\Omega_i$ ,  $i \in I$ , we have

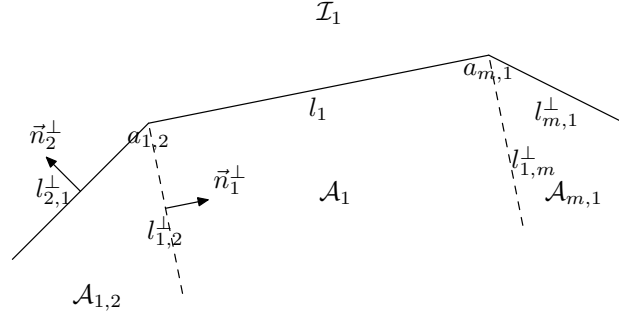
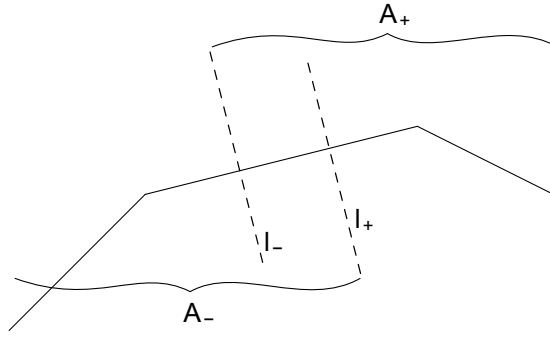
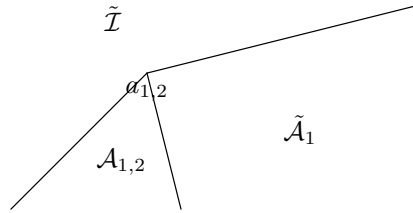
- $|p_2(x)| \leq h_2(x)$  if  $x \in \Omega_{2,1}$ ,
- $|p_1(x)| \leq \sqrt{2}|\vec{h}(x)|$  if  $x \in \Omega_{1,3}$  or  $x \in \Omega_{3,1}$ ,
- $|p_1(x) + p_2(x)| \leq \sqrt{2}|\vec{h}(x)|$  if  $x \in \Omega_{2,3}$  or  $x \in \Omega_{3,2}$ .

The case  $x \in \Omega_{1,3}$  for example, can be checked as follows (other cases are very similar). Either  $p_1(x) + h_1(x) \leq 0$ , then  $0 \leq p_1(x) \leq |h_1(x)|$  or  $p_1(x) + h_1(x) > 0$ , then  $0 \leq p_1(x) \leq -h_1(x) - h_2(x) - p_2(x) \leq |h_1(x)| + |h_2(x)| \leq \sqrt{2}|\vec{h}(x)|$ . In either case  $|p_1(x)| \leq \sqrt{2}|\vec{h}(x)|_{\mathbb{R}^2}$ .

So far we proved that  $f(\vec{p})$  is Newton differentiable from  $L^q(\Omega, \mathbb{R}^2)$  to  $L^p(\Omega)$ . Next we show that  $\vec{\lambda} = H(\vec{p})$  is also Newton differentiable as mapping from  $L^q(\Omega, \mathbb{R}^2)$  to  $L^p(\Omega, \mathbb{R}^m)$ .

Step 2: With reference to (2.10) it suffices to consider one coordinate of  $\vec{\lambda}(\vec{p})$ . Without loss of generality we show that  $\lambda_1$  is Newton differentiable as a function of  $\vec{p}$  from  $L^q(\Omega, \mathbb{R}^2)$  to  $L^p(\Omega)$ . It will be convenient to refer back to Figure 2.2, see also Figure 2.4 and to recall the formula for  $\lambda_1$ :

$$(2.13) \quad \lambda_1(\vec{p}) = \begin{cases} \vec{p} \cdot \vec{n}_1 - \alpha \psi_1 & \vec{p} \in \mathcal{A}_1 \\ \frac{1}{\sin^2(\theta_1 - \theta_2)} \begin{pmatrix} -\sin(\theta_1 - \theta_2)(p_1 \sin \theta_2 - p_2 \cos \theta_2) \\ -\alpha \psi_1 + \alpha \cos(\theta_1 - \theta_2) \psi_2 \end{pmatrix} & \vec{p} \in \mathcal{A}_{1,2} \\ \frac{1}{\sin^2(\theta_1 - \theta_m)} \begin{pmatrix} \sin(\theta_1 - \theta_m)(p_1 \cos \theta_m - p_1 \sin \theta_m) \\ -\alpha \psi_1 + \alpha \cos(\theta_1 - \theta_m) \psi_m \end{pmatrix} & \vec{p} \in \mathcal{A}_{m,1} \\ 0 & \text{otherwise} \end{cases}.$$

FIGURE 2.4. Lagrange multiplier  $\lambda_1$ FIGURE 2.5. Partition for  $\lambda_1$ FIGURE 2.6. Subdomains for  $\lambda_{1,-}$ 

We next decompose  $\lambda_1$  into two additive parts by the following construction. Choose two lines  $\hat{l}_-$  and  $\hat{l}_+$  which are perpendicular to  $l_1$ , see Figure 2.5, and denote the half-space to the right of  $\hat{l}_-$  by  $\hat{A}_+$ , and the half-space to the left of  $\hat{l}_+$  by  $\hat{A}_-$ . One can find two positive smooth functions  $\xi_+$  and  $\xi_-$  which have support in  $\hat{A}_+$  and  $\hat{A}_-$ , which satisfy  $\xi_+ + \xi_- = 1$ . We further introduce  $\lambda_{1,-}$ , see Figure

2.6, by

$$\lambda_{1,-}(\vec{p}) = \begin{cases} \vec{p} \cdot \vec{n}_1 - \alpha\psi_1 & \vec{p} \in \tilde{\mathcal{A}}_1 \\ \lambda_1 & \vec{p} \in \mathcal{A}_{1,2} \\ 0 & \vec{p} \in \tilde{\mathcal{I}} = \Omega \setminus (\tilde{\mathcal{A}}_1 \cup \mathcal{A}_{1,2}) \end{cases}.$$

where  $\tilde{\mathcal{A}}_1$  denotes the sector depicted in Figure 2.6. Further  $\lambda_{1,+}$  is defined analogously  $\lambda_{1,-}$  with  $\mathcal{A}_{1,2}$  replaced by  $\mathcal{A}_{m,1}$ .

With these preliminaries the Lagrange multiplier  $\lambda_1$  can be decomposed as

$$(2.14) \quad \lambda_1 = \lambda_1 \xi_- + \lambda_1 \xi_+ = \lambda_{1,-} \xi_- + \lambda_{1,+} \xi_+,$$

We next argue that  $\lambda_{1,-}$  and  $\lambda_{1,+}$  are affine transformation of  $f$  which was defined in step 1. Without loss of generality we focus Newton differentiability of function  $\lambda_{1,-}$ .

We first consider the case  $\theta_1 - \theta_2 \in (0, \pi/2)$ , the case  $\theta_1 - \theta_2 \in [\pi/2, \pi)$  will be treated later. We define a new variable  $\vec{q}$  according to

$$(2.15) \quad \vec{q} = \begin{pmatrix} \cos \theta_1 & \sin \theta_1 \\ -\cot(\theta_1 - \theta_2) \sin \theta_1 & \cot(\theta_1 - \theta_2) \cos \theta_1 \end{pmatrix} (\vec{p} - \alpha M_{1,2}^{-1} \vec{\psi}_{1,2}).$$

A computation shows that

$$\begin{aligned} q_1 &= \cos \theta_1 (p_1 + \frac{\alpha}{\sin(\theta_1 - \theta_2)} (\sin \theta_2 \psi_1 - \sin \theta_1 \psi_2)) \\ &\quad + \sin \theta_1 (p_2 + \frac{\alpha}{\sin(\theta_1 - \theta_2)} (-\cos \theta_2 \psi_1 + \cos \theta_1 \psi_2)) \\ &= \vec{n}_1 \cdot \vec{p} - \alpha \psi_1, \end{aligned}$$

$$q_2 = \cot(\theta_1 - \theta_2) \vec{n}_1^\perp \cdot \vec{p}_{1,2},$$

where

$$\vec{n}_1^\perp = (\cos(\theta_1 + \frac{\pi}{2}), \sin(\theta_1 + \frac{\pi}{2}))' = (-\sin \theta_1, \cos \theta_1)',$$

and

$$(2.16) \quad \vec{p}_{1,2} = \vec{p}_{1,2}(\vec{p}) = \vec{p} - \alpha M_{1,2}^{-1} \vec{\psi}_{1,2} = \begin{pmatrix} p_1 + \frac{\alpha}{\sin(\theta_1 - \theta_2)} (\sin \theta_2 \psi_1 - \sin \theta_1 \psi_2) \\ p_2 + \frac{\alpha}{\sin(\theta_1 - \theta_2)} (-\cos \theta_2 \psi_1 + \cos \theta_1 \psi_2) \end{pmatrix}.$$

We also define

$$\vec{n}_2^\perp = (\cos(\theta_2 - \frac{\pi}{2}), \sin(\theta_2 - \frac{\pi}{2}))' = (\sin \theta_2, -\cos \theta_2)',$$

then from Figure 2.4, it is noted that

$$l_{1,2}^\perp = \{\vec{p} : \vec{n}_1^\perp \cdot \vec{p}_{1,2}(\vec{p}) = 0\}, \quad l_{2,1}^\perp = \{\vec{p} : \vec{n}_2^\perp \cdot \vec{p}_{1,2}(\vec{p}) = 0\},$$

and

$$(2.17) \quad \begin{cases} \vec{n}_1^\perp \cdot \vec{p}_{1,2}(\vec{p}) > 0 & \text{for } \vec{p} \in \mathcal{A}_1, \\ \vec{n}_1^\perp \cdot \vec{p}_{1,2}(\vec{p}) \leq 0, \text{ and } \vec{n}_2^\perp \cdot \vec{p}_{1,2}(\vec{p}) < 0 & \text{for } \vec{p} \in \mathcal{A}_{1,2}. \end{cases}$$

One can check:

1. For  $\vec{p} \in \tilde{\mathcal{A}}_1$ , we have  $q_1 > 0$  and  $q_2 = \cot(\theta_1 - \theta_2) \vec{n}_1^\perp \cdot \vec{p}_{1,2} > 0$ . Hence  $\vec{q} \in I_1$  (see

Figure 2.3) and  $\lambda_{1,-}(\vec{p}) = q_1 = f(\vec{q})$ .

2. For  $\vec{p} \in \mathcal{A}_{1,2}$ , we have  $q_2 \leq 0$  and

$$(2.18) \quad \begin{aligned} q_1 + q_2 &= (\vec{n}_1 + \cot(\theta_1 - \theta_2)\vec{n}_1^\perp)\vec{p}_{1,2}(\vec{p}) + \alpha(\vec{n}_1 \cdot M_{1,2}^{-1}\vec{\psi}_{1,2} - \psi_1) \\ &= -\frac{1}{\sin(\theta_1 - \theta_2)}\vec{n}_2^\perp \cdot \vec{p}_{1,2}(\vec{p}) > 0, \end{aligned}$$

hence  $\vec{q} \in I_2$ . By (2.13), and (2.16), we find

$$\lambda_{1,2}(\vec{p}) = q_1 + q_2 = f(\vec{q}).$$

3. For  $\vec{p} \in \tilde{\mathcal{A}}$ , similarly  $\vec{q} \in I_3$  and  $\lambda_{1,2}(\vec{p}) = 0 = f(\vec{q})$ .

Therefore for  $0 < \theta_1 - \theta_2 < \frac{\pi}{2}$  we established that

$$\lambda_{1,-}(\vec{p}) = f(\vec{q}(\vec{p})),$$

and by the chain rule, [9], page 238,  $\vec{p} \rightarrow \lambda_{1,-}(\vec{p})$  is Newton-differentiable.

The case  $\theta_1 - \theta_2 \in [\frac{\pi}{2}, \pi)$  can be treated similarly. We define variable  $\vec{s}$  as

$$\vec{s} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} \vec{n}_1 \cdot \vec{p}_{1,2} + \cot(\theta_1 - \theta_2)\vec{n}_1^\perp \cdot \vec{p}_{1,2} \\ -\cot(\theta_1 - \theta_2)\vec{n}_1^\perp \cdot \vec{p}_{1,2} \end{pmatrix}.$$

Using (2.17) we find:

1. For  $\vec{p} \in \tilde{\mathcal{A}}_1$ ,  $s_2 \leq 0$  and  $s_1 + s_2 > 0$ , hence  $\vec{s} \in I_2$  (see Figure 2.3) and  $\lambda_{1,2}(\vec{p}) = s_1 + s_2 = f(\vec{s})$ .

2. For  $\vec{p} \in \mathcal{A}_{1,2}$ ,  $s_2 > 0$ , and (2.18) implies that  $s_1 = -\frac{1}{\sin(\theta_1 - \theta_2)}\vec{n}_2^\perp \cdot \vec{p}_{1,2}(\vec{p}) > 0$ , hence  $\vec{q} \in I_1$  and  $\lambda_{1,2}(\vec{p}) = s_1 = f(\vec{s})$ .

3. For  $\vec{p} \in \tilde{\mathcal{A}}$ , similarly  $\vec{s} \in I_3$  and  $\lambda_{1,2}(\vec{p}) = 0 = f(\vec{s})$ .

Again by the chain rule  $\vec{p} \rightarrow f(\vec{s}(\vec{p})) = \lambda_{1,-}(\vec{p})$  is Newton-differentiable.

Combining above results, Newton differentiability of  $\vec{p} \rightarrow \lambda_1(\vec{p})$  from  $L^p(\Omega, \mathbb{R}^2)$  to  $L^q(\Omega)$  now follows from (2.14). □

### 3. SEMI-SMOOTH NEWTON ALGORITHM

From (2.3) and Lemma 2.5 the first order optimality condition has the form:

$$(3.1) \quad \begin{aligned} \alpha\Lambda\vec{y}^* &= \mathcal{B}(\mathcal{B}^*\vec{p}^* - M^T\vec{\lambda}^*), \quad \vec{y}^*|_{\partial\Omega} = \vec{0}, \\ \Lambda^*\vec{p}^* &= \vec{y}_d - \vec{y}^*, \quad \vec{p}^*|_{\partial\Omega} = \vec{0}, \\ \vec{\lambda}^* &= H(\mathcal{B}^*\vec{p}^*). \end{aligned}$$

Clearly the triple  $(\vec{y}^*, \vec{p}^*, \vec{\lambda}^*)$  is unique, and hence this condition is also a sufficient condition for Problem 1.1.

We next aim at solving (3.1) by a Newton-type method. For this purpose we introduce  $F : (H^2(\Omega, \mathbb{R}^2) \cap H_0^1(\Omega, \mathbb{R}^2))^2 \times L^2(\tilde{\Omega}, \mathbb{R}^m) \rightarrow (L^2(\Omega, \mathbb{R}^2))^2 \times L^2(\tilde{\Omega}, \mathbb{R}^m)$  by

$$(3.2) \quad F(x) = \begin{pmatrix} \alpha\Lambda\vec{y} - \mathcal{B}(\mathcal{B}^*\vec{p} - M^T\vec{\lambda}) \\ \Lambda^*\vec{p} - \vec{y}_d + \vec{y} \\ \vec{\lambda} - H(\mathcal{B}^*\vec{p}) \end{pmatrix},$$

where  $x = (\vec{y}, \vec{p}, \vec{\lambda})$ . The Newton derivative of the nonlinear operator  $F(x)$  can be expressed as

$$(3.3) \quad G_N F(x) = \begin{pmatrix} \alpha\Lambda & -\mathcal{B}\mathcal{B}^* & \mathcal{B}M^T \\ I & \Lambda^* & 0 \\ 0 & -G_N H(\mathcal{B}^* \vec{p}) \mathcal{B}^* & I \end{pmatrix},$$

where the specific form of the Newton derivative  $G_N H$  will be given below.

Then the semi-smooth Newton algorithm for (3.1) consist in iteratively solving

$$(3.4) \quad G_N F(x^k)(x^{k+1} - x^k) = -F(x^k)$$

for  $x^{k+1}$ . Except possibly for initialization this is equivalent to solving

$$(3.5) \quad \begin{cases} \alpha\Lambda \vec{y}^{k+1} = \mathcal{B}\mathcal{B}^* \vec{p}^{k+1} - \mathcal{B}M^T \vec{\lambda}^{k+1}, & \vec{y}^{k+1}|_{\partial\Omega} = \vec{0}, \\ \Lambda^* \vec{p}^{k+1} = \vec{y}_d - \vec{y}^{k+1}, & \vec{p}^{k+1}|_{\partial\Omega} = \vec{0}, \\ \vec{\lambda}^{k+1} = G_N H(\mathcal{B}^* \vec{p}^k)[\mathcal{B}^*(\vec{p}^{k+1} - \vec{p}^k)] + H(\mathcal{B}^* \vec{p}^k). \end{cases}$$

The complete algorithm is given in Algorithm 1. One can observe that if  $\vec{p}^{k+1} = \vec{p}^k$ , then Newton step (3.5) coincides with the optimality system (3.1), and uniqueness implies that  $\vec{p}^k = \vec{p}^*$ . It is therefore reasonable to choose  $\|\vec{p}^{k+1} - \vec{p}^k\| \leq \epsilon$  for a given small  $\epsilon$  as stopping rule. In practice, we observe that the algorithm often converges in finitely many step with the exact (discretized) solution.

---

**Algorithm 1** Semismooth Newton Algorithm

---

- 1: Set  $k=0$ , initialize  $\vec{p}^0$ .
  - 2: Solve (3.5) for  $(\vec{y}^{k+1}, \vec{p}^{k+1}, \vec{\lambda}^{k+1})$ .
  - 3: Stop or update  $k = k + 1$ , and go to 2.
- 

Denote the error function by  $\vec{z}^k = \vec{y}^k - \vec{y}^*$ ,  $\vec{q}^k = \vec{p}^k - \vec{p}^*$ , and  $\vec{\xi}^k = \vec{\lambda}^k - \vec{\lambda}^*$ , we have

$$(3.6) \quad \begin{cases} \alpha\Lambda \vec{z}^{k+1} = \mathcal{B}\mathcal{B}^* \vec{q}^{k+1} - \mathcal{B}M^T \vec{\xi}^{k+1}, & \vec{z}^{k+1}|_{\partial\Omega} = \vec{0}, \\ \Lambda^* \vec{q}^{k+1} = -\vec{z}^{k+1}, & \vec{q}^{k+1}|_{\partial\Omega} = \vec{0}, \\ \vec{\xi}^{k+1} = G_N H(\mathcal{B}^* \vec{p}^k) \mathcal{B}^* \vec{q}^{k+1} + \vec{r}^k, \end{cases}$$

where  $\vec{r}^k = H(\mathcal{B}^* \vec{p}^k) - H(\mathcal{B}^* \vec{p}^*) - G_N(\mathcal{B}^* \vec{p}^k) \mathcal{B}^* \vec{q}^k$ . By Lemma 2.6, we have

$$\|\vec{r}^k\|_{L^2(\tilde{\Omega})} = o(\|\vec{q}^k\|_{L^q(\Omega)}), \text{ for } q > 2.$$

Substitute  $\vec{\xi}^{k+1}$  into the first error equation, we obtain

$$(3.7) \quad \begin{cases} \alpha\Lambda \vec{z}^{k+1} - \mathcal{B}\mathcal{B}^* \vec{q}^{k+1} + \mathcal{B}M^T G_N H(\mathcal{B}^* \vec{p}^k) \mathcal{B}^* \vec{q}^{k+1} = -\mathcal{B}M^T \vec{r}^k, & \vec{z}^{k+1}|_{\partial\Omega} = \vec{0}, \\ \Lambda^* \vec{q}^{k+1} = -\vec{z}^{k+1}, & \vec{q}^{k+1}|_{\partial\Omega} = \vec{0}. \end{cases}$$

To obtain the well-posedness of Newton iteration and error estimates, we need the following Lemma.



**Lemma 3.1.** *For any given function  $\vec{p} \in L^2(\Omega, \mathbb{R}^2)$ , the operator  $\mathcal{C} = \mathcal{B}\mathcal{B}^* - \mathcal{B}M^T G_N H(\mathcal{B}^* \vec{p}) \mathcal{B}^*$  is nonnegative in the sense that*

$$(\vec{v}, [\mathcal{B}\mathcal{B}^* - \mathcal{B}M^T G_N H(\mathcal{B}^* \vec{p}) \mathcal{B}^*] \vec{v}) \geq 0, \quad \text{for all } \vec{v} \in L^2(\Omega, \mathbb{R}^2).$$

*Proof.* From Lemma 2.6, we can obtain the explicit form of  $G_N H(\vec{q})$  for any vector  $\vec{q}$ . We will show that for any  $\vec{q}$ , the matrix  $G = I - M^T G_N H(\vec{q})$  is symmetric semi-definite. Recalling the notation in (2.8), we consider three characteristic locations for the vector  $\vec{q}$ , which are the inactive one, and one, respectively two active components.

Case 1: If  $\vec{q} \in \mathcal{I}$ ,  $G_N H(\vec{q}) = 0$ , which implies that  $G = I$ .

Case 2: If  $\vec{q} \in \mathcal{A}_1$ , then

$$\vec{\lambda} = (\vec{q} \cdot \vec{n}_1 - \alpha \psi_1, 0, \dots, 0)^T$$

which implies that

$$G_N H(\vec{q}) = \begin{pmatrix} \cos \theta_1 & \sin \theta_1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}, \quad M^T G_N H(\vec{q}) = \begin{pmatrix} \cos^2 \theta_1 & \cos \theta_1 \sin \theta_1 \\ \cos \theta_1 \sin \theta_1 & \sin^2 \theta_1 \end{pmatrix}.$$

Hence the matrix

$$I - M^T G_N H(\vec{q}) = \begin{pmatrix} \sin^2 \theta_1 & -\cos \theta_1 \sin \theta_1 \\ -\cos \theta_1 \sin \theta_1 & \cos^2 \theta_1 \end{pmatrix}$$

is symmetric semi-definite.

Case 3: If  $\vec{q} \in \mathcal{A}_{1,2}$ , then  $\vec{\lambda} = (\lambda_1, \lambda_2, 0, \dots, 0)^T$ , where

$$\lambda_1 = \frac{1}{\sin^2(\theta_1 - \theta_2)} [-\sin(\theta_1 - \theta_2)(q_1 \sin \theta_2 - q_2 \cos \theta_2) - \alpha \psi_1 + \alpha \cos(\theta_1 - \theta_2) \psi_2],$$

$$\lambda_2 = \frac{1}{\sin^2(\theta_1 - \theta_2)} [-\sin(\theta_1 - \theta_2)(q_2 \cos \theta_1 - q_1 \sin \theta_1) - \alpha \psi_2 + \alpha \cos(\theta_1 - \theta_2) \psi_1].$$

This implies that

$$G_N H(\vec{q}) = \frac{1}{\sin(\theta_2 - \theta_1)} \begin{pmatrix} \sin \theta_2 & -\cos \theta_2 \\ -\sin \theta_1 & \cos \theta_1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}, \quad M^T G_N H(\vec{q}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

In this case, the matrix  $I - M^T G_N H(\vec{q})$  is zero and hence symmetric semi-definite.

Using above fact and the definition of  $\mathcal{B}^*$ , for any function  $\vec{v} \in L^2(\Omega, \mathbb{R}^2)$ , we have

$$\begin{aligned} (\vec{v}, [\mathcal{B}\mathcal{B}^* - \mathcal{B}M^T G_N H(\mathcal{B}^* \vec{p}) \mathcal{B}^*] \vec{v}) &= \int_{\Omega} \mathcal{B}^* \vec{p} \cdot (I - M^T G_N H(\mathcal{B}^* \vec{p})) (\mathcal{B}^* \vec{v}) \, dx \\ &= \int_{\tilde{\Omega}} \vec{v} \cdot (I - M^T G_N H(\vec{p})) \vec{v} \, dx \geq 0. \end{aligned}$$

□

**Corollary 3.2.** For any initialization the Newton iterates of (3.5) are well-defined and satisfy  $\vec{x}^k = (\vec{y}^k, \vec{u}^k, \vec{\lambda}^k) \in (H^2(\Omega, \mathbb{R}^2) \cap H_0^1(\Omega))^2 \times L^2(\tilde{\Omega}, \mathbb{R}^m)$ , for  $k = 1, \dots$

*Proof.* Let  $\mathcal{C}$  be the operator defined in Lemma 3.1 and set  $\vec{g} = \mathcal{B}M^T(G_N H(\mathcal{B}^* \vec{p}^k - H(\mathcal{B}^* \vec{p}^k)))$ ,  $(\vec{y}, \vec{p}) = (\vec{y}^{k+1}, \vec{p}^{k+1})$ . Then, inserting the last equation of (3.5) into the first, we find from the second that for any  $\vec{v} \in H_0^1(\Omega, \mathbb{R}^2) \cap H^2(\Omega, \mathbb{R}^2)$

$$(\Lambda^* \vec{p}, \Lambda^* \vec{v}) = (\vec{y}_\alpha - \vec{y}, \Lambda^* \vec{v}) = -\frac{1}{\alpha}(\mathcal{C}\vec{p}, \vec{v}) - \frac{1}{2}(\vec{g}, \vec{v}) + (\vec{y}_d, \Lambda^* \vec{v}),$$

and (3.5) becomes equivalent to solving

$$(3.8) \quad \begin{cases} (\Lambda^* \vec{p}, \Lambda^* \vec{v}) + \frac{1}{\alpha}(\mathcal{C}\vec{p}, \vec{v}) = (\vec{y}_d, \Lambda^* \vec{v}) - \frac{1}{\alpha}(\vec{g}, \vec{v}), \text{ for all } \vec{v} \in H_0^1 \cap H^2 \\ \alpha \Lambda^* \vec{y} = \mathcal{C}\vec{p} + \vec{g}, \end{cases}$$

and setting  $\vec{\lambda} = G_N H(\mathcal{B}^* \vec{p}^k) [\mathcal{B}^* (\vec{p} - \vec{p}^k)] + H(\mathcal{B}^* \vec{p}^k)$ . Using positive semi-definiteness of  $\mathcal{C}$ , the first equation in (3.8) admits a unique solution in  $H_0^1(\Omega, \mathbb{R}^2) \cap H^2(\Omega, \mathbb{R}^2)$  by the Lax-Milgran lemma. The second equation has a unique solution  $\vec{y} \in H_0^1(\Omega, \mathbb{R}^2) \cap H^2(\Omega, \mathbb{R}^2)$  by ellipticity of  $\Lambda$ . This concludes the proof.  $\square$

Finally we address convergence of the semi-smooth Newton method.

**Theorem 3.3.** *Algorithm 1 is locally super-linearly convergent in the sense that*

$$\|\vec{p}^{k+1} - \vec{p}^*\|_{H^2(\Omega)} + \|\vec{y}^{k+1} - \vec{y}^*\|_{H^2(\Omega)} + \|\vec{\lambda}^{k+1} - \vec{\lambda}^*\|_{L^2(\tilde{\Omega})} = o(\|\vec{p}^k - \vec{p}^*\|_{H^2(\Omega)}).$$

*Proof.* We utilize the error equation (3.6). Taking the inner product with  $\vec{q}^{k+1}$  in the first equation of (3.7) results in

$$\begin{aligned} (\vec{q}^{k+1}, [\mathcal{B}\mathcal{B}^* - \mathcal{B}M^T G_N H(\mathcal{B}^* \vec{p}) \mathcal{B}^*] \vec{q}^{k+1}) - (\mathcal{B}M^T \vec{r}^k, \vec{q}^{k+1}) &= \alpha(\Lambda \vec{z}^{k+1}, \vec{q}^{k+1}) \\ &= \alpha(\vec{z}^{k+1}, \Lambda^* \vec{q}^{k+1}) = -\alpha \|\Lambda^* \vec{q}^{k+1}\|^2. \end{aligned}$$

Using that  $\Lambda^* : H_0^1(\Omega, \mathbb{R}^2) \cap H^2(\Omega, \mathbb{R}^2) \rightarrow L^2(\Omega, \mathbb{R}^2)$  is a homeomorphism, Lemma 3.1 and Young's inequality, we obtain

$$\|\vec{q}^{k+1}\|_{H^2(\Omega)} \leq C \|\vec{r}^k\|_{L^2(\tilde{\Omega})} = o(\|\vec{q}^k\|_{H^2(\Omega)}),$$

for a constant  $C$  independent of  $k$ . Since  $\vec{\xi}^{k+1} = G_N H(\mathcal{B}^* \vec{p}^k) \mathcal{B}^* \vec{q}^{k+1} + \vec{r}^k$ , using the pointwise uniform boundedness of  $H(\mathcal{B}^* \vec{p})$ , for  $\vec{p} \in \mathbb{R}^2$ , we find that

$$\|\vec{\xi}^{k+1}\|_{L^2(\tilde{\Omega})} = o(\|\vec{q}^k\|_{H^2(\Omega)}).$$

The estimate for  $\vec{z}^k$  can be obtained from the first equation in (3.6).  $\square$

#### 4. NUMERICAL EXAMPLE

We use a simple example to validate the efficient of semi-smooth Newton algorithm in the case of polygonal constraints. The underlying control equations is an elliptic system with homogenous Dirichlet boundary condition in the unit square, and control is on the whole domain, i.e., the governing equations are as follows:

$$-\Delta y_1 + y_2 = u_1, \quad -\Delta y_2 - y_1 = u_2, \quad \text{in } \Omega \quad \vec{y}|_{\partial\Omega} = \vec{0},$$

where  $\Omega = (0, 1)^2$ . It can be verified this system is strongly elliptic which satisfies inequality (1.3). Let the polygonal domain domain  $K$  be given by

$$u_1 \geq 0, \quad u_2 \geq 0, \quad u_1 + u_2 \leq 1.$$

Using the notations introduced in Section 2 we have

$$\theta_1 = \frac{3}{2}\pi, \quad \theta_2 = \pi, \quad \theta_3 = \frac{1}{4}\pi,$$

and the constraints can be put into matrix form:  $M\vec{u} \leq \vec{\psi}$ , where

$$M = \begin{pmatrix} 0 & -1 \\ -1 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}, \quad \psi = \begin{pmatrix} 0 \\ 0 \\ \frac{\sqrt{2}}{2} \end{pmatrix}.$$

The space  $\mathbb{R}^2$  of  $\vec{p}$  can be decomposed into 7 subdomains:  $\mathcal{I}$ ,  $\mathcal{A}_1$ ,  $\mathcal{A}_2$ ,  $\mathcal{A}_3$ ,  $\mathcal{A}_{1,2}$ ,

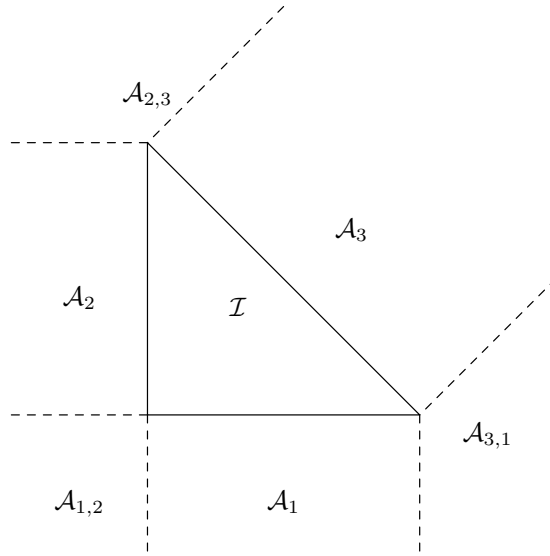


FIGURE 4.7. The decomposition of  $\vec{p}$  plane

$\mathcal{A}_{2,3}$ ,  $\mathcal{A}_{3,1}$ , as shown in Figure 4.7. The Newton derivative  $G(\vec{p}) = G_N H(\vec{p})$  is given by:

- (1)  $G(\vec{p}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$  for  $\vec{p} \in \mathcal{I}$ .
- (2)  $G(\vec{p}) = \begin{pmatrix} 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$  for  $\vec{p} \in \mathcal{A}_1$ .
- (3)  $G(\vec{p}) = \begin{pmatrix} 0 & 0 \\ -1 & 0 \\ 0 & 0 \end{pmatrix}$  for  $\vec{p} \in \mathcal{A}_2$ .
- (4)  $G(\vec{p}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$  for  $\vec{p} \in \mathcal{A}_3$ .
- (5)  $G(\vec{p}) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \\ 0 & 0 \end{pmatrix}$  for  $\vec{p} \in \mathcal{A}_{1,2}$ .

TABLE 4.1. super-linear convergence

iter. number	1	2	3	4	5
$\ y^k - y^{k-1}\ _{H^2(\Omega)}$	1.5415	0.2727	0.0091	3.1548e-005	0
$\frac{\ y^k - y^{k-1}\ _{H^2(\Omega)}}{\ y^{k-1} - y^{k-2}\ _{H^2(\Omega)}}$		0.1769	0.033	0.00347	0
$\ p^k - p^{k-1}\ _{H^2(\Omega)}$	0.0052	5.9900e-004	2.0534e-005	1.7818e-007	0
$\frac{\ p^k - p^{k-1}\ _{H^2(\Omega)}}{\ p^{k-1} - p^{k-2}\ _{H^2(\Omega)}}$		0.1152	0.0342	0.008677	0

$$(6) \quad G(\vec{p}) = \sqrt{2} \begin{pmatrix} 0 & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 0 & 1 \end{pmatrix} \quad \text{for } \vec{p} \in \mathcal{A}_{2,3}.$$

$$(7) \quad G(\vec{p}) = \sqrt{2} \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ 0 & 0 \\ 1 & 0 \end{pmatrix} \quad \text{for } \vec{p} \in \mathcal{A}_{3,1}.$$

Recall the cost functional is given by:

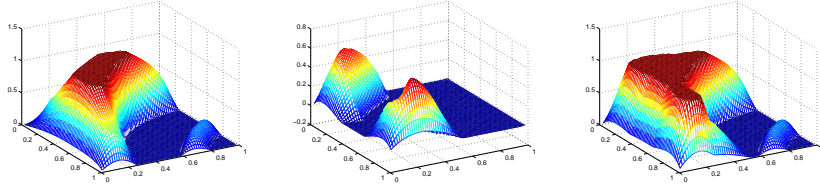
$$J(\vec{y}, \vec{u}) = \frac{1}{2} \|\vec{y} - \vec{y}_d\|^2 + \frac{\alpha}{2} \|\vec{u}\|^2.$$

Here we choose  $\vec{y}_d = (0.1 \sin(4\pi xy), 0.05(\sin(2\pi x) + \cos(2\pi y)))^T$ ,  $\alpha = 0.001$ . The choice of  $\vec{y}_d$  guarantees that all three constraints are active on some part of the domain. Superlinear and finite step convergence can be observed numerically, see Table 4.1. It should be noted the initial choice is zero function which is quite far from final solution. Since we do not know the exact solution, the error function is defined by the difference of two successive Newton solutions, i.e.,

$$e_y^k = \|\vec{y}^{k+1} - \vec{y}^k\|_{H^2(\Omega)}, \quad e_p^k = \|\vec{p}^{k+1} - \vec{p}^k\|_{H^2(\Omega)}.$$

The control variables are computed from  $\vec{u} = \frac{1}{\alpha}(\vec{p} - M^T \vec{\lambda})$  and are depicted in Figure 4.8. It can be noted that three constraints  $u_1 \geq 0$ ,  $u_2 \geq 0$  and  $u_1 + u_2 \leq 1$  are active in some part of the domain.

FIGURE 4.8. Plot for the control variables. From left to right:  $u_1$ ,  $u_2$ ,  $u_1 + u_2$ .



This numerical realisation is based on a finite difference discretization with respect to a uniform axis-parallel grid. The mesh size for the result reported in Table 4.1 is  $h = \frac{1}{64}$ . To demonstrate mesh-independence of the algorithm, we compute the same example on a series grids for mesh sizes refined by a factor of  $\frac{1}{2}$ . The number

TABLE 4.2. Number of Iteration

grid number	8	16	32	64	128
iter. number	5	5	5	5	5

of iterations before the exact finite dimensional solution is found to be fixed, see Table 4.2.

## REFERENCES

- [1] R. Adams and J. Fournier, *Sobolev Spaces*, Academic Press, 2003.
- [2] V. Barbu and T. Precupanu, *Convexity and Optimization in Banach Spaces*, Springer, 1986.
- [3] X. Chen, Z. Nashed and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal. 38(2000), 1200-1216.
- [4] J. C. de los Reyes and K. Kunisch, *A semi-smooth Newton method for control constrained optimal control of the Navier Stokes equations*, Nonlinear Analysis, 62(2005), 1289-1316.
- [5] J. C. De Los Reyes and K. Kunisch, *Optimal control of partial differential equations with affine control constraints*, Control and Cybernetics, 38 (2009), 1217-1250.
- [6] M. Hintermüller and M. Hinze, *A SQP-semismooth Newton-type algorithm applied to control of the instationary Navier-Stokes system subject to control constraints*, SIAM J. Optim. 16 (2006), 1177-1200.
- [7] M. Hintermüller, K. Ito and K. Kunisch, *The primal-dual active set strategy as a semi-Smooth Newton method*, SIAM J. Optim. 13 (2003), 865-888.
- [8] M. Hinze, *A variational discretization concept in control constrained optimization: the linear-quadratic case*, Comput. Optim. Appl. 30 (2005), 45-61.
- [9] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM Philadelphia, 2008.
- [10] K. Ito and K. Kunisch, *The primal-dual active set method for nonlinear optimal control problems with bilateral constraints*, SIAM J. Control Optim. 43 (2004), 357-376.
- [11] K. Kunisch and X. Lu, *Optimal control for an elliptic system with pointwise nonlinear control constraints*, submitted.
- [12] D.G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley Publ. Comp., Menlo Park, 1984
- [13] K. Malanowski, *Convergence of approximations vs. regularity of solutions for convex, control constrained optimal control problems*, Appl. Math. Optim., 8 (1981), 69-95.
- [14] C. Meyer and A. Rösch, *Superconvergence properties of optimal control problems*, SIAM J. Control Optim., to appear.
- [15] A. Schiela and M. Weiser, *Superlinear convergence of the control reduced interior point method for PDE constrained optimization*, Comp. Opt. and Appl. 39 (2008) 369-393.
- [16] M. Ulbrich, *Semismooth Newton methods for operator equations in function spaces*, SIAM J. Optim., 13 (2003), 805-842.
- [17] B. Vexler and W. Wollner, *Adaptive finite elements for elliptic optimization problems with control constraints*, SIAM J. Control Optim. 47 (2008), 509-534.
- [18] D. Wachsmuth, *Sufficient second-order optimality conditions for convex control constraints*, J. Math. Anal. Appl. 319 (2006), 228-247.
- [19] M. Weiser, T. Gänzler and A. Schiela, *A Control reduced primal interior point method for PDE constrained optimization*, Comp. Opt. and Appl. 41 (2008), 127-145.

INSTITUTE OF MATHEMATICS AND SCIENTIFIC COMPUTING, UNIVERSITY OF GRAZ, HEINRICH-STRASSE 36 A-8010 GRAZ, AUSTRIA

E-mail address: karl.kunisch@uni-graz.at

SCHOOL OF MATHEMATICS AND STATISTICS, WUHAN UNIVERSITY, WUHAN 430072, P.R. CHINA

E-mail address: xllv.math@whu.edu.cn