A SEMISMOOTH NEWTON METHOD FOR GENERALIZED SEMI-INFINITE PROGRAMMING PROBLEMS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

AYSUN TEZEL ÖZTURAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
MATHEMATICS

JUNE 2010

Approval of the thesis:

A SEMISMOOTH NEWTON METHOD FOR

GENERALIZED SEMI-INFINITE PROGRAMMING PROBLEMS

submitted by **AYSUN TEZEL ÖZTURAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Mathematics Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Zafer Nurlu
Head of Department, **Mathematics** _____

Prof. Dr. Bülent Karasözen
Supervisor, **Department of Mathematics, METU** _____

Prof. Dr. Oliver Stein
Co-supervisor, **Institute of Operations Research, Karlsruhe Institute of Technology, Germany** _____

Examining Committee Members:

Prof. Dr. Gerhard Wilhelm Weber
Scientific Computing, IAM, METU _____

Prof. Dr. Bülent Karasözen
Department of Mathematics, METU _____

Prof. Dr. Tanıl Ergenç
Department of Mathematics, Atılım University _____

Assoc. Prof. Ömür Uğur
Scientific Computing, IAM, METU _____

Assoc. Prof. Emre Alper Yıldırım
Department of Industrial Engineering, Bilkent University _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: AYSUN TEZEL ÖZTURAN

Signature :

ABSTRACT

A SEMISMOOTH NEWTON METHOD FOR
GENERALIZED SEMI-INFINITE PROGRAMMING PROBLEMS

Tezel Özturan, Aysun
Ph.D., Department of Mathematics
Supervisor : Prof. Dr. Bülent Karasözen
Co-Supervisor : Prof. Dr. Oliver Stein

June 2010, 119 pages

Semi-infinite programming problems is a class of optimization problems in finite dimensional variables which are subject to infinitely many inequality constraints. If the infinite index of inequality constraints depends on the decision variable, then the problem is called generalized semi-infinite programming problem (*GSIP*). If the infinite index set is fixed, then the problem is called standard semi-infinite programming problem (*SIP*).

In this thesis, convergence of a semismooth Newton method for generalized semi-infinite programming problems with convex lower level problems is investigated. In this method, using nonlinear complementarity problem functions the upper and lower level Karush-Kuhn-Tucker conditions of the optimization problem are reformulated as a semismooth system of equations. A possible violation of strict complementary slackness causes nonsmoothness. In this study, we show that the standard regularity condition for convergence of the semismooth Newton method is satisfied under natural assumptions for semi-infinite programs. In fact, under the Reduction Ansatz in the lower level problem and strong stability in the reduced upper level problem this regularity condition is satisfied. In particular, we do not have to assume strict

complementary slackness in the upper level. Furthermore, in this thesis we neither assume strict complementary slackness in the upper nor in the lower level. In the case of violation of strict complementary slackness in the lower level, the auxiliary functions of the locally reduced problem are not necessarily twice continuously differentiable. But still, we can show that a standard regularity condition for quadratic convergence of the semismooth Newton method holds under a natural assumption for semi-infinite programs. Numerical examples from, among others, design centering and robust optimization illustrate the performance of the method.

Keywords: Generalized semi-infinite optimization, semismooth Newton method, Nonlinear Complementarity function, Clarke subdifferential regularity, Reduction Ansatz.

ÖZ

GENELLEŞTİRİLMİŞ YARI SONSUZ OPTİMİZASYON
PROBLEMLERİ İÇİN YARI DÜZGÜN NEWTON YÖNTEMİ

Tezel Özturan, Aysun
Doktora, Matematik Bölümü
Tez Yöneticisi : Prof. Dr. Bülent Karasözen
Ortak Tez Yöneticisi : Prof. Dr. Oliver Stein

Haziran 2010, 119 sayfa

Yarı sonsuz programlama problemleri, sonlu değişken üzerinde sonsuz eşitsizlik kısıtlamaları olan optimizasyon problemleridir. Eşitsizlik kısıtlamalarının sonsuz indeks kümesi optimizasyonun yapıldığı değişkene bağlı ise problem genelleştirilmiş yarı sonsuz optimizasyon problemi olarak adlandırılır. Sonsuz indeks kümesi sabit bir küme ise standart yarı sonsuz optimizasyon problemi olarak adlandırılır.

Bu tezde, genelleştirilmiş yarı sonsuz optimizasyon problemlerinin konveks alt seviye problemi olanlar için yarı düzgün Newton yönteminin yakınsaklığı incelenmiştir. Bu yöntemde, lineer olmayan tamlık fonksiyonları kullanılarak optimizasyon probleminin üst ve alt seviye Karush Kuhn Tucker koşulları yarı düzgün eşitliklere dönüştürülür. Kati tamamlayıcı gevşeklik koşullarının olası ihlali düzgün olmamaya neden olur. Bu çalışmada, yarı düzgün Newton yönteminin yakınsaklığı için gerekli olan standart düzenlilik koşulunun yarı sonsuz programlamanın doğal varsayımları altında sağlandığını gösterdik. Aslında, bu düzenlilik koşulu alt seviye problemi için indirgeme yaklaşımı ve indirgenmiş üst seviye problemi için kuvvetli kararlılık koşulları altında sağlanır. Özellikle, üst seviye problemde kati tamamlayıcı

gevşeklik koşulunu varsaymak zorunda değiliz. Bu tezde, ayrıca ne üst seviyede ne de alt seviyede kati tamamlayıcı gevşeklik koşullarını varsaymadık. Alt seviyede kati tamamlayıcı gevşeklik koşulunun ihlali durumunda, yerel indirgenmiş problemin geçici fonksiyonları iki kere sürekli olarak türevlenebilir olmaz. Ama halen yarı düzgün Newton yönteminin ikinci dereceden yakınsak olması için gerekli olan standart düzenlilik koşullarının yarı sonsuz programlamanın doğal varsayımları altında gerçekleştiğini gösterebiliriz. Tasarım merkezlemenin ve güvenli en iyileme problemlerinin aralarında olduğu sayısal örnekler metodun performansını göstermektedir.

Anahtar Kelimeler: Genelleştirilmiş yarı sonsuz optimizasyon problemleri, yarı düzgün Newton metodu, lineer olmayan tamlık problemi fonksiyonları, Clarke ın genelleştirilmiş türevinin düzenliliği, İndirgeme Yaklaşımı.

To my loving mother Fatma Tezel,
whose endless love and support
have been one of my greatest
sources of strength in my life.

ACKNOWLEDGMENTS

I am very indebted to the following people who gave me the possibility to complete this thesis.

First of all, I would like to express my sincere gratitude to my supervisor Prof. Dr. Bülent Karasözen and deepest gratitude to my co-supervisor Prof. Dr. Oliver Stein for his precious support and guideness through the thesis. This thesis wouldn't have been possible without the inspiration of Prof. Dr. Oliver Stein and his encouragement. I would like to special thanks to him for his support scientifically through my visits at RWTH Aachen University and University of Karlsruhe (TH), Germany. I am grateful for his patience and continuing guidance after my visits. His scientific vision and excellent support are of considerable value to me. Also, I would like to thank Prof. Dr. Georg Still for my visits at University of Twente for his host and support scientifically. I deeply thank to Prof. Dr. Gerhard Wilhelm Weber for his valuable discussion and precious cooperation.

I would like to extend my thanks to the members of the Department of Mathematics and my friends for their suggestions during the period of writing the thesis.

My admirable parent, Fatma Tezel deserve the greatest appreciation and respect for the unconditional love, support and trust she has provided in every stage of my life. I am grateful that I have such a great mother in my life. Her support, encouragement, and constant love have sustained me throughout my life.

Last, but not the least, I would like to give my heartfelt thanks to my husband, Ahmet Özturan, for his presence, understanding, big love, patience and support to me. I would like to express my special thanks to our families for their endless supports.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

This dissertation presents a new numerical approach to solve generalized semi-infinite programming problems. Semi-infinite programming is a subfield of continuous programming which deals with finding extremum of a continuous function over a finite dimensional space subject to infinitely many continuous inequality constraints. Furthermore, if the infinite inequality constraints depend on the finite decision variable, the problem is called generalized semi-infinite programming, otherwise the problem is called standard semi-infinite programming. The purpose of this chapter is to give the outline of the dissertation. We introduce the problem addressed in this dissertation and provide a concise description of the approach introduced in this work. In the following, we point out the main contributions and define the outline of the dissertation.

A generalized semi-infinite optimization problem has the form,

$$GSIP: \qquad \text{minimize} \quad f(x) \quad \text{subject to} \quad x \in M$$

with the feasible set

$$M \;=\; \{\, x \in \mathbb{R}^n \,|\, g(x,y) \leq 0 \text{ for all } y \in Y(x) \,\}$$

and

$$Y(x) \;=\; \{\, y \in \mathbb{R}^m \,|\, v_j(x,y) \leq 0 \;\; (j \in Q) \,\}.$$

where $j \in Q = \{1, ..., q\}$.

In a *GSIP*, the possibly infinite index set $Y(x)$ of the semi-infinite inequality constraint is allowed to vary with x. As opposed to this, in a standard *SIP* the index set is fixed, that is, we have $Y(x) \equiv Y$, and if Y is described by functional constraints, then the vector function v does

not depend on x. If Y is a finite set, we arrive at the usual nonlinear programming problem. Furthermore, we assume that the set valued mapping $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is locally bounded, that is, for each $\bar{x} \in \mathbb{R}^n$ there exists a neighborhood U of $\bar{x}$ such that $\bigcup_{x \in U} Y(x)$ is bounded in $\mathbb{R}^m$ and that $Y(x) \neq \emptyset$ for all $x \in \mathbb{R}^n$.

In particular, the problem *GSIP* may have finitely many inequality constraints $g_i(x, y) \leq 0, y \in Y_i(x), i \in I$ which often arises in applications, along with finitely many equality constraints. In order to abstain from technical difficulties in this thesis we examine the case of a single semi-infinite constraint. The interested reader can be referred to [103] for more general formulations. However, the slightly more general setting of finitely many generalized semi-infinite constraints $g_i(x, y) \leq 0, y \in Y_i(x), i \in I$ would lead to almost identical formulas as the ones we develop in the sequel (with g replaced by g_i).

Semi-infinite programming has been studied and developed by researchers over the last thirty years. There are many practical applications of semi-infinite programming problems. Chebyshev and reverse Chebyshev approximation, time minimal control, minimax problems, robust optimization, design centering, optimal layout of an assembly line and disjunctive programming are some applications of semi-infinite programming.

Since the problem formulation for *GSIP* is a generalization of *SIP*, in the first studies of *GSIP* [34, 59], *GSIP* was thought to be a slight generalization of standard *SIP*. After studies, it was seen that most of the known theoretical and computational approaches to *SIP* do not generalize to *GSIP*. In [47], it is firstly recognized that *GSIP* is actually harder than *SIP*. The feasible set of *GSIP* may have topological properties that are neither known from standard semi-infinite nor from finitely constrained programing problem. In finitely constrained programming and in standard semi-infinite programming problems the feasible set is closed. This is not true for *GSIP*. In [20], it is pointed out that the feasible set of *GSIP* is not definitely a closed set and also may have a disjunctive property. These two topological properties are not known from standard *SIP*.

Example 1.0.1 ([20]) *"Re-entrant corner point"*

"For $x \in \mathbb{R}^2$ consider the index set

$$Y(x) = \{ y \in \mathbb{R} | \, y \geq x_1, \, y \geq x_2 \}$$

and put $g(x,y) = -y$. *Then we obtain*

$$\begin{aligned} M &= \{\, x \in \mathbb{R}^2 |\, g(x,y) \leq 0 \ \textit{for all}\ y \in Y(x) \,\} \\ &= \{\, x \in \mathbb{R}^2 |\, y \geq 0 \ \textit{for all}\ y \in [max(x_1, x_2), +\infty) \,\} \\ &= \{\, x \in \mathbb{R}^2 |\ max(x_1, x_2) \geq 0 \,\}. \end{aligned}$$

Figure 1.1 illustrates that M is the union of two closed halfplanes. Note that M is nonconvex, although all defining functions are linear. More precisely, M exhibits a so-called re-entrant corner point at the origin. These points are spurious points for stationary based optimality conditions."

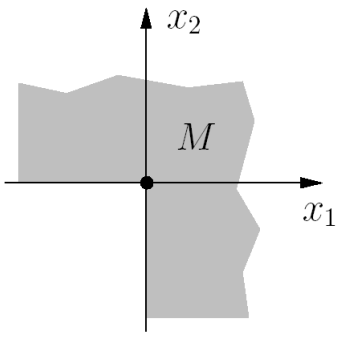

Figure 1.1: A re-entrant corner point [20].

Example 1.0.2 ([20]) *"Local nonclosedness"*

"For $x \in \mathbb{R}^2$ *consider the index set*

$$Y(x) = \{\, y \in \mathbb{R} |\, y \geq x_1,\ y \leq x_2 \}$$

and put again $g(x,y) = -y$. *Now we obtain*

$$\begin{aligned} M &= \{\, x \in \mathbb{R}^2 |\, g(x,y) \leq 0 \ \textit{for all}\ y \in Y(x) \,\} \\ &= \{\, x \in \mathbb{R}^2 |\, y \geq 0 \ \textit{for all}\ y \in [x_1, x_2] \,\} \\ &= \{\, x \in \mathbb{R}^2 |\, x_1 \leq x_2, y \geq 0 \ \textit{for all}\, y \in [x_1, x_2] \,\} \\ &\quad \cup \{\, x \in \mathbb{R}^2 |\, x_1 > x_2, y \geq 0 \, \textit{for all}\, y \in \emptyset \,\} \\ &= \{\, x \in \mathbb{R}^2 |\, x_1 \leq x_2, x_1 \geq 0 \} \cup \{\, x \in \mathbb{R}^2 |\, x_1 > x_2 \,\}. \end{aligned}$$

Figure 1.2 illustrates M which is the union of an open with a closed halfplane although all defining inequalities are nonstrict. We remark that $Y(x) = \emptyset$ for $x_1 > x_2$."

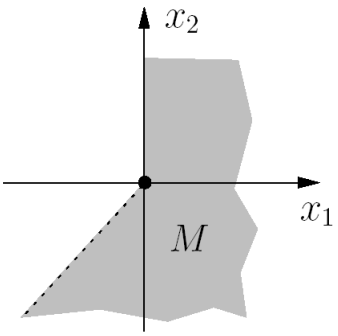

Figure 1.2: Local nonclosedness [20].

In finitely constrained and semi-infinite programming problems the local nonclosedness as in Example 1.0.2 can not occur. The re-entrant corner points as in Example 1.0.1 may occur in finitely constrained programming. The re-entrant corner points are not stable in finitely constrained programming but they are stable in *GSIP* [20]. The local nonclosedness is also stable in *GSIP*. The investigation of the semi-infinite programming shows that *GSIP* is actually harder than *SIP*. The numerical methods usually can not directly be generalized from standard *SIP* to *GSIP*. There arise serious difficulties when trying to generalize the exchange or discretization methods from standard *SIP* to *GSIP*. These difficulties are discussed and convergence results are obtained under quite general assumptions on *GSIP* in [112]. In this thesis we justified a numerical approach for *GSIP*. Our method is based on the first order necessary optimality conditions. In the following we explain the method and main contribution of the thesis.

It was known that Reduction Ansatz hold at all local minimizers of standard *SIP*. In [23], it is recently shown that generically the Reduction Ansatz holds at all local minimizers of *GSIP*. Hence, the feasible set is locally equal to the feasible set which is described by finitely many certain implicitly defined constraints. Then the optimality conditions can be obtained by using this reduced finitely constrained programming problem. A possible solution method is based on optimality conditions of the locally reduced problem. In fact, many solution methods for nonlinear programming problems are based on solving their Karush-Kuhn-Tucker (KKT) system, that is, a necessary first order optimality condition. It is well-known that the com-

plementarity conditions in the KKT system need special attention in any numerical approach. One possibility for their treatment is a reformulation by nonlinear complementarity problem functions (NCP functions), which reduces the problem to the solution of a certain system of equations which is either nonsmooth or smooth but degenerate. For special NCP functions these equations can be solved by so-called semismooth Newton methods where, in analogy to the standard Newton method, their convergence depends on a regularity condition in the solution point. It is important to note that the nonsmoothness of the system of equations stems from a possible lack of strict complementary slackness at a solution. Such KKT methods have also been suggested for standard semi-infinite programming problems, where the KKT conditions take a somewhat more complicated form [33]. In particular, they involve an upper and a lower level problem. In the article [88] it was recently suggested to use NCP functions also for a nonsmooth reformulation of the KKT conditions in standard semi-infinite programming, and a regularity condition to guarantee convergence of a certain semismooth Newton method was proposed.

It turns out, however, that strict complementarity is a part of the regularity condition from [88], in the upper as well as in the lower level problem. A numerical method which searches a point with these regularity conditions would not need to use NCP functions but, in fact, already the standard Newton method would converge under these assumptions.

One of the aims of this thesis is to point out an important pitfall in the solution of KKT systems for semi-infinite programs. A challenging problem which has not been solved yet in numerical solution of semi-infinite programming (*SIP* or *GSIP*) is finding the corresponding active index set for a feasible point. The active index set can be computed by finding all global maximizers of the lower level problem, hence we assume that the lower level problem is convex. We note that generalized semi-infinite optimization has many relevant applications with convex lower level problems. We present a regularity condition which does not assume strict complementarity in the upper level problem, thus justifying the NCP function approach for semi-infinite programs, and at the same time we transfer this approach from standard to generalized semi-infinite programming. We note that this method will merely search for KKT points of the optimization problem, whereas global optimality plays a crucial role in the treatment of the so-called lower level problem. We also complete this analysis by considering the case of strict complementarity violation in the lower level. For the convergence of a semismooth Newton method, we give an appropriate new regularity condition, thus justifying

the NCP approach for semi-infinite programs in the absence of strict complementarity. In the present case, the convergence analysis is essentially more complicated due to the lack of differentiability of the auxiliary functions of the so-called reduced problem.

In summary, this thesis contains the results of our recent research papers

Tezel, A. and Stein, O. *The semismooth approach for semi-infinite programming without strict complementarity*, SIAM Journal on Optimization, **20(2)**, 1052-1072, (2009).

Stein, O. and Tezel, A. *The semismooth approach for semi-infinite programming under the Reduction Ansatz*, Journal of Global Optimization, **41(2)**, 245-266, 2008.

In this thesis,

- the result about the semismooth approach in [88] for standard semi-infinite programming is completed,
- the semismooth approach from standard to generalized semi-infinite programming problems is extended,
- a new regularity condition which does not assume strict complementarity in the upper and lower level problems is presented,
- the convergence of a semismooth Newton method for generalized semi-infinite programming problems is proved under the natural assumptions of semi-infinite programming, and
- in order to illustrate the performance of the method numerical examples are given.

The thesis is organized as follows. Chapter 2 reviews basic facts from finite and semi-infinite programming. In Section 2.1, we briefly introduce and give some notation of unconstrained and constrained optimization. Section 2.2 deals with the introduction of semi-infinite optimization. Examples and applications, a main and well-known regularity condition in semi-infinite optimization, namely Reduction Ansatz, are given. Furthermore constraint qualifications in semi-infinite programming are mentioned. In Section 2.3, a very brief review of numerical methods in solving semi-infinite optimization problems is presented.

In Chapter 3, after giving preliminaries about Newton method, application of the method in

finite and semi-infinite programming are reviewed. Chapter 4 treats the semismooth Newton approach. The use of the method in finitely constrained programming problems is reviewed. The semismooth optimality conditions for generalized semi-infinite programming are obtained.

Chapter 5 includes our main result, convergence of semismooth Newton method for generalized semi-infinite programming is established. As we have mentioned, in this thesis we study convergence of a semismooth Newton method for generalized semi-infinite programming problems with convex lower level problems. Nonlinear complementarity problem functions (NCP) are used in order to reformulate the upper and lower level Karush-Kuhn-Tucker conditions as a semismooth system of equations. We show that the standard regularity condition for convergence of the semismooth Newton method is satisfied under natural assumptions for semi-infinite programs. We complete the result in [88] by showing convergence under the case of strict complementarity violation in the upper level, and transfer the method to generalized semi-infinite programming. We neither assume strict complementary slackness in the upper nor in the lower level in our semismooth Newton approach. In this case, the auxiliary functions of the locally reduced problem are not absolutely in C^2. But still, it is possible to show that the semi-smooth Newton method converges q-quadratically under the natural assumptions of semi-infinite programming.

In Chapter 6, computational results of the method are reported. Numerical examples from, among others, design centering and robust optimization illustrate the performance of the method. Finally we formulate some conclusions and give possible directions of future research. In Appendices, some auxiliary results about nonlinear complementarity (NCP) functions and block matrices are given.

CHAPTER 2

BACKGROUND ON SEMI-INFINITE PROGRAMMING

2.1 Finitely Constrained Programming

In this section we recall some basic facts from finite optimization problem (finitely constrained programming). We also introduce constraint qualifications in finitely constrained programming. The definitions and theorems in this section can be found in many references about finitely constrained programming, in particular we refer the reader to [2, 13, 15, 27, 44, 45, 74, 77] and references therein for this section. An unconstrained optimization problem have the form

$$P : \min_{x \in \mathbb{R}^n} f(x)$$

where f is at least twice continuously differentiable. In the sequel the row vector $\left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$ of partial derivatives of f evaluated at x will be denoted by $Df(x)$. The gradient of f is a column vector denoted by $\nabla f(x) = (Df(x))^T$. The Hessian matrix of f evaluated at x will be denoted by $D^2 f(x)$. The equation

$$Df(\bar{x}) = 0 \tag{2.1}$$

is a standard necessary first order optimality condition [45]. The points satisfying (2.1) are called *critical points* or *stationary points*. The minimizers and saddle points also satisfy equation (2.1). In order to define a sufficient condition for optimality one needs to combine (2.1) with a second order condition satisfied at $\bar{x}$, so-called second order optimality condition, the Hessian matrix is a positive definite matrix,

$$d^T D^2 f(\bar{x}) d > 0 \ \ \forall d \in \mathbb{R}^n \setminus \{0\} \tag{2.2}$$

Theorem 2.1.1 ([2]) *If $\bar{x}$ satisfies (2.1) and (2.2), then $\bar{x}$ is a strict local minimum of f.*

In the special case that f is convex (2.2) is not need to be explicitly stated.

Definition 2.1.2 ([77]) *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called convex if*

$$f(\nu x + (1-\nu)y) \leq \nu f(x) + (1-\nu) f(y)$$

for all $x, y \in \mathbb{R}^n$ and for all $\nu \in [0,1]$.

The following result shows that stationary, local optimality and global optimality are equivalent in the convex case.

Theorem 2.1.3 ([45]) *Suppose that f is a convex function and f is an at least twice continuously differentiable function. Then the following statements are equivalent.*

i) $\bar{x}$ is a global minimum.

ii) $\bar{x}$ is a local minimum.

iii) $\bar{x}$ is a critical point, i.e., $\bar{x}$ satisfies (2.1).

Stepeest descent methods, Newton method, Quasi-Newton methods etc. are some basic methods to solve unconstrained optimization problems. For more information on these and other methods we refer to [74, 77] and the references cited therein.

A finitely constrained optimization problem has the form

$$P : \min_x f(x) \quad \text{subject to} \quad g_i(x) \leq 0 \quad (i \in I), \quad h_j(x) = 0 \quad (j \in J) \tag{2.3}$$

with $I = \{1, \ldots, r\}$, $J = \{1, \ldots, s\}$. For problem P, the feasible set M_F is defined by

$$M_F = \{x \in \mathbb{R}^n | \; g_i(x) \leq 0 \; (i \in I), \; h_j(x) = 0 \; (j \in J)\}.$$

We assume that f, g_i, h_j are at least twice continuously differentiable.

There are many methods for solving constrained optimization problems which include sequential linear programming, sequential quadratic programming approach (SQP), penalty-, barrier-, interior point-, multiplier methods. It is far beyond the scope of the thesis to give details of these methods, we refer to [45, 74, 77] and the references cited therein for further reading.

As in the case of unconstrained optimization, there exist stationary conditions for constrained problems. The most popular ones are *the Karush-Kuhn-Tucker optimality conditions* and *the Fritz-John optimality conditions*.

Definition 2.1.4 ([2]) *A point $\bar{x} \in M_F$ is said to satisfy Fritz-John optimality conditions if there exist real $\kappa \in \mathbb{R}$, $\mu \in \mathbb{R}^r$, $\lambda \in \mathbb{R}^s$ all nonzero such that*

$$\begin{aligned} \kappa \nabla f(\bar{x}) + \sum_{i=1}^{r} \mu_i \nabla g_i(\bar{x}) + \sum_{j=1}^{s} \lambda_j \nabla h_j(\bar{x}) &= 0, \\ \kappa &\geq 0, \\ \mu_i g_i(\bar{x}) &= 0 \ (i \in I), \\ \mu_i &\geq 0 \ (i \in I). \end{aligned}$$

holds.

By requiring $\kappa \neq 0$ in Fritz-John optimality condition, Karush-Kuhn-Tucker optimality conditions (KKT) are obtained.

Definition 2.1.5 ([77]) *A point $\bar{x} \in M_F$ is said to satisfy Karush-Kuhn-Tucker optimality conditions if there exist real $\mu \in \mathbb{R}^r$, $\lambda \in \mathbb{R}^s$ such that*

$$\begin{aligned} \nabla f(\bar{x}) + \sum_{i=1}^{r} \mu_i \nabla g_i(\bar{x}) + \sum_{j=1}^{s} \lambda_j \nabla h_j(\bar{x}) &= 0, \\ \mu_i g_i(\bar{x}) &= 0 \ (i \in I), \\ \mu_i &\geq 0 \ (i \in I). \end{aligned} \tag{2.4}$$

holds.

The multiplicative condition in (2.4)

$$\mu_i g_i(\bar{x}) = 0 \ (i \in I), \tag{2.5}$$

is known as *complementarity conditions*, it implies the Lagrange multiplier μ_i can be strictly positive only when the corresponding constraint g_i is active. Conditions of this type play a central role in constrained optimization, a reformulation of complementarity condition is given in Section 4.2. By defining

$$\mathcal{L}(x, \mu, \lambda) = f(x) + \mu^T g(x) + \lambda^T h(x), \tag{2.6}$$

the KKT condition (2.4) can be written as

$$\begin{aligned} \nabla_x \mathcal{L}(\overline{x}, \mu, \lambda) &= 0, \\ \mu_i g_i(\overline{x}) &= 0 \ (i \in I), \\ \mu_i &\geq 0 \ (i \in I). \end{aligned}$$

The function $\mathcal{L}$ is called as *the Lagrangian function* and the vectors μ, λ are called as *the Lagrange multiplier vectors*. Let us consider a finitely constrained optimization problem without inequality constraints, i.e.,

$$P : \min_x f(x) \ \text{ subject to } \ g_i(x) \leq 0 \ (i \in I).$$

The complementarity condition gives the following result. If $g_i(\overline{x}) < 0$, (2.5) requires that $\mu_i = 0$, hence $\nabla_x \mathcal{L}(\overline{x}, \mu) = 0$ reduces to $\nabla f(\overline{x}) = 0$ (unconstrained case), and if $g_i(\overline{x}) = 0$, i.e., the inequality constraint is active at $\overline{x}$, (2.5) allows μ_i to take on a nonnegative value, so $\nabla_x \mathcal{L}(\overline{x}, \mu) = 0$ becomes equivalent to $\nabla f(\overline{x}) + \sum_{i=1}^{r} \mu_i \nabla g_i(\overline{x}) = 0$, for some $\mu_i \geq 0$.

The KKT conditions are known to be sufficient for global optimality in convex problems.

Definition 2.1.6 ([2]) *A problem P (2.3) is called convex if f and $g_i, i \in I$ are convex and $h_j, j \in J$ are affine, i.e., $h_j(x) = b_j^T x + c_j$ for some $b_j \in \mathbb{R}^n, c_j \in \mathbb{R}$.*

Theorem 2.1.7 ([2]) *Let P given in (2.3) be a convex problem and $\overline{x}$ be a KKT point of P. Then $\overline{x}$ is a global minimum of P.*

The sufficiency of KKT optimality conditions for global optimality in the convex case is stated in Theorem 2.1.7. Some regularity conditions need to be made in order to obtain necessity of KKT optimality conditions. We denote the active index set by

$$I_0(\overline{x}) = \{i \in I \mid \ g_i(\overline{x}) = 0\} \tag{2.7}$$

i.e., the index set of active constraints, at $\overline{x}$. We briefly recall famous regularity conditions, called constraint qualifications in finitely constrained programming, namely, the Mangasarian Fromovitz Constraint Qualification (MFCQ), the Linear Independence constraint Qualification (LICQ) and the Slater condition.

Definition 2.1.8 ([77]) *The Linear Independence Constraint Qualification (LICQ) holds at a point $\overline{x} \in M$, if the gradients $\nabla g_i(\overline{x}), i \in I_0(\overline{x})$ and $\nabla h_j(\overline{x}), j \in J$ are linearly independent.*

LICQ assures the following analytic stability result [110].

Lemma 2.1.9 ([110]) *Let the feasible set M_F be compact in P given in (2.3). If LICQ holds at for all $x \in M_F$, then for any small C^1 perturbation $\tilde{g}_i, \tilde{h}_j$ of the functions g_i, h_j the perturbed feasible set $\tilde{M}_F = \{x \mid \tilde{g}_i(x) \leq 0\ (i \in I),\ \tilde{h}_j(x) = 0\ (j \in J)\}$ is diffeomorphic to M_F.*

Mangasarian Fromovitz Constraint Qualification (MFCQ) is another regularity condition and it is stated in the following definition.

Definition 2.1.10 ([2]) *The Mangasarian Fromovitz Constraint Qualification (MFCQ) holds at a point $\bar{x} \in M_F$ if $\nabla h_j(\bar{x})$, $j \in J$ are linearly independent and there exist a vector d satisfying*

$$\begin{aligned} Dg_i(\bar{x})d &> 0\ (i \in I), \\ Dh_j(\bar{x})d &= 0\ (j \in J). \end{aligned}$$

MFCQ is equivalent to the following structural stability [110].

Lemma 2.1.11 ([110]) *Let the feasible set M_F be compact in P given in (2.3). Then for any small C^1 perturbation $\tilde{g}_i, \tilde{h}_j$ of the functions g_i, h_j the perturbed feasibe set $\tilde{M}_F = \{x \mid \tilde{g}_i(x) \leq 0\ (i \in I),\ \tilde{h}_j(x) = 0\ (j \in J)\}$ is (Lipschitz)-homeomorphic to M_F if and only if MFCQ is satisfied for all $x \in M_F$.*

It can be easily shown that MFCQ hold at $\bar{x} \in M_F$ with the vector d, then for any $\tau > 0$ small enough the points $\bar{x} + \tau d$ are interior points of M_F [110].

The following constraint qualification is defined only for convex problems.

Definition 2.1.12 ([13]) *Let P given in (2.3) be a convex problem. The Slater condition is said to be satisfied at a point $\bar{x} \in M_F$ if $\nabla h_j(\bar{x})$, $j \in J$ are linearly independent and there exists a point x such that*

$$\begin{aligned} g_i(x) &< 0\ (i \in I), \\ h_j(x) &= 0\ (j \in J). \end{aligned}$$

The following result shows that under a constraint qualification, a local optimum is a KKT point.

Theorem 2.1.13 ([2]) *Let $\bar{x}$ is a local minimum of P given in (2.3). Further suppose that MFCQ, LICQ or Slater condition is satisfied at $\bar{x}$, then $\bar{x}$ is a KKT point of P.*

By Theorem 2.1.7 and Theorem 2.1.13 we see that, under any of the regularity conditions, being a stationary point is equivalent to global optimality for convex problems.

Let $\bar{x}$ be a KKT point with Lagrange multiplier vectors λ, μ. The tangent space of M_F at $\bar{x}$ is defined as follows:

$$T(\bar{x}) = \{\rho \in R^n | \, Dg_i(\bar{x})\rho = 0, i \in I_0(\bar{x}), Dh_j(\bar{x})\rho = 0, j \in J\}.$$

The tangent cone of M_F at $\bar{x}$ is defined as follows:

$$C(\bar{x}) = \{\rho \in R^n | \, Dg_i(\bar{x})\rho = 0, i \in I_0(\bar{x}), Dg_i(\bar{x})\rho \leq 0, i \in I \setminus I_0(\bar{x}), Dh_j(\bar{x})\rho = 0, j \in J\}.$$

Theorem 2.1.14 ([77]) *Suppose that $\bar{x}$ is a local solution of P, and LICQ holds at $\bar{x}$. Let μ, λ be the corresponding Lagrange multipliers. Then, the Hessian $D_x^2\mathcal{L}(\bar{x}, \mu, \lambda)$ is positive semi-definite on $T(\bar{x})$.*

Theorem 2.1.14 defines a necessary condition involving second derivatives. If $\bar{x}$ is a local solution, then the curvature of the Lagrangian function along the directions in $C(\bar{x})$ must be nonnegative. The existence of a local solution and corresponding Lagrange multipliers is stated in Theorem 2.1.7 and Theorem 2.1.13. A more restrictive second order condition is needed for sufficiency of being a strict local solution. The second order condition to enforce being a strict local solution is that the Hessian of the Lagrangian function $\mathcal{L}$ at $\bar{x}$ is positive definite over $T(\bar{x})$.

Definition 2.1.15 ([77]) *The second order sufficiency condition (SOSC) holds at $\bar{x}$ if $D_x^2\mathcal{L}(\bar{x}, \mu, \lambda)$ is positive definite on $C(\bar{x})$.*

Strict complementarity holds at $\bar{x}$ if $\bar{\mu}_i > 0, g_i(\bar{x}) = 0$ or $\bar{\mu}_i = 0, g_i(\bar{x}) < 0$.

Definition 2.1.16 ([2]) *Strict complementary slackness (SCS) holds at $\bar{x}$ if $\mu_i > 0 \;\; (i \in I_0(\bar{x}))$.*

Definition 2.1.17 ([44]) *The KKT point $\bar{x}$ is called a nondegenerate local minimum of P if LICQ, SOSC and SCS hold at $\bar{x}$.*

It is known that a nondegenerate local optimum $\bar{x}$ is a strict local optimum of P [2]. The stability of both optimal solution $\bar{x}$ and corresponding Lagrange multipliers μ, λ under small data perturbations hold for a nondegenerate local optimum point [44]. Moreover, a nondegenerate local minimum is the unique, global optimum for a convex problem.

2.2 Semi-infinite Programming

In this section we recall some basic facts from semi-infinite programming. We give examples and applications of semi-infinite programming and a basic and well-known regularity condition, namely Reduction Ansatz, is introduced. We also discuss about constraint qualifications and bilevel structure of semi-infinite programming.

Semi-infinite programming problems is a class of optimization problems in finite dimensional decision variables which are subject to infinitely many inequality constraints, as the name *semi-infinite* actually suggests. Theoretical and numerical treatment of *SIP* can be traced back in the literature to papers from 1960's [28, 33, 48, 51, 110, 111, 118, 126]. More than a thousand papers have been published for *SIP*. Historically, *SIP* stems from applications in approximation theory. It is originally related with Chebyshev approximation, see [36]. For an excellent review, we refer to [33] and [82], for linear semi-infinite programming, we refer to [18]. For surveys about theory and methods for standard semi-infinite optimization we refer to [18, 33, 91], whereas introductions to generalized semi-infinite programming are given in [103]. We refer the reader to [20] for a recent tutorial for *GSIP*. For *GSIP* the topological structure of feasible set is investigated in [99, 111, 119], the optimality conditions are studied in [47, 55, 97, 98, 107] and some solution methods are discussed in [4, 21, 105, 109, 112, 118].

Definition 2.2.1 ([103]) *A standard semi-infinite problem has the form*

$$SIP: \qquad \textit{minimize} \quad f(x) \quad \textit{subject to} \quad x \in M$$

with the feasible set

$$M = \{ x \in \mathbb{R}^n | \, g(x,y) \leq 0 \textit{ for all } y \in Y \}$$

and

$$Y = \{ y \in \mathbb{R}^m | \, v_j(y) \leq 0 \ \ (j \in Q) \}.$$

In the definition, we write a problem with only one semi-infinite constraint. This formulation is sufficient when one wishes to study the essential features of the problem. However, in applications one usually deals with finitely many semi-infinite constraints. All involved functions $f,\ g,\ v_j,\ j \in Q = \{1, ..., q\}$, are at least twice continuously differentiable and are assumed to be real-valued on their respective domains.

A simple example for a standard semi-infinite constraint is given in [103], the description of the unit disc in $\mathbb{R}^2$,

$$D = \{\, x \in \mathbb{R}^2 |\, x_1^2 + x_2^2 \leq 1 \,\}$$

by means of infinitely many affine-linear inequality constraints:

$$D = \{\, x \in \mathbb{R}^2 |\, y^T x \leq 1 \text{ for all } y \in Y\}$$

with

$$Y = \{\, y \in \mathbb{R}^2 |\, \|y\|_2 = 1 \,\}.$$

In fact, this describes D as the intersection of infinitely many halfplanes. In Figure 2.1, three of these halfplanes are shown.

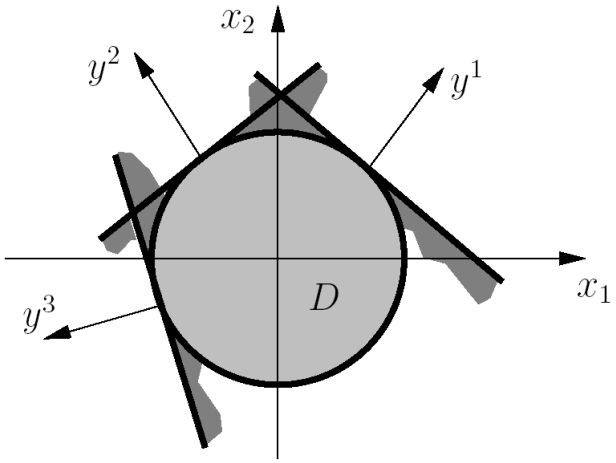

Figure 2.1: The unit disc as intersection of infinitely many halfplanes [103].

As mentioned in [103], if one has a finite description of a set, one does not necessarily search for a semi-infinite one. However, in applications one often only knows a semi-infinite description. In Subsection 2.2.1, a number of these examples will be given. We refer the reader

to Figure 2.2 for an example of feasibility for *SIP*. Here the infinite index set Y is an interval in $\mathbb{R}$. If the decision variable x is one-dimensional, then the restriction function g has two-dimensional arguments. The points x_1 and x_2 are feasible, whereas x_3 is infeasible.

On the other hand, in a *GSIP* problem the infinite index set depends also on the decision variables. We refer the reader to Figure 2.3 for feasibility under a general semi-infinite constraint. Here the restriction function g is same as in Figure 2.2, x_2 is infeasible and x_3 becomes feasible.

Definition 2.2.2 ([103]) *A generalized semi-infinite optimization problem has the form,*

$$GSIP: \quad \textit{minimize} \quad f(x) \quad \textit{subject to} \quad x \in M$$

with the feasible set

$$M = \{\, x \in \mathbb{R}^n \mid g(x,y) \leq 0 \textit{ for all } y \in Y(x) \,\}$$

and

$$Y(x) = \{\, y \in \mathbb{R}^m \mid v_j(x,y) \leq 0 \;\; (j \in Q) \,\}.$$

Here, $Y(x)$ is allowed to vary with x. In a *SIP* the index set is fixed, that is, we have $Y(x) \equiv Y$. We assume that the set valued mapping $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is locally bounded. The problem

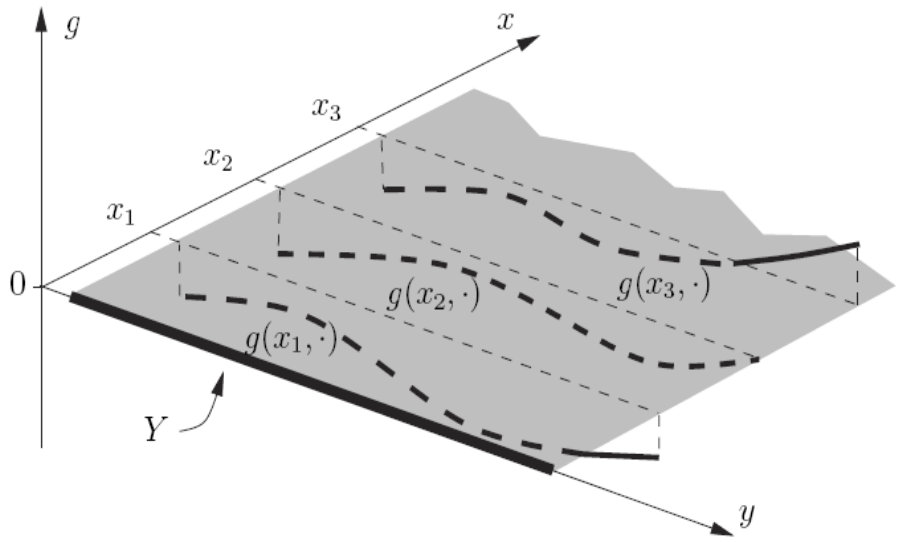

Figure 2.2: Feasibility under a standard semi-infinite constraint [103].

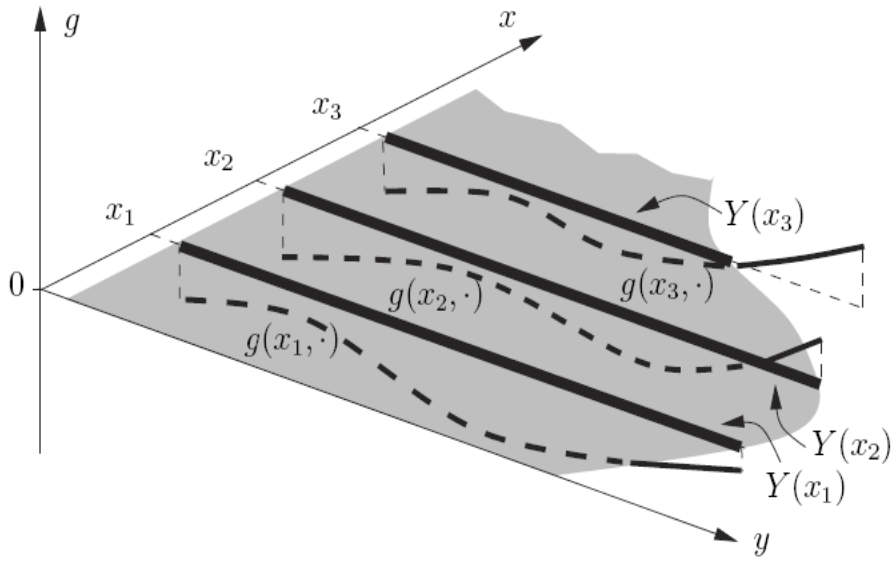

Figure 2.3: Feasibility under a general semi-infinite constraint [103].

formulation of *GSIP* is certainly a generalization of *SIP*, but it is not possible to generalize most of the known theoretical and computational approaches of *SIP* to *GSIP*. We refer to [103] for detailed information.

Definition 2.2.3 ([110]) *A feasible point $\bar{x} \in M$ is called a local minimizer of SIP if there is some $\epsilon > 0$ such that*

$$f(x) - f(\bar{x}) \geq 0 \text{ for all } x \in M \text{ with } \|x - \bar{x}\| < \epsilon.$$

The minimizer $\bar{x}$ is said to be global if this relation holds for any $\epsilon > 0$. For a feasible point $\bar{x}$ of *SIP* we will denote *its set of active indices* by

$$Y_0(\bar{x}) = \{ y \in Y \mid g(\bar{x}, y) = 0 \}$$

Lemma 2.2.4 ([110]) *Let $\bar{x} \in M$ be a local minimizer of SIP. Then, there cannot exist a strictly feasible descent direction d., i.e., a vector $d \in \mathbb{R}^n$ such that*

$$Df(\bar{x})d < 0,\ D_x g(\bar{x}, y)d > 0 \text{ for all } y \in Y_0(\bar{x}).$$

2.2.1 Examples and applications

The examples and figures in this subsection are mainly taken from the book [103].

SIP and *GSIP* problems naturally arise in approximation theory, optimal control and numerous engineering applications. There are a lot of real-life applications for standard and general semi-infinite programming. Examples include Chebyshev approximation and reverse Chebyshev approximation problems, minimax problems, robust optimization, design centering [37, 75, 81], and further problems that are not covered by these problem classes, like defect minimization for operator equations [33, 36, 103], disjunctive programming [101], the optimal layout of an assembly line [54, 118], engineering design [80], or time minimal control [55, 58, 118]. There are other real-world applications of *SIP* such as the shape optimization problem [8], gene-environment networks [121, 122] and optimal control [92].

Example 2.2.5 ([103]) ***Chebyshev Approximation***

SIP is originally related with Chebyshev approximation. Given a continuous function $F : Z \to \mathbb{R}$ with nonempty compact domain $Z \subset \mathbb{R}^M$. The aim is to approximate F by simpler functions $a(p, \cdot)$ with parameter $p \in P \subset \mathbb{R}^N$. In many applications minimizing the maximal deviation is needed, i.e. the Chebyshev norm is used instead of the Euclidean norm (cf. Figure 2.4).

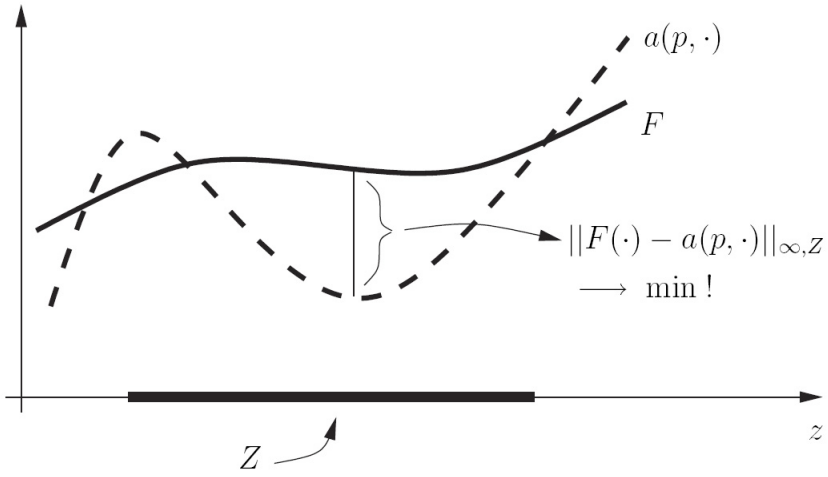

Figure 2.4: Chebyshev approximation [103].

The Chebyshev approximation is a nonsmooth problem of the following form

$$CA: \quad \min_{p\in P} \|F(\cdot) - a(p,\cdot)\|_{\infty,Z} = \min_{p\in P} \max_{z\in Z} |F(z) - a(p,z)|.$$

It is possible to rewrite CA by using the epigraph reformulation

$$\min_{(p,q)\in P\times\mathbb{R}} \quad q \quad \text{such that} \quad \max_{z\in Z} |F(z) - a(p,z)| \leq q,$$

this problem can be written as follows

$$SIP_{CA}: \qquad \min_{(p,q)\in P\times\mathbb{R}} \quad q \qquad \text{such that} \quad |F(z) - a(p,z)| \leq q, \quad \forall z \in Z.$$

SIP_{CA} is a smooth optimization problem if all defining functions are smooth, whereas CA is intrinsically nonsmooth. The price to pay for smoothness is, of course, the presence of infinitely many inequality constraints.

Example 2.2.6 ([103]) ***Reverse Chebyshev Approximation***

A modification of Chebyshev approximation called as reverse Chebyshev approximation occur in some engineering applications, e.g., the construction of low pass filters in digital filtering theory [55]. Consider the continuous function $F : Z(q) \to \mathbb{R}$ on a non-empty compact domain $Z(q) \subset \mathbb{R}^M$ with a parameter $q \in Q$. Suppose that an approximating family of functions $a(p,\cdot)$ and a desired precision $e(p,q)$ are given. Parameter vectors p, q will be found such that the domain $Z(q)$ is as large as possible without exceeding the approximation error $e(p,q)$ (cf. Figure 2.5).

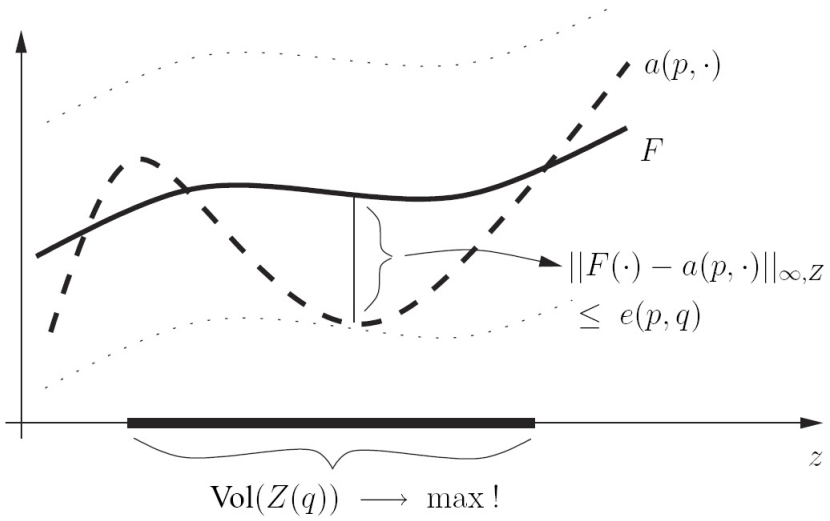

Figure 2.5: Reverse Chebyshev approximation [103].

The Reverse Chebyshev approximation problem can be stated as follows:

$$RCA: \quad \max_{(p,q)\in P\times Q} \quad \mathrm{Vol}(Z(q)) \quad \text{such that} \quad \|F(\cdot) - a(p,\cdot)\|_{\infty,Z} \leq e(p,q),$$

Here, $Vol(Z(q))$ is the M-dimensional volume of $Z(q)$. The *RCA* problem is a nonsmooth optimization problem and it is possible to rewrite this problem by using general semi-infinite constraints. The result is the following *GSIP*:

$$GSIP_{RCA}: \max_{(p,q)\in P\times Q} \quad \mathrm{Vol}(Z(q)) \quad \text{such that} \quad \pm(F(z)-a(p,z)) \leq e(p,q), \forall z \in Z(q).$$

Example 2.2.7 ([103]) ***Minimax problems***

For a non-empty and compact set $X \subset \mathbb{R}^n$, a set-valued mapping Y from $\mathbb{R}^n$ to $\mathbb{R}^m$ and a continuous real-valued function G a problem of the type

$$MM: \min_{x\in X} \max_{y\in Y(x)} G(x,y)$$

is called a minimax problem. The Chebyshev and reverse Chebyshev approximation problems from Example 2.2.5 and Example 2.2.6 are special cases of minimax problems as long as G is only assumed to be continuous. On the other hand, if G is a smooth function, then the epigraph reformulation of MM as a smooth *GSIP* takes a simpler form than in the situation of Chebyshev approximation:

$$GSIP_{MM}: \min_{(x,z)\in X\times\mathbb{R}} \quad z \quad \text{such that} \quad G(x,y) \leq z, \forall y \in Y(x).$$

Possible applications include robust optimization problems (cf. Example 2.2.8) with parameter-dependent objective functions, or the evaluation of formulas for directional derivatives of optimal value functions (cf. Section 4.2 in [103]).

Example 2.2.8 ([103]) ***Robust Optimization***

Let the finitely constrained programming problem depend on an unknown parameter $y \in Y(x)$. Then we have the following formulation:

$$P(y): \quad \min_{x\in\mathbb{R}^n} f(x,y) \quad \text{such that} \quad g_i(x,y) \leq 0 \ \ (i \in I),$$

where $|I| < \infty$. A posteriori approach to solve this problem is based on finding a solution of $P(y_0)$ where y_0 is some nominal choice of parameters and analyze change of optimal value and point for $y \approx y_0$, apparently stability and sensitivity analysis is needed. A priori approach can be used in constrast to this. Pessimistic way to deal with the constraint with some unknown parameter vector is to use its worst case reformulation, which is that the constraint holds for

all values of parameter set. Similarly if the objective function also depends on the unknown parameter, then in the worst case the aim is to minimize the maximal objective value.

Let $Y(x)$ be uncertainty set with $y \in Y(x)$.

$$RP : \min_{x\in\mathbb{R}^n} \max_{y\in Y(x)} f(x,y) \quad \text{such that} \quad g_i(x,y) \leq 0 \ \ (i \in I, y \in Y(x)).$$

The following *GSIP* is obtained:

$$GSIP_{RO} : \min_{(x,z)\in\mathbb{R}^n\times\mathbb{R}} z \quad \text{such that} \quad \begin{array}{r} f(x,y) \leq z \ \ (y \in Y(x)), \\ g_i(x,y) \leq 0 \ \ (y \in Y(x), i \in I). \end{array}$$

We refer to [3] robust optimization models in economics. In Chapter 6 we will treat examples from robust optimization numerically.

Example 2.2.9 ([103]) ***Design Centering***

In a general design centering problem, some functional, e.g. volume, of a parametrized body insribed in a second fixed body, is maximized. The parametrized body is usually called as design, $B(x)$ where x is the parameter and the second fixed body is called as container C.

$$DC : \max_{x\in\mathbb{R}^n} f(x) \ \text{ such that } \ B(x) \subset C.$$

The container C often has a complicated structure in applications and one wishes to find a lower bound on its volume by inscribing a simpler body $B(x)$. Problems of this type have been studied extensively, see e.g. [37] or [81] and the references therein. Let C be described by the inequality constraint $c(y) \leq 0$. It is easy to see that the inclusion

$$B(x) \subset C = \{y \in \mathbb{R}^m \mid c(y) \leq 0\}$$

is equivalent to the following general semi-infinite constraint

$$c(y) \leq 0 \qquad \text{for all} \quad y \in B(x).$$

So, the design centering problem is equivalent to the following *GSIP*:

$$GSIP_{DC} : \max_{x\in\mathbb{R}^n} \text{Vol}(B(x)) \quad \text{such that} \quad c(y) \leq 0 \quad \forall\, y \in B(x).$$

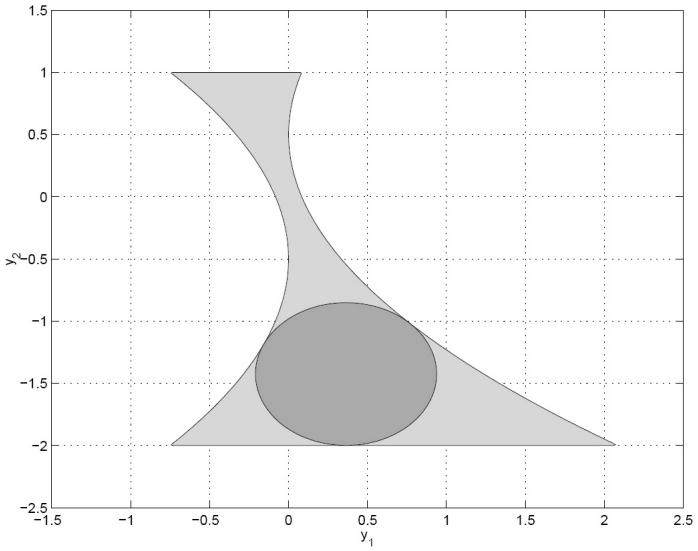

Figure 2.6: A disk with maximal area in a nonconvex container [104].

An example of a design centering problem is plotted in Figure 2.6. In this example the design $B(x)$ is a disk in $\mathbb{R}^2$, parametrized by its midpoint and its radius. Find the disc $B(x)$ having maximal area in C, i.e., find the parameter vector $x \in \mathbb{R}^3$ corresponding the maximal area. A special design centering problem is called as maneuverability problem of a robot in [19, 24, 33]. This problem leads to the one of the first formulations of a *GSIP* [35]. The explanations for maneuverability problem of a robot are from [38]. A robot can be schematically represented as a connection of links. For example the robot in Figure 2.7 has two links. Each link has a length and mass associated with it.
The optimization problem results from minimizing a certain performance index (time, energy consumption, etc.) while ensuring that the robot can follow the resulting trajectory.

Minimize total travel time:

Model 1: joint velocity, acceleration and jerk constraints are imposed.

Model 2: torque constraints are imposed.
A change of variables that relates time to the parameter τ: $t = h(\tau)$ $(\tau \in [0, 1])$ is considered.

For Model 2, minimizing total travel time (*SIP*)

$$\min\ h(1) = \int_0^1 h'(\tau) d\tau$$

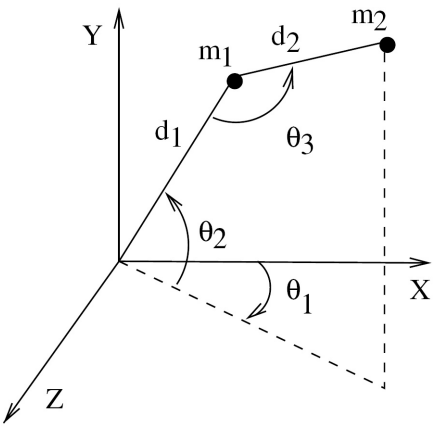

Figure 2.7: Three degrees of freedom robot [38].

$$\text{such that } h(0) = 0,$$

$$h'(\tau) > 0, \ |F_i(\tau)| \leq C_i \ (i = 1, 2, \ldots, l) \ \forall \tau \in [0, 1].$$

For practical purposes, the unknown function $h'(\tau)$ is approximated by a B-spline. AMPL modeling language for codification of problem and NSIPS as a nonlinear semi-infinite programming solver are used. Discretization methods [32, 90] are implemented in the solver. For numerical results we refer to [38].

Another special design centering problem recently studied and also applied in industry is lapidary cutting problems [128, 129]. The lapidary cutting problem deals with the maximization of the volume of a faceted coloured gemstone (the design), which is cut from an irregularly shaped rough stone. There are several basis shapes for the design. In this application, only convex designs are considered. The faceted gemstone is parameterized by its geometrical position and orientation within the rough stone as well as by some other shape parameters such as height, radius or the length-width ratio. For container constraints, rough stone data is given in the form of triangulated mesh rather than in functional constraint form. An approximate smooth container representation satisfying convexity assumption is obtained from mesh data. The lapidary cutting problem can be classified as a special case of a design centering problem, which is formulated as *GSIP* and solved by an interior point method developed in [103, 106]. In Figure 2.8 an example of four views of the optimal faceted lapidaries inscribed into the functional constraint approximation of the rough stone surface are showed [129].

The last example of a recently revisited design centering problem we want to mention is air pollution control [39]. We refer to [37, 75, 81] for design centering problems for further reading. For simple numerical examples of design centering problems, see Chapter 6.

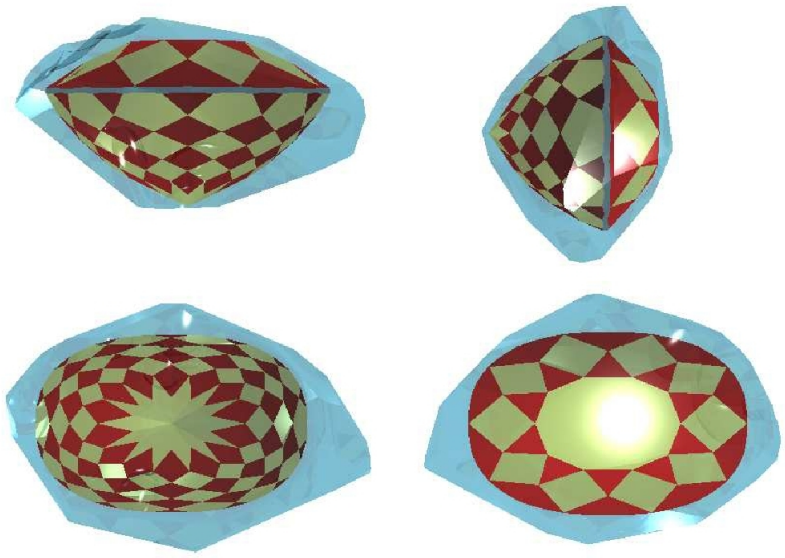

Figure 2.8: The optimal faceted lapidaries inscribed into the functional constraint approximation of the rough stone surface [129].

2.2.2 The Reduction Ansatz

In finitely constrained programming problems it is well known that around a feasible point $\bar{x} \in \mathbb{R}^n$ there exists a neighborhood $N(\bar{x})$ for which the feasible set can be described by (usually at most n) constraints which are active at $\bar{x}$. If $\bar{x}$ is a solution of the finitely constrained programming problem locally, then it is also a solution of the problem locally at which all inactive constraints at $\bar{x}$ are dropped and conversely [91].

The feasible set of an *SIP* or *GSIP* normally can not be locally represented by (usually finitely many) active constraints only. However, under proper assumptions (Reduction Ansatz), for $\bar{x} \in \mathbb{R}^n$ there exist a neighborhood of $\bar{x}$ and a finite number of certain implicitly defined inequality constraints such that the feasible set defined by these constraints coincides with the feasible set of *SIP* or *GSIP*. Hence under Reduction Ansatz, *SIP* or *GSIP* can be locally reduced to a finitely constrained programming problem at least conceptually (local reduction) for *SIP* we refer the reader to [28].

Reduction Ansatz was originally formulated for standard *SIP* under weaker regularity assumptions [28, 126]. It was transferred to *GSIP* in [34]. The main consequence of the Reduction Ansatz is that the feasible set M of either standard *SIP* or *GSIP* can locally be described by finitely many C^2-constraints, then standard *SIP* or *GSIP* locally looks like a smooth finite programming problem. The Reduction Ansatz serves as a basic regularity condition for numerical solution methods in semi-infinite programming. For *GSIP* it is a recent result that the Reduction Ansatz generically holds at all local minimizers [23]. Reduction Ansatz is a common way to obtain optimality conditions and Newton-type methods for *SIP* and *GSIP*. We use Reduction Ansatz to obtain optimality conditions for *GSIP* and for convergence of our method. In this subsection, we use definitions as in our recent paper [108].

For a feasible point $\bar{x}$ of *GSIP* we will denote its *set of active indices* (possibly infinite) by $Y_0(\bar{x}) = \{y \in Y(\bar{x}) \mid g(\bar{x}, y) = 0\}$. Let $\mathcal{L}(\bar{x}, y, \gamma) = g(\bar{x}, y) - \gamma^T v(\bar{x}, y)$, be the *Lagrangian* associated with the so-called *lower level problem* $Q(\bar{x})$,

$$Q(x): \quad \max_{y \in \mathbb{R}^m} g(x, y) \quad \text{subject to} \quad v_j(x, y) \leq 0, \ j \in Q. \tag{2.8}$$

where γ is the Lagrangian multiplier vector. The lower level problem is a finite, parametric optimization problem, here x plays the role of an n-dimensional parameter and y is the decision variable. We refer to Section 4.3.1 for more information about lower level problem. On the other hand, in the upper level problem minimizing $f(x)$ with x is the decision variable. The set $Y_0(\bar{x})$ coincides with the set of global maximizers of the lower level problem $Q(\bar{x})$ in the case $\varphi(\bar{x}) = 0$ where $\varphi(x)$ is the optimal value function of the family of the lower level problems. For an illustration see Figure 2.9.

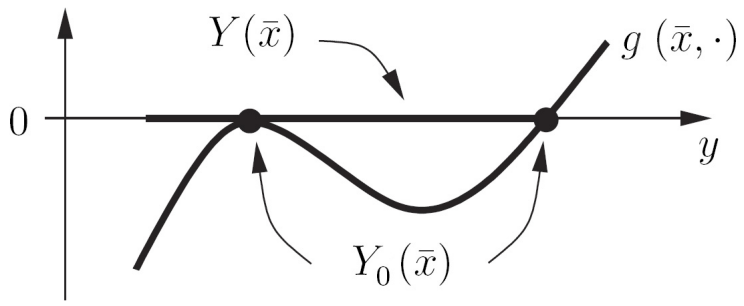

Figure 2.9: An illustration of two active indices [103].

The points $\overline{y}^i \in Y_0(\overline{x})$ satisfy system of equations for $Q(\overline{x})$:

$$\nabla_y g(\overline{x},\overline{y}^i) - \sum_{j=1}^{q} \overline{\gamma}^i_j \nabla_y v_j(\overline{x},\overline{y}^i) = 0,$$
$$\overline{\gamma}^i_j v_j(\overline{x},\overline{y}^i) = 0 \ \ (i \in P,\ j \in Q).$$

The elements of $Y_0(x)$ for varying x has to be under control for both theoretical and numerical purposes and it is achieved by implicit function theorem. The Jacobian of the system with respect to y, γ is nonsingular if linear independence constraint qualification, strict complementarity and second order sufficiency condition (with a *negative definite* Hessian) hold for $Q(\overline{x})$, this result is implied by [41]. In this case $\overline{y}^i \in Y_0(\overline{x})$ are *nondegenerate* global maximizers of $Q(\overline{x})$ in the sense of [43].

In the following we formulate first order optimality conditions as well as corresponding second order optimality conditions of finitely constrained programming from Section 2.1 to a parametric finitely constrained programming problem $Q(\overline{x})$.

Suppose that the following conditions (Q I)-(Q III) hold at some $\overline{y} \in Y_0(\overline{x})$ in $Q(\overline{x})$:

(Q-I) The *linear independence constraint qualification*:

$$(LI)_{Q(\overline{x})} \ \ \{\nabla_y v_j(\overline{x},\overline{y}) \mid j \in Q_0(\overline{x},\overline{y})\} \text{ is a linearly independent family,} \qquad (2.9)$$

where $Q_0(\overline{x},\overline{y}) = \{j \in Q \mid v_j(\overline{x},\overline{y}) = 0\}$ is the set of lower level active indices at $\overline{y} \in Y(\overline{x})$.

Because of Q-I, we have the following lower level Karush-Kuhn-Tucker conditions (See Sections 2.1 and 2.4): there exists a unique vector of Lagrange multipliers $\overline{\gamma} \in \mathbb{R}^q$ such that

$$(KKT)_{Q(\overline{x})} \left\{ \begin{array}{c} \nabla_y g(\overline{x},\overline{y}) - \sum_{j=1}^{q} \overline{\gamma}_j \nabla_y v_j(\overline{x},\overline{y}) = 0 \\ v_j(\overline{x},\overline{y}) \le 0 \\ \overline{\gamma}_j \ge 0 \\ \overline{\gamma}_j v_j(\overline{x},\overline{y}) = 0,\ j \in Q. \end{array} \right\} \qquad (2.10)$$

(Q-II) *Strict complementarity*:

for each $j \in Q$: $\overline{\gamma}_j > 0, v_j(\overline{x},\overline{y}) = 0$ or $\overline{\gamma}_j = 0, v_j(\overline{x},\overline{y}) < 0$.

(Q-III) The *second order sufficiency condition*:

$$(SOSC)_{Q(\overline{x})} \left\{ \begin{array}{c} \eta^T \nabla^2_y \mathcal{L}(\overline{x},\overline{y},\overline{\gamma})\eta < 0 \ \text{ for all } \eta \in G_{Q(\overline{x})} \setminus \{0\}, \text{where} \\ G_{Q(\overline{x})} = \{\eta \in \mathbb{R}^m \mid D_y v_j(\overline{x},\overline{y})\eta = 0,\ j \in Q_0(\overline{x},\overline{y})\} \end{array} \right\} \qquad (2.11)$$

with $\mathcal{L}(\bar{x}, y, \gamma) = g(\bar{x}, y) - \sum_{j=1}^{q} \gamma_j v_j(\bar{x}, y)$, the *Lagrangian* associated with $Q(\bar{x})$.

The conditions (Q-I) to (Q-III) state that $\bar{y}$ is a nondegenerate global maximizer of the lower level problem in the sense of Jongen/Jonker/Twilt [43]. The *Reduction Ansatz* is said to hold at $\bar{x} \in M$ if all elements $\bar{y} \in Y_0(\bar{x})$ are nondegenerate for the lower level problem.

Suppose that Reduction Ansatz holds at $\bar{x} \in M$. Then we can reduce *GSIP* locally (in a neighborhood of $\bar{x}$) to a smooth finite optimization problem $GSIP_{red}$, the so-called reduced *GSIP*, as given in the next theorem.

Theorem 2.2.10 ([28, 30, 33, 34, 36, 110, 118, 126]) *Let the Reduction Ansatz be satisfied at a feasible point $\bar{x}$ of GSIP. Then,*

(a) The active index set is finite, $Y_0(\bar{x}) = \{\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^p\}$, and there exist neighborhoods $U_{\bar{x}}$ of $\bar{x}$ and $V_{\bar{y}^i}$ of $\bar{y}^i$ and unique C^1-functions

$y^i : U_{\bar{x}} \to V_{\bar{y}^i}$ *, where* $y^i(\bar{x}) = \bar{y}^i$,

$\gamma^i : U_{\bar{x}} \to \mathbb{R}^q$, *where* $\gamma^i(\bar{x}) = \bar{\gamma}^i$,

such that for every $x \in U_{\bar{x}}$ the value $y^i(x)$ is the unique local maximizer of $Q(x)$ in $V_{\bar{y}^i}$ with corresponding Lagrange multiplier vector $\gamma^i(x)$.

(b) The following finite reduction holds: $\bar{x}$ is a solution of GSIP, locally in a neighborhood $U_{\bar{x}}$ of $\bar{x}$, if and only if $\bar{x}$ is a local solution of the so-called reduced problem

$$GSIP_{red}: \min_{x \in U_{\bar{x}}} f(x) \quad \text{subject to} \quad \varphi_i(x) = g(x, y^i(x)) \leq 0, \quad \text{for all } i = 1, 2, \ldots, p.$$

(c) The functions φ_i from part (b) are of class C^2, and for all $x \in U_{\bar{x}}$ their gradients satisfy

$$D_x \varphi_i(x) = D_x \mathcal{L}(x, y^i(x), \gamma^i(x)). \tag{2.12}$$

Proof. We refer to [28]. The functions $y^i(x)$ and $\gamma^i(x)$ in part (a) is a result of the well-known Implicit Function Theorem.

Remark 2.2.11 ([108]) *For standard SIP the formula in Theorem 2.2.10(c) simplifies to*

$$D_x \varphi_i(x) = D_x g(x, y^i(x)).$$

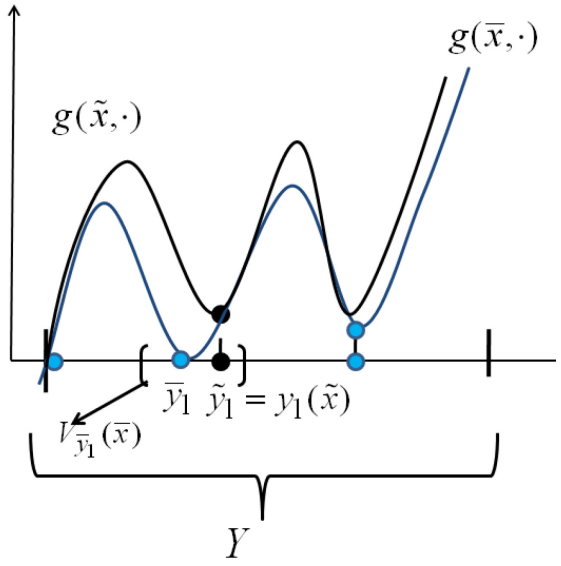

Figure 2.10: Minimizing constraint function, $g(\bar{x},\cdot)$ under the Reduction Ansatz [79].

As explained in [79], in Theorem 2.2.10 for a given feasible point $\bar{x}$, infinitely many constraints are reduced to finitely many constraints by solving the lower level problem $Q(\bar{x})$; see Figure 2.10. In Figure 2.10, $\tilde{x}$ represents the small perturbation of $\bar{x}$, i.e., $\bar{x} \to \tilde{x}$. Finitely many local minima of $Q(\bar{x})$ lead to finitely many active inequality constraints for the upper level problem. Solving the reduced problem is equivalent to solving *SIP* or *GSIP* locally. The drawback of Theorem 2.2.10 is that the indices implicitly depend on the variable x.

2.2.3 Constraint qualifications

In the following we recall the Fritz John and the Karush-Kuhn-Tucker optimality conditions in semi-infinite programming. For $\bar{x} \in M$ recall that

$$Y_0(\bar{x}) = \{y \in Y(\bar{x}) \mid g(\bar{x},y) = 0\}$$

is the active index set at $\bar{x}$. Let $\mathcal{L}(\bar{x},y,\gamma) = g(\bar{x},y) - \gamma^T v(\bar{x},y)$ be the *Lagrangian* associated with the *lower level problem* $Q(\bar{x})$.

(i) The *Extended Linear Independence Constraint Qualification* (ELICQ) is said to hold at $\bar{x} \in M$ if the vectors

$$D_x\mathcal{L}(\bar{x},\bar{y}^j,\bar{\gamma}^j),\ \ (\bar{y}^j \in Y_0(\bar{x}))\ \text{ are linearly independent as a family.} \tag{2.13}$$

(ii) The *Extended Mangasarian-Fromovitz Constraint Qualification (EMFCQ)* holds at $\bar{x}$, if there is a vector $d \in \mathbb{R}^n$ such that

$$D_x \mathcal{L}(\bar{x}, \bar{y}^j, \bar{\gamma}^j) d < 0 \quad \text{for all} \quad \bar{y}^j \in Y_0(\bar{x}). \tag{2.14}$$

For standard *SIP* (i) and (ii) reduce to

(i*) $D_x g(\bar{x}, \bar{y}^j)$, $(\bar{y}^j \in Y_0(\bar{x}))$ are linearly independent as a family.

(ii*) $D_x g(\bar{x}, \bar{y}^j) d < 0 \quad \text{for all} \quad \bar{y}^j \in Y_0(\bar{x})$

with $Y_0(\bar{x}) = \{y \in Y | \, g(\bar{x}, y) = 0\}$.

We will use these conditions in Subsection 5.1.1. It is well-known that ELICQ implies EMFCQ, i.e., ELICQ is a stronger constraint qualification.

Theorem 2.2.12 ([112]) *Let $\bar{x} \in M$. Suppose, at any $\bar{y} \in Y_0(\bar{x})$ the MFCQ is satisfied for $Q(\bar{x})$. Then, the following holds: There exists $\bar{y}^j \in Y_0(\bar{x}), \bar{\gamma}^j \geq 0, i = 1, \ldots, p$ and multipliers $\bar{\mu}_0, \bar{\mu}_1, \ldots, \bar{\mu}_p \geq 0$ such that*

$$\bar{\mu}_0 \nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x \mathcal{L}(\bar{x}, \bar{y}^j, \bar{\gamma}^j) = 0. \tag{2.15}$$

Proof. For a short proof we refer to [97]. For standard *SIP*, we have $\nabla_x \mathcal{L}(\bar{x}, \bar{y}^j, \bar{\gamma}^j) = \nabla_x g(\bar{x}, \bar{y}^j)$.

Theorem 2.2.13 ([108]) *Let $\bar{x}$ be a local minimizer of GSIP at which Reduction Ansatz and MFCQ for reduced problem $GSIP_{red}$ is satisfied. Then there exist a $p \in \{0, \ldots, n\}$ and multipliers $\bar{\mu}_i \geq 0$ and active indices $\bar{y}^j \in Y_0(\bar{x})$, $i \in \{1, 2, \ldots, p\}$, such that*

$$\nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x \mathcal{L}(\bar{x}, \bar{y}^j, \bar{\gamma}^j) = 0. \tag{2.16}$$

For standard *SIP* we have the following theorem for KKT conditions.

Theorem 2.2.14 ([42]) *Let $\bar{x}$ be a local minimizer of SIP at which EMFCQ is satisfied. Then there are a $p \in \{0, \ldots, n\}$, multipliers $\bar{\mu}_i \geq 0$ and active indices $\bar{y}^j \in Y_0(\bar{x})$, $i \in \{1, 2, \ldots, p\}$, such that*

$$\nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x g(\bar{x}, \bar{y}^j) = 0. \tag{2.17}$$

We refer to [28, 126] for first order and second order sufficient and necessary optimality conditions for standard *SIP*.

2.3 A Review of Numerical Methods for Semi-infinite Programming

For a review of numerical methods for standard semi-infinite programming see [33, 91, 112]. Natural ways to solve a standard *SIP* problem are either replacing the infinite index set Y by a finite one or solving a sequence of finite subproblems, where the infinite index set Y is replaced by finite approximations.

A brief review of numerical methods used to solve *SIP*

- Discretization methods,
- Exchange methods,
- Methods based on local reduction.

The discretization, exchange methods and methods based on local reduction basically replace *SIP* by (a sequence of) finitely constrained programming problems, i.e., problems with only a finite number of constraints. These are solved by applying appropriate linear or nonlinear programming algorithms, for which we refer to an extensive literature. Furthermore, these methods are examples of superlinearly convergent methods to compute numerically a solution of *SIP*, under the additional smoothness of the constraint, $g(x, y)$, with respect to x. Only the method based on local reduction can be extended from *SIP* to *GSIP* without encountering any difficulty. But there arise serious difficulties when trying to generalize the exchange or discretization methods from *SIP* to *GSIP*. These difficulties are discussed and convergence results are derived under quite general assumptions for *GSIP* in [112]. Other numerical methods for special structure, in particular for g linear in x are: Simplex-like methods, Cutting plane methods, etc. [91].

Discretization Method for *SIP*

Discretization method is an obvious numerical method for *SIP*. It's one of the earliest and most frequently used methods for solving *SIP* problems in engineering applications [33]. It is based on a discretization of the infinite index set of inequality constraints. The infinite index set Y is approximated by a sequence of finite subsets $\{Y_k\}$ such that Y_k becomes denser and denser in Y as k goes to infinity in a discretization method (cf. Figure 2.11). Then the *SIP* problem is approximated by a sequence of nonlinear programming problems. Thus, instead of *SIP* solve

$$SIP^k: \quad \min_{x\in\mathbb{R}^n} \quad f(x) \quad \text{such that} \quad g(x,y) \leq 0, \forall y \in Y^k$$

with $Y^k \subset Y, |Y^k| < \infty$

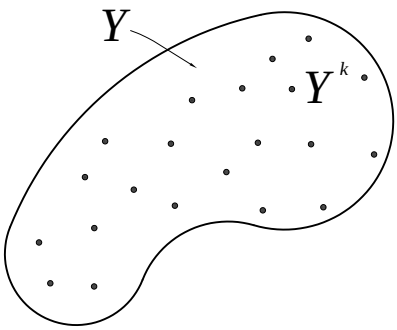

Figure 2.11: An example of a simple discretization.

Algorithm 2.3.1 ([110]) *Conceptual discretization method*

Step k: Given discretization Y^k of Y

i) Compute a solution x^k of SIP^k

ii) Terminate if x^k is approximately feasible.(i.e. $g(x^k, y) \leq \epsilon, \forall y \in Y$ with $\epsilon > 0$)

Otherwise choose finer discretization $Y^{k+1} \subset Y$

However, in a general discretization method, the subset $Y_k \subset Y$ must be sufficiently dense in Y when k is sufficiently large. This makes the algorithm computationally very expensive.

The time needed to verify feasibility and to solve this problem increases dramatically as the cardinality of Y_k grows. For accuracy cardinality of Y^k must be increased but this results an unbounded number of constraints, hence the problem can not be solved for too large k, standard nonlinear programming solvers work for $k < 100.000$.

For numerical implementation the generation of finite subset (a priori or by an implicit rule) and the choice of nonlinear programming solvers are basic elements. A general convergence result for discretization method was obtained [110] under a compactness assumption on the feasible sets.

Exchange Method for *SIP* ([110])

This class can be indeed considered as a subclass of discretization methods since it basically depends on the discretization methods. Given a discretization, the reduced problem of *SIP* is solved, and in a next iteration, discretization points become updated, until the algorithm terminates according to some stopping criterion. In exchange method, there is an exchange of contraints, in every step a number of new constraints are added and some of old constraints may be deleted.

Algorithm 2.3.2 ([110]) *Conceptual Exchange method*

Step k: Given discretization Y^k of Y and $\epsilon > 0$

i) Compute a solution x^k of SIP^k.

ii) Calculate local solutions y_i^k, $i = 1, \ldots, i_k$ of $Q(x^k)$ such that one of them, say y_1^k is a global solution.

iii) Terminate if $g(x^k, y_1^k) \leq \epsilon$ with a solution $\bar{x} \approx x^k$. Otherwise put

$$Y^{k+1} = Y^k \cup \{y_i^k, i = 1, \ldots, i_k\}$$

We observe that the substep ii) in Algorithm 2.3.2 is very costly as it requires a global search for minima of $Q(x^k)$. One must avoid an execution of this step in the overall process as much as possible. Substep ii) assumes that there are only finitely many minima of the lower level problem, if it does not hold, another method, e.g., discretization, should be used. The convergence theorem of the exchange method is presented in [110]. We refer to [29, 31, 33] for a

detailed explanation of discretization and exchange methods.

Methods based on local reduction for *SIP*

These methods are based on the local reduction of the *SIP* (see, Reduction Ansatz Subsection 2.2.2).

Algorithm 2.3.3 ([33]) *Conceptual reduction method*

Step k: One is given x^k not necessarily feasible.

i) Determine all local maxima $y^1, \ldots, y^{r_k}$ of $Q(x^k)$.

ii) Apply m_k steps of a finitely constrained programming algorithm to the reduced problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \textit{such that} \quad g^l(x) = g(x, y^l(x)) \leq 0 \ \ (l = 1, 2, \ldots, r_k).$$

Let $x^{k,j}$, $j = 1, \ldots, m_k$ be the iterates.

iii) Set $x^{k+1} = x^{k,m_k}$.

Substep i) is very costly as it requires a global search for maxima of $Q(x^k)$. In substep ii) finite programming methods such as SQP methods could be used.

We want to emphasize that the numerical solution of a generalized semi-infinite programming problem might be much more difficult than the solution of standard *SIP*. For *GSIP* numerical methods are basically based on the followings [20]:

- An explicit or implicit transformation of *GSIP* into a *SIP*.
- An generalization of methods for *SIP* to the *GSIP*.

In the following we present numerical methods proposed for *GSIP* in the literature. In [68] a branch and bound approach was proposed for *GSIP* where upper level is concave and lower level is linear, however these are very strong and restrictive assumptions. In [21] an algorithm based on Newton-SQP approach (which works well in standard *SIP*) was proposed to terminal variational problems (give rise to *GSIP*) without numerical results. In [118] it was shown that *GSIP* can locally be transformed to *SIP*, but the transformation function and new index

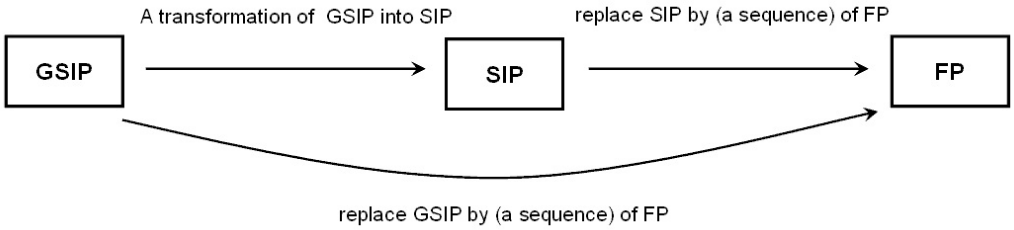

Figure 2.12: A possible scheme of numerical methods for *SIP* and *GSIP*.

set might be too expensive to determine explicitly for especially in the case of multidimensional index sets [103]. In [112] the difficulty in generalizing discretization methods from *SIP* to *GSIP* was emphasized. A conceptual method based on discretization methods (well-known methods for *SIP*) was proposed for *GSIP* and convergence is shown if x-dependent grid points are chosen such that they depend continuously on x in [112], but again it is difficult to implement this method for especially in the case of multidimensional index sets. In [106], *GSIP* with convex lower level problems was transformed to equivalent nonlinear programs, first numerical results for *GSIP* was obtained, this algorithm was implemented in [128] to the gemstone cutting problem. In [17] a conceptual method for solving *GSIP* via global optimization by exact discontinuous penalties but it is also incapable of providing numerical procedure. In [67] a branch and bound framework and uses discretization coupled with convexification for lower level problem and interval constrained reformulation for upper level problem was proposed.

Since we give optimality conditions (both in classical and semismooth form) based on local reduction, Newton and Semismooth Newton methods can be considered as a subclass of methods based on local reduction.

It is far beyond the scope of the thesis to give details of these methods. We refer to [29, 31, 33, 91, 110] and the references cited therein for further reading.

CHAPTER 3

SMOOTH NEWTON METHOD

3.1 Preliminaries on the Newton Method

Recall that Newton method,

$$z^{k+1} = z^k - (DF(z^k))^{-1}F(z^k) \tag{3.1}$$

is a classical method for solving the nonlinear equation,

$$F(z) = 0, \tag{3.2}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable function and the initial point, z_0, is given.

The interpretation of (3.1) is that we model F at the current iterate z^k with a linear function

$$M_k(z) := F(z^k) + DF(z^k)(z - z^k) \tag{3.3}$$

where the root of M_k is the next iteration z^{k+1}. Here, M_k is called the *local linear model*. If $DF(z^k)$ is nonsingular, then $M_k(z^{k+1}) = 0$ is equivalent to (3.1). The geometric interpretation of (3.1) and (3.3) in case $n = 1$ is sketched in Figure 3.1, the point z^{k+1} is the intersection of $z-$axis with the tangent line to the graph of F at the point $(z^k, F(z^k))$. We see that z_{k+1} is a better approximation than z_k for the root $\bar{z}$ of the function F.

It is well-known that if

- the equation (3.2) has a solution $\bar{z}$,
- $DF : \Omega \subset \mathbb{R}^n \to \mathbb{R}^{n\times n}$ is Lipschitz continuous, and
- $DF(\bar{z})$ is nonsingular,

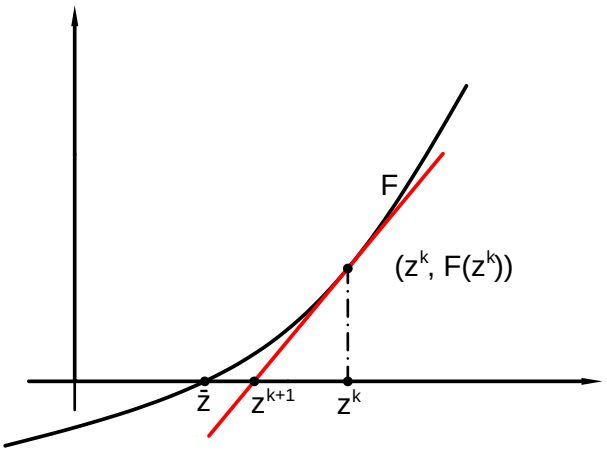

Figure 3.1: An illustration of one iteration of Newton method.

then, if z_0 is sufficiently close to $\bar{z}$, the iteration (3.1) locally converges q-quadratically (see Definition 4.1.9) to $\bar{z}$.

For more information about Newton method we refer the reader to [9, 56, 57] and references therein.

3.2 Newton Method for Finitely Constrained Programming

Newton method was applied to smooth form of KKT conditions in finitely constrained programming [1, 71, 126]. The results of the present section is based on [110]. The KKT-approach for solving a finitely constrained programming problem is based on the system of Karush-Kuhn-Tucker conditions. Let us consider the following finitely constrained programming problem (without equality constraints),

$$P : \min_{x} f(x) \quad \text{subject to} \quad g_i(x) \leq 0 \ \ (i \in I)$$

with $I = \{1, \ldots, r\}$. Let the active index set (2.7) be $I_0(\bar{x}) = \{1, \ldots, k\}$. At a local minimizer $\bar{x}$ under LICQ (see Definition 2.1.8), the following KKT-equations (compare KKT conditions

(2.4) in Section 2.1)

$$\begin{aligned} \nabla f(\bar{x}) + \sum_{j=1}^{k} \bar{\mu}_j \nabla g_j(\bar{x}) &= 0 \\ g_j(\bar{x}) &= 0 \ \ (j \in I_0(\bar{x})) \end{aligned} \tag{3.4}$$

hold with the multipliers $\bar{\mu}_j \geq 0$, $j \in I_0(\bar{x})$. Let Lagrangian be the function

$$\mathcal{L}(x, \mu) = f(x) + \sum_{j=1}^{k} \mu_j g_j(x)$$

Let $G := [g_j \ \ (j \in I_0(\bar{x}))]$. Then KKT condition (3.4) can be written as

$$\begin{aligned} \nabla_x \mathcal{L}(\bar{x}, \mu) &= 0 \\ G(\bar{x}) &= 0. \end{aligned} \tag{3.5}$$

Rewriting one obtains the Jacobian of the system (3.5) at a solution:

$$J(\bar{x}, \bar{\mu}) = \begin{pmatrix} \nabla_x^2 \mathcal{L}(\bar{x}, \bar{\mu}) & \nabla^T G(\bar{x}) \\ \nabla G(\bar{x}) & 0 \end{pmatrix}.$$

(3.5) represents a nonlinear system of $n + k$ equations in as many as unknowns x_i, μ_j. Newton method can be applied to solve (3.5).

It is well-known that the Newton method is q-quadratically convergent locally if the Jacobian is nonsingular at a solution (see [56],[63]). The well-known second order sufficient optimality conditions for finitely constrained programming problems at a local minimizer $\bar{x}$ imply that the Jacobian $J(\bar{x}, \bar{\mu})$ is nonsingular (cf. Example 12.30 in [13]). Under LICQ and SOSC (see Definition 2.1.15) Jacobian is nonsingular by Lemma B.0.37 in Appendix B. We can expect that in the general case (i.e., generically) this regularity condition is satisfied. By the following theorem, it is shown that Newton method is generically applicable.

Let $\mathcal{P} := C^\infty(\mathbb{R}^n, \mathbb{R})^{1+m}$ denote the set of problem functions $P = (f, g_1, \ldots, g_m)$. The function space $C^\infty(\mathbb{R}^n, \mathbb{R})^{1+m}$ is assumed to be endowed with the so-called Withney topology. The C^r Whitney (or strong) topology is a topology assigned to the space $C^r(M, N)$ of mappings from a C^r manifold M to a C^r manifold N having r continuous derivatives. It gives a notion of proximity of two C^r mappings, and it allows us to speak of *robustness* of properties of a mapping. A subset which is dense and open in $\mathcal{P}$ is named as generic subset. The Newton method is generically applicable as stated in the next theorem.

Theorem 3.2.1 ([44]) *There is an open and dense subset $\mathcal{P}_0 \subseteq \mathcal{P}$ such that for all finite programs $P \in \mathcal{P}_0$, LICQ holds at each feasible point x and at each solution $(\bar{x}, \bar{\mu})$ of (3.5) the Jacobian $J(\bar{x}, \bar{\mu})$ is nonsingular.*

3.3 Newton Method for Semi-infinite Programming (*GSIP*)

We recall the results in [112] in this section. In [112] smooth Newton method is applied to the *GSIP* problem from Section 2.2. In [22] a Newton method is originally proposed for the case of linear problems. A Newton method is proposed for nonlinear Chebyshev approximation in [26] and later on [30] a Newton method is proposed for solving standard *SIP*. In this section, we recall the result in [112] about convergence of Newton method for *GSIP*. Newton method is classified under continuous methods (methods based on local reduction) in [29]. In [117] both theoretical and numerical treatment of Newton method applied to standard *SIP* is considered.

Applying Newton method to the necessary optimality conditions is a common method for solving *SIP* [33, 91] . We need some theoretical considerations in order to derive optimality conditions for *SIP* and *GSIP*. The Reduction approach is a common way to obtain optimality conditions and Newton-type methods for *SIP* [33]. The idea here is to locally transform *SIP* and *GSIP* into finite parametric optimization problems. By Theorem 2.2.10, the problem *GSIP* is equivalent to $GSIP_{red}(\bar{x})$ locally near $\bar{x}$, here $GSIP_{red}(\bar{x})$ is a finitely constrained programming problem. Hence, standard optimality conditions of finitely constrained programming can be used to obtain optimality conditions for *GSIP*.

Under LICQ, at a local minimizer $\bar{x} \in M, \bar{y}^l \in Y_0(\bar{x})$ where $Y_0(\bar{x})$ denotes the set of active indices, the following KKT condition is fulfilled (compare KKT from Q-I in Subsection 2.2.2), i.e., there exists a multiplier vector $\bar{\gamma} \geq 0, \bar{\gamma} \in \mathbb{R}^{\bar{q}}$ such that

$$\nabla_y g(\bar{x}, \bar{y}^l) - \sum_{j=1}^{\bar{q}} \bar{\gamma}_j^l \nabla_y v_j(\bar{x}, \bar{y}^l) = 0 \tag{3.6}$$

Here, we let $Q_0(\bar{x}, \bar{y}^l) = \{1, \ldots, \bar{q}\}$, where $Q_0(\bar{x}, \bar{y}^l) = \{j \in Q \mid v_j(\bar{x}, \bar{y}^l) = 0\}$ is the set of lower level active indices at $\bar{y}^l \in Y(\bar{x})$. Let $\bar{x}$ be a local minimizer of *GSIP*. Under Reduction Ansatz at $\bar{x}$, suppose *EMFCQ* holds at $\bar{x}$, then by Theorem 2.2.13, the following KKT condition is fulfilled (compare with (4.34))

$$\nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x \mathcal{L}(\bar{x}, \bar{y}^l, \bar{\gamma}^l) = 0. \tag{3.7}$$

where $\mathcal{L}(\bar{x}, y, \gamma) = g(\bar{x}, y) - \sum_{j=1}^{\bar{q}} \gamma_j v_j(\bar{x}, y)$, the Lagrangian function where $\bar{\gamma}^l > 0$.

Recall from Subsection 2.2.2, (Q-III) that a sufficient second order condition is given by:

$$\begin{aligned} &\eta^T \nabla_y^2 \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i)\eta < 0 \ \text{ for all } \eta \in G_{Q(\bar{x})} \setminus \{0\}, \text{where} \\ &G_{Q(\bar{x})} = \{\eta \in \mathbb{R}^m \mid D_y v_j(\bar{x}, \bar{y}^i)\eta = 0,\ j \in Q_0(\bar{x}, \bar{y}^i)\}. \end{aligned} \tag{3.8}$$

Assume that ELICQ (2.13) is fulfilled at $\bar{x} \in M$, as well as the KKT condition (3.7) holds and the second order condition,

$$\xi^T \bar{M}_0 \xi > 0 \ \text{ for all } \ \xi \in T \setminus \{0\}, \tag{3.9}$$

with the tangent space $T = \{\xi \in \mathbb{R}^n \mid D_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i)\xi = 0, i = 1, \ldots, p\}$ and

$$\begin{aligned} \bar{M}_0 := D^2 f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i D_x^2 \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) + \sum_{i=1}^{p} \bar{\mu}_i D^T y^i(\bar{x}) D_y^2 L(\bar{x}, \bar{y}^i, \bar{\gamma}^i) D y^i(\bar{x}) \\ - \sum_{i=1}^{p} \bar{\mu}_i \sum_{j=1}^{\bar{q}} \left(D^T \gamma_j^i(\bar{x}) D_x v_j(\bar{x}, \bar{y}^i) + D_x^T v_j(\bar{x}, \bar{y}^i) D \gamma_j^i(\bar{x}) \right). \end{aligned} \tag{3.10}$$

Compare $\bar{M}_0$ by $\nabla_x^2 L(\bar{x}, \bar{y}, \bar{\mu}, \bar{\gamma})$ stated in GSIP-II in Subsection 5.1.1.

Then, $\bar{x}$ is a local minimizer of *GSIP*. Under Reduction Ansatz in Theorem 2.2.10 (see Subsection 2.2.2), the C^1 functions $y^i(x)$ and $\gamma^i(x)$ are obtained. Consider $\bar{x} \in M$ such that at any point in $Y_0(\bar{x})$ the conditions (2.9), (3.8) are satisfied, also let (2.13) and (3.9) be fulfilled, i.e., first and second order conditions holds for both lower level problem and reduced upper level reduced problem. Then, necessarily $\bar{x}, \bar{\mu}, \bar{y}^i, \bar{\gamma}^i$ $(i = 1, \ldots, p)$ is a solution of the following system of KKT equations of *GSIP* and the lower level problem $Q(\bar{x})$ (compare with (4.35) in Subsection 4.3.3):

$$\begin{aligned} \nabla f(x) + \sum_{i=1}^{p} \mu_i \nabla_x \mathcal{L}(x, y^i, \gamma^i) = 0, \\ g(x, y^i) - \sum_{j=1}^{\bar{q}} \gamma_j^i v_j(x, y^i) = 0 \ \ (i = 1, 2, \ldots, p) \end{aligned} \tag{3.11}$$

and for $i = 1, \ldots, p$

$$\begin{aligned} \nabla_y g(x, y^i) - \sum_{j=1}^{\bar{q}} \gamma_j^i \nabla_y v_j(x, y^i) = 0, \\ v_j(x, y^i) = 0 \ \ (j \in Q_0(\bar{x}, \bar{y}^i)). \end{aligned}$$

The system (3.11) consists of $K = n + p + \sum_{i=1}^{p}(m + \bar{q})$ equations with equally many unknowns $x \in \mathbb{R}^n$, $\mu_i \in \mathbb{R}$, $y^i \in \mathbb{R}^m$, $\gamma^i \in \mathbb{R}^{\bar{q}}$. Since, the number of equations and unknowns is equal, Newton-type methods can be used. For *SIP* problems, since v_j does not depend on x, in the first equation of 3.11, the sum over $\gamma_i^i D_x v_j(x, y^i)$ vanishes (appearing in $\nabla_x \mathcal{L}(x, y^i, \gamma^i)$).

The following lemma shows that, under these assumptions, the Jacobian of the system (3.11) is nonsingular at the solution. This, in particular, implies that the Newton method applied to (3.11) will locally converge q-quadratically.

Lemma 3.3.1 ([112]) *Let $\bar{x} \in M$ be given such that at any point $\bar{y}^i \in Y_0(\bar{x}),\ i = 1, \ldots, p$ the conditions LICQ (2.9) and second order sufficiency condition for lower level problem (3.8) are satisfied and let ELICQ (2.13) and second order condition for reduced upper level problem (3.9) be fulfilled. Then, the Jacobian of (3.11) at $\bar{x}, \bar{\mu}, \bar{y}^i, \bar{\gamma}^i,\ i = 1, \ldots, p$, is nonsingular.*

Proof. We refer to [112].

For *SIP* a globally convergent algorithm from finitely constrained programming to the locally reduced problems SIP_{red} is described in [33], Algorithm 7.4, which can directly be generalized to *GSIP*.

CHAPTER 4

SEMISMOOTH NEWTON METHOD

4.1 Preliminaries on the Semismooth Newton Approach

For this section we refer the reader to our recent papers [108, 115]. For a locally Lipschitzian vector valued function $F : \mathbb{R}^n \to \mathbb{R}^m$ let $\partial F(x)$ denote Clarke's generalized Jacobian at x [6]. Let us write F in terms of component functions as $F(x) = (f^1(x), f^2(x), \ldots, f^m(x))^T$. We assume that each f^i (and, hence, F) is Lipschitz near a given point x of interest. Rademacher's Theorem asserts that F is differentiable (i.e., each f^i is differentiable) almost everywhere on any neighbourhood of x in which F is Lipschitz. Let us denote the set of points in $\mathbb{R}^n$ at which F fails to be differentiable by Ω_F. We shall write $JF(y)$ for the usual $(m \times n)$ Jacobian matrix of partial derivatives whenever y is a point at which the necessary partial derivatives exist.

Definition 4.1.1 ([6]) *The generalized Jacobian of F at x, denoted by $\partial F(x)$, is the convex hull of all $(m \times n)$ matrices Z obtained as the limit of a sequence of the form $JF(x_i)$, where $x_i \to x$ and $x_i \notin \Omega_F$*

$$\partial F(x) := co\{\lim_{i \to \infty} JF(x_i) \mid x_i \to x \ (i \to \infty), x_i \notin \Omega_F\}. \tag{4.1}$$

If F is continuously differentiable then $\partial F(x) = \{JF(x)\}$.

4.1.1 Semismooth functions

Definition 4.1.2 ([87]) *A locally Lipschitzian vector valued function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called semismooth at $x \in \mathbb{R}^n$ if F is directionally differentiable at x and if for all $V \in \partial F(x + d)$ and*

$d \to 0$ *we have*

$$F'(x; d) = Vd + o(\|d\|).$$

In Definition 4.1.2, $\|\ \|$ denotes Euclidean distance. If F is continuously differentiable at x, then it is semismooth at x. In some sense, semismoothness is equivalent to the uniform convergence of directional derivatives in all directions [87]. Semismoothness was originally introduced by Mifflin for functionals [72]. In [87], the definition of semismooth functions was extended to $F : \mathbb{R}^n \to \mathbb{R}^m$. It was proved that F is semismooth at x if and only if all its component functions are semismooth. Scalar products and sums of semismooth functions are still semismooth functions [72].

The importance of semismooth equations is that although the mapping is in general nonsmooth, Newton method is still applicable and converges *q-superlinearly* to a regular solution. The rate of convergence of the semismooth Newton method can be improved if an estimate of higher order is available. This leads to the following definition of higher order semismoothness.

Definition 4.1.3 ([87]) *A locally Lipschitzian vector valued function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called p-order semismooth or at x if F is semismooth at x and if for all $V \in \partial F(x + d)$ and $d \to 0$ we have*

$$Vd - F'(x; d) = O(\|d\|^{p+1}),$$

where $0 < p \leq 1$. In particular if $p = 1$, F is called 1-order semismooth.

1-order semismoothness is renamed as strongly semismoothness in [85]. In this thesis, we also use strongly semismoothness for 1-order semismooth functions. For other definitions and properties of semismoothness we refer to [87, 116].

Examples of Semismooth Functions ([116])

Example 1- The Euclidean Norm

The Euclidean norm $f : \mathbb{R}^n \to \mathbb{R}$, defined by

$$f(x) = \|x\|_2 = (x^T x)^{1/2}$$

is an important example of a strongly semismooth function that arises, e.g., as the nonsmooth part of the Fischer Burmeister function which is defined in Example 2.
Obviously, f is Lipschitz continuous on $\mathbb{R}^n$, and C^1 on $\mathbb{R}^n \setminus \{0\}$ with

$$f'(x) = \frac{1}{2}\frac{x^T}{\|x\|_2}.$$

Therefore

$$\partial f(x) = \left\{ \frac{x^T}{\|x\|_2} \right\} \text{ for } x \neq 0,$$
$$\partial f(0) = \{v^T \mid v \in \mathbb{R}^n, \|v\|_2 \leq 1\}.$$

f is strongly semismooth at $x = 0$, so it is strongly semismooth on $\mathbb{R}^n$.

Example 2- Fischer Burmeister and Min functions

Definition 4.1.4 ([60]) *A scalar valued function $\psi : \mathbb{R}^2 \to \mathbb{R}$ is called an NCP-function if*

$$\psi(a, b) = 0 \quad \textit{if and only if} \quad a \geq 0, \;\; b \geq 0 \;\; \textit{and} \;\; ab = 0.$$

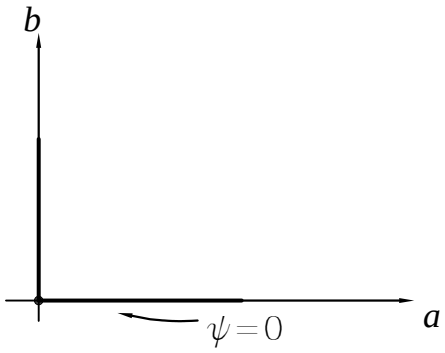

Figure 4.1: The zero set of an NCP function, i.e., $\psi(a, b)$.

Here, NCP stands for *nonlinear complementarity problem*. An important example of NCP-function is the *Fischer-Burmeister* function

$$\psi_{FB}(a, b) = \sqrt{a^2 + b^2} - a - b. \tag{4.2}$$

The Fischer Burmeister Function was introduced by Fischer [14]. Many modern algorithms for finite dimensional NCPs are based on reformulations by means of the Fischer Burmeister

NCP function. This function is Lipschitz continuous and strongly semismooth on $\mathbb{R}^2$. Further, ψ_{FB} is C^1 on $\mathbb{R}^2 \setminus \{0\}$, and $(\psi_{FB})^2$ is continuously differentiable on $\mathbb{R}^2$. ψ_{FB} is the difference of the strongly semismooth function $\|x\|_2$ and linear function $h(x) = x_1 + x_2$. Therefore, ψ_{FB} is Lipschitz continuous and strongly semismooth on $\mathbb{R}^2$. Another important example of an NCP-function is the *min* function

$$\psi_{min}(a, b) = -\min\{a, b\}. \tag{4.3}$$

ψ_{min} is also strongly semismooth on $\mathbb{R}^2$.

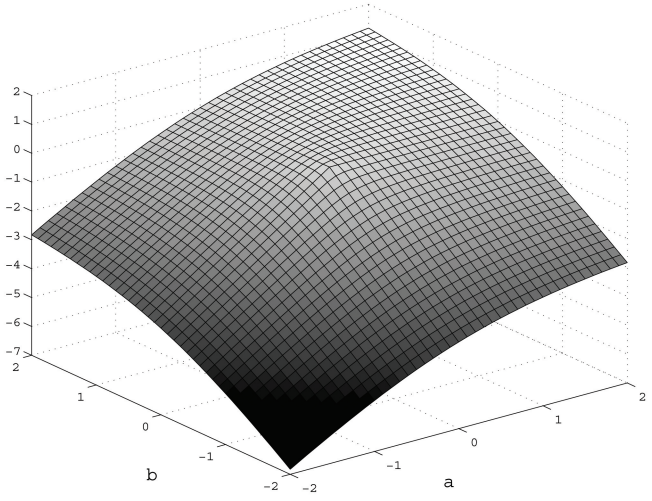

Figure 4.2: The graph of minus Fischer Burmeister function, i.e., $-\psi_{FB}$.

We use both Fischer Burmeister NCP-function and min NCP function in this thesis. See Appendix A for subdifferentials of the convex functions ψ_{FB} and ψ_{min}.

Example 3- Piecewise Differentiable Functions

Piecewise continuously differentiable functions are an important subclass of semismooth functions.

Definition 4.1.5 ([100]) *A function $f : V \to \mathbb{R}^m$ defined on the open set $V \subset \mathbb{R}^n$ is called PC^k–function (piecewise C^k function), $1 \leq k \leq \infty$, if f is continuous and if at every point $x_0 \in V$ there exist a neighborhood $W \subset V$ of x_0 and a finite collection of C^k–functions $f^i : W \to \mathbb{R}^m, i = 1, \ldots, N$ such that*

$$f(x) \in \{f^1(x), \ldots, f^N(x)\} \text{ for all } x \in W.$$

f is called a continuous selection of $\{f^1, \ldots, f^N\}$ on W. The class of PC^k functions is closed under composition, finite summation and multiplication. The functions $f_1 : \mathbb{R} \to \mathbb{R}$, $f_1(x) = |x|$, $f_2 : \mathbb{R}^2 \to \mathbb{R}$, $f_2(x) = \max\{x_1, x_2\}$ and $f_3 : \mathbb{R}^2 \to \mathbb{R}$, $f_3(x) = \min\{x_1, x_2\}$ are $PC^\infty-$ functions.

Proposition 4.1.6 ([116]) *Let $f : V \to \mathbb{R}^m$ be a PC^1- function on the open set $V \subset \mathbb{R}^n$. Then f is semismooth. If f is a PC^2-function, then f is strongly semismooth.*

We refer to [100, 116] for more information and properties of piecewise differentiable functions.

4.1.2 Semismooth Newton method

In analogy to the standard Newton method, the basic iteration of the semismooth Newton approach for solving the equation $F(z) = 0$ is [87]

$$z^{k+1} = z^k - (W^k)^{-1} F(z^k) \tag{4.4}$$

with $W^k \in \partial F(z^k)$.

To study convergence properties of the semismooth Newton method, the concept of CD regularity was introduced. Here, CD stands for the Clarke subdifferential [85].

Let vector valued function $F : \mathbb{R}^n \to \mathbb{R}^n$ be semismooth. Then F is called *CD-regular* at a point $\bar{x}$, if all matrices in $\partial F(\bar{x})$ are nonsingular [85].

Proposition 4.1.7 ([87]) *If all $V \in \partial F(x)$ are nonsingular, then there is a neighbourhood $N(x)$ of x and a constant C such that for any $y \in N(x)$ and any $V \in \partial F(y)$, V is nonsingular and*

$$\left\|V^{-1}\right\| \leq C.$$

Proof. If the conclusion is not true, then there is a sequence $y^k \to x\,(k \to \infty)$, $V_k \in \partial F(y^k)$, such that either all V_k are singular or $\left\|V_k^{-1}\right\| \to \infty\,(k \to \infty)$. Since F is locally Lipschitzian, ∂F is bounded in a neighbourhood of x. By passing to a subsequence, we may assume that

$V_k \to V\,(k \to \infty)$. Then V must be singular, a contradiction to the assumption of nonsingularity of $V \in \partial F(x)$. This completes the proof. •

In the following we give the definition of convergence rate.

Definition 4.1.8 ([56, 77]) *Suppose that the sequence $\{z_k\}$ converges to the number ξ. We say that this sequence converges linearly to ξ, if there exists a number $\mu \in [0, 1]$ such that*

$$\lim_{k\to\infty} \frac{|z_{k+1} - \xi|}{|z_k - \xi|} = \mu.$$

The number μ is called the rate of convergence. If the above holds with $\mu = 0$, then the sequence is said to converge q-superlinearly. The next definition is used to distinguish superlinear rates of convergence.

Definition 4.1.9 ([56, 57]) *We say that the sequence converges with order q for $q > 1$ to ξ if*

$$\lim_{k\to\infty} \frac{|z_{k+1} - \xi|}{|z_k - \xi|^q} = \mu \text{ with } \mu > 0.$$

In particular, convergence with order 2 is called q-quadratic convergence. The q stands for quotient, because the definition uses the quotient between two successive terms. In numerical analysis, usually the term quadratic convergence is used for q-quadratic convergence. There is also a slightly weaker form of convergence, characterized by the prefix R (for root). It is concerned with the overall rate of decrease in the error, rather than the decrease over a single step of the algorithm, see [77] for R-quadratic convergence. However, most convergence analyses of optimization algorithms are concerned with q-convergence.

Theorem 4.1.10 ([87]) *Suppose that $\bar{x}$ is a solution of $F(x) = 0$, and F is semismooth and CD-regular at $\bar{x}$. Then the iteration method (4.4) is well defined and $\{x^k\}$, the sequence generated by (4.4), converges to $\bar{x}$ q-superlinearly in a neighborhood of $\bar{x}$. If in addition F is p-order semismooth at $\bar{x}$, then the convergence is of order $1 + p$. In particular, if F is strongly semismooth at $\bar{x}$, then the convergence is q-quadratic.*

Proof. By Proposition 4.1.7, (4.4) is well-defined in a neighbourhood of $\bar{x}$ for the first step

$k = 0$. Now

$$\begin{aligned} \left\|x^{k+1} - \bar{x}\right\| &= \left\|x^k - V_k^{-1}F(x^k) - \bar{x}\right\| \\ &= \left\|x^k - \bar{x} - V_k^{-1}F(x^k)\right\| \\ &\leq \left\|V_k^{-1}[F(x^k) - F(\bar{x}) - F'(\bar{x}; x^k - \bar{x})]\right\| \\ &+ \left\|V_k^{-1}[V_k(x^k - \bar{x}) - F'(\bar{x}; x^k - \bar{x})]\right\| \\ &= o(\left\|x^k - \bar{x}\right\|). \end{aligned} \tag{4.5}$$

This completes the proof. •

4.2 Semismooth Newton Method for Finitely Constrained Programming

In this section, a review of nonsmooth Newton method applied to the problem of finitely constrained programming as in [85], and also the convergence properties are investigated.

Let us consider, the finitely constrained optimization problem (2.3) from Section 2.1, i.e.,

$$P : \min_x f(x) \quad \text{subject to} \quad g_i(x) \leq 0 \ \ (i \in I), \quad h_j(x) = 0 \ \ (j \in J), \tag{4.6}$$

where f, g_i, h_j are continuously differentiable functions with $I = \{1, \ldots, r\}$, $J = \{1, \ldots, s\}$. Let $N = n + r + s$. Let $g(x) = (g_1(x), \ldots, g_r(x))^T$, and $h(x) = (h_1(x), \ldots, h_s(x))^T$. As given in Section 2.1, the KKT system for this problem is:

$$\begin{gathered} \nabla f(x) + \sum_{j=1}^{r} \mu_j \nabla g_j(x) + \sum_{j=1}^{s} \lambda_j \nabla h_j(x) = 0 \\ \mu \geq 0, \ \ g(x) \leq 0, \\ \mu^T g(x) = 0, \\ h(x) = 0. \end{gathered} \tag{4.7}$$

We denote $z^T = (x^T, \mu^T, \lambda^T)$. The KKT system plays a central role in the theory and algorithms for problems of nonlinear programming. In this section, we assume f, g and h are twice continuously differentiable and $D^2 f, D^2 g$ and $D^2 h$ are locally Lipschitzian.

Many iterative methods have been developed to solve KKT systems. We refer to [15] for a comphrehensive treatment of these methods. In [85] an approach was developed to construct generalized Newton method for solving these nonsmooth KKT equations.

(4.7) is reformulated by using nonlinear complementarity problem functions (NCP-functions) given in Definition 4.1.4. By using NCP-functions, the system becomes

$$T(z) = \begin{pmatrix} \nabla f(x) + \sum_{j=1}^{r} \mu_j \nabla g_j(x) + \sum_{j=1}^{s} \lambda_j \nabla h_j(x) \\ \psi(\mu, -g(x)) \\ h(x) \end{pmatrix} \tag{4.8}$$

Two typical versions of KKT equations are used, by using the NCP-functions *Fisher Burmeister function* given in (4.2) and *min function* given in (4.3), which are equivalent to (4.7). These versions of KKT equations are strongly semismooth. As in Section 2.1, we denote the Lagrangian of (4.6) by

$$L(x, \mu, \lambda) = f(x) + \mu^T g(x) + \lambda^T h(x)$$

and denote its gradient with respect to x by

$$\nabla_x L(x, \mu, \lambda).$$

By using *min* NCP-function, the KKT system becomes

$$T_1(z) = \begin{pmatrix} \nabla f(x) + \sum_{j=1}^{r} \mu_j \nabla g_j(x) + \sum_{j=1}^{s} \lambda_j \nabla h_j(x) \\ -\min\{\mu, -g(x)\} \\ h(x) \end{pmatrix}. \tag{4.9}$$

By using *Fischer-Burmeister* NCP-function, the KKT system becomes

$$T_2(z) = \begin{pmatrix} \nabla f(x) + \sum_{j=1}^{r} \mu_j \nabla g_j(x) + \sum_{j=1}^{s} \lambda_j \nabla h_j(x) \\ \sqrt{\mu_1^2 + g_1(x)^2} + g_1(x) - \mu_1 \\ \vdots \\ \sqrt{\mu_r^2 + g_r(x)^2} + g_r(x) - \mu_r \\ h(x) \end{pmatrix}. \tag{4.10}$$

In Subsection 4.1.1, it is mentioned that Fischer-Burmeister NCP function and min NCP function is strongly semismooth. Here a main result based on this is given.

Theorem 4.2.1 ([85]) *Both T_1 and T_2 defined by (4.9) and (4.10) are strongly semismooth.*

Proof. We refer to [85].

Now, we discuss about convergence of the nonsmooth version of Newton method (4.4), i.e., see Section 4.1, to solve the equations $T_1(z) = 0$ and $T_2(z) = 0$. By Theorem 4.2.1 we see that T_1 and T_2 are strongly semismooth functions.

Hence, the key thing is to identify the conditions such that all elements in the generalized Jacobians at the solution point $\bar{z}$ of T_1 and T_2 are nonsingular. This is equivalent to the condition both T_1 and T_2 are CD-regular (see Section 4.1) at a KKT-point z^*.

Definition 4.2.2 (implied by [85]) *Suppose that* $z^T = (x^T, \mu^T, \lambda^T) \in \mathbb{R}^N$. *Let* $R = \{1, 2, \ldots, r\}$,

$$I(z) = \{j \mid j \in R, g_j(x) = 0\},$$

$$I_0(z) = \{j \in I(z) \mid \mu_j = 0\},$$

$$I_1(z) = \{j \in I(z) \mid \mu_j > 0\},$$

and

$$G(z) := \{d \in \mathbb{R}^n \mid Dg_j(x)d = 0 \ \ (j \in I_1(z)),\ Dh(x)d = 0\}.$$

A point $z \in \mathbb{R}^N$ *is said to satisfy the strong second-order sufficiency condition for (4.6) if it satisfies the first-order KKT condition (4.7) and if* $d^T V d > 0$ *for all* $d \in G(z) \setminus \{0\}$ *and* $V = \nabla^2_x L(x, \mu, \lambda)$. *We say that a KKT point z of (4.6) satisfies the linear independence condition if* $\{\nabla g_j(x) \ \ (j \in I(z)), \nabla h_j(x) \ \ (j \in J)\}$ *are linearly independent.*

A point $z \in \mathbb{R}^N$ *is said to satisfy the Robinson condition if it satisfies both the linear indepence condition and the strong second-order sufficiency condition.*

Note that the strong second order sufficiency condition implies that x is a strict local minimum of (4.6). We will reformulate and then use Robinson condition (Definition 4.2.2) for $GSIP_{red}$ in Section 5.1.

Theorem 4.2.3 ([85]) *Suppose that* $\bar{z} \in \mathbb{R}^N$ *is a KKT point of (4.6) and satisfies the Robinson condition. Then, both* T_1 *and* T_2 *are CD-regular at* $\bar{z}$.

Proof. Let $W \in \partial T_1(\bar{z})$. Then,

$$W = \begin{pmatrix} V & Dg_{I_1}(\bar{x}) & Dg_{I_0}(\bar{x}) & Dg_J(\bar{x}) & Dh(\bar{x}) \\ \nabla g_{I_1}(\bar{x}) & 0 & 0 & 0 & 0 \\ \Lambda_0 Dg_{I_0}(\bar{x}) & 0 & -\Gamma_0 & 0 & 0 \\ 0 & 0 & 0 & -E_J & 0 \\ \nabla h(\bar{x}) & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{4.11}$$

where $V = \nabla_x^2 L(\bar{x}, \bar{\mu}, \bar{\lambda}), I_1(\bar{z}) \subseteq I_1 \subseteq I(\bar{z}), I_0 \subseteq (I_0(\bar{z}) \setminus I_1), J = R \setminus (I_1 \cup I_0), E_J$ is the identity matrix of dimension $|J|$, Λ_0 and Γ_0 are negative definite diagonal matrices of dimension $|I_0|$ and diagonal elements $\lambda_j \in [0, 1]$ and $\gamma_j = 1 - \lambda_j$, respectively, for $j \in I_0$, the order of $j \in R$ is reordered to separate I_1 and I_0 and J. Suppose that

$$W \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} = 0, \tag{4.12}$$

where $d_1 \in \mathbb{R}^n, d_2 \in \mathbb{R}^{|I_1|}, d_3 \in \mathbb{R}^{|I_0|}, d_4 \in \mathbb{R}^{|J|}$. Note that the jth element of d_3 is equal to 0 by (4.12) if $\lambda_j = 0$. Without loss of generality, we may assume that $\lambda_j \in (0, 1]$. Then (4.12) implies $d_4 = 0$,

$$Vd_1 + Dg_{I_1}(\bar{x})d_2 + Dg_{I_0}(\bar{x})d_3 + Dh(\bar{x})d_5 = 0, \tag{4.13}$$

$$\nabla g_{I_1}(\bar{x})d_1 = 0, \tag{4.14}$$

$$\nabla g_{I_0}(\bar{x})d_1 = \Lambda_0^{-1}\Gamma_0 d_3. \tag{4.15}$$

and

$$\nabla h(\bar{x})d_1 = 0, \tag{4.16}$$

Suppose that d_1 satisfies (4.14). Then $d_1 \in G(\bar{z})$. Multiplying (4.13) with d_1^T, by (4.14) and (4.15), we have

$$d_1^T V d_1 + d_3^T \Gamma_0 \Lambda_0^{-1} d_3 = 0.$$

By the strong second-order sufficiency condition and positive definiteness of $\Gamma_0\Lambda_0^{-1}, d_1 = 0$ and $d_3 = 0$. Now, (4.13) yields

$$Dg_{I_1}(\bar{x})d_2 + Dh(\bar{x})d_5 = 0.$$

By the linear independence condition, $d_2 = 0$ and $d_5 = 0$. Hence, $d = 0$. This shows that W is nonsingular. Therefore, T_1 is CD-regular at $\bar{z}$.

For any $z \in \mathbb{R}^N$, T_2 is differentiable at z if and only if $\nabla_x^2 L(x, \mu, \lambda)$ exists and $\mu_j^2 + g_j(x)^2 > 0$ for all $j \in R$. For these points z,

$$\nabla T_2(z) = \begin{pmatrix} \nabla_x^2 L(x, \mu, \lambda) & Dg(x) & Dh(x) \\ \Lambda \nabla g(x) & \Gamma & 0 \\ \nabla h(x) & 0 & 0 \end{pmatrix}, \tag{4.17}$$

where $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_r\}, \Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_r\}$,

$$\lambda_j = \frac{g_j(x)}{\sqrt{\mu_j^2 + g_j(x)^2}} + 1 \tag{4.18}$$

and

$$\gamma_j = \frac{g_j(x)}{\sqrt{\mu_j^2 + g_j(x)^2}} - 1 \tag{4.19}$$

for $j \in R$. By (4.18) and (4.19), we have

$$(\lambda_j - 1)^2 + (\gamma_j + 1)^2 = 1. \tag{4.20}$$

By (4.17), (4.20) and the definition of the generalized Jacobian of T_2, if $W \in \partial T_2(\overline{z})$, we have,

$$W = \begin{pmatrix} V & Dg(\overline{x}) & Dh(\overline{x}) \\ \Lambda \nabla g(\overline{x}) & \Gamma & 0 \\ \nabla h(\overline{x}) & 0 & 0 \end{pmatrix}. \tag{4.21}$$

where $V = \nabla_x^2 L(\overline{x}, \overline{\mu}, \overline{\lambda})$ and again $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_r\}, \Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_r\}$, and

$$(\lambda_j - 1)^2 + (\gamma_j + 1)^2 \leq 1. \tag{4.22}$$

Suppose that

$$W \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = 0, \tag{4.23}$$

where $d_1 \in \mathbb{R}^n, d_2 \in \mathbb{R}^r$. Use d_{2_j} to denote the components of d_2. Then, (4.23) implies

$$Vd_1 + Dg(\overline{x})d_2 + Dh(\overline{x})d_3 = 0, \tag{4.24}$$

$$\lambda_j \nabla g_j(\overline{x})d_1 + \gamma_j d_{2_j} = 0, \tag{4.25}$$

for $j \in R$ and

$$\nabla h(\overline{x})d_1 = 0, \tag{4.26}$$

Let $R_1 := \{j \in R, \lambda_j > 0, \gamma_j < 0\}, R_2 := \{j \in R, \lambda_j = 0\}$ and $R_3 := \{j \in R, \gamma_j = 0\}$. Then $I_1(\overline{z}) \subseteq R_3$. By (4.25) and (4.22), $d_{2_j} = 0$ if $j \in R_2$,

$$\nabla g_j(\overline{x})d_1 = 0 \tag{4.27}$$

if $j \in R_3$, and

$$\nabla g_j(\bar{x})d_1 = \upsilon_j d_{2_j}, \tag{4.28}$$

where $\upsilon_j = -\gamma_j/\lambda_j > 0$ if $j \in R_1$. Multiplying (4.24) with d_1^T, by (4.27) and (4.28),

$$d_1^T V d_1 + \sum_{j \in R_1} \upsilon_j d_{2_j}^2 = 0.$$

Since $I_1(\bar{z}) \subseteq R_3$, by (4.27), $d_1 \in G(\bar{z})$. Since $\upsilon_j > 0$ for $j \in R_1$, by the strong second-order sufficiency condition, $d_1 = 0$ and $d_{2_j} = 0$ for $j \in R_1$. Now, (4.24) yields

$$\sum_{j \in R_3} Dg_j(\bar{x})^T d_{2_j} = 0.$$

Notice $R_3 \subseteq I(\bar{z})$. By linear independence condition, $d_{2_j} = 0$ for $j \in R_3$. Hence, $d = 0$. This shows that W is nonsingular. Therefore, T_2 is CD-regular at $\bar{z}$. This completes the proof. •

By Theorems 4.1.10, 4.2.1 and 4.2.3, we have the following theorem which shows that semismooth Newton method is q-quadratically convergent if Robinson condition holds at the solution point.

Theorem 4.2.4 ([85]) *Let $T = T_1$ or $T = T_2$. Consider the semismooth Newton method (4.4). Suppose that $\bar{z}$ is a solution of (4.7) and satisfies Robinson condition. Then the iterative method (4.4) is well-defined and $\{z^k\}$, the sequence generated by (4.4), converges to $\bar{z}$ q-quadratically in a neighbourhood of $\bar{z}$.*

Note that in [60] it is shown that, for finite optimization problems with C^2 data, CD-regularity of the Kojima formulation of the KKT system is equivalent to the strong stability of a KKT point. From another point of view, in [85] it is shown that the Robinson condition (Definition 4.2.2) at the KKT point of a C^2 problem implies CD-regularity of the KKT system, in which the complementarity conditions are reformulated using the *min* or *Fischer Burmeister* NCP functions. In Subsection 5.1.1 we will show an analogous result for *GSIP*, involving weaker smoothness assumptions.

We refer the reader to [85] some other semismooth KKT equations.

4.3 Semismooth Optimality Conditions in Semi-infinite Programming

In this subsection we recall the results obtained in our recent paper [108]. In [88], recently the semismooth optimality conditions for standard *SIP* was suggested. In this section, we recall

semismooth opimality conditions for standard *SIP* in Subsection 4.3.2 and we derive them for *GSIP* in Subsection 4.3.3.

4.3.1 The bilevel structure of semi-infinite programming

As we stated in our recent paper [108], the theoretical and numerical treatment of *GSIP* is closely related to the bilevel structure of semi-infinite programming. In fact, under our assumptions the semi-infinite constraint in *GSIP* is equivalent to

$$\varphi(x) \;=\; \max_{y \in Y(x)} g(x,y) \;\leq\; 0.$$

The feasible set M of *GSIP* is the lower level set of some optimal value function:

$$M \;=\; \{\, x \in \mathbb{R}^n \mid \varphi(x) \leq 0 \,\}.$$

An example of feasible set defined by the optimal value function is illustrated in Figure 4.3. The function φ is the optimal value function of the lower level problem defined in (2.8)

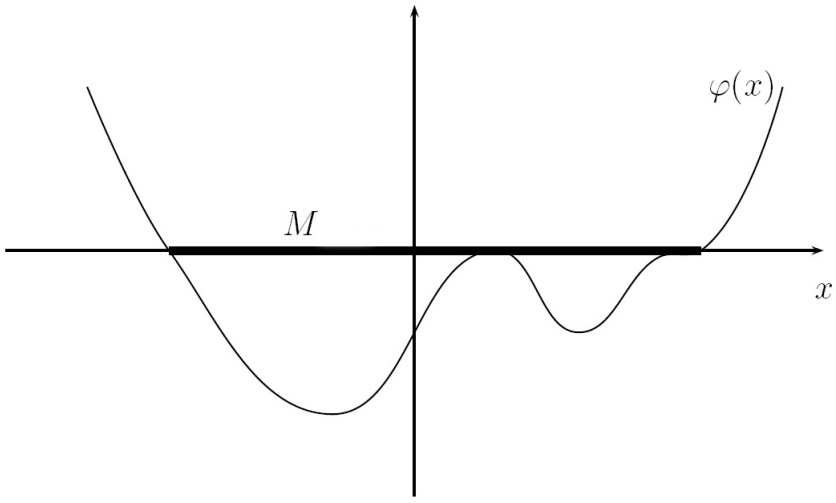

Figure 4.3: The feasible set defined by optimal value function [128].

$$Q(x): \quad \max_{y \in \mathbb{R}^m} g(x,y) \quad \text{subject to} \quad v_j(x,y) \;\leq\; 0 \;\; (j \in Q).$$

In the upper level problem the aim is to minimize f over M where x is the decision variable, however, in the lower level problem x plays the role of an $n-$dimensional parameter, and y is the decision variable.

In semi-infinite programming the main numerical problem is the following: the lower level problem has to be solved to *global optimality*, even if we want to find a stationary point of the upper level problem. In fact, standard nonlinear programming (NLP) solvers can only be expected to produce a *local* maximizer y_{loc} of $Q(\bar{x})$ which is not necessarily a global maximizer y_{glob}. Even if $g(\bar{x}, y_{\text{loc}}) \leq 0$ holds, $\bar{x}$ might be infeasible since $g(\bar{x}, y_{\text{loc}}) \leq 0 < \varphi(\bar{x}) = g(\bar{x}, y_{\text{glob}})$ cannot be ruled out in general.

Since, in the following, we aim to use the approach from [88] and replace the lower level problem by its KKT conditions, we must make sure that a solution of the KKT system is a global maximizer. We emphasize that otherwise one might compute *infeasible points* for the semi-infinite problem, which is a major pitfall of the approach at hand. In particular, the concept of *substationary points* from [88] may entail infeasibility.

A natural assumption under which a solution of the KKT conditions leads to a global maximizer is the convexity of the lower level problem, that is, for each $x \in \mathbb{R}^n$ the function $g(x, \cdot)$ is concave, and the set $Y(x)$ is convex. We thus make the following assumption throughout the remainder of the thesis.

Assumption 4.3.1 ([108, 115]) *For all $x \in \mathbb{R}^n$ the lower level problem $Q(x)$ is convex.*

In *GSIP* many relevant applications have convex lower level problems [103] (see also Chapter 6). On the other hand, in standard semi-infinite optimization this situation is rather rare. For recent solution approaches to standard *SIP* with nonconvex lower level problems, we refer the reader to [16, 76].

In the remainder of the thesis, we will not use Assumption 4.3.1 explicitly, but develop the theory in the more general setting without convex lower level problems. In particular we will deal with finitely many lower level maximizers, instead of a unique one in the convex regular case. However, the slightly more general setting of finitely many generalized semi-infinite constraints $g_i(x, y) \leq 0,\ y \in Y(x),\ i \in I$, each with a convex lower level problem, would lead to almost identical formulas as the ones we develop in the sequel (with g replaced by g_i).

4.3.2 Semismooth optimality conditions for *SIP*

Let us first consider the standard semi-infinite case. As mentioned in Section 2.2 a standard semi-infinite problem has the form

$$SIP: \qquad \text{minimize} \quad f(x) \quad \text{subject to} \quad x \in M$$

with

$$M \;=\; \{\, x \in \mathbb{R}^n \,|\, g(x,y) \le 0 \text{ for all } y \in Y \,\}$$

and

$$Y \;=\; \{\, y \in \mathbb{R}^m \,|\, v_j(y) \le 0 \;\; (j \in Q) \,\}.$$

To formulate first and second order optimality conditions, we use the following notation. For a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, we denote the gradient in row form by $Df(x) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]$ and in column form by $\nabla f(x) = D^T f(x)$. For a continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^r$ we denote the Jacobian of F at $x \in \mathbb{R}^n$ by $DF(x)$ whereas the transposed Jacobian is $\nabla F(x)$. For a function $g : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ we denote by $\nabla_x g(x,y)$ the gradient of g at (x,y) with respect to x and by $\nabla^2_{xx} g(x,y)$, $\nabla^2_{xy} g(x,y) = D_y \nabla_x g(x,y)$ and $\nabla^2_{yy} g(x,y)$, the respective $n \times n$, $n \times r$ and $r \times r$ matrices of second order partial derivatives of g at (x,y).

Recall that by Theorem 2.2.13 from Subsection 2.2.3, for a local minimizer under EMFCQ (2.14), the following first order optimality condition holds,

$$\nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x g(\bar{x}, \bar{y}^i) = 0. \tag{4.29}$$

Next we complement the upper level first order condition from Theorem 2.2.13 by a lower level first order condition. In fact, since the active indices $\bar{y}^i \in Y_0(\bar{x})$, $i \in P$, are global solutions of $Q(\bar{x})$, under some constraint qualification like Slater's condition
(see Definition 2.1.12) in the lower level problem (2.8), there exist vectors of Lagrange multipliers $\bar{\gamma}^i \in \mathbb{R}^q$ such that

$$\begin{aligned} &\nabla_y g(\bar{x}, \bar{y}^i) - \sum_{j=1}^{q} \bar{\gamma}^i_j \nabla_y v_j(\bar{y}^i) = 0 \;\; (i \in P), \\ &\bar{\gamma}^i_j \ge 0, \quad v_j(\bar{y}^i) \le 0, \quad \bar{\gamma}^i_j v_j(\bar{y}^i) = 0 \;\; (i \in P, j \in Q). \end{aligned} \tag{4.30}$$

By (4.29) and (4.30), we arrive at the following equalities and inequalities:

$$\nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x g(\bar{x}, \bar{y}^i) = 0,$$

$$\bar{\mu}_i \geq 0, \;\; g(\bar{x}, \bar{y}^i) = 0 \;\; (i \in P),$$

$$\nabla_y g(\bar{x}, \bar{y}^i) - \sum_{j=1}^{q} \bar{\gamma}_j^i \nabla_y v_j(\bar{y}^i) = 0 \;\; (i \in P), \tag{4.31}$$

$$\bar{\gamma}_j^i \geq 0, v_j(\bar{y}^i) \leq 0, \; \bar{\gamma}_j^i v_j(\bar{y}^i) = 0 \;\; (i \in P, j \in Q).$$

With any NCP function ψ (4.31) can be reformulated as the following system of equations:

$$\nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x g(\bar{x}, \bar{y}^i) = 0,$$

$$\psi(\bar{\mu}_i, -g(\bar{x}, \bar{y}^i)) = 0 \;\; (i \in P),$$

$$\nabla_y g(\bar{x}, \bar{y}^i) - \sum_{j=1}^{q} \bar{\gamma}_j^i \nabla_y v_j(\bar{y}^i) = 0 \;\; (i \in P, j \in Q), \tag{4.32}$$

$$\psi(\bar{\gamma}_j^i, -v_j(\bar{y}^i)) = 0 \;\; (i \in P, j \in Q).$$

As observed in [88], the system of equations (4.32) is not directly equivalent to (4.31). The system (4.32) allows the case that

$$\bar{\mu}_i = 0, \;\; g(\bar{x}, \bar{y}^i) < 0.$$

However, if there is an $n + (m + q + 1)p$ dimensional vector, say $(\bar{x}, \bar{\mu}, \bar{y}, \bar{\gamma})$, satisfying (4.32), the variables indexed by i with $\bar{\mu}_i = 0$ may be dropped. Thus, we get a solution of (4.31). It is easy to see that a solution of (4.31) satisfies (4.32). Hence, (4.31) and (4.32) are *equivalent*,

and finding the solution of (4.31) amounts to finding a zero of the function $T : \mathbb{R}^N \to \mathbb{R}^N$

$$T(z) = \begin{pmatrix} \nabla f(x) + \sum_{i=1}^{p} \mu_i \nabla_x g(x, y^i) \\ \psi(\mu_1, -g(x, y^1)) \\ \vdots \\ \psi(\mu_p, -g(x, y^p)) \\ \nabla_y g(x, y^1) - \sum_{j=1}^{q} \gamma_j^1 \nabla_y v_j(y^1) \\ \psi(\gamma_1^1, -v_1(y^1)) \\ \vdots \\ \psi(\gamma_q^1, -v_q(y^1)) \\ \vdots \\ \nabla_y g(x, y^p) - \sum_{j=1}^{q} \gamma_j^p \nabla_y v_j(y^p) \\ \psi(\gamma_1^p, -v_1(y^p)) \\ \vdots \\ \psi(\gamma_q^p, -v_q(y^p)) \end{pmatrix}, \tag{4.33}$$

where $N = n + (m + q + 1)p$ with $z = (x^T, \mu^T, y^T, \gamma^T)^T \in \mathbb{R}^{n+(m+q+1)p}$, $x \in \mathbb{R}^n, \mu \in \mathbb{R}^p, y \in \mathbb{R}^{mp}$ and $\gamma \in \mathbb{R}^{qp}$.

4.3.3 Semismooth optimality conditions for *GSIP*

In this subsection we derive semismooth optimality conditions for the following *GSIP* (see Section 2.2) as in our recent paper [108].

$$GSIP: \qquad \text{minimize} \quad f(x) \quad \text{subject to} \quad x \in M$$

with the feasible set

$$M \;=\; \{\, x \in \mathbb{R}^n \,|\, g(x, y) \leq 0 \text{ for all } y \in Y(x) \,\}$$

and

$$Y(x) \;=\; \{\, y \in \mathbb{R}^m \,|\, v_j(x, y) \leq 0 \;\; (j \in Q) \,\}.$$

If $\bar{x} \in M$ is a local minimizer of *GSIP* at which the Reduction Ansatz holds, then, by Theorem 2.2.10, $\bar{x}$ is also a local minimizer of the locally reduced problem $GSIP_{red}$, and necessary

optimality conditions for $\bar{x}$ in the reduced problem are also necessary optimality conditions for $\bar{x}$ in the original problem. In particular, if $\bar{x} \in M$ is a local minimizer of *GSIP* at which the Reduction Ansatz (see Subsection 2.2.2) and the Mangasarian-Fromovitz constraint qualification hold, then there exist multipliers $\bar{\mu}_i \geq 0$, $i \in P = \{1, ..., p\}$, such that

$$\nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla \varphi_i(\bar{x}) = 0. \tag{4.34}$$

Note that all constraints in $GSIP_{red}$ are active at $\bar{x}$ by construction, that is, we have $\varphi_i(\bar{x}) = 0$ for all $i \in P$.

Using $\varphi_i(\bar{x}) = g(\bar{x}, \bar{y}^i)$, the fact that each $\bar{y}^i$ is a global maximizer of the lower level problem, as well as the evaluation of (2.12) at $\bar{x}$, we arrive at the system

$$\begin{aligned}
&\nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) = 0, \\
&\bar{\mu}_i \geq 0, \;\; g(\bar{x}, \bar{y}^i) = 0 \;\; (i \in P), \\
&\nabla_y g(\bar{x}, \bar{y}^i) - \sum_{j=1}^{q} \bar{\gamma}^i_j \nabla_y v_j(\bar{x}, \bar{y}^i) = 0 \;\; (i \in P), \\
&\bar{\gamma}^i_j \geq 0, v_j(\bar{x}, \bar{y}^i) \leq 0 \;\; (i \in P, j \in Q), \\
&\bar{\gamma}^i_j v_j(\bar{x}, \bar{y}^i) = 0 \;\; (i \in P, j \in Q).
\end{aligned} \tag{4.35}$$

With any NCP function ψ, along the same lines as in Subsection 4.3.2 the solution of (4.35) is seen to be equivalent to finding a zero of the function

$$T(z) = \begin{pmatrix} \nabla f(x) + \sum_{i=1}^{p} \mu_i \nabla_x \mathcal{L}(x, y^i, \gamma^i) \\ \psi(\mu_1, -g(x, y^1)) \\ \vdots \\ \psi(\mu_p, -g(x, y^p)) \\ \nabla_y g(x, y^1) - \sum_{j=1}^{q} \gamma_j^1 \nabla_y v_j(x, y^1) \\ \psi(\gamma_1^1, -v_1(x, y^1)) \\ \vdots \\ \psi(\gamma_q^1, -v_q(x, y^1)) \\ \vdots \\ \nabla_y g(x, y^p) - \sum_{j=1}^{q} \gamma_j^p \nabla_y v_j(x, y^p) \\ \psi(\gamma_1^p, -v_1(x, y^p)) \\ \vdots \\ \psi(\gamma_q^p, -v_q(x, y^p)) \end{pmatrix}, \tag{4.36}$$

with $z = (x^T, \mu^T, y^T, \gamma^T)^T \in \mathbb{R}^N$, $N = n + (m + q + 1)p$, $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}^p$, $y \in \mathbb{R}^{mp}$ and $\gamma \in \mathbb{R}^{qp}$. Again, T is strongly semismooth under our assumptions.

We emphasize that for a standard *SIP* the term $\nabla_x \mathcal{L}(x, y^i, \gamma^i)$ in T is replaced by $\nabla_x g(x, y^i)$, and $v_j(x, y^i)$ by $v_j(y^i)$, that is, the function T from (4.36) generalizes the function T from (4.33) from *SIP* to *GSIP*. Note that for standard *SIP* it is not necessary to assume the Reduction Ansatz to derive the function T. However, the situation for *GSIP* is not essentially more restrictive, since in Subsection 5.1.1 we will anyway assume the Reduction Ansatz at the solution point for our convergence result and the Reduction Ansatz at local minimizers of *GSIP* is a weak assumption [23].

CHAPTER 5

SEMISMOOTH NEWTON APPROACH for SEMI-INFINITE PROGRAMMING

In Section 4.2, a review of semismooth Newton method applied to finitely constrained programming and the convergence of the method are given. In this section we will apply semismooth Newton method to semi-infinite programming problems.

In Section 5.1, for the convergence of the semismooth Newton method we show that a standard regularity condition is needed and indeed it is satisfied under natural assumptions for semi-infinite programming. Under the Reduction Ansatz (see Subsection 2.2.2) in the lower level and strong stability in the reduced upper level problem, i.e., Robinson condition (see Definition 4.2.2) this regularity condition is satisfied. We do not have to assume strict complementary slackness in the upper level in this section.

As we have mentioned in Chapter 1, NCP functions were used to get a nonsmooth reformulation of the KKT conditions in standard *SIP* in the article [88]. A regularity condition was suggested to guarantee convergence of a specific semismooth Newton method. However, in [88] strict complementarity is a part of the regularity condition in the upper as well as in the lower level problem. The standard Newton method would converge under these assumptions. In Section 5.1, we complement the result in [88] by showing convergence under the case strict complementarity violation in the upper level. We also transfer the method to generalized semi-infinite programming.

In Section 5.2 we consider the case strict complementarity violated in the upper and in the lower level. Now, in locally reduced problem, the auxiliary functions are then not C^2 functions. But still, it is possible to show convergence of the semi-smooth Newton method under

weak assumptions.

5.1 The Case with the Strict Complementarity in the Lower Level Problem

In this section we wish to apply the semismooth Newton approach from Section 4.1 to find a zero of the function T from (4.36). In particular, we want to use Theorem 4.1.10 and, thus, find a sufficient condition for CD-regularity of T in a solution point $\bar{z}$.

As we mention in our recent paper [108], the part of this condition concerning the lower level problem will be the Reduction Ansatz (see Subsection 2.2.2), while in the upper level problem we will assume the so-called *Robinson condition* (which is previously used for finitely constrained programming, see Definition 4.2.2) . In fact, consider $\bar{x} \in M$ and the locally reduced problem $GSIP_{red}$ where, according to the definition of T, we neglect the fact that all constraints φ_i $(i \in P)$ are active by their definition. Let $P_0(\bar{x}) = \{i \in P \,|\, \varphi_i(\bar{x}) = 0\}$ be the set of active indices at $\bar{x}$ for the upper level problem. Recall from Theorem 2.2.10(c) that the auxiliary functions $\varphi_i (i \in P)$, in the reduced problem $GSIP_{red}$ are twice continuously differentiable, so that it makes sense to impose a second order regularity condition on $GSIP_{red}$.

5.1.1 Convergence of the method

For convergence of our method, in the upper level we will assume Robinson condition. In the following we give this condition [108].

The Robinson condition is said to hold at $\bar{x}$ if the following conditions (GSIP-I) and (GSIP-II) are satisfied:

(GSIP-I) The *linear independence constraint qualification*:

$(LI)_{GSIP}$ $\{\nabla_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) \,|\, i \in P_0(\bar{x})\}$ is a linearly independent family.

If $\bar{x}$ is a local minimizer, then there exists a unique vector $\bar{\mu} \in \mathbb{R}^p$ of Lagrange multipliers with

$$\begin{aligned} (KKT)_{GSIP} \quad \nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) &= 0, \\ \bar{\mu}_i \geq 0,\ g(\bar{x}, \bar{y}^i) \leq 0,\ \bar{\mu}_i g(\bar{x}, \bar{y}^i) &= 0 \ (i \in P). \end{aligned}$$

(GSIP-II) The *strong second order sufficiency condition*:

$$(SSOSC)_{GSIP} \left\{ \begin{array}{c} \xi^T \nabla_x^2 L(\bar{x},\bar{y},\bar{\mu},\bar{\gamma})\xi > 0 \text{ for all } \xi \in G_{GSIP} \setminus \{0\} \text{ with} \\ G_{GSIP} = \{d \in \mathbb{R}^n |\, D_x \mathcal{L}(\bar{x},\bar{y}^i,\bar{\gamma}^i)d = 0 \text{ for } i \in P_+(\bar{x})\}. \end{array} \right\}$$

with $P_+(\bar{x}) = \{i \in P_0(\bar{x}) \,|\, \bar{\mu}_i > 0\}$. Here, $\nabla_x^2 L(\bar{x},\bar{y},\bar{\mu},\bar{\gamma})$ stands for the Hessian of the Lagrangian $f(x) + \sum_{i=1}^{p} \mu_i \varphi_i(x)$ of $GSIP_{red}$, that is, for $\nabla_x^2 f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x^2 \varphi_i(\bar{x})$, with

$$\nabla_x^2 \varphi_i(\bar{x}) = \nabla_x^2 \mathcal{L}(\bar{x},\bar{y}^i,\bar{\gamma}^i) - \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}_i \\ -D_x v_{Q_0^i}(\bar{x},\bar{y}^i) \end{pmatrix}^T \begin{pmatrix} D_y^2 \mathcal{L}_i & -\nabla_y v_{Q_0^i}(\bar{x},\bar{y}^i) \\ -D_y v_{Q_0^i}(\bar{x},\bar{y}^i) & 0 \end{pmatrix}^{-1} \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}_i \\ -D_x v_{Q_0^i}(\bar{x},\bar{y}^i) \end{pmatrix},$$

where $D_x v_{Q_0^i}$ stands for the matrix with rows $D_x v_j,\ j \in Q_0^i = Q_0(\bar{x},\bar{y}^i)$. $D_y v_{Q_0^i}$ is defined similarly.

Note that the Robinson condition does *not* assume strict complementary slackness. In the standard *SIP* case, $\nabla_x \mathcal{L}(\bar{x},\bar{y}^i,\bar{\gamma}^i)$ reduces to $\nabla_x g(\bar{x},\bar{y}^i)$, and in the Hessian of the Lagrangian we obtain

$$\nabla_x^2 \varphi_i(\bar{x}) = \nabla_x^2 g(\bar{x},\bar{y}^i) - \begin{pmatrix} \nabla_{yx}^2 g(\bar{x},\bar{y}^i) \\ 0 \end{pmatrix}^T \begin{pmatrix} D_y^2 \mathcal{L}_i & -\nabla v_{Q_0^i}(\bar{y}^i) \\ -D v_{Q_0^i}(\bar{y}^i) & 0 \end{pmatrix}^{-1} \begin{pmatrix} \nabla_{yx}^2 g(\bar{x},\bar{y}^i) \\ 0 \end{pmatrix}.$$

where the term

$$\begin{pmatrix} \nabla_{yx}^2 g(\bar{x},\bar{y}^i) \\ 0 \end{pmatrix}^T \begin{pmatrix} D_y^2 \mathcal{L}_i & -\nabla v_{Q_0^i}(\bar{y}^i) \\ -D v_{Q_0^i}(\bar{y}^i) & 0 \end{pmatrix}^{-1} \begin{pmatrix} \nabla_{yx}^2 g(\bar{x},\bar{y}^i) \\ 0 \end{pmatrix}$$

is called as *shift term* in [44].

Now consider a zero z of T from (4.36). Then the KKT conditions mentioned in (Q-I) and (GSIP-I) hold by definition of T. The remaining conditions in the Reduction Ansatz and the Robinson condition are algebraic conditions on the involved functions which can also be imposed independently of the fact that we deal with lower level global maximizers and upper level local minimizers. In this sense, we can make the following assumption:

Assumption 5.1.1 ([108]) *The Reduction Ansatz and the Robinson condition hold at z.*

Theorem 5.1.2 ([108]) *Suppose that $\bar{z} = (\bar{x}^T, \bar{\mu}^T, \bar{y}^T, \bar{\gamma}^T)^T$ is a zero of T from (4.36) with the choices $\psi = \psi_{FB}$ or $\psi = \psi_{min}$, and that Assumption 5.1.1 holds at $\bar{z}$. Then T is CD-regular at $\bar{z}$.*

Proof. Let $\bar{z}$ be a zero of T at which Assumption 5.1.1 holds, and let ψ denote either one of the two NCP functions ψ_{FB} and ψ_{min}. We only consider the case of two active indices, that is, $p = 2$, the general case running along the same lines. We will distinguish two cases, depending on whether upper level strict complementary slackness holds or not.

Case 1: Strict complementarity holds in the upper level problem.

We have $\psi(\bar{\mu}_i, -g(\bar{x}, \bar{y}^i)) = 0$ and $(\bar{\mu}_i, -g(\bar{x}, \bar{y}^i)) \neq 0$, so that ψ is differentiable at $(\bar{\mu}_i, -g(\bar{x}, \bar{y}^i))$ with gradient

$$D\psi(\bar{\mu}_i, -g(\bar{x}, \bar{y}^i)) = \begin{cases} (-1, 0), & i \notin P_0(\bar{x}) \\ (0, -1), & i \in P_0(\bar{x}) \end{cases}$$

for $i = 1, 2$ (see (A.3) in Appendix A).

The Reduction Ansatz in the lower level problem implies $\psi(\bar{\gamma}^i_j, -v_j(\bar{x}, \bar{y}^i)) = 0$ and $(\bar{\gamma}^i_j, -v_j(\bar{x}, \bar{y}^i)) \neq 0$, so that ψ is also differentiable at $(\bar{\gamma}^i_j, -v_j(\bar{x}, \bar{y}^i))$ with gradient

$$D\psi(\bar{\gamma}^i_j, -v_j(\bar{x}, \bar{y}^i)) = \begin{cases} (-1, 0), & j \notin Q_0(\bar{x}, \bar{y}^i) \\ (0, -1), & j \in Q_0(\bar{x}, \bar{y}^i) \end{cases}$$

for $i = 1, 2$ and $j \in Q$. Together this means that T is differentiable at $\bar{z}$, and its Jacobian is the matrix

$$DT(\bar{z}) = \begin{pmatrix} \mathcal{B}_{11} & \mathcal{B}_{12} & \mathcal{B}_{13} \\ \mathcal{B}_{21} & \mathcal{B}_{22} & 0 \\ \mathcal{B}_{31} & 0 & \mathcal{B}_{33} \end{pmatrix} \tag{5.1}$$

with the following blocks:

$$\mathcal{B}_{11} = \begin{pmatrix} D_x^2 f(\bar{x}) + \sum_{i=1}^{2} \bar{\mu}_i \nabla_x^2 \mathcal{L}_i & \nabla_x \mathcal{L}_1 & \nabla_x \mathcal{L}_2 \\ \lambda_1 D_x g(\bar{x}, \bar{y}^1) & \theta_1 & 0 \\ \lambda_2 D_x g(\bar{x}, \bar{y}^2) & 0 & \theta_2 \end{pmatrix}$$

with $\nabla_x^2 \mathcal{L}_i = \nabla_x^2 \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i)$, etc.,

$$\lambda_i = \begin{cases} 0, & i \notin P_0(\bar{x}) \\ 1, & i \in P_0(\bar{x}) \end{cases}, \quad \theta_i = \begin{cases} -1, & i \notin P_0(\bar{x}) \\ 0, & i \in P_0(\bar{x}) \end{cases}, \; i = 1, 2,$$

$$\mathcal{B}_{12} = \begin{pmatrix} \bar{\mu}_1 \nabla^2_{xy} \mathcal{L}_1 & -\bar{\mu}_1 \nabla_x v(\bar{x}, \bar{y}^1) \\ \lambda_1 D_y g(\bar{x}, \bar{y}^1) & 0 \\ 0 & 0 \end{pmatrix},$$

$$\mathcal{B}_{13} = \begin{pmatrix} \bar{\mu}_2 \nabla^2_{xy} \mathcal{L}_2 & -\bar{\mu}_2 \nabla_x v(\bar{x}, \bar{y}^2) \\ 0 & 0 \\ \lambda_2 D_y g(\bar{x}, \bar{y}^2) & 0 \end{pmatrix},$$

and with

$$\alpha^i_j = \begin{cases} 0, & j \notin Q_0(\bar{x}, \bar{y}^i) \\ 1, & j \in Q_0(\bar{x}, \bar{y}^i) \end{cases}, \quad \beta^i_j = \begin{cases} -1, & j \notin Q_0(\bar{x}, \bar{y}^i) \\ 0, & j \in Q_0(\bar{x}, \bar{y}^i) \end{cases},$$

$$\mathcal{B}_{i+1,1} = \begin{pmatrix} \nabla^2_{yx} \mathcal{L}_i & 0 & 0 \\ \operatorname{diag}(\alpha^i) D_x v(\bar{x}, \bar{y}^i) & 0 & 0 \end{pmatrix}$$

as well as

$$\mathcal{B}_{i+1,i+1} = \begin{pmatrix} \nabla^2_y \mathcal{L}_i & -\nabla_y v(\bar{x}, \bar{y}^i) \\ \operatorname{diag}(\alpha^i) D_y v(\bar{x}, \bar{y}^i) & \operatorname{diag}(\beta^i) \end{pmatrix}, \; i = 1, 2.$$

Our aim is to show that $DT(\bar{z})$ is nonsingular under Assumption 5.1.1. The main idea of the proof is to consider an appropriate Schur complement in the block matrix $DT(\bar{z})$ (see Appendix B).

Note that the matrices $\mathcal{B}_{22}$ and $\mathcal{B}_{33}$ are nonsingular under the Reduction Ansatz. In fact, by the definitions of α^1 and β^1, $\mathcal{B}_{22}$ is nonsingular if and only if the matrix

$$\begin{pmatrix} \nabla^2_y \mathcal{L}_1 & \nabla_y v_{Q_0^1}(\bar{x}, \bar{y}^1) \\ D_y v_{Q_0^1}(\bar{x}, \bar{y}^1) & 0 \end{pmatrix}$$

is nonsingular. Since under the Reduction Ansatz we have $(LI)_{Q(\bar{x})}$ and $(SOSC)_{Q(\bar{x})}$ at $\bar{y}^1$, the latter matrix is nonsingular by Lemma B.0.37. Analogously the nonsingularity of $\mathcal{B}_{33}$ is shown.

Consequently, according to Lemma B.0.35 the matrix $DT(\bar{z})$ is nonsingular if and only if the Schur complement

$$\begin{aligned} S &= DT(\bar{z}) / \begin{pmatrix} \mathcal{B}_{22} & 0 \\ 0 & \mathcal{B}_{33} \end{pmatrix} = \mathcal{B}_{11} - (\mathcal{B}_{12}, \mathcal{B}_{13}) \begin{pmatrix} \mathcal{B}_{22} & 0 \\ 0 & \mathcal{B}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \mathcal{B}_{21} \\ \mathcal{B}_{31} \end{pmatrix} \\ &= \mathcal{B}_{11} - \mathcal{B}_{12} \mathcal{B}_{22}^{-1} \mathcal{B}_{21} - \mathcal{B}_{13} \mathcal{B}_{33}^{-1} \mathcal{B}_{31} \end{aligned}$$

is nonsingular. We will show that the latter is the case under the Robinson condition.

The calculation of $\mathcal{B}_{12}\mathcal{B}_{22}^{-1}\mathcal{B}_{21}$ is next step. From the block structures of $\mathcal{B}_{12}$ and $\mathcal{B}_{21}$ it is clear that only the first and second block in the first block column of this matrix are nonzero, and we only have to calculate

$$\begin{pmatrix} \bar{\mu}_1\nabla^2_{xy}\mathcal{L}_1 & -\bar{\mu}_1\nabla_x v(\bar{x},\bar{y}^1) \\ \lambda_1 D_y g(\bar{x},\bar{y}^1) & 0 \end{pmatrix} \mathcal{B}_{22}^{-1} \begin{pmatrix} \nabla^2_{yx}\mathcal{L}_1 \\ \operatorname{diag}(\alpha^1) D_x v(\bar{x},\bar{y}^1) \end{pmatrix}. \tag{5.2}$$

Now recall that the implicit functions $y^1(x)$ and $\gamma^1(x)$ from Theorem 2.2.10 satisfy

$$\begin{pmatrix} \nabla_y \mathcal{L}(x, y^1(x), \gamma^1(x)) \\ \psi(\gamma_1^1(x), -v_1(x, y^1(x))) \\ \vdots \\ \psi(\gamma_q^1(x), -v_q(x, y^1(x))) \end{pmatrix} \equiv 0. \tag{5.3}$$

Taking derivatives with respect to x and evaluating at $\bar{x}$ yields (by the usual Chain rule)

$$\begin{pmatrix} \nabla^2_y \mathcal{L}_1 & -\nabla_y v(\bar{x},\bar{y}^1) \\ \operatorname{diag}(\alpha^1) D_y v(\bar{x},\bar{y}^1) & \operatorname{diag}(\beta^1) \end{pmatrix} \begin{pmatrix} Dy^1(\bar{x}) \\ D\gamma^1(\bar{x}) \end{pmatrix} + \begin{pmatrix} \nabla^2_{yx}\mathcal{L}_1 \\ \operatorname{diag}(\alpha^1) D_x v(\bar{x},\bar{y}^1) \end{pmatrix} = 0$$

and, thus,

$$\mathcal{B}_{22}^{-1} \begin{pmatrix} \nabla^2_{yx}\mathcal{L}_1 \\ \operatorname{diag}(\alpha^1) D_x v(\bar{x},\bar{y}^1) \end{pmatrix} = - \begin{pmatrix} Dy^1(\bar{x}) \\ D\gamma^1(\bar{x}) \end{pmatrix}.$$

Consequently, the matrix in (5.2) becomes

$$\begin{pmatrix} -\bar{\mu}_1\nabla^2_{xy}\mathcal{L}_1 Dy^1(\bar{x}) + \bar{\mu}_1\nabla_x v(\bar{x},\bar{y}^1) D\gamma^1(\bar{x}) \\ -\lambda_1 D_y g(\bar{x},\bar{y}^1) Dy^1(\bar{x}) \end{pmatrix}.$$

With an analogous calculation for $\mathcal{B}_{13}\mathcal{B}_{33}^{-1}\mathcal{B}_{31}$ we arrive at

$$S = \begin{pmatrix} \widetilde{\nabla^2_x L} & \nabla_x \mathcal{L}_1 & \nabla_x \mathcal{L}_2 \\ \lambda_1 \left(D_x g(\bar{x},\bar{y}^1) + D_y g(\bar{x},\bar{y}^1) Dy^1(\bar{x}) \right) & \theta_1 & 0 \\ \lambda_2 \left(D_x g(\bar{x},\bar{y}^2) + D_y g(\bar{x},\bar{y}^2) Dy^2(\bar{x}) \right) & 0 & \theta_2 \end{pmatrix}$$

with

$$\widetilde{\nabla^2_x L} = D^2_x f(\bar{x}) + \sum_{i=1}^{2} \bar{\mu}_i \left(\nabla^2_x \mathcal{L}_i + \nabla^2_{xy}\mathcal{L}_i Dy^i(\bar{x}) - \nabla_x v(\bar{x},\bar{y}^i) D\gamma^i(\bar{x}) \right).$$

Using (5.10) it is not hard to see that

$$\nabla^2 \varphi_i(\bar{x}) = \nabla^2_x \mathcal{L}_i + \nabla^2_{xy}\mathcal{L}_i Dy^i(\bar{x}) - \nabla_x v(\bar{x},\bar{y}^i) D\gamma^i(\bar{x})$$

holds for $i = 1, 2$, so that the matrix $\widetilde{\nabla_x^2 L}$ coincides with $\nabla_x^2 L$ from condition (GSIP-II).

Next we show

$$D_x g(\bar{x}, \bar{y}^i) + D_y g(\bar{x}, \bar{y}^i) Dy^i(\bar{x}) \;=\; D_x \mathcal{L}_i \tag{5.4}$$

for $i = 1, 2$. In fact, for $i = 1$ we have

$$D_y g(\bar{x}, \bar{y}^1) Dy^1(\bar{x}) \;=\; \sum_{j=1}^{q} \bar{\gamma}_j^1 D_y v_j(\bar{x}, \bar{y}^1) Dy^1(\bar{x}). \tag{5.5}$$

The implicit functions from Theorem 2.2.10 particularly satisfy the identity

$$\gamma_j^1(x) \cdot v_j(x, y^1(x)) \;\equiv\; 0 \;\; (j \in Q).$$

Taking derivatives with respect to x yields

$$D\gamma_j^1(x)\, v_j(x, y^1(x)) + \gamma_j^1(x) \left(D_x v_j(x, y^1(x)) + D_y v_j(x, y^1(x)) Dy^1(x) \right) \;\equiv\; 0$$

for $j \in Q$, where the first term vanishes at $\bar{x}$ for $j \in Q_0(\bar{x}, \bar{y}^1)$ due to $v_j(\bar{x}, \bar{y}^1) = 0$, and for $j \notin Q_0(\bar{x}, \bar{y}^1)$ because of $\gamma_j^1(x) \equiv 0$ and, thus, $D\gamma_j^1(\bar{x}) = 0$. Evaluating the remaining terms at $\bar{x}$ implies

$$\bar{\gamma}_j^1 D_y v_j(\bar{x}, \bar{y}^1) Dy^1(\bar{x}) \;=\; -\bar{\gamma}_j^1 D_x v_j(\bar{x}, \bar{y}^1) \;\; (j \in Q), \tag{5.6}$$

and a combination of (5.5) and (5.6) yields

$$D_y g(\bar{x}, \bar{y}^1) Dy^1(\bar{x}) \;=\; -\sum_{j=1}^{q} \bar{\gamma}_j^1 D_x v_j(\bar{x}, \bar{y}^1).$$

This shows (5.4) for $i = 1$, and analogously for $i = 2$. As a consequence, the Schur complement simplifies further to

$$S \;=\; \begin{pmatrix} \nabla_x^2 L & \nabla_x \mathcal{L}_1 & \nabla_x \mathcal{L}_2 \\ \lambda_1 D_x \mathcal{L}_1 & \theta_1 & 0 \\ \lambda_2 D_x \mathcal{L}_2 & 0 & \theta_2 \end{pmatrix}.$$

By the definitions of λ_i and θ_i, $i = 1, 2$, S is nonsingular if and only if the matrix

$$\begin{pmatrix} \nabla_x^2 L & \nabla_x \mathcal{L}_{P_0(\bar{x})} \\ D_x \mathcal{L}_{P_0(\bar{x})} & 0 \end{pmatrix}$$

is nonsingular. Under the Robinson condition the latter is true in view of Lemma B.0.37. This completes the proof for Case 1.

Case 2: Strict complementarity is violated in the upper level problem.

In this case we have $\bar{\mu}_i = g(\bar{x}, \bar{y}^j) = 0$ for at least one $i \in \{1, 2\}$. Here we only consider the case that strict complementarity is violated at $i = 1$ with $P_0(\bar{x}) = \{1, 2\}$, the general case running along the same lines.

In the present case, T is *not* differentiable at $\bar{z}$ since ψ is not differentiable at the origin. The Clarke (in fact, convex) subdifferentials [6, 95] of ψ_{FB} and ψ_{min} are given in Lemma A.0.33 in the Appendix A. Moreover, the generalized Jacobian of $\psi(\mu_1, -g(x, y^1))$ can be computed by the Chain Rule II and Proposition 2.3.6 from [6], noting that convex functions are regular. In fact, its generalized Jacobian with respect to (x, μ_1, y^1) is

$$\left\{\left(\lambda_1 D_x g(x, y^1),\ \theta_1\,,\ \lambda_1 D_y g(x, y^1)\right) \mid (-\lambda_1, \theta_1) \in \partial\psi(0, 0)\right\}. \tag{5.7}$$

This means that the elements of $\partial T(\bar{z})$ can be parameterized by

$$\{W(\lambda_1, \theta_1) \mid (-\lambda_1, \theta_1) \in \partial\psi(0, 0)\},$$

where $W(\lambda_1, \theta_1)$ is a matrix of exactly the form from (5.1) where the blocks $\mathcal{B}_{11}, \mathcal{B}_{12}, \mathcal{B}_{13}$, etc. are defined as previously. Consequently, proving CD-regularity of T at $\bar{z}$ amounts to showing nonsingularity of all matrices $W(\lambda_1, \theta_1)$ with $(-\lambda_1, \theta_1) \in \partial\psi(0, 0)$.

Choose any (λ_1, θ_1) with $(-\lambda_1, \theta_1) \in \partial\psi(0, 0)$. With the same arguments as in Case 1 we find that $W(\lambda_1, \theta_1)$ is nonsingular if and only if the matrix

$$S(\lambda_1, \theta_1) = \begin{pmatrix} \nabla_x^2 L & \nabla_x \mathcal{L}_1 & \nabla_x \mathcal{L}_2 \\ \lambda_1 D_x \mathcal{L}_1 & \theta_1 & 0 \\ \lambda_2 D_x \mathcal{L}_2 & 0 & \theta_2 \end{pmatrix}$$

is nonsingular. The latter, however, is true by Theorem 4.2.3 i.e., by [85, Theorem 4.2] for finitely constrained programming problems. This completes the proof. •

Remark 5.1.3 ([108]) *In the special case of SIP, an explicit proof of Theorem 5.1.2 would be shorter due to the simplifications that* $\nabla_x \mathcal{L}_i$ *is replaced by* $\nabla_x g(\bar{x}, \bar{y}^j)$, $\nabla_{xy} \mathcal{L}_i$ *by* $\nabla_{xy} g(\bar{x}, \bar{y}^j)$, *and* $\nabla_x v(\bar{x}, \bar{y}^j)$ *vanishes.*

Remark 5.1.4 ([108]) *Assumption 5.1.1 is a weak assumption in the following sense: the Reduction Ansatz holds generically at all local minimizers of GSIP, and they are even non-degenerate critical points of the locally reduced problem* $GSIP_{red}$. *That is, generically even*

upper level strict complementarity and the Robinson condition hold. While this fact has been known for standard SIP for some time [102], it is a recent result for GSIP [23]. In view of this genericity, one can expect Assumption 5.1.1 to be satisfied in practical applications.

Altogether, in Subsection 5.1.1 we have shown that under the weak Assumption 5.1.1, in view of Theorem 4.1.10 and Theorem 5.1.2, the semismooth Newton method from (4.4) converges q-quadratically.

5.2 The Case without the Strict Complementarity in the Lower Level Problem

Section 5.1 shows that the semismooth Newton method for semi-infinite programming can actually handle nonsmoothness, since there the result from [88] is extended to the case of violated strict complementarity in the *upper level* problem. Moreover, we transferred the semismooth approach from standard to generalized semi-infinite programming.

This section completes this analysis by considering the case of strict complementarity violation in the *lower level*. For the convergence of a semismooth Newton method we give an appropriate new regularity condition thus justifying the NCP approach for semi-infinite programs in the absence of strict complementarity. In the present case, the convergence analysis is essentially more complicated due to the lack of differentiability of the auxiliary functions of the so-called reduced problem (see Section 5.2.1.1). We refer the reader to our recent paper [115] for this section.

5.2.1 Convergence of the method

In this section we wish to find a sufficient condition for CD-regularity of the function T from (4.36) at a zero $\overline{z}$, so that the semismooth Newton approach from Section 4.1 may be applied to identify $\overline{z}$.

The Reduction Ansatz without strict complementarity given in Subsubsection 5.2.1.1 is the part of this condition concerning the lower level problem. In the upper level problem we will assume a natural *generalization* of the so-called *Robinson condition* [93, 108] for the reduced problem (see Section 5.2.1.3). However, the auxiliary functions φ_i $(i \in P)$, in the reduced problem $GSIP_{red}$ are not twice continuously differentiable (see Theorem 5.2.1(c)), so that

we cannot state a standard second order regularity condition for $GSIP_{red}$. Instead, first we compute the generalized Hessians of φ_i , $i \in P$, at $\bar{x}$.

5.2.1.1 The Reduction Ansatz without strict complementarity

As stated in our paper [115], here we consider a feasible point $\bar{x}$ of *GSIP* and its set of active indices

$$Y_0(\bar{x}) = \{y \in Y(\bar{x}) \mid g(\bar{x}, y) = 0\}.$$

All defining functions of *GSIP* are at least twice continuously differentiable. Consider the following conditions (Q^+-I) and (Q^+-II) at $\bar{y} \in Y_0(\bar{x})$ in $Q(\bar{x})$:

(Q^+-I) The *linear independence constraint qualification*:

$$(LI)_{Q(\bar{x})} \quad \{\nabla_y v_j(\bar{x}, \bar{y}) \mid j \in Q_0(\bar{x}, \bar{y})\} \text{ is a linearly independent family,}$$

where $Q_0(\bar{x}, \bar{y}) = \{j \in Q \mid v_j(\bar{x}, \bar{y}) = 0\}$ is the set of lower level active indices at $\bar{y} \in Y(\bar{x})$.

Because of (Q^+-I) we have the following lower level KKT conditions: there exists a unique vector of Lagrange multipliers $\bar{\gamma} \in \mathbb{R}^q$ such that

$$(KKT)_{Q(\bar{x})} \left\{ \begin{array}{c} \nabla_y g(\bar{x}, \bar{y}) - \sum\limits_{j=1}^{q} \bar{\gamma}_j \nabla_y v_j(\bar{x}, \bar{y}) = 0 \\ v_j(\bar{x}, \bar{y}) \leq 0 \\ \bar{\gamma}_j \geq 0 \\ \bar{\gamma}_j v_j(\bar{x}, \bar{y}) = 0, \ j \in Q. \end{array} \right\}. \tag{5.8}$$

(Q^+-II) The *strong second order sufficiency condition*:

$$(SSOSC)_{Q(\bar{x})} \left\{ \begin{array}{c} \eta^T \nabla_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) \eta < 0 \ \text{ for all } \eta \in G^+_{Q(\bar{x})} \setminus \{0\}, \text{where} \\ G^+_{Q(\bar{x})} := \{\eta \in \mathbb{R}^m \mid D_y v_j(\bar{x}, \bar{y}) \eta = 0, \ j \in Q_0^+(\bar{x}, \bar{y})\} \end{array} \right\} \tag{5.9}$$

with $Q_0^+(\bar{x}, \bar{y}) := \{j \in Q_0(\bar{x}, \bar{y}) \mid \bar{\gamma}_j > 0\}$ and $\mathcal{L}(\bar{x}, y, \gamma) := g(\bar{x}, y) - \sum\limits_{j=1}^{q} \gamma_j v_j(\bar{x}, y)$, the *Lagrangian* associated with $Q(\bar{x})$.

The *Reduction Ansatz without strict complementarity* is said to hold at $\bar{x} \in M$ if all elements of $Y_0(\bar{x})$ satisfy the conditions (Q^+-I) and (Q^+-II). It generalizes the standard Reduction Ansatz

in [108] to the case of possibly violated strict complementary slackness. In the case of strict complementary slackness, the Reduction Ansatz without strict complementarity coincides with the standard Reduction Ansatz.

If the Reduction Ansatz without strict complementarity holds at some $\bar{x} \in M$, we can locally reduce *GSIP* to $GSIP_{red}$, the so-called reduced generalized semi-infinite problem. Note that, as stated in [96], the Reduction Ansatz without strict complementarity is a sufficient condition for strongly stable lower level stationary points in the sense of Kojima [60].

Reduction Ansatz without strict complementarity was stated for *SIP* in [52], for *GSIP* the following theorem was stated.

Theorem 5.2.1 ([34]) *Let the Reduction Ansatz without strict complementarity be satisfied at a feasible point $\bar{x}$ of GSIP and all defining functions of GSIP are assumed to be at least twice continuously differentiable. Then,*

(a) The active index set is finite, $Y_0(\bar{x}) = \{\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^p\}$, and there exist neighborhoods $U_{\bar{x}}$ of $\bar{x}$ and $V_{\bar{y}^i}$ of $\bar{y}^i$ and Lipschitz continuous functions

$y^i : U_{\bar{x}} \to V_{\bar{y}^i}$ *, where* $y^i(\bar{x}) = \bar{y}^i$,

$\gamma^i : U_{\bar{x}} \to \mathbb{R}^q$, *where* $\gamma^i(\bar{x}) = \bar{\gamma}^i$,

such that for every $x \in U_{\bar{x}}$ the value $y^i(x)$ is the unique local maximizer of $Q(x)$ in $V_{\bar{y}^i}$ with corresponding Lagrange multiplier vector $\gamma^i(x)$.

(b) The following finite reduction holds: $\bar{x}$ is a solution of GSIP, locally in a neighborhood $U_{\bar{x}}$ of $\bar{x}$, if and only if $\bar{x}$ is a local solution of the so-called reduced problem

$$GSIP_{red}: \min_{x \in U_{\bar{x}}} f(x) \quad \text{subject to} \quad \varphi_i(x) = g(x, y^i(x)) \leq 0, \quad \text{for all } i = 1, 2, \ldots, p.$$

(c) The functions φ_i from part (b) are of class C^1 and for all $x \in U_{\bar{x}}$ their gradients satisfy

$$D\varphi_i(x) = D_x \mathcal{L}(x, y^i(x), \gamma^i(x)). \tag{5.10}$$

Remark 5.2.2 ([115]) *For standard SIP the formula (5.10) simplifies to*

$$D\varphi_i(x) = D_x g(x, y^i(x)).$$

5.2.1.2 The generalized Hessian of the lower level optimal value function

The definitions are from our recent paper [115]. Let the Reduction Ansatz without strict complementarity hold at $\bar{x} \in M$, and let $\bar{y} \in Y_0(\bar{x})$. According to Theorem 5.2.1, in the case of violated strict complementarity in the lower level, the functions $y(\cdot)$ and $\gamma(\cdot)$ are at least Lipschitz continuous, so that also $D\varphi(x) = D_x\mathcal{L}(x, y(x), \gamma(x))$ is at least Lipschitz continuous. Moreover, due to a result in [52], the function $\varphi(x) = g(x, y(x))$ is not only C^1 with Lipschitz continuous gradient $D\varphi(x)$, but it is even piecewise C^2. The generalized Hessian for φ can thus be defined as the convex hull of the Hessians on the C^2 pieces.

To define the appropriate C^2 pieces, recall that the Reduction Ansatz with strict complementarity basically means that the 'KKT function'

$$\begin{pmatrix} \nabla_y \mathcal{L}(x, y, \gamma) \\ v_{Q_0(\bar{x},\bar{y})}(x, y) \end{pmatrix}$$

has a nonsingular Jacobian with respect to (y, γ) at $(\bar{x}, \bar{y}, \bar{\gamma})$. Here, $v_{Q_0(\bar{x},\bar{y})}$ denotes the vector of functions v_j with $j \in Q_0(\bar{x}, \bar{y})$. Since the inactive constraints v_j, $j \in Q_0^c(\bar{x}, \bar{y})$, locally remain inactive for continuity reasons, the corresponding multipliers γ_j vanish identically. Thus, the above function may as well be extended to

$$\begin{pmatrix} \nabla_y \mathcal{L}(x, y, \gamma) \\ v_{Q_0(\bar{x},\bar{y})}(x, y) \\ -\gamma_{Q_0^c(\bar{x},\bar{y})} \end{pmatrix}.$$

For our subsequent arguments it will be useful to rewrite this function as

$$\mathcal{G}(x, y, \gamma; \alpha, \beta) \;=\; \begin{pmatrix} \nabla_y \mathcal{L}(x, y, \gamma) \\ \mathrm{diag}(\alpha) v(x, y) - \mathrm{diag}(\beta)\gamma \end{pmatrix}$$

with

$$\begin{aligned} (\alpha_j, \beta_j) &= (1, 0), \; j \in Q_0(\bar{x}, \bar{y}), \\ (\alpha_j, \beta_j) &= (0, 1), \; j \in Q_0^c(\bar{x}, \bar{y}). \end{aligned}$$

If the Reduction Ansatz without strict complementarity holds at $\bar{x}$, the index set Q is further partitioned into the three sets

$$\begin{aligned} Q_0^0(\bar{x}, \bar{y}) &= \{j \in Q |\; v_j(\bar{x}, \bar{y}) = 0, \; \bar{\gamma}_j = 0\}, \\ Q_0^+(\bar{x}, \bar{y}) &= \{j \in Q |\; v_j(\bar{x}, \bar{y}) = 0, \; \bar{\gamma}_j > 0\}, \\ Q_0^c(\bar{x}, \bar{y}) &= \{j \in Q |\; v_j(\bar{x}, \bar{y}) < 0, \; \bar{\gamma}_j = 0\}. \end{aligned}$$

With q_0 denoting the cardinality of $Q_0^0(\bar{x},\bar{y})$, we may now consider the 2^{q_0} auxiliary KKT systems

$$\mathcal{G}(x,y,\gamma;\alpha,\beta) \;=\; 0, \quad (\alpha,\beta)\in A,$$

with $(\alpha,\beta)\in A = A(\bar{x},\bar{y})$ if and only if

$$\left.\begin{array}{lcll} (\alpha_j,\beta_j) & \in & \{(1,0),(0,1)\}, & j\in Q_0^0(\bar{x},\bar{y}),\\ (\alpha_j,\beta_j) & = & (1,0), & j\in Q_0^+(\bar{x},\bar{y}),\\ (\alpha_j,\beta_j) & = & (0,1), & j\in Q_0^c(\bar{x},\bar{y}). \end{array}\right\} \tag{5.11}$$

The results in the next lemma are well known even in a more general context [5]. To clarify the connections to our subsequent arguments we state them for our setting and briefly sketch the proofs.

Lemma 5.2.3 ([115]) *Suppose that conditions (Q^+-I) and (Q^+-II) are satisfied at $\bar{y}\in Y_0(\bar{x})$ in $Q(\bar{x})$, where all defining functions of GSIP are at least twice continuously differentiable. Then the following assertions hold:*

(i) For each $(\alpha,\beta)\in A$, the Jacobian

$$D_{(y,\gamma)}\mathcal{G}(\bar{x},\bar{y},\bar{\gamma};\alpha,\beta) \;=\; \begin{pmatrix} \nabla_y^2\mathcal{L} & -\nabla_y v\\ \operatorname{diag}(\alpha)D_y v & -\operatorname{diag}(\beta)\end{pmatrix}$$

is nonsingular, where in the right hand side the obvious arguments have been dropped for ease of notation.

(ii) For each $(\alpha,\beta)\in A$, there exist locally defined C^1 functions $y(x;\alpha,\beta)$ and $\gamma(x;\alpha,\beta)$ which are the locally unique zeros of $\mathcal{G}(x,y,\gamma;\alpha,\beta)$ around $(\bar{x},\bar{y},\bar{\gamma})$. The gradient of $\varphi(x;\alpha,\beta) = g(x,y(x;\alpha,\beta))$ is given by $\nabla\varphi(x;\alpha,\beta) = \nabla_x\mathcal{L}(x,y(x;\alpha,\beta),\gamma(x;\alpha,\beta))$, and $\nabla\varphi(\bar{x};\alpha,\beta) = \nabla_x\mathcal{L}(\bar{x},\bar{y},\bar{\gamma})$ does not depend on (α,β).

(iii) For each $(\alpha,\beta)\in A$, the Hessian of $\varphi(x;\alpha,\beta) = g(x,y(x;\alpha,\beta))$ at $\bar{x}$ exists and is given by

$$D^2\varphi(\bar{x};\alpha,\beta) \;=\; \nabla_x^2\mathcal{L} - \begin{pmatrix}\nabla_{xy}^2\mathcal{L}\\ -D_x v\end{pmatrix}^T S(\alpha,\beta)\begin{pmatrix}\nabla_{xy}^2\mathcal{L}\\ -D_x v\end{pmatrix}$$

with

$$S(\alpha,\beta) \;=\; \begin{pmatrix} \nabla_y^2\mathcal{L} & -\nabla_y v\\ \operatorname{diag}(\alpha)D_y v & -\operatorname{diag}(\beta))\end{pmatrix}^{-1}\begin{pmatrix} I & 0\\ 0 & -\operatorname{diag}(\alpha)\end{pmatrix}. \tag{5.12}$$

(iv) The generalized Hessian of φ at $\bar{x}$ is

$$\partial\nabla\varphi(\bar{x}) \;=\; \operatorname{conv}\left(\left\{\nabla_x^2\mathcal{L} - \begin{pmatrix}\nabla_{xy}^2\mathcal{L}\\ -D_xv\end{pmatrix}^T S(\alpha,\beta)\begin{pmatrix}\nabla_{xy}^2\mathcal{L}\\ -D_xv\end{pmatrix}\,\middle|\,(\alpha,\beta)\in A\right\}\right),$$

where conv *denotes the convex hull.*

Proof. Part (i) is well known (see, e.g., [52]). It also follows from the more general arguments of the subsequent Lemma 5.2.6. Part (i) and the implicit function theorem yield the existence of locally unique C^1 functions $y(x;\alpha,\beta)$ and $\gamma(x;\alpha,\beta)$ with $\mathcal{G}(x,y(x;\alpha,\beta),\gamma(x;\alpha,\beta)) \equiv 0$. Taking first derivatives of the latter equation leads to

$$\begin{pmatrix}Dy(\bar{x};\alpha,\beta)\\ D\gamma(\bar{x};\alpha,\beta)\end{pmatrix} \;=\; -\left(D_{(y,\gamma)}\mathcal{G}\right)^{-1} D_x\mathcal{G} \;=\; -S(\alpha,\beta)\begin{pmatrix}\nabla_{xy}^2\mathcal{L}\\ -D_xv\end{pmatrix}. \tag{5.13}$$

Furthermore, standard techniques of parametric optimization [45] yield that $\varphi(x;\alpha,\beta)$ is a C^2 function with

$$\nabla\varphi(x;\alpha,\beta) \;=\; \nabla_x\mathcal{L}(x,y(x;\alpha,\beta),\gamma(x;\alpha,\beta)) \tag{5.14}$$

locally around $\bar{x}$. Since the implicit functions satisfy $y(\bar{x};\alpha,\beta) = \bar{y}$ and $\gamma(\bar{x};\alpha,\beta) = \bar{\gamma}$, the assertion of part (ii) follows.

Taking derivatives of (5.14) and plugging in (5.13) leads to the assertion of part (iii). Part (iv) immediately follows from (iii) and the fact that $y(x)$ and $\gamma(x)$ are pieced together from $y(x;\alpha,\beta)$ and $\gamma(x;\alpha,\beta)$, respectively, with $(\alpha,\beta)\in A$ [52]. •

Remark 5.2.4 ([115]) *For standard SIP the formula in Lemma 5.2.3(iv) reduces to*

$$\partial\nabla\varphi(\bar{x}) \;=\; \operatorname{conv}\left(\left\{\nabla_x^2 g - \begin{pmatrix}\nabla_{xy}^2 g\\ 0\end{pmatrix}^T S(\alpha,\beta)\begin{pmatrix}\nabla_{xy}^2 g\\ 0\end{pmatrix}\,\middle|\,(\alpha,\beta)\in A\right\}\right),$$

and $S(\alpha,\beta)$ may as well be defined as

$$S(\alpha,\beta) \;=\; \begin{pmatrix}\nabla_y^2\mathcal{L} & -\nabla_y v\\ \operatorname{diag}(\alpha)D_y v & -\operatorname{diag}(\beta)\end{pmatrix}^{-1}.$$

5.2.1.3 A generalized Robinson condition for the upper level problem

In this subsection we derive a generalized Robinson condition for the upper level problem as in our recent paper [115]. Let $P_0(\bar{x}) = \{i\in P \mid \varphi_i(\bar{x}) = 0\}$ be the set of active indices at $\bar{x}$ for the upper level problem.

An appropriate *generalized Robinson condition* is said to hold at $\bar{x}$ if the following conditions ($GSIP^+$-I) and ($GSIP^+$-II) are satisfied:

($GSIP^+$-I) The *linear independence constraint qualification*:

$$(LI)_{GSIP} \quad \{\nabla_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) \,|\, i \in P_0(\bar{x})\ \} \text{ is a linearly independent family.}$$

If $\bar{x}$ is a local minimizer, then there exists a unique vector $\bar{\mu} \in \mathbb{R}^p$ of Lagrange multipliers with

$$\begin{aligned} (KKT)_{GSIP} \quad \nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) &= 0 \\ \bar{\mu}_i \geq 0,\ g(\bar{x}, \bar{y}^i) \leq 0,\ \bar{\mu}_i g(\bar{x}, \bar{y}^i) &= 0 \ (i \in P). \end{aligned}$$

($GSIP^+$-II) The *generalized strong second order sufficiency condition*:

$$(GSSOSC)_{GSIP} \quad \left\{ \begin{array}{c} \xi^T W \xi > 0 \text{ for all } \xi \in G_{GSIP} \setminus \{0\} \text{ and for all } W \text{given in (5.15)} \\ \text{with } G_{GSIP} := \{d \in \mathbb{R}^n |\ D_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) d = 0 \text{ for } i \in P_0^+(\bar{x})\} \end{array} \right\}$$

with $P_0^+(\bar{x}) := \{i \in P_0(\bar{x}) \,|\, \bar{\mu}_i > 0\}$. Here,

$$W \in D_x^2 f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \partial \nabla \varphi_i(\bar{x}), \tag{5.15}$$

where $\partial \nabla \varphi_i(\bar{x})$ is given by Lemma 5.2.3(iv). Note that, in view of the well known calculus rules for Clarke subdifferentials (see [6]), the expression on the right-hand side of (5.15) is an overestimate for generalized Hessian of the Lagrangian $f(x) + \sum_{i=1}^{p} \mu_i \varphi_i(x)$ of $GSIP_{red}$.

In the case of strict complementary slackness, the generalized Robinson condition obviously coincides with the (standard) Robinson condition.

5.2.1.4 A sufficient condition for CD-regularity

In this subsection we state a sufficient condition for CD-regularity of generalized Jacobian as in our paper [115]. Consider a zero $\bar{z}$ of T from (4.36). The KKT conditions mentioned

in (Q^+-I) and ($GSIP^+$-I) then hold by definition of T. The remaining conditions in the Reduction Ansatz without strict complementarity and in the generalized Robinson condition are algebraic conditions on the involved functions which can also be imposed independently of the fact that we deal with lower level global maximizers and upper level local minimizers. In this sense, we can make the following assumption:

Assumption 5.2.5 ([115]) *The Reduction Ansatz without strict complementarity and the generalized Robinson condition hold at $\bar{z}$.*

For proving the main result of the present article, we need the following lemmata, in which $\mathbb{H}^2$ will denote the nonnegative quadrant in $\mathbb{R}^2$. Note that, according to Lemma A.0.33, the generalized Jacobians at the origin of the convex NCP functions ψ_{FB} and ψ_{min} both are subsets of $-\mathbb{H}^2 \setminus \{0\}$.

For the special case $\psi = \psi_{FB}$ the following result can be found in a stronger version in [12]. It is actually the counterpart of results in [60] for the approach by Kojima's function instead of NCP functions. An elegant connection between Newton methods for both approaches is presented in [83].

Lemma 5.2.6 ([115]) *Suppose that conditions (Q^+-I) and (Q^+-II) are satisfied at $\bar{y} \in Y_0(\bar{x})$ in $Q(\bar{x})$. Let the KKT system of $Q(x)$ be described by*

$$\mathcal{H}(x,y,\gamma) \;=\; \begin{pmatrix} \nabla_y g(x,y) - \sum\limits_{j=1}^{q} \gamma_j \nabla_y v_j(x,y) \\ \psi(\gamma_1, -v_1(x,y)) \\ \vdots \\ \psi(\gamma_q, -v_q(x,y)) \end{pmatrix} \;=\; 0$$

with some convex NCP function ψ which satisfies $-\partial\psi(0,0) \in \mathbb{H}^2 \setminus \{0\}$. Then the following assertions hold:

(i) The generalized Jacobian of $\mathcal{H}$ at $(\bar{x},\bar{y},\bar{\gamma})$ with respect to (y,γ) satisfies

$$\partial\mathcal{H}_{(y,\gamma)}(\bar{x},\bar{y},\bar{\gamma}) \;\subseteq\; \{\mathcal{B}(\alpha,\beta) |\; (\alpha,\beta) \in B\}$$

with

$$\mathcal{B}(\alpha,\beta) \;=\; \begin{pmatrix} \nabla_y^2 \mathcal{L} & -\nabla_y v \\ \operatorname{diag}(\alpha) D_y v & -\operatorname{diag}(\beta) \end{pmatrix}$$

and $(\alpha,\beta) \in B = B(\bar{x},\bar{y})$ *if and only if*

$$\left.\begin{array}{rcll} (\alpha_j,\beta_j) & \in & -\partial\psi(0,0), & j \in Q_0^0(\bar{x},\bar{y}), \\ (\alpha_j,\beta_j) & = & (1,0), & j \in Q_0^+(\bar{x},\bar{y}), \\ (\alpha_j,\beta_j) & = & (0,1), & j \in Q_0^c(\bar{x},\bar{y}). \end{array}\right\}. \tag{5.16}$$

(ii) All elements of $\partial\mathcal{H}_{(y,\gamma)}(\bar{x},\bar{y},\bar{\gamma})$ *are nonsingular.*

Proof. Due to the violated strict complementarity, exactly the components $\psi(\gamma_j, -v_j(x,y))$, $j \in Q_0^0(\bar{x},\bar{y})$, of $\mathcal{H}$ are nondifferentiable at $(\bar{x},\bar{y},\bar{\gamma})$. Their generalized Jacobians can be computed by Chain Rule II and Proposition 2.3.6 from [6], noting that convex functions are regular. In fact, the generalized Jacobian with respect to (y,γ_j) is

$$\left\{\left(\alpha_j D_y v_j(x,y)\,,\ -\beta_j\right) \mid (\alpha_j,\beta_j) \in -\partial\psi(0,0)\right\}.$$

In view of (A.3) in Appendix A, the Jacobian of $\psi(\gamma_j, -v_j(x,y))$ with $j \in Q_0^+ \cup Q_0^c$ is

$$\left(\alpha_j D_y v_j(x,y)\,,\ -\beta_j\right)$$

with

$$(\alpha_j,\beta_j) \ = \ \begin{cases} (1,0), & j \in Q_0^+(\bar{x},\bar{y}) \\ (0,1), & j \in Q_0^c(\bar{x},\bar{y}) \end{cases}$$

This shows part (i). To see part (ii), choose any $(\alpha,\beta) \in B$. We will show that $\mathcal{B}(\alpha,\beta)$ is nonsingular, entailing the assertion in view of part (i).

In the following, α_0 denotes the vector with entries α_j, $j \in Q_0^0(\bar{x},\bar{y})$, and V_0 the matrix with columns $-\nabla_y v_j(\bar{x},\bar{y})$, $j \in Q_0^0(\bar{x},\bar{y})$, etc. After rearranging its rows and columns, $\mathcal{B}(\alpha,\beta)$ is nonsingular if and only if

$$\begin{pmatrix} \nabla_y^2\mathcal{L} & V_+ & V_0 & V_c \\ -\mathrm{diag}(\alpha_+)V_+^T & -\mathrm{diag}(\beta_+) & 0 & 0 \\ -\mathrm{diag}(\alpha_0)V_0^T & 0 & -\mathrm{diag}(\beta_0) & 0 \\ -\mathrm{diag}(\alpha_c)V_c^T & 0 & 0 & -\mathrm{diag}(\beta_c) \end{pmatrix}$$

is nonsingular. Due to $\alpha_c = 0, \beta_+ = 0, \alpha_+ = e$ and $\beta_c = e$, the all ones vector, the latter matrix is nonsingular if and only if

$$\begin{pmatrix} \nabla_y^2\mathcal{L} & V_+ & V_0 \\ -V_+^T & 0 & 0 \\ -\mathrm{diag}(\alpha_0)V_0^T & 0 & -\mathrm{diag}(\beta_0) \end{pmatrix}$$

is nonsingular. Let α_0^+ be the subvector of α_0 consisting of nonvanishing entries, if any, and α_0^0 the subvector of vanishing entries, if any. The matrices V_0^+ and V_0^0 are defined accordingly. As $0 \notin -\partial\psi(0,0)$, the matrix $-\mathrm{diag}(\beta_0^0)$ is nonsingular. Along with $\alpha_0^0 = 0$, and after an elementary row transformation, we obtain that the above matrix is nonsingular if and only if

$$\begin{pmatrix} \nabla_y^2 \mathcal{L} & V_+ & V_0^+ \\ V_+^T & 0 & 0 \\ (V_0^+)^T & 0 & \mathrm{diag}(\alpha_0^+)^{-1}\mathrm{diag}(\beta_0^+) \end{pmatrix}$$

is nonsingular. Since the columns of V_+ are linearly independent by (Q^+-I), and $\nabla_y^2 \mathcal{L}$ is negative definite on $\mathrm{Ker}(V_+^T)$ by (Q^+-II), the block matrix

$$\begin{pmatrix} \nabla_y^2 \mathcal{L} & V_+ \\ V_+^T & 0 \end{pmatrix}$$

is nonsingular by Lemma B.0.37. We can thus take its Schur complement (see Definition B.0.34) in the latter matrix and, in view of Lemma B.0.35, obtain that the above matrix is nonsingular if and only if

$$\mathrm{diag}(\alpha_0^+)^{-1}\mathrm{diag}(\beta_0^+) - \begin{pmatrix} V_0^+ \\ 0 \end{pmatrix}^T \begin{pmatrix} \nabla_y^2 \mathcal{L} & V_+ \\ V_+^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} V_0^+ \\ 0 \end{pmatrix}$$

is nonsingular. The matrix $\mathrm{diag}(\alpha_0^+)^{-1}\mathrm{diag}(\beta_0^+)$ is a diagonal matrix with nonnegative diagonal entries, due to $-\partial\psi(0,0) \subset \mathbb{H}^2$. In the following we will show that

$$W = -\begin{pmatrix} V_0^+ \\ 0 \end{pmatrix}^T \begin{pmatrix} \nabla_y^2 \mathcal{L} & V_+ \\ V_+^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} V_0^+ \\ 0 \end{pmatrix}$$

is positive definite, so that also the above matrix is positive definite, showing the nonsingularity of $\mathcal{B}(\alpha, \beta)$.

In fact, we may write

$$W = \begin{pmatrix} \nabla_y^2 \mathcal{L} & V_+ & V_0^+ \\ V_+^T & 0 & 0 \\ (V_0^+)^T & 0 & 0 \end{pmatrix} / \begin{pmatrix} \nabla_y^2 \mathcal{L} & V_+ \\ V_+^T & 0 \end{pmatrix},$$

so that a combination of Lemma B.0.36 and Lemma B.0.37 (here, $In(A)$ denotes the inertia-triple of A, that is, the number of negative, positive and vanishing eigenvalues of A,

respectively) yields

$$\begin{aligned}\mathrm{In}(W) &= \mathrm{In}\begin{pmatrix} \nabla_y^2\mathcal{L} & V_+ & V_0^+ \\ V_+^T & 0 & 0 \\ (V_0^+)^T & 0 & 0 \end{pmatrix} - \mathrm{In}\begin{pmatrix} \nabla_y^2\mathcal{L} & V_+ \\ V_+^T & 0 \end{pmatrix} \\ &= \mathrm{In}(\nabla_y^2\mathcal{L}|_{\mathrm{Ker}((V_+,V_0^+)^T)}) + (q_+ + q_0^+, q_+ + q_0^+, 0) - (\mathrm{In}(\nabla_y^2\mathcal{L}|_{\mathrm{Ker}(V_+^T)}) + (q_+, q_+, 0)),\end{aligned}$$

where q_+ denotes the number of columns of V_+, etc. As $\nabla_y^2\mathcal{L}$ is negative definite on $\mathrm{Ker}(V_+^T)$, it is also negative definite on the smaller linear space $\mathrm{Ker}((V_+, V_0^+)^T)$, so that

$$\begin{aligned}\mathrm{In}(\nabla_y^2\mathcal{L}|_{\mathrm{Ker}(V_+^T)}) &= (m - q_+, 0, 0), \\ \mathrm{In}(\nabla_y^2\mathcal{L}|_{\mathrm{Ker}((V_+,V_0^+)^T)}) &= (m - q_+ - q_0^+, 0, 0).\end{aligned}$$

We thus arrive at

$$\mathrm{In}(W) = (0, q_0^+, 0),$$

showing the assertion of part (ii). •

The following is the central result on which the proof of the subsequent Theorem 5.2.8 is based.

Lemma 5.2.7 ([115]) *Suppose that conditions (Q^+-I) and (Q^+-II) hold at $\bar{y} \in Y_0(\bar{x})$ in $Q(\bar{x})$, and let ψ be some NCP function which satisfies $-\partial\psi(0,0) \in \mathbb{H}^2 \setminus \{0\}$, where $\mathbb{H}^2$ denotes nonnegative quadrant. Let $S(\alpha,\beta)$ denote the matrix from (5.12), and let the sets A and B be defined by the conditions in (5.11) and (5.16), respectively. Then the inclusion*

$$\{\, S(\alpha,\beta) \mid (\alpha,\beta) \in B \,\} \subset \mathrm{conv}(\{\, S(\alpha,\beta) \mid (\alpha,\beta) \in A \,\})$$

holds.

Proof. The assertion is trivial for $Q_0^0(\bar{x},\bar{y}) = \emptyset$. Thus let $Q_0^0(\bar{x},\bar{y}) \neq \emptyset$ and choose some $j \in Q_0^0(\bar{x},\bar{y})$. Without loss of generality we assume $q \in Q_0^0(\bar{x},\bar{y})$ and $j = q$. Since otherwise, i.e., if $q \notin Q_0^0(\bar{x},\bar{y})$, either $q \in Q_0^+(\bar{x},\bar{y})$ or $q \in Q_0^c(\bar{x},\bar{y})$, in both cases the result is trivial.

Let $(\alpha,\beta) \in B$ be given. For $(\sigma,\tau) \in -\partial\psi(0,0)$ we define the functions $\alpha(\sigma)$ and $\beta(\tau)$ by $\alpha_j(\sigma) \equiv \alpha_j$, $\beta_j(\tau) \equiv \beta_j$ for $j < q$ and $\alpha_q(\sigma) = \sigma$, $\beta_q(\tau) = \tau$. To show the assertion, it is sufficient to prove it in the single component q, that is,

$$S(\,\alpha(\sigma),\beta(\tau)\,) \subset \mathrm{conv}(\{\, S(\alpha(1),\beta(0)),\, S(\alpha(0),\beta(1)) \,\}). \tag{5.17}$$

In fact, let $(\alpha^1,\beta^1) = (\alpha(1),\beta(0))$ and $(\alpha^2,\beta^2) = (\alpha(0),\beta(1))$. Then $(\alpha^1_q,\beta^1_q) = (1,0)$ and $(\alpha^2_q,\beta^2_q) = (0,1)$ are identical to the corresponding entries in A, respectively. In the case $Q^0_0(\bar{x},\bar{y}) = \{q\}$ this shows the assertion. If $Q^0_0(\bar{x},\bar{y})$ contains another index, we assume without loss of generality that it is $j = q-1$, since otherwise by a change of rows and columns we can make the index $j = q-1$. Using (5.17) for the $(q-1)$-th entry of (α^1,β^1) we obtain that $S(\alpha^1,\beta^1)$ lies in the convex hull of two matrices $S(\alpha^{1,1},\beta^{1,1})$ and $S(\alpha^{1,2},\beta^{1,2})$ with $(\alpha^{1,1}_q,\beta^{1,1}_q) = (\alpha^{1,2}_q,\beta^{1,2}_q) = (1,0)$ and $(\alpha^{1,1}_{q-1},\beta^{1,1}_{q-1}) = (1,0)$, $(\alpha^{1,2}_{q-1},\beta^{1,2}_{q-1}) = (0,1)$. Analogously, $S(\alpha^2,\beta^2)$ lies in the convex hull of two matrices $S(\alpha^{2,1},\beta^{2,1})$ and $S(\alpha^{2,2},\beta^{2,2})$ with $(\alpha^{2,1}_q,\beta^{2,1}_q) = (\alpha^{2,2}_q,\beta^{2,2}_q) = (0,1)$ and $(\alpha^{2,1}_{q-1},\beta^{2,1}_{q-1}) = (1,0)$, $(\alpha^{2,2}_{q-1},\beta^{2,2}_{q-1}) = (0,1)$. Combining these facts, $S(\alpha,\beta)$ is an element of the convex hull of the four matrices $S(\alpha^{1,1},\beta^{1,1})$, $S(\alpha^{1,2},\beta^{1,2})$, $S(\alpha^{2,1},\beta^{2,1})$ and $S(\alpha^{2,2},\beta^{2,2})$. Continuing this argument for all elements of $Q^0_0(\bar{x},\bar{y})$ yields the assertion.

To see (5.17) we define

$$\begin{aligned} b &= \begin{pmatrix} -\nabla_y v_q \\ 0 \end{pmatrix}, \\ C &= \begin{pmatrix} \nabla^2_y \mathcal{L} & -\nabla_y v_{Q\setminus\{q\}} \\ \operatorname{diag}(\alpha_1,...,\alpha_{q-1}) D_y v_{Q\setminus\{q\}} & -\operatorname{diag}(\beta_1,...,\beta_{q-1}) \end{pmatrix}, \\ D &= \begin{pmatrix} I & 0 \\ 0 & -\operatorname{diag}(\alpha_1,...,\alpha_{q-1}) \end{pmatrix}. \end{aligned}$$

Then we have

$$S(\,\alpha(\sigma),\beta(\tau)\,) = \begin{pmatrix} C & b \\ -\sigma b^T & -\tau \end{pmatrix}^{-1} \begin{pmatrix} D & 0 \\ 0 & -\sigma \end{pmatrix}.$$

Since the conditions (Q^+-I) and (Q^+-II) at $\bar{y} \in Y_0(\bar{x})$ do not change when the restriction $v_q(x,y) \le 0$ is removed from problem $Q(x)$ (recall that $\bar{\gamma}_q = 0$), the matrix C is nonsingular in view of Lemma 5.2.6(ii). We may thus use Lemma B.0.38 to compute

$$S(\,\alpha(\sigma),\beta(\tau)\,) = \begin{pmatrix} C^{-1}D & 0 \\ 0 & 0 \end{pmatrix} + \frac{\sigma}{\sigma\, b^T C^{-1} b - \tau} \begin{pmatrix} C^{-1} b b^T C^{-1} D & -C^{-1} b \\ -b^T C^{-1} D & 1 \end{pmatrix}.$$

In this formulation it is easy to see that (5.17) is equivalent to

$$\frac{\sigma\, b^T C^{-1} b}{\sigma\, b^T C^{-1} b - \tau} \in [0,1].$$

Since $(\sigma,\tau) \in -\partial\psi(0,0) \subset \mathbb{H}^2 \setminus \{0\}$, the latter follows if $b^T C^{-1} b < 0$.

In order to show negativity of $b^TC^{-1}b$ we interpret this number as a Schur complement. In fact, thanks to the zero part in the vector b, the same matrix manipulations as in the proof of Lemma 5.2.6(ii) may be used to rewrite $b^TC^{-1}b$.

Define $\overline{\alpha} = (\alpha_1, ..., \alpha_{q-1}, 1)^T$ and $\overline{\beta} = (\beta_1, ..., \beta_{q-1}, 0)^T$. Let $\overline{\alpha}_0^+$ stand again for the vector with entries $\overline{\alpha}_j$, $j \in Q_0^0(\overline{x}, \overline{y}) \cap \{j | \overline{\alpha}_j > 0\}$, and note that $j = q$ is among these indices. Let $\tilde{\alpha}_0^+$ denote the vector with entries $\overline{\alpha}_j$, $j \in Q_0^0(\overline{x}, \overline{y}) \cap \{j | \overline{\alpha}_j > 0\} \setminus \{q\}$, let the matrices V_0^+ and $\tilde{V}_0^+$ be defined accordingly, put $v = -\nabla_y v_q(\overline{x}, \overline{y})$, and let V_+ stand again for $-\nabla_y v_{Q_0^+(\overline{x},\overline{y})}(\overline{x}, \overline{y})$.

Then, using the mentioned matrix manipulations, one can show

$$b^TC^{-1}b \;=\; \begin{pmatrix} v \\ 0 \\ 0 \end{pmatrix}^T \begin{pmatrix} \nabla_y^2\mathcal{L} & V_+ & \tilde{V}_0^+ \\ V_+^T & 0 & 0 \\ (\tilde{V}_0^+)^T & 0 & \mathrm{diag}(\tilde{\alpha}_0^+)^{-1}\mathrm{diag}(\tilde{\beta}_0^+) \end{pmatrix}^{-1} \begin{pmatrix} v \\ 0 \\ 0 \end{pmatrix}$$

so that

$$-b^TC^{-1}b \;= \begin{pmatrix} \nabla_y^2\mathcal{L} & V_+ & V_0^+ \\ V_+^T & 0 & 0 \\ (V_0^+)^T & 0 & \mathrm{diag}(\alpha_0^+)^{-1}\mathrm{diag}(\beta_0^+) \end{pmatrix} / \begin{pmatrix} \nabla_y^2\mathcal{L} & V_+ & \tilde{V}_0^+ \\ V_+^T & 0 & 0 \\ (\tilde{V}_0^+)^T & 0 & \mathrm{diag}(\tilde{\alpha}_0^+)^{-1}\mathrm{diag}(\tilde{\beta}_0^+) \end{pmatrix}.$$

In view of Lemma B.0.36 we obtain

$$\mathrm{In}(-b^TC^{-1}b) \;= \mathrm{In}\begin{pmatrix} \nabla_y^2\mathcal{L} & V_+ & V_0^+ \\ V_+^T & 0 & 0 \\ (V_0^+)^T & 0 & \mathrm{diag}(\alpha_0^+)^{-1}\mathrm{diag}(\beta_0^+) \end{pmatrix} - \mathrm{In}\begin{pmatrix} \nabla_y^2\mathcal{L} & V_+ & \tilde{V}_0^+ \\ V_+^T & 0 & 0 \\ (\tilde{V}_0^+)^T & 0 & \mathrm{diag}(\tilde{\alpha}_0^+)^{-1}\mathrm{diag}(\tilde{\beta}_0^+) \end{pmatrix},$$

and in the proof of Lemma 5.2.6(ii) we have seen that the Schur complements of $\begin{pmatrix} \nabla_y^2\mathcal{L} & V_+ \\ V_+^T & 0 \end{pmatrix}$ in both matrices of the latter right hand side are positive definite. According to Lemma B.0.36 we arrive at

$$\mathrm{In}(-b^TC^{-1}b) \;=\; (0, q_0^+, 0) - (0, \tilde{q}_0^+, 0) \;=\; (0, 1, 0)$$

which implies negativitiy of $b^TC^{-1}b$, and hence we prove our main assertion about inclusion.

•

Theorem 5.2.8 ([115]) *Suppose that $\overline{z} = (\overline{x}^T, \overline{\mu}^T, (\overline{y}^1)^T, (\overline{\gamma}^1)^T, \ldots, (\overline{y}^p)^T, (\overline{\gamma}^p)^T)^T$ is a zero of T from (4.36) with the choices $\psi = \psi_{FB}$ or $\psi = \psi_{min}$, and that Assumption 5.2.5 holds at $\overline{z}$. Then T is CD-regular at $\overline{z}$.*

Proof. Let $\bar{z}$ be a zero of T at which Assumption 5.2.5 holds, and let ψ denote either one of the two NCP functions ψ_{FB} and ψ_{min}. We only consider the case of two active indices, that is, $p = 2$, the general case running along the same lines. Recall that the case of satisfied lower level strict complementarity is treated in [108]. The generalized Jacobian of T at $\bar{z}$ is a subset of Cartesian product of generalized gradients of component functions as computed in Lemma 5.2.6(i). In fact, the elements of $\partial T(\bar{z})$ are of the form

$$\mathcal{A}(\lambda,\theta,\alpha^1,\beta^1,\alpha^2,\beta^2) = \begin{pmatrix} \mathcal{B}_{11}(\lambda,\theta) & \mathcal{B}_{12}(\lambda,\theta) & \mathcal{B}_{13}(\lambda,\theta) \\ \mathcal{B}_{21}(\alpha^1,\beta^1) & \mathcal{B}_{22}(\alpha^1,\beta^1) & 0 \\ \mathcal{B}_{31}(\alpha^2,\beta^2) & 0 & \mathcal{B}_{33}(\alpha^2,\beta^2) \end{pmatrix} \tag{5.18}$$

with the following blocks:

$$\mathcal{B}_{11}(\lambda,\theta) = \begin{pmatrix} D_x^2 f(\bar{x}) + \sum_{i=1}^{2} \bar{\mu}_i \nabla_x^2 \mathcal{L}^i & \nabla_x \mathcal{L}^1 & \nabla_x \mathcal{L}^2 \\ \lambda_1 D_x g^1 & -\theta_1 & 0 \\ \lambda_2 D_x g^2 & 0 & -\theta_2 \end{pmatrix}$$

with $\nabla_x^2 \mathcal{L}^i = \nabla_x^2 \mathcal{L}(\bar{x},\bar{y}^i,\bar{\gamma}^i)$, etc., and with $(\lambda,\theta) \in \Theta$ if and only if

$$\begin{aligned} (\lambda_i,\theta_i) &\in -\partial\psi(0,0), && i \in P_0^0(\bar{x}), \\ (\lambda_i,\theta_i) &= (1,0), && i \in P_0^+(\bar{x}), \\ (\lambda_i,\theta_i) &= (0,1), && i \in P_0^c(\bar{x}), \end{aligned}$$

$$\mathcal{B}_{12}(\lambda,\theta) = \begin{pmatrix} \bar{\mu}_1 \nabla_{xy}^2 \mathcal{L}^1 & -\bar{\mu}_1 \nabla_x v^1 \\ \lambda_1 D_y g^1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\mathcal{B}_{13}(\lambda,\theta) = \begin{pmatrix} \bar{\mu}_2 \nabla_{xy}^2 \mathcal{L}^2 & -\bar{\mu}_2 \nabla_x v^2 \\ 0 & 0 \\ \lambda_2 D_y g^2 & 0 \end{pmatrix},$$

$$\mathcal{B}_{i+1,1}(\alpha^i,\beta^i) = \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}^i & 0 & 0 \\ \mathrm{diag}(\alpha^i) D_x v^i & 0 & 0 \end{pmatrix} \quad (i = 1,2),$$

as well as

$$\mathcal{B}_{i+1,i+1}(\alpha^i,\beta^i) = \begin{pmatrix} \nabla_y^2 \mathcal{L}^i & -\nabla_y v^i \\ \mathrm{diag}(\alpha^i) D_y v^i & -\mathrm{diag}(\beta^i) \end{pmatrix} \quad (i = 1,2),$$

with $(\alpha^i,\beta^i) \in B^i$, where $B^i = B(\bar{x},\bar{y}^i)$ is chosen according to (5.16), $i = 1,2$. Our aim is to show CD-regularity of T at $\bar{z}$. For this aim we will show nonsingularity of all matrices

$\mathcal{A}(\lambda,\theta,\alpha^1,\beta^1,\alpha^2,\beta^2)$ with $(\lambda,\theta)\in\Theta$, $(\alpha^i,\beta^i)\in B^i$ $(i=1,2)$. The main idea of the proof is the same as in [108], namely to consider an appropriate Schur complement in the block matrix $\mathcal{A}(\lambda,\theta,\alpha^1,\beta^1,\alpha^2,\beta^2)$.

Choose arbitrary $(\bar\lambda,\bar\theta)\in\Theta$ and $(\bar\alpha^i,\bar\beta^i)\in B^i$, $i=1,2$. For conciseness, let us write $\mathcal{B}_{12}=\mathcal{B}_{12}(\bar\lambda,\bar\theta)$, $\mathcal{B}_{22}=\mathcal{B}_{22}(\bar\alpha^1,\bar\beta^1)$, $\mathcal{A}=\mathcal{A}(\bar\lambda,\bar\theta,\bar\alpha^1,\bar\beta^1,\bar\alpha^2,\bar\beta^2)$, and so on. Then, due to Lemma 5.2.6(ii), the matrices $\mathcal{B}_{22}$ and $\mathcal{B}_{33}$ are nonsingular.

According to Lemma B.0.35 the matrix $\mathcal{A}$ is nonsingular if and only if the Schur complement

$$\begin{aligned} R &= \mathcal{A}/\begin{pmatrix} \mathcal{B}_{22} & 0 \\ 0 & \mathcal{B}_{33} \end{pmatrix} \\ &= \mathcal{B}_{11}-(\mathcal{B}_{12},\,\mathcal{B}_{13})\begin{pmatrix} \mathcal{B}_{22} & 0 \\ 0 & \mathcal{B}_{33} \end{pmatrix}^{-1}\begin{pmatrix} \mathcal{B}_{21} \\ \mathcal{B}_{31} \end{pmatrix} \\ &= \mathcal{B}_{11}-\mathcal{B}_{12}\mathcal{B}_{22}^{-1}\mathcal{B}_{21}-\mathcal{B}_{13}\mathcal{B}_{33}^{-1}\mathcal{B}_{31} \end{aligned}$$

is nonsingular. We will show that the latter is the case under the generalized Robinson condition.

When we write down the term $\mathcal{B}_{12}\mathcal{B}_{22}^{-1}\mathcal{B}_{21}$ more explicitly, a matrix of the form $S(\alpha,\beta)$ from (5.12) reappears. In fact, the part $\mathcal{B}_{22}^{-1}\mathcal{B}_{21}$ becomes

$$\begin{pmatrix} \nabla_y^2\mathcal{L}^1 & -\nabla_y v^1 \\ \operatorname{diag}(\bar\alpha^1)D_y v^1 & -\operatorname{diag}(\bar\beta^1) \end{pmatrix}^{-1}\begin{pmatrix} \nabla_{yx}^2\mathcal{L}^1 & 0 & 0 \\ \operatorname{diag}(\bar\alpha^1)D_x v^1 & 0 & 0 \end{pmatrix}$$

$$= S(\bar\alpha^1,\bar\beta^1)\begin{pmatrix} \nabla_{yx}^2\mathcal{L}^1 & 0 & 0 \\ -D_x v^1 & 0 & 0 \end{pmatrix},$$

so that $\mathcal{B}_{12}\mathcal{B}_{22}^{-1}\mathcal{B}_{21}$ coincides with

$$\begin{pmatrix} \bar\mu_1\nabla_{xy}^2\mathcal{L}^1 & -\bar\mu_1\nabla_x v^1 \\ \bar\lambda_1 D_y g^1 & 0 \\ 0 & 0 \end{pmatrix} S(\bar\alpha^1,\bar\beta^1)\begin{pmatrix} \nabla_{yx}^2\mathcal{L}^1 & 0 & 0 \\ -D_x v^1 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \bar\mu_1\begin{pmatrix} \nabla_{yx}^2\mathcal{L}^1 \\ -D_x v^1 \end{pmatrix}^T S(\bar\alpha^1,\bar\beta^1)\begin{pmatrix} \nabla_{yx}^2\mathcal{L}^1 \\ -D_x v^1 \end{pmatrix} & 0 & 0 \\ \bar\lambda_1\begin{pmatrix} \nabla_y g^1 \\ 0 \end{pmatrix}^T S(\bar\alpha^1,\bar\beta^1)\begin{pmatrix} \nabla_{yx}^2\mathcal{L}^1 \\ -D_x v^1 \end{pmatrix} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Rewriting $\mathcal{B}_{13}\mathcal{B}_{33}^{-1}\mathcal{B}_{31}$ analogously we arrive at

$$R = \begin{pmatrix} D_x^2 f(\bar{x}) + \sum_{i=1}^{2} \bar{\mu}_i \Lambda(\bar{\alpha}^i, \bar{\beta}^i) & \nabla_x \mathcal{L}^1 & \nabla_x \mathcal{L}^2 \\ \bar{\lambda}_1 \Gamma(\bar{\alpha}^1, \bar{\beta}^1) & -\bar{\theta}_1 & 0 \\ \bar{\lambda}_2 \Gamma(\bar{\alpha}^2, \bar{\beta}^2) & 0 & -\bar{\theta}_2 \end{pmatrix}$$

with

$$\Lambda(\bar{\alpha}^i, \bar{\beta}^i) = \nabla_x^2 \mathcal{L}^i - \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}^i \\ -D_x v^i \end{pmatrix}^T S(\bar{\alpha}^i, \bar{\beta}^i) \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}^i \\ -D_x v^i \end{pmatrix} \ (i = 1, 2)$$

and

$$\Gamma(\bar{\alpha}^i, \bar{\beta}^i) = D_x g^i - \begin{pmatrix} \nabla_y g^i \\ 0 \end{pmatrix}^T S(\bar{\alpha}^i, \bar{\beta}^i) \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}^i \\ -D_x v^i \end{pmatrix} \ (i = 1, 2).$$

Applying the crucial Lemma 5.2.7, in the following we will show

$$\Lambda(\bar{\alpha}^i, \bar{\beta}^i) \in \partial \nabla \varphi_i(\bar{x}) \ (i = 1, 2) \tag{5.19}$$

and

$$\Gamma(\bar{\alpha}^i, \bar{\beta}^i) = D_x \mathcal{L}^i \ (i = 1, 2). \tag{5.20}$$

In fact, let $A^i = A(\bar{x}, \bar{y}^i)$ $(i = 1, 2)$, be defined according to (5.11). Since $(\bar{\alpha}^i, \bar{\beta}^i) \in B^i$, Lemma 5.2.7 yields

$$S(\bar{\alpha}^i, \bar{\beta}^i) \in \operatorname{conv}(\{ S(\alpha, \beta) | \ (\alpha, \beta) \in A^i \}) \ (i = 1, 2). \tag{5.21}$$

Together with Lemma 5.2.3 (iv), (5.21) implies

$$\begin{aligned} \Lambda(\bar{\alpha}^i, \bar{\beta}^i) &\in \operatorname{conv}\left(\left\{ \nabla_x^2 \mathcal{L}^i - \begin{pmatrix} \nabla_{xy}^2 \mathcal{L}^i \\ -D_x v^i \end{pmatrix}^T S(\alpha, \beta) \begin{pmatrix} \nabla_{xy}^2 \mathcal{L}_i \\ -D_x v^i \end{pmatrix} \middle| \ (\alpha, \beta) \in A^i \right\} \right) \\ &= \partial \nabla \varphi_i(\bar{x}) \ (i = 1, 2), \end{aligned}$$

and thus (5.19).

Furthermore, (5.21) implies

$$\Gamma(\bar{\alpha}^i, \bar{\beta}^i) \in \operatorname{conv}\left(\left\{ D_x g^i - \begin{pmatrix} \nabla_y g^i \\ 0 \end{pmatrix}^T S(\alpha, \beta) \begin{pmatrix} \nabla_{xy}^2 \mathcal{L}_i \\ -D_x v^i \end{pmatrix} \middle| \ (\alpha, \beta) \in A^i \right\} \right) \ (i = 1, 2). \tag{5.22}$$

In view of (5.13), for $i = 1, 2$ and any $(\alpha, \beta) \in A^i$ we may write

$$\begin{aligned} D_x g^i - \begin{pmatrix} \nabla_y g^i \\ 0 \end{pmatrix}^T S(\alpha, \beta) \begin{pmatrix} \nabla_{xy}^2 \mathcal{L}_i \\ -D_x v^i \end{pmatrix} &= D_x g^i + D_y g^i D y^i(\bar{x}; \alpha, \beta) \\ &= D\varphi(\bar{x}; \alpha, \beta), \end{aligned}$$

so that Lemma 5.2.3(ii) yields

$$\Gamma(\bar{\alpha}^i, \bar{\beta}^i) \in \operatorname{conv}\left(\{D_x\mathcal{L}^i(\bar{x}, \bar{y}^i, \bar{\gamma}^i) | (\alpha, \beta) \in A^i\}\right) = \{D_x\mathcal{L}^i\},$$

that is, (5.20).

Under (5.19) and (5.20) we find

$$R \in \begin{pmatrix} D_x^2 f(\bar{x}) + \sum_{i=1}^{2} \bar{\mu}_i \partial\nabla_x \varphi_i(\bar{x}) & \nabla_x\mathcal{L}^1 & \nabla_x\mathcal{L}^2 \\ \bar{\lambda}_1 D_x\mathcal{L}^1 & -\bar{\theta}_1 & 0 \\ \bar{\lambda}_2 D_x\mathcal{L}^2 & 0 & -\bar{\theta}_2 \end{pmatrix}$$

so that there exists some matrix

$$W \in D_x^2 f(\bar{x}) + \sum_{i=1}^{2} \bar{\mu}_i \partial\nabla\varphi_i(\bar{x})$$

with

$$R = \begin{pmatrix} W & \nabla_x\mathcal{L}^1 & \nabla_x\mathcal{L}^2 \\ \bar{\lambda}_1 D_x\mathcal{L}^1 & -\bar{\theta}_1 & 0 \\ \bar{\lambda}_2 D_x\mathcal{L}^2 & 0 & -\bar{\theta}_2 \end{pmatrix}.$$

Under the generalized Robinson condition, the latter matrix can be seen to be nonsingular with the same arguments as in Lemma 5.2.6 (ii) (compare also [85, Theorem 4.2]). So we show that T is CD-regular at $\bar{z}$ where $\bar{z}$ is a zero of T. We get this result under the Reduction Ansatz without strict complementarity and generalized Robinson condition at $\bar{z}$. •

Remark 5.2.9 ([115]) *In the special case of SIP, an explicit proof of Theorem 5.2.8 would be slightly shorter due to the simplifications that* $\nabla_x\mathcal{L}_i$ *reduces to* $\nabla_x g(\bar{x}, \bar{y}^i)$, $\nabla_{xy}\mathcal{L}_i$ *to* $\nabla_{xy} g(\bar{x}, \bar{y}^i)$, *and* $\nabla_x v(\bar{x}, \bar{y}^i)$ *vanishes,* $i \in P$.

In Theorem 5.2.8, we have shown that Assumption 5.2.5 implies CD-regularity of the semismooth reformulation of the upper and lower KKT conditions of *GSIP* at a zero of this reformulation, where Assumption 5.2.5 neither needs strict complementarity in the upper nor in the lower level problem. In view of Theorem 4.1.10, the semismooth Newton method from (4.4) hence converges q-quadratically (because of strong semismoothness of T at $\bar{z}$, see Section 4.3.3) to a zero. For preliminary numerical tests of the semismooth Newton method for *GSIP* see Chapter 6 where sample problems with violated upper and lower level strict complementarity are given.

CHAPTER 6

NUMERICAL RESULTS

In this chapter we give some numerical results from design centering, robust optimization and other examples in order to show performance of the semismooth Newton method. The examples are mainly from our recent paper [108].

We consider also numerical examples, for which the strict complementarity is violated in the upper or in the lower level or in both levels simultaneously. We use the generalized damped semismooth Newton approach proposed in [88].

We want to solve $T(z) = 0$ where the system of nonlinear equations for *SIP* is defined in 4.33 and for *GSIP* it is defined in 4.36 with $z \in \mathbb{R}^N$. Since this is a system of nonlinear equations, we need to define a merit function which is a scalar valued function whose values indicates whether a new candidate iterate is better or worse than the current iterate, in the sense of making progress toward a root of T. The most widely used merit function is the sum of squares, defined by (see [74], [77])

$$\theta(z) = \frac{1}{2} T(z)^T T(z).$$

We use this as a merit function.

If ψ_{FB} is used, then θ is C^1 with the following gradient

$$\nabla\theta(z) = W^T T(z),$$

where $W \in \partial T(z)$, is the generalized Jacobian of T at z. In the case of ψ_{min}, the merit function is not continuously differentiable. We use ψ_{FB} whenever the gradient of the merit function is needed in the algorithm.

Algorithm 6.0.10 ([88])

Step 1. Let $z^0 \in \mathbb{R}^N,\ \sigma, \rho \in (0,1),\ \eta > 0,\ a > 2$ *and* $k = 0$.

Step 2. If $T(z^k) = 0$*, stop. Otherwise, let* d^k *be a solution of*

$$T(z^k) + W^k d = 0, \tag{6.1}$$

where $W^k \in \partial T(z^k)$.

If (6.1) is not solvable, or if

$$\nabla\theta(z^k)^T d^k > -\eta \left\|d^k\right\|^a,$$

set $d^k = -\nabla\theta(z^k)$

Step 3. Find a minimum nonnegative integer, say, m_k*, such that*

$$\theta(z^k + \rho^{m_k} d^k) \leq \theta(z^k) + \sigma\rho^{m_k}\nabla\theta(z^k)^T d^k,$$

Let $\alpha_k = \rho^{m_k}$.

Step 4. Let $z^{k+1} = z^k + \alpha_k d^k$ *and* $k = k + 1$*. Go to Step 2.*

This algorithm is a standard damped Newton algorithm. It uses Newton direction as a search direction d^k if possible, if it is not possible to solve the linear system of equation for d^k and if a standard sufficient decrease condition is not satisfied, it passes to steepest descent direction for the merit function. As a globalization strategy it uses a standard Armijo rule line search method in order to find step size α_k and update the current iterate by the step size α_k and search direction d^k. For the implementation of the algorithm, at iterates where T is differentiable, we do not use the Jacobian matrix DT in the form as in the proof of Theorem 5.1.2, since the iterates cannot be expected to be zeros of T. Thus we may not use the simplified gradients of the NCP functions from (A.3), but the ones from (A.1), (A.2). For ψ_{FB} this results in replacing $\lambda_i,\ \theta_i,\ \alpha^i_j,\ \beta^i_j,\ i \in P,\ j \in Q$, by

$$\lambda_i = \frac{g(\overline{x}, \overline{y}^i)}{\sqrt{\overline{\mu}_i^2 + g(\overline{x}, \overline{y}^i)^2}} + 1,\ \theta_i = \frac{\overline{\mu}_i}{\sqrt{\overline{\mu}_i^2 + g(\overline{x}, \overline{y}^i)^2}} - 1, \tag{6.2}$$

and

$$\alpha^i_j = \frac{v_j(\overline{x}, \overline{y}^i)}{\sqrt{(\overline{\gamma}^i_j)^2 + v_j(\overline{x}, \overline{y}^i)^2}} + 1,\ \beta^i_j = \frac{\overline{\gamma}^i_j}{\sqrt{(\overline{\gamma}^i_j)^2 + v_j(\overline{x}, \overline{y}^i)^2}} - 1. \tag{6.3}$$

whereas for ψ_{min} the gradients in (A.2) and (A.3) coincide.

At nondifferentiability points of T we choose the element W from the generalized Jacobian of T which corresponds to the midpoints of the subdifferentials of the NCP functions. In view of (5.7) and Lemma A.0.33 this means that for ψ_{FB} we use $\lambda_i = 1,\ \theta_i = -1$, and for ψ_{min} we use $\lambda_i = 1/2,\ \theta_i = -1/2$.

The Algorithm 6.0.10 is implemented in MATLAB 7.3. Throughout the computational experiments, the parameters used in the algorithm are $\rho = 0.5,\ a = 2.1,\ \eta = 10^{-8}$ and $\sigma = 0.1$. The algorithm is terminated when $\|T(z^k)\| < 10^{-6}$.

In the numerical examples we test both ψ_{FB} and ψ_{min} as the NCP function. However, due to the mentioned smoothness properties, in the merit function we use ψ_{FB} for both cases. In the ψ_{min} case, we use the gradient of the merit function in the ψ_{FB} case whenever the gradient is needed.

Example 6.0.11 *In the following examples we solve standard semi-infinite optimization problems.*

Problem 6.0.12 ([7])

We consider the following *SIP*:

$$\min\ \ f(x) = 1.21e^{x_1} + e^{x_2}\ \ \text{such that}\ x \in M = \{x \in \mathbb{R}^2 |\ g(x,y) \leq 0,\ y \in Y\},$$

where

$$g(x,y) = y - e^{x_1+x_2}$$

and

$$Y = \{y \in \mathbb{R}\,|\,v(y) = y^2 - y \leq 0\}.$$

In the ψ_{FB} case, with the starting point $x^0 = (0,0)$ the semismooth Newton method obtains the optimal value 2.2 with $\bar{x} = (-0.0953, 0.0953)$ and $\bar{y} = 1$ for the optimal point. We have $\|T(\bar{z})\| = 2.8275^{-12}$ after 6 iterations within 0.032 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 8 iterations with $\|T(\bar{z})\| = 1.4035^{-8}$ within 0.016 seconds of CPU time.

Problem 6.0.13 ([7])

Let us consider the following *SIP*:

$$\min \ \ f(x) = x_1^2 + x_2^2 + x_3^2 \ \ \text{such that } x \in M = \{x \in \mathbb{R}^2 |\ g(x,y) \leq 0,\ y \in Y\},$$

where

$$g(x,y) = x_1 + x_2 e^{x_3 y} + e^{2y} - 2\sin(4y)$$

and

$$Y = \{y \in \mathbb{R} \,|\, v(y) = y^2 - y \leq 0\}.$$

In the ψ_{FB} case, with the starting point $x^0 = (1,1,1)$ the semismooth Newton method obtains the optimal value 5.33 with $\bar{x} = (-0.213, -1.361, 1.854)$ and $\bar{y} = 1$ for the optimal point. We have $||T(\bar{z})|| = 4.6578^{-11}$ after 12 iterations within 0.03 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 9 iterations with $||T(\bar{z})|| = 8.2055^{-7}$ within again 0.03 seconds of CPU time.

Problem 6.0.14 ([7])

We consider the following *SIP*:

$$\min \ \ f(x) = (x_1 - 2x_2 + 5x_2^2 - x_2^3 - 13)^2 + (x_1 - 14x_2 + x_2^2 + x_2^3 - 29)^2$$

$$\text{such that } x \in M = \{x \in \mathbb{R}^2 |\ g(x,y) \leq 0,\ y \in Y\},$$

where

$$g(x,y) = x_1^2 + 2x_2 y^2 + e^{x_1 + x_2} - e^y$$

and

$$Y = \{y \in \mathbb{R} \,|\, v(y) = y^2 - y \leq 0\}.$$

In the ψ_{FB} case, with the starting point $x^0 = (0.4, -1.1)$ the semismooth Newton method obtains the optimal value 97.15 with $\bar{x} = (0.719, -1.45)$ and $\bar{y} = 0$ for the optimal point. We have $||T(\bar{z})|| = 1.1072^{-7}$ after 7 iterations within 0.06 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 6 iterations with $||T(\bar{z})|| = 6.8452^{-7}$ within again 0.02 seconds of CPU time.

Problem 6.0.15 ([33])

Let us consider the following *SIP*:

$$\min \ \ f(x) = -x_1^2 - (x_2 + 5)^2 \ \ \text{such that } x \in M = \{x \in \mathbb{R}^2 |\ g(x, y) \leq 0,\ y \in Y\},$$

where

$$g(x, y) = 2yx_1 + x_2 - y^2$$

and

$$Y = \{y \in \mathbb{R} \,|\, v(y) = y^2 - 1 \leq 0\}.$$

In the ψ_{FB} case, with the starting point $x^0 = (1, 1)$ the semismooth Newton method obtains the optimal value -25 with $\bar{x} = (0, 5.1416^{-10})$ and $\bar{y} = 0$ for the optimal point. We have $||T(\bar{z})|| = 5.1416^{-10}$ after 3 iterations within 0.015 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in again 3 iterations with $||T(\bar{z})|| = 6.4701^{-8}$ within again 0.016 seconds of CPU time.

In the following examples we solve generalized semi-infinite optimization problems as well as standard semi-infinite optimization problems. The test problems in Examples 6.0.16 and 6.0.21 are taken from [103].

Example 6.0.16 ([103], [108]) *Design Centering (GSIP)*

We consider the following *GSIP* reformulation of a design centering problem:

$$\max_{x \in \mathbb{R}^n} \text{Vol}(B(x)) \quad \text{such that} \quad B(x) \subset G.$$

Let $G = \{y \in \mathbb{R}^2 \,|\, g(y) \leq 0\}$ with

$$g(y) = \begin{pmatrix} -y_1 - y_2^2 \\ y_1/4 + y_2 - 3/4 \\ -y_2 - 1 \end{pmatrix}.$$

We refer the reader to the Figure 6.1 for container G. The *GSIP* formulation of the general design centering problem is as follows:

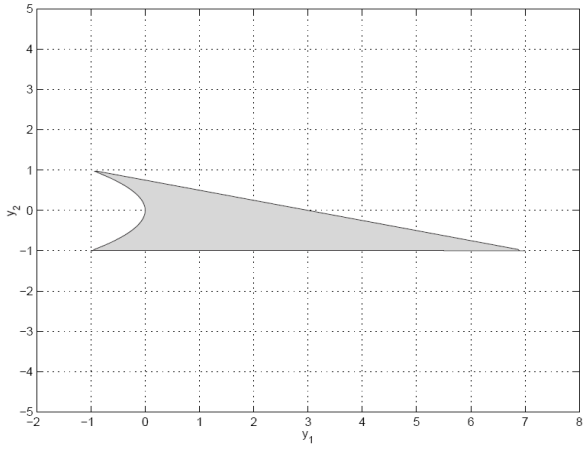

Figure 6.1: The container G, [103].

$$\max_{x\in\mathbb{R}^n} \text{Vol}(B(x)) \quad \text{such that} \quad g(y) \leq 0 \text{ for all } y \in B(x).$$

We consider the following design centering problems.

Problem 6.0.17 ([103], [108]) *The disc with maximal area with free center and radius inscribed in G (GSIP)*

The aim is to find the largest disc with free center and radius inscribed in G. We then have $n = 3$ and

$B(x) = \{y \in \mathbb{R}^2 \,|\, (y_1 - x_1)^2 + (y_2 - x_2)^2 - x_3^2 \leq 0\}$, $\text{Vol}(B(x)) = \pi x_3^2$.

In the ψ_{FB} case, with the starting point $x^0 = (0, 0, 1)$ the semismooth Newton method obtains the optimal value 1.8606 with $\bar{x} = (0.749, -0.230, 0.770)$ and $\bar{y}^1 = (-0.008, -0.091)$, $\bar{y}^2 = (0.935, 0.516)$, $\bar{y}^3 = (0.749, -1)$ for the optimal point. We have $\|T(\bar{z})\| = 7.1239^{-10}$ after 4 iterations within 0.23 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 4 iterations with $\|T(\bar{z})\| = 3.3466^{-13}$ within 0.42 seconds of CPU time. For the solution we refer the reader to the Figure 6.2.

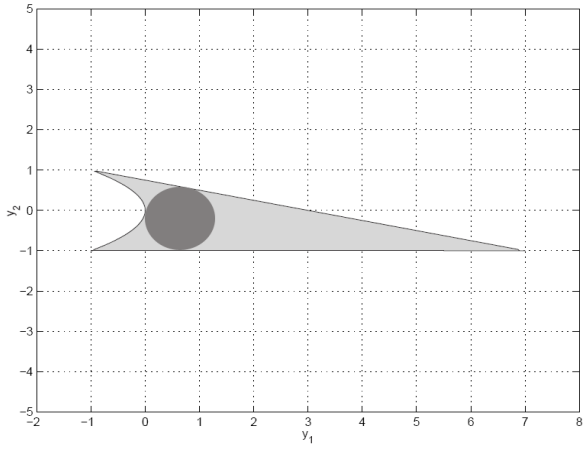

Figure 6.2: The maximal ball inscribed in the container G, [103].

Problem 6.0.18 ([103], [108]) *The largest ellipse with free center and axis lengths inscribed in G (GSIP)*

Our aim here is to find the largest ellipse with free center and axis lengths inscribed in G. We have $n = 4$ and

$$B(x) = \{y \in \mathbb{R}^2 \mid \frac{(y_1-x_1)^2}{x_3^2} + \frac{(y_2-x_2)^2}{x_4^2} - 1 \leq 0\}, \text{Vol}(B(x)) = \pi x_3 x_4.$$

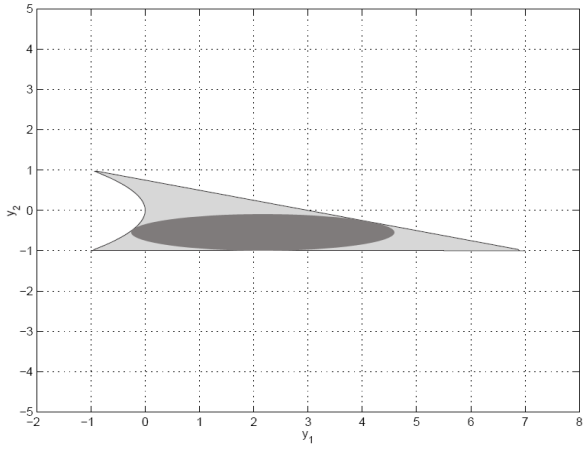

Figure 6.3: The maximal ellipse inscribed in the container G, [103].

In the ψ_{FB} case, with the starting point $x^0 = (0, 0, 1, 1)$ the semismooth Newton method obtains the optimal value 3.484 with $\bar{x} = (2.013, -0.5, 2.217, 0.5)$ and $\bar{y}^1 = (-0.167, -0.408)$, $\bar{y}^2 =$

$(3.658, -0.165)$, $\bar{y}^3 = (2.013, -1)$ for the optimal point. We have $\|T(\bar{z})\| = 3.3603^{-10}$ after 6 iterations within 0.34 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 6 iterations with $\|T(\bar{z})\| = 2.9269^{-11}$ within 0.57 seconds of CPU time. For the solution we refer the reader to the Figure 6.3.

$$\max_{x,z} z \text{ such that } z - y^T x \leq 0 \text{ for all } y \in Y, \sum_{i=1}^{K} x_i = 1, x \geq 0.$$

In fact, this is a linear semi-infinite programming problem.

Problem 6.0.19 *The largest ellipsoid with free center and axis lengths inscribed in a container (GSIP)*

We consider another example of design centering problem in $\mathbb{R}^3$. Our aim here is to find the largest ellipsoid with free center and axis lengths inscribed in container G. We let container be the simple polyhedron $G = \{y \in \mathbb{R}^3 \,|\, g(y) \leq 0\}$ with

$$g(y) = \begin{pmatrix} y_1 + y_2 + y_3 - 90 \\ 14 - y_1 \\ y_1 - 60 \\ 9 - y_2 \\ y_2 - 60 \\ -y_3 \\ y_3 - 60 \end{pmatrix}.$$

Here $n = 6$ and

$$B(x) = \{y \in \mathbb{R}^3 \,|\, \tfrac{(y_1-x_1)^2}{x_4^2} + \tfrac{(y_2-x_2)^2}{x_5^2} + \tfrac{(y_3-x_3)^2}{x_6^2} - 1 \leq 0\}, \text{Vol}(B(x)) = \tfrac{4}{3}\pi x_4 x_5 x_6.$$

The columns of the Tables 6.1 and 6.2 are labeled as follows: `initialp.center` center of ellipsoid for initial point, `ov` is the optimal value, $\|T(\bar{z})\|$ is the Euclidean norm of T at the last iteration point, `iter` is the number of iterations and `CPU time` denotes the CPU time for iterations in seconds. Both ψ_{FB} and ψ_{min} are tested. For the solution we refer the reader to the Figure 6.4. Table 6.1 contains numerical results.

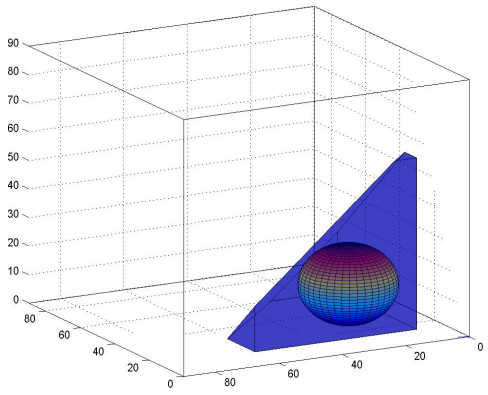

Figure 6.4: The maximal ellipsoid inscribed in the polyhedron G.

	ψ_{FB}	ψ_{min}
`initial p.center`	(20,20,20)	(20,20,20)
`ov`	11890	11890
$\|\|T(\bar{z})\|\|$	4.0134e-7	8.8400e-9
`iter`	7	7
`CPU time`	0.35	0.27

Table 6.1: The Maximal Ellipsoid in the Polyhedron.

Problem 6.0.20 *The largest simple diamond inscribed in container (GSIP)*

Our aim here is to find the largest simple diamond inscribed in container G (test problem from [128]). We let container be the simple polyhedron in Problem 6.0.19. For the simple diamond shape we refer the reader to the Figure 6.5 and for the solution we refer to the Figure 6.6. Table 6.2 contains numerical results.

	ψ_{FB}	ψ_{min}
`ov`	1.398	1.398
$\|\|T(\bar{z})\|\|$	9.8146e-7	4.2024e-9
`iter`	4	9
`CPU time`	0.52	0.98

Table 6.2: The Maximal Simple Diamond in the Polyhedron.

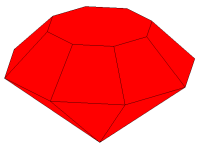

Figure 6.5: A simple diamond as a design.

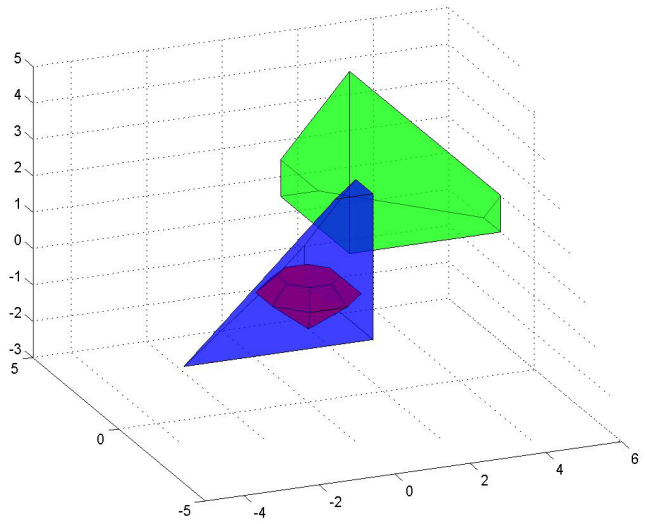

Figure 6.6: The maximal diamond inscribed in the polyhedron.

Example 6.0.21 ([103], [108]) *Robust Optimization*

As explained in Example 2.2.8 in robust optimization problems the data are uncertain and only known to belong to some uncertainty set which may be taken as infinite index set in semi-infinite programming.

Let 1 Euro be invested in a portfolio comprised of K shares. At the end of a given period the return of share i is $y_i > 0$. The goal is to determine the amount x_i to be invested in share i, $i = 1, ..., K$, so as to maximize the end-of-period portfolio value $y^T x$.

Since y is uncertain, the assumption that y varies in some non-empty compact set $Y \subset \mathbb{R}^K$ leads us to the following semi-infinite programming problem:

Problem 6.0.22 ([103], [108]) *(SIP)*

Let the uncertainty set Y be in the form:

$$Y = \left\{ y \in \mathbb{R}^K \;\middle|\; \sum_{i=1}^{K} \frac{(y_i - \bar{y_i})^2}{\sigma_i^2} \leq \theta^2 \right\},$$

where $\bar{y_i}$ is some nominal value of y_i, σ_i is scaling parameter and θ measures the risk aversion. With the particular choices from [3]

$$\bar{y_i} = 1.15 + i\frac{0.05}{K} \;\; (i = 1, 2, \ldots, K),$$
$$\sigma_i = \frac{0.05}{\theta K}\sqrt{\frac{K(K+1)i}{2}} \;\; (i = 1, 2, \ldots, K),$$
$$\theta = 1.5,$$

the optimal value is 1.15 for any K. The optimal policy in this situation is to invest equally in all shares and $x_i = 1/K$, $i = 1, ..., K$. We use the starting point $x^0 = (1, 0, \ldots, 0)^T$ in $\mathbb{R}^{K+1}$.

The columns of the Tables 6.3-6.7 are labeled as follows: `K` is the number of shares, `ov` is the optimal value, $\|T(\bar{z})\|$ is the Euclidean norm of T at the last iteration point, `CPU time` is the CPU time for iterations in seconds, `iter` is the number of iterations. Note that this optimization problem is convex so that the computed KKT point is even a global maximizer. We refer the reader to Table 6.3 and Table 6.4 for numerical results.

K	ov	$\|T(\bar{z})\|$	CPU time	iter
10	1.15	5.2928^{-8}	0.36	11
50	1.15	5.7463^{-7}	1.43	11
100	1.15	1.7676^{-10}	8.61	11
150	1.15	4.2121^{-10}	25.17	12

Table 6.3: Optimal Portfolio, Problem 6.0.22 with ψ_{FB}.

In the next problem, we will consider an example of *GSIP* in portfolio optimization.

K	ov	$\|T(\bar{z})\|$	CPU time	iter
10	1.15	3.2058^{-7}	0.54	10
50	1.15	2.5966^{-13}	1.39	11
100	1.15	7.5748^{-10}	5.58	11
150	1.15	1.0056^{-12}	15.35	12

Table 6.4: Optimal Portfolio, Problem 6.0.22 with ψ_{min}.

Problem 6.0.23 ([103], [108]) *(GSIP)*

Let the uncertainty set Y depend on x in which the risk aversion of the decision maker depends on the point x. Replacing θ by $\Theta(x)$, $Y(x)$ is given in the form [103]:

$$Y(x) = \left\{ y \in \mathbb{R}^K \;\middle|\; \sum_{i=1}^{K} \frac{(y_i - \bar{y}_i)^2}{\sigma_i^2} \leq \Theta(x)^2 \right\},$$

with

$$\Theta(x) = \theta \left(1 + \sum_{i=1}^{K} (x_i - \frac{1}{N})^2 \right).$$

In this case we have an example for *generalized* semi-infinite programming problems. We use the starting $x^0 = (1, 0, \ldots, 0)^T$ in $\mathbb{R}^{K+1}$. Since the optimization problem is not convex, we have no guarantee that the computed KKT point is a global maximizer. We refer the reader to Table 6.5 and Table 6.6 for numerical results.

K	ov	$\|T(\bar{z})\|$	CPU time	iter
10	0.7033	4.2379^{-8}	0.28	5
50	0.9638	2.3920^{-9}	0.72	7
100	1.0259	4.0606^{-7}	2.87	7
150	1.0535	8.7426^{-10}	8.76	8

Table 6.5: Optimal Portfolio, Problem 6.0.23 with ψ_{FB}.

K	ov	$\|T(\bar{z})\|$	CPU time	iter
10	0.7033	1.8731^{-11}	0.51	4
50	0.9638	9.2578^{-9}	1.43	6
100	1.0259	4.2917^{-10}	7.06	7
150	1.0535	9.9549^{-10}	34.4	11

Table 6.6: Optimal Portfolio, Problem 6.0.23 with ψ_{min}.

Example 6.0.24 *Strict complementarity violated in the upper level or in the lower level or in the both levels simultaneously.*

It can be checked that in the problems in Example 6.0.16 and 6.0.21 strict complementarity holds in the upper and lower level problems, so that we actually have a smooth system. Now, for an illustration of the case that strict complementarity is violated in the upper level, we give the following example.

Problem 6.0.25 ([108]) *Strict complementarity violated in the upper level:*

Let us consider the following *SIP* :

$$\min \;\; f(x) = (x_1 - 1)^2 + (x_2 - 1)^2 \;\; \text{such that } g(x, y) \leq 0 \;\; \forall\, y \in Y,$$

where

$$g(x, y) = (y_1 - x_1) + (y_2 - x_2)$$

and

$$Y = \{y \in \mathbb{R}^2 \,|\, v_1(y) = y_1^2 - 1 \leq 0,\; v_2(y) = y_2^2 - 1 \leq 0\}.$$

The feasible set is $M = \{x \in \mathbb{R}^2 |\; x_1 + x_2 \geq 2\}$, so that strict complementarity is violated at the solution $\bar{x} = (1, 1)$. In the ψ_{FB} case with the starting point $x^0 = (1, 2)$ the semismooth Newton method obtains the optimal value 0 with $\bar{x} = (1, 1)$ and $\bar{y} = (1, 1)$ for the optimal point. We have $\|T(\bar{z})\| = 1.6391^{-10}$ after 7 iterations within 0.14 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 6 iterations with $\|T(\bar{z})\| = 5.3765^{-10}$ within 0.28 seconds of CPU time. This shows that the method also works well for this problem where strict complementarity is violated in the upper level.

Finally, we also test the method for problems with violated strict complementarity in the lower level. At the corresponding nondifferentiability points the values of α_j^i and β_j^i are chosen according to the same rule as explained above for λ_i and θ_i . The theoretical foundation of the following examples is derived in Section 5.2.

Problem 6.0.26 ([7]) *Strict complementarity violated in the lower level*

We consider the following *SIP*:

$$\min \ \ f(x) = \tfrac{1}{3}x_1^2 + \tfrac{1}{2}x_1 + x_2^2 \ \ \text{such that } x \in M = \{x \in \mathbb{R}^2 |\ g(x,y) \leq 0,\ y \in Y\},$$

where

$$g(x,y) = (1 - x_1^2 y^2)^2 - x_1 y^2 - x_2^2 + x_2$$

and

$$Y = \{y \in \mathbb{R} \,|\, v(y) = y^2 - y \leq 0\}.$$

In the ψ_{FB} case, with the starting point $x^0 = (0,0)$ the semismooth Newton method obtains the optimal value 0.1945 with $\bar{x} = (-0.75, -0.618)$ and $\bar{y} = (1,0)$ for the optimal point. We have $||T(\bar{z})|| = 2.4356^{-10}$ after 4 iterations within 0.032 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 7 iterations with $||T(\bar{z})|| = 2.2765^{-11}$ within again 0.015 seconds of CPU time. This shows that the method also works well for this problem where strict complementarity is violated in the lower level.

Problem 6.0.27 ([110]) *Strict complementarity violated in the lower level*

Let us consider the following *SIP*: ($g(x,y) = \cos y + 2$ is taken instead of $\cos y$).

$$\min \ \ f(x) = x_1 \ \ \text{such that } x \in M = \{x \in \mathbb{R}^2 |\ g(x,y) \leq 0,\ y \in Y\},$$

where

$$g(x,y) = -x_1(\cos y + 2) - x_2 \sin y + 1$$

and

$$Y = \{y \in \mathbb{R} \,|\, v(y) = (y - \pi)(y - 3\pi/2) \leq 0\}.$$

In the ψ_{FB} case, with the starting point $x^0 = (0, 0)$ the semismooth Newton method obtains the optimal value 1 with $\bar{x} = (1, 0)$ and $\bar{y} = \pi$ for the optimal point. We have $\|T(\bar{z})\| = 2.8077^{-9}$ after 4 iterations within 0.047 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 5 iterations with $\|T(\bar{z})\| = 1.844^{-8}$ within again 0.047 seconds of CPU time.

Problem 6.0.28 ([84]) *Strict complementarity violated in the lower level:*

We consider the following *SIP* :

$$\min \ \ f(x) = (x_1 - 2)^2 + x_2^2 \ \ \text{such that } g(x, y) \leq 0 \ \ \forall\, y \in Y,$$

where

$$g(x, y) = x_1 \cos y + x_2 \sin y - 1$$

and

$$Y = \{y \in \mathbb{R} \,|\, v(y) = y^2 - \pi y \leq 0\}, \|x\|_\infty \leq 1.$$

In this problem strict complementarity is violated at the global maximizer of lower level problem $\bar{x} = (1, 0)$. In the ψ_{FB} case with the starting point $x^0 = (1, 1)$ the semismooth Newton method obtains the optimal value 1 with $\bar{x} = (1, 0)$ and $\bar{y} = 0$ for the optimal point. We have $\|T(\bar{z})\| = 4.4181 \times 10^{-8}$ after 5 iterations within 0.03 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 7 iterations with $\|T(\bar{z})\| = 3.5023 \times 10^{-7}$ within 0.02 seconds of CPU time.

Problem 6.0.29 *Strict complementarity violated in the lower level:*

Let us consider the following *GSIP*:

$$\min \ \ f(x) = x_1^2 + x_2^2 \ \ \text{such that } x \in M = \{x \in [-1, 0] \,|\, g(x, y) \leq 0,\, y \in Y(x)\},$$

where

$$g(x, y) = -(y_1 - x_1)^2 - (y_2 - x_2)^2$$

and

$$Y(x) = \{y \in \mathbb{R}^2 \,|\, v_1(x,y) = y_1 - x_1 \leq 0,\ v_2(x,y) = y_2 - x_2 \leq 0\}.$$

The unique unconstrained minimum of the objective function $\bar{x} = (0,0)$ is feasible and therefore optimal for this problem. Its active index set only contains the point $(\bar{x}, \bar{y})$ with $\bar{y} = (0,0)$, and lower level strict complementarity is violated there. In the ψ_{FB} case with the starting point $x^0 = (1,1)$ the semismooth Newton method obtains the optimal value 0 with $\bar{x} = (0,0)$ and $\bar{y} = (0,0)$ for the optimal point. We have $\|T(\bar{z})\| = 5.4022 \times 10^{-13}$ after 2 iterations within 0.03 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 4 iterations with $\|T(\bar{z})\| = 3.5403 \times 10^{-8}$ within 0.05 seconds of CPU time.

In the following example we test ψ_{FB} as the NCP function. The test problem 6.0.30 is taken from [103].

Problem 6.0.30 ([103]) *Strict complementarity violated in the lower level in Robust Optimization:*

In Example 6.0.21 let the uncertainty set Y be in the form

$$Y_\delta = \{y \in \mathbb{R}^N \,|\, \|\text{diag}(\sigma)^{-1}(y - \bar{y})\|_\delta \leq \theta\},$$

where $\bar{y_i}$ is some nominal value of y_i, σ_i is scaling parameter and θ measures the risk aversion. With the particular choice of $\delta = 10$, Y_δ becomes a non-ellipsoidal set. Moreover, we have violated strict complementarity in the lower level. We use the starting point $x^0 = (1, 1, \ldots, 1)^T$ in $\mathbb{R}^{K+1}$. We refer the reader to Table 6.7 for numerical results.

K	ov	$\|T(\bar{z})\|$	CPU time	iter
10	1.1190	1.5176×10^{-7}	1.70	28
50	1.1155	8.9342×10^{-9}	1.72	25
100	1.1151	1.3327×10^{-7}	14.24	44
150	1.1150	9.0689×10^{-7}	25.17	24

Table 6.7: Optimal Portfolio, Problem 6.0.30 with ψ_{FB}.

Problem 6.0.31 ([52], [108]) *Strict complementarity violated in the lower level:*

Another example for the case of violated strict complementarity in the lower level is given:

$$\min \ \ f(x) = x_1^2 + 3x_2^2 + x_3 \text{ such that } g(x,y) \leq 0 \ \forall\, y \in Y,$$

where

$$g(x,y) = -\frac{1}{2}(y_1 - x_1)^2 - (y_2 - x_2)^2 - x_3$$

and

$$Y = \{y \in \mathbb{R}^2 \,|\, v_1(y) = -y_1 - y_2 \leq 0,\ v_2(y) = -y_2 \leq 0,\ v_3(y) = y_1^2 + y_2^2 - 1 \leq 0\}.$$

In the ψ_{FB} case with the starting point $x^0 = (1,1,1)^T$ the semismooth Newton method obtains the optimal value 0 with $\bar{x} = (0,0,0)$ and $\bar{y} = (0,0)$ for the optimal point. We have $\|T(\bar{z})\| = 5.4371 \times 10^{-9}$ after 10 iterations within 0.16 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 16 iterations with $\|T(\bar{z})\| = 2.3642 \times 10^{-17}$ within 0.44 seconds of CPU time.

Problem 6.0.32 ([108]) *Strict complementarity violated in the upper level and in the lower level simultaneously :*

For the case of violated strict complementarity simultaneously in the upper and lower level problems, consider the following variation of Problem 6.0.31:

$$\min \ \ f(x) = x_1^2 + x_2^2 + x_3^2 \text{ such that } g(x,y) \leq 0 \ \forall\, y \in Y,$$

where

$$g(x,y) = -\frac{1}{2}(y_1 - x_1)^2 - (y_2 - x_2)^2 - x_3$$

and

$$Y = \{y \in \mathbb{R}^2 \,|\, v_1(y) = -y_1 - y_2 \leq 0,\ v_2(y) = -y_2 \leq 0,\ v_3(y) = y_1^2 + y_2^2 - 1 \leq 0\}.$$

In the ψ_{FB} case with the starting point $x^0 = (1,1,1)$ the semismooth Newton method obtains the optimal value 0 with $\bar{x} = (0,0,0)$ and $\bar{y} = (0,0)$ for the optimal point.

We have $\|T(\bar{z})\| = 2.0792 \times 10^{-7}$ after 8 iterations within 0.17 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 5 iterations with $\|T(\bar{z})\| = 0$ within 0.29 seconds of CPU time. The method works also for this problem where strict complementarity is violated simultaneously in the upper and lower level problems.

In all of the examples, if ψ_{FB} is replaced by ψ_{min}, the performance of the method does not change meaningly in both systems (4.33) for *SIP* and (4.36) for *GSIP*.

General advantages of the semismooth Newton method can be listed as follows, only a system of linear equations needs to be solved at each iteration and the convergence rate is *q-quadratic* from good initial guess if (generalized) Jacobian is nonsingular. Disadvantages of the semismooth Newton method can be listed as follows: it is not globally convergent, it requires computation of (generalized) Jacobian at each iteration and each iteration requires the solution of a system of linear equations that may be singular or ill-conditioned.

CHAPTER 7

CONCLUSION

In this thesis, we have developed and justified a numerical method, namely semismooth Newton method for solving generalized semi-infinite programming problems. The semismooth Newton method is based on the reformulated Karush-Kuhn-Tucker (KKT) conditions of generalized semi-infinite programming. The complementarity conditions in the KKT system need special attention in any numerical approach. One possibility for their treatment is a reformulation by nonlinear complementarity problem functions (NCP functions). It was suggested to use NCP functions for a nonsmooth reformulation of the KKT conditions in finitely constrained programming problems, also in standard semi-infinite programming problems [88]. But the result in [88] for standard semi-infinite programming problems was incomplete. In this thesis we corrected and completed the result of semismooth Newton method for standard semi-infinite programming problems and we also transferred the semismooth approach from standard to generalized semi-infinite programming problems.

The study in this thesis can be basically divided into two parts. The first part of our study was given in Section 5.1. Section 5.2 includes the second part of the study.

In the first part of this study, we completed the result in [88] by presenting a regularity condition which does not assume strict complementarity in the upper level problem, thus justifying the NCP function approach for semi-infinite programs. We also pointed out that global optimality of lower level problem is needed in the solution of KKT systems for semi-infinite programs. The first part of this study was published in our recent paper [108] and shows that the semismooth Newton method for semi-infinite programming can actually handle nonsmoothness, since there the result from [88] was extended to the case of violated strict complementarity in the upper level problem. Moreover, we transferred the semismooth approach

from standard semi-infinite programming problems to generalized semi-infinite programming problems.

The second part of the study completes the first part by considering the case of strict complementarity violation in the lower level. This result was published in our second paper [115]. There, we gave an appropriate new regularity condition for the convergence of a semismooth Newton method, thus justifying the NCP approach for semi-infinite programs in the absence of strict complementarity in both upper level and lower level problems. In the present case, the convergence analysis was essentially more complicated due to the lack of differentiability of the auxiliary functions of the so-called reduced problem.

Our main result in the second part was given in the Section 5.2.1, a sufficient condition for quadratic convergence of the semismooth Newton method for generalized semi-infinite programming problems where strict complementarity neither has to be assumed in the upper nor in the lower level problem.

As a future study, another globalization method (ex: trust-region interior method) can be used and the problems faced while applying the method to gemstone cutting problem (ill-conditioned of Jacobian) can be fixed (use of some scaling matrix/preconditioner) and the use of the method in genetic networks under uncertainty [121, 122, 123, 124, 125] can be considered.

REFERENCES

[1] Abadie, J. *Generalized reduced gradient and global Newton methods*, R. Conti, E. D. Giorgi and F. Giannessi, eds., *Optimization and Related Fields*, Springer-Verlag, Berlin, 1-20, (1986).

[2] Bazaraa, M. S., Sherali, H. D. and Shetty, C. M. *Nonlinear Programming: Theory and Algorithms*, Wiley, (1993).

[3] Ben-Tal, A. and Nemirovski, A. *Robust solutions of uncertain linear programs*, Operations Research Letters, **25**, 1-13, (1999).

[4] Bhattacharjee, B., Lemonidis, P., Green, W. H. and Barton, P. I. *Global solution of semi-infinite programs*, *Mathematical Programming*, **103**, 283-307, (2005).

[5] Bonnans, J. F. and Shapiro, A. *Perturbation Analysis of Optimization Problems*, Springer, New York, (2000).

[6] Clarke, F. H. *Optimization and Nonsmooth Analysis*, Wiley, New York, (1983).

[7] Coope, I.D. and Watson, G.A. *A projected Lagrangian algorithm for semi-infinite programming*, Math. Programming, **32**, 337-356, (1985).

[8] Dambrine, M. and Pierre, M. *About stability of equilibrium shapes*, Mathem. Modelling and Num. Analysis, **34(4)**, 811-834, (2000).

[9] Deuflhard, P. *Newton Methods for Nonlinear Problems Affine Invariance and Adaptive Algorithms*, Springer, (2004).

[10] Dong-Hui, L., Yamashita, N. and Fukushima, M. *Nonsmooth equation based BFGS method for solving KKT Systems in mathematical programming*, J. of Opt. Theory and Appl. 1, **109**, 123-167, (2001).

[11] Dong-Hui, L., Qi, L., Tam, J. and Wu, S.-Y. *A smoothing Newton method for semi-infinite programming*, J. of Global Optim., **30**, 169-194, (2004).

[12] Facchinei, F., Fischer, A. and Kanzow, C. *Regularity properties of a semismooth reformulation of variational inequalities*, SIAM Journal on Optimization, **8**, 850-869, (1998).

[13] Faigle, U., Kern, W. and Still, G. *Algorithmic Principles of Mathematical Programming*, Kluwer, Dordrecht, (2002).

[14] Fischer, A. *A special Newton type optimization method*, Optimization, **24**, 269-284, (1992).

[15] Fletcher, R. *Practical Methods of Optimization*, John Wiley and Sons, England, (1987).

[16] Floudas, C. A. and Stein, O. *The adaptive convexification algorithm: a feasible point method for semi-infinite programming*, SIAM J. Optim., **18(4)**, 1187-1208, (2007).

[17] Geletu, A. and Hoffmann, A. *A conceptual method for solving generalized semi-infinite programming problems via global optimization by exact discontinuous penalization*, European Journal of Operations Research, **157**, 3-15, (2004).

[18] Goberna, M. A. and Lopez, M. A. *Linear Semi-infinite Optimization*, Wiley, Chichester, (1998).

[19] Graettinger, T.J. and Krogh, B.H. *The acceleration radius: a global performance measure for robotic manipulators*, IEEE Journal of Robotics and Automation, **4**, 60-69, (1988).

[20] Guerra Vazquez, F., Rückmann, J.-J., Stein, O. and Still, G. *Generalized semi-infinite programming: a tutorial*, Journal of Computational and Applied Mathematics, **217**, 394-419, (2008).

[21] Gugat, M. *Semi-infinite terminal problems: a Newton-type method*, Optimization, **44**, 25-48, (1998).

[22] Gustafson, S.A. and Kortanek, K.O. *Numerical treatment of a class of semi-infinite programming problems*, Nav. Res. Log. Quart., **20**, 477-504, (1973).

[23] Günzel, H., Jongen, H. Th. and Stein, O. *Generalized semi-infinite programming: on generic local minimizers*, Journal of Global Optimization, **42**, 413-421, (2008).

[24] Haaren-Retagne, E. *A semi-infinite programming algorithm for robot trajectory planning*, Ph.D. thesis, University Trier, (1992).

[25] Han, S.P., Pang, J. and Rangaraj, N. *Globally convergent Newton methods for nonsmooth equations*, Math. of Op.Res., **3(17)**, 586-607, (1992).

[26] Hettich, R. *A Newton method for nonlinear Chebyshev approximation, in: Approximation Theory*, Lect. notes in Math. 556, Scaback, R., Scherer, K., eds., Springer, Berlin-Heidelberg-New York, 222-236, (1976).

[27] Hettich, R. and Jongen, H. Th. *On first and second order conditions for local optima for optimization problems in finite dimensions*, Operations Research Verfahren XXIII, Tag. Oberwolfach, (1976).

[28] Hettich, R. and Jongen H. Th. *Semi-infinite programming: conditions of optimality and applications*, In J. Stoer, editor, Optimization Techniques 2, Lecture notes in Control and Information Sci. Springer, Berlin, Heidelberg, New York, (1978).

[29] Hettich, R. *A comparison of some numerical methods for semi-infinite programming*, In Balakrishnan A.v., Thoma M., eds., Lecture notes in control and information science, **15**, 113-125, (1978).

[30] Hettich, R. and van Honstede, W. *On quadratically convergent methods for semi-infinite programming*, Semi-infinite Programming, **15**, 97-111, (1979).

[31] Hettich, R. *A Review of numerical methods for semi-infinite optimization*, Semi-infinite programming and applications, **215**, 158-178, (1981).

[32] Hettich, R. *An implementation of a discretization method for semi-infinite programming*, Mathematical Programming, **34(3)**, 354-361, (1986).

[33] Hettich, R. and Kortanek, K. *Semi-infinite programming: theory, methods and applications*, SIAM Review, **35**, 380-429, (1993).

[34] Hettich, R. and Still, G. *Second Order Optimality Conditions for Generalized Semi-infinite Programming Problems*, Optimization, **34**, 195-211, (1995).

[35] Hettich, R. and Still, G. *Semi-infinite programming models in robotics*, in: J. Guddat, H.Th. Jongen, B.Kummer, F.Nozicka (eds): Parametric Optimization and RelatedTopics II, Akademie Verlag, Berlin, 112-118, (1991).

[36] Hettich, R. and Zencke, P. *Numerische Methoden der Approximation und semi-infiniten Optimierung*, Tuebner, Stuttgart, (1982).

[37] Horst, R. and Tuy, H. *Global Optimization*, Springer, Berlin, (1990).

[38] Vaz, Ismael F. A., Fernandes, Edite M. G. P. and Paula S. F. Gomes, M. *Robot trajectory planning with semi-infinite programming*, European Journal of Operational Research, **153**, 607-617, (2004).

[39] Vaz, Ismael F. A., Ferreira and Eugénio C. *Air pollution control with semi-infinite programming*, Applied Mathematical Modelling, **33**, 1957-1969, (2009).

[40] Jiang, H. and Qi, L. *A new nonsmooth equations approach to nonlinear complementarity problems*, SIAM J. Cont. Optim., **35**, 178-193, (1997).

[41] Jittorntrum, K. *Solution point differentiability without strict complementarity in nonlinear programming*, Math.Programming, **21**, 127-138, (1984).

[42] John, F. *Extremum problems with inequalities as subsidiary conditions*, in: Studies and Essays, R. Courant Anniversary Volume, Interscience, New York, 187-204, (1948).

[43] Jongen, H. Th., Jonker, P. and Twilt, F. *Critical sets in parametric optimization*, Mathematical Programming, **34**, 333-353, (1986).

[44] Jongen, H.Th., Jonker, P. and Twilt, F. *Nonlinear Optimization in Finite Dimensions*, Kluwer, Netherlands, (2000).

[45] Jongen, H. Th., Meer, K. and Triesch, E. *Optimization Theory*, Kluwer Academic Publishers, (2004).

[46] Jongen, H. Th., Möbert, T., Rückmann, J.-J. and Tammer, T. *On inertia and Schur complement in optimization*, Lin. Alg. Appl., **95**, 97-109, (1987).

[47] Jongen, H.Th., Rückmann, J.-J. and Stein, O. *Generalized semi-infinite optimization: a first order optimality condition and examples*, Mathematical Programming, **83**, 145-158, (1998).

[48] Jongen, H.Th., Rückmann, J.-J. and Weber, G.-W. *One-parametric semi-infinite optimization: On the stability of the feasible set*, SIAM Journal of Optimization, **4**, 637-648, (1994) .

[49] Jongen, H.Th., Twilt, F. and Weber, G.-W. *Semi-infinite optimization: structure and stability of the feasible set*, J. Optim. Theory Applic., **72**, 529-552, (1992).

[50] Jongen, H.Th. and Weber, G.-W. *Nonlinear optimization: characterization of structural stability*, J.Global Optim., **1**, 47-64, (1991).

[51] Jongen, H.Th. and Weber, G.-W. *On parametric nonlinear programming*, Annals of Op.Res., **27**, 253-284, (1990).

[52] Jongen, H. Th., Wetterling, W. and Zwier, G. *On sufficient conditions for local optimality in semi-infinite programming*, Optimization, **18**, 165-178, (1987).

[53] Jongen, H.Th. and Zwier, G. *On the local structure of the feasible set in semi-infinite optimization*, Optimization, **72**, 185-202, (1984).

[54] Kaiser, C. and Krabs, W. *Ein Problem der semi-infiniten optimierung im maschinenbau und seine verallgemeinerung*, Working paper, Darmstadt University of Technology, Department of Mathematics, (1986).

[55] Kaplan, A. and Tichatschke, R. *On a class of terminal variational problems*, in: J. Guddat, H.Th. Jongen, F. Nozicka, G. Still, F. Twilt (eds): Parametric Optimization and Related Topics IV, Peter Lang, Frankfurt a.M., 185-199, (1997).

[56] Kelley, C.T. *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, (1995).

[57] Kelley, C.T. *Solving Nonlinear Equations with Newton's Method*, SIAM, Philedelphia, (2003).

[58] Krabs, W. *On time-minimal heating or cooling of a ball*, in: International Series of Numerical Mathematics, **81**, Birkhauser, Basel, 121-131, (1987).

[59] Klatte, D. *Stability of stationary solutions in semi-infinite optimization via the reduction approach*, in: W. Oettli, D. Pallaschke (eds): Advances in Optimization, Lecture Notes in Economics and Mathematical Systems, **382**, Springer, Berlin, 155-170, (1992).

[60] Klatte, D. and Kummer, B. *Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications*, Springer Verlag, New York, (2002).

[61] Kocvara, M., Outrata, J. and Zowe, J. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory,Applications and Numerical Results*, Kluwer, Dordrecht, (1998).

[62] Kojima, M. *Strongly stable stationary solutions in nonlinear programming*, S.M. Robinson, ed., Analysis and Computation of Fixed Points, Academic Press, New York, 93-138, (1980).

[63] Kress, R. *Numerical Analysis*, Springer Verlag, Newyork, (1998).

[64] Kummer, B. *Generalized Newton and NCP-methods: convergence,regularity,actions*, Control and Optimization, **20**, 209-244, (2000).

[65] Kummer, B. *Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis*, Advances in Optimization, 171-194, (1992).

[66] Kummer, B. *Newton's Method for Nondifferentiable Functions*, Advances in Math.Optim., 114-125, (1998).

[67] Lemonidis, P. *Global Optimization Algorithms for Semi-infinite and Generalized Semi-infinite Programs*, PhD Thesis, Massachusetts Institute of Technology, (2008).

[68] Levitin, E. and Tichatschke, R. *A branch-and-bound approach for solving a class of generalized semi-infinite programming problems*, Journal of Global Optimization, **13**, 299-315, 1998.

[69] Levitin, E. *Reduction of generalized semi-infinite programming problems to semi-infinite or piece-wise smooth programming problems*, Preprint No. 8-2001, University of Trier, (2001).

[70] Mangasarian, O.L. *Unconstrained Lagrangians in nonlinear programming*, SIAM J. Control, **13**, 772-791, (1975).

[71] Mangasarian, O. L. *Equivalence of the complementarity problem to a system of nonlinear equations*, SIAM J. Applied Mathematics, **31**, 89-92, (1976).

[72] Mifflin, R. *Semismooth and semiconvex functions in constrained optimization*, SIAM J.Control Optim., **15**, 959-972, (1997).

[73] Mifflin, R. *An Algorithm for constrained optimization with semismooth functions*, Math.of Op.Res., 191-207, (1977).

[74] Nash, S.G. and Sofer, A. *Linear and Nonlinear Programming*, McGraw-Hill, Newyork, (1998).

[75] Nguyen, V.H. and Strodiot, J.J. *Computing a global optimal solution to a design centering problem*, Mathematical Programming, **53**, 111-123, (1992).

[76] Ni, Q., Ling, L., Qi, L. and Teo, K.L. *A truncated projected Newton-type algorithm for large scale semi-infinite programming*, SIAM Journal on Optimization, **16**, 1137-1154 (2006).

[77] Nocedal, J. and Wright, S.J. *Numerical Optimization*, Springer, (1999).

[78] Ouellette, D.V. *Schur complements and statistics*, Lin.Alg.Appl., **36**, 187-295, (1981).

[79] Özöğür-Akyüz, S. *A Mathematical contribution of statistical learning and continuous optimization using infinite and semi-infinite programming, to computational statistics*, PhD Thesis, Institute of Applied Mathematics, METU, (2008).

[80] Polak, E. *Semi-infinite optimization in engineering design*, Beckmann M., Krelle W., eds., Lecture notes in economics and mathematical systems, **215**, 237-248, (1981).

[81] Polak, E. *An implementable algorithm for the optimal design centering, tolerancing and tuning problem*, Journal of Optimization Theory and Applications, **37**, 45-67, (1982).

[82] Polak, E. *On the mathematical foundation of nondifferentiable optimization in engineering design*, SIAM Rev., **29(1)**, 21-89, (1997).

[83] Ponomarenko, A. *Lösungsmethoden für Variationsungleichungen*, Doctoral Thesis, Humboldt-Universität zu Berlin, (2003).

[84] Price, C.J. and Coope, I.D. *Numerical experiments in semi-infinite programming*, Computational Optimization and Applications, **6**, 169-189, (1996).

[85] Qi, L. and Jiang, H. *Semismooth KKT equations and convergence analysis of Newton and Quasi-Newton methods for solving these equations*, Math. of Op.Research, **22**, 301-325, (1997).

[86] Qi, L. and Martinez, J.M. *Inexact Newton methods for solving nonsmooth equations*, J.Comput. Appl. Math., **60**, 127-145, (1995).

[87] Qi, L. and Sun, J. *A Nonsmooth version of Newton's method*, Math. Programming, **58**, 353-367, (1993).

[88] Qi, L., Wu, S.-Y. and Zhou, G. *Semismooth Newton methods for solving semi-infinite problems*, Journal of Global Optimization, **27**, 215-232, (2003).

[89] Qi, L. *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., **18**, 227-244, (1993).

[90] Reemtsen, R. *Discretization methods for the solution of semi-infinite programming problems*, Journal of Optimization Theory and Applications, **71(1)**, 85-103, (1991).

[91] Reemtsen, R. and Görner, S. *Numerical methods for semi-infinite programming: a survey*, in: R. Reemtsen, J.-J. Rückmann (eds.): Semi-Infinite Programming, Kluwer, Boston, 195-275, (1998).

[92] Reemtsen, R. and Rückmann, J.-J. *Semi-Infinite Programming*, Springer, (1998).

[93] Robinson, S. M. *Generalized Equations and their solutions, Part II, Applications to nonlinear programming*, Math. Programming Stud., **19**, 200-221, (1982).

[94] Robinson, S.M. *Newton's Method for a class of nonsmooth functions*, Industrial Engineering Working Paper, University of Wisconsin, (1998).

[95] Rockafellar, R. T. *Convex Analysis*, Princeton University Press, New Jersey, (1970).

[96] Rückmann, J.-J. *On existence and uniqueness of stationary points in semi-infinite optimization*, Math. Programming, **86**, 387-415, (1999).

[97] Rückmann, J.-J. and Shapiro A. *On first order optimality conditions in generalized semi-infinite programming*, J. optim. theory Appl., **101(3)**, 677-691, (1999).

[98] Rückmann, J.-J. and Shapiro A. *Second order optimality conditions in generalized semi-infinite programming*, Set-Valued Analysis, **9**, 169-186, (2001).

[99] Rückmann, J.-J. and Stein, O. *On linear and linearized generalized semi-infinite optimization problems*, Annals of Op. Res., **11**, 191-208, (2001).

[100] Scholtes, S. *Introduction to piecewise differentiable equations*, Habilitationsschrift, Institut für Statistik und Mathematische Wirtschaftstheorie, Universitat Karlsruhe, Karlsruhe, Germany, (1994).

[101] Sherali, H.D. *Disjunctive programming*, in: Encyclopedia of Optimization, Kluwer, Dordrecht, 466-470, (2000).

[102] Stein, O. *On Parametric Semi-infinite Optimization*, PhD Thesis, University of Trier, Department of Mathematics, (1997).

[103] Stein, O. *Bi-Level Strategies in Semi-infinite Programming*, Kluwer, (2003).

[104] Stein, O. *A semi-infinite approach to design centering*, Optimization with Multivalued Mappings, Springer, 209-228, (2006).

[105] Stein, O. and Still, G. *On generalized semi-infinite optimization and bilevel optimization*, European Journal of Operational Research, **142**, 444-462, (2002).

[106] Stein, O. and Still, G. *Solving semi-infinite optimization problems with interior point techniques*, SIAM J.Control Optim., **42(3)**, 769-788, (2003).

[107] Stein, O. and Still, G. *On optimality conditions for generalized semi-infinite programming problems*, J.of Optim. Theory and Appl., **104(2)**, 443-458, (2000).

[108] Stein, O. and Tezel, A. *The Semismooth Approach for semi-infinite programming under the Reduction Ansatz*, Journal of Global Optimization, **41**, 245-266, (2008).

[109] Still, G. *Discretization in semi-infinite programming: The rate of convergence*, Math. Program., **91**, 53-69, (2001).

[110] Still, G. *Semi-infinite programming: An introduction, preliminary version*, Technical report, University of Twente Department of Applied Mathematics, Enschede, The Netherlands, (2004).

[111] Still, G. *Generalized semi-infinite programming: theory and methods*, European J. of Op. Res., **119**, 301-313, (1999).

[112] Still, G. *Generalized semi-infinite programming: numerical aspects*, Optimization, **49**, 223-242, (2001).

[113] Sun, D. and Han, J. *Newton and Quasi-Newton methods for a class of nonsmooth equations and related topics*, SIAM J. Optim., **7**, 463-480, (1997).

[114] Sun, D. and Qi, L. *On NCP-Functions*, Comp. Optim. and Appl., **13**, 201-220, (1999).

[115] Tezel, A. and Stein O. *The semismooth approach for semi-infinite programming without strict complementarity*, SIAM J. Optim., **20(2)**, 1052-1072, (2009).

[116] Ulbrich, M. *Nonsmooth Newton-like Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, Habilitation Thesis, Technical University of Munich, (2002).

[117] Watson, G.A. *Lagrangian Methods for semi-infinite programming Problems, Infinite programming*, Proc. Int. Symp., Cambridge UK, 90-107, (1984).

[118] Weber, G.-W. *Generalized Semi-infinite Optimization and Related Topics*, Heldermann Verlag, (2003).

[119] Weber, G.-W. *Generalized semi-infinite optimization: on some foundations*, J. of Comp. Tech., **4**, 41-61, (1999).

[120] Weber, G.-W. and Karasözen, B. Numerical Optimization lecture notes, IAM, 2003.

[121] Weber, G.-W. and Tezel, A. *New views: Generalized semi-infinite optimization of genetic networks*, TOP, the Operational Research journal of SEIO (Spanish Statistics and Operations Research Society), **14(1)**, 48-55, June 2006.

[122] Weber, G.-W. and Tezel, A. *On generalized semi-infinite optimization of genetic networks*, TOP, the Operational Research journal of SEIO (Spanish Statistics and Operations Research Society), **15(1)**, 65-77, (2007).

[123] Weber, G.-W., Tezel, A., Taylan, P., Söyler, A. and Çetin, M. *Mathematical contributions to dynamics and optimization of gene-environment Networks*, Special Issue: In Celebration of Prof. Dr. Dr. Hubertus Th. Jongen's 60th Birthday, D. Pallaschke, O. Stein, guest eds., of Optimization, **57(2)**, 353-377, (April 2008).

[124] Weber, G.-W., Uğur, Ö., Taylan, P. and Tezel, A. *On optimization, dynamics and uncertainty: a tutorial for gene-environment Networks*, Special issue Networks in Computational Biology of Discrete Applied Mathematics, **157(10)**, 2494-2513, (May 2009).

[125] Weber, G.-W., Kropat, E., Tezel, A. and Belen, S. *Optimization applied on regulatory and eco-finance networks - survey and new developments*, to appear in the special issue of Pacific Journal of Optimization (PJO), **6(3)**, in honour of Professor Rubinov, guest eds.: M. Fukushima et al., (September 2010).

[126] Wetterling, W.W.E. *Definitheitsbedingungen für relative extrema bei optimierungsund approximationsaufgaben*, Numer. Math., **12**, 122-136, (1970).

[127] Wierzbicki, A.P. *Note on the equivalence of Kuhn-Tucker complementarity conditions to an equation*, J. Optim. Theory Appl., **37**, 401-405, (1982).

[128] Winterfeld, A. *Large-scale Semi-infinite Optimization Applied to Industrial Gemstone Cutting*, PhD Thesis, University of Kaiserslautern, (2007).

[129] Winterfeld, A. *Application of general semi-infinite programming to lapidary cutting problems*, European Journal of Operational Research, **191(3)**, 838-854, (2008).

[130] Yamashita, N. *Properties of restricted NCP functions for nonlinear complementarity problems*, J.of Optim.Theory and Appl., **98(3)**, 701-717, (1998).

APPENDIX A

NCP FUNCTIONS

Appendix A is based on our recent paper [108].

Both functions $\psi_{FB}(a,b) = \sqrt{a^2+b^2} - a - b$ and $\psi_{min}(a,b) = -\min\{a,b\}$ are convex on $\mathbb{R}^2$, and are differentiable in $\mathbb{R}^2$ except for the origin and the diagonal $\{(a,b) \in \mathbb{R}^2 | \ a = b\}$, respectively. For $(a,b) \neq 0$ we have

$$D\psi_{FB}(a,b) = \left(\frac{a}{\sqrt{a^2+b^2}} - 1, \ \frac{b}{\sqrt{a^2+b^2}} - 1 \right) \tag{A.1}$$

and for $a \neq b$

$$D\psi_{min}(a,b) = \begin{cases} (-1,0), & a < b \\ (0,-1), & a > b \end{cases} . \tag{A.2}$$

In their (identical) zero set, the only point of nondifferentiability for either of the two functions is the origin, and their gradients for $(a,b) \neq 0$ coincide:

$$D\psi_{FB}(a,b) = D\psi_{min}(a,b) = \begin{cases} (-1,0), & a = 0 \\ (0,-1), & b = 0 \end{cases} . \tag{A.3}$$

At the origin we calculate subdifferentials of the convex functions ψ_{FB} and ψ_{min} as follows. We denote the usual *directional derivative* of ψ in the direction d at $\bar{x}$ by $\psi'(\bar{x};d)$ and the *generalized* directional derivative of ψ (in the sense of Clarke) in the direction d at $\bar{x}$ by $\psi^0(\bar{x};d)$. For completeness, we give the proofs of the following well-known results.

Lemma A.0.33 ([108]) *The following assertions hold:*

(i) $\psi'_{FB}(0;d) = \psi^0_{FB}(0;d) = \psi_{FB}(d)$ *for any* $d \in \mathbb{R}^2$,

(ii) $\psi'_{min}(0;d) = \psi^0_{min}(0;d) = \psi_{min}(d)$ *for any* $d \in \mathbb{R}^2$,

(iii) $\partial\psi_{FB}(0) = \{s \in \mathbb{R}^2 \,|\, (s_1+1)^2 + (s_2+1)^2 \leq 1\}$,

(iv) $\partial\psi_{min}(0) = \text{conv}\{(-1,0)^T, (0,-1)^T\}$.

Proof. Parts (i) and (ii) can easily be checked by calculation. In part (iii), by definition of the Clarke subdifferential for real-valued functions, we have

$$\begin{aligned}
\partial\psi_{FB}(0) &= \{s \in \mathbb{R}^2 |\, \psi^0(0;d) \geq d^T s \text{ for all } d \in \mathbb{R}^2\} \\
&= \{\tilde{s} \in \mathbb{R}^2 |\, d^T\tilde{s} \leq \|d\|_2, \text{ for all } d \in \mathbb{R}^2\} - (1,1)^T \\
&= \{\tilde{s} \in \mathbb{R}^2 |\, \max_{d \in \partial B(0,1)} d^T\tilde{s} \leq 1\} - (1,1)^T
\end{aligned}$$

where $\tilde{s} = s + (1,1)^T$ and $\partial B(0,1)$ denotes the boundary of the unit ball. The Cauchy-Schwarz inequality $d^T\tilde{s} \leq \|d\|_2\|\tilde{s}\|_2$ implies that $\max_{d\in\partial B(0,1)} d^T\tilde{s} \leq \|\tilde{s}\|_2$ for all $\tilde{s} \in \mathbb{R}^2$. On the other hand, with $\bar{d} = \tilde{s}/\|\tilde{s}\|_2 \in \partial B(0,1)$, we find $\max_{d\in\partial B(0,1)} d^T\tilde{s} \geq \bar{d}^T\tilde{s} = \|\tilde{s}\|_2$ for all $\tilde{s} \in \mathbb{R}^2 \setminus \{0\}$, the case $\tilde{s} = 0$ being trivial. This shows

$$\partial\psi_{FB}(0) = \{\tilde{s} \in \mathbb{R}^2 |\, \|\tilde{s}\|_2 \leq 1\} - (1,1)^T$$

and completes the proof of part (iii).

To see part (iv), note that by definition of the Clarke subdifferential for vector-valued functions we have

$$\partial\psi_{min}(0) \;=\; \text{conv}\{\lim_{x\to 0} \nabla\psi_{min}(x) \,|\, x \in D_\psi\} \;=\; \text{conv}\{(-1,0)^T, (0,-1)^T\}$$

where D_ψ is set of differentiability points of ψ_{min}. •

APPENDIX B

BLOCK MATRICES

Appendix B is based on our recent papers [108, 115].

Definition B.0.34 ([78]) *Consider the quadratic block matrix*

$$A = \begin{pmatrix} E & F \\ G & H \end{pmatrix}$$

where H is quadratic and nonsingular. Then the matrix

$$S = A/H = E - FH^{-1}G$$

is called the Schur complement *of H in A.*

Lemma B.0.35 ([78]) *Let a block matrix A be given as in Definition B.0.34 and let H be nonsingular. Then*

$$det(A) = det(H) \cdot det(A/H).$$

In particular, A is nonsingular if and only if A/H is nonsingular.

Lemma B.0.36 ([78]) *Let a block matrix A be given as in Definition B.0.34 with E nonsingular, and let A be symmetric. Then,*

$$\mathrm{In}(A) = \mathrm{In}(E) + \mathrm{In}(A/E),$$

where In(A) denotes the inertia-triple *of A, that is, the number of negative, positive and vanishing eigenvalues, respectively.*

Lemma B.0.37 ([46]) *For $A \in S^N$ and $B \in \mathbb{R}_K^{N\times M}$ we have*

$$In\begin{pmatrix} A & B \\ B^T & 0_{M\times M} \end{pmatrix} = In(A|_{Ker(B^T)}) + (K, K, M-K). \tag{B.1}$$

Here, S^N denotes the set of symmetric matrices in $\mathbb{R}^{N\times N}$ and $\mathbb{R}_K^{N\times M}$ denotes the set of matrices in $\mathbb{R}^{N\times M}$ with rank K. The null space of a matrix A is denoted by $\mathrm{Ker}(A)$.

By Lemma B.0.37, we have that $A|_{\mathrm{Ker}(B^T)}$ is nonsingular and the columns of B are linearly independent if and only if the matrix

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix}$$

is nonsingular.

Lemma B.0.38 ([78]) *Let a block matrix A be given as in Definition B.0.34 with E nonsingular. Let $S = A/E := H - GE^{-1}F$ be the Schur complement of E in A. Then,*

$$A^{-1} = \begin{pmatrix} E^{-1} + E^{-1}FS^{-1}GE^{-1} & -E^{-1}FS^{-1} \\ -S^{-1}GE^{-1} & S^{-1} \end{pmatrix}.$$

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Tezel Özturan, Aysun

Date and Place of Birth: 22 August 1978, Ordu

email: atezel@metu.edu.tr

EDUCATION

M.Sc. METU, Mathematics, January 2005.

B.S. METU, Mathematics Education, June 2001 (CPGA: 3.79/4.00).

B.S. METU, Mathematics, June 2001 (CPGA: 3.69/4.00).

High School, Amasya Anadolu Öğretmen Lisesi, June 1996.

WORK EXPERIENCE

2002-Present, Research Assistant, Department of Mathematics, METU.

RESEARCH VISITS

1. 7 November 2006-28 October 2007, University of Karlsruhe, Karlsruhe, Germany, supervised by Prof. Dr. Oliver Stein.
2. 19 July 2006-19 August 2006, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, USA, supervised by Prof. Dr. Lorenz T. Biegler.
3. 15 January 2006-15 February 2006, RWTH Aachen University, Aachen, Germany, supervised by Prof. Dr. Oliver Stein.
4. 15 August 2004-15 September 2004, RWTH Aachen University, Aachen, Germany supervised by Prof. Dr. Oliver Stein and University of Twente, Enschede, The Netherlands, supervised by Prof. Dr. Georg Still.

FOREIGN LANGUAGES

English (High level), German (Basic level)

PUBLICATIONS

1. **Tezel, A.** and Stein, O. *The semismooth approach for semi-infinite programming without strict complementarity*, SIAM Journal on Optimization, **20(2)**, 1052-1072, (2009).
2. Weber, G.-W., Uğur, Ö., Taylan, P. and **Tezel, A.** *On optimization, dynamics and uncertainty: a tutorial for gene-environment Networks*, Special issue Networks in Computational Biology of Discrete Applied Mathematics, **157(10)**, 2494-2513, (May 2009).
3. Stein, O. and **Tezel, A.** *The semismooth approach for semi-infinite programming under the Reduction Ansatz*, Journal of Global Optimization, **41(2)**, 245-266, 2008.
4. Weber, G.-W., **Tezel, A.**, Taylan, P., Söyler, A. and Çetin, M. *Mathematical contributions to dynamics and optimization of gene-environment Networks*, Special Issue: In Celebration of Prof. Dr. Dr. Hubertus Th. Jongen's 60th Birthday, D. Pallaschke, O. Stein, guest eds., of Optimization, **57(2)**, 353-377, (April 2008).
5. Weber, G.-W. and **Tezel, A.** *On generalized semi-infinite optimization of genetic networks*, TOP, the Operational Research journal of SEIO (Spanish Statistics and Operations Research Society), **15(1)**, 65-77, (2007).
6. Weber, G.-W. and **Tezel, A.** *New views: Generalized semi-infinite optimization of genetic Networks*, TOP, the Operational Research journal of SEIO (Spanish Statistics and Operations Research Society), **14(1)**, 48-55, (June 2006).
7. Weber, G.-W., Kropat, E., **Tezel, A.** and Belen, S. *Optimization applied on regulatory and eco-finance networks - survey and new developments*, to appear in the special issue of Pacific Journal of Optimization (PJO), **6(3)**, in honour of Professor Rubinov, guest eds.: M. Fukushima et al., (September 2010).

PRESENTATIONS IN INTERNATIONAL SCIENTIFIC MEETINGS

1. *The Semismooth Approach for Generalized Semi-infinite Programming*, EURO XXII, 22nd European Conference on Operational Research, Prague, Czech Republic, July 2007.
2. *Semismooth Newton Method for Semi-infinite Programming Problems*, EURO XXI, 21st European Conference on Operational Research, Reykjavik, Iceland, July 2006.
3. *On Semi-Infinite Optimization of Anticipatory Systems and their Modern Applications*, 8th SIAM Conference in Optimization, Stockholm, Sweden, May 2005.

PARTICIPATION IN INTERNATIONAL SCIENTIFIC MEETINGS

1. 6th International Congress on Industrial and Applied Mathematics, Zurich, Switzerland, ICIAM 07, 16-20 July 2007.
2. 22nd European Conference on Operational Research, Prague, Czech Republic, July 8-11, 2007.
3. 21st European Conference on Operational Research, Reykjavik, Iceland, July 2-5, 2006.
4. Workshop on Advances in Continuous Optimization, Reykjavik, Iceland, June 30-July 1, 2006.
5. 8th SIAM Conference in Optimization, Stockholm, Sweden, 15-19 May, 2005.
6. EURO Summer Institute, Optimization and Data Mining, Ankara Turkey, July 9-25, 2004.
7. Workshop on Advances in Continuous Optimization, Istanbul, Turkey, July 4-5, 2003.

ORGANIZATION OF SCIENTIFIC EVENTS

1. Member of organizing commitee (stream organizer) of *EUROPT-OMS* 2007, Prague, July 2007.
2. Member of organizing commitee of *Workshop on Advances in Continuous Optimization*, Reykjavik, Iceland, June 30-July 1, 2006.