Primal and Dual Gap Functions for Generalized Nash Equilibrium Problems and Quasi-Variational Inequalities

Dissertation zur Erlangung
des naturwissenschaftlichen Doktorgrades
der Julius-Maximilians-Universität Würzburg

vorgelegt von

NADJA HARMS

aus
Prostornoje

Würzburg, 2014

Eingereicht am: 16. Juli 2014
1. Gutachter: Prof. Dr. Christian Kanzow, Universität Würzburg
2. Gutachter: Prof. Dr. Andreas Fischer, Technische Universität Dresden

Acknowledgments

The last three years passed by so quickly. And when I look back, I remember the long way that brought me so far. Many steps had to be taken during my graduation race.

All the time Christian Kanzow has been playing an essential role as an outstanding trainer with clear mindset and strong knowledge. He helped me to reach the finish line – by coaching and motivating, especially at the critical milestones, but also by leaving me enough space for my own research and to have fun with well-known and at the same time well-unknown optimization fields.

Whereas the research activities in science are happening on the edge between known and unknown things, the science itself is built up by the cornerstones like networking, collaboration and team work. Herewith I take the chance to thank all my colleagues and friends, in particular, Sebastian Karl and Martin Sprengel, for always being there for me. Furthermore, I thank Oliver Stein from Karlsruhe Institute of Technology for his cooperation with Christian Kanzow and me on two research papers. Additionally, I thank my colleague Tim Hoheisel for our research teamwork on two other papers with Christian Kanzow.

Next I would particularly like to thank my colleague Axel Dreves. He helped me to start my journey and to travel through unknown mathematical areas. Axel backed me up at the beginning as well as at the end of my dissertation.

Of course, there is also a young woman with great attitude to be mentioned here – Alexandra Schwartz has become my friend during the last three years. Her sharp mind and strong focus on results were benefits for many colleagues here in Würzburg. Thank you for ice cream and sweets, too!

The science progress is something that does not happen suddenly. Working day by day and step by step, exploring the unknown and focusing on the results – that is the cornerstone of the success in science. I am very grateful to DFG (Deutsche Forschungsgemeinschaft) and to Elite Network of Bavaria for my grants and for being a participant of the international doctorate program “Identification, Optimization and Control with Application in Modern Technologies”. Both organizations have laid the foundation for my success. Their financial support really helped me to concentrate on my investigations. Moreover, they enabled me to join a great international conference in Lisbon and an exciting summer school in Cetraro (Italy).

My family was and is a part of my life. My parents, Alexander Unrainchenko and Lilia Harms, and three sisters, Helena, Irina, and Anastasia, are giving me important emotional support during my whole life with all ups and downs and also during my doctorate period. I thank them all.

On top of this, my special thanks go to my fiancé, Siegfried Schalm. Since more than ten years he has been supporting and motivating me in all my professional destinations and giving me the desired balance. You are my life and sunshine.

Contents

1. Introduction **1**

2. Preliminaries **3**
2.1. Tools from Variational Analysis 3
2.2. A Piecewise Smoothness Result for a Convex Parametric Nonlinear Program . . 7

I. Generalized Nash Equilibrium Problems 13

3. Background on Generalized Nash Equilibrium Problems **15**
3.1. Definition and Overview 15
3.2. Reformulations as a Constrained or Unconstrained Optimization Problem 18

4. Smoothness Properties of a Primal Gap Function for Generalized Nash Equilibrium Problems **23**
4.1. Continuity Properties and Domain 23
4.2. Differentiability Properties 30

5. Smoothness Properties of a Dual Gap Function for Generalized Nash Equilibrium Problems **41**
5.1. A Smooth Dual Gap Function 41
5.2. Second-Order Properties 58

6. Numerical Results for Generalized Nash Equilibrium Problems **61**
6.1. Primal Gap Function Approach 61
6.2. Dual Gap Function Approach 64
6.3. Comparing the Two Approaches 69

II. Quasi-Variational Inequalities 71

7. Background on Quasi-Variational Inequalities **73**
7.1. Definition and Overview 73
7.2. Preliminaries on Gap Functions 75
7.2.1. Gap Functions for Variational Inequalities 75
7.2.2. Gap Functions for Quasi-Variational Inequalities 76

8. Special Classes of Quasi-Variational Inequalities **81**
8.1. Quasi-Variational Inequalities with Generalized Moving Sets 81
8.2. Quasi-Variational Inequalities with Set-valued Mappings in Product Form 86
8.3. Application to Generalized Nash Equilibrium Problems 88

9. Smoothness Properties of a Primal Gap Function for Quasi-Variational Inequalities **91**
9.1. Continuity Properties and Domain 91
9.2. Differentiability Properties 94

10. Smoothness Properties of a Dual Gap Function for Quasi-Variational Inequalities **103**
10.1. A Smooth Dual Gap Function 103
10.2. Second-Order Properties 118

11. Numerical Results for Quasi-Variational Inequalities **123**
11.1. Primal Gap Function Approach 123
11.2. Dual Gap Function Approach 126

Final Remarks **129**

Bibliography **131**

Abbreviations

ACQ	Abadie constraint qualification
CRCQ	constant rank constraint qualification
DC	difference of two convex functions
CG	conjugate gradient (method)
GNEP	generalized Nash equilibrium problem
GSG	global spectral gradient (method)
isc	inner semicontinuous (set-valued mapping)
KKT	Karush-Kuhn-Tucker (conditions)
LICQ	linear independence constraint qualification
LP	linear program
lsc	lower semicontinuous (function)
MFCQ	Mangasarian Fromovitz constraint qualification
NEP	Nash equilibrium problem
osc	outer semicontinuous (set-valued mapping)
PC^1	piecewise smooth (function)
QVI	quasi-variational inequality
SLQP	sequential linear-quadratic programming
SMFC	strict Mangasarian Fromovitz condition
VI	variational inequality

Notation

Spaces

$\mathbb{N}$	the natural numbers
$\mathbb{R}$	the real numbers
$\mathbb{R}_{\geq}$	the nonnegative real numbers
$\mathbb{R}_{>}$	the positive real numbers
$\mathbb{R}^n$	the n-dimensional real vector space
$\mathbb{R}^n_{\geq}$	the nonnegative orthant in $\mathbb{R}^n$
$\mathbb{R}^n_{>}$	the positive orthant in $\mathbb{R}^n$

Sets

$\{x\}$	the set consisting of the vector x
$\lvert S \rvert$	the cardinality of S
$S_1 \subseteq S_2$	S_1 is a subset of S_2
$S_1 \backslash S_2$	the set of elements contained in S_1 but not in S_2
$S_1 \cap S_2$	the intersection of S_1 and S_2
$S_1 \cup S_2$	the union of S_1 and S_2
$S_1 \times S_2$	the cartesian product of S_1 and S_2
$D \cdot \Omega$	$D \cdot \Omega := \{Dw \mid w \in \Omega\}$ for a set $\Omega \subseteq \mathbb{R}^n$ and a matrix $D \in \mathbb{R}^{m \times n}$
$]x_1, x_2[$	an open interval in $\mathbb{R}$
$[x_1, x_2]$	a closed interval in $\mathbb{R}$

Vectors and Matrices

$x \in \mathbb{R}^n$	a column vector in $\mathbb{R}^n$
(x, y)	the column vector $(x^T, y^T)^T$
x_i	the i-th component of x
x_J	the vector in $\mathbb{R}^{\lvert J \rvert}$ with $J \subseteq \{1, \ldots, n\}$ consisting of the components x_i, $i \in J$
supp x	the support of a vector $x \in \mathbb{R}^n$, supp $x = \{i \mid x_i \neq 0\} \subseteq \{1, \ldots, n\}$
$x \geq y$	componentwise comparison $x_i \geq y_i$, $i = 1, \ldots, n$
$x > y$	componentwise comparison $x_i > y_i$, $i = 1, \ldots, n$
$\max\{x, y\}$	the vector whose i-th component is $\max\{x_i, y_i\}$
$\lVert x \rVert$	the Euclidean norm of x, $\lVert x \rVert := \lVert x \rVert_2 = \sqrt{\sum_{i=1}^n x_i^2}$
$I_n \in \mathbb{R}^{n \times n}$	the identity matrix of size $n \times n$
diag (x)	the diagonal matrix with diagonal elements x_i for a vector $x \in \mathbb{R}^n$

Cones

$\mathcal{T}_X(x)$	the Bouligand tangent cone to X in x
$\mathcal{L}_X(x)$	the linearized tangent cone to X in x

Functions

$f : \mathbb{R}^n \to \mathbb{R}^m$	a function that maps $\mathbb{R}^n$ to $\mathbb{R}^m$
$f_i : \mathbb{R}^n \to \mathbb{R}$	the i-th component of f
$\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$	a set-valued mapping that maps $\mathbb{R}^n$ to subsets of $\mathbb{R}^m$
$\operatorname{dom} F$	the domain of a function or set-valued mapping F
$\operatorname{epi} f$	the epigraph of a function f
$\operatorname{gph} \Phi$	the graph of the set-valued mapping Φ
$\nabla f(x)$	the gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ at the point $x \in \mathbb{R}^n$, column vector
$\nabla_x f(x, y)$	the gradient of f with respect to x only
$Df(x)$	the Jacobian of $f : \mathbb{R}^n \to \mathbb{R}^m$ at $x \in \mathbb{R}^n$
$\nabla f(x)$	the transposed Jacobian of $f : \mathbb{R}^n \to \mathbb{R}^m$ at $x \in \mathbb{R}^n$
$\nabla^2 f(x)$	the Hessian of $f : \mathbb{R}^n \to \mathbb{R}$ at x
$\nabla^2_{xx} f(x, y)$	the Hessian of $f : \mathbb{R}^n \to \mathbb{R}$ with respect to x only
$\partial f(x)$	the subdifferential of a convex function $f : \mathbb{R}^n \to \mathbb{R}$ at the point $x \in \mathbb{R}^n$
$P_X(x)$	the Euclidean projection of a vector x onto the set X
δ_X	the indicator function of a set X

Sequences

$\{x^k\} \subseteq \mathbb{R}^n$	a sequence in R^n
$x^k \to \bar{x}$	a convergent sequence with limit $\bar{x}$
$\lim_{k\to\infty} x^k$	limit of the convergent sequence x^k
$\{t_k\} \subseteq \mathbb{R}$	a real-valued sequence
$t_k \downarrow \bar{t}$	a convergent sequence with limit $\bar{t}$ and $t_k > \bar{t}$ for all $k \in \mathbb{N}$

Quantifiers and Logical Connectives

$\forall$	for all
$\exists$	there exists
$\exists!$	there exists exactly one
$\wedge$	conjunction (and)

1. Introduction

In this thesis we investigate smoothness properties of primal and dual reformulations of two problem classes: generalized Nash equilibrium problems (GNEPs) and quasi-variational inequality problems (QVIs). GNEPs have widespread applications in various fields such as economics, engineering, computer science, operations research, telecommunications, and deregulated markets (see [47] for references), whereas QVIs provide a generalization of GNEPs and can be very helpful by modeling complex equilibrium situations that occur not only in the context of GNEPs but also in many other fields such as mechanics, physics, statistics, transportation, and biology (see [50] for references).

In this thesis we first analyze player convex GNEPs in Part I and then their generalization QVIs in Part II. Before that, we present some background material from convex and variational analysis and obtain a smoothness result for a class of parametric optimization problems that fits into our framework in Chapter 2.

The two main parts of this thesis have similar structure: After the definitions and a literature overview on the particular problem classes, we consider reformulations of player convex GNEPs and QVIs as possibly nonsmooth constrained or unconstrained minimization problems based on well-known primal gap functions, see Chapters 3 and 7, respectively. Additionally, we analyze three special classes of QVIs in Chapter 8: The first class is a generalization of QVIs with 'moving sets', the second are QVIs with set-valued mappings in product form, and the third are QVIs as an important application to GNEPs. In Chapters 4 and 9 we investigate the continuity of a corresponding primal gap function for player convex GNEPs or QVIs and relate the points at which these functions are continuous to interior points of the domain of these functions, where subsequently the differentiability properties of these functions are studied. Our main result is that, apart from special cases, the primal gap functions are differentiable at all local minimizers of the respective reformulation for player convex GNEPs or QVIs.

In Chapters 5 and 10, respectively, we study smoothness properties of an unconstrained optimization reformulation of a class of GNEPs or QVIs arising from a corresponding dual gap function. These dual functions are based on an idea by Dietrich [30] and developed by rewriting the primal gap functions, which are analyzed in Chapters 4 and 9, as a difference of two strongly convex functions and employing suitable duality theory to these two functions. These dual gap functions are continuously differentiable and, under suitable assumptions, have piecewise smooth gradients.

The results in Chapters 4 and 5 as well as in Chapters 8, 9, and 10 motivate to use certain smooth optimization techniques for both the primal and dual reformulations, and therefore we present some numerical results in Chapters 6 and 11 based on primal and dual approaches.

We conclude this thesis by summarizing the main results, discussing open questions, and giving some suggestions on future research topics.

The main results of this thesis are illustrated by accompanying examples and figures. Note that more detailed summaries of all chapters of this thesis are given at the beginning of the respective chapter.

The results of this thesis have already been published in [72, 73, 74, 75]. All of them are joint work with my supervisor Christian Kanzow. Furthermore, the papers [74, 75] arose under great support of our DFG project partner Oliver Stein, and the papers [72, 73] were developed in cooperation with my colleague Tim Hoheisel.

2. Preliminaries

This chapter is divided into two sections: First, the basic variational tools needed for the subsequent analysis in this thesis are provided in Section 2.1. Secondly, an auxiliary result from parametric optimization is proven in Section 2.2. This result has been shown in the paper [73] and will be used in Section 5.2.

2.1. Tools from Variational Analysis

In this section we review certain concepts from variational and convex analysis employed in the sequel. The notation and terminology is, in large parts, based on [117].

We first restate some definitions for set-valued mappings, see, for example, [117, Chapter 5] for more details.

Definition 2.1 *Let $X \subseteq \mathbb{R}^n$, and $\Phi : X \rightrightarrows \mathbb{R}^m$ be a set-valued mapping. Then Φ is called*

(a) inner semicontinuous (isc) *at $\bar{x} \in X$ relative to X if for all sequences $\left\{x^k\right\} \subseteq X$ with $x^k \to \bar{x}$ and all $\bar{z} \in \Phi(\bar{x})$ there exists a number $k_0 \in \mathbb{N}$ and a sequence $z^k \to \bar{z}$ such that $z^k \in \Phi(x^k)$ for all $k \geq k_0$;*

(b) outer semicontinuous (osc) *at $\bar{x} \in X$ relative to X if for all sequences $\left\{x^k\right\} \subseteq X$ with $x^k \to \bar{x}$ and all sequences $z^k \to \bar{z}$ with $z^k \in \Phi(x^k)$ for all $k \in \mathbb{N}$ sufficiently large we have $\bar{z} \in \Phi(\bar{x})$;*

(c) continuous *at $\bar{x} \in X$ relative to X if it is isc and osc at $\bar{x} \in X$ relative to X;*

(d) isc, osc *or* continuous *on X relative to X if it is isc, osc or continuous at every $x \in X$ relative to X, respectively;*

(e) graph-convex *if its graph*

$$\operatorname{gph} \Phi := \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^m \mid z \in \Phi(x)\}$$

is a convex set.

In the literature, isc or osc set-valued mappings are also called *open* or *closed* set-valued mappings, respectively (see [82]). Alternative concepts to inner and outer semicontinuity are the concepts of lower or upper semicontinuity in the sense of Berge (see [17, 18]). Note that the terms of inner semicontinuity and lower semicontinuity are equivalent, but the terms of outer semicontinuity and upper semicontinuity differ from each other.

In our further analysis we use the concept of continuity of a function relative to a set.

Definition 2.2 *Given a set $X \subseteq \mathbb{R}^n$ and a function $f : X \to \mathbb{R}^m$, we say that the function f is* continuous at $\bar{x} \in X$ relative to X *if $f(x^k) \to f(\bar{x})$ holds for all sequences $\{x^k\} \subset X$ converging to the vector $\bar{x}$.*

Note that relative properties of functions and set-valued mappings are meant relative to $\mathbb{R}^n$ if not stated otherwise. The next result follows immediately from [82, Corollaries 8.1 and 9.1] and will be used in Sections 5.1, 8.1, 9.1 and 10.1.

Lemma 2.3 *Let $X \subseteq \mathbb{R}^n$ be an arbitrary set, and $v : (z, x) \in \mathbb{R}^m \times X \mapsto \mathbb{R}$ be concave in z for fixed x and continuous on $\mathbb{R}^m \times X$. Let $\Phi : X \rightrightarrows \mathbb{R}^m$ be a set-valued mapping that is osc on a neighborhood of $\bar{x}$ and isc at $\bar{x}$ relative to X, and the set $\Phi(x)$ be convex in a neighborhood of $\bar{x}$. Define*

$$Z(x) := \left\{ \zeta \in \Phi(x) \;\middle|\; \sup_{z \in \Phi(x)} v(z, x) = v(\zeta, x) \right\},$$

and assume that $Z(\bar{x})$ is a singleton. Then the set-valued mapping $x \mapsto Z(x)$ is continuous at $\bar{x}$ relative to X.

The following properties of an osc and graph-convex set-valued mapping will also be used in our subsequent analysis.

Lemma 2.4 *Let $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be an osc and graph-convex set-valued mapping. Then the following statements hold:*

(a) The sets $\Phi(x)$ are closed and convex (possibly empty).

(b) For all $x_1, x_2 \in \mathbb{R}^n$ with $\Phi(x_i) \neq \emptyset$ for $i = 1, 2$, and all $t \in [0, 1]$, we have

$$t\Phi(x_1) + (1 - t)\Phi(x_2) \subseteq \Phi(tx_1 + (1 - t)x_2),$$

in particular, the set on the right-hand side is nonempty.

(c) The set $\operatorname{gph} \Phi$ *is closed and convex.*

All statements are well known and follow directly from the respective definitions; regarding assertion (b), see [117, p. 155].

We next introduce some important concepts for extended real-valued functions, more precisely, for functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Handy tools for the analysis of such a function are its *epigraph*

$$\operatorname{epi} f := \{(x, \gamma) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \gamma\}$$

and its *domain*

$$\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}.$$

Note that a function f is called *proper* if $\operatorname{dom} f \neq \emptyset$. The important concepts for extended real-valued functions are summarized in the next definition.

Definition 2.5 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper.*

(a) f is called lower semicontinuous (lsc) *if* epi f *is closed.*

(b) f is called convex *if* epi f *is convex.*

(c) f is called strongly convex *with modulus $c > 0$ if $f - \frac{c}{2}\|\cdot\|^2$ is convex.*

(d) If f is convex and $\bar{x} \in \mathbb{R}^n$ then the (possibly empty) set

$$\partial f(\bar{x}) := \left\{ s \in \mathbb{R}^n \;\middle|\; f(\bar{x}) + s^T(x - \bar{x}) \le f(x) \quad \forall x \in \mathbb{R}^n \right\}$$

is called the subdifferential *of f at $\bar{x}$.*

(e) The conjugate *of f is the function $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$f^*(y) := \sup_{x \in \mathbb{R}^n} [x^T y - f(x)] = \sup_{x \in \operatorname{dom} f} [x^T y - f(x)].$$

Note that, in view of its definition, an lsc function is often called *closed*. Further note that the subdifferential $\partial f(\bar{x})$ of a proper and convex function f is nonempty if x lies in the (relative) interior of dom f, and that we have $\partial f(x) = \{\nabla f(x)\}$ for a convex and differentiable function f, see the monograph [81].

Given a set $X \subseteq \mathbb{R}^n$, a very prominent extended real-valued function is the *indicator function* $\delta_X : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\delta_X(x) := \begin{cases} 0, & \text{if } x \in X, \\ +\infty, & \text{if } x \notin X. \end{cases} \tag{2.1}$$

It is easily verified that δ_X is lsc if and only if X is closed, and convex if and only if X is convex.

The following result summarizes some well-known properties of the conjugate function.

Lemma 2.6 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function. Then the following statements hold:*

(a) The conjugate f^* of f is convex and lsc.

(b) The bi-conjugate function $f^{**} := (f^*)^*$ is convex and lsc.

(c) The inequality $f^{**}(x) \le f(x)$ holds for all $x \in \mathbb{R}^n$.

(d) The equality $f^{**}(x) = f(x)$ holds for all $x \in \mathbb{R}^n$ if and only if f is a (convex and) lsc function.

(e) The Fenchel inequality $f(x) + f^*(y) \ge x^T y$ holds for all $x, y \in \mathbb{R}^n$.

(f) The equality $f(\bar{x}) + f^*(\bar{y}) = \bar{x}^T \bar{y}$ holds if and only if $\bar{y} \in \partial f(\bar{x})$.

All statements can be found in [81, Chapter E]. Another useful observation on the conjugate function is restated in the following result, cf. [117, Proposition 12.60].

Lemma 2.7 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper, lsc, and convex. Then f is strongly convex with modulus $c > 0$ if and only if f^* is differentiable with ∇f^* Lipschitz continuous with modulus $\frac{1}{c}$.*

Besides of the standard concept of differentiability, we use also other concepts in Sections 4.2 and 9.2 summarized in the following definition.

Definition 2.8 *Let $U \subseteq \mathbb{R}^n$ be an open set. A function $f : U \to \mathbb{R}$ is*

(a) directionally differentiable *at a point $x \in U$ if the limit*

$$\lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}$$

exists for all directions $d \in \mathbb{R}^n$. The directional derivative *$f'(x, d)$ at x along the vector d is then defined by the corresponding limit.*

(b) directionally differentiable in the Hadamard sense *or simply* Hadamard directionally differentiable *at $x \in U$ if the limit*

$$\lim_{t \downarrow 0,\ d' \to d} \frac{f(x + td') - f(x)}{t}$$

exists for all directions $d \in \mathbb{R}^n$. The Hadamard directional derivative *$f'(x, d)$ at x along the vector d is then defined by the corresponding limit.*

(c) Gâteaux differentiable *if it is directionally differentiable and if the directional derivative is a linear function of the direction.*

Note that Hadamard directional differentiability implies the usual directional differentiability. Furthermore, if a function $f : U \to \mathbb{R}$ with an open set U is Gâteaux differentiable on U and the partial derivatives of f are continuous at $\bar{x} \in U$, then f is continuously differentiable at $\bar{x}$.

For the objective function

$$F(x) := \inf_{z \in S(x)} f(z, x) \quad \Big(\text{or } \ G(x) := \sup_{z \in S(x)} f(z, x)\Big) \tag{2.2}$$

with a function $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and a set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, we write

$$F(x) = \min_{z \in S(x)} f(z, x) \quad \Big(\text{or } \ G(x) = \max_{z \in S(x)} f(z, x)\Big)$$

for $x \in X \subseteq \mathbb{R}^n$ if the infimum (or supremum) is attained in (2.2) for all $x \in X$. We can write minimum (or maximum) instead of infimum (or supremum) immediately if the function f is strongly convex (or strongly concave) in z for each fixed $x \in X$ and the set $S(x)$ is nonempty, convex, and closed for all $x \in X$, since in such cases the minimization (or maximization) problem in F (or in G) with respect to $S(x)$ has a (unique) solution for all $x \in X$.

2.2. A Piecewise Smoothness Result for a Convex Parametric Nonlinear Program

A number of profound results for parametric optimization can be found in the monographs [11, 21, 27, 58, 90]. In this section we analyze smoothness properties of the solution mapping for a class of strongly convex parametric optimization problems, where the parameter only occurs in the objective function. The main result in this section might be known, but we could not find an explicit reference. The difference to the existing literature is that we assume the objective function to be strongly convex (not just convex) for each fixed parameter, which is, of course, a very restrictive assumption, but this assumption will be satisfied automatically in our applications. On the other hand, if for each fixed parameter the objective function of a convex parametric nonlinear program is strongly convex and the feasible set is nonempty, closed, and convex, this problem has a unique solution for any parameter. Additionally, if the constraints are independent of the parameter, it turns out that this solution depends continuously on the parameter even without the Mangasarian Fromovitz constraint qualification (MFCQ) or Slater condition. Note that, in general, this observation does not hold, cf. [27, 82, 114].

In this section we will see that the solution function of our parametric optimization problem is, under some standard assumptions, *piecewise smooth*. The analysis is carried out in the spirit of the results from [114, 82] and the piecewise smoothness result for the projection mapping from [54, 107]. We commence by introducing the concept of piecewise smoothness, see [54, 119] for comprehensive accounts on the topic.

Definition 2.9 *A continuous function $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is called* piecewise smooth *or* PC^1 *near $\bar{x} \in D$ if there exists an open neighborhood $U \subseteq D$ of $\bar{x}$ and a finite family of continuously differentiable functions $f_i : U \to \mathbb{R}^m$, $i = 1, \dots, l$, such that $f(x) \in \{f_1(x), \dots, f_l(x)\}$ for all $x \in U$.*

Now, for a parameter $v \in \mathbb{R}^n$, consider the optimization problem

$$\min_{u\in\mathbb{R}^m} \phi(u, v) \quad \text{subject to} \quad c_j(u) \leq 0 \quad (j = 1, \dots, p), \tag{$P(v)$}$$

where $\phi : (u, v) \in \mathbb{R}^m \times \mathbb{R}^n \mapsto \mathbb{R}$ is strongly convex in u for each fixed $v \in \mathbb{R}^n$ and continuous on $\mathbb{R}^m \times \mathbb{R}^n$, and the functions $c_j : \mathbb{R}^m \to \mathbb{R}$, $j = 1, \dots, p$, are convex and continuous. Let

$$\mathcal{F} := \left\{u \in \mathbb{R}^m \;\middle|\; c_j(u) \leq 0 \quad \forall j = 1, \dots, p\right\} \tag{2.3}$$

denote the feasible set, which is closed and convex, and is supposed to be nonempty. Under the above assumptions, the next lemma shows that the solution mapping of problem $P(v)$ is continuous.

Lemma 2.10 *The solution mapping $u^* : \mathbb{R}^n \to \mathbb{R}^m$ of the problem $P(v)$ given by*

$$u^*(v) := \operatorname*{argmin}_{u\in\mathcal{F}} \phi(u, v) \tag{2.4}$$

is well-defined and continuous.

Proof. Under the assumptions on the functions ϕ and c_j, $j = 1, \dots, p$, the objective function is strongly convex in u and the feasible set $\mathcal{F}$ is nonempty, closed, and convex. Hence the problem $P(v)$ is uniquely solvable for all $v \in \mathbb{R}^n$. Therefore, for each $v \in \mathbb{R}^n$, there exists a unique vector $u^*(v)$ solving (2.4). Therefore, the solution mapping u^* is well-defined. The continuity of the mapping u^* follows from Lemma 2.3, which is based on [82, Corollaries 8.1 and 9.1]. □

Now, for $u \in \mathcal{F}$, we define the active set

$$J_0(u) := \left\{ j \in \{1, \dots, p\} \;\middle|\; c_j(u) = 0 \right\}. \tag{2.5}$$

Due to Lemma 2.10, for all $v \in \mathbb{R}^n$, the sets

$$J(v) := J_0(u^*(v))$$

are well-defined.

For the remainder of this section, we assume that all functions defining $P(v)$ are, in addition to the convexity property, twice continuously differentiable. As a reminder and a reference point, all of the demanded properties are summarized below.

Assumption 2.11 *The functions ϕ and c_j, $j = 1, \dots, p$, defining $P(v)$ are assumed to have the following properties:*

(a) The objective function $\phi : (u, v) \in \mathbb{R}^m \times \mathbb{R}^n \mapsto \mathbb{R}$ is strongly convex in u for each fixed $v \in \mathbb{R}^n$ and twice continuously differentiable on $\mathbb{R}^m \times \mathbb{R}^n$.

(b) The constraints $c_j : \mathbb{R}^m \to \mathbb{R}$, $j = 1, \dots, p$, are convex and twice continuously differentiable.

(c) The feasible set $\mathcal{F} := \{u \in \mathbb{R}^m \mid c_j(u) \leq 0 \quad \forall j = 1, \dots, p\}$ is nonempty.

For $v \in \mathbb{R}^n$ and a subset $J \subseteq J(v)$, we define $H^J(\cdot, v, \cdot) : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^{m+p}$ by

$$H^J(u, v, \lambda) := \begin{pmatrix} \nabla_u \phi(u, v) + \sum_{j \in J} \lambda_j \nabla c_j(u) \\ c_J(u) \\ \lambda_{\hat{J}} \end{pmatrix}$$

with $\hat{J} := \{1, \dots, p\} \setminus J$. Then the following result is easily proven.

Lemma 2.12 *Let Assumption 2.11 hold, let $v \in \mathbb{R}^n$, and let $J \subseteq J(v)$ such that the vectors $\nabla c_j(u)$ $(j \in J)$ are linearly independent. Then the Jacobian $D_{(u,\lambda)} H^J(u, v, \lambda)$ is nonsingular for all $\lambda_J \geq 0$.*

Proof. After reordering the components of λ accordingly, we obtain

$$D_{(u,\lambda)} H^J(u, v, \lambda) = \begin{pmatrix} \nabla^2_{uu} \phi(u, v) + \sum_{j \in J} \lambda_j \nabla^2 c_j(u) & Dc_J(u)^T & 0 \\ Dc_J(u) & 0 & 0 \\ 0 & 0 & I_{|\hat{J}|} \end{pmatrix}.$$

Since the functions c_j, $j = 1, \dots, p$, are convex and ϕ is strongly convex in the first variable for each fixed $v \in \mathbb{R}^n$, the matrix $\nabla^2_{uu}\phi(u, v) + \sum_{j\in J} \lambda_j \nabla^2 c_j(u)$ is positive definite for all $\lambda_J \geq 0$. Hence the assertion follows from the linear independence of the vectors $\nabla c_j(u)$ $(j \in J)$. □

We next introduce the *constant rank constraint qualification* due to [85], which occurs as a standard assumption in the context of parametric optimization and piecewise smoothness results, see, for example, [28, 107, 114].

Definition 2.13 *Let $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, p$, be continuously differentiable, and let $\mathcal{F}$ and $J_0(u)$ be defined by* (2.3) *and* (2.5)*, respectively. We say that the* constant rank constraint qualification (CRCQ) *holds at $\bar{u} \in \mathcal{F}$ (with respect to the set $\mathcal{F}$) if there exists a neighborhood U of $\bar{u}$ such that for every $J \subseteq J_0(\bar{u})$ the set $\{\nabla c_j(u) \mid j \in J\}$ has constant rank (depending on J) for all $u \in U$.*

Note that CRCQ is a local property of the feasible set $\mathcal{F}$ in the sense that if CRCQ holds at $\bar{u}$, it also holds at u for all $u \in \mathcal{F}$ sufficiently close to $\bar{u}$.

CRCQ allows us to prove the next theorem on the piecewise smoothness of the solution mapping of the program $P(v)$, which is the main result of this section.

Theorem 2.14 *Let $\bar{v} \in \mathbb{R}^n$, and suppose that Assumption 2.11 is fulfilled. Then there exists a neighborhood $\bar{V}$ of $\bar{v}$ such that the function $u^* : \mathbb{R}^n \to \mathbb{R}^m$ defined in* (2.4) *is PC^1 on $\bar{V}$, provided CRCQ holds at $\bar{u} := u^*(\bar{v}) \in \mathcal{F}$.*

Proof. The argumentation in this proof is similar to those in [54, 107, 114].

For $v \in \mathbb{R}^n$ we define

$$M(v) := \left\{\lambda \in \mathbb{R}^p \;\middle|\; (u^*(v), \lambda) \text{ is a KKT point of } P(v)\right\}$$

as the set of KKT multipliers for $P(v)$ at $u^*(v)$. Since CRCQ at $\bar{u} = u^*(\bar{v})$ is inherited to a whole neighborhood and because u^* is continuous by Lemma 2.10, there exists a neighborhood V of $\bar{v}$ such that CRCQ holds at $u^*(v)$ for all $v \in V$. In particular, since CRCQ yields KKT multipliers at a local minimizer (see [85, Proposition 2.3]), we have $M(v) \neq \emptyset$ for all $v \in V$. Hence it follows from [80, Lemma 3.2] that the set

$$B(v) := \left\{J \subseteq J(v) \;\middle|\; \nabla c_j(u^*(v)) \ (j \in J) \text{ linearly independent } \wedge \ \exists \lambda \in M(v) : \operatorname{supp} \lambda \subseteq J\right\}$$

is nonempty for all $v \in V$. Moreover, from [80, Lemma 3.3] it follows that

$$\forall v \in V,\ J \in B(v)\ \exists!\ \lambda^{*,J}(v) \in M(v) :\ H^J(u^*(v), v, \lambda^{*,J}(v)) = 0. \tag{2.6}$$

Note that, necessarily, $\operatorname{supp} \lambda^{*,J}(v) \subseteq J$, and that $\lambda^{*,J}(v)$ is nonnegative.

Now, we have by (2.6) for $J \in B(\bar{v})$ a uniquely determined $\bar{\lambda}^J := \lambda^{*,J}(\bar{v})$ such that $\bar{\lambda}^J \in M(\bar{v})$ and $H^J(\bar{u}, \bar{v}, \bar{\lambda}^J) = 0$ hold. As $J \in B(\bar{v})$, the vectors $\nabla c_j(\bar{u})$ $(j \in J)$ are linearly independent, hence Lemma 2.12 together with $\bar{\lambda}^J \geq 0$ implies that $D_{(u,\lambda)}H^J(\bar{u}, \bar{v}, \bar{\lambda}^J)$ is nonsingular. Thus, the implicit function theorem (see, e.g., [4, Theorem 8.2]) yields neighborhoods V^J of $\bar{v}$ and N^J of $(\bar{u}, \bar{\lambda}^J)$, and a continuously differentiable function $(u^J, \lambda^J) : V^J \to N^J$ such that

$$u^J(\bar{v}) = \bar{u}, \quad \lambda^J(\bar{v}) = \bar{\lambda}^J, \quad \text{and} \quad H^J(u^J(v), v, \lambda^J(v)) = 0 \quad \forall v \in V^J, \tag{2.7}$$

and for all $v \in V^J$ the vector $(u^J(v), \lambda^J(v))$ is the unique solution of

$$H(u, v, \lambda) \overset{!}{=} 0, \quad (u, \lambda) \in N^J.$$

Note that, without loss of generality, we can assume that $V^J \subseteq V$.

Now, set

$$\bar{V} := \bigcap_{J \in B(\bar{v})} V^J \subseteq V.$$

Since $B(\bar{v})$ is finite, $\bar{V}$ is still a neighborhood of $\bar{v}$. Moreover, in view of [80, Lemma 3.5 (b)], we can assume without loss of generality that $B(v) \subseteq B(\bar{v})$ for all $v \in \bar{V}$. We will now prove that, with a possibly smaller neighborhood of $\bar{v}$, which we still denote by $\bar{V}$, we have

$$u^*(v) \in \left\{ u^J(v) \;\middle|\; J \in B(\bar{v}) \right\} \quad \forall v \in \bar{V}. \tag{2.8}$$

Then it follows that $u^* : \bar{V} \to \mathbb{R}^m$ is in fact PC^1, as $\{u^J : \bar{V} \to \mathbb{R}^m \mid J \in B(\bar{v})\}$ is a finite family of continuously differentiable functions, and u^* is continuous by Lemma 2.10. The desired inclusion in (2.8) follows immediately if we can show that

$$\forall v \in \bar{V},\ \forall J \in B(v) :\ u^*(v) = u^J(v) \tag{2.9}$$

since $B(v) \subseteq B(\bar{v})$ for all $v \in \bar{V}$. Note that this does not imply that $u^* = u^J$ holds locally (which would imply u^* to be smooth) since the index set J also depends on v.

For these purposes, let $v \in \bar{V} (\subseteq V)$ and $J \in B(v)$. Due to (2.6), there exists a unique multiplier $\lambda^{*,J}(v) \in M(v)$ such that $H^J(u^*(v), v, \lambda^{*,J}(v)) = 0$. As it was done after (2.6) we can once again use the implicit function theorem to show the existence of neighborhoods V^J of $\bar{v}$ and N^J of $(\bar{u}, \bar{\lambda}^J)$ as well as a continuously differentiable function $(u^J, \lambda^J) : V^J \to N^J$ such that (2.7) holds. Moreover, for all $v \in V^J$, the vector $(u^J(v), \lambda^J(v))$ is the unique solution of

$$H(u, v, \lambda) \overset{!}{=} 0, \ \ (u, \lambda) \in N^J.$$

Hence, in order to prove (2.9), it suffices to show that

$$\forall v \in \bar{V} \text{ sufficiently close to } \bar{v},\ \forall J \in B(v) :\ (u^*(v), \lambda^{*,J}(v)) \in N^J.$$

Suppose that this is not true: Then there exists a convergent sequence $v^k \to \bar{v}$ with $v^k \in \bar{V}$ and a sequence of index sets $J_k \in B(v^k)$ such that

$$(u^*(v^k), \lambda^{*,J_k}(v^k)) \notin N^{J_k} \quad \forall k \in \mathbb{N}. \tag{2.10}$$

As $B(\bar{v})$ is finite and $B(v^k) \subseteq B(\bar{v})$ for all $k \in \mathbb{N}$, we can assume without loss of generality that $J_k = \bar{J}$ for all $k \in \mathbb{N}$. From (2.6) we infer that

$$0 = \nabla_u \phi(u^*(v^k), v^k) + \sum_{j \in \bar{J}} (\lambda^{*,\bar{J}}(v^k))_j \nabla c_j(u^*(v^k)) \quad \forall k \in \mathbb{N}. \tag{2.11}$$

By continuity of all functions involved, the linear independence of the vectors $\nabla c_j(u^*(v^k))$ $(j \in \bar{J})$ for all $k \in \mathbb{N}$ together with the assumed CRCQ condition and the fact that supp $\lambda^{*,\bar{J}}(v^k) \subseteq \bar{J}$, we infer that the sequence $\lambda^{*,\bar{J}}(v^k)$ is convergent, that is, there exists $\lambda^{*,\bar{J}}$ such that $\lambda^{*,\bar{J}}(v^k) \to \lambda^{*,\bar{J}}$ with supp $\lambda^{*,\bar{J}} \subseteq \bar{J}$. Hence passing to the limit in (2.11) yields

$$0 = \nabla_u \phi(\bar{u}, \bar{v}) + \sum_{j \in \bar{J}} (\lambda^{*,\bar{J}})_j \nabla c_j(\bar{u}).$$

On the other hand, also $\bar{\lambda}^{\bar{J}}$ solves the above equation. Due to the linear independence of the gradients $\nabla c_j(\bar{u})$ $(j \in \bar{J})$, and the fact that supp $\bar{\lambda}^{\bar{J}} \cup$supp $\lambda^{*,\bar{J}} \subseteq \bar{J}$, we have $\lambda^{*,\bar{J}} = \bar{\lambda}^{\bar{J}}$. Therefore, we infer that $\lambda^{*,J_k}(v^k) \to \bar{\lambda}^{\bar{J}}$. In view of $u^*(v^k) \to \bar{u}$ by continuity, we obtain that $(u^*(v^k), \lambda^{*,J_k}(v^k)) \in N^{\bar{J}}$ for all k sufficiently large contradicting (2.10). Hence the proof is complete. □

Part I.

Generalized Nash Equilibrium Problems

3. Background on Generalized Nash Equilibrium Problems

Chapter 3 contains some preparations for the analysis in Part I. First we give a background on generalized Nash equilibrium problems in Section 3.1. Then we review some reformulations of a generalized Nash equilibrium problem as a constrained or unconstrained minimization problem in Section 3.2, which are relevant to our subsequent analysis. The results of Section 3.2 are based on Section 2 in [39].

3.1. Definition and Overview

A generalized Nash equilibrium problem (GNEP) consists of $N \in \mathbb{N}$ players $\nu = 1, \ldots, N$. Each player ν controls his decision variable $x^\nu \in \mathbb{R}^{n_\nu}$, $n_\nu \in \mathbb{N}$, such that the vector $x = (x^1, \ldots, x^N) \in \mathbb{R}^n$ with $n = n_1 + \ldots + n_N$ describes the decision vector of all players. In order to emphasize the role of player ν's variable x^ν within the vector x, we often write $x = (x^\nu, x^{-\nu})$. Furthermore, each player ν has a cost function $\theta_\nu : \mathbb{R}^n \to \mathbb{R}$ and a strategy space $X_\nu(x^{-\nu}) \subseteq \mathbb{R}^{n_\nu}$ defined by the set-valued mapping $X_\nu : \mathbb{R}^{n-n_\nu} \rightrightarrows \mathbb{R}^{n_\nu}$. Both the cost function and the strategy space can depend on the other players' decisions $x^{-\nu}$. Define the set-valued mapping $\Omega : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ by

$$\Omega(x) := X_1(x^{-1}) \times \ldots \times X_N(x^{-N}). \tag{3.1}$$

Then the GNEP consists in finding a vector $\bar{x} = (\bar{x}^1, \ldots, \bar{x}^N) \in \Omega(\bar{x})$ such that for each index $\nu \in \{1, \ldots, N\}$ the vector $\bar{x}^\nu$ solves

$$Q^\nu(\bar{x}^{-\nu}) : \qquad \min_{x^\nu} \ \theta_\nu(x^\nu, \bar{x}^{-\nu}) \quad \text{subject to} \quad x^\nu \in X_\nu(\bar{x}^{-\nu}). \tag{3.2}$$

A solution point $\bar{x}$ of a GNEP is called a *generalized Nash equilibrium*. If the set $X_\nu(x^{-\nu})$ is independent of $x^{-\nu}$ for all $\nu = 1, \ldots, N$, that is, $X_\nu(x^{-\nu}) = X_\nu$ for all strategies $x \in \mathbb{R}^n$ and all players ν with some constant sets X_ν, then the GNEP reduces to the so-called *Nash equilibrium problem* (NEP).

A natural assumption to make GNEPs numerically tractable is the convexity of the problems $Q^\nu(\bar{x}^{-\nu})$, $\nu = 1, \ldots, N$, in the respective players' variable x^ν. Assumption 3.1 will be a standing assumption throughout this part of the thesis.

Assumption 3.1 *(a) The cost functions $\theta_\nu(\cdot, x^{-\nu})$, $\nu = 1, \ldots, N$, are convex for each fixed vector $x^{-\nu} \in \mathbb{R}^{n-n_\nu}$.*

(b) The strategy spaces $X_\nu(x^{-\nu})$, $\nu = 1, \ldots, N$, are closed and convex.

GNEPs satisfying Assumption 3.1 are called *player convex*. Apart from very few exceptions, see [33, 108, 109], it is the most general form of a GNEP studied in the literature. Note that Assumption 3.1 (a) provides the continuity of $\theta_\nu(\cdot, x^{-\nu})$ on $\mathbb{R}^{n_\nu}$ for each fixed vector $x^{-\nu} \in \mathbb{R}^{n-n_\nu}$, and that Assumption 3.1 (b) is satisfied if, for example, the strategy spaces X_ν are defined by

$$X_\nu(x^{-\nu}) \;:=\; \{x^\nu \in \mathbb{R}^{n_\nu} \mid g_i^\nu(x^\nu, x^{-\nu}) \leq 0 \quad \forall i = 1, \ldots, m_\nu\} \tag{3.3}$$

with functions $g_i^\nu : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m_\nu$, $\nu = 1, \ldots, N$, that are continuous on $\mathbb{R}^n$ and convex in x^ν for each fixed $x^{-\nu} \in \mathbb{R}^{n-n_\nu}$. We will assume this representation of the strategy spaces in Chapter 4 and Section 5.2 and use the general definition of these sets otherwise. In cases with the representation (3.3), we will also use the notation $g^\nu := (g_1^\nu, \ldots, g_{m_\nu}^\nu)$, $g^\nu : \mathbb{R}^n \to \mathbb{R}^{m_\nu}$, for the constraint functions.

Player convex GNEPs have a widely studied subclass, which is called *jointly convex* GNEPs and characterized by the existence of a fixed convex set $X \subseteq \mathbb{R}^n$ such that the strategy space of each player ν, $\nu = 1, \ldots, N$, is given by

$$X_\nu(x^{-\nu}) = \;\{x^\nu \in \mathbb{R}^{n_\nu} \mid (x^\nu, x^{-\nu}) \in X\}.$$

In this case, the set-valued mappings X_ν from (3.3) are all defined by the same continuous constraint function $g^1 = g^2 = \ldots = g^N =: g$ and the components of the constraint function g are convex in the whole vector $x = (x^1, \ldots, x^N)$. Although we will not study this problem class in detail in this thesis, Remark 4.27 in Section 4.2 will summarize the implications of joint convexity for our approaches from Chapter 4.

GNEPs were formally introduced in 1952/54 by Arrow and Debreu [26, 6] as a generalization of NEPs defined in 1950/51 by Nash [94, 95]. The roots of NEPs go back to the concept of equilibrium in the context of an oligopolistic economy observed by Cournot [25] already in 1838. A detailed historical overview for GNEPs can be found in [47, 24]. GNEPs have widespread applications in various fields such as economics, engineering, mathematics, computer science, operations research, telecommunications, and deregulated markets. Many solution methods for GNEPs exist in the meantime, which work under different sets of assumptions. The interested reader is referred to [47, 56, 24, 60, 40, 41] and references therein for a detailed survey of applications, theory, and algorithms up to the year 2014.

There are several reformulations of a GNEP as

- a quasi-variational inequality (QVI) problem,
- an optimization problem,
- a fixed point problem, and
- a constrained system of equations.

These reformulations are often the basis for both theoretical and algorithmic researches. The first reformulation, which employs the known equivalence between GNEPs with continuously differentiable cost functions and QVIs (see [13, 71, 45]), allows to adapt the existing gap functions for QVIs like those discussed in [31, 63, 67, 127] and Section 8.3 to the setting of GNEPs. This has been done, for example, in [9, 87]. However, this yields a nondifferentiable optimization reformulation of a GNEP except for some special cases. We will consider such a special case in Section 8.3, where the strategy spaces of a player convex GNEP are 'generalized moving sets'.

Another approach to a suitable optimization reformulation of GNEPs is to use the Nikaido-Isoda function introduced in [97] or a regularized version of the Nikaido-Isoda function introduced in [70] and later explored in [78, 38] in order to reformulate the jointly convex GNEP either as a constrained or unconstrained optimization problem. This work was extended to the larger class of player convex GNEPs, see [39]. A major drawback of the corresponding optimization problems, however, is the fact that they typically have nonsmooth objective functions. The aim of Chapter 4 is therefore to have a closer look at the smoothness properties of these objective functions. Preliminary results of this kind, especially regarding the continuity and piecewise smoothness, can already be found in [39]. Subsequently, we show further structural properties in Section 4.2, in particular, our main result indicates that, apart from some degenerate points, the objective functions are differentiable in (local or global) minima. Note that this result is also of some importance for jointly convex GNEPs, although there exist differentiable optimization formulations of this class of problems, cf. [78]. However, the solutions of these differentiable formulations do not characterize the full solution set of jointly convex GNEPs (only so-called *normalized* solutions can be obtained), whereas here we consider reformulations characterizing all solutions of both jointly convex and player convex GNEPs.

Nikaido-Isoda functions additionally supply the basis for reformulations of GNEPs as fixed point problems, see, for example, [86, 133, 79].

Furthermore, we obtain a reformulation of certain GNEPs as a smooth and unconstrained dual optimization problem using conjugate functions and Toland-Singer duality theory [123, 130, 131] in Chapter 5. Another application of conjugate duality for solving GNEPs was considered in [3].

The reformulation of player convex GNEPs (with continuously differentiable cost functions and continuously differentiable inequality constraints defining the players' strategy spaces) as a constrained system of equations is based on the concatenated Karush-Kuhn-Tucker (KKT) conditions of optimization problem (3.2) for all players. More details of these approaches can be found, for example, in [34, 35, 36, 43, 44, 46].

Based on the previously mentioned reformulations of GNEPs, there are many methods to solve these problems such as

- decomposition methods,
- penalty methods,
- methods based on Nikaido-Isoda functions,
- methods based on KKT conditions.

Using decomposition methods like Jacobi-type or Gauss-Seidel-type methods is very natural for solving GNEPs because of their special structure. In the process each player updates his own strategy at each iteration by solving his optimization problem in (3.11) or its regularized version based on simultaneously or sequentially calculated strategies of other players. These methods are easy to implement, but there are only very few convergence results for some special GNEPs, see [110, 57].

The basic idea of penalty methods in [55, 48, 49] is to reduce a GNEP to a standard NEP by penalizing of players' cost functions with an additional term consisting of a penalty parameter and respective constraints. Then such a penalized NEP yields a subproblem that is to be solved at each iteration of a penalty method. A partial penalization, which includes only (difficult) constraints that are dependent on other players' decisions, is also possible, see [53, 64]. Penalty methods are globally convergent if the penalty parameters remain finite during the iteration. In order to guarantee this fact, some conditions on the constraints are necessary. Moreover, subproblems of penalty methods may be difficult to solve.

A method, which can be applied to an unconstrained optimization reformulation of GNEPs based on the regularized Nikaido-Isoda function, is the robust gradient sampling algorithm from [22]. This method was applied in [38] for jointly convex GNEPs and in [39] for player convex GNEPs. There are also other methods based on the regularized Nikaido-Isoda function that provide results for NEPs or jointly convex GNEPs and can only find normalized solutions. Examples for such methods are gradient methods from [78, 37], fixed point methods from [86, 133], a relaxation method of the fixed point iteration in [79] as well as nonsmooth Newton-type minimization or fixed point methods considered in [77, 80, 37].

There are many solution methods for GNEPs in the literature based on KKT conditions and a closely related reformulation of GNEPs as a constrained system of equations, for example, Newton-type methods in [42, 43, 46], Levenberg-Marquardt methods in [59, 61], an LP-Newton method in [44], and a potential reduction method in [34, 36]. The potential reduction algorithm from [36] is a very robust and under suitable assumptions globally convergent interior point method, which is based on minimization of a potential function. In order to obtain also fast local convergence properties (Q-quadratic rate) under suitable assumptions, this algorithm was combined with the LP-Newton method from [44] in a hybrid algorithm (see [35]). As a last point in this overview we refer to [33], where a nonsmooth projection method was obtained in order to calculate Fritz-John points of a GNEP.

3.2. Reformulations as a Constrained or Unconstrained Optimization Problem

This section is mostly based on [39]. Here we review how a GNEP can be equivalently replaced by a (possibly nonsmooth) constrained or unconstrained optimization problem. This is achieved using a gap function for GNEPs. Basically, a function f is called a *gap function* for a mathematical program if the function f is nonnegative and a point is a solution of the corresponding mathematical program if and only the objective function f is zero at this point.

First we define the *fixed point set*

$$W := \{x \in \mathbb{R}^n \mid x \in \Omega(x)\} \tag{3.4}$$

of the set-valued mapping Ω defined in (3.1), which is also called the *feasible set* of the corresponding GNEP. Note that the feasible set has the representation

$$W = \{x \in \mathbb{R}^n \mid \; g^\nu(x) \le 0 \quad \forall \nu = 1, \ldots, N\} \tag{3.5}$$

if the set-valued mappings X_ν, $\nu = 1, \ldots, N$, are defined by (3.3).

Now, we consider the so-called *Nikaido-Isoda function* ([97])

$$\psi(z, x) \; := \; \sum_{\nu=1}^{N} [\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(z^\nu, x^{-\nu})]$$

and the optimal value function

$$V(x) := \sup_{z \in \Omega(x)} \; \psi(z, x), \tag{3.6}$$

with $V(x) = -\infty$ exactly for $x \notin \operatorname{dom} \Omega$, where

$$\operatorname{dom} \Omega := \{x \in \mathbb{R}^n \mid \Omega(x) \neq \emptyset\} \tag{3.7}$$

is the *domain* of the set-valued mapping Ω. Note that the supremum in (3.6) may be a nonuniquely attained maximum or not attained at all on $\operatorname{dom} \Omega$ even in the player convex case, where $\Omega(x)$ is a closed and convex set for any $x \in \mathbb{R}^n$ (as a Cartesian product of closed and convex sets $X_\nu(x^{-\nu})$) but the Nikaido-Isoda function ψ is, in general, just concave in z for each fixed $x \in \mathbb{R}^n$. Nevertheless, given a GNEP (also not player convex), it is easily verified that V is nonnegative for all $x \in \Omega(x)$ and that $\bar{x}$ is a generalized Nash equilibrium of this GNEP if and only if $\bar{x} \in \Omega(\bar{x})$ and $V(\bar{x}) = 0$. Since $x \in \Omega(x)$ holds if and only if $x \in W$, the optimal value function $V : W \to \mathbb{R}_{\geq} \cup \{+\infty\}$ is a gap function for the corresponding GNEP. Therefore, solving a GNEP is equivalent to finding a solution of the constrained minimization problem

$$\min \; V(x) \quad \text{subject to} \quad x \in W \tag{3.8}$$

with zero as the optimal value.

In order to guarantee the existence of unique maximal points in (3.6) on $\operatorname{dom} \Omega$ for player convex GNEPs, we replace the function ψ by the *regularized Nikaido-Isoda function* ([70])

$$\psi_\alpha(z, x) := \sum_{\nu=1}^{N} [\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(z^\nu, x^{-\nu})] - \frac{\alpha}{2} \|x - z\|^2$$

where $\alpha > 0$ denotes a given parameter (and $\psi_0 = \psi$). In view of Assumption 3.1 (a), the function ψ_α is strongly concave in z for each fixed $x \in \mathbb{R}^n$. Hence, for all $x \in \operatorname{dom} \Omega$ there exists a unique solution $z_\alpha(x)$ of the maximization problem

$$\max_z \; \psi_\alpha(z, x) \quad \text{subject to} \quad z \in \Omega(x).$$

Therefore, the optimal value function

$$V_\alpha(x) := \sup_{z \in \Omega(x)} \psi_\alpha(z, x) = \sum_{\nu=1}^{N} \theta_\nu(x^\nu, x^{-\nu}) - \inf_{z \in \Omega(x)} \left(\sum_{\nu=1}^{N} \theta_\nu(z^\nu, x^{-\nu}) + \frac{\alpha}{2}\|x - z\|^2 \right) \tag{3.9}$$

takes the value $-\infty$ exactly for $x \notin \operatorname{dom}\Omega$ and is real-valued for all $x \in \operatorname{dom}\Omega$. Further properties of V_α are given in the following result, the proof of which can be found in [39].

Lemma 3.2 *Under Assumption 3.1, the following statements hold:*

(a) $x \in \Omega(x)$ *if and only if* $x \in W$*; in particular, we have* $W \subseteq \operatorname{dom}\Omega$ *and* V_α *is real-valued on* W.

(b) $V_\alpha(x) \geq 0$ *for all* $x \in W$.

(c) $\bar{x}$ *is a generalized Nash equilibrium if and only if* $\bar{x} \in W$ *and* $V_\alpha(\bar{x}) = 0$.

(d) For all $x \in \operatorname{dom}\Omega$ *there exists a unique vector* $z_\alpha(x)$ *such that*

$$z_\alpha(x) = \operatorname*{argmin}_{z \in \Omega(x)} \left(\sum_{\nu=1}^{N} \theta_\nu(z^\nu, x^{-\nu}) + \frac{\alpha}{2}\|x - z\|^2 \right). \tag{3.10}$$

(e) $\bar{x}$ *is a generalized Nash equilibrium if and only if* $\bar{x} = z_\alpha(\bar{x})$ *holds, that is,* $\bar{x}$ *is a fixed point of the mapping* $x \mapsto z_\alpha(x)$.

It follows from Lemma 3.2 (a)–(c) that the optimal value function $V_\alpha : W \to [0, +\infty[$ is a gap function for player convex GNEPs. Hence solving a player convex GNEP is equivalent to finding a solution of the constrained minimization problem

$$P: \qquad \min\ V_\alpha(x) \quad \text{subject to} \quad x \in W \tag{3.11}$$

or, alternatively, using the indicator function δ_W of W defined in (2.1), to solving the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^n}\ [V_\alpha(x) + \delta_W(x)] \tag{3.12}$$

with zero optimal value in both reformulations and with the convention $\eta + \infty = +\infty$ for all $\eta \in \mathbb{R} \cup \{\pm\infty\}$ in the unconstrained reformulation. This convention makes sense, since the objective function from (3.12) should take the function value $+\infty$ on the complement W^c of the set W and, in particular, on $(\operatorname{dom}\Omega)^c \subseteq W^c$. We will study the reformulation (3.12) in Chapter 5 for a class of the player convex GNEPs and consider a dual gap function for these GNEPs applying the duality theory by Toland and Singer [123, 130, 131]. Therefore, we call the optimal value functions V and V_α also the *primal* gap functions for the player convex GNEPs.

Theoretical and numerical results on the solutions of the optimization problems (3.8), (3.11), and (3.12) obviously depend on the structure of the objective functions V and V_α as well as of the feasible set W. While not much can be said about the structure of the set W, basic continuity and differentiability properties of V_α on W for player convex GNEPs were studied in [39]. Chapter 4 will complement these properties with additional useful characteristics of V_α at its points of nondifferentiability. The continuity and differentiability properties of V_α are applicable to the optimal value function V if the maximum in (3.6) is uniquely attained on $\operatorname{dom}\Omega$ without adding the regularization term with the parameter α to the function ψ. In such cases, it is simpler to calculate the function V than the function V_α and therefore to allow the choice $\alpha = 0$, that is, the employment of the original Nikaido-Isoda function. Hence we will illustrate most results of Chapter 4 on examples of the GNEPs where the maximization problem in (3.6) is uniquely solvable on $\operatorname{dom}\Omega$. Note that if the maximum in (3.6) is not uniquely attained, the function V may have worse continuity and differentiability properties than the function V_α.

4. Smoothness Properties of a Primal Gap Function for Generalized Nash Equilibrium Problems

Recall that the primal gap function V_α lays the foundation for the reformulation (3.11) of a player convex GNEP. The aim of Chapter 4 is to study differentiability properties of this primal gap function. First we review from [39] a result on the continuity of the primal gap function V_α in Section 4.1 and relate the continuity points of V_α to interior points of the domain of V_α. Subsequently, differentiability properties of V_α are studied in Section 4.2. Our main result is stated in Theorem 4.23 in Section 4.2 and treats differentiability of the function V_α at local minimizers of the reformulation in (3.11). The results of this chapter were published in [74].

4.1. Continuity Properties and Domain

First we recall in Section 4.1 sufficient conditions for the continuity of the primal gap function V_α for player convex GNEPs from [39] and then we analyze at which points of the feasible set W we can study differentiability properties of V_α.

Throughout Chapter 4, we assume that the strategy spaces X_ν, $\nu = 1, \ldots, N$, are given by the representation (3.3). Therefore, we adjust Assumption 3.1 to the following initial assumption of this chapter.

Assumption 4.1 *(a) The cost functions θ_ν, $\nu = 1, \ldots, N$, are continuous on $\mathbb{R}^n$ and convex in x^ν for each fixed $x^{-\nu} \in \mathbb{R}^{n-n_\nu}$.*

(b) The strategy spaces $X_\nu(x^{-\nu})$ are defined by

$$X_\nu(x^{-\nu}) \ := \ \{x^\nu \in \mathbb{R}^{n_\nu} \mid g_i^\nu(x^\nu, x^{-\nu}) \leq 0 \quad \forall i = 1, \ldots, m_\nu\}$$

with functions $g_i^\nu : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m_\nu$, $\nu = 1, \ldots, N$, that are continuous on $\mathbb{R}^n$ and convex in x^ν for each fixed $x^{-\nu} \in \mathbb{R}^{n-n_\nu}$.

To study the structural properties of the reformulation (3.11) for player convex GNEPs, we rewrite the primal gap function V_α for any $x \in W$ as

$$V_\alpha(x) \ = \ \sup_{z \in \Omega(x)} \psi_\alpha(z, x) \ = \ \sum_{\nu=1}^{N} \left[\theta_\nu(x) - \varphi_\alpha^\nu(x)\right] \tag{4.1}$$

with the optimal value functions

$$\varphi_\alpha^\nu(x) \ := \inf_{z^\nu \in X_\nu(x^{-\nu})} \left[\theta_\nu(z^\nu, x^{-\nu}) + \frac{\alpha}{2} \|x^\nu - z^\nu\|^2 \right] \tag{4.2}$$

of the strongly convex and therefore uniquely solvable problems

$$Q_\alpha^\nu(x): \qquad \min_{z^\nu} \theta_\nu(z^\nu, x^{-\nu}) + \frac{\alpha}{2}\|x^\nu - z^\nu\|^2 \quad \text{subject to} \quad g^\nu(z^\nu, x^{-\nu}) \leq 0$$

for $\nu = 1, \dots, N$. Note that the choice $\alpha = 0$ is allowed only if the problems Q_0^ν are uniquely solvable on $\operatorname{dom} \Omega$. As mentioned at the end of Section 3.2, in such cases the continuity and differentiability properties of V_α are applicable to the optimal value function V, where $V := V_0$ is easier to calculate than the function V_α.

Let $S_\alpha^\nu(x) = \{z_\alpha^\nu(x)\}$ denote the (singleton) set of optimal points of the optimization problems $Q_\alpha^\nu(x)$ for $x \in W$. Hence we can rewrite the optimal value functions φ_α^ν as

$$\varphi_\alpha^\nu(x) \ = \ \theta_\nu(z_\alpha^\nu(x), x^{-\nu}) + \frac{\alpha}{2}\|x^\nu - z_\alpha^\nu(x)\|^2.$$

It is easy to see that $(z_\alpha^1(x), \dots, z_\alpha^N(x))$ coincides with the unique maximizer $z_\alpha(x)$ of $\psi_\alpha(\cdot, x)$ on $\Omega(x)$. Clearly, the structural properties of V_α heavily depend on the structural properties of the functions φ_α^ν.

Now, we briefly recall sufficient conditions for the continuity of the primal gap function V_α on the feasible set W from [39, Lemma 3.4].

Lemma 4.2 *Let Assumption 4.1 hold. Additionally, let $X_\nu(\bar{x}^{-\nu})$ satisfy the Slater condition for $\bar{x} \in W$ and for all $\nu \in \{1, \dots, N\}$, that is, there exists some $\bar{z}^\nu \in \mathbb{R}^{n_\nu}$ with $g^\nu(\bar{z}^\nu, \bar{x}^{-\nu}) < 0$ for all $\nu \in \{1, \dots, N\}$. Then the functions z_α^ν, φ_α^ν, $\nu = 1, \dots, N$, and V_α are continuous at $\bar{x}$.*

Lemma 4.2 guarantees continuity of V_α on $W \setminus D_1$ with the 'degenerate point set'

$$D_1 \ := \ \{x \in W \mid \text{for some } \nu \in \{1, \dots, N\} \text{ the set } X_\nu(x^{-\nu}) \text{ violates the Slater condition}\}.$$

As explained in [39] and illustrated in Example 4.3 below, one has to expect that the set D_1 is nonempty. This was the motivation to develop a weaker sufficient condition for continuity of V_α on W relative to W (see Definition 2.2), for which the interested reader is referred to the paper [39, Theorem 3.5].

The following example illustrates the continuity properties of V_α for a GNEP and will also serve to illustrate differentiability properties of V_α below.

Example 4.3 Consider a player convex GNEP with $N = 2$, $n_1 = n_2 = 1$, the variables x_1 and x_2 controlled by player 1 and 2, respectively, the cost functions $\theta_1(x) := x_1$ and $\theta_2(x) := x_2$, the constraint $g_1^1(x) := -2x_1 + x_2 \leq 0$ for the first player and the constraints $g_1^2(x) = x_1^2 + x_2^2 - 1 \leq 0$ and $g_2^2(x) = -x_1 - x_2 \leq 0$ for the second player. Then it is easy to see that the problems $Q_\alpha^1(x)$ and $Q_\alpha^2(x)$ with the choice $\alpha = 0$ are uniquely solvable for all $x \in W$ with the set

$$W \ = \ \left\{ x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1, \ -x_1 \leq x_2 \leq 2x_1 \right\},$$

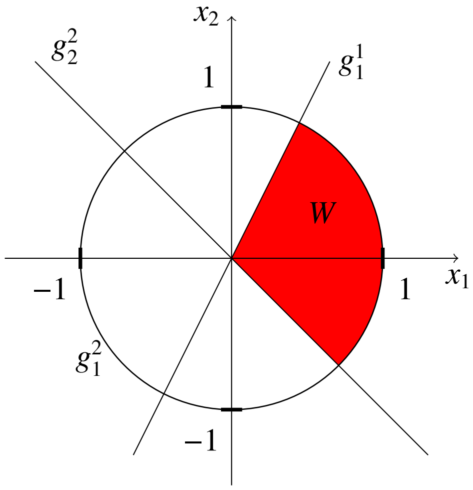


Figure 4.1.: Illustration of the set W from Example 4.3

which is illustrated in Figure 4.1.

In fact, for $x \in W$ we obtain the strategy spaces

$$\begin{aligned} X_1(x_2) &= \left[\frac{x_2}{2}, +\infty\right[, \\ X_2(x_1) &= \left[\max\left\{-x_1, -\sqrt{1-x_1^2}\right\}, \sqrt{1-x_1^2}\right], \end{aligned}$$

so that the optimal points, which are equal to the optimal values of $Q_0^1(x)$ and $Q_0^2(x)$, are

$$\begin{aligned} z_0^1(x) &= \varphi_0^1(x) &= \frac{x_2}{2}, \\ z_0^2(x) &= \varphi_0^2(x) &= \max\left\{-x_1, -\sqrt{1-x_1^2}\right\}. \end{aligned}$$

Due to (3.6) and (4.1), this results in

$$V(x) := V_0(x) = \theta_1(x) + \theta_2(x) - \varphi_0^1(x) - \varphi_0^2(x) = x_1 + x_2 + \min\left\{x_1, \sqrt{1-x_1^2}\right\} - \frac{x_2}{2}$$

for all $x \in W$. Note that, in spite of player convexity, V is a concave function as the minimum of two smooth concave functions (compare also Figure 4.2 and Remark 4.12 below).

For all $x \in W$ the set $X_1(x_2)$ obviously satisfies the Slater condition. However, the set $X_2(x_1)$ satisfies the Slater condition only for values $x_1 \neq 1$, whereas $X_2(1) = \{0\}$ is a singleton. This results in $D_1 = \{(1, 0)\}$ and shows that D_1 can easily be nonempty. Lemma 4.2 then yields continuity of V only on $W \setminus D_1$. Furthermore, direct inspection shows that V is continuous even on all of W relative to W where, however, V has 'infinite slope' at the point $x = (1, 0)$. We remark that the improvement of Lemma 4.2 by [39, Theorem 3.5] also yields continuity of V at $(1, 0)$ relative to W for the present example.

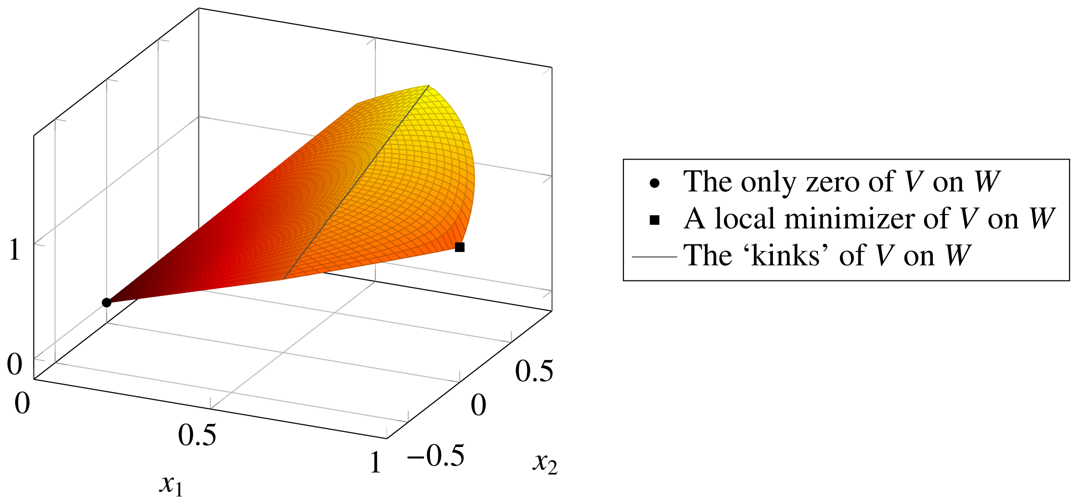


Figure 4.2.: Function V on W from Example 4.3

It is not hard to see that the function V has a unique global minimizer on W at $\bar{x} := (0, 0)$. The function value at $\bar{x}$ is zero, so that $\bar{x}$ is the unique generalized Nash equilibrium in view of Lemma 3.2. However, it can be also shown that V has a local minimizer on W at $(1, 0)$ with value one. We point out that the point $(1, 0)$ is an element of D_1, so that D_1 is not only nonempty but also contains a 'structurally relevant' point in the present example. ◇

Section 4.2 will show that the set D_1 also plays a crucial role for differentiability properties of V_α. In fact, we shall study differentiability of V_α at points in the topological interior of the domain of V_α, where we have seen in Section 3.2 that the domain

$$\operatorname{dom} V_\alpha \;=\; \{x \in \mathbb{R}^n \mid V(x) \in \mathbb{R}\}$$

of V_α coincides with the domain of the set-valued mapping $\Omega(x)$ defined in (3.1) and (3.7), respectively. Hence their topological interiors satisfy

$$\operatorname{int}\operatorname{dom} V_\alpha \;=\; \operatorname{int}\operatorname{dom}\Omega. \tag{4.3}$$

The following example shows that we cannot expect the inclusion $W \subseteq \operatorname{int}\operatorname{dom} V_\alpha$ to hold despite the fact that we have $W \subseteq \operatorname{dom}\Omega = \operatorname{dom} V_\alpha$.

Example 4.4 Example 4.3 yields $\operatorname{int}\operatorname{dom} V = \operatorname{int}\operatorname{dom}\Omega =]-1/\sqrt{2}, 1[\times \mathbb{R}$, and therefore

$$W \setminus \{(1, 0)\} \;=\; W \cap \operatorname{int}\operatorname{dom} V.$$

Together with $D_1 = \{(1, 0)\}$ we arrive at $W \setminus D_1 = W \cap \operatorname{int}\operatorname{dom} V$. ◇

The next result guarantees that the inclusion $W \setminus D_1 \subseteq W \cap \operatorname{int}\operatorname{dom} V_\alpha$ is also true in general.

Lemma 4.5 *Let Assumption 4.1 hold. Then we have*

$$W \setminus D_1 \;\subseteq\; W \cap \operatorname{int}\operatorname{dom} V_\alpha. \tag{4.4}$$

Proof. In view of (4.3), the assertion is shown if we can prove the relation

$$W \setminus D_1 \subseteq W \cap \operatorname{int} \operatorname{dom} \Omega.$$

Choose $\bar{x} \in W \setminus D_1$. Then we have $\bar{x} \in W$, and for all $\nu = 1, \dots, N$ there exists some $\bar{z}^\nu \in \mathbb{R}^{n_\nu}$ with $g^\nu(\bar{z}^\nu, \bar{x}^{-\nu}) < 0$. Due to the continuity of the functions g^ν, we can choose a neighborhood U of $\bar{x}$ such that for all $x \in U$ and $\nu = 1, \dots, N$ also $g^\nu(\bar{z}^\nu, x^{-\nu}) < 0$ is satisfied. In particular, for all $x \in U$ each set $X_\nu(x^{-\nu})$, $\nu = 1, \dots, N$, is nonempty, and U is therefore contained in $\operatorname{dom} \Omega$. This shows the assertion. □

Lemma 4.5 will allow us to study differentiability properties of V_α on the set $W \setminus D_1$ in Section 4.2. Clearly, Lemma 4.5 does not exclude that also some elements of D_1 are contained in $\operatorname{int} \operatorname{dom} V_\alpha$. However, under mild additional assumptions, we conjecture that actually equality holds in (4.4). This fact will be shown in the sequel under certain conditions, including Assumption 4.6 below, though we believe that this assumption can be relaxed.

Assumption 4.6 *All functions g_i^ν, $i = 1, \dots, m_\nu$, $\nu = 1, \dots, N$, are convex.*

Obviously, under Assumption 4.6 of joint constraint convexity the set W is convex. Note that we will use this assumption only in Theorem 4.11 below. Furthermore, we will assume in the remainder of Chapter 4 that all defining functions of GNEPs are at least continuously differentiable.

Assumption 4.7 *The functions θ_ν and g_i^ν, $i = 1, \dots, m_\nu$, are continuously differentiable for each $\nu \in \{1, \dots, N\}$.*

Furthermore, for each $\nu \in \{1, \dots, N\}$ we denote by

$$I^\nu = \{1, \dots, m_\nu\}$$

the index set of inequality constraints of player ν. Additionally, we put

$$W_\nu := \{x \in \mathbb{R}^n \mid g_i^\nu(x) \le 0 \quad \forall i \in I^\nu\}, \tag{4.5}$$

and we define the active index set

$$I_0^\nu(x) := \{i \in I^\nu \mid g_i^\nu(x) = 0\} \tag{4.6}$$

for $x \in W_\nu$. Note that W_ν coincides with the graph $\operatorname{gph} X_\nu$ of the set-valued mapping X_ν, and that we obviously have $W = \bigcap_{\nu=1}^N W_\nu$.

In the following we distinguish between three Mangasarian Fromovitz constraint qualifications.

Definition 4.8 *(a) Let $\nu \in \{1, \dots, N\}$. The* Mangasarian Fromovitz constraint qualification (MFCQ$_\nu$) *holds at $x \in W_\nu$ if there exists some vector $d \in \mathbb{R}^n$ (typically depending on the index ν) with*

$$Dg_i^\nu(x)\, d < 0 \quad \forall i \in I_0^\nu(x).$$

(b) *The* Mangasarian Fromovitz constraint qualification (MFCQ) *holds at $x \in W$ if there exists some vector $d \in \mathbb{R}^n$ (independent of the index ν) with*

$$Dg_i^\nu(x)\, d \;<\; 0 \quad \forall i \in I_0^\nu(x),\ \nu = 1, \ldots, N.$$

(c) *For player $\nu \in \{1, \ldots, N\}$ and a point $x \in W_\nu$ the* player Mangasarian Fromovitz constraint qualification (player MFCQ) *holds at $x^\nu \in X_\nu(x^{-\nu})$ if there exists some vector $d^\nu \in \mathbb{R}^{n_\nu}$ with*

$$D_{x^\nu} g_i^\nu(x^\nu, x^{-\nu})\, d^\nu \;<\; 0 \quad \forall i \in I_0^\nu(x^\nu, x^{-\nu}),$$

where the active index set $I_0^\nu(x^\nu, x^{-\nu})$ of x^ν in $X_\nu(x^{-\nu})$ coincides with the active index set of $(x^\nu, x^{-\nu})$ in W_ν defined in (4.6).

Note that, under Assumption 4.6, MFCQ$_\nu$ is satisfied at one point $x \in W_\nu$ if and only if the Slater condition holds for this set W_ν, and therefore MFCQ$_\nu$ is satisfied at one point $x \in W_\nu$ if and only if it is valid at all points $x \in W_\nu$. We need only Definition 4.8 (a) for the next assumption about joint MFCQ$_\nu$ used in Theorem 4.11.

Assumption 4.9 *MFCQ$_\nu$ holds everywhere in W_ν for each $\nu \in \{1, \ldots, N\}$.*

Under Assumption 4.6 and using the comment after Definition 4.8, Assumption 4.9 is equivalent to the Slater condition for each W_ν, $\nu = 1, \ldots, N$.

Remark 4.10 We stress that Assumption 4.9 is unrelated to the assumption of MFCQ everywhere in the set W. In fact, assuming MFCQ at points in W does not allow conclusions about points in $W_\nu \setminus W$ for any ν. In particular, MFCQ$_\nu$ may be violated at some $\bar{x} \in W_\nu \setminus W$, so that Assumption 4.9 does not hold. On the other hand, consider a two player game with variables $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$ controlled by player 1 and 2, respectively, and the constraint functions $g^1(x) = (x_1 - 1)^2 + x_2^2 - 1$ and $g^2(x) = (x_1 + 1)^2 + x_2^2 - 1$. Then Assumption 4.9 holds, but MFCQ is violated in the set $W = \{(0, 0)\}$. ◇

Now, we summarize the sufficient assumptions for the validity of the equality in (4.4).

Theorem 4.11 *Let Assumptions 4.1, 4.6, 4.7, and 4.9 hold. Then we have*

$$W \setminus D_1 \;=\; W \cap \operatorname{int} \operatorname{dom} V.$$

Proof. In view of Lemma 4.5 and (4.3), the assertion is shown if we can prove the relation

$$W \cap\ \operatorname{int} \operatorname{dom} \Omega \;\subseteq\; W \setminus D_1.$$

Let $\bar{x} \in D_1$. We show that $\bar{x}$ lies in the set complement $(\operatorname{int} \operatorname{dom} \Omega)^c$. Using the relation $D_1 \subseteq W \subseteq \operatorname{dom} \Omega$, it suffices to guarantee that any neighborhood of $\bar{x}$ contains points from the set complement $(\operatorname{dom} \Omega)^c$.

Choose some $\nu \in \{1, \dots, N\}$ such that $X_\nu(\bar{x}^{-\nu})$ violates the Slater condition. Then player MFCQ is violated at any element of $X_\nu(\bar{x}^{-\nu})$ and, in particular, at $\bar{x}^\nu$. By Gordan's theorem (see, e.g., [120]), there exist multipliers $\gamma_i \geq 0$ ($i \in I_0^\nu(\bar{x})$) with $\sum_{i \in I_0^\nu(\bar{x})} \gamma_i = 1$ such that

$$d^\nu \;:=\; \sum_{i \in I_0^\nu(\bar{x})} \gamma_i \nabla_{x^\nu} g_i^\nu(\bar{x}) \;=\; 0. \tag{4.7}$$

We use the same multipliers to define

$$d^{-\nu} \;:=\; \sum_{i \in I_0^\nu(\bar{x})} \gamma_i \nabla_{x^{-\nu}} g_i^\nu(\bar{x})$$

as well as $d := (d^\nu, d^{-\nu})$. We claim that d is nonzero. In fact, if we had $d = 0$, we would obtain $\sum_{i \in I_0^\nu(\bar{x})} \gamma_i \nabla g_i^\nu(\bar{x}) = 0$. Hence, noting that $\bar{x} \in W_\nu$, it would follow from Assumption 4.9 and the fact that MFCQ$_\nu$ is equivalent to the positive linear independence of the corresponding vectors that $\gamma_i = 0$ holds for all $i \in I_0^\nu(\bar{x})$. This is a contradiction to $\sum_{i \in I_0^\nu(\bar{x})} \gamma_i = 1$. Consequently, $d \neq 0$ holds and, in view of (4.7), we then also know that $d^{-\nu}$ cannot vanish.

We now define the ray

$$x^{-\nu}(t) \;:=\; \bar{x}^{-\nu} + t d^{-\nu}$$

and show that for all $t > 0$ the set $X_\nu(x^{-\nu}(t))$ is empty. To this end, we note that Assumption 4.6 implies

$$0 \;\geq\; Dg_i^\nu(\bar{x})(x - \bar{x}) \quad \forall i \in I_0^\nu(\bar{x}),\; x \in W_\nu.$$

Taking the convex combination of the latter inequalities with the above coefficients γ_i and using $d^\nu = 0$ yields that all $x \in W_\nu$ also satisfy

$$0 \;\geq\; d^\intercal(x - \bar{x}) \;=\; (d^{-\nu})^\intercal(x^{-\nu} - \bar{x}^{-\nu}). \tag{4.8}$$

The relation in (4.8) holds for all $x^{-\nu} \in \operatorname{dom} X_\nu$, since for each $x^{-\nu} \in \operatorname{dom} X_\nu$ there exists some $x^\nu \in \mathbb{R}^{n_\nu}$ with $x \in \operatorname{gph} X_\nu = W_\nu$. As (4.8) does not depend on x^ν, this means

$$\operatorname{dom} X_\nu \;\subseteq\; \{x^{-\nu} \in \mathbb{R}^{n-n_\nu} |\, (d^{-\nu})^\intercal(x^{-\nu} - \bar{x}^{-\nu}) \leq 0\}.$$

On the other hand, for any $t > 0$ the point $x^{-\nu}(t)$ satisfies

$$(d^{-\nu})^\intercal(x^{-\nu}(t) - \bar{x}^{-\nu}) \;=\; t\|d^{-\nu}\|^2 \;>\; 0,$$

and therefore $x^{-\nu}(t) \notin \operatorname{dom} X_\nu$, that is, $X_\nu(x^{-\nu}(t)) = \emptyset$. This shows the assertion. □

Note that the assumptions of Theorem 4.11 are satisfied in Example 4.3.

As a last point in this section, we consider a case in which the optimal value function V_α turns out to be concave.

Remark 4.12 Under Assumption 4.6 and the additional assumption of (affine) linear functions θ_ν, $\nu = 1, \dots, N$, one can show along the lines of the proof of [124, Prop. 3.1.26] that the functions φ_α^ν, $\nu = 1, \dots, N$, are convex on W, and that V_α is concave on W. Example 4.3 illustrates this situation. ◇

4.2. Differentiability Properties

The next step is to study differentiability properties of V_α on $W \setminus D_1$, as motivated by Lemma 4.5 and Theorem 4.11. Assumptions 4.1 and 4.7 are blanket assumptions in this section.

For the following theorem, recall that $S^\nu_\alpha(x)$ denotes the set of optimal points of $Q^\nu_\alpha(x)$, let

$$L^\nu_\alpha(x, z^\nu, \lambda^\nu) \;:=\; \theta_\nu(z^\nu, x^{-\nu}) + \frac{\alpha}{2}\|x^\nu - z^\nu\|^2 + (\lambda^\nu)^\top g^\nu(z^\nu, x^{-\nu}) \tag{4.9}$$

denote the Lagrange function of $Q^\nu_\alpha(x)$, and let

$$KKT^\nu_\alpha(x) \;:=\; \{\lambda^\nu \in \mathbb{R}^{m_\nu} \mid \nabla_{z^\nu} L^\nu_\alpha(x, z^\nu, \lambda^\nu) = 0,\; \lambda^\nu \geq 0,\; (\lambda^\nu)^\top g^\nu(z^\nu, x^{-\nu}) = 0\}$$

be the set of Karush-Kuhn-Tucker multipliers for $z^\nu \in S^\nu_\alpha(x)$. Note that the convex polyhedron $KKT^\nu_\alpha(x)$ does not depend on z^ν as $Q^\nu_\alpha(x)$ is a convex problem (this statement can be easily shown using the well-known Saddle Point Theorem for convex optimization problems, see [69]), and that $KKT^\nu_\alpha(x)$ is a nonempty convex polytope if and only if $X_\nu(x^{-\nu})$ satisfies the Slater condition [65].

In this section we will use besides the standard concept of differentiability also several differentiability concepts (directional differentiability, Hadamard directional differentiability, and Gâteaux differentiability) summarized in Definition 2.8.

Theorem 4.13 *Let Assumptions 4.1 and 4.7 hold, and let $x \in W \setminus D_1$. Then V_α is Hadamard directionally differentiable at x with*

$$V'_\alpha(x, d) \;=\; \sum_{\nu=1}^{N} \left[D\theta_\nu(x)\, d - \max_{\lambda^\nu \in KKT^\nu_\alpha(x)} D_x L^\nu_\alpha(x, z^\nu_\alpha(x), \lambda^\nu)\, d \right]$$

for any direction $d \in \mathbb{R}^n$.

Proof. In view of (4.1), we have

$$V_\alpha(x) = \sum_{\nu=1}^{N} [\theta_\nu(x) - \varphi^\nu_\alpha(x)]$$

with the differentiable functions θ_ν and possibly nondifferentiable functions φ^ν_α from (4.2). Since $x \in W \setminus D_1$, the sets $X_\nu(x^{-\nu})$ satisfy the Slater condition for all $\nu = 1, \dots, N$, hence, by a standard result (see, e.g., [69, 83, 116]), the functions φ^ν_α are Hadamard directionally differentiable, and their directional derivatives are given by

$$(\varphi^\nu_\alpha)'(x, d) \;=\; \min_{z^\nu \in S^\nu_\alpha(x)} \max_{\lambda^\nu \in KKT^\nu_\alpha(x)} D_x L^\nu_\alpha(x, z^\nu, \lambda^\nu)\, d \tag{4.10}$$

for any directions $d \in \mathbb{R}^n$. Taking into account that $S^\nu_\alpha(x) = \{z^\nu_\alpha(x)\}$ is actually a singleton in our case, the desired statement follows. □

Obviously, the formula for the directional derivative of φ^ν_α from (4.10) simplifies further if not only $S^\nu_\alpha(x)$ but also $KKT^\nu_\alpha(x)$ is a singleton.

Proposition 4.14 *Let Assumptions 4.1 and 4.7 hold. Additionally, for $\nu \in \{1,\dots,N\}$ and $x \in W_\nu$, let $X_\nu(x^{-\nu})$ satisfy the Slater condition, and let $KKT_\alpha^\nu(x)$ be the singleton $\{\lambda_\alpha^\nu(x)\}$. Then φ_α^ν is Gâteaux differentiable at x with*

$$(\varphi_\alpha^\nu)'(x,d) \;=\; D_x L_\alpha^\nu(x, z_\alpha^\nu(x), \lambda_\alpha^\nu(x))\, d$$

for any $d \in \mathbb{R}^n$.

The previous result motivates to define a second 'degenerate point set',

$$D_2 \;:=\; \{x \in W \mid \text{for some } \nu = 1,\dots,N \text{ the set } KKT_\alpha^\nu(x) \text{ is not a singleton}\}.$$

Note that the formulation '$KKT_\alpha^\nu(x)$ is not a singleton' allows not only the case where $KKT_\alpha^\nu(x)$ contains more than one element but also the case $KKT_\alpha^\nu(x) = \emptyset$. Before we characterize the points $x \in D_2^c$, we define three linear independence constraint qualifications and a strict Mangasarian Fromovitz condition for GNEPs.

Definition 4.15 *(a) Let $\nu \in \{1,\dots,N\}$. Then the* linear independence constraint qualification (LICQ$_\nu$) *holds at $x \in W_\nu$ if the gradients*

$$\nabla g_i^\nu(x) \quad (i \in I_0^\nu(x))$$

are linearly independent.

(b) The linear independence constraint qualification (LICQ) *holds at $x \in W$ if the gradients*

$$\nabla g_i^\nu(x) \quad (i \in I_0^\nu(x),\ \nu = 1,\dots,N)$$

are linearly independent.

(c) For player $\nu \in \{1,\dots,N\}$ and a point $x \in W_\nu$ the player linear independence constraint qualification (player LICQ) *holds at $x^\nu \in X_\nu(x^{-\nu})$ if the gradients*

$$\nabla_{x^\nu} g_i^\nu(x^\nu, x^{-\nu}) \quad (i \in I_0^\nu(x^\nu, x^{-\nu}))$$

are linearly independent.

(d) For player $\nu \in \{1,\dots,N\}$ let $x \in W_\nu$ with $KKT_\alpha^\nu(x) \neq \emptyset$ for the optimal point $z_\alpha^\nu(x)$. Then the player strict Mangasarian Fromovitz condition (player SMFC) *holds at $z_\alpha^\nu(x) \in X_\nu(x^{-\nu})$ with a multiplier $\lambda^\nu \in KKT_\alpha^\nu(x)$ if the gradients*

$$\nabla_{x^\nu} g_i^\nu(z_\alpha^\nu(x), x^{-\nu}) \quad (i \in I_{0+}^\nu(z_\alpha^\nu(x), x^{-\nu}))$$

are linearly independent, and there exists some vector $d^\nu \in \mathbb{R}^{n_\nu}$ with

$$\begin{aligned} D_{x^\nu} g_i^\nu(z_\alpha^\nu(x), x^{-\nu})\, d^\nu &= 0 \quad \forall i \in I_{0+}^\nu(z_\alpha^\nu(x), x^{-\nu}),\\ D_{x^\nu} g_i^\nu(z_\alpha^\nu(x), x^{-\nu})\, d^\nu &< 0 \quad \forall i \in I_{00}^\nu(z_\alpha^\nu(x), x^{-\nu}), \end{aligned}$$

where

$$\begin{aligned} I_{0+}^\nu(x) &= \{i \in I_0^\nu(x) \mid \lambda_i^\nu > 0\},\\ I_{00}^\nu(x) &= \{i \in I_0^\nu(x) \mid \lambda_i^\nu = 0\}. \end{aligned}$$

We point out that player SMFC is unentitled to be called a constraint qualification as sometimes done in the literature, since this condition relies on the existence of Lagrange multipliers and depends (indirectly) on the objective function φ_α^ν, see [134].

A sufficient condition for $x \in W$ to lie in D_2^c is that player LICQ holds at $z_\alpha^\nu(x) \in X_\nu(x^{-\nu})$ for all $\nu = 1, \ldots, N$. In fact, it is well known that player LICQ at the optimal point $z_\alpha^\nu(x)$ entails a unique KKT multiplier $\lambda_\alpha^\nu(x)$ corresponding to $z_\alpha^\nu(x)$. Note that if the set $I_0^\nu(z_\alpha^\nu(x), x^{-\nu})$ is empty for $x \in W_\nu$ and some $\nu \in \{1, \ldots, N\}$, then player LICQ holds at $z_\alpha^\nu(x)$, and therefore the set $KKT_\alpha^\nu(x)$ is a singleton.

Furthermore, by a result from [88], a characterization for $x \in D_2^c \cap W$ is given by the fact that $KKT_\alpha^\nu(x) \neq \emptyset$ and the player SMFC holds at $z_\alpha^\nu(x) \in X_\nu(x^{-\nu})$ with a multiplier $\lambda^\nu \in KKT_\alpha^\nu(x)$ for all $\nu = 1, \ldots, N$. This yields

$$\begin{aligned} D_2 \;=\; \{x \in W \mid & \text{ either } KKT_\alpha^\nu(x) = \emptyset \text{ or player SMFC is violated} \\ & \text{at } z_\alpha^\nu(x) \in X_\nu(x^{-\nu}) \text{ for some } \nu \in \{1, \ldots, N\}\} \end{aligned} \tag{4.11}$$

and allows us to prove the following relation between D_1 and D_2.

Lemma 4.16 *Under Assumptions 4.1 and 4.7, the degenerate point sets satisfy $D_1 \subseteq D_2$.*

Proof. Choose any point $x \in D_2^c$. The assertion is trivial if $x \in W^c$. Otherwise, in view of (4.11), $KKT_\alpha^\nu(x) \neq \emptyset$ and player SMFC holds at $z^\nu(x)$ in $X_\nu(x^{-\nu})$ for all $\nu = 1, \ldots, N$. Since player SMFC implies the ordinary player MFCQ at the latter point, $X_\nu(x^{-\nu})$ also satisfies the Slater condition for all $\nu = 1, \ldots, N$. This means that x lies in D_1^c and shows the assertion. □

In view of Lemma 4.16, the Slater condition may be dropped in Proposition 4.14, and we arrive at the following theorem.

Theorem 4.17 *Let Assumptions 4.1 and 4.7 hold, and let $x \in W \setminus D_2$ and $\lambda_\alpha^\nu(x)$, $\nu = 1, \ldots, N$, denote the unique KKT multipliers. Then V_α is Gâteaux differentiable at x with*

$$V_\alpha'(x, d) \;=\; \left(\sum_{\nu=1}^{N} \left[D\theta_\nu(x) - D_x L_\alpha^\nu(x, z_\alpha^\nu(x), \lambda_\alpha^\nu(x)) \right] \right) d$$

for any $d \in \mathbb{R}^n$.

Clearly, player LICQ, or even player SMFC, cannot be expected to hold at $z_\alpha^\nu(x)$ in $X_\nu(x^{-\nu})$ for all $x \in W$, even if LICQ is fulfilled at each $x \in W$. To begin with, the point $(z_\alpha^\nu(x), x^{-\nu}) \in W_\nu$ does not even have to belong to W. However, violation of player SMFC is in some sense exceptional, as the following example illustrates.

Example 4.18 Again we go back to Example 4.3. In order to prevent the confusion between indices of elements and a power of a term, we use the notations $z_1 := z^1$, $z_2 := z^2$, $z_1(x) := z_0^1(x)$ as well as $z_2(x) := z_0^2(x)$, since each player controls only a single variable, and we denote $\lambda_1 := \lambda^1$, $\lambda_1(x) := \lambda_0^1(x)$, $(\lambda_2, \lambda_3) := \lambda^2$ and $(\lambda_2(x), \lambda_3(x)) := \lambda_0^2(x)$, since there are three constraints (one for the first player and two for the second). Recall that in Example 4.3 the set $X_2(x_1)$ violates the

Slater condition for $x \in D_1 = \{(1,0)\}$ and, thus, player SMFC is violated at this single element. In the following, we check for points in $D_2 \setminus D_1$.

The Lagrangian of player 1 is

$$L_0^1(x, z_1, \lambda_1) = z_1 + \lambda_1(-2z_1 + x_2).$$

The optimal point $z_1(x) = x_2/2$ has the active index set $I_0^1(z_1(x), x_2) = \{1\}$, and we obtain the multiplier set

$$KKT_0^1(x) = \{\lambda_1 \in \mathbb{R} \mid 1 - 2\lambda_1 = 0,\ \lambda_1 \geq 0\} = \{1/2\}.$$

In particular, the multiplier $\lambda_1(x) = 1/2$ is unique for any $x \in W$.

For player 2, the Lagrangian is

$$L_0^2(x, z_2, (\lambda_2, \lambda_3)) = z_2 + \lambda_2(x_1^2 + z_2^2 - 1) + \lambda_3(-x_1 - z_2).$$

For $x \in W \setminus D_1$ with $x_1 > 1/\sqrt{2}$, the optimal point is $z_2(x) = -\sqrt{1 - x_1^2}$ with active index set $I_0^2(x_1, z_2(x)) = \{1\}$, and we obtain the multiplier set

$$\begin{aligned} KKT_0^2(x) &= \left\{(\lambda_2, \lambda_3) \in \mathbb{R}^2 \;\middle|\; 1 + 2\lambda_2 z_2(x) - \lambda_3 = 0,\ \lambda_2 \geq 0,\ \lambda_3 = 0\right\} \\ &= \left\{\left(\left(2\sqrt{1 - x_1^2}\right)^{-1}, 0\right)\right\}. \end{aligned}$$

For vectors $x \in W$ with $x_1 < 1/\sqrt{2}$, the optimal point is $z_2(x) = -x_1$ with active index set $I_0^2(x_1, z_2(x)) = \{2\}$, and we obtain the multiplier set

$$KKT_0^2(x) = \{(\lambda_2, \lambda_3) \in \mathbb{R}^2 \mid 1 + 2\lambda_2 z_2(x) - \lambda_3 = 0,\ \lambda_2 = 0,\ \lambda_3 \geq 0\} = \{(0, 1)\}.$$

Altogether, for all $x \in W \setminus D_1$ with $x_1 \neq 1/\sqrt{2}$ the multipliers $\lambda_2(x)$ and $\lambda_3(x)$ are unique.

On the other hand, for $x \in W$ with $x_1 = 1/\sqrt{2}$, the active index set of $z_2(x) = -1/\sqrt{2}$ is $I_0^2(x_1, z_2(x)) = \{1, 2\}$, and we obtain the nonunique multiplier set

$$\begin{aligned} KKT_0^2(x) &= \left\{(\lambda_2, \lambda_3) \in \mathbb{R}^2 \;\middle|\; 1 + 2\lambda_2 z_2(x) - \lambda_3 = 0,\ (\lambda_2, \lambda_3) \geq 0\right\} \\ &= \left\{(\lambda_2, \lambda_3) \geq 0 \;\middle|\; \lambda_3 = 1 - \sqrt{2}\lambda_2\right\} \\ &= \left\{\left(t, 1 - \sqrt{2}t\right) \;\middle|\; t \in \left[0, 1/\sqrt{2}\right]\right\}. \end{aligned}$$

Thus, we arrive at

$$D_2 = D_1 \cup \left\{x \in W \;\middle|\; x_1 = 1/\sqrt{2}\right\}$$

and, in view of Theorem 4.17, V is Gâteaux differentiable on $W \setminus D_2$.

Finally, we check for differentiability properties of V on $D_2 \setminus D_1$. Recall that we cannot study differentiability of V on $D_1 = \{(1,0)\}$, as the point $(1,0)$ is not an interior point of dom V. In fact, as mentioned already in Example 4.3, V actually has 'infinite slope' at $(1,0)$.

For all $x \in D_2 \setminus D_1$, that is, $x \in W$ with $x_1 = 1/\sqrt{2}$, Theorem 4.13 yields directional differentiability of V with

$$\begin{aligned} V'(x,d) &= D\theta_1(x)\,d + D\theta_2(x)\,d - D_x L_0^1(x, z_1(x), \lambda_1(x))\,d \\ &\quad - \max_{(\lambda_2,\lambda_3)\in KKT_0^2(x)} D_x L_0^2(x, z_2(x), (\lambda_2,\lambda_3))\,d \\ &= d_1 + d_2 - \frac{1}{2}d_2 - \max_{t\in[0,1/\sqrt{2}]} \left(2t/\sqrt{2} - \left(1 - \sqrt{2}t\right)\right) d_1 \\ &= d_1 + \frac{1}{2}d_2 + \min_{t\in[0,1/\sqrt{2}]} \left(1 - 2\sqrt{2}t\right) d_1 \\ &= \begin{cases} 2d_1 + \frac{1}{2}d_2, & \text{if } d_1 < 0, \\ \frac{1}{2}d_2, & \text{if } d_1 \geq 0, \end{cases} \end{aligned}$$

for all $d \in \mathbb{R}^2$. This corresponds to the 'concave kink' in the graph of V, see Figure 4.2. ◇

The observed differentiability properties in Example 4.18 guarantee that any local minimizer $\bar{x}$ of V on W lies in D_1, or the function V is Gâteaux differentiable at $\bar{x}$. In Theorem 4.23 we will show that, under mild assumptions, a similar assertion holds in the general case. In Section 6.1 we will discuss the essential implications of this fact for the design of numerical methods to solve the optimization problem P in (3.11). We begin now with a preliminary result which gives a representation for the gradient of the Lagrangian L_α^ν from (4.9).

Lemma 4.19 *Let Assumptions 4.1 and 4.7 hold, and let L_α^ν be the Lagrangian of $Q_\alpha^\nu(x)$. Then the gradient with respect to all variables $x = (x^1, \ldots, x^N)$, evaluated at a point $(\bar{x}, \bar{z}^\nu, \lambda^\nu)$ with $\bar{x} \in W$, $\bar{z}^\nu := z_\alpha^\nu(\bar{x})$ and $\lambda^\nu \in KKT_\alpha^\nu(\bar{x})$, has the representation*

$$\nabla_x L_\alpha^\nu(\bar{x}, \bar{z}^\nu, \lambda^\nu) = \nabla\theta_\nu(\bar{z}^\nu, \bar{x}^{-\nu}) + \sum_{i\in I_0^\nu(\bar{z}^\nu,\bar{x}^{-\nu})} \lambda_i^\nu \nabla g_i^\nu(\bar{z}^\nu, \bar{x}^{-\nu}).$$

Proof. The definitions of $L_\alpha^\nu(x, z^\nu, \lambda^\nu)$ and $KKT_\alpha^\nu(x)$ immediately imply that

$$\nabla_{x^\mu} L_\alpha^\nu(\bar{x}, \bar{z}^\nu, \lambda^\nu) = \nabla_{x^\mu}\theta_\nu(\bar{z}^\nu, \bar{x}^{-\nu}) + \sum_{i\in I_0^\nu(\bar{z}^\nu,\bar{x}^{-\nu})} \lambda_i^\nu \nabla_{x^\mu} g_i^\nu(\bar{z}^\nu, \bar{x}^{-\nu}) \tag{4.12}$$

holds for all $\mu \in \{1, \ldots, N\} \setminus \{\nu\}$. Moreover, the combination of

$$\nabla_{x^\nu} L_\alpha^\nu(\bar{x}, \bar{z}^\nu, \lambda^\nu) = \alpha(\bar{x}^\nu - \bar{z}^\nu)$$

and, by the definition of $KKT_\alpha^\nu(\bar{x})$,

$$0 = \nabla_{z^\nu} L_\alpha^\nu(\bar{x}, \bar{z}^\nu, \lambda^\nu) = \nabla_{x^\nu}\theta_\nu(\bar{z}^\nu, \bar{x}^{-\nu}) - \alpha(\bar{x}^\nu - \bar{z}^\nu) + \sum_{i\in I_0^\nu(\bar{z}^\nu,\bar{x}^{-\nu})} \lambda_i^\nu \nabla_{x^\nu} g_i^\nu(\bar{z}^\nu, \bar{x}^{-\nu})$$

shows that (4.12) also holds for $\mu = \nu$ (independently of α). □

The following result will be used in order to show that there exist feasible descent directions for V_α at certain points from W. To this end, we exclude some degenerate points from D_2 for the proof of Theorem 4.23. In fact, we define D_3 to be the set of points in D_2, such that we have

$$\begin{aligned} &\operatorname{span}\{\nabla g_i^\nu(x) \quad (i \in I_0^\nu(x),\ \nu = 1, \dots, N)\} \cap \\ &\operatorname{span}\left\{\nabla g_i^\mu(z_\alpha^\mu(x), x^{-\mu}) \quad (i \in I_0^\mu(z_\alpha^\mu(x), x^{-\mu}))\right\} \neq \{0\} \end{aligned} \tag{4.13}$$

whenever $KKT_\alpha^\mu(x) = \emptyset$ or player SMFC is violated at $z_\alpha^\mu(x)$ for some $\mu \in \{1, \dots, N\}$, that is,

$$\begin{aligned} D_3 := \{x \in D_2 \mid &\text{(4.13) holds at least for one index } \mu \in \{1, \dots, N\} \text{ where} \\ &KKT_\alpha^\mu(x) = \emptyset \text{ or player SMFC is violated at } z_\alpha^\mu(x) \in X_\mu(x^{-\mu})\}. \end{aligned}$$

Note that at least the *unconstrained* points $x \in D_2$, that is, the ones with $I_0^\nu(x) = \emptyset$, $\nu = 1, \dots, N$, do not lie in the set D_3, since $\operatorname{span}\emptyset = \{0\}$ holds. More generally, one may expect in the case $\sum_{\nu=1}^N |I_0^\nu(x)| + |I_0^\mu(z_\alpha^\mu(x), x^{-\mu})| \le n$ that x does not lie in D_3, as the involved gradients are evaluated at different arguments. On the other hand, for $\sum_{\nu=1}^N |I_0^\nu(x)| + |I_0^\mu(z_\alpha^\mu(x), x^{-\mu})| > n$ and linearly independent gradients, x will definitely lie in D_3. Furthermore, under mild conditions, a generalized Nash equilibrium x lying in D_2 is necessarily an element of D_3. In fact, $KKT_\alpha^\nu(x) = \emptyset$ or SMFC is violated at $(z_\alpha^\nu(x), x^{-\nu})$ in $X_\nu(x^{-\nu})$ for $x \in D_2$ and some $\nu \in \{1, \dots, N\}$ so that, in particular, $I_0^\nu(z_\alpha^\nu(x), x^{-\nu})$ is nonempty. Moreover, the points $z_\alpha^\nu(x)$ and x^ν coincide for a generalized Nash equilibrium x by Lemma 3.2 (e). This means that the vectors $\nabla g_i^\nu(x)$ ($i \in I_0^\nu(x)$) appear in both spans in (4.13) if $I_0^\nu(x)$ is nonempty. Thus, except for the case where all these vectors vanish, the intersection of the two spans is strictly larger than $\{0\}$.

With regard to linear independence, we will actually need the following assumption about joint LICQ$_\nu$, which strengthens Assumption 4.9 and is based on Definition 4.15 (a).

Assumption 4.20 *LICQ$_\nu$ holds everywhere in W_ν for each $\nu \in \{1, \dots, N\}$.*

Proposition 4.21 *Let $\bar{x} \in D_2 \setminus (D_1 \cup D_3)$ and Assumptions 4.1, 4.7, and 4.20 hold. Then there exists a vector $d \in \mathbb{R}^n$ solving the system*

$$V_\alpha'(\bar{x}, d) < 0, \quad Dg_i^\nu(\bar{x})\, d \le 0 \quad (i \in I_0^\nu(\bar{x}),\ \nu = 1, \dots, N). \tag{4.14}$$

Proof. Assume that (4.14) does not possess a solution $d \in \mathbb{R}^n$. By Theorem 4.13, we have

$$V_\alpha'(\bar{x}, d) \;=\; \sum_{\nu=1}^N \left[D\theta_\nu(\bar{x})\, d - \max_{\lambda^\nu \in KKT_\alpha^\nu(\bar{x})} D_x L_\alpha^\nu(\bar{x}, \bar{z}^\nu, \lambda^\nu)\, d \right]$$

for all $d \in \mathbb{R}^n$, where we put $\bar{z}^\nu = z_\alpha^\nu(\bar{x})$ for $\nu = 1, \dots, N$. Hence the inconsistency of (4.14) implies that also the system

$$\begin{aligned} \left(\sum_{\nu=1}^N \left[D\theta_\nu(\bar{x}) - D_x L_\alpha^\nu(\bar{x}, \bar{z}^\nu, \lambda^\nu) \right] \right) d \;&<\; 0, \\ Dg_i^\nu(\bar{x})\, d \;&\le\; 0 \quad (i \in I_0^\nu(\bar{x}),\ \nu = 1, \dots, N), \end{aligned}$$

is inconsistent for any choice of

$$\lambda := (\lambda^1, \ldots, \lambda^N) \in KKT^1_\alpha(\bar{x}) \times \ldots \times KKT^N_\alpha(\bar{x}). \tag{4.15}$$

Note that $KKT^\nu_\alpha(\bar{x})$, $\nu = 1, \ldots, N$, are nonempty sets, since $\bar{x} \in W \setminus D_1$ and, thus, $X_\nu(\bar{x}^{-\nu})$ satisfies the Slater condition for all $\nu = 1, \ldots, N$. By Farkas' lemma (see, e.g., [120]), the latter system is inconsistent if and only if there exist scalars $\gamma^\nu_i(\lambda) \geq 0$ for all $i \in I^\nu_0(\bar{x})$, $\nu = 1, \ldots, N$, with

$$\sum_{\nu=1}^N [\nabla\theta_\nu(\bar{x}) - \nabla_x L^\nu_\alpha(\bar{x}, \bar{z}^\nu, \lambda^\nu)] + \sum_{\nu=1}^N \sum_{i \in I^\nu_0(\bar{x})} \gamma^\nu_i(\lambda) \nabla g^\nu_i(\bar{x}) \;=\; 0.$$

After rearranging terms, we find that for any choice λ with (4.15) there exist multipliers $\gamma^\nu_i(\lambda) \geq 0$ with

$$\sum_{\nu=1}^N \Big(\nabla\theta_\nu(\bar{x}) + \sum_{i \in I^\nu_0(\bar{x})} \gamma^\nu_i(\lambda) \nabla g^\nu_i(\bar{x})\Big) \;=\; \sum_{\nu=1}^N \nabla_x L^\nu_\alpha(\bar{x}, \bar{z}^\nu, \lambda^\nu).$$

Using Lemma 4.19 to replace the expression for the gradient on the right-hand side, we conclude that for any choice λ with (4.15) there exist multipliers $\gamma^\nu_i(\lambda) \geq 0$ with

$$\sum_{\nu=1}^N \Big(\nabla\theta_\nu(\bar{x}) + \sum_{i \in I^\nu_0(\bar{x})} \gamma^\nu_i(\lambda) \nabla g^\nu_i(\bar{x})\Big) \;=\; \sum_{\nu=1}^N \Big(\nabla\theta_\nu(\bar{z}^\nu, \bar{x}^{-\nu}) + \sum_{i \in I^\nu_0(\bar{z}^\nu, \bar{x}^{-\nu})} \lambda^\nu_i \nabla g^\nu_i(\bar{z}^\nu, \bar{x}^{-\nu})\Big). \tag{4.16}$$

Next, we use that $\bar{x}$ was chosen from D_2 so that a multiplier λ exists and player SMFC is violated at $\bar{z}^\mu$ at least for one $\mu \in \{1, \ldots, N\}$, say for $\mu = 1$. Then $KKT^1_\alpha(\bar{x})$ contains two different multipliers $\hat{\lambda}^1 \neq \tilde{\lambda}^1$. For $\nu = 2, \ldots, N$, we choose any $\lambda^\nu \in KKT^\nu_\alpha(\bar{x})$ and put $\hat{\lambda} := (\hat{\lambda}^1, \lambda^2, \ldots, \lambda^N)$ as well as $\tilde{\lambda} := (\tilde{\lambda}^1, \lambda^2, \ldots, \lambda^N)$. Equation (4.16) then holds with $\lambda = \hat{\lambda}$ as well as with $\lambda = \tilde{\lambda}$. Subtracting these two equations leads to

$$\sum_{\nu=1}^N \sum_{i \in I^\nu_0(\bar{x})} (\gamma^\nu_i(\hat{\lambda}) - \gamma^\nu_i(\tilde{\lambda})) \nabla g^\nu_i(\bar{x}) \;=\; \sum_{i \in I^1_0(\bar{z}^1, \bar{x}^{-1})} (\hat{\lambda}^1_i - \tilde{\lambda}^1_i) \nabla g^1_i(\bar{z}^1, \bar{x}^{-1}),$$

where the left hand side is some element of

$$\operatorname{span}\{\nabla g^\nu_i(\bar{x}) \quad (i \in I^\nu_0(\bar{x}),\ \nu = 1, \ldots, N)\},$$

and the right hand side is some element of

$$\operatorname{span}\Big\{\nabla g^1_i(\bar{z}^1, \bar{x}^{-1}) \quad (i \in I^1_0(\bar{z}^1, \bar{x}^{-1}))\Big\},$$

which, in addition, cannot vanish due to $\hat{\lambda}^1 \neq \tilde{\lambda}^1$, $(\bar{z}^1, \bar{x}^{-1}) \in W_1$ and Assumption 4.20. Therefore, the intersection of these two spans is nontrivial. However, since $\bar{x}$ was taken from $D_2 \setminus D_3$, this is a contradiction. Consequently, contrary to our assumption, (4.14) must be consistent. This shows the assertion. □

Before we state the main result of this section, we need one more assumption. To this end, we first recall that the *tangent* (or *contingent* or *Bouligand*) *cone* to W at $\bar{x}$ is defined by

$$\mathcal{T}_W(\bar{x}) \; := \; \left\{ d \in \mathbb{R}^n \;\middle|\; \exists t_k \downarrow 0,\ d^k \to d :\ \bar{x} + t_k d^k \in W \text{ for all } k \in \mathbb{N} \right\},$$

and that the *linearization cone* to W at $\bar{x}$ is given by

$$\mathcal{L}_W(\bar{x}) \; := \; \left\{ d \in \mathbb{R}^n \;\middle|\; Dg_i^\nu(\bar{x})\, d \leq 0 \quad \forall i \in I_0^\nu(\bar{x}),\ \nu = 1, \ldots, N \right\}.$$

The inclusion $\mathcal{T}_W(\bar{x}) \subseteq \mathcal{L}_W(\bar{x})$ always holds (cf., e.g., [125]). The *Abadie constraint qualification* (ACQ) is said to hold at $\bar{x} \in W$ if both cones actually coincide.

Assumption 4.22 *ACQ holds everywhere in* W.

ACQ is typically considered a very weak constraint qualification. Nevertheless, we point out that the example from Remark 4.10 shows that neither Assumption 4.9 nor Assumption 4.20 imply Assumption 4.22.

The following is the main result of this section.

Theorem 4.23 *Let Assumptions 4.1, 4.7, 4.20, and 4.22 hold. Then any local minimizer $\bar{x}$ of V_α on W lies in $D_1 \cup D_3$, or the function V_α is Gâteaux differentiable at $\bar{x}$.*

Proof. By Theorem 4.17, V_α is Gâteaux differentiable everywhere on $W \setminus D_2$. Choose any $\bar{x} \in D_2 \setminus (D_1 \cup D_3)$. The assertion is shown if we can prove that $\bar{x}$ is not a local minimizer of V_α on W. The main idea of the proof is to show this by guaranteeing the existence of a first order feasible descent direction for V_α on W in $\bar{x}$.

In view of Proposition 4.21, there exists a vector $d \in \mathbb{R}^n$ solving the system (4.14). In particular, d belongs to the linearization cone $\mathcal{L}_W(\bar{x})$. In view of Assumption 4.22, d also lies in the tangent cone $\mathcal{T}_W(\bar{x})$, that is, there exist sequences $t_k \downarrow 0$ and $d^k \to d$ with $\bar{x} + t_k d^k \in W$ for all $k \in \mathbb{N}$.

Assume that $\bar{x}$ is a local minimizer of V_α on W. This implies $V_\alpha(\bar{x} + t_k d^k) \geq V_\alpha(\bar{x})$ and, thus,

$$\frac{V_\alpha(\bar{x} + t_k d^k) - V_\alpha(\bar{x})}{t_k} \; \geq \; 0 \tag{4.17}$$

for all sufficiently large $k \in \mathbb{N}$. By Theorem 4.13, V_α is Hadamard directionally differentiable at $\bar{x}$ so that the limit of the left-hand side in (4.17) exists and equals $V_\alpha'(\bar{x}, d)$ (note that here we exploit the fact that V_α is actually Hadamard directionally differentiable and not just directionally differentiable in the ordinary sense). However, since the implication $V_\alpha'(\bar{x}, d) \geq 0$ contradicts (4.14), $\bar{x}$ cannot be a local minimizer of V_α on W. □

It is well-known (cf., e.g., [125, Prop. 3.2]) that ACQ holds everywhere in W in the case where all constraints g_i^ν are linear. This immediately leads to the following result.

Corollary 4.24 *Let Assumptions 4.1, 4.7, and 4.20 hold, and assume that all constraint functions g_i^ν are linear. Then any local minimizer $\bar{x}$ of V_α on W lies in $D_1 \cup D_3$, or the function V_α is Gâteaux differentiable at $\bar{x}$.*

We remark that in Example 4.18 the two constrained elements of D_2 belong to D_3, but first order feasible descent directions for V_α on W still exist in these points. This indicates that it should be possible to weaken the assumptions of Theorem 4.23.

With respect to Remark 4.12, we note that, coarsely speaking, 'concavity property prevails in the kinks of V_α' although V_α is not necessarily concave under the general assumptions of Theorem 4.23.

For completeness, we emphasize that Gâteaux differentiability of V_α at $\bar{x}$ in Theorem 4.17 can be replaced by continuous differentiability if V_α is Gâteaux differentiable on a neighborhood of $\bar{x}$ and the partial derivatives of V_α are continuous. The next corollary states a sufficient condition for this situation.

Corollary 4.25 *Let Assumptions 4.1 and 4.7 hold. Additionally, for $\bar{x} \in W \setminus D_2$ let player LICQ hold at $\bar{z}^\nu := z_\alpha^\nu(\bar{x})$ with the unique multiplier $\bar{\lambda}^\nu := \lambda_\alpha^\nu(\bar{x})$ for each $\nu \in \{1, \dots, N\}$. Then V_α is continuously differentiable on a neighborhood of $\bar{x}$ with*

$$\nabla V_\alpha(\bar{x}) \;=\; \sum_{\nu=1}^{N} \left[\nabla\theta_\nu(\bar{x}) - \nabla_x L_\alpha^\nu(\bar{x}, \bar{z}^\nu, \bar{\lambda}^\nu) \right].$$

Proof. First, in view of Lemmata 4.5 and 4.16, we have $\bar{x} \in \operatorname{int}\operatorname{dom} V_\alpha$, and therefore it makes sense to study differentiability of V_α at $\bar{x}$. In the following let $\nu \in \{1, \dots, N\}$. As in the proof of Lemma 4.5, one can show that the function $z_\alpha^\nu(x)$ is defined on a whole neighborhood U of $\bar{x}$ and that U does not contain points from D_1. Thus, by Lemma 4.2, the function z_α^ν is continuous on the whole set U. As player LICQ is stable under perturbations, U can also be chosen such that player LICQ holds at $z_\alpha^\nu(x)$ for each $x \in U$. Consequently, the set-valued mapping $KKT_\alpha^\nu(x)$ is single-valued on U, say $KKT_\alpha^\nu(x) \equiv \{\lambda_\alpha^\nu(x)\}$ with $\lambda_\alpha^\nu(\bar{x}) = \bar{\lambda}^\nu$.

Since not only the function z_α^ν but also the function λ_α^ν is continuous on U by [83, Lemma 2] for each $\nu \in \{1, \dots, N\}$, the partial derivatives of V_α from Theorem 4.17 are continuous on U. This shows continuous differentiability of V_α on U. □

Under the assumptions of Corollary 4.25, Gâteaux differentiability can be replaced by continuous differentiability also in Theorem 4.23. Thus, if under the assumptions of Theorem 4.23 each $x \in W \setminus D_2$ satisfies the assumptions of Corollary 4.25, then each local minimizer $\bar{x}$ of V_α on W lies in $D_1 \cup D_3$, or V_α is continuously differentiable on a neighborhood of $\bar{x}$. Note that Example 4.3 satisfies these assumptions.

The next remark shows that the assumption of player LICQ in Corollary 4.25 can be replaced by the assumption of stable player SMFC.

Remark 4.26 Let $\bar{x} \in W \setminus D_2$. Then the assumption of player LICQ at $z_\alpha^\nu(\bar{x}) \in X_\nu(\bar{x}^{-\nu})$ for each $\nu \in \{1, \dots, N\}$ in Corollary 4.25 can be replaced in cases when player SMFC is stable at $z_\alpha^\nu(\bar{x}) \in X_\nu(\bar{x}^{-\nu})$ for each $\nu \in \{1, \dots, N\}$. Using the characterization of the set D_2 from (4.11), player SMFC already holds at $z_\alpha^\nu(\bar{x}) \in X_\nu(\bar{x}^{-\nu})$ for all $\nu = 1, \dots, N$. Hence if additionally the sets $I_{00}^\nu(\bar{x}) = \{i \in I_0^\nu(\bar{x}) \mid \bar{\lambda}_i^\nu = 0\}$, $\nu = 1, \dots, N$, remain constant under small perturbations of $\bar{x}$ (this is true if, for example, the sets $I_{00}^\nu(\bar{x})$, $\nu = 1, \dots, N$, are empty, that is, the so-called *strict complementarity slackness* holds), then continuity arguments show that SMFC is stable at $z_\alpha^\nu(\bar{x})$

for each $\nu \in \{1, \dots, N\}$ under sufficiently small perturbations of $\bar{x}$. After this observation, one can show continuous differentiability of V_α on a neighborhood of $\bar{x}$ along the lines of the proof of Corollary 4.25. ◇

The previous results of Chapter 4 can also be used for jointly convex GNEPs, but the following remark proposes a better way how to use the approaches from Chapter 4 for this class of GNEPs.

Remark 4.27 Recall that in the jointly convex case one assumes identical constraints for all players, $g^1 = g^2 = \ldots = g^N =: g$, and that the components of g are convex in the whole vector $x = (x^1, \dots, x^N)$. Thus, the strategy spaces have the representation

$$X_\nu(x^{-\nu}) \;=\; \{x^\nu \in \mathbb{R}^{n_\nu} \mid g(x^\nu, x^{-\nu}) \le 0\} \quad (\nu = 1, \dots, N),$$

so that $x \in \Omega(x)$ holds if and only if x lies in the set

$$\widetilde{W} \;:=\; \{x \in \mathbb{R}^n \mid g(x) \le 0\}\,. \tag{4.18}$$

Note that, in contrast to the player convex case, $\widetilde{W}$ is necessarily convex (and Assumption 4.6 automatically holds). An important observation is that the definition of $\widetilde{W}$ is slightly different from the definition of W in (3.5). In fact, while the geometries of both sets coincide, their functional descriptions are different as, formally, W is described by N identical inequalities $g(x) \le 0$ in the jointly convex case. With regard to constraint qualifications like LICQ, the latter description of W is necessarily degenerate at boundary points, while the description of $\widetilde{W}$ from (4.18) may enjoy nondegeneracy properties.

In particular, all sets W_ν, $\nu = 1, \dots, N$, from (4.5) coincide with $\widetilde{W}$, and Assumption 4.9 coincides with the assumption of MFCQ everywhere in $\widetilde{W}$. Of course, the latter is equivalent to the Slater condition for $\widetilde{W}$ and implies the Abadie constraint qualification for $\widetilde{W}$. Furthermore, Assumption 4.20 of LICQ$_\nu$ at each point of each set W_ν, $\nu = 1, \dots, N$, can be replaced by the assumption of LICQ at each point of $\widetilde{W}$ which, in turn, implies the Slater condition for $\widetilde{W}$. In any case, Assumption 4.22 is superfluous in Theorem 4.23 for jointly convex problems. Hence, in contrast to the player convex case, our assumptions on constraint qualifications are highly interrelated in the jointly convex case. We also remark that, in the jointly convex case, D_3 is defined to be the set of points $x \in D_2$ such that at least for one index $\mu \in \{1, \dots, N\}$ for which $KKT^\mu_\alpha(x) = \emptyset$ or player SMFC is violated at $z^\mu_\alpha(x) \in X_\mu(x^{-\mu})$ we have

$$\operatorname{span}\{\nabla g_i(x) \quad (i \in I_0(x))\} \,\cap\, \operatorname{span}\{\nabla g_i(z^\mu_\alpha(x), x^{-\mu}) \quad (i \in I_0(z^\mu_\alpha(x), x^{-\mu}))\} \;\neq\; \{0\}. \tag{4.19}$$

We emphasize that so-called shared constraints as in the jointly convex case lead to numerical difficulties in all established numerical methods for the solution of GNEPs. In fact, while repeating identical constraints can lead to degeneracies in any numerical approach, dropping redundant constraints basically leads to underdetermined systems and, thus, alternative numerical problems in all approaches, which we are aware of. In contrast to this, dropping redundant constraints in the present approach by switching from W to $\widetilde{W}$ does not introduce numerical problems. ◇

The following example illustrates for the jointly convex case that a generalized Nash equilibrium can fall in any category mentioned in Theorem 4.23.

Example 4.28 We slightly modify Example 4.3 by setting $N = 2$, $n_1 = n_2 = 1$, $\theta_1(x) = -x_1$, $\theta_2(x) = x_2$, $g^1_1(x) = g^2_1(x) = -2x_1 + x_2$, $g^1_2(x) = g^2_2(x) = x_1^2 + x_2^2 - 1$, $g^1_3(x) = g^2_3(x) = -x_1 - x_2$. Then for all $x \in \widetilde{W} = \{x \in \mathbb{R}^2 |\ x_1^2 + x_2^2 \leq 1,\ -x_1 \leq x_2 \leq 2x_1\}$ (cf. Figure 4.1) the problems $Q^1_\alpha(x)$ and $Q^2_\alpha(x)$ are easily seen to be uniquely solvable for $\alpha = 0$. Furthermore, it is not hard to see that the assumptions of Theorem 4.23 with the modifications from Remark 4.27 are satisfied.

As to be expected for problems with shared constraints, the set of generalized Nash equilibria is not a singleton, but it is formed by the (closed) boundary arc of $\widetilde{W}$ connecting the points $\bar{x} := 1/\sqrt{2}\,(1, -1)$ and $\hat{x} := (1, 0)$. Both $\bar{x}$ and $\hat{x}$ are elements of D_1, and $\bar{x}$ also lies in D_3, since player SMFC is violated at the point $z^2_0(\bar{x}) = -1/\sqrt{2} \in X_2\left(1/\sqrt{2}\right)$ and the intersection of spans in (4.19) is $\mathbb{R}^2$.

On the other hand, by direct inspection or as shown in Theorem 4.23, the resulting function

$$V(x) := V_0(x) = -x_1 + x_2 + \sqrt{1 - x_2^2} + \min\{x_1, \sqrt{1 - x_1^2}\}$$

with $x \in \widetilde{W}$ is differentiable at all generalized Nash equilibria except for $\bar{x}$ and $\hat{x}$. ◇

The statements of this section motivate the application of certain smooth optimization techniques on the optimization problem P from (3.11). Therefore, we will draw on this fact and present some numerical results in Section 6.1.

5. Smoothness Properties of a Dual Gap Function for Generalized Nash Equilibrium Problems

The reformulations of GNEPs in (3.8) and (3.11) as constrained minimization problems have, in general, nonsmooth objective functions V and V_α. Consequently, also the primal gap function $V_\alpha + \delta_W$ for the reformulation in (3.12) is, in general, nonsmooth. However, based on an idea by Dietrich [30], this primal gap function may be viewed as a difference of two convex functions, which then allows the application of the Toland-Singer duality theory [123, 130, 131] in order to obtain a smooth and unconstrained optimization reformulation of certain GNEPs. In Section 5.1 we develop such a smooth and unconstrained dual optimization reformulation for a class of player convex GNEPs. In Section 5.2 we then apply a smoothness result from parametric optimization observed in Section 2.2 to our particular setting and therefore obtain second-order properties of our unconstrained objective function. The results of Chapter 5 come from the paper [73].

5.1. A Smooth Dual Gap Function

It is possible to obtain a smooth reformulation of certain GNEPs in case that the optimal value function $V_\alpha + \delta_W$ of the the unconstrained minimization problem in (3.12) can be rewritten as a difference of two strongly convex and lsc functions (see Definition 2.5). A class of player convex GNEPs satisfying the next assumption has this property.

Assumption 5.1 *(a) The feasible set $W = \{x \in \mathbb{R}^n \mid x \in \Omega(x)\}$ of the GNEP* (3.2) *is nonempty.*

(b) The cost functions $\theta_\nu : \mathbb{R}^n \to \mathbb{R}$, $\nu = 1, \dots, N$, are convex on $\mathbb{R}^n$.

(c) The set-valued mappings $X_\nu : \mathbb{R}^{n-n_\nu} \rightrightarrows \mathbb{R}^{n_\nu}$, $\nu = 1, \dots, N$, are graph-convex and osc on $\mathbb{R}^{n-n_\nu}$ (see Definition 2.1).

Since these assumptions play a central role within our analysis in Chapter 5, we would like to add a few comments. Assumption 5.1 (a) is rather natural, since otherwise the corresponding GNEP is not solvable. Assumption 5.1 (c) is satisfied if, for example, X_ν, $\nu = 1, \dots, N$, are defined by (3.3) with functions $g^\nu : \mathbb{R}^n \to \mathbb{R}^{m_\nu}$, $\nu = 1, \dots, N$, which are convex in the whole variable $x = (x^\nu, x^{-\nu})$ on $\mathbb{R}^n$. In particular, Assumption 5.1 (c) therefore holds for the class of jointly convex GNEPs, where $g^1 = g^2 = \dots = g^N =: g$ and the components of g are convex in the whole variable, cf. [47] for more details. However, we do not use the representation of X_ν with

g^ν here. Finally, Assumption 5.1 (b) is probably the most restrictive condition, since it requires all cost functions to be convex in the entire vector x. Note that this assumption provides the continuity of all cost functions on $\mathbb{R}^n$. However, we will see later that the convexity assumption on the cost functions can be relaxed considerably, see the discussion following Lemma 5.7. In order to avoid any technical discussion, it is convenient to assume this condition to formulate and prove the subsequent results.

Assumption 5.1 (c) yields the following result.

Lemma 5.2 *Let Assumption 5.1 (c) hold. Then the set-valued mapping* $\Omega : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *defined by* (3.1) *is graph-convex and osc on* $\mathbb{R}^n$.

Proof. First we verify the graph-convexity of Ω. By the definition of Ω in (3.1) and Assumption 5.1 (c), it holds that

$$\begin{aligned}
\operatorname{gph}\Omega &= \{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n \mid z \in \Omega(x)\} \\
&= \{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n \mid z^\nu \in X_\nu(x^{-\nu}) \;\; \forall \nu = 1,\ldots,N\} \\
&= \bigcap_{\nu=1}^{N} \{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n \mid z^\nu \in X_\nu(x^{-\nu})\} \\
&= \bigcap_{\nu=1}^{N} \{(x^\nu, x^{-\nu}, z^\nu, z^{-\nu}) \in \mathbb{R}^{n_\nu} \times \mathbb{R}^{n-n_\nu} \times \mathbb{R}^{n_\nu} \times \mathbb{R}^{n-n_\nu} \mid (x^{-\nu}, z^\nu) \in \operatorname{gph} X_\nu\} \\
&=: \bigcap_{\nu=1}^{N} C_\nu
\end{aligned}$$

with the convex sets $\operatorname{gph} X_\nu := \{(x^{-\nu}, z^\nu) \in \mathbb{R}^{n-n_\nu} \times \mathbb{R}^{n_\nu} \mid z^\nu \in X_\nu(x^{-\nu})\}$ for all $\nu \in \{1,\ldots,N\}$. Furthermore, the sets C_ν, $\nu = 1,\ldots,N$, are convex as suitable Cartesian products of the convex sets $\operatorname{gph} X_\nu$ and $\mathbb{R}^{n_\nu} \times \mathbb{R}^{n-n_\nu}$. Then the set $\operatorname{gph}\Omega$ is convex as an intersection of convex sets. Therefore, the set-valued mapping Ω is graph-convex.

Next, we show that the set-valued mapping Ω is osc on $\mathbb{R}^n$. Let $\bar{x} \in \mathbb{R}^n$ be arbitrarily given. Since the set-valued mappings $X_\nu : \mathbb{R}^{n-n_\nu} \rightrightarrows \mathbb{R}^{n_\nu}$, $\nu = 1,\ldots,N$, are osc on $\mathbb{R}^{n-n_\nu}$, for all sequences $\left\{x^{k,-\nu}\right\} \subseteq \mathbb{R}^{n-n_\nu}$ with $x^{k,-\nu} \to \bar{x}^{-\nu}$ and all sequences $z^{k,\nu} \to \bar{z}^\nu$ with $z^{k,\nu} \in X_\nu(x^{k,-\nu})$ for all $k \in \mathbb{N}$ sufficiently large, we have $\bar{z}^\nu \in X_\nu(\bar{x}^{-\nu})$. Then for all sequences $\left\{x^k\right\} \subseteq \mathbb{R}^n$ with $x^k \to \bar{x}$ and $z^k \to \bar{z}$ with $z^k \in X_1(x^{k,-1}) \times \ldots \times X_N(x^{k,-N}) = \Omega(x^k)$ for all $k \in \mathbb{N}$ sufficiently large, we have $\bar{z} \in X_1(\bar{x}^{-1}) \times \ldots \times X_N(\bar{x}^{-N}) = \Omega(\bar{x})$. Consequently, Ω is osc at $\bar{x}$. Since $\bar{x} \in \mathbb{R}^n$ was arbitrarily chosen, the set-valued mapping Ω is osc on $\mathbb{R}^n$. □

The next result, which follows from Lemma 5.2 together with Assumption 5.1 (a), is critical for the further analysis.

Lemma 5.3 *Let Assumptions 5.1 (a) and (c) hold. Then:*

(a) The feasible set W of the GNEP (3.2) *defined in* (3.4) *is nonempty, closed, and convex.*

(b) The domain $\operatorname{dom}\Omega$ *from* (3.7) *of the set-valued mapping* Ω *is nonempty and convex.*

Proof. (a) In view of Assumption 5.1 (a), the set W is nonempty. Furthermore, let $\{x^k\} \subseteq W$ be an arbitrary convergent sequence with a limit $\bar{x} \in \mathbb{R}^n$. Then $x^k \in \Omega(x^k)$ for all $k \in \mathbb{N}$. Since the set-valued mapping $\Omega : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is osc by Lemma 5.2, it follows that $\bar{x} \in \Omega(\bar{x})$. Therefore, $\bar{x} \in W$ so that the set W is closed.

Next, we show that W is convex. To this end, let $x_1, x_2 \in W$ and $t \in [0, 1]$ be arbitrarily given. Then $x_1 \in \Omega(x_1)$ and $x_2 \in \Omega(x_2)$. By Lemmata 5.2 and 2.4 (b), it follows that Ω is graph-convex and $tx_1+(1-t)x_2 \in \Omega\big(tx_1+(1-t)x_2\big)$, that is, $tx_1+(1-t)x_2 \in W$. Hence the set W is convex.

(b) The set $\operatorname{dom}\Omega$ is nonempty, since $\operatorname{dom}\Omega$ contains the nonempty set W in view of Lemma 3.2 (a). The convexity of $\operatorname{dom}\Omega$ follows immediately from graph-convexity of Ω, see Lemma 2.4 (b). □

The subsequent example illustrates that even for a graph-convex and osc set-valued mapping Ω its domain $\operatorname{dom}\Omega$ is not necessarily closed.

Example 5.4 Consider a GNEP with two players having arbitrary cost functions and each controlling a single variable, which, for simplicity of notation, we call x_1 and x_2, respectively. Furthermore, let $\Omega : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ be given by $\Omega(x) = X_1(x_2) \times X_2(x_1)$ with the set-valued mappings X_1, $X_2 : \mathbb{R} \rightrightarrows \mathbb{R}$ defined by

$$X_1(x_2) := \begin{cases} \left\{x_1 \in \mathbb{R} \,\middle|\, x_1 \geq \frac{1}{x_2}\right\}, & \text{if } x_2 > 0, \\ \emptyset, & \text{if } x_2 \leq 0, \end{cases} \qquad \text{and} \qquad X_2(x_1) := [0, \infty[.$$

These set-valued mappings X_1 and X_2 are obviously graph-convex. Furthermore, X_1 and X_2 are osc on $\mathbb{R}$, since, if $x_2^k \downarrow 0$, all sequences $\{z_1^k\}$ with $z_1^k \in X_1(x_2^k)$ are divergent, and all other cases are unproblematic. In view of Lemma 5.2, the set-valued mapping Ω is also graph-convex and osc on $\mathbb{R}^2$. On the other hand, $\operatorname{dom}\Omega = \mathbb{R} \times \mathbb{R}_>$ is not closed. ◇

In order to obtain a differentiable reformulation of GNEPs satisfying Assumption 5.1, we rewrite the unconstrained objective function from (3.12) as a difference of two strongly convex and lsc functions and apply the duality theory by Toland [130] and Singer [123] to this DC minimization problem. Note that a problem is called *DC minimization problem* if it consists of the minimization of a difference of two convex functions. For a survey of DC programming, we refer to [84].

Using the reformulation of the optimal value function V_α in (3.9) with a parameter $\alpha > 0$ and the specific structure of this function, we first observe

$$V_\alpha(x) + \delta_W(x) = \varrho(x) - \varphi_\alpha(x) \tag{5.1}$$

with the functions $\varrho : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $\varphi_\alpha : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\varrho(x) := \sum_{\nu=1}^{N} \theta_\nu(x) + \delta_W(x) \quad \text{and} \quad \varphi_\alpha(x) := \inf_{z \in \Omega(x)} \left(\sum_{\nu=1}^{N} \theta_\nu(z^\nu, x^{-\nu}) + \frac{\alpha}{2}\|x - z\|^2 \right), \tag{5.2}$$

where the infimum is uniquely attained at $z_\alpha(x)$ defined in (3.10) for all $x \in \operatorname{dom}\Omega$ and takes the value $+\infty$ for all $x \notin \operatorname{dom}\Omega$. The functions ϱ and φ_α are lsc and convex. Lower semicontinuity and convexity of ϱ are easily verified, since θ_ν, $\nu = 1, \dots, N$, is continuous and convex on $\mathbb{R}^n$ by Assumption 5.1 (b) and the set W is closed and convex by Lemma 5.3 (a). For the proof of lower semicontinuity and convexity of the function φ_α we need the following result.

Lemma 5.5 *Let $\Omega : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be graph-convex and osc on $\mathbb{R}^n$. Then the function*

$$\Psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, \quad \Psi(z, x) := \delta_{\Omega(x)}(z)$$

is lsc and convex in (z, x).

Proof. First we show that Ψ is lsc: To this end, let $\{(z^k, x^k, \gamma_k)\} \subseteq \operatorname{epi}\Psi$ such that (z^k, x^k, γ_k) converges to $(\bar{z}, \bar{x}, \bar{\gamma})$. In particular, it holds that $\gamma_k \geq 0$, hence $\bar{\gamma} \geq 0$. On the other hand, we have $\delta_{\Omega(x^k)}(z^k) \leq \gamma_k < \infty$, hence it necessarily follows from the definition of the indicator function that $\delta_{\Omega(x^k)}(z^k) = 0$, that is, $z^k \in \Omega(x^k)$ holds for all $k \in \mathbb{N}$. Since Ω is osc, we therefore have $\bar{z} \in \Omega(\bar{x})$ and thus, $\delta_{\Omega(\bar{x})}(\bar{z}) = 0 \leq \bar{\gamma}$, hence $(\bar{z}, \bar{x}, \bar{\gamma}) \in \operatorname{epi}\Psi$. It follows that $\operatorname{epi}\Psi$ is closed, that is, Ψ is lsc.

It remains to prove that Ψ is convex in (z, x): For these purposes, let $(z, x, \gamma), (z', x', \gamma') \in \operatorname{epi}\Psi$ and $t \in [0, 1]$. Similar to the first part of the proof, it follows that $z \in \Omega(x)$ and $z' \in \Omega(x')$. Consequently, we have $tz \in t\Omega(x)$ and $(1 - t)z' \in (1 - t)\Omega(x')$ and hence, due to the graph-convexity of Ω, we observe $tz + (1 - t)z' \in \Omega(tx + (1 - t)x')$, cf. Lemma 2.4 (b). Hence

$$\Psi\left(tz + (1 - t)z', tx + (1 - t)x'\right) = 0 \leq t\gamma + (1 - t)\gamma',$$

and thus, $\operatorname{epi}\Psi$ is convex, that is, Ψ is convex in (z, x). □

Note that the proof of Lemma 5.5 does not come from the paper [73] but has been already published in the former paper [72, Lemma 3.3]. Using Lemma 5.5, we are now in position to verify lower semicontinuity and convexity of the function φ_α defined in (5.2).

Lemma 5.6 *Let Assumption 5.1 hold. Then the function φ_α is proper, lsc, and convex on $\mathbb{R}^n$.*

Proof. In view of (5.2), we rewrite φ_α as

$$\varphi_\alpha(x) = \inf_{z \in \mathbb{R}^n} \tau_\alpha(z, x),$$

where $\tau_\alpha : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\tau_\alpha(z, x) := \sum_{\nu=1}^{N} \theta_\nu(z^\nu, x^{-\nu}) + \frac{\alpha}{2}\|x - z\|^2 + \delta_{\Omega(x)}(z).$$

By Assumption 5.1 and Lemma 5.5, each summand of τ_α is convex and (at least) lsc on $\mathbb{R}^n \times \mathbb{R}^n$. Hence τ_α is lsc and convex. Furthermore, the function τ_α is proper, since $\operatorname{dom}\tau_\alpha = \operatorname{dom}\Omega \neq \emptyset$. Additionally, the mapping

$$\operatorname*{argmin}_{z \in \mathbb{R}^n} \ \tau_\alpha(z, x) = \{z_\alpha(x)\}$$

is single-valued for all $x \in \operatorname{dom}\Omega$. Thus, the assertions follow from [117, Corollary 3.32]. □

Since the functions ϱ and φ_α are lsc and convex, the representation in (5.1) is an lsc DC formulation of the unconstrained objective function from (3.12). For the purpose of a differentiable dual reformulation of GNEPs satisfying Assumption 5.1, we add to both functions ϱ and φ_α the same strongly convex quadratic term. This alteration leads to the following DC decomposition of the optimal value function from (3.12):

$$V_\alpha(x) + \delta_W(x) = f_\alpha(x) - h_\alpha(x)$$

with two functions $f_\alpha, h_\alpha : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f_\alpha(x) := \frac{\alpha}{2}\|x\|^2 + \sum_{\nu=1}^{N} \theta_\nu(x) + \delta_W(x) = \frac{\alpha}{2}\|x\|^2 + \varrho(x), \tag{5.3}$$

$$h_\alpha(x) := \frac{\alpha}{2}\|x\|^2 + \inf_{z\in\Omega(x)} \left(\sum_{\nu=1}^{N} \theta_\nu(z^\nu, x^{-\nu}) + \frac{\alpha}{2}\|x - z\|^2 \right) = \frac{\alpha}{2}\|x\|^2 + \varphi_\alpha(x). \tag{5.4}$$

In principle, we could have used another parameter for the quadratic term in the functions f_α and h_α than for the quadratic term in the function φ_α. Furthermore, the quadratic term $\frac{\alpha}{2}\|x\|^2$ could be replaced by any strongly convex function without really changing the subsequent theory.

Some elementary properties of the above DC decomposition are summarized in the following result.

Lemma 5.7 *Let Assumption 5.1 hold, and let f_α and h_α be defined as in* (5.3) *and* (5.4)*, respectively. Then the following statements hold:*

(a) The function f_α is lsc and strongly convex on $\mathbb{R}^n$ and has the domain W.

(b) The function h_α is lsc and strongly convex on $\mathbb{R}^n$ and has the domain $\operatorname{dom}\Omega$.

(c) $\bar{x}$ is a solution of the GNEP if and only if it is a solution of the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} [f_\alpha(x) - h_\alpha(x)]$$

with optimal function value equal to zero.

Note that the previous result still holds for certain classes of nonconvex but continuous cost functions θ_ν, that is, for functions not satisfying Assumption 5.1 (b). This follows directly from the definitions of f_α and h_α, since these functions may become strongly convex even for nonconvex and continuous functions θ_ν by adding a suitable strongly convex term. For example, for quadratic cost functions θ_ν, it can be achieved by adding the strongly convex quadratic term $\frac{\alpha}{2}\|x\|^2$ with a sufficiently large parameter α. This observation will be exploited in our numerical section in order to compute a suitable parameter α.

Before we apply the duality theory by Toland and Singer to this DC decomposition, we consider the required conjugate functions of f_α and h_α in the next two results. The definition of a conjugate function is given in Definition 2.5 (e).

Lemma 5.8 *Let Assumption 5.1 hold. Then the following statements hold for the conjugate f_α^* of f_α:*

(a) f_α^* is given by

$$f_\alpha^*(y) = x_\alpha^{f^*}(y)^T y - \frac{\alpha}{2}\left\|x_\alpha^{f^*}(y)\right\|^2 - \sum_{\nu=1}^N \theta_\nu(x_\alpha^{f^*}(y))$$

where $x_\alpha^{f^*}(y)$ denotes the unique solution of the maximization problem

$$\max_x \left[x^T y - \frac{\alpha}{2}\|x\|^2 - \sum_{\nu=1}^N \theta_\nu(x)\right] \quad \textit{subject to} \quad x \in W.$$

(b) f_α^* has the domain $\operatorname{dom} f_\alpha^* = \mathbb{R}^n$.

(c) f_α^* is differentiable with Lipschitz gradient given by $\nabla f_\alpha^*(y) = x_\alpha^{f^*}(y)$.

Proof. Application of Definition 2.5 (e) leads to

$$f_\alpha^*(y) = \sup_{x\in W}\left[x^T y - \frac{\alpha}{2}\|x\|^2 - \sum_{\nu=1}^N \theta_\nu(x)\right] =: \max_{x\in W} F_\alpha(x,y). \tag{5.5}$$

The function F_α is continuous on $\mathbb{R}^n \times \mathbb{R}^n$ and strongly concave in x for each fixed $y \in \mathbb{R}^n$. Since the set W is nonempty, closed, and convex by Lemma 5.3 (a), the maximization problem in (5.5) has a unique solution $x_\alpha^{f^*}(y)$ for each fixed $y \in \mathbb{R}^n$, so that $\operatorname{dom} f_\alpha^* = \mathbb{R}^n$. This proves statements (a) and (b).

Furthermore, the function F_α is continuously differentiable in the second variable for each fixed $x \in \mathbb{R}^n$, and the mapping $y \mapsto x_\alpha^{f^*}(y)$ is continuous on $\mathbb{R}^n$ in view of Lemma 2.3 based on [82, Corollaries 8.1 and 9.1]. Due to Danskin's Theorem (see, e.g., [8, Chapter 4, Theorem 1.7] or [19]), the function f_α^* is continuously differentiable with

$$\nabla f_\alpha^*(y) = \nabla_y F_\alpha(x,y)\Big|_{x=x_\alpha^{f^*}(y)} = x_\alpha^{f^*}(y).$$

In view of Lemma 2.7, this gradient ∇f_α^* is even Lipschitz. This completes the proof. □

In a similar way as for the function f_α we consider the conjugate function of h_α.

Lemma 5.9 *Let Assumption 5.1 hold. Then the following statements hold for the conjugate h_α^* of h_α:*

(a) $h_\alpha^*(y)$ is given by

$$h_\alpha^*(y) = x_\alpha^{h^*}(y)^T y - \frac{\alpha}{2}\left\|x_\alpha^{h^*}(y)\right\|^2 - \sum_{\nu=1}^N \theta_\nu(z_\alpha^{h^*}(y)^\nu, x_\alpha^{h^*}(y)^{-\nu}) - \frac{\alpha}{2}\left\|x_\alpha^{h^*}(y) - z_\alpha^{h^*}(y)\right\|^2$$

where $(x_\alpha^{h^*}, z_\alpha^{h^*})(y)$ is the unique solution of the maximization problem

$$\max_{(x,z)}\left[x^T y - \frac{\alpha}{2}\|x\|^2 - \sum_{\nu=1}^N \theta_\nu(z^\nu, x^{-\nu}) - \frac{\alpha}{2}\|x-z\|^2\right] \quad \textit{subject to} \quad (x,z) \in \operatorname{gph}\Omega.$$

(b) $h_\alpha^*(y)$ *has the domain* $\operatorname{dom} h_\alpha^* = \mathbb{R}^n$.

(c) $h_\alpha^*(y)$ *is differentiable with Lipschitz gradient given by* $\nabla h_\alpha^*(y) = x_\alpha^{h^*}(y)$.

Proof. Due to Definition 2.5 (e), we obtain

$$h_\alpha^*(y) = \sup_{x\in\mathbb{R}^n} \left[x^T y - \frac{\alpha}{2}\|x\|^2 - \inf_{z\in\Omega(x)} \left(\sum_{\nu=1}^N \theta_\nu(z^\nu, x^{-\nu}) + \frac{\alpha}{2}\|x-z\|^2 \right) \right] \tag{5.6}$$

$$= \max_{(x,z)\in\operatorname{gph}\Omega} \left[x^T y - \frac{\alpha}{2}\|x\|^2 - \sum_{\nu=1}^N \theta_\nu(z^\nu, x^{-\nu}) - \frac{\alpha}{2}\|x-z\|^2 \right] =: \max_{(x,z)\in\operatorname{gph}\Omega} H_\alpha(x,z,y). \tag{5.7}$$

The function H_α is continuous on $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$, strongly concave in (x,z) for each fixed $y \in \mathbb{R}^n$, and continuously differentiable in the third variable for each fixed $(x,z) \in \mathbb{R}^n \times \mathbb{R}^n$. Since $\operatorname{gph}\Omega$ is nonempty, closed and convex by Assumption 5.1 (a) and (c), the proof of all statements of Lemma 5.9 is analogous to the proof of Lemma 5.8. Here, it holds that

$$\nabla h_\alpha^*(y) = \nabla_y H_\alpha(x,z,y)\Big|_{(x,z)=(x_\alpha^{h^*}, z_\alpha^{h^*})(y)} = x_\alpha^{h^*}(y)$$

where $(x_\alpha^{h^*}, z_\alpha^{h^*})(y)$ is the unique maximal point in (5.7). □

The following simple example illustrates the two previous results.

Example 5.10 Consider a GNEP satisfying Assumption 5.1 with $N = 2$, $n_1 = n_2 = 1$, the variables x_1 and x_2 controlled by player 1 and 2, respectively, the cost functions $\theta_1(x) := x_1^2$, $\theta_2(x) := x_2$, the constraints $g_1^2(x) := x_1 - x_2 \le 0$ and $g_2^2(x) := -x_1 - x_2 \le 0$ for the second player, and without constraints for the first player for simplicity. Then we have the feasible set $W = \left\{ x \in \mathbb{R}^2 \mid x_1 - x_2 \le 0,\ -x_1 - x_2 \le 0 \right\}$, which is illustrated in Figure 5.1 (a). Let $\alpha = 2$. The optimal points of the minimization problems in

$$V_2(x) = x_1^2 + x_2 - \min_{z_1\in\mathbb{R}} \left[z_1^2 + (x_1 - z_1)^2 \right] - \min_{z_2\in[|x_1|,+\infty[} \left[z_2 + (x_2 - z_2)^2 \right]$$

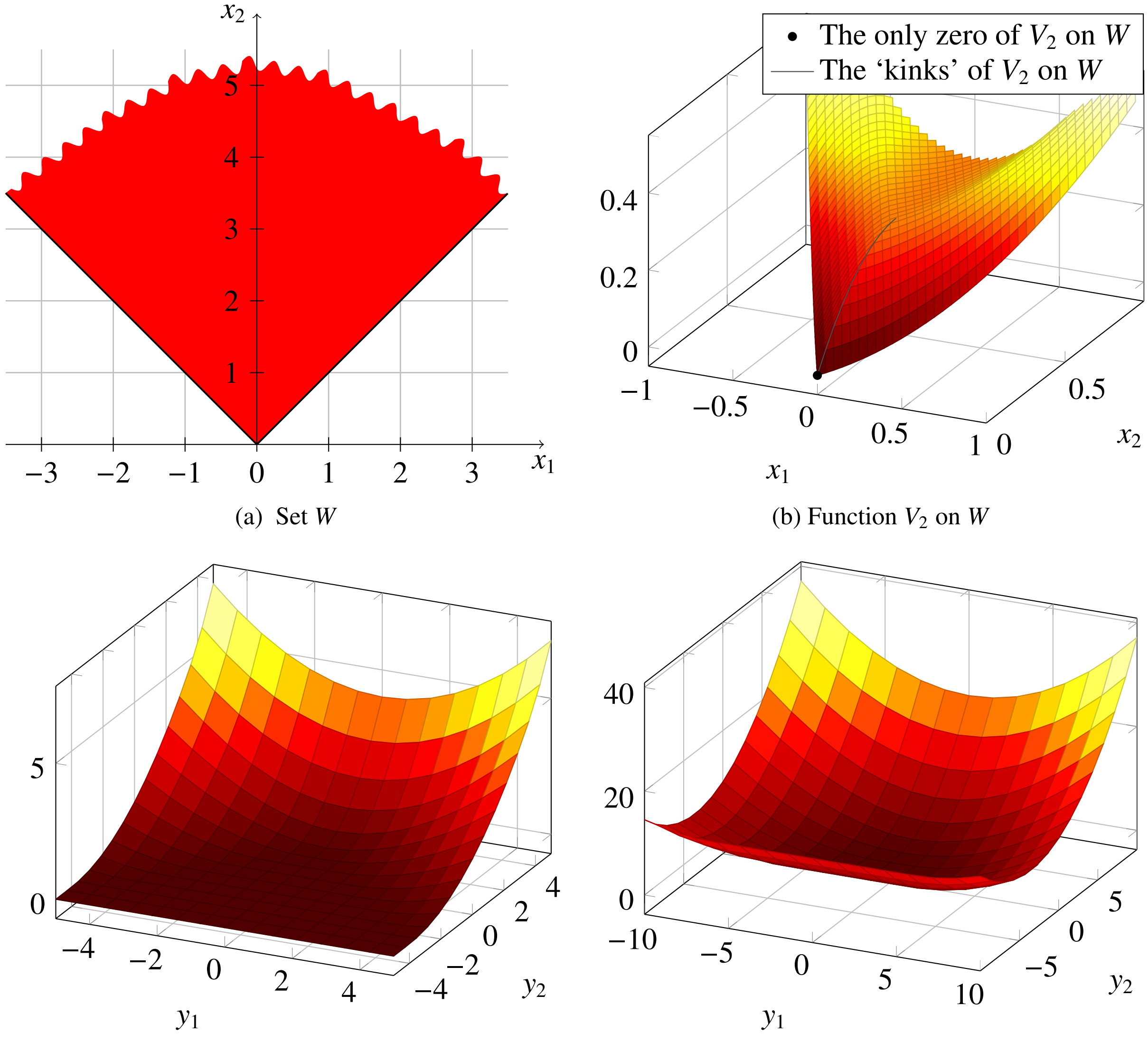
are

$$(z_2(x))_1 = \tfrac{1}{2}x_1 \ \text{ for all } x \in \mathbb{R}^2, \qquad (z_2(x))_2 = \begin{cases} x_2 - \frac{1}{2}, & \text{if } x_2 - \frac{1}{2} \ge |x_1|, \\ |x_1|, & \text{else.} \end{cases}$$

Hence we obtain the optimal value function

$$V_2(x) = \begin{cases} \frac{1}{2}x_1^2 + \frac{1}{4}, & \text{if } x_2 - \frac{1}{2} \ge |x_1|, \\ \frac{1}{2}x_1^2 + x_2 - |x_1| - (x_2 - |x_1|)^2, & \text{else.} \end{cases}$$

The function V_2 is nondifferentiable at the points $x \in \mathbb{R}^2$ with $x_1 = 0$ and $x_2 < \frac{1}{2}$ (see 'kinks' in Figure 5.1 (b)), which, in particular, include the unique solution $\bar{x} = 0$ of the GNEP. This solution



(a) Set W (b) Function V_2 on W

(c) Function f_2^* (d) Function h_2^*

Figure 5.1.: Illustrations for Example 5.10

can be verified using Lemma 3.2 (e). On the other hand, we consider the DC decomposition of V_2 with the functions

$$f_2(x) = 2x_1^2 + x_2^2 + x_2 + \delta_W(x) \quad \text{and} \quad h_2(x) = \frac{3}{2}x_1^2 + x_2^2 + \begin{cases} x_2 - \frac{1}{4}, & \text{if } x_2 - \frac{1}{2} \geq |x_1|, \\ |x_1| + (x_2 - |x_1|)^2, & \text{else.} \end{cases}$$

Using Definition 2.5 (e) yields

$$\begin{aligned} f_2^*(y) &= \sup_{x \in W} \left[x_1 y_1 + x_2 y_2 - 2x_1^2 - x_2^2 - x_2 \right] \\ &= \frac{1}{4}y_1^2 + \frac{1}{4}y_2^2 - \min_{x \in W} \left[x_1^2 + x_2 + \left(x_1 - \frac{1}{2}y_1 \right)^2 + \left(x_2 - \frac{1}{2}y_2 \right)^2 \right] \end{aligned} \tag{5.8}$$

with a strongly convex and differentiable minimization problem in (5.8). This problem has the following Lagrange function:

$$L_2^{f^*}(x,\lambda,y) = x_1^2 + x_2 + \left(x_1 - \frac{1}{2}y_1\right)^2 + \left(x_2 - \frac{1}{2}y_2\right)^2 + \lambda_1(x_1 - x_2) + \lambda_2(-x_1 - x_2).$$

Then the KKT conditions of this problem for $x \in W$ are

$$\nabla_x L_2^{f^*}(x,\lambda,y) = \begin{pmatrix} 4x_1 - y_1 + \lambda_1 - \lambda_2 \\ 1 + 2x_2 - y_2 - \lambda_1 - \lambda_2 \end{pmatrix} = 0 \quad \text{and} \quad \begin{cases} \lambda_1 \geq 0,\ \lambda_1(x_1 - x_2) = 0, \\ \lambda_2 \geq 0,\ \lambda_2(-x_1 - x_2) = 0. \end{cases}$$

There are four possibilities for a solution of the minimization problem in (5.8): to be unconstrained, to be constrained by exactly one out of two constraint, or to be constrained by both constraints:

Case 1: $\begin{cases} x_1 - x_2 < 0, \\ -x_1 - x_2 < 0, \end{cases}$ Case 2: $\begin{cases} x_1 - x_2 = 0, \\ -x_1 - x_2 < 0, \end{cases}$ Case 3: $\begin{cases} x_1 - x_2 < 0, \\ -x_1 - x_2 = 0, \end{cases}$ Case 4: $\begin{cases} x_1 - x_2 = 0, \\ -x_1 - x_2 = 0. \end{cases}$

Using the KKT conditions of the problem in (5.8) for $x \in W$, we obtain

- the first case for the optimal point $x_2^{f^*}(y) = \frac{1}{4}(y_1, 2y_2 - 2)$ if $y_2 > 1 + \frac{1}{2}|y_1|$ holds,
- the second case for $x_2^{f^*}(y) = \frac{1}{6}(y_1 + y_2 - 1, y_1 + y_2 - 1)$ if $1 - y_1 < y_2 \leq 1 + \frac{1}{2}y_1$ holds,
- the third case for $x_2^{f^*}(y) = \frac{1}{6}(y_1 - y_2 + 1, -y_1 + y_2 - 1)$ if $1 + y_1 < y_2 \leq 1 - \frac{1}{2}y_1$ holds,
- the fourth case for $x_2^{f^*}(y) = (0,0)$ if $y_2 \leq 1 - |y_1|$ holds.

The second and third case can be combined such that the function f_2 has the following conjugate

$$f_2^*(y) = \begin{cases} \frac{1}{8}\left(y_1^2 + 2(y_2 - 1)^2\right), & \text{if } y_2 > 1 + \frac{1}{2}|y_1|, \\ \frac{1}{12}(1 - |y_1| - y_2)^2, & \text{if } 1 - |y_1| < y_2 \leq 1 + \frac{1}{2}|y_1|, \\ 0, & \text{if } y_2 \leq 1 - |y_1|, \end{cases}$$

which is illustrated in Figure 5.1 (c).

In a similar way, we observe that

$$\begin{aligned} h_2^*(y) &= \sup_{(x,z)\in \operatorname{gph}\Omega} \left[x_1y_1 + x_2y_2 - x_1^2 - x_2^2 - z_1^2 - z_2 - (x_1 - z_1)^2 - (x_2 - z_2)^2\right] \\ &= \frac{1}{4}\|y\|^2 - \min_{(x,z)\in \operatorname{gph}\Omega} \left[\left(x_1 - \frac{1}{2}y_1\right)^2 + \left(x_2 - \frac{1}{2}y_2\right)^2 + z_1^2 + z_2 + (x_1 - z_1)^2 + (x_2 - z_2)^2\right], \end{aligned} \tag{5.9}$$

where in this example

$$\operatorname{gph}\Omega = \left\{(x,z) \in \mathbb{R}^2 \times \mathbb{R}^2 \;\middle|\; x_1 - z_2 \leq 0,\ -x_1 - z_2 \leq 0\right\}.$$

The Lagrange function of the strongly convex and differentiable minimization problem in (5.9) is given by

$$L_2^{h^*}(x,z,\lambda,y) = \left(x_1 - \frac{1}{2}y_1\right)^2 + \left(x_2 - \frac{1}{2}y_2\right)^2 + z_1^2 + z_2 + (x_1 - z_1)^2 + (x_2 - z_2)^2 + \\ + \lambda_1(x_1 - z_2) + \lambda_2(-x_1 - z_2).$$

Then the corresponding KKT conditions for $(x,z) \in \operatorname{gph}\Omega$ are

$$\nabla_{(x,z)} L_2^{h^*}(x,z,\lambda,y) = \begin{pmatrix} 4x_1 - y_1 - 2z_1 + \lambda_1 - \lambda_2 \\ 4x_2 - y_2 - 2z_2 \\ 4z_1 - 2x_1 \\ 1 + 2z_2 - 2x_2 - \lambda_1 - \lambda_2 \end{pmatrix} = 0 \quad \text{and} \quad \begin{cases} \lambda_1 \geq 0, \ \lambda_1(x_1 - z_2) = 0, \\ \lambda_2 \geq 0, \ \lambda_2(-x_1 - z_2) = 0. \end{cases}$$

Again we obtain four possibilities for a solution of the minimization problem in (5.9) depending on the following cases:

Case 1:	Case 2:	Case 3:	Case 4:
$\begin{cases} x_1 - z_2 < 0, \\ -x_1 - z_2 < 0, \end{cases}$	$\begin{cases} x_1 - z_2 = 0, \\ -x_1 - z_2 < 0, \end{cases}$	$\begin{cases} x_1 - z_2 < 0, \\ -x_1 - z_2 = 0, \end{cases}$	$\begin{cases} x_1 - z_2 = 0, \\ -x_1 - z_2 = 0. \end{cases}$

Using the KKT conditions of the problem in (5.9) for $(x,z) \in \operatorname{gph}\Omega$, we observe

- the first case for the optimal points $x_2^{h^*}(y) = \frac{1}{6}(2y_1, 3y_2 - 3)$ and $z_2^{h^*}(y) = \frac{1}{6}(y_1, 3y_2 - 6)$ under the condition $y_2 > 2 + \frac{2}{3}|y_1|$,
- the second case for the optimal point $x_2^{h^*}(y) = \frac{1}{16}(4y_1 + 2y_2 - 4, 2y_1 + 5y_2 - 2)$ together with $z_2^{h^*}(y) = \frac{1}{16}(2y_1 + y_2 - 2, 4y_1 + 2y_2 - 4)$ under the conditions $2 - 2y_1 < y_2 \leq 2 + \frac{2}{3}y_1$,
- the third case for the optimal point $x_2^{h^*}(y) = \frac{1}{16}(4y_1 - 2y_2 + 4, -2y_1 + 5y_2 - 2)$ together with $z_2^{h^*}(y) = \frac{1}{16}(2y_1 - y_2 + 2, -4y_1 + 2y_2 - 4)$ under the conditions $2 + 2y_1 < y_2 \leq 2 - \frac{2}{3}y_1$,
- the fourth case for $x_2^{h^*}(y) = \left(0, \frac{1}{4}y_2\right)$ and $z_2^{h^*}(y) = (0,0)$ under the condition $y_2 \leq 2 - 2|y_1|$.

Again the second and third case can be combined such that the function h_2 has the following conjugate:

$$h_2^*(y) = \begin{cases} \frac{1}{12}\left(2y_1^2 + 3(y_2 - 1)^2 + 3\right), & \text{if } y_2 > 2 + \frac{2}{3}|y_1|, \\ \frac{1}{32}\left((2 - 2|y_1| - y_2)^2 + 4y_2^2\right), & \text{if } 2 - 2|y_1| < y_2 \leq 2 + \frac{2}{3}|y_1|, \\ \frac{1}{8}y_2^2, & \text{if } y_2 \leq 2 - 2|y_1|, \end{cases}$$

see Figure 5.1 (d). The continuous differentiability of both functions f_2^* and h_2^* can be shown by simple calculations or, alternatively, follows directly from Lemmata 5.8 and 5.9, respectively. ◇

Finally, by applying the duality theory by Toland and Singer [131, Theorem 2.2], we obtain the next theorem, which is the main result of this section. An essential finding of this duality theory is the following statement: It holds that

$$\inf_{x\in\mathbb{R}^n} [f(x) - h(x)] = \inf_{y\in\mathbb{R}^n} [h^*(y) - f^*(y)] \tag{5.10}$$

for all functions $f, h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with h convex and lsc and without conditions on f.

Theorem 5.11 *Let Assumption 5.1 hold, and define the* dual gap function

$$d_\alpha^* := h_\alpha^* - f_\alpha^*$$

with the functions f_α^* and h_α^* given by Lemmata 5.8 and 5.9, respectively. Then the following statements hold:

(a) The function d_α^* is differentiable with Lipschitz gradient.

(b) If $\bar{y}$ is a solution of the unconstrained minimization problem

$$\min_{y\in\mathbb{R}^n} d_\alpha^*(y) \tag{5.11}$$

with $d_\alpha^*(\bar{y}) = 0$, then $\bar{x} := \nabla f_\alpha^*(\bar{y})$ is a solution of the GNEP.

(c) Conversely, if $\bar{x}$ is a solution of the GNEP and $\bar{y} \in \partial h_\alpha(\bar{x})$, then $\bar{y}$ is a solution of (5.11) *with $d_\alpha^*(\bar{y}) = 0$.*

Proof. This result follows directly from the duality theory by Toland [130, 131] and Singer [123]. Nevertheless, for the sake of completeness, we elaborate on all details in this proof.

(a) The definition of the dual gap function d_α^* and Lemmata 5.8 (c) and 5.9 (c) imply the continuous differentiability of d_α^*.

(b) Let $\bar{y}$ be a solution of the unconstrained and differentiable minimization problem (5.11) with

$$0 = d_\alpha^*(\bar{y}) = h_\alpha^*(\bar{y}) - f_\alpha^*(\bar{y}). \tag{5.12}$$

Then we obtain $\nabla d_\alpha^*(\bar{y}) = 0$. This implies that $\nabla f_\alpha^*(\bar{y}) = \nabla h_\alpha^*(\bar{y})$ holds. Therefore, the two subdifferentials $\partial f_\alpha^*(\bar{y})$ and $\partial h_\alpha^*(\bar{y})$ also coinside, since the functions f_α^* and h_α^* are convex and differentiable and hence $\partial f_\alpha^*(\bar{y}) = \{\nabla f_\alpha^*(\bar{y})\}$ and $\partial h_\alpha^*(\bar{y}) = \{\nabla h_\alpha^*(\bar{y})\}$ hold. Furthermore, Lemma 2.6 (d) and Lemma 5.7 (a) and (b) show that $f_\alpha^{**} = f_\alpha$ and $h_\alpha^{**} = h_\alpha$. Hence defining $\bar{x} := \nabla f_\alpha^*(\bar{y})$ and applying Lemma 2.6 (f) leads to

$$f_\alpha(\bar{x}) + f_\alpha^*(\bar{y}) = \bar{x}^T\bar{y} \quad \text{and} \quad h_\alpha(\bar{x}) + h_\alpha^*(\bar{y}) = \bar{x}^T\bar{y}. \tag{5.13}$$

Consequently, we obtain

$$f_\alpha(\bar{x}) - h_\alpha(\bar{x}) = h_\alpha^*(\bar{y}) - f_\alpha^*(\bar{y}). \tag{5.14}$$

Due to (5.12), the vector $\bar{x}$ minimizes the nonnegative function $f_\alpha - h_\alpha$ with function value equal to zero. Thus, the vector $\bar{x}$ is a solution of the GNEP in view of Lemma 5.7 (c).

(c) Now, let $\bar{x}$ be a solution of the GNEP. Then Lemma 5.7 (c) implies

$$0 = V_\alpha(\bar{x}) = \min_{x \in \mathbb{R}^n} [f_\alpha(x) - h_\alpha(x)] \tag{5.15}$$

and therefore

$$f_\alpha(\bar{x}) - h_\alpha(\bar{x}) \le f_\alpha(x) - h_\alpha(x) \quad \forall x \in \mathbb{R}^n.$$

In view of $\bar{y} \in \partial h_\alpha(\bar{x})$, we have

$$\bar{y}^T(x - \bar{x}) \le h_\alpha(x) - h_\alpha(\bar{x})$$

for all $x \in \mathbb{R}^n$. Consequently, we conclude

$$\bar{y}^T(x - \bar{x}) \le h_\alpha(x) - h_\alpha(\bar{x}) \le f_\alpha(x) - f_\alpha(\bar{x}).$$

This means that not only the subdifferential $\partial h_\alpha(\bar{x})$ but also the subdifferential $\partial f_\alpha(\bar{x})$ contains the element $\bar{y}$. Together with Lemma 2.6 (f) we obtain the equations (5.13) and (5.14) successively. Due to (5.15), the left-hand side of (5.14) is equal to zero. Using the relation (5.10) yields that the vector $\bar{y}$ minimizes the dual gap function d_α^* with function value equal to zero. □

We recall that, basically, a function f is called a *gap function* for a mathematical program if the function f is nonnegative and a point is a solution of the corresponding mathematical program if and only if the objective function f is zero at this point. Therefore, the name ‘dual gap function’ for the nonnegative function d_α^* is justified only if the corresponding GNEP supplies the function h_α with nonempty subdifferential ∂h_α for all generalized Nash equilibriums of this GNEP. Nevertheless, we still use the name ‘dual gap function’ for the function d_α^* as Dietrich did in [30].

In order to illustrate the results of Theorem 5.11, we go back to Example 5.10.

Example 5.12 Consider the GNEP with the unique solution $\bar{x} = 0$ from Example 5.10. Since

$$h_2(x) = \frac{3}{2}x_1^2 + x_2^2 + \begin{cases} x_2 - \frac{1}{4}, & \text{if } x_2 - \frac{1}{2} \ge |x_1|, \\ x_1^2 + x_2^2 + (1 - 2x_2)|x_1|, & \text{else,} \end{cases}$$

the function h_2 has in $(0, 0)$ the subdifferential $\partial h_2(0, 0) = \left\{ s \in \mathbb{R}^2 \;\middle|\; s_1 \in [-1, 1],\ s_2 = 0 \right\}$. Due to Theorem 5.11 (c), all vectors $\bar{y} \in \partial h_2(0, 0)$ are solutions of the dual minimization problem (5.11) with zero as the optimal value. Simple calculations of global minima of the dual gap function $d_2^* = h_2^* - f_2^*$ confirm this assertion, see Figure 5.2. Furthermore, Theorem 5.11 (b) states that $\bar{x} = \nabla f_2^*(\bar{y}) = 0$ is a solution of the GNEP. This fact was already mentioned in Example 5.10. ◇

Note that the points $\bar{y} \in \mathbb{R}^2$ with $\bar{y}_1 = 0$ and $\bar{y}_2 \ge 2$ are stationary points or local minima of the dual gap function d_2^* in Example 5.12, which are not solutions of the corresponding GNEP, see Figure 5.2. This example, which has fairly nice properties, points to the fact that it might be difficult to find sufficient conditions for optimality of stationary points. The following proposition is only a partial result in this direction.

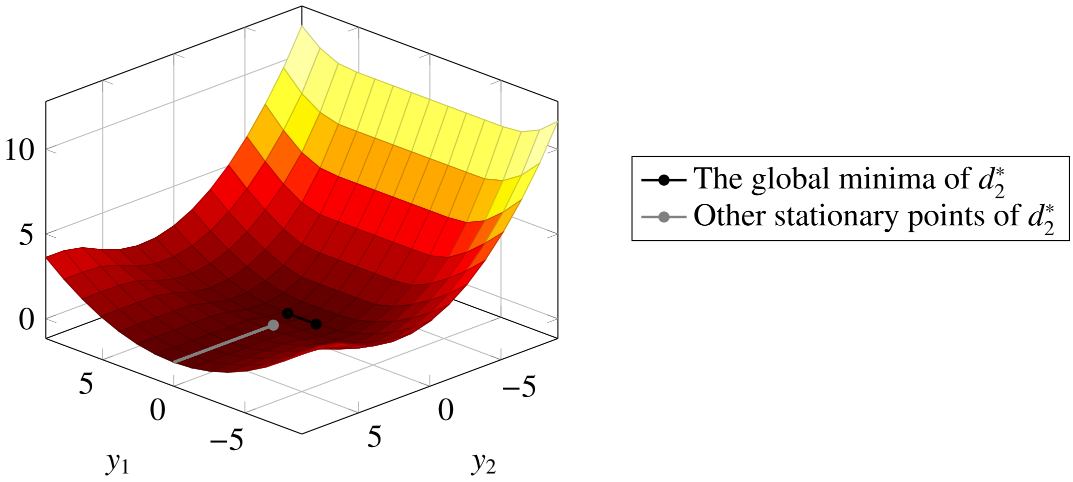


Figure 5.2.: Dual gap function d_2^* from Example 5.12

Proposition 5.13 *Let Assumption 5.1 hold, let $d_\alpha^* = h_\alpha^* - f_\alpha^*$ be the dual gap function, and let $x_\alpha^{f^*}(y)$ and $x_\alpha^{h^*}(y)$, $z_\alpha^{h^*}(y)$ denote the vectors defined in Lemmata 5.8 and 5.9, respectively. Then the following statements are equivalent:*

(a) $x_\alpha^{f^*}(\bar{y}) = x_\alpha^{h^*}(\bar{y}) = z_\alpha^{h^*}(\bar{y})$*;*

(b) $d_\alpha^*(\bar{y}) = 0$.

Proof. Assume that $x_\alpha^{f^*}(\bar{y}) = x_\alpha^{h^*}(\bar{y}) = z_\alpha^{h^*}(\bar{y}) =: \bar{x}$ holds. Using Lemma 5.8 (a) and Lemma 5.9 (a) leads to

$$f_\alpha^*(\bar{y}) = \bar{x}^T \bar{y} - \frac{\alpha}{2}\|\bar{x}\|^2 - \sum_{\nu=1}^N \theta_\nu(\bar{x})$$

and

$$h_\alpha^*(\bar{y}) = \bar{x}^T \bar{y} - \frac{\alpha}{2}\|\bar{x}\|^2 - \sum_{\nu=1}^N \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) - \frac{\alpha}{2}\|\bar{x} - \bar{x}\|^2 = \bar{x}^T \bar{y} - \frac{\alpha}{2}\|\bar{x}\|^2 - \sum_{\nu=1}^N \theta_\nu(\bar{x}).$$

Therefore, we have $d_\alpha^*(\bar{y}) = h_\alpha^*(\bar{y}) - f_\alpha^*(\bar{y}) = 0$.

Conversely, assume that $d_\alpha^*(\bar{y}) = 0$ holds. Applying Theorem 5.11 yields that $\bar{x} := \nabla f_\alpha^*(\bar{y})$ is a solution of the GNEP and that $\bar{y}$ is a global minimum of d_α^* on $\mathbb{R}^n$. The second part implies $\nabla d_\alpha^*(\bar{y}) = 0$. Apart from that, we have

$$\nabla d_\alpha^*(\bar{y}) = \nabla h_\alpha^*(\bar{y}) - \nabla f_\alpha^*(\bar{y}) = x_\alpha^{h^*}(\bar{y}) - x_\alpha^{f^*}(\bar{y})$$

in view of Lemma 5.8, Lemma 5.9, and the definition of d_α^*. These all add up to

$$\bar{x} = x_\alpha^{f^*}(\bar{y}) = x_\alpha^{h^*}(\bar{y}). \tag{5.16}$$

Additionally, it follows from (3.9) that

$$V_\alpha(\bar{x}) = \sum_{\nu=1}^N \theta_\nu(\bar{x}) - \sum_{\nu=1}^N \theta_\nu(z_\alpha(\bar{x})^\nu, \bar{x}^{-\nu}) - \frac{\alpha}{2}\left\|\bar{x} - z_\alpha(\bar{x})\right\|^2$$

with the uniquely defined minimum

$$z_\alpha(\bar{x}) = \underset{z\in\Omega(\bar{x})}{\operatorname{argmin}} \left(\sum_{\nu=1}^{N} \theta_\nu(z^\nu, \bar{x}^{-\nu}) + \frac{\alpha}{2}\|\bar{x} - z\|^2 \right).$$

Since $\bar{x}$ is a solution of the GNEP, we obtain with help of Lemma 3.2 (c) and Lemma 3.2 (e) that $\bar{x} \in W$ and $V_\alpha(\bar{x}) = 0$ as well as $z_\alpha(\bar{x}) = \bar{x}$. Furthermore, due to the representation (5.6) of the function $h^*_\alpha(\bar{y})$, it holds that $z_\alpha(\bar{x})$ is identical to $z^{h^*}_\alpha(\bar{y})$. Thus, we also have $z^{h^*}_\alpha(\bar{y}) = \bar{x}$. Together with (5.16) this completes the proof. □

Proposition 5.13 states that $x^{f^*}_\alpha(\bar{y}) = x^{h^*}_\alpha(\bar{y}) = z^{h^*}_\alpha(\bar{y}) =: \bar{x}$ implies $d^*_\alpha(\bar{y}) = 0$ and, consequently, that $\bar{x}$ is a solution of the GNEP. Since it is not difficult to see that $x^{f^*}_\alpha(\bar{y}) = x^{h^*}_\alpha(\bar{y})$ holds at any stationary point of d^*_α, it remains to provide conditions under which these two vectors are equal to $z^{h^*}_\alpha(\bar{y})$. However, we have to leave this question open and therefore also the question in which cases stationary points of the dual gap function d^*_α provide solutions of a GNEP. On the other hand, we know the optimal value of d^*_α, so this disadvantage might not be that strong, since the function value itself tells us whether we are in a solution or not. Note that, in Example 5.12, we have $x^{f^*}_\alpha(\bar{y}) = x^{h^*}_\alpha(\bar{y}) \neq z^{h^*}_\alpha(\bar{y})$ for all stationary points $\bar{y} \in \mathbb{R}^2$ with $\bar{y}_1 = 0$ and $\bar{y}_2 \geq 2$ as well as $d^*_\alpha(\bar{y}) = \frac{1}{4} \neq 0$. This function value alone shows us that none of these stationary points provides a solution of the corresponding GNEP.

Theorem 5.11 treats the relation between the solutions of the GNEP and the global minima of the dual gap function d^*_α. More precisely, it shows that every solution of the unconstrained optimization problem (5.11) provides a solution of the GNEP, but the converse is not necessarily true, because statement (c) of Theorem 5.11 assumes (implicitly) that the subdifferential $\partial h_\alpha(\bar{x})$ is nonempty at a solution $\bar{x}$ of the GNEP. In fact, this subdifferential could be empty, and a global minimum of the function d^*_α could be nonexistent although the corresponding GNEP is solvable. The next example illustrates this assertion.

Example 5.14 Consider a GNEP satisfying Assumption 5.1 with $N = 2$, $n_1 = n_2 = 1$, the variables x_1 and x_2 controlled by player 1 and 2, respectively, the cost functions $\theta_1(x) := (x_1 - 1)^2$, $\theta_2(x) := (x_2 + 4)^2$ and the constraint $g^2_1(x) := x_1^2 + x_2^2 - 1 \leq 0$ for the second player and without constraints for the first player for simplicity. Then the feasible set W is given by

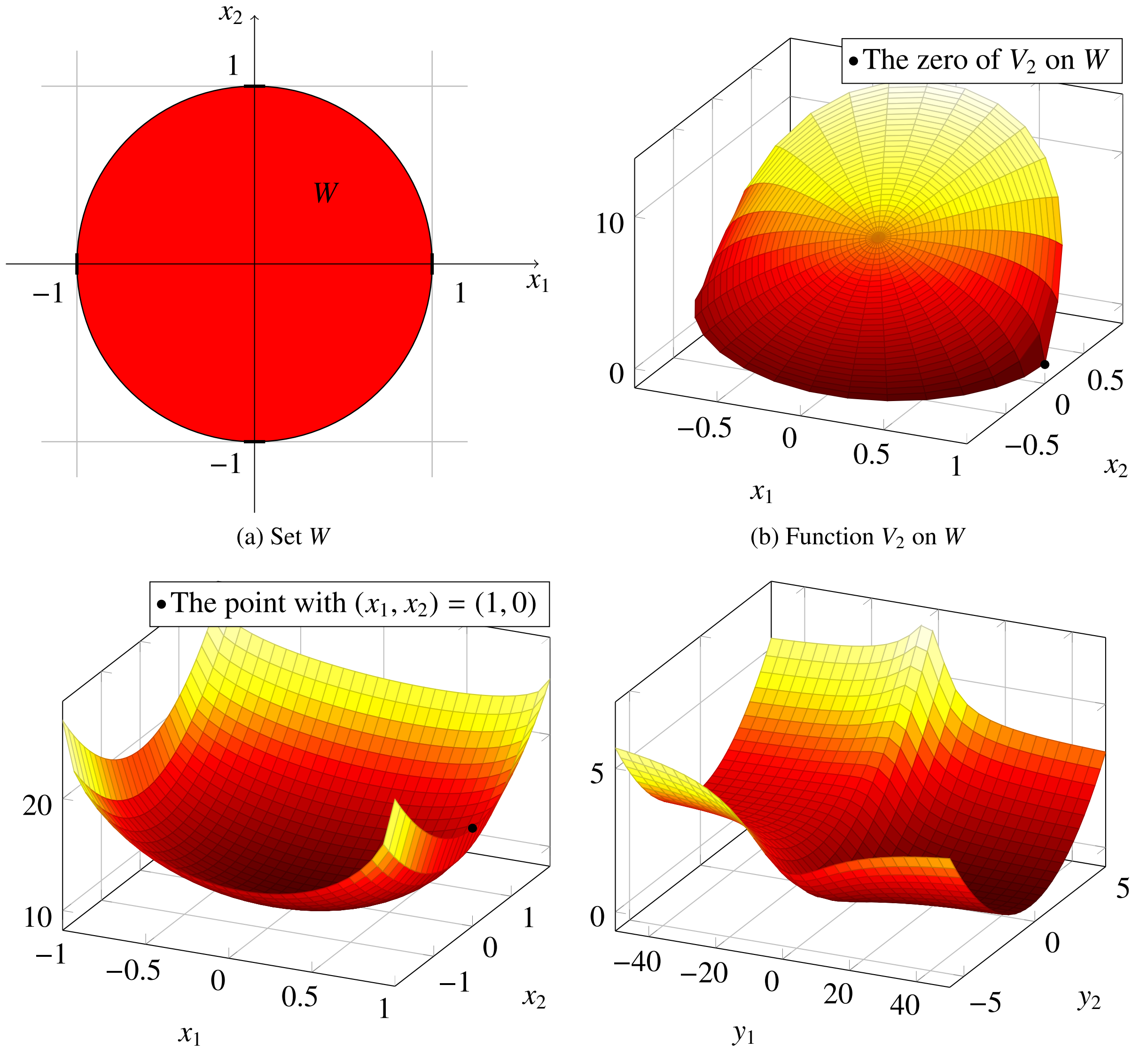
$$W = \left\{ x \in \mathbb{R}^2 \;\middle|\; x_1^2 + x_2^2 \leq 1 \right\},$$

which is illustrated in Figure 5.3 (a), and the graph of the set-valued mapping Ω is given by

$$\operatorname{gph}\Omega = \left\{ (x, z) \in \mathbb{R}^2 \times \mathbb{R}^2 \;\middle|\; x_1^2 + z_2^2 \leq 1 \right\}.$$

Let $\alpha = 2$, then we have

$$\begin{aligned} V_2(x) \;=\;& (x_1 - 1)^2 + (x_2 + 4)^2 - \min_{z_1\in\mathbb{R}} \left[(z_1 - 1)^2 + (x_1 - z_1)^2 \right] + \\ & - \inf_{z_2\in\left[-\sqrt{1-x_1^2},\, \sqrt{1-x_1^2}\right]} \left[(z_2 + 4)^2 + (x_2 - z_2)^2 \right]. \end{aligned}$$



(a) Set W (b) Function V_2 on W

(c) Function h_2 (d) Dual gap function d_2^*

Figure 5.3.: Illustrations for Example 5.14

The optimal points of the minimization problems in V_2 are

$$(z_2(x))_1 = \frac{1}{2}x_1 + \frac{1}{2} \quad \text{for all } x \in \mathbb{R}^2, \quad (z_2(x))_2 = \begin{cases} \frac{1}{2}x_2 - 2, & \text{if } x_1^2 + \frac{1}{4}(x_2 - 4)^2 < 1, \\ -\sqrt{1 - x_1^2}, & \text{if } \frac{1}{2}x_2 - 2 \leq -\sqrt{1 - x_1^2}, \\ \sqrt{1 - x_1^2}, & \text{if } \frac{1}{2}x_2 - 2 \geq \sqrt{1 - x_1^2}, \\ \text{nonexistent}, & \text{if } |x_1| > 1. \end{cases}$$

Thus, we obtain

$$V_2(x) = \frac{1}{2}(x_1-1)^2 + (x_2+4)^2 - \begin{cases} \frac{1}{2}(x_2+4)^2, & \text{if } x_1^2 + \frac{1}{4}(x_2-4)^2 < 1, \\ \left(4 - \sqrt{1-x_1^2}\right)^2 + \left(x_2 + \sqrt{1-x_1^2}\right)^2, & \text{if } \frac{1}{2}x_2 - 2 \le -\sqrt{1-x_1^2}, \\ \left(4 + \sqrt{1-x_1^2}\right)^2 + \left(x_2 - \sqrt{1-x_1^2}\right)^2, & \text{if } \frac{1}{2}x_2 - 2 \ge \sqrt{1-x_1^2}, \\ \infty, & \text{if } |x_1| > 1. \end{cases}$$

Note that the inequality $\frac{1}{2}x_2 - 2 \le -\sqrt{1-x_1^2}$ holds for all $x \in W$. The graph of the optimal value function V_2 on the set W is illustrated in Figure 5.3 (b). On the other hand, we consider the DC decomposition of V_2 with the functions

$$f_2(x) = x_1^2 + x_2^2 + (x_1-1)^2 + (x_2+4)^2 + \delta_W(x)$$

and

$$h_2(x) = x_1^2 + x_2^2 + \frac{1}{2}(x_1-1)^2 + \begin{cases} \frac{1}{2}(x_2+4)^2, & \text{if } x_1^2 + \frac{1}{4}(x_2-4)^2 < 1, \\ \left(4 - \sqrt{1-x_1^2}\right)^2 + \left(x_2 + \sqrt{1-x_1^2}\right)^2, & \text{if } \frac{1}{2}x_2 - 2 \le -\sqrt{1-x_1^2}, \\ \left(4 + \sqrt{1-x_1^2}\right)^2 + \left(x_2 - \sqrt{1-x_1^2}\right)^2, & \text{if } \frac{1}{2}x_2 - 2 \ge \sqrt{1-x_1^2}, \\ \infty, & \text{if } |x_1| > 1. \end{cases}$$

Using Definition 2.5 (e) yields

$$\begin{aligned} f_2^*(y) &= \sup_{x\in W} \left[x_1y_1 + x_2y_2 - x_1^2 - x_2^2 - (x_1-1)^2 - (x_2+4)^2\right] \\ &= \frac{1}{4}y_1^2 + \frac{1}{4}y_2^2 - \min_{x\in W}\left[(x_1-1)^2 + (x_2+4)^2 + \left(x_1 - \frac{1}{2}y_1\right)^2 + \left(x_2 - \frac{1}{2}y_2\right)^2\right] \end{aligned} \tag{5.17}$$

with a strongly convex and differentiable minimization problem in (5.17). This problem has the following Lagrange function:

$$L_2^{f^*}(x,\lambda,y) = (x_1-1)^2 + (x_2+4)^2 + \left(x_1 - \frac{1}{2}y_1\right)^2 + \left(x_2 - \frac{1}{2}y_2\right)^2 + \lambda(x_1^2 + x_2^2 - 1).$$

Then the KKT conditions of this problem are

$$\begin{aligned} \nabla_x L_2^{f^*}(x,\lambda,y) = \begin{pmatrix} 4x_1 - 2 - y_1 + 2\lambda x_1 \\ 4x_2 + 8 - y_2 + 2\lambda x_2 \end{pmatrix} = 0, \\ \lambda \ge 0,\ x_1^2 + x_2^2 - 1 \le 0,\ \lambda(x_1^2 + x_2^2 - 1) = 0. \end{aligned}$$

There are two possibilities for a solution of the minimization problem in (5.17): to be unconstrained ($x_1^2 + x_2^2 - 1 < 0$) or to be constrained ($x_1^2 + x_2^2 - 1 = 0$). With the help of the KKT conditions, we obtain that the first case holds for the optimal point $x_2^{f^*}(y) = \frac{1}{4}(y_1 + 2, y_2 - 8)$ under the condition $(y_1 + 2)^2 + (y_2 - 8)^2 < 16$, otherwise the second case holds for

$$x_2^{f^*}(y) = \left(\frac{y_1 + 2}{\sqrt{(y_1 + 2)^2 + (y_2 - 8)^2}}, \frac{y_2 - 8}{\sqrt{(y_1 + 2)^2 + (y_2 - 8)^2}} \right).$$

Then the function f_2 has the following continuously differentiable conjugate

$$f_2^*(y) = \begin{cases} \frac{1}{8}((y_1 + 2)^2 + (y_2 - 8)^2) - 17, & \text{if } (y_1 + 2)^2 + (y_2 - 8)^2 < 16, \\ \sqrt{(y_1 + 2)^2 + (y_2 - 8)^2} - 19, & \text{else.} \end{cases}.$$

In a similar way, we observe

$$\begin{aligned} h_2^*(y) &= \sup_{(x,z)\in \operatorname{gph}\Omega} \left[x_1 y_1 + x_2 y_2 - x_1^2 - x_2^2 - (z_1 - 1)^2 - (z_2 + 4)^2 - (x_1 - z_1)^2 - (x_2 - z_2)^2 \right] \\ &= \frac{1}{4}\|y\|^2 - \min_{(x,z)\in \operatorname{gph}\Omega} \left[\left(x_1 - \frac{1}{2}y_1\right)^2 + \left(x_2 - \frac{1}{2}y_2\right)^2 + (z_1 - 1)^2 + (z_2 + 4)^2 + (x_1 - z_1)^2 + (x_2 - z_2)^2 \right]. \end{aligned}$$

The Lagrange function of the strongly convex and differentiable minimization problem for the calculation of $h_2^*(y)$ is given by

$$\begin{aligned} L_2^{h^*}(x, z, \lambda, y) = &\left(x_1 - \frac{1}{2}y_1\right)^2 + \left(x_2 - \frac{1}{2}y_2\right)^2 + (z_1 - 1)^2 + (z_2 + 4)^2 + (x_1 - z_1)^2 + (x_2 - z_2)^2 + \\ &+ \lambda(x_1^2 + z_2^2 - 1). \end{aligned}$$

Then the corresponding KKT conditions are

$$\nabla_{(x,z)} L_2^{h^*}(x, z, \lambda, y) = \begin{pmatrix} 4x_1 - y_1 - 2z_1 + 2\lambda x_1 \\ 4x_2 - y_2 - 2z_2 \\ 4z_1 - 2 - 2x_1 \\ 4z_2 + 8 - 2x_2 + 2\lambda z_2 \end{pmatrix} = 0,$$

$$\lambda \geq 0, \ x_1^2 + z_2^2 - 1 \leq 0, \ \lambda(x_1^2 + z_2^2 - 1) = 0.$$

For a solution of the minimization problem for the calculation of $h_2^*(y)$ we obtain the two possibilities $x_1^2 + z_2^2 - 1 < 0$ and $x_1^2 + z_2^2 - 1 = 0$. Again with the help of the KKT conditions, we observe the first case for the optimal points $x_2^{h^*}(y) = \frac{1}{3}(y_1 + 1, y_2 - 4)$ and $z_2^{h^*}(y) = \frac{1}{6}(y_1 + 4, y_2 - 16)$ if the condition $(y_1 + 1)^2 + \frac{1}{4}(y_2 - 16)^2 < 9$ holds and otherwise the second case for

$$x_2^{h^*}(y) = \left(\frac{y_1 + 1}{\sqrt{(y_1 + 1)^2 + \frac{1}{4}(y_2 - 16)^2}}, \frac{1}{4}y_2 + \frac{y_2 - 16}{4\sqrt{(y_1 + 1)^2 + \frac{1}{4}(y_2 - 16)^2}} \right)$$

and

$$z_2^{h^*}(y) = \left(\frac{1}{2} + \frac{y_1 + 1}{2\sqrt{(y_1 + 1)^2 + \frac{1}{4}(y_2 - 16)^2}}, \frac{y_2 - 16}{2\sqrt{(y_1 + 1)^2 + \frac{1}{4}(y_2 - 16)^2}} \right).$$

Then the function h_2 has the following continuously differentiable conjugate:

$$h_2^*(y) = \begin{cases} \frac{1}{6}((y_1+1)^2 + (y_2-4)^2 - 51), & \text{if } (y_1-1)^2 + \frac{1}{4}(y_2-16)^2 < 9, \\ \frac{1}{2}\sqrt{4(y_1-1)^2 + (y_2-16)^2} - 18 + \frac{1}{8}y_2^2, & \text{else.} \end{cases}$$

Using Lemma 3.2 (e), we obtain that the GNEP has the unique solution $(\bar{x}_1, \bar{x}_2) = (1, 0)$. At this point, the function h_2, which is illustrated in Figure 5.3 (c), has 'infinite slope', and $\partial h_2(\bar{x}) = \emptyset$. Therefore, Theorem 5.11 is not applicable to determine a solution of the corresponding dual problem (5.11). Furthermore, the function $d_2^* = h_2^* - f_2^*$ is positive on $\mathbb{R}^2$, and it holds that $\lim_{y_1 \to \infty} d_2^*(y_1, 0) = 0$, see Figure 5.3 (d). Thus, the dual problem (5.11) has the infimum zero, but does not attain its infimum, hence it has no solution. ◇

Based on the approaches of this section and especially on Theorem 5.11, we will present some numerical results in Section 6.2.

5.2. Second-Order Properties

In this section, we analyze second-order properties of the dual gap function d_α^*. More precisely, we show piecewise smoothness of the gradient mapping ∇d_α^* under certain conditions on the strategy spaces and the cost functions of GNEPs. This piecewise smoothness result follows from our parametric optimization result stated in Theorem 2.14. To this end, we use the following assumptions.

Assumption 5.15 *(a) The set-valued mappings $X_\nu : \mathbb{R}^{n-n_\nu} \to \mathbb{R}^{n_\nu}$ are given by*

$$X_\nu(x^{-\nu}) := \{x^\nu \in \mathbb{R}^{n_\nu} \mid g_i^\nu(x^\nu, x^{-\nu}) \le 0 \quad \forall i = 1, \dots, m_\nu\} \quad (\nu = 1, \dots, N) \tag{5.18}$$

with convex and twice continuously differentiable functions $g_i^\nu : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m_\nu$, $\nu = 1, \dots, N$.

(b) The feasible set

$$W = \{x \in \mathbb{R}^n \mid g_i^\nu(x^\nu, x^{-\nu}) \le 0 \quad \forall i = 1, \dots, m_\nu, \ \nu = 1, \dots, N\}$$

of the GNEP (3.2) *is nonempty.*

(c) The cost functions θ_ν, $\nu = 1, \dots, N$, are convex and twice continuously differentiable.

Then, in particular, we observe

$$\operatorname{gph} \Omega = \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^n \mid g_i^\nu(z^\nu, x^{-\nu}) \le 0 \quad \forall i = 1, \dots, m_\nu, \ \nu = 1, \dots, N\}.$$

We start our analysis by showing that the gradient of the conjugate function f_α^* is piecewise smooth under a suitable CRCQ assumption, see Definition 2.13.

Lemma 5.16 *Let Assumption 5.15 hold, and let $\bar{y} \in \mathbb{R}^n$ such that CRCQ holds at $\bar{x} := x_\alpha^{f^*}(\bar{y})$ with respect to the feasible set W. Then there exists a neighborhood U of $\bar{y}$ such that $\nabla f_\alpha^* = x_\alpha^{f^*}$ is piecewise smooth on U.*

Proof. We define $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$\phi(x, y) := \frac{\alpha}{2}\|x\|^2 + \sum_{\nu=1}^{N} \theta_\nu(x) - x^T y.$$

Then ϕ is twice continuously differentiable and $\phi(\cdot, y)$ strongly convex in x. Furthermore, we define the function $c : \mathbb{R}^n \to \mathbb{R}^p$ with $p := \sum_{\nu=1}^N m_\nu$ by

$$c(x) := (g_i^\nu(x))_{(i=1,\dots,m_\nu,\ \nu=1,\dots,N)}.$$

Then each function c_j $(j = 1, \dots, p)$ is convex and twice continuously differentiable by Assumption 5.15 (a), and the assertion follows from Theorem 2.14 with $U := \bar{V}$. □

We obtain a similar result for the gradient of the conjugate function h_α^*.

Lemma 5.17 *Let Assumption 5.15 hold, and let $\bar{y} \in \mathbb{R}^n$ be given such that CRCQ holds at $(\bar{x}, \bar{z}) := (x_\alpha^{h^*}(\bar{y}), z_\alpha^{h^*}(\bar{y}))$ with respect to* gph Ω*. Then there exists a neighborhood U of $\bar{y}$ such that $\nabla h_\alpha^* = x_\alpha^{h^*}$ is piecewise smooth on U.*

Proof. The assertion follows from Theorem 2.14 similar to the proof of Lemma 5.16. □

The following theorem is the main result of this section and an immediate consequence of the two foregoing lemmata.

Theorem 5.18 *Let the assumptions of Lemmata 5.16 and 5.17 hold at $\bar{y} \in \mathbb{R}^n$. Then the function ∇d_α^* is PC^1 near $\bar{y}$.*

Proof. The proof follows immediately from Lemma 5.16 and 5.17 together with the fact that $\nabla d_\alpha^* = \nabla h_\alpha^* - \nabla f_\alpha^*$. □

We conclude this section with a simple corollary, which covers the case where the constraint functions from (5.18) are affine-linear and hence CRCQ is satisfied automatically.

Corollary 5.19 *Let Assumption 5.15 hold, and let the functions g_i^ν, $\nu = 1, \dots, N$, $i = 1, \dots, m_\nu$, from* (5.18) *be affine-linear. Then the function ∇d_α^* is PC^1 on $\mathbb{R}^n$.*

6. Numerical Results for Generalized Nash Equilibrium Problems

In this chapter we presents some numerical results for a number of GNEPs based on our approaches in Section 4.2 and Section 5.1. More precisely, in Section 6.1 we adapt a feasible direction-type method to the reformulation of a GNEP as the constrained optimization problem in (3.11) based on the primal gap function V_α defined in (3.9), and in Section 6.2 we apply a global spectral gradient method to the reformulation of a GNEP as the unconstrained optimization problem based on the dual gap function d_α^* defined in Theorem 5.11. Additionally, we compare the two approaches in Section 6.3. The results from Section 6.1 were published in [74], whereas Section 6.2 contains the results prepared for the paper [73].

6.1. Primal Gap Function Approach

In view of Proposition 3.2, we know that the computation of a generalized Nash equilibrium is equivalent to solving the constrained optimization problem

$$P: \qquad \min\ V_\alpha(x) \quad \text{subject to} \quad x \in W$$

with the primal gap function V_α defined in (3.9). This function V_α is, in general, nondifferentiable. However, Section 4.2 indicates that, on the one hand, the set of nondifferentiable points is exceptional and, on the other hand (and more importantly), we may expect differentiability of V_α at any solution of a GNEP. Hence we may view problem P essentially as a smooth optimization problem. Note that the primal gap function V_α may not be defined outside the feasible set W, hence any suitable algorithm applied to problem P should guarantee that all iterates stay feasible. We therefore decided to apply a feasible direction-type method to problem P.

The class of feasible direction methods was introduced by Zoutendijk [135]. A variant is due to Topkins and Veinott [132], which, in turn, is the basis of the method presented by Birge et al. in [20]. The latter method uses a convex quadratic program at each iteration and will be used in order to solve our problem P. We adapt this method to the setting of problem P in the following algorithm.

Algorithm 6.1 *(Feasible direction-type method from [20])*

(S.0) *Choose* $x^0 \in W$, $H_0 \in \mathbb{R}^{n\times n}$ *symmetric positive definite,* $\beta, \sigma \in]0,1[$, $c^0_{\nu,i} > 0$ *for all* $i = 1,\dots,m_\nu$, $\nu = 1,\dots,N$, $c^0_{V_\alpha} > 0$, *and set* $k := 0$.

(S.1) *If a suitable termination criterion holds: STOP.*

(S.2) Compute a solution $(d^k, \delta^k) \in \mathbb{R}^n \times \mathbb{R}$ *of*

$$\begin{aligned} \min \quad & \delta + \frac{1}{2} d^T H_k d \\ \textit{subject to} \quad & \nabla V_\alpha(x^k)^T d \leq c^k_{V_\alpha} \delta, \\ & g_i^\nu(x^k) + \nabla g_i^\nu(x^k)^T d \leq c^k_{\nu,i} \delta \quad \forall i = 1, \dots, m_\nu, \nu = 1, \dots, N. \end{aligned} \tag{6.1}$$

If $(d^k, \delta^k) = (0, 0)$*: STOP. Otherwise go to (S.3).*

(S.3) Compute a stepsize $t_k = \max\{\beta^l \mid l = 0, 1, 2, \dots\}$ *such that the following conditions hold:*

$$V_\alpha(x^k + t_k d^k) \leq V_\alpha(x^k) + \sigma t_k \nabla V_\alpha(x^k)^T d^k$$

and

$$g_i^\nu(x^k + t_k d^k) \leq 0 \quad \forall i = 1, \dots, m_\nu, \ \nu = 1, \dots, N.$$

(S.4) Choose $c^{k+1}_{\nu,i} > 0$ $(i = 1, \dots, m_\nu, \nu = 1, \dots, N)$*,* $c^{k+1}_{V_\alpha} > 0$*,* $H_{k+1} \in \mathbb{R}^{n\times n}$ *symmetric positive definite, set* $x^{k+1} := x^k + t_k d^k$*,* $k \leftarrow k + 1$*, and go to (S.1).*

This algorithm can be fitted to QVIs which have not only inequality constraints defining strategy spaces but also affine-linear equality constraints $h_j^\nu : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \dots, p_\nu$, $\nu = 1, \dots, N$, with $p_\nu \in \mathbb{N}$. For such QVIs the equality constraints

$$\nabla h_j^\nu(x)^T d = 0 \quad \forall j = 1, \dots, p_\nu, \ \nu = 1, \dots, N,$$

have to be added to the inequality constraints of the optimization problem in (6.1). Then for a feasible vector $x^k \in \mathbb{R}^n$, any stepsize $t_k \in \mathbb{R}$ and a solution $d^k \in \mathbb{R}^n$ of the modified optimization problem (6.1), the vector $x^{k+1} := x^k + t_k d^k$ fulfills the affine-linear equality constraints such that these equality constraints do not affect the choice of the stepsize in $(S.3)$.

The main termination criterion used in (S.1) is

$$V_\alpha(x^k) \leq N \cdot \varepsilon \quad \text{with} \quad \varepsilon := 10^{-5}.$$

The factor N in front of ε comes from the fact that V_α is the sum of N terms, see (4.1), and the basic idea is that each term should be less than ε, hence our termination criterion is, in some way, independent of the number of players. Additionally, the parameter ε should not be taken too small, since the feasible direction method used here is not a locally fast convergent method.

The computation of the matrix H_k was done in the following way: We begin with $H_0 := I_n$ and compute H_{k+1} as the BFGS-update of H_k whenever this gives a symmetric positive definite matrix, whereas we simply take $H_{k+1} := I_n$ otherwise. Furthermore, the parameters $c^k_{\nu,i}$, $c^k_{V_\alpha}$ are chosen in the following way: We always use $c^k_{\nu,i} := 1$ for all i, ν, and for all iterations k (including $k = 0$), whereas we take $c^0_{V_\alpha} := 10$ in Step (S.0) and update this parameter in Step (S.4) by $c^{k+1}_{V_\alpha} := 5 \cdot c^k_{V_\alpha}$ whenever $t_k < 1$ had to be chosen in (S.3); otherwise we set $c^{k+1}_{V_\alpha} := c^k_{V_\alpha}$. We also note that for Algorithm 6.1, the stepsize $t_k = 1$ is not necessarily a natural choice; therefore, we also allow a larger stepsize whenever this is possible, that is, when $t_k = 1$ satisfies the criteria

Example	N	n	x^0	k	V_α^{opt}
A1	10	10	$(0.01,\ldots,0.01)$	6	6.4089e-05
			$(0.1,\ldots,0.1)$	2	4.9819e-06
			$(1,\ldots,1)$	2	4.9819e-06
A3	3	7	$(0,\ldots,0)$	14	2.6397e-05
			$(1,\ldots,1)$	14	2.9342e-06
			$(10,\ldots,10)$	29	2.3998e-06
A4	3	7	$(10,\ldots,10)$	62	0.0000e+00
A5	3	7	$(0,\ldots,0)$	13	1.5470e-05
			$(1,\ldots,1)$	29	2.5578e-05
			$(10,\ldots,10)$	55	1.9534e-05
A6	3	7	$(0,\ldots,0)$	92	2.1231e-05
			$(1,\ldots,1)$	33	2.4189e-05
			$(10,\ldots,10)$	89	2.7651e-05
A7	4	20	$(0,\ldots,0)$	77	2.2178e-05
			$(1,\ldots,1)$	65	2.4832e-05
			$(10,\ldots,10)$	153	2.5712e-05
A8	3	3	$(0,0,0)$	15	1.1013e-05
			$(1,1,1)$	13	1.2011e-05
			$(10,10,10)$	13	1.2010e-05
A9a	7	56	$(0,\ldots,0)$	119	5.4673e-05
A9b	7	112	$(0,\ldots,0)$	395	4.1144e-05
A11	2	2	$(0,0)$	6	1.0419e-05
A12	2	2	(2,0)	15	1.4158e-05
A13	3	3	$(0,0,0)$	10	8.1729e-06
A14	10	10	$(0,\ldots,0)$	3	6.2243e-05
A15	3	6	$(0,\ldots,0)$	24	1.0849e-05
A16a	5	5	$(10,\ldots,10)$	22	0.0000e+00
A16b	5	5	$(10,\ldots,10)$	11	0.0000e+00
A16c	5	5	$(10,\ldots,10)$	10	0.0000e+00
A16d	5	5	$(10,\ldots,10)$	13	2.3575e-05
A17	2	3	$(0,0,0)$	23	1.0266e-05
A18	2	12	$(0,\ldots,0)$	44	9.1178e-06
			$(1,\ldots,1)$	57	2.8294e-06
			$(10,\ldots,10)$	57	1.6509e-05
Ex. 4.3	2	2	(1,1)	8	0.0000e+00
			(0.5,0.5)	8	1.7113e-05
			(0.5,0)	6	1.8224e-05
			(0.9,0)	9	1.3397e-05

Table 6.1.: Numerical results for different GNEPs from the collection in [48] and Example 4.3 using the primal gap function V_α

from (S.3), we test $t_k = 1/\beta$ and so on, until one of the conditions is violated for the first time. Finally, the values $\beta = 0.5$ and $\sigma = 10^{-4}$ were chosen for all test runs.

The numerical results obtained in this way are summarized in Table 6.1. All results are based on the choice $\alpha = 0.01$ for the regularization parameter α in the definition of V_α. The test examples called A1–A18 are taken from the collection of the test problems provided in the report-version [48] of the paper [49]. We choose the same starting points as in [48], which, however, do not necessarily belong to the feasible set W. Hence we first project these starting points onto W and then begin our iteration with these projected starting points. In some cases, this projection was already the solution of the underlying GNEP, and we therefore do not present results for these GNEPs; the reader therefore does not find the results for problem A2 for all starting points from [48] and problem A4 for the first and second starting points from [48] in Table 6.1. Furthermore, we note that problem A10 is not included in Table 6.1 and is the only problem considered in this section that contains affine-linear equality constaints; our method fails to solve this example since the stepsize becomes too small during the iteration.

For each test example, Table 6.1 contains the following data: the name of the example, the number of players N, the total number of variables n, the starting point x^0, the number of iterations k needed until convergence, and the final value of the objective function V_α in column V_α^{opt}.

The results from Table 6.1 may be difficult to interpret, however, it can be noted that Algorithm 6.1 solves all test examples reported there (i.e., all test examples with the exception of problem A10), whereas, for example, the penalty method from [48] has two failures on this set, namely on Examples A7 and A8, when using the third starting point. Furthermore, the number of iterations is quite reasonable and typically better than the corresponding number of iterations reported in [39] for an unconstrained optimization reformulation, especially because each function evaluation of the objective function in this unconstrained optimization reformulation comes with the cost of the solution of two maximization problems, whereas in our case we only need to solve one maximization problem in order to compute $V_\alpha(x)$.

Moreover, we would like to draw the attention to the results for Example 4.3. Table 6.1 shows that the sequence computed with our method converges to the unique solution of the GNEP and hence to the global minimum of problem P. This is particularly appealing, since one can verify (similar to Example 4.3 where $\alpha = 0$ was chosen) that the function V_α still has a strict local minimum at $(1, 0)$ with a positive function value $V_\alpha(1, 0) = 1 - \frac{\alpha}{2}$ (at least for all $\alpha < 0.5$), so that this minimum does not correspond to a solution of the GNEP. The sequence computed with our method even converges to the global minimum when the starting point is chosen close to the local minimum. In this respect, please recall that we cannot start exactly at the local minimum, since V_α is not differentiable in this point.

Finally, we stress that we had to use a method for the solution of problem P that generates feasible iterates, since otherwise V_α might not be well-defined. On the other hand, this also has the advantage that we may apply our method to problems where the functions θ_ν of the players ν are not defined outside of W due to some logarithmic terms, for example. This is in contrast to other existing methods, which assume that the functions θ_ν are defined on the whole space $\mathbb{R}^n$.

6.2. Dual Gap Function Approach

Theorem 5.11 motivates to tackle a GNEP by solving the corresponding dual unconstrained minimization problem

$$D : \quad \min_{y \in \mathbb{R}^n} d_\alpha^*(y)$$

with the dual gap function d^*_α as the objective function. This dual gap function d^*_α is, however, relatively expensive to calculate, since two convex constrained optimization problems are to solve for each dual gap function evaluation. On the other hand, our previous results show that each function evaluation also provides the gradient automatically. We therefore use the global spectral gradient (GSG) method from [115], which is a global version of the spectral gradient method proposed by Barzilai and Borwein in [12]. The GSG method has the advantage that only first order information is required and that, typically, no line search with extra function evaluations are needed. We adapt this GSG method to the setting of problem D in the following algorithm.

Algorithm 6.2 *(Global spectral gradient method from [115])*

(S.0) Choose $y^0 \in \mathbb{R}^n$, $t_0 > 0$, *integer* $M > 0$, $\gamma \in]0,1[$, $0 < \sigma_1 < \sigma_2 < 1$, $0 < \varepsilon < 1$. *Set* $k := 0$.

(S.1) If a suitable termination criterion holds: STOP.

(S.2) Choose $\delta_k > 0$. *If* $t_k \geq \frac{1}{\varepsilon}$ *or* $t_k \leq \varepsilon$, *then set* $t_k := \delta_k$.

(S.3) (Nonmonotone line search) If the inequality

$$d^*_\alpha\left(y^k - t_k \nabla d^*_\alpha(y^k)\right) \leq \max_{0 \leq j \leq \min\{k,M\}} \left[d^*_\alpha(y^{k-j}) - \gamma t_k \|\nabla d^*_\alpha(y^k)\|^2\right]$$

holds, then set $y^{k+1} := y^k - t_k \nabla d^*_\alpha(y^k)$, *and go to (S.5). Otherwise go to (S.4).*

(S.4) Choose $\sigma_k \in [\sigma_1, \sigma_2]$, *set* $t_k := \sigma_k t_k$, *and go to (S.3).*

(S.5) Set $r^k := \nabla d^*_\alpha(y^{k+1}) - \nabla d^*_\alpha(y^k)$ *and* $t_{k+1} := -\dfrac{t_k \|\nabla d^*_\alpha(y^k)\|^2}{\left(\nabla d^*_\alpha(y^k)\right)^T r_k}$, $k \leftarrow k+1$, *and go to (S.1).*

Additionally, we use the same setting as in [115] with the parameters $t_0 := 1$, $M := 10$, $\gamma := 10^{-4}$, $\sigma_1 := 0.1$, $\sigma_2 := 0.5$, $\varepsilon := 10^{-10}$,

$$\delta_k := \begin{cases} 1, & \text{if } \|\nabla d^*_\alpha(y^k)\| > 1, \\ \|\nabla d^*_\alpha(y^k)\|, & \text{if } 10^{-5} \leq \|\nabla d^*_\alpha(y^k)\| \leq 1, \\ 10^{-5}, & \text{if } \|\nabla d^*_\alpha(y^k)\| < 10^{-5}, \end{cases}$$

and the quadratic interpolation of σ_k described in [29, p.127] with

$$\sigma_k := \begin{cases} \sigma_1, & \text{if } \sigma_{\min,k} < \sigma_1, \\ \sigma_{\min,k}, & \text{if } \sigma_1 \leq \sigma_{\min,k} \leq \sigma_2, \\ \sigma_2, & \text{if } \sigma_{\min,k} > \sigma_2, \end{cases}$$

where

$$\sigma_{\min,k} := \frac{\|\nabla d^*_\alpha(y^k)\|^2}{2\left(d^*_\alpha\,(y^k - t_k \nabla d^*_\alpha(y^k)) - d^*_\alpha(y^k) + \|\nabla d^*_\alpha(y^k)\|^2\right)}.$$

Example	N	n	α	$\lambda_{min,F}$	$\lambda_{min,H}$	y^0	k	$\#d^*_\alpha$	$d^*_\alpha(y^k)$	$\|\nabla d^*_\alpha(y^k)\|$
Ex. 5.10	2	2	2	2.00	0.76	$(0,0)$	0	1	0.0000e+00	0.0000e+00
						$(1,1)$	5	6	1.0856e-07	1.9022e-04
						$(10,10)$	3	4	2.5000e-01	7.9040e-07
Ex. 5.14	2	2	1	3.00	1.38	$(0,0)$	83	141	1.7000e+01	9.9880e-07
						$(1,1)$	64	106	1.7000e+01	9.9216e-07
						$(10,10)$	64	108	1.7000e+01	9.3826e-07
A3	3	7	63	13.86	0.83	$(0,\dots,0)$	970	1377	7.4902e-08	1.0444e-05
						$(1,\dots,1)$	1973	2809	4.0346e-07	5.5712e-05
						$(10,\dots,10)$	1499	2133	4.7735e-07	3.3865e-04
A5	3	7	14	4.76	0.61	$(0,\dots,0)$	20	22	6.7906e-07	1.7271e-04
						$(1,\dots,1)$	17	18	7.0005e-07	1.1031e-04
						$(10,\dots,10)$	32	35	3.3339e-07	1.7712e-04
A7	4	20	117	4.85	0.63	$(0,\dots,0)$	15	16	5.0958e-08	2.1327e-05
						$(1,\dots,1)$	15	16	5.2761e-08	2.1706e-05
						$(10,\dots,10)$	15	16	6.0704e-08	2.3002e-05
A8	3	3	2	2.00	0.76	$(0,0,0)$	11	12	1.4745e-07	1.3367e-04
						$(1,1,1)$	5	6	4.7894e-07	2.8679e-04
						$(10,10,10)$	9	10	1.2500e-01	3.3832e-07
A11	2	2	1	3.00	1.38	$(0,0)$	4	5	4.9960e-15	3.7652e-08
						$(1,1)$	0	1	0.0000e+00	5.6501e-10
						$(10,10)$	2	3	1.4334e-11	2.0243e-06
A12	2	2	2	2.00	1.00	$(2,0)$	4	5	2.5693e-10	6.0585e-06
						$(0,0)$	2	3	1.8645e-11	2.4924e-06
						$(1,1)$	2	3	2.4218e-10	6.9597e-06
						$(10,10)$	2	3	3.9645e-10	8.9048e-06
A13	3	3	2	2.02	0.78	$(0,0,0)$	13	15	9.9189e-07	2.0379e-04
						$(1,1,1)$	14	16	2.2737e-13	4.6215e-07
						$(10,10,10)$	20	25	-2.2737e-13	5.7849e-07
A15	3	6	3	3.02	0.60	$(0,\dots,0)$	1094	1433	9.9879e-07	1.0102e-05
						$(1,\dots,1)$	1099	1470	9.1871e-07	1.1981e-04
						$(10,\dots,10)$	1443	1922	3.3720e-07	3.7221e-04
A17	2	3	2	1.63	0.79	$(0,0,0)$	7	8	8.9804e-09	3.8128e-05
						$(1,1,1)$	8	9	0.0000e+00	2.7595e-08
						$(10,10,10)$	8	9	2.2402e-07	1.9045e-04
A18	2	12	2	2.00	0.76	$(0,\dots,0)$	23	25	9.3132e-10	2.6248e-06
						$(1,\dots,1)$	36	37	2.9153e-07	2.3941e-04
						$(10,\dots,10)$	27	32	1.4095e-08	4.8729e-05
T1	2	2400	2	2.76	0.57	$(0,\dots,0)$	18	19	3.2072e-07	3.0320e-04
						$(1,\dots,1)$	20	21	3.0623e-07	1.9514e-04
						$(10,\dots,10)$	21	22	3.3050e-07	2.2579e-04
T2	2	4800	2	2.76	0.57	$(0,\dots,0)$	19	20	1.7288e-08	5.3392e-05
						$(1,\dots,1)$	20	21	2.3871e-07	2.5285e-04
						$(10,\dots,10)$	21	22	6.2399e-07	2.9929e-04

Table 6.2.: Numerical results for some GNEPs with the global spectral gradient (GSG) method from [115] using the dual gap function d^*_α

k	$\#d_\alpha^*$	t_k	$d_\alpha^*(y^k)$	$\|\nabla d_\alpha^*(y^k)\|$
0	1	0.000	2.5322e+05	2.9233e+02
1	2	1.000	1.7608e+05	2.3741e+02
2	3	5.134	2.3860e+04	9.7272e+01
3	4	5.009	9.1171e+03	7.2845e+01
4	5	3.295	2.7760e+03	3.1505e+01
5	6	2.585	1.3002e+03	1.3528e+01
6	7	3.039	8.1620e+02	1.0461e+01
7	8	11.499	1.1249e+02	5.7856e+00
8	9	12.677	4.4560e+02	1.8486e+01
9	10	3.599	9.7405e+01	9.0002e+00
10	11	2.464	1.4541e+00	6.4820e-01
11	12	2.373	7.4622e-01	3.5923e-01
12	13	4.092	3.2655e-01	2.2632e-01
13	14	9.962	1.6841e-02	5.1157e-02
14	15	12.666	7.0349e-04	2.3781e-02
15	16	12.320	1.1471e-02	9.8163e-02
16	17	2.420	4.8805e-05	5.1082e-03
17	18	2.371	7.7907e-06	1.9112e-03
18	19	3.509	3.2072e-07	3.0320e-04

Table 6.3.: Numerical results for Example T1 with starting point $(0, \dots, 0)$ in each iteration of the global spectral gradient (GSG) method from [115]

Furthermore, we terminate the iteration if either $\|\nabla d_\alpha^*(y^k)\| \leq 10^{-6}$ or $d_\alpha^*(y^k) \leq 10^{-6}$ holds.

For the computation of the conjugate functions of f_α and h_α from Lemmata 5.8 and 5.9, respectively, we use the TOMLAB/SNOPT solver with settings `Prob.SOL.optPar(9)=` 10^{-8}, `Prob.SOL.optPar(11)=` 10^{-8} and `Prob.SOL.optPar(12)=` 10^{-8}, see the TOMLAB/SNOPT User's Guide on the web site [1] for more information about the TOMLAB/SNOPT solver.

The test problems used here are: Examples 5.10 and 5.14 from Section 5.1, a class of test examples indicated by a capital T, which are GNEP reformulations of a discrete approximation of a transportation problem defined as a generalized quasi-variational inequality problem in [121] with details given in Example 6.3, and a subset of test problems from the report version [48] of the paper [49], indicated by a capital A. All these test examples satisfy Assumptions 5.1 (a) and (c), whereas the requirement (b) of this assumption is violated except for Examples 5.10, 5.14, A8, and A11. For our method to work also on the remaining examples, we used the strategy for the choice of the parameter α outlined after the statement of Lemma 5.7.

More precisely, in our implementation, whenever possible, we first choose for each example the parameter α as the smallest integer such that the minimal eigenvalues $\lambda_{min,F}$ and $\lambda_{min,H}$ of the Hessians $\nabla^2_{xx}(-F_\alpha)$ and $\nabla^2_{(x,z)(x,z)}(-H_\alpha)$ with the functions F_α and H_α defined in (5.5) and (5.7), respectively, are larger than 0.5. This choice of α guarantees that the functions f_α and h_α have all the desired properties. Such a suitable choice was easily possible for the two transportation problems T1 and T2 based on Example 6.3 below as well as for all test problems from [48] with quadratic cost functions, whereas the other test problems from that collection were excluded from our test set. Note that, without this particular choice of α, we usually get much worse results and often do not even converge to a solution.

The details of Examples T1 and T2 are given in the next example.

Example 6.3 Consider a discrete approximation of a transportation problem defined as a generalized quasi-variational inequality problem in [121]. There are two players controlling $2T$ variables in the GNEP reformulation, where T gives the number of the partitions of the time interval $[0, 1]$. The optimization problems of the first player is given by

$$\begin{aligned} \min_{x^1} \quad & \sum_{i=1}^{T} \left(\frac{1}{2} \begin{pmatrix} x^1_{2i-1} & x^1_{2i} \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 2 & 10 \end{pmatrix} \begin{pmatrix} x^1_{2i-1} \\ x^1_{2i} \end{pmatrix} + \begin{pmatrix} x^1_{2i-1} & x^1_{2i} \end{pmatrix} \left[\begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} x^2_{2i-1} \\ x^2_{2i} \end{pmatrix} + \begin{pmatrix} 40 \\ 30 \end{pmatrix} \right] \right) \\ \text{subject to} \quad & 0 \le \begin{pmatrix} x^1_{2i-1} \\ x^1_{2i} \end{pmatrix} \le p_i + 12, \\ & x^1_{2i-1} + x^1_{2i} = -10 p_i + \frac{2}{3} x^2_{2i} + 11 \text{ for all } i = 1, \dots, T, \text{ with } p_i = \frac{i-1}{T-1}, \end{aligned}$$

and for the second player by

$$\begin{aligned} \min_{x^2} \quad & \sum_{i=1}^{T} \left(\frac{1}{2} \begin{pmatrix} x^2_{2i-1} & x^2_{2i} \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 2 & 10 \end{pmatrix} \begin{pmatrix} x^2_{2i-1} \\ x^2_{2i} \end{pmatrix} + \begin{pmatrix} x^2_{2i-1} & x^2_{2i} \end{pmatrix} \left[\begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} x^1_{2i-1} \\ x^1_{2i} \end{pmatrix} + \begin{pmatrix} 40 \\ 30 \end{pmatrix} \right] \right) \\ \text{subject to} \quad & 0 \le \begin{pmatrix} x^2_{2i-1} \\ x^2_{2i} \end{pmatrix} \le p_i + 12, \\ & x^2_{2i-1} + x^2_{2i} = -4 p_i + \frac{1}{2} x^2_{2i-1} + 7 \text{ for all } i = 1, \dots, T, \text{ with } p_i = \frac{i-1}{T-1}. \end{aligned}$$

The exact solution of this GNEP given in [121] is composed of

$$\left(x^1_{2i-1}, x^1_{2i}, x^2_{2i-1}, x^2_{2i} \right) = \left(-\frac{134}{15} p_i + \frac{601}{60}, -\frac{26}{15} p_i + \frac{77}{30}, -\frac{112}{15} p_i + \frac{289}{30}, -p_i + \frac{19}{8} \right)$$

for all $i = 1, \dots, T$, with $p_i = \frac{i-1}{T-1}$. For Examples T1 and T2 we choose $T = 600$ and $T = 1200$, respectively. ◇

The numerical results obtained with the GSG method are summarized in Table 6.2. This table contains the following data: the name of the example, the number of players N, the number of variables n, the value of the chosen parameter α, the eigenvalues $\lambda_{min,F}$ and $\lambda_{min,H}$ of the corresponding Hessians $\nabla^2_{xx}(-F_\alpha)$ and $\nabla^2_{(x,z)(x,z)}(-H_\alpha)$, respectively, the starting point $y^0 \in \mathbb{R}^n$, the number of iterations k, the cumulated number of dual gap function evaluations $\#d^*_\alpha$ until termination, the final value of the dual gap function $d^*_\alpha(y^k)$, and the final value of the gradient norm $\|\nabla d^*_\alpha(y^k)\|$.

The calculations with the GSG method were quite successful for most instances, except for Example 5.14 with all starting points and Examples 5.10 and A8 with the third starting point. The function value in these examples is not small enough and the iteration is terminated, since the norm of the gradient gets small, hence we are close to a non-optimal stationary point of the function d^*_α. Note that the failure in Example 5.14 was expected based on the considerations in this example. Additionally, there are two cases, namely Example 5.10 with the first starting point and Example A11 with the second starting point, where the starting point already provides a solution of the dual unconstrained minimization problem (5.11). Furthermore, the bad convergence speed in Examples A3 and A15 leads to a large number of iterations although the calculations of solutions for all starting points were successful. In all other test examples, we observed far better convergence properties. The iteration is terminated after a relatively small number of iterations, in particular, taking into account that the GSG method is just a first-order gradient method.

Table 6.3 gives some details on the iteration history for Example T1. The cumulated number of function evaluations $\#d_\alpha^*$ is, at each iteration k, equal to $k+1$, which indicates that no line search had to be applied at any of the iterations. This is quite typical for most of the test problems where the method was able to solve the example within a relatively small number of iterations.

6.3. Comparing the Two Approaches

The methods from Sections 6.1 and 6.2 have been tested partially on the same problems. Nevertheless, it is difficult to compare the results of the two approaches. First, the objective functions, which have to be minimized in both methods (constrained in Section 6.1 or unconstrained in Section 6.2), are different. Second, the starting points are not interconnected although they have equal values.

Anyway, we would like to compare the corresponding methods and the obtained results. To this end, we compute for the common examples the values of the primal gap function $V_\alpha(x^k)$ for $\alpha = 0.01$ and the corresponding vector $x^k := \nabla f_\alpha^*(y^k)$ after the termination of the GSG method, since $\alpha = 0.01$ has been chosen for the calculations of the primal gap function V_α in Section 6.1. Furthermore, for the common test examples we give the cumulated number of the primal gap function evaluations until the termination of the feasible direction-type method from Section 6.1. In order to allow an easier comparison, we combine these results with the results from Sections 6.1 and 6.2 in Table 6.4.

Table 6.4 contains the following data: the name of the example, the starting points x^0 and y^0, the number of iterations k for both methods, the cumulated number of the primal gap function evaluations $\#V_\alpha$ until the termination of the feasible direction-type method from Section 6.1 as well as the corresponding primal gap function value V_α^{opt}, the cumulated number of dual gap function evaluations $\#d_\alpha^*$ until termination of the GSG method as well as the corresponding primal gap function value $V_\alpha(x^k)$.

Now, we compare the fundamental points of both methods as far as possible. For the evaluation of the primal gap function we need to solve only one constrained convex optimization problem, whereas solving two constrained convex optimization problems are necessary for the evaluation of the dual gap function. Nevertheless, the GSG method yields lower cumulated numbers of function evaluations on average even if we double these numbers. Additionally, in each iteration of the feasible direction-type method one evaluation of the gradient ∇V_α is necessary, whereas the evaluation of the function d_α^* provides the gradient ∇d_α^* automatically. Furthermore, the average of the function evaluations per iteration in the GSG method is one or lies between one and two, whereas the average of the function evaluations per iteration in the method from Section 6.1 is bigger than two, less than nine, and on average bigger than four. The bad outliers for the GSG

		The feasible direction-type method from Section 6.1			The GSG method from Section 6.2		
Example	$x^0,\ y^0$	k	$\#V_\alpha$	V_α^{opt}	k	$\#d_\alpha^*$	$V_\alpha(x^k)$
A3	$(0,\ldots,0)$	14	123	2.6397e-05	970	1377	5.1765e-07
	$(1,\ldots,1)$	14	90	2.9342e-06	1973	2809	1.4069e-06
	$(10,\ldots,10)$	29	170	2.3998e-06	1499	2133	2.9298e-06
A5	$(0,\ldots,0)$	13	33	1.5470e-05	20	22	5.1974e-07
	$(1,\ldots,1)$	29	86	2.5578e-05	17	18	1.9621e-06
	$(10,\ldots,10)$	55	186	1.9534e-05	32	35	3.8092e-07
A7	$(0,\ldots,0)$	77	322	2.2178e-05	15	16	0.0000e+00
	$(1,\ldots,1)$	65	291	2.4832e-05	15	16	0.0000e+00
	$(10,\ldots,10)$	153	619	2.5712e-05	15	16	0.0000e+00
A8	$(0,0,0)$	15	48	1.1013e-05	11	12	9.3428e-08
	$(1,1,1)$	13	44	1.2011e-05	5	6	6.3017e-07
	$(10,10,10)$	13	44	1.2010e-05	9	10	0.2488e+00
A11	$(0,0)$	6	14	1.0419e-05	4	5	0.0000e+00
A12	(2,0)	15	89	1.4158e-05	4	5	4.4736e-10
A13	$(0,0,0)$	10	26	8.1729e-06	13	15	0.0000e+00
A15	$(0,\ldots,0)$	24	157	1.0849e-05	1094	1433	8.0144e-05
A17	$(0,0,0)$	23	66	1.0266e-05	7	8	0.0000e+00
A18	$(0,\ldots,0)$	44	218	9.1178e-06	23	25	0.0000e+00
	$(1,\ldots,1)$	57	262	2.8294e-06	36	37	0.0000e+00
	$(10,\ldots,10)$	57	282	1.6509e-05	27	32	0.0000e+00

Table 6.4.: Data for comparing the two approaches from Sections 6.1 and 6.2

method are problems A3 and A15, in which very slow convergence speed was observed. Another advantage of the GSG method is that the primal gap function values in most problems in Table 6.4 are closer to zero than the values of this function provided by the method from Section 6.1 or are even ‘exactly’ zero, and the value zero of the primal gap function indicates that a solution of the corresponding GNEP is found, see Lemma 3.2. The bad outlier for the GSG method in this context is problem A8 with the third starting point and the primal gap function value bigger than 0.2. Here we were close to a non-optimal stationary point of the dual gap function d_α^*.

Altogether, the GSG method fares better than the method from Section 6.1 if this method is applicable to the corresponding GNEP. Nevertheless, even if the strategy spaces fulfill the Assumptions 5.1 (a) and (c), it can be difficult to find the required DC reformulation of the primal gap function, except the cost functions of all players are quadratic or convex in the whole variable. For example, problems A1, A2, A4, A14, and A16 fulfill Assumptions 5.1 (a) and (c), but we could not find a DC reformulation of these problems and therefore not apply the GSG method to these problems, whereas we do not have to prepare the ground for the calculations with the method from Section 6.1. Anyway, if a GNEP fulfills Assumptions 5.1 (a) and (c) and we have a DC reformulation of this GNEP, it is promising to obtain the corresponding unconstrained dual reformulation, since the dual gap function is continuously differentiable, has, under suitable assumptions, a piecewise smooth gradient, and therefore allows the application of second-order Newton-type methods, whereas the primal gap function has, in general, nondifferentiable points, supplies a constrained reformulation, and may not be defined outside the feasible set W.

Part II.

Quasi-Variational Inequalities

7. Background on Quasi-Variational Inequalities

In Chapter 7 we provide the basis for the analysis in Part II. This chapter is organized in the following way: In Section 7.1 we give the definition of quasi-variational inequality problems and a literature overview for these problems, and in Section 7.2 we recall the definitions of some gap functions for variational inequalities and quasi-variational inequalities.

7.1. Definition and Overview

Given a function $F : \mathbb{R}^n \to \mathbb{R}^n$ and a set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ such that $S(x)$ is a closed and convex (possibly empty) set for any $x \in \mathbb{R}^n$, the finite-dimensional *quasi-variational inequality problem* (QVI) consists in finding a vector $x \in S(x)$ such that

$$F(x)^T(z - x) \geq 0 \qquad \forall z \in S(x). \tag{7.1}$$

If the set $S(x)$ is independent of x, that is, $S(x) = S$ for all $x \in \mathbb{R}^n$ with a constant, nonempty, closed, and convex set $S \subseteq \mathbb{R}^n$, then the QVI reduces to the so-called *variational inequality* (VI) problem, cf. the monograph [54] for an extensive discussion of VIs.

The QVI was formally introduced in a series of papers [14, 15, 16] by Bensoussan et al. It has soon become a powerful modeling tool for many different problems both in the finite and in the infinite-dimensional setting. An early summary may be found in the article by Mosco [93]. The infinite-dimensional problem with several mechanical and engineering applications is discussed in the monograph [10] by Baiocchi and A. Capelo. For several other applications, we refer the reader to the list of references in the recent paper [52]. In the meantime, several applications coming from totally different origins can also be found in a test problem collection whose details are given in [51].

Unfortunately, QVIs turn out to be a difficult class of problems, and the numerical solution of QVIs is still a challenging task. To the best of our knowledge, the first method was proposed by Chan and Pang in [23]. They consider a projection-type algorithm and prove a global convergence result under certain assumptions for the class of QVIs where the set-valued mapping S is given by $S(x) = c(x) + K$ for a suitable function $c : \mathbb{R}^n \to \mathbb{R}^n$ and a fixed, closed, and convex set $K \subseteq \mathbb{R}^n$. This particular class of problems is sometimes called a QVI with a *moving set* $S(x)$ since the fixed set K moves along the mapping $c(x)$. There are a number of subsequent extensions of this approach, see, for example, [5, 89, 96, 98, 100, 101, 118, 122], which all use a projection-type or fixed point iteration and essentially deal with the moving set case only in order to obtain suitable global convergence results. More recently, Pang and Fukushima [106] suggested a multiplier-penalty-type approach where they have to solve a sequence of (standard) VIs. They obtain a global convergence result for a class of problems not restricted to the moving set case, but their VI-subproblems are in general non-monotone and therefore difficult to solve. A recent interior point method by Facchinei et al. [52] applies a potential-reduction-type method to the corresponding KKT conditions and proves global convergence results for some classes of QVIs that go beyond the moving set case. Besides these (more or less) globally convergent approaches, there also exist some semismooth or nonsmooth Newton-type methods based on suitable reformulations of a QVI by Outrata et al., see, in particular, [103, 104, 105]. These Newton-type methods are locally fast convergent under appropriate assumptions. However, it is difficult to globalize these methods in comparison to another semismooth Newton method for the solution of the KKT system of a QVI that was considered and also globalized recently by Facchinei et al. in [50]. This paper investigates the theoretical basis for such semismooth methods regarding the nonsingularity of certain matrices and guarantees at least the superlinear local convergence for some important classes of QVIs under certain assumptions. Furthermore, the numerical results presented in [50] show that the globalized semismooth Newton method for the solution of the KKT system of a QVI has a very fast local convergence and can achieve a high accuracy, but is less robust than the interior point method in [52]. More comparisons between these two methods are given in [50]. Another fast local method, which can be applied to the KKT conditions of QVIs, is the LP-Newton method that was originally developed in [43, 44] for solving nonsmooth systems of equations with nonisolated solutions. Therefore, the LP-Newton method provides an approach even for solving QVIs in which nonisolated solutions and nonsmoothness occur. Such QVIs arise, for example, from GNEPs with nonisolated solutions, see [32, 33] for generic properties of solutions of GNEPs. We recall that, in the context of GNEPs, a combination of the LP-Newton method and the interior point method was studied in [35]. The disadvantage of the LP-Newton method is that the computational cost at each iteration can become immensely high for problems with large dimensions.

Apart from the previous classes of methods, there exist a number of different gap functions for QVIs, cf. [9, 31, 63, 67, 127] and the corresponding discussion in Section 7.2. These gap functions allow a reformulation of the QVI as an optimization problem and therefore the application of standard software. However, the disadvantage is that these gap functions are usually nonsmooth in the QVI setting, so that the previous literature concentrates on error bound results or the local Lipschitz continuity and directional differentiability of these gap functions. In particular, also the regularized gap function in the QVI setting that was originally introduced by Fukushima [62] in the context of standard variational inequalities is typically nonsmooth. Exceptions for QVIs where this regularized gap function turns out to be continuously differentiable everywhere are the cases of QVIs with the feasible sets of the moving-set-type [31] or a suitable generalization of it, see Section 8.1. Additionally, we show in Chapter 9 that, except for some pathological cases for general QVIs, this regularized gap function is continuously differentiable at all minimizers. Furthermore, for a class of QVIs that are different from the moving set case, Dietrich observed in [30] that this regularized gap function may be viewed as a difference of two convex functions and can therefore be used, by means of a suitable duality theory, to obtain a dual gap function. This dual gap function then gives a smooth reformulation for this class of

QVIs. Therefore, Chapter 10 aims at elaborating further on this dual approach. In particular, we get rid of the (implicit) assumption from [30] that the set-valued mapping defining the QVI is always nonempty-valued, since, in many practical instances, this is indeed frequently violated. Furthermore, we verify some stronger smoothness properties in Section 10.2 and present some numerical results obtained by the dual gap function approach in Chapter 11.

7.2. Preliminaries on Gap Functions

There exist several gap functions for QVIs. All these gap functions were originally introduced for standard VIs and then extended to QVIs. Therefore, we first recall the definitions of the relevant gap functions for VIs in Subsection 7.2.1 and then present their counterparts for QVIs in Subsection 7.2.2, together with some elementary properties of one of these gap functions that plays a central role in our subsequent analysis. Note that there exist other gap functions for VIs and QVIs, which, however, do not play any role in our context, see, for example, [99].

7.2.1. Gap Functions for Variational Inequalities

Recall that the (standard) variational inequality consists in finding a solution $x \in S$ such that

$$F(x)^T(z-x) \geq 0 \quad \forall z \in S$$

holds, where $S \subseteq \mathbb{R}^n$ is a nonempty, closed, and convex set, and $F : \mathbb{R}^n \to \mathbb{R}^n$ denotes a function, which we assume to be continuously differentiable in this subsection. The classical *gap function* for VI is defined by

$$g(x) := -\inf_{z\in S} F(x)^T(z-x)$$

and was introduced by Auslender [8], see also Hearn [76] and, for example, the paper [91] for an algorithmic application. The gap function is nonnegative on S, and $g(\bar{x}) = 0$ for some $\bar{x} \in S$ holds if and only if $\bar{x}$ solves the VI. Hence the VI is equivalent to the constrained optimization problem

$$\min\ g(x) \quad \text{subject to} \quad x \in S \tag{7.2}$$

with zero as the optimal value. However, unless S is compact, the objective function g is typically extended-valued, moreover, g is usually nondifferentiable.

In order to avoid these problems, Fukushima [62] and Auchmuty [7] independently developed the *regularized gap function*

$$g_\alpha(x) := -\min_{z\in S}\left[F(x)^T(z-x) + \frac{\alpha}{2}\|z-x\|^2\right],$$

where $\alpha > 0$ denotes a given parameter. Note that the strongly convex minimization problem in g_α with respect to the nonempty, closed, and convex set S is uniquely solvable for all $x \in \mathbb{R}^n$ and therefore we may use minimum istead of infimum in the definition of the regularized gap function. Similar to the gap function, one can show that also the regularized gap function is nonnegative on S, and $g_\alpha(\bar{x}) = 0$ for some $\bar{x} \in S$ holds if and only if $\bar{x}$ solves the VI. Moreover, g_α is finite-valued and continuously differentiable everywhere (by Danskin's Theorem, see, e.g., [8, Chapter 4, Theorem 1.7]). Hence the VI is equivalent to a smooth optimization problem of the form (7.2) with g being replaced by g_α. This fact has been exploited, for example, in the paper [129], which presents a simple globalization of the standard Josephy-Newton method based on the regularized gap function.

The main computational burden of the regularized gap function is the fact that the evaluation of $g_\alpha(x)$ is quite expensive for nonlinear (non-polyhedral) sets S since then one has to solve a convex optimization problem with a nonlinear feasible set, which is practically impossible. Motivated by this observation, Taji and Fukushima [128] introduced the following modification of the regularized gap function:

$$\tilde{g}_\alpha(x) := -\min_{z\in T(x)} \left[F(x)^T(z-x) + \frac{\alpha}{2}\|z-x\|^2 \right],$$

where $T(x)$ denotes the polyhedral approximation of S at x defined by

$$T(x) := \left\{ z \in \mathbb{R}^n \;\middle|\; s_i(x) + \nabla s_i(x)^T(z-x) \le 0 \;\; \forall i = 1,\dots,m \right\}$$

and where we assume that the feasible set S has the representation

$$S = \{x \in \mathbb{R}^n \mid s_i(x) \le 0 \;\; \forall i = 1,\dots,m\}$$

for some convex functions $s_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1,\dots,m$. It was shown in [128] that, once again, the VI is equivalent to a constrained optimization problem like (7.2) with $\tilde{g}_\alpha$ replacing g, and with zero objective function value at the solution. However, in contrast to the regularized gap function g_α, the mapping $\tilde{g}_\alpha$ is, in general, not differentiable.

7.2.2. Gap Functions for Quasi-Variational Inequalities

In the context of QVIs, the *fixed point* set of S,

$$X := \{x \in \mathbb{R}^n \mid x \in S(x)\} \tag{7.3}$$

plays a special role and is sometimes called the *feasible set* of the underlying QVI. In case of a VI, this set is equal to the constant set S and therefore justifies this terminology. Furthermore, the (*effective*) *domain* of S,

$$M := \operatorname{dom} S := \{x \in \mathbb{R}^n \mid S(x) \neq \emptyset\} \tag{7.4}$$

will also play a central role in Part II. Clearly, the relation

$$X \subseteq M \tag{7.5}$$

holds.

Consider the QVI from (7.1). A direct extension of the classical gap function from VIs to QVIs seems to be due to Giannessi [67], who defines the mapping

$$g(x) := -\inf_{z\in S(x)} F(x)^T(z-x)$$

and shows that

- $g(x) \geq 0$ for all $x \in X$;
- $g(\bar{x}) = 0$ for some $\bar{x} \in X$ if and only if $\bar{x}$ solves the QVI,

where X denotes the feasible set of a QVI from (7.3). Hence the QVI is equivalent to the constrained optimization problem

$$\min \; g(x) \quad \text{subject to} \quad x \in X$$

with zero as the optimal value. However, the optimal value function g is nondifferentiable and possibly extended-valued (both $g(x) = -\infty$ and $g(x) = +\infty$ may occur if $S(x) = \emptyset$ or g is unbounded from above). Further note that the set X might have a complicated structure.

An extension of the regularized gap function to QVIs is due to Taji [127] and was, in fact, introduced earlier by Dietrich [30] in the infinite-dimensional setting, see also the later paper [9] by Aussel et al. This *regularized gap function* for QVIs is defined by

$$g_\alpha(x) := -\inf_{z \in S(x)} \left[F(x)^T(z - x) + \frac{\alpha}{2}\|z - x\|^2 \right] \tag{7.6}$$

where $\alpha > 0$ denotes a given parameter. In view of Assumption 9.1, the function

$$\varphi_\alpha(z, x) := F(x)^T(z - x) + \frac{\alpha}{2}\|z - x\|^2 \tag{7.7}$$

is strongly convex in z for each fixed $x \in \mathbb{R}^n$. Therefore, the following remark holds.

Remark 7.1 For any $x \in M$ (the domain of S) the minimum in (7.6) is uniquely attained by the solution $z_\alpha(x)$ of the optimization problem

$$\min_z \; \varphi_\alpha(z, x) \quad \text{subject to} \quad z \in S(x). \tag{7.8}$$

In particular, we have $g_\alpha(x) = -\varphi_\alpha(z_\alpha(x), x) \in \mathbb{R}$. Note, however, that $g_\alpha(x) = -\infty$ holds for $x \notin M$, so that g_α is real-valued exactly on M. Consequently, due to (7.5), the optimal value function g_α is real-valued on X. ◇

The following result, whose proof may be found in [127], clarifies the relation between the regularized gap function g_α and the QVI (7.1). Recall once again that the set X in this result denotes the feasible set from (7.3).

Lemma 7.2 *The following statements hold:*

(a) $g_\alpha(x) \geq 0$ *for all* $x \in X$.

(b) $g_\alpha(\bar{x}) = 0$ *for some* $\bar{x} \in X$ *if and only if* $\bar{x}$ *is a solution of the QVI.*

(c) A point $\bar{x} \in X$ *is a solution of the QVI if and only if* $z_\alpha(\bar{x}) = \bar{x}$, *where*

$$z_\alpha(x) := \operatorname*{argmin}_{z \in S(x)} \varphi_\alpha(z, x).$$

Lemma 7.2 (a) and (b) shows that solving the QVI is equivalent to finding an optimal point $\bar{x}$ of the constrained optimization problem

$$\min\ g_\alpha(x) \quad \text{subject to} \quad x \in X \tag{7.9}$$

with $g_\alpha(\bar{x}) = 0$. Unfortunately, and in contrast to standard VIs, simple examples show that the objective function of this problem is nondifferentiable in general, and for infeasible points $x \notin X$, it might also take the value $-\infty$ (cf. Remark 7.1). Additionally, we can rewrite the constrained optimization problem (7.9) as the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} [g_\alpha(x) + \delta_X(x)] \tag{7.10}$$

with the indicator function δ_X of X defined in (2.1) and the convention $\eta + \infty = +\infty$ for all $\eta \in \mathbb{R} \cup \{\pm\infty\}$. Note that this convention makes sense in our case since the objective function from (7.10) should take the function value $+\infty$ outside of X, in particular, we would like to have $g_\alpha(x) + \delta_X(x) = +\infty$ also for all $x \notin X \cup M = M$. This unconstrained reformulation will be considered in Chapter 10 in order to obtain a smooth dual unconstrained optimization reformulation for a class of QVIs. Therefore, we call the regularized gap function g_α sometimes also the *primal gap function*.

Based on the reformulation (7.9), it seems natural to replace g_α by the counterpart of the modified regularized gap function $\tilde{g}_\alpha$ from the previous subsection in cases where the set-valued mapping S is defined by

$$S(x) = \{z \in \mathbb{R}^n \mid s_i(z, x) \le 0 \ \ \forall i = 1, \dots, m\}$$

with functions $s_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$, which are continuous on $\mathbb{R}^n \times \mathbb{R}^n$ and convex in z for each fixed $x \in \mathbb{R}^n$. In fact, this was done by Fukushima [63], but we skip the corresponding details here, mainly because it turns out that the regularized gap function has better differentiability properties. In fact, in an important special case to be discussed in the following chapter, the regularized gap function from (7.6) turns out to be continuously differentiable, whereas the modified regularized gap function from [63] would still be nonsmooth in general.

We conclude this subsection by introducing an example, which not only illustrates Lemma 7.2 but will also serve to illustrate continuity and differentiability properties of g_α on X in Sections 9.1 and 9.2, respectively.

Example 7.3 Consider the QVI with $n = 2$, $F(x) := (1, 1)^T$, and the set-valued mapping S defined by $S(x) := \{z \in \mathbb{R}^2 \mid s_i(z, x) \le 0 \ \ \forall i = 1, 2, 3\}$, where

$$s_1(z, x) := -2z_1 + x_2, \quad s_2(z, x) := x_1^2 + z_2^2 - 1, \quad s_3(z, x) := -x_1 - z_2.$$

Then for $x \in M$ we have $S(x) = S_1(x) \times S_2(x)$ with

$$\begin{aligned} S_1(x) &= \{z_1 \in \mathbb{R} \mid -2z_1 + x_2 \le 0\} = \left[\frac{x_2}{2}, +\infty\right[, \\ S_2(x) &= \left\{z_2 \in \mathbb{R} \;\middle|\; x_1^2 + z_2^2 - 1 \le 0,\ -x_1 - z_2 \le 0\right\} = \left[\max\left\{-x_1, -\sqrt{1 - x_1^2}\right\}, \sqrt{1 - x_1^2}\right], \end{aligned}$$

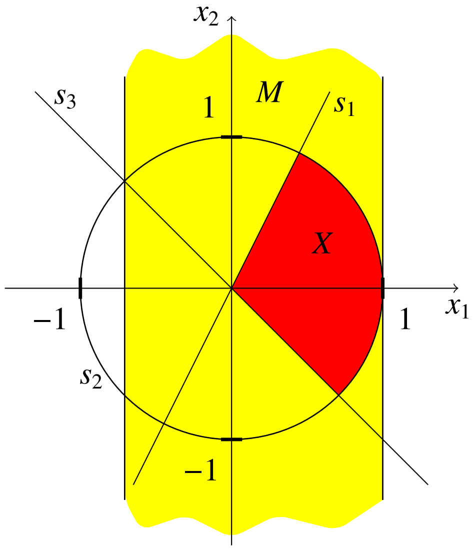


Figure 7.1.: Illustration of the sets X and M from Example 7.3

so that $M = \left[-1/\sqrt{2}, 1\right] \times \mathbb{R}$ and

$$X = \left\{x \in \mathbb{R}^2 \;\middle|\; -2x_1 + x_2 \leq 0,\ x_1^2 + x_2^2 - 1 \leq 0,\ -x_1 - x_2 \leq 0\right\},$$

see Figure 7.1. For the regularized gap function with $\alpha > 0$ we obtain

$$\begin{aligned} g_\alpha(x) &= -\inf_{z \in S(x)} \left[F(x)^T(z-x) + \frac{\alpha}{2}\|z-x\|^2\right] \\ &= x_1 + x_2 - \min_{z_1 \in S_1(x)} \left(z_1 + \frac{\alpha}{2}(z_1 - x_1)^2\right) - \inf_{z_2 \in S_2(x)} \left(z_2 + \frac{\alpha}{2}(z_2 - x_2)^2\right). \end{aligned} \tag{7.11}$$

For $x \in M$ the two components of $z_\alpha(x)$ are the unique optimal points corresponding to the two optimal values in (7.11). In fact, for $x \in M$, with

$$\varrho_1(x) := x_1 - \frac{x_2}{2}, \quad \varrho_2(x) := x_2 + \min\left\{x_1, \sqrt{1 - x_1^2}\right\}, \quad \varrho_3(x) := x_2 - \sqrt{1 - x_1^2},$$

we have

$$(z_\alpha(x))_1 = \begin{cases} x_1 - \varrho_1(x), & \text{if } \varrho_1(x) \leq \frac{1}{\alpha}, \\ x_1 - \frac{1}{\alpha}, & \text{if } \frac{1}{\alpha} < \varrho_1(x), \end{cases}$$

and

$$(z_\alpha(x))_2 = \begin{cases} x_2 - \varrho_2(x), & \text{if } \varrho_2(x) \leq \frac{1}{\alpha}, \\ x_2 - \frac{1}{\alpha}, & \text{if } \varrho_3(x) < \frac{1}{\alpha} < \varrho_2(x), \\ x_2 - \varrho_3(x), & \text{if } \frac{1}{\alpha} \leq \varrho_3(x). \end{cases}$$

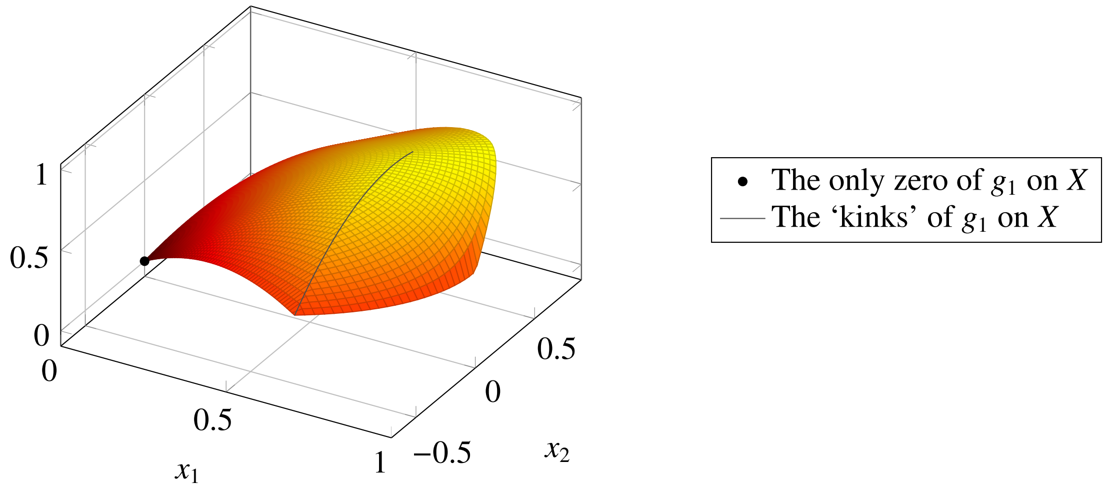


Figure 7.2.: Regularized gap function on X for $\alpha = 1$ from Example 7.3

Inserting these optimal values into (7.11) leads to

$$g_\alpha(x) = \begin{cases} \varrho_1(x) - \frac{\alpha}{2}\varrho_1^2(x), & \text{if } \varrho_1(x) \le \frac{1}{\alpha}, \\ \frac{1}{2\alpha}, & \text{if } \frac{1}{\alpha} < \varrho_1(x), \end{cases} + \begin{cases} \varrho_2(x) - \frac{\alpha}{2}\varrho_2^2(x), & \text{if } \varrho_2(x) \le \frac{1}{\alpha}, \\ \frac{1}{2\alpha}, & \text{if } \varrho_3(x) < \frac{1}{\alpha} < \varrho_2(x), \\ \varrho_3(x) - \frac{\alpha}{2}\varrho_3^2(x), & \text{if } \frac{1}{\alpha} \le \varrho_3(x), \end{cases}$$

for all $x \in M$. Figure 7.2 illustrates the graph of the regularized gap function on the set X for $\alpha = 1$. It can be shown that $\bar{x} := 0$ is the unique global minimizer of g_α on X with $g_\alpha(\bar{x}) = 0$. Therefore, $\bar{x}$ is the unique solution of the QVI by Lemma 7.2. ◇

8. Special Classes of Quasi-Variational Inequalities

In Chapter 8 we discuss three special classes of QVIs, namely QVIs with a generalization of the moving set case for which the regularized gap function turns out to be continuously differentiable under suitable assumptions in Section 8.1, further QVIs with set-valued mappings in product form in Section 8.2, and, finally, QVIs as an important application to generalized Nash equilibrium problems in Section 8.3. The results of this chapter were published in [75, Section 3].

8.1. Quasi-Variational Inequalities with Generalized Moving Sets

Many papers dealing with QVIs do not consider the general setting from (7.1), see, for example, [23, 31, 96]. They only discuss the particular case where the nonempty, closed, and convex set $S(x)$ has the form $S(x) = c(x) + K$ for some function $c : \mathbb{R}^n \to \mathbb{R}^n$ and a fixed nonempty, closed, and convex set $K \subseteq \mathbb{R}^n$. This class of QVIs is often called the *moving set case* for reasons that should be clear from Figure 8.1 (a). We assume in this section that the function c is continuous. If the set K is described as

$$K = \{y \in \mathbb{R}^n \mid k_i(y) \le 0,\ i = 1, \dots, m\}$$

with convex functions $k_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$, then the set-valued mapping S is defined by

$$S(x) = \{z \in \mathbb{R}^n \mid s_i(z, x) := k_i(z - c(x)) \le 0 \quad \forall i = 1, \dots, m\},$$

where the functions s_i are continuous on $\mathbb{R}^n \times \mathbb{R}^n$ and convex in z for each fixed $x \in \mathbb{R}^n$, such that the sets $S(x)$ are nonempty, closed, and convex for all $x \in \mathbb{R}^n$.

Here we consider a generalization of this case. To this end, let c be given as before and let $K \subseteq \mathbb{R}^p$ be a nonempty, closed, and convex set with $p \le n$. In addition, assume that we have a matrix $Q(x) \in \mathbb{R}^{n \times p}$ of full (column) rank for all $x \in \mathbb{R}^n$. Then we consider the case where the set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ has the form

$$S(x) = c(x) + Q(x)K := \{c(x) + Q(x)y \mid y \in K\}. \tag{8.1}$$

Note that $S(x) \neq \emptyset$ holds in this case for any $x \in \mathbb{R}^n$, that is, we have $M = \mathbb{R}^n$. We call a QVI with the mapping S defined in this way the *generalized moving set case*. In the special case $p = n$ and $Q(x) = I_n$ for all $x \in \mathbb{R}^n$ we re-obtain the moving set case. Our generalization of this case

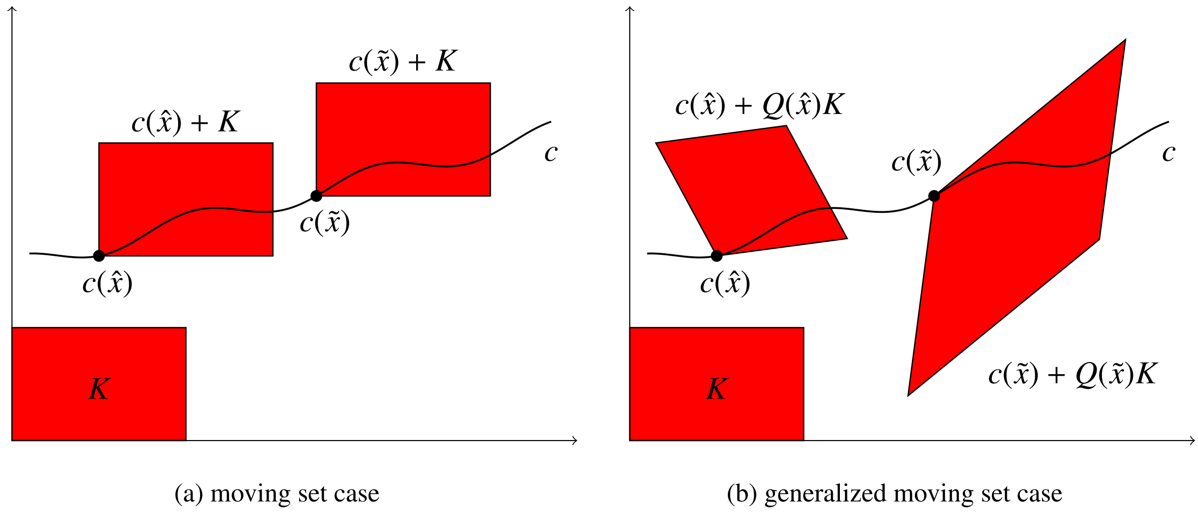


Figure 8.1.: Examples for a moving set case and a generalized moving set case

for $p = n$ actually allows any x-dependent affine transformation $T(K, x) = c(x) + Q(x)K$ of K instead of just translation, that is, also scaling, rotation, reflection, and shearing, as shown in Figure 8.1 (b) for $p = n = 2$. A further generalization of this approach is obtained in Remark 8.3.

Note that if c and Q are continuous and if K is described as

$$K = \{y \in \mathbb{R}^p \mid k_i(y) \leq 0,\ i = 1, \ldots, m\}$$

with convex functions $k_i : \mathbb{R}^p \to \mathbb{R}$, $i = 1, \ldots, m$. Then for $p = n$ the set-valued mapping S is defined by

$$S(x) = \left\{z \in \mathbb{R}^n \;\middle|\; s_i(z, x) := k_i\left(Q(x)^{-1}(z - c(x))\right) \leq 0 \quad \forall i = 1, \ldots, m\right\},$$

For $p < n$ and any $x \in \mathbb{R}^n$ one may choose some matrix $B(x) \in \mathbb{R}^{n\times(n-p)}$ whose columns form a basis of the null space of $Q(x)^T$, with the function B being continuous, cf. [68], and then set

$$S(x) = \Big\{z \in \mathbb{R}^n \;\Big|\; s_i(z, x) := k_i\left(\left(Q(x)^T Q(x)\right)^{-1} Q(x)^T (z - c(x))\right) \leq 0 \quad \forall i = 1, \ldots, m,$$
$$\bar{s}(z, x) := B(x)^T (z - c(x)) = 0\Big\}$$

In both cases, $p = n$ and $p < n$, the functions s_i are continuous on $\mathbb{R}^n \times \mathbb{R}^n$ and convex in z for each fixed $x \in \mathbb{R}^n$. Furthermore, the function $\bar{s}$ is continuous on $\mathbb{R}^n \times \mathbb{R}^n$ and linear in z for each fixed $x \in \mathbb{R}^n$. Consequently, the sets $S(x)$ are nonempty, closed, and convex for all $x \in \mathbb{R}^n$.

With the exception of the recent paper [52], the QVIs with moving sets are essentially the only case that has been investigated in papers dealing with the numerical solution of QVIs, and for which a more or less complete convergence theory is available. For example, Dietrich [31] considers QVIs with moving sets only and notes that the regularized gap function is continuously differentiable in this case. It seems that this observation has been widely overlooked in the subsequent literature.

In this section, we want to generalize this observation by showing that the regularized gap function g_α from (7.6) is still smooth in the case where the set $S(x)$ is given by (8.1) with continuously differentiable functions c and Q. To this end, we first reformulate the minimization problem from (7.8) as

$$\begin{aligned} &\min_z \;\; \varphi_\alpha(z,x) \quad \text{subject to} \quad z \in S(x) \\ &\quad \Longleftrightarrow \min_z \;\; \varphi_\alpha(z,x) \quad \text{subject to} \quad z \in c(x) + Q(x)K \\ &\quad \Longleftrightarrow \min_z \;\; \varphi_\alpha(z,x) \quad \text{subject to} \quad \exists\, y \in K: \; z = c(x) + Q(x)y \\ &\quad \Longleftrightarrow \min_{y,z} \;\; \varphi_\alpha(z,x) \quad \text{subject to} \quad z = c(x) + Q(x)y, \; y \in K \\ &\quad \Longleftrightarrow \min_y \;\; \psi_\alpha(y,x) \quad \text{subject to} \quad y \in K, \end{aligned} \tag{8.2}$$

where

$$\begin{aligned} \psi_\alpha(y,x) &:= \varphi_\alpha(c(x) + Q(x)y, x) \\ &= F(x)^T(c(x)-x) + \frac{\alpha}{2}\|c(x)-x\|^2 + \big(F(x) + \alpha(c(x)-x)\big)^T Q(x)y + \frac{\alpha}{2} y^T Q(x)^T Q(x) y \end{aligned}$$

is convex quadratic in y for each x. Note that the full rank of $Q(x)$ is actually not needed for the reformulation (8.2), but that under this assumption, for each fixed $x \in \mathbb{R}^n$, the function $\psi_\alpha(\cdot,x)$ is strongly convex with respect to y because

$$\nabla^2_{yy}\psi_\alpha(y,x) = \alpha Q(x)^T Q(x)$$

is uniformly positive definite (in y). Therefore, problem (8.2) has a unique solution $y_\alpha(x)$ for all $x \in \mathbb{R}^n$, and we obtain

$$g_\alpha(x) = -\min_{z \in S(x)} \varphi_\alpha(z,x) = -\min_{y\in K} \psi_\alpha(y,x) = -\psi_\alpha(y_\alpha(x),x).$$

The function $x \mapsto y_\alpha(x)$ turns out to be continuous on $\mathbb{R}^n$.

Proposition 8.1 *Let F be continuous on $\mathbb{R}^n$. Consider a QVI with $S(x)$ being defined by* (8.1) *with $p \le n$, $K \subseteq \mathbb{R}^p$ being nonempty, closed, and convex, c and Q being continuous, and $Q(x)$ having full rank for each fixed $x \in \mathbb{R}^n$. Then the function $x \mapsto y_\alpha(x)$ is continuous on $\mathbb{R}^n$.*

Proof. First recall that $-\psi_\alpha(\cdot,x)$ is concave for each fixed $x \in \mathbb{R}^n$ and continuous on $\mathbb{R}^p \times \mathbb{R}^n$. Since K is a closed set, the constant set-valued mapping $x \mapsto K$ is continuous on $\mathbb{R}^n$, see Definition 2.1. Moreover, the set K is convex. Furthermore, the set

$$Y_\alpha(x) := \left\{ \zeta \in K \;\middle|\; \max_{y\in K}\,(-\psi_\alpha(y,x)) = -\psi_\alpha(\zeta,x) \right\}$$

is a singleton for all $x \in \mathbb{R}^n$ since the function $\psi_\alpha(\cdot, x)$ is strongly convex for each fixed $x \in \mathbb{R}^n$, and the set K is nonempty, closed, and convex. Therefore, Lemma 2.3 implies that the (single-valued) set-valued mapping $x \mapsto Y_\alpha(x) = \{y_\alpha(x)\}$ is continuous on $\mathbb{R}^n$. Hence the function $x \mapsto y_\alpha(x)$ is continuous on $\mathbb{R}^n$. □

Since we minimize the function $\psi_\alpha(\cdot, x)$ with respect to a fixed set K, we may apply Danskin's Theorem and Proposition 8.1 and immediately obtain the following result.

Proposition 8.2 *Let F be continuously differentiable on $\mathbb{R}^n$. Consider a QVI with $S(x)$ being defined by* (8.1) *with $p \leq n$, $K \subseteq \mathbb{R}^p$ being nonempty, closed, and convex, c and Q being continuously differentiable, and $Q(x)$ having full rank for each fixed $x \in \mathbb{R}^n$. Then g_α is continuously differentiable with gradient*

$$\begin{aligned}\nabla g_\alpha(x) &= -\nabla_x \psi_\alpha(y, x)\Big|_{y=y_\alpha(x)} \\ &= \Big[\nabla F(x)(x - c(x) - Q(x)y) + \\ &\quad +(I_n - \nabla c(x) - \nabla_x (Q(x)y))(\alpha(c(x) + Q(x)y - x) + F(x))\Big]_{y=y_\alpha(x)}, \end{aligned} \tag{8.3}$$

where $y_\alpha(x)$ denotes the unique solution of problem (8.2).

As mentioned above, a further generalization of the generalized moving set case is possible and is specified in the next remark.

Remark 8.3 A careful analysis of the above proofs shows that the introduced generalized moving set case with a nonempty, closed and convex set K can be further generalized to the case $S(x) = T(K, x)$ with any continuously differentiable nonlinear mapping $T : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}^n$ such that $\psi_\alpha(y, x) = \varphi_\alpha(T(y, x), x)$ is strongly convex in y for all fixed $x \in \mathbb{R}^n$. In applications, however, it might be cumbersome to check the strong convexity assumption on the function ψ_α. ◇

The following example illustrates the results of this section.

Example 8.4 Let $p = n = 2$, $K = \mathbb{R}^2_\geq$, and F be continuously differentiable on $\mathbb{R}^2$. On K we simultaneously impose the translation $c(x) := x$, the scaling $\gamma(x) > 0$ and the rotation by the angle $\omega(x)$ for $x \in \mathbb{R}^2$ with continuously differentiable functions $\gamma : \mathbb{R}^2 \to \mathbb{R}$ and $\omega : \mathbb{R}^2 \to \mathbb{R}$. Then we may set $Q(x) := \gamma(x)R(x)$ with the rotation matrix

$$R(x) := \begin{pmatrix} \cos(\omega(x)) & -\sin(\omega(x)) \\ \sin(\omega(x)) & \cos(\omega(x)) \end{pmatrix}$$

and $S(x) = x + Q(x)K$. Clearly, $Q(x)$ is nonsingular for all $x \in \mathbb{R}^2$, and we obtain

$$\psi_\alpha(y, x) = F(x)^T Q(x)y + \frac{\alpha\gamma^2(x)}{2} y^T y.$$

For a given $x \in \mathbb{R}^2$ the unconstrained minimizer of $\psi_\alpha(\cdot, x)$ is

$$y^*_\alpha(x) = -\frac{1}{\alpha\gamma(x)} R(x)^T F(x).$$

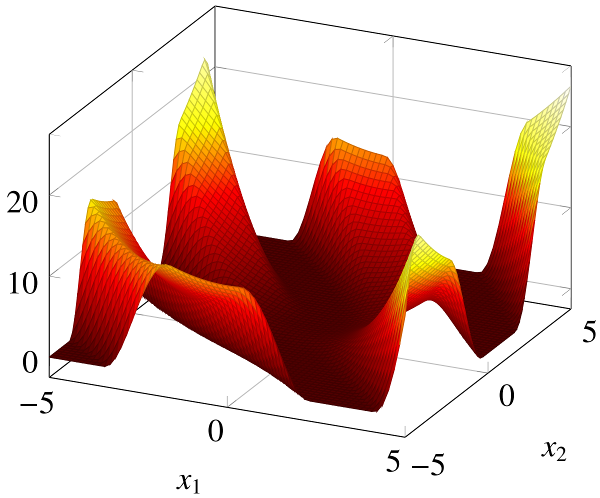


Figure 8.2.: Regularized gap function with $\alpha = 1$ from Example 8.4

Therefore, the minimizer of $\psi_\alpha(\cdot, x)$ on $K = \mathbb{R}^2_\geq$ is

$$y_\alpha(x) = \max\left\{0, -\frac{1}{\alpha\gamma(x)}R(x)^T F(x)\right\}$$

(with the maximum taken componentwise) for all $x \in \mathbb{R}^2$. The function y_α is obviously continuous on $\mathbb{R}^2$, and the regularized gap function

$$g_\alpha(x) = -\psi_\alpha(y_\alpha(x), x) = \frac{\alpha\gamma^2(x)}{2}\|y_\alpha(x)\|^2 = \frac{1}{2\alpha}\left\|\max\{0, -R(x)^T F(x)\}\right\|^2$$

is also known to be continuously differentiable on $\mathbb{R}^2$. By Propositions 8.1 and 8.2, we achieve the same results for this particular example. Note that the regularized gap function g_α does not depend on the scaling function γ.

Due to $0 \in K$ we have $x \in S(x) = x + Q(x)K$ for all x, so that $X = \mathbb{R}^2$ and, by Lemma 7.2, the solutions of the QVI are exactly the unconstrained minimizers of g_α with value zero, that is, the vectors $x \in \mathbb{R}^2$ with

$$\max\{0, -R(x)^T F(x)\} = 0.$$

Thus, the solutions of the QVI are formed by the set

$$\{x \in \mathbb{R}^2 \mid R(x)^T F(x) \geq 0\}.$$

A plot of the regularized gap function with the special choices $F(x) := x$, $\omega(x) := x_1 + x_2$ and $\alpha = 1$ is illustrated in Figure 8.2. ◇

In Section 9.2, we will investigate the smoothness properties of the regularized gap function g_α in general cases.

8.2. Quasi-Variational Inequalities with Set-valued Mappings in Product Form

Motivated by Example 7.3 (and Section 8.3 below), let us consider QVIs with a set-valued mapping S in product form. We say that a set-valued mapping S is in *product form* if there exist some $N \in \mathbb{N}$, $n_\nu \in \mathbb{N}$, $\nu = 1, \dots, N$, with $n_1 + n_2 + \dots + n_N = n$, and set-valued mappings $S_\nu : \mathbb{R}^n \rightrightarrows \mathbb{R}^{n_\nu}$, $\nu = 1, \dots, N$ such that

$$S(x) = S_1(x) \times S_2(x) \times \dots \times S_N(x)$$

holds for all $x \in \mathbb{R}^n$. After partitioning the variables $x = (x^1, \dots, x^N)$ and $z = (z^1, \dots, z^N)$ as well as the function $F(x) = (F^1(x), \dots, F^N(x))$, we may use the separability with respect to z of the function φ_α from (7.7) to obtain

$$\begin{aligned} g_\alpha(x) &= - \inf_{z \in S(x)} \left[F(x)^T (z - x) + \frac{\alpha}{2} \|z - x\|^2 \right] \\ &= - \sum_{\nu=1}^{N} \inf_{z^\nu \in S_\nu(x)} \left[F^\nu(x)^T (z^\nu - x^\nu) + \frac{\alpha}{2} \|z^\nu - x^\nu\|^2 \right] = \sum_{\nu=1}^{N} g_\alpha^\nu(x) \end{aligned} \tag{8.4}$$

with

$$g_\alpha^\nu(x) := - \inf_{z^\nu \in S_\nu(x)} \left[F^\nu(x)^T (z^\nu - x^\nu) + \frac{\alpha}{2} \|z^\nu - x^\nu\|^2 \right], \quad \nu = 1, \dots, N. \tag{8.5}$$

As the function g_α, the functions g_α^ν are nonnegative on X.

Lemma 8.5 *For all $x \in X$ and $\nu \in \{1, \dots, N\}$, we have $g_\alpha^\nu(x) \geq 0$.*

Proof. For any $\nu \in \{1, \dots, N\}$ choose some $x \in X$. Then we have

$$(x^1, x^2, \dots, x^N) \in S_1(x) \times S_2(x) \times \dots \times S_N(x)$$

and, in particular, $x^\nu \in S_\nu(x)$. Consequently, a lower bound of $g_\alpha^\nu(x)$ is the value

$$- \left[F^\nu(x)^T (z^\nu - x^\nu) + \frac{\alpha}{2} \|z^\nu - x^\nu\|^2 \right]_{z^\nu := x^\nu} = 0.$$

This shows the assertion. □

The combination of Lemma 7.2 (a) and (b), the reformulation (8.4), and Lemma 8.5 immediately yield the following separation result.

Theorem 8.6 *A point $\bar{x}$ solves a QVI with set-valued mapping in product form if and only if $\bar{x}$ is the global minimizer of g_α^ν on X with value zero for all $\nu = 1, \dots, N$.*

Next, we combine the ideas of generalized moving sets from Section 8.1 with set-valued mappings in product form. In fact, the product form and the resulting separability allow each set $S_\nu(x)$, $\nu = 1, \dots, N$, to be written as an *independent* generalized moving set, that is,

$$S_\nu(x) = \{c^\nu(x) + Q^\nu(x) y \mid y \in K^\nu\} \tag{8.6}$$

where, for $p_\nu \leq n_\nu$, the set $K^\nu \subseteq \mathbb{R}^{p_\nu}$ is nonempty, closed, and convex, the functions $c^\nu : \mathbb{R}^n \to \mathbb{R}^{n_\nu}$ and $Q^\nu : \mathbb{R}^n \to \mathbb{R}^{n_\nu \times p_\nu}$ are continuous, and $Q^\nu(x)$ has full (column) rank for all $x \in \mathbb{R}^n$. Then the proof of the assertion in Proposition 8.2 translates word by word to a proof of the assertion that, under continuous differentiability assumptions on F, c^ν and Q^ν, the function g_α^ν from (8.5) is continuously differentiable for each $\nu = 1, \dots, N$ with known gradient.

To prepare the statement of this result note that, for $\nu = 1, \dots, N$, we may rewrite the function g_α^ν from (8.5) as

$$g_\alpha^\nu(x) = -\inf_{z^\nu \in S_\nu(x)} \varphi_\alpha^\nu(z^\nu, x)$$

with

$$\varphi_\alpha^\nu(z^\nu, x) := F^\nu(x)^T (z^\nu - x^\nu) + \frac{\alpha}{2} \|z^\nu - x^\nu\|^2$$

for all $x \in X$. In analogy to (8.2), upon defining

$$\psi_\alpha^\nu(y^\nu, x) := \varphi_\alpha^\nu(c^\nu(x) + Q^\nu(x) y^\nu, x)$$

one can show that also

$$g_\alpha^\nu(x) = -\min_{y^\nu \in K^\nu} \psi_\alpha^\nu(y^\nu, x)$$

as well as

$$\nabla g_\alpha^\nu(x) \quad = \quad -\nabla_x \psi_\alpha^\nu(y^\nu, x)\Big|_{y^\nu = y_\alpha^\nu(x)} \tag{8.7}$$

hold, where $y_\alpha^\nu(x)$ denotes the unique solution of the problem

$$\min_{y^\nu} \psi_\alpha^\nu(y^\nu, x) \quad \text{subject to} \quad y^\nu \in K^\nu.$$

Consequently, (8.4) yields the following result.

Theorem 8.7 *Consider a QVI with set-valued mapping in product form and generalized moving sets of the form* (8.6) *where, for $p_\nu \leq n_\nu$, the set $K^\nu \subseteq \mathbb{R}^{p_\nu}$ is nonempty, closed, and convex, the functions F, c^ν, and Q^ν are continuously differentiable, and $Q^\nu(x)$ has full rank for all $x \in \mathbb{R}^n$, $\nu = 1, \dots, N$. Then the function g_α is continuously differentiable with $\nabla g_\alpha(x) = \sum_{\nu=1}^N \nabla g_\alpha^\nu(x)$ and $\nabla g_\alpha^\nu(x)$ given by* (8.7)*.*

Note that, under the above assumptions, $S(x)$ can be written as a generalized moving set in the form $S(x) = c(x) + Q(x)K$ with the nonempty, closed, and convex set $K = K^1 \times \ldots \times K^N$ in the product form as well as

$$c(x) = \begin{pmatrix} c^1(x) \\ \vdots \\ c^N(x) \end{pmatrix} \quad \text{and} \quad Q(x) = \begin{pmatrix} Q^1(x) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Q^N(x) \end{pmatrix},$$

where $Q(x)$ has full rank for all $x \in \mathbb{R}^n$.

8.3. Application to Generalized Nash Equilibrium Problems

In this section we tackle some player convex GNEPs. The corresponding definitions, a literature overview, and some reformulations of these problems are given in Chapter 3. Throughout this section, we make the following smoothness and convexity assumptions.

Assumption 8.8 *(a) The functions $\theta_\nu : \mathbb{R}^n \to \mathbb{R}$, $\nu = 1, \dots, N$, are continuously differentiable.*

(b) The functions $\theta_\nu(\cdot, x^{-\nu})$, $\nu = 1, \dots, N$, are convex for all fixed $x^{-\nu} \in \mathbb{R}^{n-n_\nu}$.

(c) The sets $X_\nu(x^{-\nu})$, $\nu = 1, \dots, N$, are closed and convex for all $x \in \mathbb{R}^n$.

Under Assumption 8.8, it is well-known, see, for example, [47, 71], that a GNEP is equivalent to a QVI in the sense that $\bar{x}$ is a solution of the GNEP if and only if $\bar{x}$ solves the corresponding QVI with F being defined by

$$F^{GNEP}(x) := F(x) := \begin{pmatrix} \nabla_{x^1}\theta_1(x) \\ \vdots \\ \nabla_{x^N}\theta_N(x) \end{pmatrix}$$

and $S(x)$ having the product structure (cf. Section 8.2)

$$S(x) := X_1(x^{-1}) \times \ldots \times X_N(x^{-N}).$$

For $x \in M$ $(= \operatorname{dom} S)$ the regularized gap function of this particular QVI therefore reads

$$\begin{aligned} g_\alpha(x) &= -\inf_{z\in S(x)} \left[F^{GNEP}(x)^T(z-x) + \frac{\alpha}{2}\|z-x\|^2 \right] \\ &= -\inf_{z\in S(x)} \left[\sum_{\nu=1}^N \left(\nabla_{x^\nu}\theta_\nu(x^\nu, x^{-\nu})^T(z^\nu - x^\nu) + \frac{\alpha}{2}\|z^\nu - x^\nu\|^2 \right) \right] \\ &= -\sum_{\nu=1}^N \inf_{z^\nu \in X_\nu(x^{-\nu})} \left[\nabla_{x^\nu}\theta_\nu(x^\nu, x^{-\nu})^T(z^\nu - x^\nu) + \frac{\alpha}{2}\|z^\nu - x^\nu\|^2 \right]. \end{aligned}$$

Taking into account the convexity of θ_ν as a function of x^ν, it follows that

$$g_\alpha(x) \geq -\sum_{\nu=1}^N \inf_{z^\nu \in X_\nu(x^{-\nu})} \left[\theta_\nu(z^\nu, x^{-\nu}) - \theta_\nu(x^\nu, x^{-\nu}) + \frac{\alpha}{2}\|z^\nu - x^\nu\|^2 \right] = -\inf_{z\in S(x)} \Phi_\alpha(z,x) \;=:\; V_\alpha(x),$$

where

$$\Phi_\alpha(z,x) := \sum_{\nu=1}^N \left(\theta_\nu(z^\nu, x^{-\nu}) - \theta_\nu(x^\nu, x^{-\nu}) + \frac{\alpha}{2}\|z^\nu - x^\nu\|^2 \right). \tag{8.8}$$

The functions $-\Phi_\alpha$ and V_α defined in this way are the regularized Nikaido-Isoda function and the corresponding optimal value function, respectively, which are known from theoretical and numerical considerations in the context of GNEPs, see, for example, Section 3.2. If g_α and V_α are considered to be extended-valued, then for all $x \notin M$ we trivially have $g_\alpha(x) = V_\alpha(x) = -\infty$. We summarize the previous discussion in the following result.

Lemma 8.9 *Let Assumption 8.8 hold. Consider a QVI arising from a player convex GNEP, and let g_α and V_α be the corresponding regularized gap function and optimal value function, respectively. Then $g_\alpha(x) \geq V_\alpha(x)$ holds for all $x \in \mathbb{R}^n$.*

The previous result implies, for example, that any error bound result for V_α also gives an error bound result for the regularized gap function g_α, whereas the converse might not be true.

Next, we also study the differentiability properties of the optimal value function V_α of player convex GNEPs in the generalized moving set case $S(x) = c(x) + Q(x)K$ defined by (8.1). In fact, due to the inherent product structure of $S(x)$ in the GNEP case, we have

$$S(x) = S_1(x) \times \ldots \times S_N(x)$$

with $S_\nu(x) = X_\nu(x^{-\nu})$, $\nu = 1, \ldots, N$, so that we may use independent generalized moving sets for each player as defined in (8.6):

$$X_\nu(x^{-\nu}) = \{c^\nu(x^{-\nu}) + Q^\nu(x^{-\nu})y \mid y \in K^\nu\} \tag{8.9}$$

where, for $p_\nu \leq n_\nu$, the sets $K^\nu \subseteq \mathbb{R}^{p_\nu}$, $\nu = 1, \ldots, N$, are nonempty, closed, and convex, the functions $c^\nu : \mathbb{R}^{n-n_\nu} \to \mathbb{R}^{n_\nu}$ and $Q^\nu : \mathbb{R}^{n-n_\nu} \to \mathbb{R}^{n_\nu \times p_\nu}$ are continuous, and $Q^\nu(x^{-\nu})$ has full (column) rank for all $x^{-\nu} \in \mathbb{R}^{n-n_\nu}$. The following example is a class of player convex GNEP with (generalized) moving set which fits into this framework.

Example 8.10 Consider a GNEP with N players, $n_\nu = 1$ for all $\nu = 1, \ldots, N$, and the strategy sets $X_\nu(x^{-\nu}) := \{x^\nu \in \mathbb{R} \mid \sum_{\nu=1}^N \alpha_\nu x^\nu \leq \beta\}$, where β, $\alpha_\nu \in \mathbb{R}$ with $\alpha_\nu > 0$ for all ν. Then all sets $X_\nu(x^{-\nu})$ are (generalized) moving sets with nonempty, closed, and convex sets $K^\nu :=]-\infty, 0]$ and continuous functions $c^\nu : \mathbb{R}^{N-1} \to \mathbb{R}$, $c^\nu(x^{-\nu}) := \frac{1}{\alpha_\nu}(\beta - \sum_{\mu \neq \nu} \alpha_\mu x^\mu)$, and $Q^\nu : \mathbb{R}^{N-1} \to \mathbb{R}$, $Q^\nu(x^{-\nu}) := 1$, hence $Q^\nu(x^{-\nu})$ has full rank for all $x^{-\nu} \in \mathbb{R}^{N-1}$. In fact, we have

$$X_\nu(x^{-\nu}) = \left\{ x^\nu \in \mathbb{R} \,\middle|\, x^\nu \leq \frac{1}{\alpha_\nu}\left(\beta - \sum_{\mu \neq \nu} \alpha_\mu x^\mu\right) \right\} = \{c^\nu(x^{-\nu}) + y \mid y \in]-\infty, 0]\},$$

from which the above observation follows immediately. ◇

Note that, under the additional assumption of continuous differentiability on the functions F^{GNEP} (that is, the assumption of twice continuous differentiability on the functions θ_ν), c^ν and Q^ν, $\nu = 1, \ldots, N$, the regularized gap function g_α is continuously differentiable with known gradient by Theorem 8.7. The corresponding analysis for the optimal value function V_α is similar to the one given in Section 8.2. A first difference is that in the description

$$V_\alpha(x) = -\inf_{z \in S(x)} \Phi_\alpha(z, x)$$

the function Φ_α from (8.8) is not separable with respect to all components of z, while the function φ_α from (7.7) in the description

$$g_\alpha(x) = -\inf_{z \in S(x)} \varphi_\alpha(z, x)$$

is. However, Φ_α obviously is separable with respect to the vectors $z^1, \ldots, z^N$ which suffices to mimic the proof of continuous differentiability of g_α in Theorem 8.7 in order to show continuous differentiability of V_α. In fact, the separability allows us to write $V_\alpha(x) = \sum_{\nu=1}^N V_\alpha^\nu(x)$ with

$$V_\alpha^\nu(x) := -\inf_{z^\nu \in X_\nu(x^{-\nu})} \Phi_\alpha^\nu(z^\nu, x)$$

and

$$\Phi_\alpha^\nu(z^\nu, x) := \theta_\nu(z^\nu, x^{-\nu}) - \theta_\nu(x^\nu, x^{-\nu}) + \frac{\alpha}{2}\|z^\nu - x^\nu\|^2$$

or, equivalently,

$$V_\alpha^\nu(x) = -\min_{y^\nu \in K^\nu} \Psi_\alpha^\nu(y^\nu, x)$$

with

$$\Psi_\alpha^\nu(y^\nu, x) := \Phi_\alpha^\nu(c^\nu(x^{-\nu}) + Q^\nu(x^{-\nu})y^\nu, x)$$

for $\nu = 1, \ldots, N$. As a second difference to the analysis of the gap function, the strong convexity of Ψ_α^ν in y^ν is slightly less apparent. In fact, the convexity of Φ_α^ν in z^ν implies the convexity of Ψ_α^ν in y^ν. Moreover, by the full rank of $Q^\nu(x^{-\nu})$, the matrix

$$\nabla^2_{y^\nu y^\nu}\Psi_\alpha^\nu(y^\nu, z) = Q^\nu(x^{-\nu})^T \left(\nabla^2_{z^\nu z^\nu}\Phi_\alpha^\nu(z^\nu, x)|_{z^\nu = c^\nu(x^{-\nu}) + Q^\nu(x^{-\nu})y^\nu}\right) Q^\nu(x^{-\nu})$$

with

$$\nabla^2_{z^\nu z^\nu}\Phi_\alpha^\nu(z^\nu, x) = \nabla^2_{z^\nu z^\nu}\theta_\nu(z^\nu, x^{-\nu}) + \alpha I_{n_\nu}$$

is uniformly positive definite (in y^ν), so that Ψ_α^ν even is strongly convex in y^ν. Therefore, for each $\nu \in \{1, \ldots, N\}$ the problem

$$\min_{y^\nu} \Psi_\alpha^\nu(y^\nu, x) \quad \text{subject to} \quad y^\nu \in K^\nu$$

has a unique solution $y_\alpha^\nu(x)$, and along the lines of Section 8.2 we obtain that V_α^ν is continuously differentiable with

$$\nabla V_\alpha^\nu(x) = -\nabla_x \Psi_\alpha^\nu(y^\nu, x)|_{y^\nu = y_\alpha^\nu(x)} \tag{8.10}$$

where

$$\begin{aligned}
\nabla_{x^\nu}\Psi_\alpha^\nu(y^\nu, x) &= -\nabla_{x^\nu}\theta(x^\nu, x^{-\nu}) - \alpha\left(c^\nu(x^{-\nu}) + Q^\nu(x^{-\nu})y^\nu - x^\nu\right),\\
\nabla_{x^{-\nu}}\Psi_\alpha^\nu(y^\nu, x) &= \Big[\nabla_{x^{-\nu}}\theta_\nu(z^\nu, x^{-\nu}) - \nabla_{x^{-\nu}}\theta_\nu(x^\nu, x^{-\nu}) + \\
&\quad + \left(\nabla_{x^{-\nu}}c^\nu(x^{-\nu}) + \nabla_{x^{-\nu}}(Q^\nu(x^{-\nu})y^\nu)\right)\left(\nabla_{x^\nu}\theta_\nu(z^\nu, x^{-\nu}) + \alpha(z^\nu - x^\nu)\right)\Big]_{z^\nu = c^\nu(x^{-\nu}) + Q^\nu(x^{-\nu})y^\nu}.
\end{aligned}$$

The following theorem summarizes the previous discussion.

Theorem 8.11 *Consider a GNEP with strategy spaces of generalized moving set form* (8.9) *where, for $p_\nu \leq n_\nu$, the sets $K^\nu \subseteq \mathbb{R}^{p_\nu}$ are nonempty, closed, and convex, the functions θ_ν are twice continuously differentiable, the functions c^ν and Q^ν are continuously differentiable, and $Q^\nu(x^{-\nu})$ has full rank for all $x^{-\nu} \in \mathbb{R}^{n-n_\nu}$, $\nu = 1, \ldots, N$. Then V_α is continuously differentiable with $\nabla V_\alpha(x) = \sum_{\nu=1}^N \nabla V_\alpha^\nu(x)$ and $\nabla V_\alpha^\nu(x)$ given by* (8.10).

9. Smoothness Properties of a Primal Gap Function for Quasi-Variational Inequalities

In Chapter 9 we turn back to the general QVI, where the regularized gap function is typically nonsmooth. Hence we investigate its continuity properties in Section 9.1 under suitable assumptions. Then we discuss the differentiability properties of the gap function in Section 9.2. Our main result of Section 9.2 is that, apart from special cases, all local minimizers of the reformulation are differentiability points of the gap function. The results and analysis of this chapter are similar to that in Chapter 4 and are published in [75].

We assume in Chapter 9 that $S(x)$ has the representation

$$S(x) = \{z \in \mathbb{R}^n \mid \; s_i(z, x) \leq 0 \;\; \forall i = 1, \ldots, m\}$$

with suitable functions $s_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$. Then the feasible set X is given by

$$X = \{x \in \mathbb{R}^n \mid s_i(x, x) \leq 0 \;\; \forall i = 1, \ldots, m\}.$$

Throughout this chapter, we make the following smoothness and convexity assumptions.

Assumption 9.1 *(a) The function F is continuous on $\mathbb{R}^n$.*

(b) The functions s_i, $i = 1, \ldots, m$, are continuous on $\mathbb{R}^n \times \mathbb{R}^n$.

(c) The functions $s_i(\cdot, x)$, $i = 1, \ldots, m$, are convex for each fixed $x \in \mathbb{R}^n$.

Note that, in particular, Assumptions 9.1 (b), (c) guarantee that $S(x)$ is indeed a closed and convex (possibly empty) set for any given $x \in \mathbb{R}^n$.

9.1. Continuity Properties and Domain

In the first part of this section we show that the solution function z_α of the problem (7.8) is continuous at $\bar{x} \in M = \operatorname{dom} S$ if $S(\bar{x})$ satisfies the Slater condition, that is, if there exists some $\bar{z} \in \mathbb{R}^n$ satisfying $s_i(\bar{z}, \bar{x}) < 0$ for all $i = 1, \ldots, m$. We therefore define the 'degenerate point set'

$$D_1 := \{x \in M \mid \text{the set } S(x) \text{ violates the Slater condition}\}.$$

Note that continuity of z_α at $\bar{x}$, in particular, implies the continuity of the regularized gap function g_α at $\bar{x}$. The corresponding analysis is similar to the one given in [38] and in Chapter 4 for certain objective functions arising in the context of jointly and player convex GNEPs, respectively. After two generalizations of our main result, we then study a topological property of the set $M \setminus D_1$.

In this section we use Definition 2.1 where the definitions of an inner semicontinuous (isc) and an outer semicontinuous (osc) set-valued mapping are given.

Theorem 9.2 *Let Assumption 9.1 hold, and let the set-valued mapping S be isc at $\bar{x} \in M$. Then the functions z_α and g_α are continuous at $\bar{x}$.*

Proof. Recall that $\varphi_\alpha(\cdot, x)$ is convex for each fixed $x \in \mathbb{R}^n$ and continuous on $\mathbb{R}^n \times \mathbb{R}^n$. Therefore, the function $-\varphi_\alpha(\cdot, x)$ is concave for each fixed $x \in \mathbb{R}^n$ and continuous on $\mathbb{R}^n \times \mathbb{R}^n$.

The set-valued mapping S is osc, since its graph

$$\operatorname{gph} S := \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^n \mid s_i(z, x) \leq 0 \ \ \forall i = 1, \dots, m\}$$

is a closed set in view of continuity of s_i, $i = 1, \dots, m$, see [82, Theorem 2]. Due to Assumption 9.1, $S(x)$ is convex for all $x \in \mathbb{R}^n$. Additionally, the set

$$Z_\alpha(x) = \left\{\zeta \in S(x) \,\middle|\, \max_{z \in S(x)} \left(-\varphi_\alpha(z, x)\right) = -\varphi_\alpha(\zeta, x)\right\}$$

is a singleton with the unique element $z_\alpha(x)$ for all $x \in M$, see Remark 7.1. Therefore, Lemma 2.3 implies that the set-valued mapping $x \mapsto \{z_\alpha(x)\}$ is continuous at $\bar{x}$. Hence the function $x \mapsto z_\alpha(x)$ is continuous at $\bar{x}$. Furthermore, $g_\alpha(x) = -\varphi_\alpha(z_\alpha(x), x)$ is continuous at $\bar{x}$ as a composition of continuous functions. □

As an immediate consequence of Theorem 9.2, we obtain the following result.

Corollary 9.3 *Let Assumption 9.1 hold. Then z_α and g_α are continuous on $M \setminus D_1$.*

Proof. Let $\bar{x} \in M \setminus D_1$. Due to Assumption 9.1 and the Slater condition for $S(\bar{x})$, the set-valued mapping S is isc at $\bar{x}$ (see [82, Theorem 12]). Therefore, Theorem 9.2 implies that the functions z_α and g_α are continuous at $\bar{x}$. □

Let us illustrate the previous result in the context of Example 7.3.

Example 9.4 Note that Assumption 9.1 is fulfilled in Example 7.3. In the situation of Example 7.3, for $x \in M$ the set $S(x) = S_1(x) \times S_2(x)$ satisfies the Slater condition if and only if $S_1(x)$ as well as $S_2(x)$ possess a Slater point. Clearly, $S_1(x)$ satisfies the Slater condition for all $x \in M$. On the other hand, $S_2(x)$ violates the Slater condition exactly for all x with $x_1 = -1/\sqrt{2}$ and for all x with $x_1 = 1$. Hence we obtain $D_1 = \left(\{-1/\sqrt{2}\} \cup \{1\}\right) \times \mathbb{R}$ and, by Corollary 9.3, the functions z_α and g_α are continuous on $M \setminus D_1 =]-1/\sqrt{2}, 1[\times \mathbb{R}$. ◇

Direct inspection of the functions z_α and g_α from Example 7.3 shows that they are actually continuous at least on all of X ($\subseteq M$) relative to X, see Definition 2.2. This motivates to relax the assumption of Corollary 9.3 in the spirit of [39, Theorem 3.5] for generalized Nash equilibrium problems. To this end, let us define the set

$$D_1' := \{x \in M \mid \text{the set } S(x) \text{ violates the Slater condition and is not a singleton}\}.$$

The following result shows that z_α and hence also g_α are continuous on the set $X \setminus D_1'$ (relative to X), that is, they are continuous at every point $x \in X$ (relative to X) where $S(x)$ either satisfies the Slater condition or reduces to a single point. Note that the latter degenerate case occurs quite frequently, for example, in the context of GNEPs.

Theorem 9.5 *Let Assumption 9.1 hold. Then z_α and g_α are continuous on $X \setminus D_1'$ (relative to X).*

Proof. Let $\bar{x} \in X \setminus D_1'$. In view of $X \subseteq M$ and Corollary 9.3, we only have to consider the case that $S(\bar{x})$ is a singleton. Due to $\bar{x} \in X$, we actually have $S(\bar{x}) = \{\bar{x}\}$. Choose any sequence $\{x^k\} \subseteq X$ with $x^k \to \bar{x}$. Then for each $k \in \mathbb{N}$ we have $x^k \in S(x^k)$, so that S turns out to be isc at $\bar{x}$ (relative to X). Theorem 9.2 now yields the assertion. □

Unfortunately, in Example 7.3 we obtain $X \cap D_1 = X \cap D_1' = \{(1, 0)\}$ as $S(1, 0) = [0, +\infty[\times \{0\}$ violates the Slater condition while not being a singleton. Hence Theorem 9.5 may not be evoked to show continuity of z_α and g_α on all of X (relative to X). However, the product form of the set-valued mapping S in Example 7.3 justifies to modify the assumptions of Theorem 9.5. Let us consider the general case of a set-valued mapping in product form (cf. Section 8.2)

$$S(x) = S_1(x) \times S_2(x) \times \ldots \times S_N(x)$$

and define

$$\begin{aligned} D_1'' := \{x \in M \mid &\text{for some } \nu \in \{1, \ldots, N\} \text{ the set } S_\nu(x) \text{ violates the Slater condition} \\ &\text{and is not a singleton}\}. \end{aligned}$$

Recall that this product structure of $S(x)$ arises quite naturally in the GNEP context (see Section 8.3).

Theorem 9.6 *Let Assumption 9.1 hold, and let S be given in product form. Then the functions z_α and g_α are continuous on $X \setminus D_1''$ (relative to X).*

Proof. Let $\bar{x} \in X \setminus D_1''$. Then for each $\nu \in \{1, \ldots, N\}$ the set $S_\nu(\bar{x})$ either satisfies the Slater condition or coincides with the singleton $\{\bar{x}^\nu\}$. Choose any sequence $\{x^k\} \subseteq X$ with $x^k \to \bar{x}$ and any $\bar{z} \in S(\bar{x})$, that is, we have $x^{\nu,k} \to \bar{x}^\nu$ and $\bar{z}^\nu \in S_\nu(\bar{x})$, $\nu = 1, \ldots, N$. For those $\nu \in \{1, \ldots, N\}$ with $S_\nu(\bar{x})$ satisfying the Slater condition, the set-valued mapping S_ν is isc at $\bar{x}$, so that for sufficiently large k a sequence $z^{\nu,k} \in S_\nu(x^k)$ with $z^{\nu,k} \to \bar{z}^\nu$ exists. On the other hand, for each ν with $S_\nu(\bar{x}) = \{\bar{x}^\nu\}$, we may choose $z^{\nu,k} := x^{\nu,k} \in S_\nu(x^k)$ and obtain $z^{\nu,k} = x^{\nu,k} \to \bar{x}^\nu = \bar{z}^\nu$ as in the proof of Theorem 9.5. This shows the inner semicontinuity of S at $\bar{x}$ (relative to X), and Theorem 9.2 yields the assertion. □

Note that we have $X \setminus D_1'' = X$ in Example 7.3, so that Theorem 9.6 finally yields the continuity of z_α and g_α on all of X (relative to X).

Let us return to the set D_1 which is also important in our analysis of the differentiability properties of g_α. In fact, in Section 9.2 we shall study differentiability properties of g_α at points from the topological interior of the domain of g_α where, in view of Remark 7.1,

$$\operatorname{dom} g_\alpha = \{x \in \mathbb{R}^n \mid g_\alpha(x) \in \mathbb{R}\}$$

coincides with M. Therefore, their topological interiors also coincide:

$$\operatorname{int} \operatorname{dom} g_\alpha = \operatorname{int} M. \tag{9.1}$$

The following result relates the set $M \setminus D_1$ to the interior of the domain of g_α.

Lemma 9.7 *Let Assumption 9.1 hold. Then the set $M \setminus D_1$ is open and satisfies*

$$M \setminus D_1 \subseteq \operatorname{int} \operatorname{dom} g_\alpha.$$

Proof. Let $\bar{x} \in M \setminus D_1$. Then there exists some $\bar{z} \in \mathbb{R}^n$ satisfying $s_i(\bar{z}, \bar{x}) < 0$ for all $i = 1, \ldots, m$. Due to continuity of the functions s_i, $i = 1, \ldots, m$, we can choose a neighborhood U of $\bar{x}$ such that for all $x \in U$ also $s_i(\bar{z}, x) < 0$ is satisfied for all $i = 1, \ldots, m$. Therefore, for all $x \in U$ the set $S(x)$ satisfies the Slater condition, that is, we have $x \in M \setminus D_1$. This shows that $M \setminus D_1$ is open. In particular, U is contained in $\operatorname{dom} S = M$. This implies $\bar{x} \in \operatorname{int} M$ and, due to (9.1), shows the second assertion. □

As a last point of this section, we remark that Lemma 9.7 lays the foundation for the analysis in Section 9.2.

Remark 9.8 Lemma 9.7 guarantees that the set $\operatorname{dom} g_\alpha \setminus D_1$ is an open subset of $\operatorname{int} \operatorname{dom} g_\alpha$, so that we will be able to study differentiability of g_α on $\operatorname{dom} g_\alpha \setminus D_1$ in Section 9.2. We point out that, under stronger convexity and regularity assumptions, along the lines of Theorem 4.11 one can also show the reverse inclusion in Lemma 9.7, that is, the topological boundary of $\operatorname{dom} g_\alpha$ coincides with D_1. An illustration of this result is given in Example 9.4. ◇

9.2. Differentiability Properties

Assumption 9.1 and the following Assumption 9.9 are the blanket assumptions for this section.

Assumption 9.9 *The functions F and s_i, $i = 1, \ldots, m$, are continuously differentiable.*

In this section we want to study differentiability properties of g_α. To this end, we have to make sure that we consider differentiability only at points in the interior of the domain of g_α, since otherwise it makes no sense to talk about differentiability. In view of Lemma 9.7, it is reasonable to investigate the differentiability of the function g_α on the set $M \setminus D_1$. To this end, consider once again the convex optimization problem from (7.8). In view of Remark 7.1, this problem has a unique optimal point $z_\alpha(x)$ for all $x \in M$, in particular, for all $x \in M \setminus D_1$. Let

$$L_\alpha(x, z, \lambda) := \varphi_\alpha(z, x) + \sum_{i=1}^{m} \lambda_i s_i(z, x) = F(x)^T (z - x) + \frac{\alpha}{2}\|z - x\|^2 + \sum_{i=1}^{m} \lambda_i s_i(z, x)$$

denote the Lagrange function of the optimization problem (7.8), and let

$$KKT_\alpha(x) := \Big\{\lambda \in \mathbb{R}^m \;\Big|\; F(x) + \alpha(z_\alpha(x) - x) + \sum_{i=1}^m \lambda_i \nabla_z s_i(z_\alpha(x), x) = 0,$$
$$\lambda_i \geq 0,\ \lambda_i s_i(z_\alpha(x), x) = 0 \ \ \forall i = 1, \ldots, m\Big\}$$

be the set of Karush-Kuhn-Tucker multipliers for $z_\alpha(x) \in S(x)$. Note that the convex polyhedron $KKT_\alpha(x)$ is a convex and nonempty polytope if and only if $S(x)$ satisfies the Slater condition [65], that is, for $x \in M \setminus D_1$. Furthermore,

$$\mathcal{I}_\alpha(x) := \{i \in \{1, \ldots, m\} \mid s_i(z_\alpha(x), x) = 0\}$$

will denote the set of active indices of $z_\alpha(x) \in S(x)$.

Besides the standard concept of differentiability, this section is based on several differentiability concepts (directional differentiability, Hadamard directional differentiability, and Gâteaux differentiability) summarized in Definition 2.8.

Theorem 9.10 *Let Assumptions 9.1 and 9.9 hold, and let $x \in M \setminus D_1$. Then the regularized gap function g_α is Hadamard directionally differentiable at x with*

$$g'_\alpha(x, d) = \min_{\lambda \in KKT_\alpha(x)} \left[\left(F(x) - (\nabla F(x) - \alpha I_n)(z_\alpha(x) - x) - \sum_{i=1}^m \lambda_i \nabla_x s_i(z_\alpha(x), x) \right)^T d \right] \tag{9.2}$$

for any $d \in \mathbb{R}^n$.

Proof. Since $x \in M \setminus D_1$, the set $S(x)$ satisfies the Slater condition. A standard result from parametric optimization (see, e.g., [69, 83, 116]) then states that the optimal value function $-g_\alpha$ of the optimization problem (7.8) is Hadamard directionally differentiable at x with

$$\begin{aligned}(-g_\alpha)'(x, d) &= \max_{\lambda \in KKT_\alpha(x)} \left(\nabla_x L_\alpha(x, z, \lambda)|_{z = z_\alpha(x)} \right)^T d \\ &= \max_{\lambda \in KKT_\alpha(x)} \left[d^T \Big(- F(x) + (\nabla F(x) - \alpha I_n)(z - x) + \sum_{i=1}^m \lambda_i \nabla_x s_i(z, x) \Big)\Big|_{z = z_\alpha(x)} \right]\end{aligned}$$

for any $d \in \mathbb{R}^n$. This shows the assertion. □

Remark 9.11 Note that, in the assertion of Theorem 9.10 and in the following, for any $x \in M \setminus D_1$ and any $\lambda \in KKT_\alpha(x)$ one may replace the term $\sum_{i=1}^m \lambda_i \nabla_x s_i(z_\alpha(x), x)$ by $\sum_{i \in \mathcal{I}_\alpha(x)} \lambda_i \nabla_x s_i(z_\alpha(x), x)$, since $\lambda_i = 0$ for all $i \in \{1, \ldots, m\} \setminus \mathcal{I}_\alpha(x)$. ◇

The formula (9.2) for the directional derivative of g_α at some $x \in M \setminus D_1$ simplifies if not only the optimal point set $\{z_\alpha(x)\}$ of (7.8) but also the set $KKT_\alpha(x)$ is a singleton. This motivates to define a next 'degenerate point set'

$$D_2 := \{x \in M \mid \text{the set } KKT_\alpha(x) \text{ is not a singleton}\}.$$

As mentioned before, the convex polyhedron $KKT_\alpha(x)$ is a convex and nonempty polytope if and only if $x \in M \setminus D_1$. Hence for $x \in D_1$ the set $KKT_\alpha(x)$ is either empty or unbounded but certainly not a singleton. This shows the relation

$$D_1 \subseteq D_2. \tag{9.3}$$

Theorem 9.10 and (9.3) lead to the following result.

Corollary 9.12 *Let Assumptions 9.1 and 9.9 hold, and let $x \in M \setminus D_2$ with $KKT_\alpha(x) = \{\lambda_\alpha(x)\}$. Then the regularized gap function g_α is Gâteaux differentiable at x with*

$$g'_\alpha(x, d) = \Big(F(x) - (\nabla F(x) - \alpha I_n)(z_\alpha(x) - x) - \sum_{i=1}^{m} (\lambda_\alpha(x))_i \nabla_x s_i(z_\alpha(x), x)\Big)^T d \tag{9.4}$$

for any $d \in \mathbb{R}^n$.

For algebraic characterizations of the sets D_1 and D_2 we use the following definitions of some conditions on QVIs.

Definition 9.13 *(a) The* Mangasarian Fromovitz constraint qualification (MFCQ) *holds at $z_\alpha(x) \in S(x)$ if there exists a vector $d \in \mathbb{R}^n$ satisfying $\nabla_z s_i(z_\alpha(x), x)^T d < 0$ for all $i \in \mathcal{I}_\alpha(x)$.*

(b) The strict Mangasarian Fromovitz condition (SMFC) *holds at $z_\alpha(x)$ in $S(x)$ with a multiplier $\lambda \in KKT_\alpha(x)$ if the gradients*

$$\nabla_z s_i(z_\alpha(x), x) \quad (i \in \mathcal{I}^+_\alpha(x) = \{i \in \mathcal{I}_\alpha(x) \mid \lambda_i > 0\}),$$

are linearly independent, and there exists a vector $d \in \mathbb{R}^n$ satisfying

$$\begin{aligned} \nabla_z s_i(z_\alpha(x), x)^T d < 0 \quad & \forall i \in \mathcal{I}^0_\alpha(x) = \{i \in \mathcal{I}_\alpha(x) \mid \lambda_i = 0\}, \\ \nabla_z s_i(z_\alpha(x), x)^T d = 0 \quad & \forall i \in \mathcal{I}^+_\alpha(x). \end{aligned}$$

(c) The linear independence constraint qualification (LICQ) *holds at $z_\alpha(x) \in S(x)$ if the vectors $\nabla_z s_i(z_\alpha(x), x)$ $(i \in \mathcal{I}_\alpha(x))$ are linearly independent.*

Note that, because of the convexity of the functions $s_i(\cdot, x)$ $(i = 1, \ldots, m)$ for each fixed x, MFCQ holds at $z_\alpha(x)$ if and only if the Slater condition is satisfied for $S(x)$. Hence we have the characterization

$$D_1 = \{x \in M \mid \text{MFCQ is violated at } z_\alpha(x) \in S(x)\}.$$

Furthermore, it is known from [88] that SMFC at $z_\alpha(x) \in S(x)$ characterizes a unique KKT multiplier $\lambda_\alpha(x)$ at the optimal point $z_\alpha(x)$. Therefore, we arrive at

$$D_2 = \{x \in M \mid \text{either } KKT_\alpha(x) = \emptyset \text{ or SMFC is violated at } z_\alpha(x) \in S(x)\}.$$

Finally, as LICQ implies SMFC at $z_\alpha(x) \in S(x)$, the set

$$D_3 := \{x \in M \mid \text{LICQ is violated at } z_\alpha(x) \in S(x)\}$$

satisfies

$$D_1 \subseteq D_2 \subseteq D_3. \tag{9.5}$$

For the proof of the next result recall that if a function $f : U \to \mathbb{R}$ with an open domain U is Gâteaux differentiable on U and the partial derivatives of f are continuous at $\bar{x} \in U$, then f is continuously differentiable at $\bar{x}$.

Theorem 9.14 *Let Assumptions 9.1 and 9.9 hold, and let $\bar{x} \in M \setminus D_3$ with $KKT_\alpha(\bar{x}) = \{\lambda_\alpha(\bar{x})\}$. Then $KKT_\alpha(x) = \{\lambda_\alpha(x)\}$ holds in a neighborhood of $\bar{x}$, and the regularized gap function g_α is continuously differentiable on this neighborhood of $\bar{x}$ with*

$$\nabla g_\alpha(x) = F(x) - (\nabla F(x) - \alpha I_n)(z_\alpha(x) - x) - \sum_{i=1}^{m} (\lambda_\alpha(x))_i \nabla_x s_i(z_\alpha(x), x).$$

Proof. First, due to (9.5) and Lemma 9.7, $\bar{x}$ is an interior point of $\operatorname{dom} g_\alpha$, and there is some neighborhood U of $\bar{x}$ such that for all $x \in U$ the optimal point $z_\alpha(x) \in S(x)$ satisfies the Slater condition. By Corollary 9.3, the function z_α is actually continuous on U. Consequently, since LICQ is stable under perturbations, U may be chosen such that LICQ holds at $z_\alpha(x) \in S(x)$ for each $x \in U$. This implies that KKT_α is single-valued on U, say $KKT_\alpha(x) = \{\lambda_\alpha(x)\}$ for $x \in U$. Corollary 9.12 thus guarantees that g_α is Gâteaux differentiable on U with (9.4). By [83, Lemma 2] the set-valued mapping KKT_α is locally bounded and osc on U. As it is also single-valued in our case, the function λ_α is continuous on U, so that the partial derivatives of g_α are continuous at $\bar{x}$. This shows continuous differentiability of g_α at $\bar{x}$ with the asserted gradient. Since the partial derivatives of g_α actually are continuous on all of U, also continuous differentiability of g_α on U follows. □

The next remark shows that the assumption $\bar{x} \in M \setminus D_3$ in Theorem 9.14 can be replaced by $\bar{x} \in M \setminus D_2$ and the additional assumption of stable SMFC.

Remark 9.15 The main reason to use D_3 instead of the smaller set D_2 in the assumption of Theorem 9.14 is the lack of stability of SMFC (cf. also Example 9.16 below). On the other hand, a different sufficient condition for continuous differentiability of g_α on a neighborhood of $\bar{x} \in M \setminus D_2$ can be obtained in cases when SMFC is stable at $z_\alpha(\bar{x}) \in S(\bar{x})$. Note that, since $\bar{x} \in M \setminus D_2$, SMFC already holds at $z_\alpha(\bar{x}) \in S(\bar{x})$ and yields the unique multiplier $\lambda_\alpha(\bar{x})$. If additionally the set $\mathcal{I}^0_\alpha(\bar{x}) = \left\{ i \in \mathcal{I}_\alpha(\bar{x}) \mid (\lambda_\alpha(\bar{x}))_i = 0 \right\}$ remains constant under small perturbations of $\bar{x}$ (this is true if, for example, the set $\mathcal{I}^0_\alpha(\bar{x})$ is empty, that is, the so-called *strict complementarity slackness* holds), then continuity arguments show that SMFC is stable at $z_\alpha(x)$ under sufficiently small perturbations of $\bar{x}$. After this observation, one can show continuous differentiability of g_α on a neighborhood of $\bar{x}$ along the lines of the proof of Theorem 9.14. ◇

Let us illustrate our results for the QVI from Example 7.3 and check differentiability properties of the regularized gap function g_α on $X \setminus D_1$.

Example 9.16 Consider the QVI from Example 7.3. Note that Assumptions 9.1 and 9.9 hold for this example. By Theorem 9.14, g_α is continuously differentiable at each $x \in X \setminus D_3$ with known gradient. In the following, we will determine the sets $X \cap (D_3 \setminus D_1)$ and $X \cap (D_2 \setminus D_1)$ as well as the corresponding directional derivatives of g_α.

By definition of D_3 one has

$$X \cap (D_3 \setminus D_1) = \{x \in X \setminus D_1 \mid \text{LICQ is violated at } z_\alpha(x) \text{ in } S(x)\}$$

so that we have to check for violation of LICQ. The involved gradients are

$$\nabla_z s_1(z_\alpha(x), x) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \quad \nabla_z s_2(z_\alpha(x), x) = \begin{pmatrix} 0 \\ 2(z_\alpha(x))_2 \end{pmatrix}, \quad \nabla_z s_3(z_\alpha(x), x) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Some tedious calculations show that the activities are characterized as follows, where we use the functions ϱ_i from Example 7.3:

$$\begin{aligned}
\{x \in X \setminus D_1 \mid 1 \in \mathcal{I}_\alpha(x)\} &= \{x \in X \setminus D_1 \mid \varrho_1(x) \le 1/\alpha\}, \\
\{x \in X \setminus D_1 \mid 2 \in \mathcal{I}_\alpha(x)\} &= \left\{x \in X \setminus D_1 \mid \varrho_2(x) \le 1/\alpha,\ x_1 \ge 1/\sqrt{2}\right\} \\
&\quad \cup \{x \in X \setminus D_1 \mid 1/\alpha \le \varrho_3(x)\}, \\
\{x \in X \setminus D_1 \mid 3 \in \mathcal{I}_\alpha(x)\} &= \left\{x \in X \setminus D_1 \mid \varrho_2(x) \le 1/\alpha,\ x_1 \le 1/\sqrt{2}\right\}.
\end{aligned}$$

In particular, if $2 \in \mathcal{I}_\alpha(x)$, then for all $x \in X \setminus D_1$ with $\varrho_2(x) \le 1/\alpha$ and $x_1 \ge 1/\sqrt{2}$ we find

$$\nabla_z s_2(z_\alpha(x), x) = \begin{pmatrix} 0 \\ -2\sqrt{1 - x_1^2} \end{pmatrix} \ne 0,$$

and for all $x \in X \setminus D_1$ with $1/\alpha \le \varrho_3(x)$

$$\nabla_z s_2(z_\alpha(x), x) = \begin{pmatrix} 0 \\ 2\sqrt{1 - x_1^2} \end{pmatrix} \ne 0,$$

so that

$$X \cap (D_3 \setminus D_1) = \{x \in X \setminus D_1 \mid \{2, 3\} \subseteq \mathcal{I}_\alpha(x)\}.$$

As $\varrho_3(x) < \varrho_2(x)$ holds for all $x \in X \setminus D_1$, this implies

$$\begin{aligned}
X \cap (D_3 \setminus D_1) &= \left\{x \in X \setminus D_1 \,\middle|\, \varrho_2(x) \le \frac{1}{\alpha},\ x_1 = \frac{1}{\sqrt{2}}\right\} \\
&= \left\{x \in X \setminus D_1 \,\middle|\, x_2 + \frac{1}{\sqrt{2}} \le \frac{1}{\alpha},\ x_1 = \frac{1}{\sqrt{2}}\right\} \\
&= \left\{\frac{1}{\sqrt{2}}\right\} \times \left[-\frac{1}{\sqrt{2}}, \min\left\{\frac{1}{\sqrt{2}}, \frac{1}{\alpha} - \frac{1}{\sqrt{2}}\right\}\right].
\end{aligned}$$

For $\alpha \le 1/\sqrt{2}$ this results in

$$X \cap (D_3 \setminus D_1) = \left\{x \in X \setminus D_1 \mid x_1 = 1/\sqrt{2}\right\}$$

and corresponds to a 'concave kink in the graph of g_α on X along the line segment connecting the boundary points $\left(1/\sqrt{2}, -1/\sqrt{2}\right)$ and $\left(1/\sqrt{2}, 1/\sqrt{2}\right)$ of X'. This fact becomes apparent below.

The example exhibits a more interesting feature, however, for $\alpha > 1/\sqrt{2}$ when

$$X \cap (D_3 \setminus D_1) = \left\{\frac{1}{\sqrt{2}}\right\} \times \left[-\frac{1}{\sqrt{2}}, \frac{1}{\alpha} - \frac{1}{\sqrt{2}}\right].$$

In the following we will see that this corresponds to a 'concave kink in the graph of g_α on X along the line segment connecting the boundary point $\left(1/\sqrt{2}, -1/\sqrt{2}\right)$ and the *interior* point $\left(1/\sqrt{2}, -1/\sqrt{2} + 1/\alpha\right)$ of X'. For $\alpha = 1$ $\left(> 1/\sqrt{2}\right)$, this 'kink' is visualized in Figure 7.2. For simplicity, in the remainder of this example, let us focus on the case $\alpha = 1$ with

$$X \cap (D_3 \setminus D_1) = \left\{x(t) := \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} + t\right) \;\middle|\; t \in [0, 1]\right\}.$$

To identify the set $X \cap (D_2 \setminus D_1)$, we compute the sets $KKT_1(x(t))$ for $t \in [0, 1]$. It is not hard to see that $1 \in \mathcal{I}_1(x(t))$ if and only if $t \geq 3/\sqrt{2} - 2$. Some more computations show that

$$KKT_1(x(t)) = \left\{(1-s)\begin{pmatrix} 0 \\ \frac{1-t}{\sqrt{2}} \\ 0 \end{pmatrix} + s\begin{pmatrix} 0 \\ 0 \\ 1-t \end{pmatrix} \;\middle|\; s \in [0, 1]\right\}$$

for all $t \in \left[0, 3/\sqrt{2} - 2\right[$, and

$$KKT_1(x(t)) = \left\{(1-s)\begin{pmatrix} \frac{1}{2}\left(1 - \frac{3}{2\sqrt{2}} + \frac{t}{2}\right) \\ \frac{1-t}{\sqrt{2}} \\ 0 \end{pmatrix} + s\begin{pmatrix} \frac{1}{2}\left(1 - \frac{3}{2\sqrt{2}} + \frac{t}{2}\right) \\ 0 \\ 1-t \end{pmatrix} \;\middle|\; s \in [0, 1]\right\}$$

for all $t \in \left[3/\sqrt{2} - 2, 1\right]$. Hence $KKT_1(x(t))$ contains more than one multiplier for all $t \in [0, 1[$, whereas $KKT_1(x(1))$ is a *singleton*. In other words, for $t = 1$, that is, at 'the interior end point of the kink' $x(1) = \left(1/\sqrt{2}, 1 - 1/\sqrt{2}\right)$, SMFC holds at $y_1(x(1))$ in $S(x(1))$ while LICQ is violated. We arrive at

$$X \cap (D_2 \setminus D_1) = \left\{x(t) := \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} + t\right) \;\middle|\; t \in [0, 1[\right\}.$$

In particular, by Corollary 9.12, g_1 is Gâteaux differentiable at $x(1)$, but SMFC is *unstable* at $y_1(x(1))$ in $S(x(1))$, as it is violated at $y_1(x(t))$ in $S(x(t))$ with $t < 1$. In the following we shall see that, indeed, g_1 is not Gâteaux differentiable at $x(t)$ with $t < 1$. To this end, we compute the Hadamard directional derivatives of g_1 at $x(t)$ with the formula from Theorem 9.10. The appearing derivatives are

$$\nabla F(x) = 0,\ \nabla_x s_1(x, y_1(x)) = \begin{pmatrix} 0 \\ 1 \end{pmatrix},\ \nabla_x s_2(x, y_1(x)) = \begin{pmatrix} 2x_1 \\ 0 \end{pmatrix},\ \nabla_x s_3(x, y_1(x)) = \begin{pmatrix} -1 \\ 0 \end{pmatrix},$$

and for $d \in \mathbb{R}^n$, we obtain

$$g_1'(x(t), d) = (1-t)\,(d_2 - |d_1|)$$

for all $t \in \left[0, 3/\sqrt{2} - 2\right[$ as well as

$$g_1'(x(t), d) = \left(1 - \frac{3}{2\sqrt{2}} + \frac{t}{2}\right) d_1 - \frac{1}{2}\left(1 - \frac{3}{2\sqrt{2}} + \frac{t}{2}\right) d_2 + (1 - t)(d_2 - |d_1|)$$

for all $t \in \left[3/\sqrt{2} - 2, 1\right[$. This shows that g_1 is not Gâteaux differentiable at $x(t)$ with $t < 1$, but that a 'concave kink' occurs in the graph of g_1 along $X \cap (D_2 \setminus D_1)$. Additionally, at $x(1)$ we have

$$g_1'(x(1), d) = \frac{3}{2}\left(1 - \frac{1}{\sqrt{2}}\right)\left(d_1 - \frac{d_2}{2}\right)$$

for all $d \in \mathbb{R}^n$.

We point out that the main argument in the proof of Theorem 9.14 needs Gâteaux differentiability of g_1 not only at the point under consideration but also on a neighborhood of this point. In the present example, Gâteaux differentiability of g_1 at $x(1)$ does not extend to a neighborhood of this point. ◇

The observed differentiability properties in Example 9.16 particularly guarantee that any local minimizer $\bar{x}$ of g_α on X lies in D_1, or g_α is at least Gâteaux differentiable at $\bar{x}$, where usually even continuous differentiability occurs at $\bar{x}$. In the sequel we will show that, under mild assumptions, this also holds in the general case.

To this end, we use the *linearization cone* to $X = \{x \in \mathbb{R}^n \mid s_i(x, x) \leq 0 \; \forall i = 1, \ldots, m\}$ at a point x, which is easily seen to be given by

$$\mathcal{L}_X(x) := \left\{d \in \mathbb{R}^n \mid (\nabla_x s_i(x, x) + \nabla_z s_i(x, x))^T d \leq 0 \quad \forall i \in I_0(x)\right\}$$

with the active index set

$$I_0(x) := \{i \in \{1, \ldots, m\} \mid s_i(x, x) = 0\}.$$

Furthermore, we define the 'degenerate point set' D_4 as a set of points in D_2 with

$$\operatorname{span}\left\{\nabla_x s_i(z_\alpha(x), x) \quad (i \in \mathcal{I}_\alpha(x))\right\} \cap \operatorname{span}\left\{\nabla_x s_i(x, x) + \nabla_z s_i(x, x) \quad (i \in I_0(x))\right\} \neq \{0\}. \tag{9.6}$$

Therefore, we have

$$D_4 := \{x \in D_2 \mid (9.6) \text{ holds for } z_\alpha(x) \in S(x)\}.$$

For the next result we need the following assumption.

Assumption 9.17 *The vectors $\nabla_x s_i(z_\alpha(x), x)$ $(i \in \mathcal{I}_\alpha(x))$ are linearly independent for all vectors $x \in D_2 \setminus (D_1 \cup D_4)$.*

Note that Assumption 9.17 is not to be confused with LICQ at $z_\alpha(x) \in S(x)$, as here the gradients are taken with respect to x.

Proposition 9.18 *Let Assumptions 9.1, 9.9, and 9.17 hold, and let $\bar{x} \in D_2 \setminus (D_1 \cup D_4)$. Then there exists a vector $d \in \mathbb{R}^n$ solving the system*

$$g_\alpha'(\bar{x}, d) < 0, \quad (\nabla_x s_i(\bar{x}, \bar{x}) + \nabla_z s_i(\bar{x}, \bar{x}))^T d \leq 0 \quad (i \in I_0(\bar{x})). \tag{9.7}$$

Proof. Assume that (9.7) does not possess a solution $d \in \mathbb{R}^n$. By Theorem 9.10, this implies the inconsistency of

$$\Big(F(\bar{x}) - (\nabla F(\bar{x}) - \alpha I_n)(z_\alpha(\bar{x}) - \bar{x}) - \sum_{i \in \mathcal{I}_\alpha(\bar{x})} \lambda_i \nabla_x s_i(z_\alpha(\bar{x}), \bar{x})\Big)^T d < 0,$$
$$(\nabla_x s_i(\bar{x}, \bar{x}) + \nabla_z s_i(\bar{x}, \bar{x}))^T d \leq 0 \quad (\, i \in I_0(\bar{x})),$$

for any $\lambda \in KKT_\alpha(\bar{x})$. By Farkas' lemma (see, e.g., [120]), this system is inconsistent if and only if there exist scalars $\gamma_i(\lambda) \geq 0$, $i \in I_0(\bar{x})$, with

$$\begin{aligned} & F(\bar{x}) - (\nabla F(\bar{x}) - \alpha I_n)(z_\alpha(\bar{x}) - \bar{x}) - \sum_{i \in \mathcal{I}_\alpha(\bar{x})} \lambda_i \nabla_x s_i(z_\alpha(\bar{x}), \bar{x}) \\ & + \sum_{i \in I_0(\bar{x})} \gamma_i\,(\lambda)\,(\nabla_x s_i(\bar{x}, \bar{x}) + \nabla_z s_i(\bar{x}, \bar{x})) = 0. \end{aligned} \tag{9.8}$$

Because of $\bar{x} \in D_2 \setminus D_1$, there exist two different multipliers $\hat{\lambda} \neq \tilde{\lambda}$ with $\hat{\lambda}, \tilde{\lambda} \in KKT_\alpha(\bar{x})$. Then equation (9.8) holds for $\lambda = \hat{\lambda}$ as well as for $\lambda = \tilde{\lambda}$. Subtracting and rearranging these two equations leads to

$$\sum_{i \in I_0(\bar{x})} \Big(\gamma_i(\hat{\lambda}) - \gamma_i(\tilde{\lambda})\Big)(\nabla_x s_i(\bar{x}, \bar{x}) + \nabla_z s_i(\bar{x}, \bar{x})) = \sum_{i \in \mathcal{I}_\alpha(\bar{x})} (\hat{\lambda}_i - \tilde{\lambda}_i) \nabla_x s_i(z_\alpha(\bar{x}), \bar{x}),$$

where the left hand side is some element of

$$\operatorname{span}\Big\{\nabla_x s_i(\bar{x}, \bar{x}) + \nabla_z s_i(\bar{x}, \bar{x}) \quad (\, i \in I_0(\bar{x}))\Big\},$$

and the right hand side is some element of

$$\operatorname{span}\Big\{\nabla_x s_i(z_\alpha(\bar{x}), \bar{x}) \quad (i \in \mathcal{I}_\alpha(\bar{x}))\Big\}.$$

The right hand side cannot be trivial in view of $\hat{\lambda} \neq \tilde{\lambda}$ and Assumption 9.17. Hence (9.6) holds, which is a contradiction to $\bar{x} \in D_2 \setminus D_4$. Therefore, our assumption is wrong, and there exists a vector $d \in \mathbb{R}^n$ solving the system (9.7). □

Before we present the main result of this section, we recall that the *tangent* (or *contingent* or *Bouligand*) *cone* to X at point x is defined by

$$\mathcal{T}_X(x) := \Big\{d \in \mathbb{R}^n \;\Big|\; \exists t_k \downarrow 0,\ d^k \to d :\ x + t_k d^k \in X \text{ for all } k \in \mathbb{N}\Big\}.$$

It is well-known that the relation $\mathcal{T}_X(x) \subseteq \mathcal{L}_X(x)$ always holds (see, e.g., [125]), and the *Abadie constraint qualification* (ACQ) is said to hold at $x \in X$ if $\mathcal{T}_X(x) = \mathcal{L}_X(x)$.

Assumption 9.19 *ACQ holds for all* $x \in D_2 \setminus (D_1 \cup D_4)$.

The following theorem is the main result of this section.

Theorem 9.20 *Let Assumptions 9.1, 9.9, 9.17, and 9.19 hold. Then any local minimizer $\bar{x}$ of g_α on X lies in $D_1 \cup D_4$, or g_α is at least Gâteaux differentiable at $\bar{x}$. If, in the latter case, LICQ holds at $z_\alpha(\bar{x}) \in S(\bar{x})$, then g_α is continuously differentiable at $\bar{x}$.*

Proof. Let $\bar{x}$ be a local minimizer of g_α on X. We distinguish the cases $\bar{x} \in D_2$ and $\bar{x} \in X \setminus D_2$.

First let $\bar{x} \in D_2$. Then $\bar{x} \in D_1 \cup D_4$ holds or, by Proposition 9.18, there exists a vector $d \in \mathbb{R}^n$ solving the system (9.7). We shall show that the latter leads to a contradiction. In fact, because of $(\nabla_x s_i(\bar{x}, \bar{x}) + \nabla_z s_i(\bar{x}, \bar{x}))^T d \leq 0$ for all $i \in I_0(\bar{x})$, this vector d is an element of the linearization cone $\mathcal{L}_X(\bar{x})$. Due to Assumption 9.19, d also belongs to the tangent cone $\mathcal{T}_X(\bar{x})$. Hence there exist sequences $t_k \downarrow 0$ and $d^k \to d$ with $\bar{x} + t_k d^k \in X$ for all $k \in \mathbb{N}$. As $\bar{x}$ is a local minimizer of g_α on X, we have $g_\alpha(\bar{x} + t_k d^k) \geq g_\alpha(\bar{x})$ and

$$\frac{g_\alpha(\bar{x} + t_k d^k) - g_\alpha(\bar{x})}{t_k} \geq 0 \tag{9.9}$$

for all sufficiently large $k \in \mathbb{N}$. By Theorem 9.10, the function g_α is Hadamard directionally differentiable at $\bar{x}$. Hence the limit of the left-hand side in (9.9) exists and is equal to $g'_\alpha(\bar{x}, d)$ (note that just directionally differentiability in the ordinary sense is not sufficient for this implication). Consequently, it holds $g'_\alpha(\bar{x}, d) \geq 0$. This is a contradiction to (9.7).

In the second case, let $\bar{x} \in X \setminus D_2$. In view of Corollary 9.12 and (7.5), g_α is Gâteaux differentiable at $\bar{x}$. This completes the proof of the first part of the assertion.

The second part immediately follows from Theorem 9.14. □

We conclude this section with a corollary, which covers the case with the affine-linear constraint functions s_i, $i = 1, \ldots, N$.

Corollary 9.21 *Let Assumptions 9.1, 9.9, 9.17 hold, and assume that all constraint functions s_i are affine-linear. Then any local minimizer $\bar{x}$ of g_α on X lies in $D_1 \cup D_4$, or the function g_α is at least Gâteaux differentiable at $\bar{x}$. If, in the latter case, LICQ holds at $z_\alpha(\bar{x}) \in S(\bar{x})$, then g_α is continuously differentiable at $\bar{x}$.*

Proof. Due to affine linearity of all constraint functions s_i, ACQ holds everywhere in X (see, e.g., [125]). Then Theorem 9.20 yields the statements. □

10. Smoothness Properties of a Dual Gap Function for Quasi-Variational Inequalities

The primal gap function $g_\alpha + \delta_X$ for the reformulation of a QVI in (7.10) is, in general, nonsmooth like the primal gap function g_α. Nevertheless, Dietrich observed in [30] that this regularized gap function may be viewed as a difference of two convex functions. This reformulation then allows the application of the Toland-Singer duality theory [123, 130, 131] in order to obtain a dual gap function that gives a smooth reformulation for a class of QVIs. Therefore, we elaborate further on this approach in Chapter 10. In Section 10.1 we remove the (implicit) assumption from [30] that the set-valued mapping defining the QVI is always nonempty-valued and derive the dual gap function and its basic properties adjusted to our QVI setting. Furthermore, we verify some stronger smoothness properties of this dual reformulation in Section 10.2. The results of this chapter have already been published in [72].

10.1. A Smooth Dual Gap Function

Recall that a QVI consists in finding a vector $x \in S(x)$ with

$$F(x)^T(z - x) \geq 0 \qquad \forall z \in S(x), \qquad \text{(defined in (7.1))}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a function and $S(x)$ is a closed and convex (possibly empty) set for any $x \in \mathbb{R}^n$, that the fixed point set of S is given by

$$X := \{x \in \mathbb{R}^n \mid x \in S(x)\}, \qquad \text{(defined in (7.3))}$$

and that the domain of S is the set

$$M = \operatorname{dom} S = \{x \in \mathbb{R}^n \mid S(x) \neq \emptyset\}. \qquad \text{(defined in (7.4))}$$

In this section we rewrite the objective function

$$g_\alpha + \delta_X = -\inf_{z \in S(x)} \left[F(x)^T(z - x) + \frac{\alpha}{2}\|z - x\|^2 \right] + \delta_X$$

of the optimization problem

$$\min_{x \in \mathbb{R}^n} [g_\alpha(x) + \delta_X(x)] \qquad \text{(defined in (7.10))}$$

as a difference of two strongly convex functions. Then we deduce from this reformulation a dual reformulation of a QVI as an unconstrained and smooth optimization problem. Furthermore, we study the basic properties of this dual reformulation.

Our first goal is to rewrite the objective function $g_\alpha + \delta_X$ of the optimization problem (7.10) as a difference of two convex functions. To this end, we have to make some assumptions on the class of QVIs that we are going to deal with. In this assumption we use the definition of a graph-convex and osc set-valued mapping given in Definition 2.1.

Assumption 10.1 *(a) The feasible set X of the QVI* (7.1) *defined in* (7.3) *is nonempty.*

(b) The function $F : \mathbb{R}^n \to \mathbb{R}^n$ is given by $F(x) = Ax + b$ with $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$.

(c) The set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is graph-convex and osc on $\mathbb{R}^n$.

Assumption 10.1 (a) does not limit the application of our theory, since otherwise the QVI would not have a solution. On the other hand, Assumptions 10.1 (b) and (c) are more restrictive in the sense that we consider only QVIs with an affine operator F and suitable set-valued mappings S. Note that Assumption 10.1 (b) can be relaxed as it will be mentioned in Remark 10.5.

There are immediate consequences of Assumption 10.1 summarized in the following result.

Lemma 10.2 *Suppose that Assumptions 10.1 (a) and (c) hold. Then:*

(a) The set X from (7.3) *is nonempty, closed, and convex.*

(b) The set M from (7.4) *is nonempty and convex.*

Proof. (a) The set X is nonempty by Assumption 10.1 (a). In order to show that X is also closed, let $\{x^k\} \subseteq X$ be an arbitrary sequence with $x^k \to \bar{x}$ for some $\bar{x} \in \mathbb{R}^n$. Then we have $x^k \in S(x^k)$ for all $k \in \mathbb{N}$. Since the set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is osc by Assumption 10.1 (c), it follows that $\bar{x} \in S(\bar{x})$. Hence $\bar{x} \in X$ so that X is a closed set.

We next show that X is also convex. To this end, let $x_1, x_2 \in X$ and $t \in [0, 1]$ be arbitrarily given. Then $x_1 \in S(x_1)$ and $x_2 \in S(x_2)$. Using the assumed graph-convexity of S together with Lemma 2.4 (b) yields $tx_1 + (1-t)x_2 \in S(tx_1 + (1-t)x_2)$. This means that $tx_1 + (1-t)x_2 \in X$, that is, X is a convex set.

(b) By Assumption 10.1 (a), there exists an element $x \in X$ which means that $x \in S(x)$, hence $x \in M$, so that M is nonempty.

Finally, we come to the convexity of M. Let $x_1, x_2 \in M$ and $t \in [0, 1]$ be given. Then $S(x_1) \neq \emptyset$ and $S(x_2) \neq \emptyset$, hence there exist elements $z_1 \in S(x_1)$ and $z_2 \in S(x_2)$. Using Assumption 10.1 (c) together with Lemma 2.4 (b), this implies

$$tz_1 + (1-t)z_2 \in S(tx_1 + (1-t)x_2).$$

Consequently, the set on the right-hand side is nonempty, that is, $tx_1 + (1-t)x_2 \in M$. □

It is worth mentioning that, even for an osc and graph-convex set-valued mapping S, its domain is not necessarily closed as illustrated by the subsequent example.

Example 10.3 Let $S : \mathbb{R} \rightrightarrows \mathbb{R}$ be given by

$$S(x) := \begin{cases} \left\{ y \in \mathbb{R} \,\middle|\, y \geq \frac{1}{x} \right\}, & \text{if } x > 0, \\ \emptyset, & \text{if } x \leq 0. \end{cases}$$

Obviously, S is graph-convex. Additionally, S is osc, since, if $x_k \downarrow 0$, a sequence $\{z_k\} \subseteq \mathbb{R}$ with $z_k \in S(x_k)$ is divergent, and all other cases are unproblematic. On the other hand, the set $M = \operatorname{dom} S = \mathbb{R}_>$ is not closed. ◇

We next follow an observation by Dietrich [30] and reformulate the unconstrained objective function from problem (7.10) explicitly as a difference of two convex functions, that is, we obtain a DC minimization problem, see [84] for a survey of DC programming. Having this DC formulation, it is pretty straightforward to obtain a reformulation as a difference of two strongly convex functions. Then we may invoke the duality theory by Toland [130] and Singer [123] in order to derive a smooth dual formulation of the original QVI.

The basic step to obtain a DC formulation is the rearrangement of the regularized gap function:

$$\begin{aligned} g_\alpha(x) &= -\inf_{z \in S(x)} \left[-\frac{1}{2\alpha}\|F(x)\|^2 + \frac{\alpha}{2}\left(\|z - x\|^2 + \frac{2}{\alpha}F(x)^T(z - x) + \frac{1}{\alpha^2}\|F(x)\|^2 \right) \right] \\ &= \frac{1}{2\alpha}\|F(x)\|^2 - \frac{\alpha}{2} \inf_{z \in S(x)} \left\| z - \left(x - \frac{1}{\alpha}F(x) \right) \right\|^2 \qquad (10.1) \\ &= \frac{1}{2\alpha}\|F(x)\|^2 - \Phi_\alpha(x) \qquad (10.2) \end{aligned}$$

with the function $\Phi_\alpha : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\begin{aligned} \Phi_\alpha(x) &:= \frac{\alpha}{2} \inf_{z \in S(x)} \left\| z - \left(x - \frac{1}{\alpha}F(x) \right) \right\|^2 \qquad (10.3) \\ &= \begin{cases} \frac{\alpha}{2} \left\| P_{S(x)}\left(x - \frac{1}{\alpha}F(x) \right) - \left(x - \frac{1}{\alpha}F(x) \right) \right\|^2, & \text{if } x \in M, \\ +\infty, & \text{if } x \notin M, \end{cases} \qquad (10.4) \end{aligned}$$

where the projection $P_{S(x)}(y)$ of y onto the set $S(x)$ is well-defined for all $x \in M = \operatorname{dom} S$, since the set $S(x)$ is nonempty, closed, and convex in view of Assumption 10.1 (c) and Lemma 2.4 (a).

Our next goal is to prove that Φ_α is a convex and lsc function, see Definition 2.5 (a). For these purposes, the auxiliary Lemma 5.5 is pivotal. This lemma shows that the function

$$\Psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, \quad \Psi(z, x) := \delta_{S(x)}(z)$$

is lsc and convex in (z, x) for a graph-convex and osc set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$.

Lemma 10.4 *Let Assumption 10.1 hold. Then the function Φ_α is lsc and convex on $\mathbb{R}^n$.*

Proof. In view of (10.3), we may rewrite Φ_α as

$$\Phi_\alpha(x) = \inf_{z \in \mathbb{R}^n} f(z, x),$$

where $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is given by

$$f(z, x) := \frac{\alpha}{2} \left\| z - \left(x - \frac{1}{\alpha} F(x) \right) \right\|^2 + \delta_{S(x)}(z).$$

The first summand of f is convex as it is the composition of the convex function $\frac{\alpha}{2} \| \cdot \|^2$ and an affine mapping, see, e.g., [117, Ex. 2.20]. Moreover, the first summand is, in particular, continuous. The second summand is lsc and convex due to Lemma 5.5, hence f is lsc and convex (and proper, since $M \neq \emptyset$). Moreover, it holds that

$$\operatorname*{argmin}_z f(z, x) = \left\{ P_{S(x)}\left(x - \frac{1}{\alpha} F(x) \right) \right\} \quad \forall x \in M$$

is single-valued. Since $M \neq \emptyset$, the assertions therefore follow from [117, Cor. 3.32]. □

Note that Lemma 10.4 exploits the definition (10.3) of the mapping Φ_α in order to verify that it is both lsc and convex. Alternatively, one might try to use the representation (10.4) to rewrite Φ_α in the form

$$\Phi_\alpha(x) = \frac{\alpha}{2} \left\| P_{S(x)}\left(x - \frac{1}{\alpha} F(x) \right) - \left(x - \frac{1}{\alpha} F(x) \right) \right\|^2 + \delta_M(x).$$

This formulation can indeed be used to show convexity of Φ_α, but the verification of the lower semicontinuity is more difficult, especially since M is not necessarily closed, hence this formulation is, in general, not the sum of two lsc functions.

The following remark is a comment on Assumption 10.1 (b).

Remark 10.5 We would like to point out that the proof of Lemma 10.4 exploits, for the first time, the assumption that $F(x) = Ax + b$ is an affine mapping, since it uses the fact that the composition of an outer convex function with an inner affine-linear function remains convex. Similar situations will also arise in the subsequent analysis, and it is clear that there exist more general classes of functions F which have this property, but in order to avoid any technical conditions and to concentrate on the main ideas of our approach, Assumption 10.1 (b) takes F as an affine-linear function. ◇

In view of Lemma 10.4 and Assumption 10.1 (b), the representation (10.2) gives an explicit formulation of the regularized gap function as a DC optimization problem. In order to obtain better smoothness properties in a corresponding dual formulation, we add and substract a simple strongly convex quadratic term. This gives us the following DC decomposition of the unconstrained objective function from (7.10):

$$g_\alpha(x) + \delta_X(x) = f_\alpha(x) - h_\alpha(x)$$

with the two functions $f_\alpha, h_\alpha : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f_\alpha(x) := \frac{\alpha}{2} \|x\|^2 + \frac{1}{2\alpha} \|F(x)\|^2 + \delta_X(x) \quad \text{and} \quad h_\alpha(x) := \frac{\alpha}{2} \|x\|^2 + \Phi_\alpha(x). \tag{10.5}$$

We summarize the previous discussion in the following result.

Lemma 10.6 *Let Assumption 10.1 hold, and let f_α, h_α be defined as in* (10.5). *Then:*

(a) The function f_α is lsc and convex on $\mathbb{R}^n$ and strongly convex on its domain $\operatorname{dom} f_\alpha = X$.

(b) The function h_α is lsc and convex on $\mathbb{R}^n$ and strongly convex on its domain $\operatorname{dom} h_\alpha = M$.

(c) The vector $\bar{x}$ is a solution of the QVI if and only if it is a solution of the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} [f_\alpha(x) - h_\alpha(x)]$$

with optimal function value equal to zero.

We next want to apply the duality theory by Toland and Singer. This theory involves the conjugates of the two functions f_α and h_α, see Definition 2.5 (e). We therefore give explicit expressions for these two conjugate functions in the next two results.

Lemma 10.7 *Let Assumption 10.1 hold. Define*

$$Q_\alpha := \alpha\left(I_n + \frac{1}{\alpha^2}A^T A\right), \quad q_\alpha := \frac{1}{\alpha}A^T b, \quad c_\alpha := \frac{1}{2\alpha}\|b\|^2, \quad \|x\|_{Q_\alpha} := \sqrt{x^T Q_\alpha x}. \tag{10.6}$$

Then the following statements hold for the conjugate f_α^* of the function f_α:

(a) f_α^* is given by

$$f_\alpha^*(y) = \frac{1}{2}\left\|Q_\alpha^{-1}(y - q_\alpha)\right\|_{Q_\alpha}^2 - \frac{1}{2}\left\|Q_\alpha^{-1}(y - q_\alpha) - x_\alpha^{f^*}(y)\right\|_{Q_\alpha}^2 - c_\alpha$$

where $x_\alpha^{f^*}(y)$ denotes the unique solution of the minimization problem

$$\min_x \frac{1}{2}\left\|Q_\alpha^{-1}(y - q_\alpha) - x\right\|_{Q_\alpha}^2 \quad \textit{subject to} \quad x \in X, \tag{10.7}$$

that is, $x_\alpha^{f^*}(y)$ is the projection of the vector $Q_\alpha^{-1}(y - q_\alpha)$ onto the set X with respect to the Q_α-norm.

(b) f_α^* has the domain $\operatorname{dom} f_\alpha^* = \mathbb{R}^n$.

(c) f_α^* is differentiable with Lipschitz gradient given by $\nabla f_\alpha^*(y) = x_\alpha^{f^*}(y)$.

Proof. Using Definition 2.5 (e) and the notation from (10.6), we obtain

$$\begin{aligned} f_\alpha^*(y) &= \sup_{x \in \mathbb{R}^n}\left[x^T y - \frac{\alpha}{2}\|x\|^2 - \frac{1}{2\alpha}\|F(x)\|^2 - \delta_X(x)\right] \\ &= \sup_{x \in \mathbb{R}^n}\left[x^T y - \frac{\alpha}{2}x^T\left(I_n + \frac{1}{\alpha^2}A^T A\right)x - \frac{1}{\alpha}b^T A x - \frac{1}{2\alpha}\|b\|^2 - \delta_X(x)\right] \\ &= \sup_{x \in \mathbb{R}^n}\left[x^T(y - q_\alpha) - \frac{1}{2}\|x\|_{Q_\alpha}^2 - c_\alpha - \delta_X(x)\right] \end{aligned}$$

$$= \sup_{x \in X} \left[\frac{1}{2} \left\| Q_\alpha^{-1}(y - q_\alpha) \right\|_{Q_\alpha}^2 - c_\alpha - \frac{1}{2} \left\| Q_\alpha^{-1}(y - q_\alpha) - x \right\|_{Q_\alpha}^2 \right]$$
$$= \frac{1}{2} \left\| Q_\alpha^{-1}(y - q_\alpha) \right\|_{Q_\alpha}^2 - c_\alpha - \frac{1}{2} \min_{x \in X} \left\| Q_\alpha^{-1}(y - q_\alpha) - x \right\|_{Q_\alpha}^2 .$$

Since the set X is nonempty, closed, and convex by Lemma 10.2 (a), and taking into account that the matrix Q_α is positive definite, the minimization problem (10.7) has a unique solution $x_\alpha^{f^*}(y)$ for all $y \in \mathbb{R}^n$. By definition, this solution is simply the projection of the vector $Q_\alpha^{-1}(y - q_\alpha)$ onto the set X with respect to the Q_α-norm and therefore known to be well-defined for all $y \in \mathbb{R}^n$, so that $\operatorname{dom} f_\alpha^* = \mathbb{R}^n$. This proves statements (a) and (b).

Part (c) can be derived as follows: Using the continuity of the projection operator, it follows that the mapping $y \mapsto x_\alpha^{f^*}(y)$ is continuous. Therefore, application of Danskin's Theorem gives that f_α^* is continuously differentiable and directly yields $\nabla f_\alpha^*(y) = x_\alpha^{f^*}(y)$, cf. also the subsequent proof where a similar statement is carried out in some more detail. The fact that ∇f_α^* is even Lipschitz follows directly from Lemma 2.7. □

The following result computes the conjugate function of h_α and states some additional properties in the same spirit as in the previous result for the function f_α.

Lemma 10.8 *Let Assumption 10.1 hold. Then the following statements hold for the conjugate h_α^* of the function h_α:*

(a) $h_\alpha^*(y)$ is given by

$$h_\alpha^*(y) = \frac{1}{2\alpha} \|y\|^2 - \frac{\alpha}{2} \left\| x_\alpha^{h^*}(y) - \frac{1}{\alpha} y \right\|^2 - \frac{\alpha}{2} \left\| z_\alpha^{h^*}(y) - \left(x_\alpha^{h^*}(y) - \frac{1}{\alpha} F\left(x_\alpha^{h^*}(y) \right) \right) \right\|^2 \tag{10.8}$$

where $(x_\alpha^{h^*}, z_\alpha^{h^*})(y)$ is the unique solution of the minimization problem

$$\min_{(x,z)} \left[\left\| x - \frac{1}{\alpha} y \right\|^2 + \left\| z - \left(x - \frac{1}{\alpha} F(x) \right) \right\|^2 \right] \quad \textit{subject to} \quad (x, z) \in \operatorname{gph} S.$$

(b) $h_\alpha^*(y)$ has the domain $\operatorname{dom} h_\alpha^* = \mathbb{R}^n$.

(c) $h_\alpha^*(y)$ is differentiable with Lipschitz gradient given by $\nabla h_\alpha^*(y) = x_\alpha^{h^*}(y)$.

Proof. Using Definition 2.5 (e), we have

$$h_\alpha^*(y) = \sup_{x \in \mathbb{R}^n} \left[x^T y - \frac{\alpha}{2} \|x\|^2 - \Phi_\alpha(x) \right]$$
$$= \sup_{x \in \mathbb{R}^n} \left[\frac{1}{2\alpha} \|y\|^2 - \frac{\alpha}{2} \left(\|x\|^2 - \frac{2}{\alpha} x^T y + \frac{1}{\alpha^2} \|y\|^2 \right) - \Phi_\alpha(x) \right]$$
$$= \frac{1}{2\alpha} \|y\|^2 - \inf_{x \in \mathbb{R}^n} \frac{\alpha}{2} \left[\left\| x - \frac{1}{\alpha} y \right\|^2 + \inf_{z \in S(x)} \left\| z - \left(x - \frac{1}{\alpha} F(x) \right) \right\|^2 \right] \tag{10.9}$$
$$= \frac{1}{2\alpha} \|y\|^2 - \min_{(x,z) \in \operatorname{gph} S} \frac{\alpha}{2} \left[\left\| x - \frac{1}{\alpha} y \right\|^2 + \left\| z - \left(x - \frac{1}{\alpha} F(x) \right) \right\|^2 \right]. \tag{10.10}$$

Recall that, by Assumption 10.1 (a) and (c) and Lemma 2.4 (c), the set gph S is nonempty, closed, and convex. Furthermore, the mapping

$$\varphi_\alpha(x,y,z) := \left\| x - \frac{1}{\alpha} y \right\|^2 + \left\| z - \left(x - \frac{1}{\alpha} F(x) \right) \right\|^2$$

is strongly convex in (x,z) (uniformly in y), since

$$\nabla^2_{(x,z)(x,z)} \varphi(x,y,z) = 2 \begin{pmatrix} 2I_n + \frac{1}{\alpha^2} A^T A - \frac{1}{\alpha}(A^T + A) & -I_n + \frac{1}{\alpha} A^T \\ -I_n + \frac{1}{\alpha} A & I_n \end{pmatrix} =: B_\alpha$$

is positive definite in (x,z) (uniformly in y) because we have

$$\begin{pmatrix} v^T & w^T \end{pmatrix} B_\alpha \begin{pmatrix} v \\ w \end{pmatrix} = 2 \left[\|v\|^2 + \left\| w - v + \frac{1}{\alpha} A v \right\|^2 \right] \geq 0 \quad \text{and}$$

$$\begin{pmatrix} v^T & w^T \end{pmatrix} B_\alpha \begin{pmatrix} v \\ w \end{pmatrix} = 0 \text{ if and only if } (v,w) = 0.$$

Hence the infimum in (10.10) is uniquely attained for all $y \in \mathbb{R}^n$. We denote this unique solution by $(x_\alpha^{h^*}, z_\alpha^{h^*})(y)$ and obtain (10.8) and dom $h_\alpha^* = \mathbb{R}^n$. This proves statements (a) and (b).

Furthermore, the continuous differentiability of the conjugate convex function h_α^* follows from Lemma 2.7. Alternatively, we may invoke Lemma 2.3 based on [82, Corollaries 8.1 and 9.1] to see that the mapping $y \mapsto (x_\alpha^{h^*}, z_\alpha^{h^*})(y)$ is continuous, which together with Danskin's Theorem can be used to see that h_α^* is indeed continuously differentiable, with gradient given by

$$\nabla h_\alpha^*(y) = \frac{1}{\alpha} y - \frac{\alpha}{2} \nabla_y \varphi(x,y,z) \Big|_{(x,z)=(x_\alpha^{h^*}, z_\alpha^{h^*})(y)} = \frac{1}{\alpha} y + x_\alpha^{h^*}(y) - \frac{1}{\alpha} y = x_\alpha^{h^*}(y).$$

The fact that ∇h_α^* is Lipschitz is due to Lemma 2.7. This completes the proof. □

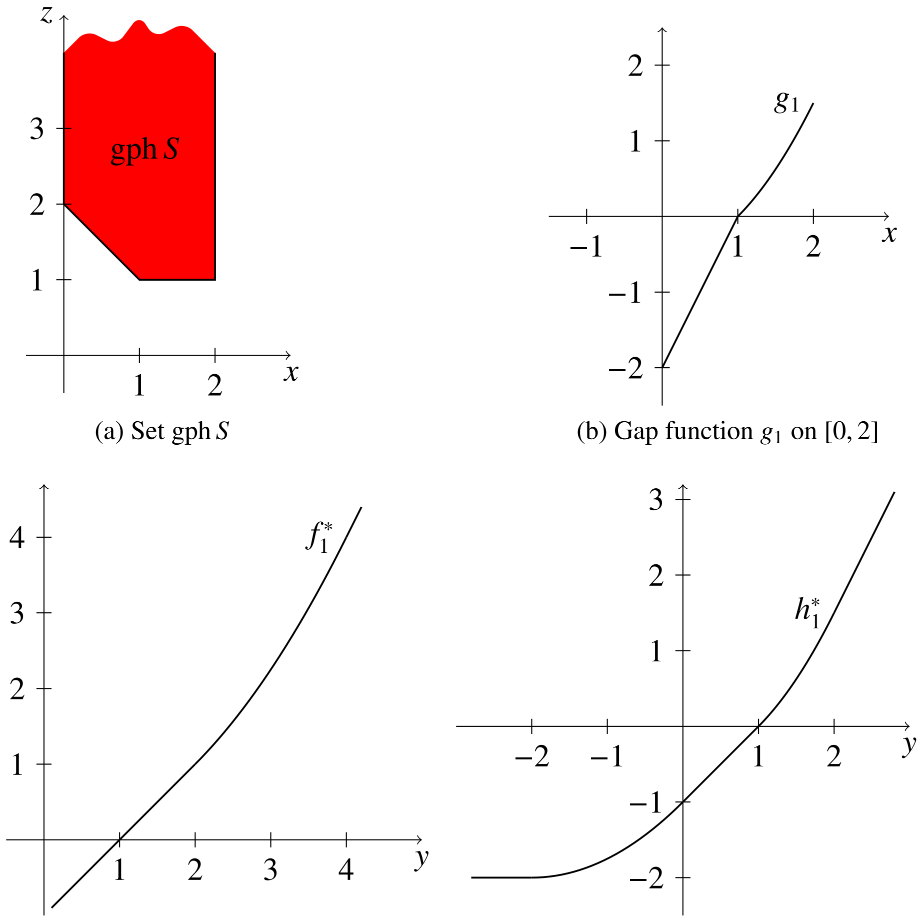
In order to illustrate the two previous and the subsequent results, we consider a simple example.

Example 10.9 Consider the QVI with $n = 1$, $F(x) = x$, and

$$S(x) = \begin{cases} [-x+2, \infty[, & \text{if } x \in [0,1], \\ [1, \infty[, & \text{if } x \in \,]1,2], \\ \emptyset, & \text{if } x \notin [0,2]. \end{cases}$$

Note that $M = [0,2]$ is the domain of S in this example, that $X = [1,2]$ is the feasible set, and that all conditions from Assumption 10.1 are satisfied. Let $\alpha = 1$. Using (10.2), we may write the corresponding regularized gap function g_1 as

$$g_1(x) = \frac{1}{2} x^2 - \frac{1}{2} \inf_{z \in S(x)} z^2 = \begin{cases} 2x - 2, & \text{if } x \in [0,1], \\ \frac{1}{2}(x^2 - 1), & \text{if } x \in \,]1,2], \\ -\infty, & \text{if } x \notin [0,2], \end{cases}$$



(a) Set gph S (b) Gap function g_1 on $[0, 2]$

(c) Conjugate function f_1^* (d) Conjugate function h_1^*

Figure 10.1.: Illustrations for Example 10.9

since

$$z_1(x) = \begin{cases} -x + 2, & \text{if } x \in [0, 1], \\ 1, & \text{if } x \in]1, 2], \\ \text{nonexistent}, & \text{if } x \notin [0, 2], \end{cases}$$

holds. The graph of the set-valued mapping S and the graph of the function g_1 on $\operatorname{dom} g_1 = [0, 2]$ are illustrated in Figures 10.1 (a) and (b), respectively. We see that the function g_1 is zero only at $x = 1$, hence this point is the unique solution of the QVI, but g_1 has a 'kink' precisely at this solution point. On the other hand, we consider the DC decomposition of g_1 with the functions

$$f_1(x) = x^2 + \delta_X(x) \quad \text{and} \quad h_1(x) = \frac{1}{2}x^2 + \frac{1}{2} \inf_{z \in S(x)} z^2.$$

Using Definition 2.5 (e) yields the conjugate function of f_1 illustrated in Figure 10.1 (c):

$$f_1^*(y) = \sup_{x\in[1,2]} \left[xy - x^2\right] = \begin{cases} y-1, & \text{if } y < 2,\\ \frac{1}{4}y^2, & \text{if } y \in [2,4],\\ 2y-4, & \text{if } y > 4, \end{cases} \quad \text{since} \quad x_1^{f^*}(y) = \begin{cases} 1, & \text{if } y < 2,\\ \frac{1}{2}y, & \text{if } y \in [2,4],\\ 2, & \text{if } y > 4. \end{cases}$$

In order to obtain the conjugate function of h_2, we rewrite the set-valued mapping S of this QVI with the help of inequality constraints

$$\begin{aligned} S(x) = \{z \in \mathbb{R} \mid s_1(z,x) &:= -x \le 0,\\ s_2(z,x) &:= -z - x + 2 \le 0,\\ s_3(z,x) &:= -z + 1 \le 0,\\ s_4(z,x) &:= x - 2 \le 0\}. \end{aligned} \tag{10.11}$$

Due to Definition 2.5 (e), we have

$$h_1^*(y) = \sup_{(x,z)\in \operatorname{gph} S} \left[xy - \frac{1}{2}x^2 - \frac{1}{2}z^2\right] = \frac{1}{2}y^2 - \min_{(x,z)\in \operatorname{gph} S} \left[\frac{1}{2}(x-y)^2 + \frac{1}{2}z^2\right], \tag{10.12}$$

where the minimization problem in (10.12) is strongly convex and differentiable. This problem has the following Lagrange function:

$$L_1^{h^*}(x,z,\lambda,y) = \frac{1}{2}(x-y)^2 + \frac{1}{2}z^2 - \lambda_1 x + \lambda_2(-z-x+2) + \lambda_3(-z+1) + \lambda_4(x-2).$$

Then the KKT conditions of this problem for $(x,z) \in \operatorname{gph} S$ are

$$\nabla_{(x,z)} L_1^{h^*}(x,z,\lambda,y) = \begin{pmatrix} x - y - \lambda_1 - \lambda_2 + \lambda_4 \\ z - \lambda_2 - \lambda_3 \end{pmatrix} = 0 \quad \text{and} \quad \begin{cases} \lambda_1 \ge 0,\ \lambda_1 x = 0,\\ \lambda_2 \ge 0,\ \lambda_2(-z-x+2) = 0,\\ \lambda_3 \ge 0,\ \lambda_3(-z+1) = 0,\\ \lambda_4 \ge 0,\ \lambda_4(x-2) = 0. \end{cases}$$

There are five possibilities for a solution of the minimization problem in (10.12):

Case 1: $\begin{cases} -x = 0,\\ z + x = 2,\\ -z + 1 < 0,\\ x - 2 < 0, \end{cases}$ Case 2: $\begin{cases} -x < 0,\\ z + x = 2,\\ -z + 1 < 0,\\ x - 2 < 0, \end{cases}$ Case 3: $\begin{cases} -x < 0,\\ z + x = 2,\\ -z + 1 = 0,\\ x - 2 < 0, \end{cases}$ Case 4: $\begin{cases} -x < 0,\\ z + x > 2,\\ -z + 1 = 0,\\ x - 2 < 0, \end{cases}$ Case 5: $\begin{cases} -x < 0,\\ z + x > 2,\\ -z + 1 = 0,\\ x - 2 = 0. \end{cases}$

Using the KKT conditions of the minimization problem in (10.12) for $(x,z) \in \operatorname{gph} S$, we obtain

$$h_1^*(y) = \begin{cases} -2, & \text{if } y \le -2,\\ \frac{1}{4}y^2 + y - 1, & \text{if } y \in\ ]-2,0[,\\ y - 1, & \text{if } y \in [0,1],\\ \frac{1}{2}(y^2 - 1), & \text{if } y \in\ ]1,2[,\\ 2y - \frac{5}{2}, & \text{if } y \ge 2, \end{cases} \quad \text{since} \quad \left(x_1^{h^*}(y), z_1^{h^*}(y)\right) = \begin{cases} (0,2), & \text{if } y \le -2,\\ \frac{1}{2}(2+y, 2-y), & \text{if } y \in\ ]-2,0[,\\ (1,1), & \text{if } y \in [0,1],\\ (y,1), & \text{if } y \in\ ]1,2[,\\ (2,1), & \text{if } y \ge 2, \end{cases}$$

see Figure 10.1 (d). Simple calculations show that the functions f_1^* and h_1^* are continuously differentiable on $\mathbb{R}$ with gradients

$$\nabla f_1^*(y) = \begin{cases} 1, & \text{if } y < 2, \\ \frac{1}{2}y, & \text{if } y \in [2,4], \\ 2, & \text{if } y > 4, \end{cases} \quad \text{and} \quad \nabla h_1^*(y) = \begin{cases} 0, & \text{if } y \leq -2, \\ \frac{1}{2}y+1, & \text{if } y \in]-2,0[, \\ 1, & \text{if } y \in [0,1], \\ y, & \text{if } y \in]1,2[, \\ 2, & \text{if } y \geq 2. \end{cases}$$

The same results follow from Lemmata 10.7 and 10.8, respectively. ◇

We now apply Toland's and Singer's duality theory [131, Theorem 2.2] which states that

$$\inf_{x\in\mathbb{R}^n} [f(x) - h(x)] = \inf_{y\in\mathbb{R}^n} [h^*(y) - f^*(y)] \tag{10.13}$$

for all functions $f, h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with h convex and lsc. Hence this duality fits perfectly within our framework and allows us to state the following main result of this section.

Theorem 10.10 *Let Assumption 10.1 hold, and define the dual gap function*

$$d_\alpha^* := h_\alpha^* - f_\alpha^*$$

with the functions f_α^* and h_α^* given by Lemmata 10.7 and 10.8, respectively. Then the following statements hold:

(a) The function d_α^* is differentiable with Lipschitz gradient.

(b) If $\bar{y}$ is a solution of the unconstrained minimization problem

$$\min_{y\in\mathbb{R}^n} d_\alpha^*(y) \tag{10.14}$$

with $d_\alpha^*(\bar{y}) = 0$, then $\bar{x} := \nabla f_\alpha^*(\bar{y})$ is a solution of the QVI.

(c) Conversely, if $\bar{x}$ is a solution of the QVI and $\bar{y} \in \partial h_\alpha(\bar{x})$, then $\bar{y}$ is a solution of (10.14) *with $d_\alpha^*(\bar{y}) = 0$.*

Proof. The result is essentially an application of the duality theory by Toland [130, 131] and Singer [123]. The proof of this result is analogous to the proof of Theorem 5.11.

(a) This assertion follows immediately from the definition of the dual gap function d_α^* together with Lemmata 10.7 (c) and 10.8 (c).

(b) Let $\bar{y}$ be a solution of the differentiable unconstrained minimization problem (10.14) with

$$0 = d_\alpha^*(\bar{y}) = h_\alpha^*(\bar{y}) - f_\alpha^*(\bar{y}). \tag{10.15}$$

Then we have $\nabla d_\alpha^*(\bar{y}) = 0$ and therefore $\nabla f_\alpha^*(\bar{y}) = \nabla h_\alpha^*(\bar{y})$. This also shows that the subdifferentials $\partial f_\alpha^*(\bar{y})$ and $\partial h_\alpha^*(\bar{y})$ are equal, since the functions f_α^* and h_α^* are convex and differentiable and therefore $\partial f_\alpha^*(\bar{y}) = \{\nabla f_\alpha^*(\bar{y})\}$ and $\partial h_\alpha^*(\bar{y}) = \{\nabla h_\alpha^*(\bar{y})\}$ hold. We denote the single element of the two subdifferential by the vector $\bar{x} := \nabla f_\alpha^*(\bar{y})$. Since we also have $f_\alpha^{**} = f_\alpha$ and $h_\alpha^{**} = h_\alpha$ by Lemma 2.6 (d) and Lemma 10.6 (a), (b), we obtain from Lemma 2.6 (f) that

$$f_\alpha(\bar{x}) + f_\alpha^*(\bar{y}) = \bar{x}^T\bar{y} \quad \text{and} \quad h_\alpha(\bar{x}) + h_\alpha^*(\bar{y}) = \bar{x}^T\bar{y}. \tag{10.16}$$

Subtracting and rearranging these two equations yields

$$f_\alpha(\bar{x}) - h_\alpha(\bar{x}) = h_\alpha^*(\bar{y}) - f_\alpha^*(\bar{y}). \tag{10.17}$$

The right-hand side of this equation is equal to zero in view of (10.15). Hence $\bar{x}$ is a minimum of the nonnegative function $f_\alpha - h_\alpha$ with function value equal to zero. Therefore Lemma 10.6 (c) implies that $\bar{x}$ is a solution of the QVI.

(c) Finally, let $\bar{x}$ be a solution of the QVI. Consequently, we have

$$0 = g_\alpha(\bar{x}) = \min_{x\in\mathbb{R}^n} [f_\alpha(x) - h_\alpha(x)]$$

in view of Lemma 10.6 (c). Hence

$$f_\alpha(x) - h_\alpha(x) \geq f_\alpha(\bar{x}) - h_\alpha(\bar{x}) \quad \forall x \in \mathbb{R}^n$$

and, using $\bar{y} \in \partial h_\alpha(\bar{x})$,

$$h_\alpha(x) - h_\alpha(\bar{x}) \geq \bar{y}^T(x - \bar{x}) \quad \forall x \in \mathbb{R}^n.$$

Combining these two inequalities yields the relation

$$\bar{y}^T(\bar{x} - x) \geq h_\alpha(\bar{x}) - h_\alpha(x) \geq f_\alpha(\bar{x}) - f_\alpha(x).$$

This relation shows that the element $\bar{y}$ from the subdifferential $\partial h_\alpha(\bar{x})$ also belongs to the subdifferential $\partial f_\alpha(\bar{x})$. Using these facts, we obtain from Lemma 2.6 (f) that (10.16) holds which, in turn, implies that (10.17) is also true. But this time, the left-hand side of (10.17) is equal to zero. Consequently, the right-hand side is also equal to zero, meaning that $\bar{y}$ is a solution of the minimization problem (10.14) with $d_\alpha^*(\bar{y}) = 0$ because of (10.13). □

In order to illustrate the results of Theorem 10.10, we return to Example 10.9.

Example 10.11 Consider once again the setting from Example 10.9. Calculating the difference of $h_1^* - f_1^*$, we obtain

$$d_1^*(y) = h_1^*(y) - f_1^*(y) = \begin{cases} -y-1, & \text{if } y \leq -2, \\ \frac{1}{4}y^2, & \text{if } y \in\,]-2, 0[, \\ 0, & \text{if } y \in [0, 1], \\ \frac{1}{2}y^2 - y + \frac{1}{2}, & \text{if } y \in\,]1, 2[, \\ -\frac{1}{4}y^2 + 2y - \frac{5}{2}, & \text{if } y \in [2, 4], \\ \frac{3}{2}, & \text{if } y > 4. \end{cases}$$

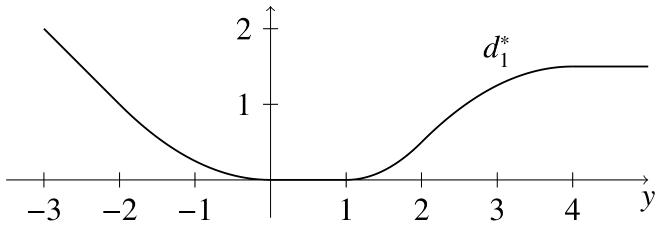


Figure 10.2.: Dual gap function d_1^* from Example 10.11

This function is illustrated in Figure 10.2. Due to the observations in Example 10.9, the corresponding QVI has the unique solution $\bar{x} = 1$. Furthermore, it holds that $\partial h_1(1) = [0, 1]$, since

$$h_1(x) = \frac{1}{2}x^2 + \frac{1}{2}\inf_{z \in S(x)} z^2 = \begin{cases} x^2 - 2x + 2, & \text{if } x \in [0, 1], \\ \frac{1}{2}(x^2 + 1), & \text{if } x \in]1, 2], \\ +\infty, & \text{if } x \notin [0, 2]. \end{cases}$$

In view of Theorem 10.10 (c), all $\bar{y} \in [0, 1]$ solve the dual problem (10.14), and this statement is consistent with the graph of the dual problem shown in Figure 10.2. Furthermore, given any solution $\bar{y} \in [0, 1]$ of (10.14), Theorem 10.10 (b) states that $\bar{x} = \nabla f_1^*(\bar{y})$ is a solution of the QVI. Since, in our case, we obtain $\nabla f_1^*(\bar{y}) = 1$ for all $\bar{y} \in [0, 1]$, it follows that $\bar{x} = 1$ solves the QVI. This confirms a corresponding observation given in Example 10.9. ◇

Note that the dual gap function in Example 10.11 has stationary points or local minima that do not provide solutions of the QVI. Since this example has relatively nice properties, this indicates that it might be difficult to obtain a result which says that, under suitable conditions, a stationary point is already a global minimum of the dual gap function. In fact, we were not able to derive such a result, but we have a partial result in this direction that is based on the following proposition.

Proposition 10.12 *Let Assumption 10.1 hold, let $d_\alpha^* = h_\alpha^* - f_\alpha^*$ be the dual gap function, and let $x_\alpha^{f^*}(y)$ and $x_\alpha^{h^*}(y)$, $z_\alpha^{h^*}(y)$ denote the vectors defined in Lemmata 10.7 and 10.8, respectively. Then the following statements are equivalent:*

(a) $x_\alpha^{f^*}(\bar{y}) = x_\alpha^{h^*}(\bar{y}) = z_\alpha^{h^*}(\bar{y})$.

(b) $d_\alpha^*(\bar{y}) = 0$.

Proof. The proof of this result is analogous to the proof of Proposition 5.13.

Assume that $x_\alpha^{f^*}(\bar{y}) = x_\alpha^{h^*}(\bar{y}) = z_\alpha^{h^*}(\bar{y})$ holds. For simplicity of notation, let us denote this common vector by $\bar{x}$. Then, in particular, we have $\bar{x} \in X$, hence the definition of f_α^* yields

$$f_\alpha^*(\bar{y}) = \bar{x}^T\bar{y} - \frac{\alpha}{2}\|\bar{x}\|^2 - \frac{1}{2\alpha}\|F(\bar{x})\|^2,$$

whereas the definition of h_α^* implies

$$h_\alpha^*(\bar{y}) = \frac{1}{2\alpha}\|\bar{y}\|^2 - \frac{\alpha}{2}\left(\left\|\bar{x} - \frac{1}{\alpha}\bar{y}\right\|^2 + \left\|\frac{1}{\alpha}F(\bar{x})\right\|^2\right) = \bar{x}^T\bar{y} - \frac{\alpha}{2}\|\bar{x}\|^2 - \frac{1}{2\alpha}\|F(\bar{x})\|^2.$$

This immediately gives $d_\alpha^*(\bar{y}) = h_\alpha^*(\bar{y}) - f_\alpha^*(\bar{y}) = 0$.

Conversely, assume that $d_\alpha^*(\bar{y}) = 0$ holds. Then, in view of Theorem 10.10, $\bar{y}$ is a global minimum of the unconstrained optimization problem (10.14). Hence we have $\nabla d_\alpha^*(\bar{y}) = 0$. On the other hand, the definition of d_α^* together with Lemmata 10.7 and 10.8 yields

$$\nabla d_\alpha^*(\bar{y}) = \nabla h_\alpha^*(\bar{y}) - \nabla f_\alpha^*(\bar{y}) = x_\alpha^{h^*}(\bar{y}) - x_\alpha^{f^*}(\bar{y}).$$

Hence we obtain

$$x_\alpha^{f^*}(\bar{y}) = x_\alpha^{h^*}(\bar{y}). \tag{10.18}$$

Furthermore, $d_\alpha^*(\bar{y}) = 0$ and Theorem 10.10 together imply that $\bar{x} := \nabla f_\alpha^*(\bar{y})$ is a solution of the QVI. Note that (10.18) and Lemma 10.7 yield

$$\bar{x} = x_\alpha^{f^*}(\bar{y}) = x_\alpha^{h^*}(\bar{y}). \tag{10.19}$$

The vector $\bar{x}$ being a solution of the QVI means that $\bar{x} \in X$ and $g_\alpha(\bar{x}) = 0$, where g_α denotes the regularized gap function, cf. Lemma 7.2. In view of (10.1), we may rewrite this regularized gap function as

$$\begin{aligned} g_\alpha(\bar{x}) &= \frac{1}{2\alpha}\|F(\bar{x})\|^2 - \frac{\alpha}{2} \inf_{z \in S(\bar{x})} \left\| z - \left(\bar{x} - \frac{1}{\alpha}F(\bar{x})\right)\right\|^2 \\ &= \frac{1}{2\alpha}\|F(\bar{x})\|^2 - \frac{\alpha}{2} \left\| z_\alpha(\bar{x}) - \left(\bar{x} - \frac{1}{\alpha}F(\bar{x})\right)\right\|^2 \end{aligned}$$

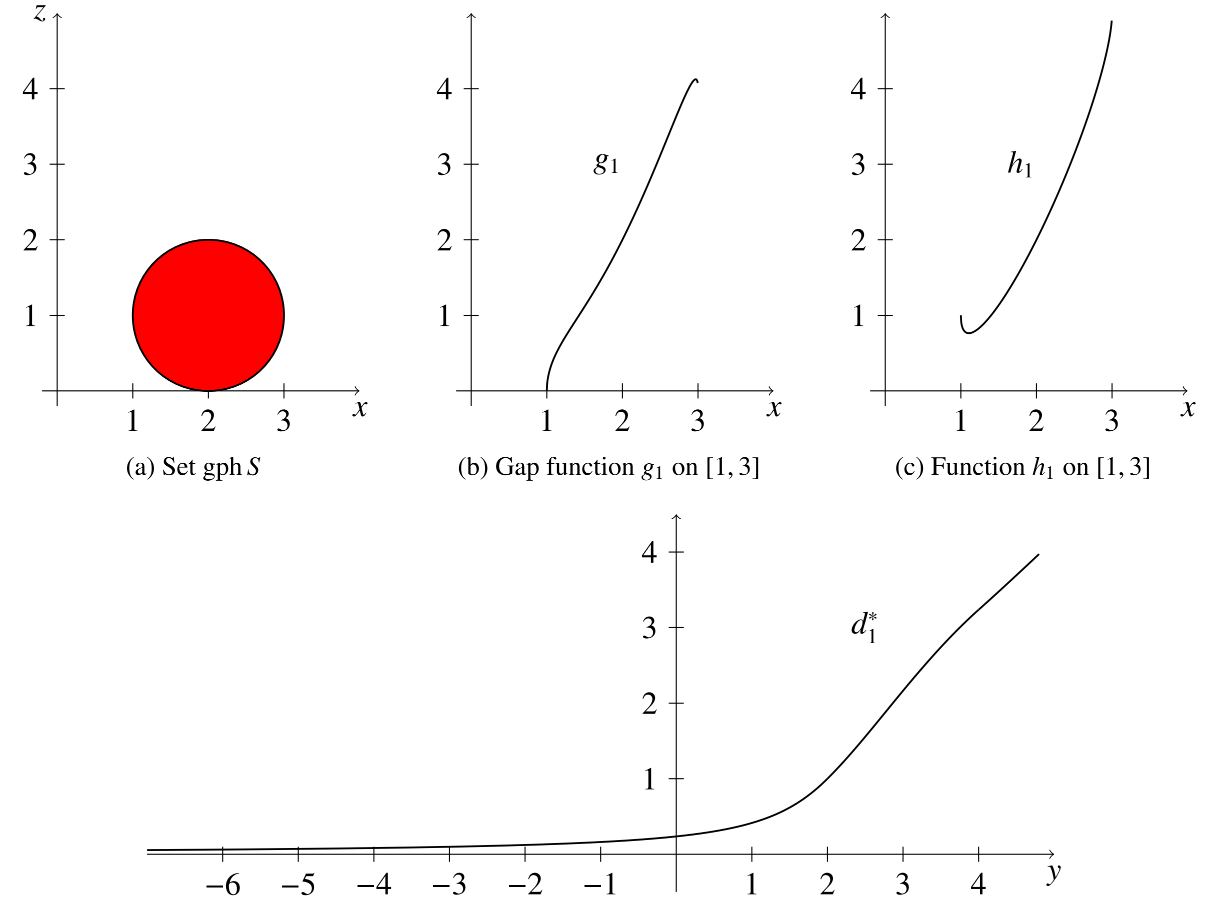
with the uniquely defined minimum

$$z_\alpha(\bar{x}) := \operatorname*{argmin}_{z \in S(\bar{x})} \left\| z - \left(\bar{x} - \frac{1}{\alpha}F(\bar{x})\right)\right\|^2.$$

According to a result of Taji [127] given in Lemma 7.2 (c), $\bar{x}$ being a solution of the QVI is equivalent to $z_\alpha(\bar{x}) = \bar{x}$. However, in view of the representation (10.9) of the function $h_\alpha^*(\bar{y})$, it follows that $z_\alpha(\bar{x})$ is identical to $z_\alpha^{h^*}(\bar{y})$. Consequently, we have $z_\alpha^{h^*}(\bar{y}) = \bar{x}$. Together with (10.19), this completes the proof. □

Proposition 10.12 shows that $x_\alpha^{f^*}(\bar{y}) = x_\alpha^{h^*}(\bar{y}) = z_\alpha^{h^*}(\bar{y}) =: \bar{x}$ implies $d_\alpha^*(\bar{y}) = 0$ and, therefore, that $\bar{x}$ is a solution of the QVI. This sufficient condition for a solution is partially satisfied at any stationary point of the dual gap function, since, as noted in the previous proof, we always have $x_\alpha^{f^*}(\bar{y}) = x_\alpha^{h^*}(\bar{y})$ at a stationary point $\bar{y}$ of d_α^*. The missing part is therefore to verify that these two vectors are also equal to $z_\alpha^{h^*}(\bar{y})$ which seems to be the difficult part that is not satisfied in Example 10.9 for all $y \geq 4$. However, we could not verify a sufficient condition under which stationary points of the dual gap function d_α^* already provide solutions of a QVI. On the other hand, since we know the optimal value of d_α^*, this disadvantage might not be that strong, since the function value itself tells us whether we have a solution or not.

Theorem 10.10 gives, more or less, a one-to-one correspondence between the solutions of the QVI and the global minima of the dual gap function d_α^*. In fact, it shows that every solution of the optimization problem (10.14) yields a solution of the QVI, but the converse is not necessarily



(a) Set gph S (b) Gap function g_1 on $[1,3]$ (c) Function h_1 on $[1,3]$
(d) Dual gap function d_1^*

Figure 10.3.: Illustrations for Example 10.13

true, because statement (c) of Theorem 10.10 assumes (implicitly) that the subdifferential $\partial h_\alpha(\bar{x})$ is nonempty. As illustrated by the following counterexample, this subdifferential could be empty, and the infimum in the relation (10.13) is not necessarily attained.

Example 10.13 Consider the QVI with $n = 1$, $F(x) = x$, and $S(x) = \{z \in \mathbb{R} \mid (x-2)^2 + (z-1)^2 \leq 1\}$ such that

$$S(x) = \begin{cases} \left[1 - \sqrt{1-(x-2)^2}, 1 + \sqrt{1-(x-2)^2}\right], & \text{if } x \in [1,3], \\ \emptyset, & \text{if } x \notin [1,3], \end{cases}$$

see Figure 10.3 (a). Note that $M = [1,3]$ is the domain of S in this example, that $X = [1,2]$ is the feasible set, and that all conditions in Assumption 10.1 are satisfied. Let $\alpha = 1$. The corresponding regularized gap function

$$g_1(x) = \frac{1}{2}x^2 - \frac{1}{2}\inf_{z \in S(x)} z^2 = \begin{cases} \frac{1}{2}x^2 - \frac{1}{2}\left(1 - \sqrt{1-(x-2)^2}\right)^2, & \text{if } x \in [1,3], \\ -\infty, & \text{if } x \notin [1,3], \end{cases}$$

is illustrated in Figure 10.3 (b). This function has the DC decomposition with the functions

$$f_1(x) = x^2 + \delta_X(x) \quad \text{and} \quad h_1(x) = \frac{1}{2}x^2 + \frac{1}{2}\inf_{z\in S(x)} z^2 \quad \text{(see Figure 10.3 (c)).}$$

The conjugate function of f_1 is the same as in Example 10.9:

$$f_1^*(y) = \sup_{x\in[1,2]} \left[xy - x^2\right] = \begin{cases} y-1, & \text{if } y < 2,\\ \frac{1}{4}y^2, & \text{if } y \in [2,4],\\ 2y-4, & \text{if } y > 4. \end{cases}$$

As in Example 10.9, we obtain

$$h_1^*(y) = \frac{1}{2}y^2 - \min_{(x,z)\in \operatorname{gph} S} \left[\frac{1}{2}(x-y)^2 + \frac{1}{2}z^2\right], \tag{10.20}$$

where the minimization problem in (10.20) is again strongly convex and differentiable. This problem has the following Lagrange function:

$$L_1^{h^*}(x,z,\lambda,y) = \frac{1}{2}(x-y)^2 + \frac{1}{2}z^2 + \lambda\left((x-2)^2 + (z-1)^2 - 1\right).$$

Then the KKT conditions of this problem are

$$\nabla_{(x,z)} L_1^{h^*}(x,z,\lambda,y) = \begin{pmatrix} x - y + 2\lambda(x-2) \\ z + 2\lambda(z-1) \end{pmatrix} = 0,$$
$$\lambda \geq 0,\ (x-2)^2 + (z-1)^2 - 1 \leq 0,\ \lambda\left((x-2)^2 + (z-1)^2 - 1\right) = 0.$$

Since the unconstrained minimum $(\bar{x},\bar{z}) = (y,0)$ of the minimization problem in (10.20) is not an interior point of $\operatorname{gph} S$ for all $y \in \mathbb{R}$, any solution of this problem has to be constrained, that is, $\left(x_1^{f^*}(y) - 2\right)^2 + \left(z_1^{f^*}(y) - 1\right)^2 - 1 = 0$ has to hold. Using the KKT conditions of the problem in (10.20), we obtain

$$\left(x_1^{f^*}(y), z_1^{f^*}(y)\right) = \left(\frac{y - 2 + 2\sqrt{1+(y-2)^2}}{\sqrt{1+(y-2)^2}}, \frac{-1 + \sqrt{1+(y-2)^2}}{\sqrt{1+(y-2)^2}}\right)$$

and therefore

$$h_1^*(y) = 2y + \sqrt{1+(y-2)^2} - 3.$$

We see that the function g_1 is zero only at $x = 1$, hence this point is the unique solution of the QVI. At this point, the slope of h_1 is infinite, and $\partial h_1(1) = \emptyset$. Hence, for this example, we cannot apply Theorem 10.10 to determine the solutions for the dual problem (10.14). We further note that $d_1^*(y) = h_1^*(y) - f_1^*(y) > 0$ holds for all $y \in \mathbb{R}$ and $\lim_{y\to-\infty} d_1^*(y) = 0$, see Figure 10.3 (d). Therefore, zero is the infimum but not the minimum of the unconstrained minimization problem (10.14), which does not have a solution. ◇

Based on the continuously differentiable dual reformulation of QVIs described in Theorem 10.10, we will present some numerical results in Section 11.2.

10.2. Second-Order Properties

The dual gap function d_α^* turned out to be piecewise smooth (see Definition 2.9) in all previous examples. The aim of this section is therefore to show that this observation is true in a rather general setting.

Piecewise smooth functions arise naturally in the context of Euclidian projections onto convex sets. To this end, let us assume that we have a set $\Omega \subseteq \mathbb{R}^n$ described by

$$\Omega := \{x \in \mathbb{R}^n \mid c_i(x) \le 0 \quad \forall i = 1, \dots, m\}, \tag{10.21}$$

with $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$, convex and twice continuously differentiable. Furthermore, for $x \in \Omega$ we define the active index set

$$I(x) := \{i \in \{1, \dots, m\} \mid c_i(x) = 0\}$$

Recall from Definition 2.13, which goes back to Janin [85], that CRCQ is satisfied at $\bar{x} \in \Omega$ (with respect to the set Ω) if there exists a neighborhood U of $\bar{x}$ such that for all $K \subseteq I(\bar{x})$, the family of gradients $\{\nabla c_i(x) \mid i \in K\}$ has constant rank (depending on the set K) for all $x \in U$.

Let Ω be the set from (10.21). Recall that the unique solution of the strongly convex minimization problem

$$\min_{w \in \Omega} \frac{1}{2}\|w - v\|^2$$

is called the *Euclidean projection* of a given vector $v \in \mathbb{R}^n$ onto the set Ω, denoted by $P_\Omega(v)$. The mapping $v \mapsto P_\Omega(v)$ is then called the *projection mapping*. It is well-known that this mapping is piecewise smooth under the CRCQ assumption. More precisely, the following result holds, see, for example, [54, Theorem. 4.5.2].

Theorem 10.14 *Let Ω be the set defined in* (10.21) *with twice continuously differentiable and convex functions c_i. Let $\bar{v} \in \mathbb{R}^n$ be given such that CRCQ holds at $\bar{w} := P_\Omega(\bar{v}) \in \Omega$. Then the projection mapping P_Ω is a PC^1 function near $\bar{v}$.*

In order to apply this result to our case, recall that our two conjugate functions f_α^* and h_α^* also involve projections, but not with respect to the Euclidean norm. Instead, we are dealing with scaled projection problems of the form

$$\min_{w \in \Omega} \frac{1}{2}\|Dw - v\|^2, \tag{10.22}$$

where Ω denotes again the set from (10.21) and $D \in \mathbb{R}^{n \times n}$ is a nonsingular matrix. Then problem (10.22) is equivalent to the standard (Euclidean) projection problem

$$\min_{u \in D \cdot \Omega} \frac{1}{2}\|u - v\|^2$$

in the sense that the optimal values are equal with

$$\operatorname*{argmin}_{u \in D \cdot \Omega} \frac{1}{2}\|u - v\|^2 = P_{D \cdot \Omega}(v) \quad \text{and} \quad \operatorname*{argmin}_{w \in \Omega} \frac{1}{2}\|Dw - v\|^2 = D^{-1} P_{D \cdot \Omega}(v).$$

We are interested in the smoothness properties of the mapping $v \mapsto D^{-1}P_{D\cdot\Omega}(v)$. To this end, we first state the following result, which says that CRCQ still holds if the set is transformed in a simple way. The transformation is precisely the one that will be used in order to deal with projection-like problems as in (10.22).

Lemma 10.15 *Let $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$, be convex and continuously differentiable,*

$$\Omega := \{x \in \mathbb{R}^n \mid c_i(x) \leq 0 \quad \forall i = 1, \dots, m\}$$

and $\bar{v} \in \mathbb{R}^n$ such that CRCQ holds at $\bar{w} := D^{-1}P_{D\cdot\Omega}(\bar{v}) \in \Omega$. Then CRCQ holds at $\bar{u} := D\bar{w} \in D\cdot\Omega$.

Proof. First note that, setting $\tilde{c}_i(u) := c_i(D^{-1}u)$ for all $u \in \mathbb{R}^n$ and $i = 1, \dots, m$, we have

$$D \cdot \Omega = \{u \in \mathbb{R}^n \mid \tilde{c}_i(u) \leq 0 \quad \forall i = 1, \dots, m\},$$

and thus,

$$I(\bar{u}) = \{i \in \{1, \dots, m\} \mid \tilde{c}_i(\bar{u}) = 0\} = \{i \in \{1, \dots, m\} \mid c_i(\bar{w}) = 0\} = I(\bar{w}).$$

By assumption, there exists a neighborhood W of $\bar{w}$ such that for all $K \subseteq I(\bar{w})$ the family of gradients $\{\nabla c_i(w) \mid i \in K\}$ has constant rank for all $w \in W$. Since D is nonsingular, the set $U := D \cdot W$ is a neighborhood of $\bar{u}$. Now, let $u, u' \in U$ and $K \subseteq I(\bar{u})$ be given, in particular, there exist $w, w' \in W$ such that $u = Dw$ and $u' = Dw'$. Furthermore, we obtain

$$\{\nabla \tilde{c}_i(u) \mid i \in K\} = D^{-1} \cdot \{\nabla c_i(w) \mid i \in K\}$$

and

$$\{\nabla \tilde{c}_i(u') \mid i \in K\} = D^{-1} \cdot \{\nabla c_i(w') \mid i \in K\}.$$

Doe to the arguments above, both sets have the same rank, which concludes the proof. □

The previous result allows us to formulate the PC^1 property for the solution mapping of problems in the form (10.22).

Proposition 10.16 *Let the assumptions of Lemma 10.15 hold. Assume, in addition, that the functions c_i, $i = 1, \dots, l$, are twice continuously differentiable. Then there exists a neighborhood V of $\bar{v}$ such that $v \mapsto D^{-1}P_{D\cdot\Omega}(v)$ is PC^1 on V.*

Proof. From Lemma 10.15 we infer that CRCQ holds at $\bar{u} := D\bar{w} = P_{D\cdot\Omega}(\bar{v}) \in D \cdot \Omega$. Hence, in view of Theorem 10.14, we conclude that there is a neighborhood U of $\bar{u}$ on which $v \mapsto P_{D\cdot\Omega}(v)$ is PC^1. Hence the function $v \mapsto D^{-1}P_{D\cdot\Omega}(v)$ is PC^1 on $V := D \cdot U$ (recall that V is indeed a neighborhood of $\bar{v}$ due to the nonsingularity of the matrix D). □

Now we want to apply the previous result in order to show that the gradient ∇d_α^* of the function d_α^* from Theorem 10.10 is PC^1. For these purposes, we assume throughout that the set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ has the form

$$S(x) := \{z \in \mathbb{R}^n \mid s_i(z, x) \leq 0 \quad \forall i = 1, \dots, m\}, \tag{10.23}$$

where the functions $s_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$, are twice continuously differentiable and convex in (z, x). Note that Assumption 10.1 (c) automatically holds in this case. Then we have

$$X = \{x \in \mathbb{R}^n \mid s_i(x, x) \leq 0 \quad \forall i = 1, \dots, m\}.$$

In order to verify the piecewise smoothness of the gradient of the dual gap function d_α^*, we show that both ∇h_α^* and ∇f_α^* are piecewise smooth. We begin with the mapping h_α^*.

Lemma 10.17 *Let Assumption 10.1 hold, and let $\bar{y} \in \mathbb{R}^n$ such that CRCQ holds at the point $(\bar{x}, \bar{z}) := D_h^{-1} P_{D_h \cdot \operatorname{gph} S}(\bar{y}, -b) \in \operatorname{gph} S$, where*

$$D_h := \begin{pmatrix} \alpha I_n & 0 \\ A - \alpha I_n & \alpha I_n \end{pmatrix}.$$

Then ∇h_α^* is PC^1 near $\bar{y}$.

Proof. Due to Lemma 10.8, we have $\nabla h_\alpha^*(y) = x_\alpha^{h^*}(y)$ for all $y \in \mathbb{R}^n$, where

$$\begin{aligned}
\begin{pmatrix} x_\alpha^{h^*}(y) \\ z_\alpha^{h^*}(y) \end{pmatrix} &= \operatorname*{argmin}_{(x,z) \in \operatorname{gph} S} \left\{ \left\| x - \frac{1}{\alpha} y \right\|^2 + \left\| z - \left(x - \frac{1}{\alpha} F(x) \right) \right\|^2 \right\} \\
&= \operatorname*{argmin}_{(x,z) \in \operatorname{gph} S} \left\| \begin{pmatrix} \alpha I_n & 0 \\ A - \alpha I_n & \alpha I_n \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} - \begin{pmatrix} y \\ -b \end{pmatrix} \right\|^2 \\
&= D_h^{-1} P_{D_h \cdot \operatorname{gph} S}(y, -b).
\end{aligned}$$

Hence the assertion follows immediately from Proposition 10.16. □

Similar to the previous result, we prove in the next lemma that also the function ∇f_α^* is piecewise smooth under a suitable CRCQ assumption.

Lemma 10.18 *Let Assumption 10.1 hold, and let $\bar{y} \in \mathbb{R}^n$ such that CRCQ holds at the point $\bar{x} := D_f^{-1} P_{D_f \cdot X}(D_f^{-1}(\bar{y} - q_\alpha)) \in X$, where $q_\alpha = \frac{1}{\alpha} A^T b$ and $D_f := Q_\alpha^{\frac{1}{2}}$ denotes the matrix square root of the matrix $Q_\alpha = \alpha(I_n + \frac{1}{\alpha^2} A^T A)$ from* (10.6)*. Then ∇f_α^* is PC^1 near $\bar{y}$.*

Proof. Due to Lemma 10.7, we have

$$\begin{aligned}
\nabla f_\alpha^*(y) &= \operatorname*{argmin}_{x \in X} \left\| Q_\alpha^{-1}(y - q_\alpha) - x \right\|_{Q_\alpha}^2 \\
&= \operatorname*{argmin}_{x \in X} \left\| Q_\alpha^{-\frac{1}{2}}(y - q_\alpha) - Q_\alpha^{\frac{1}{2}} x \right\|^2 \\
&= D_f^{-1} P_{D_f \cdot X}\left(D_f^{-1}(y - q_\alpha) \right)
\end{aligned}$$

for all $y \in \mathbb{R}^n$. Hence the assertion follows immediately from Proposition 10.16. □

Summarizing the previous results, we obtain the following main result of this section.

Theorem 10.19 *Let Assumption 10.1 hold, and let $\bar{y} \in \mathbb{R}^n$ such that the assumptions of Lemmata 10.17 and 10.18 hold for $\bar{y}$. Then the gradient of the dual gap function ∇d_α^* is PC^1 near $\bar{y}$.*

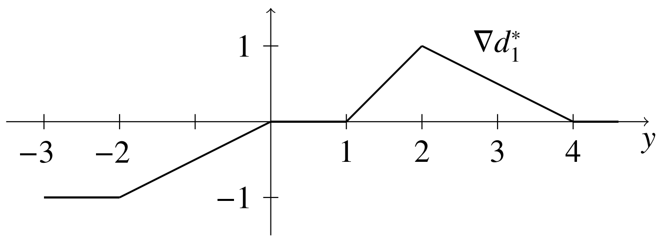


Figure 10.4.: Gradient ∇d_1^* from Example 10.21

Note that the two CRCQ conditions used in Lemmata 10.17 and 10.18 are independent of each other. A simple, but still important, case where the constant rank assumption holds, is the linear one. This yields the following consequence.

Corollary 10.20 *Let the functions s_i, $i = 1, \dots, m$, in* (10.23) *be affine-linear. Then the gradient of the dual gap function ∇d_α^* is a PC^1 mapping (in fact, it is piecewise affine-linear).*

In order to illustrate the results of Corollary 10.20, we return to Examples 10.9 and 10.11.

Example 10.21 Consider once again the QVI from Example 10.9. The set-valued mapping S of this QVI has the form

$$S(x) = \{z \in \mathbb{R} \mid s_i(z, x) \leq 0 \quad \forall i = 1, \dots, m\}$$

with affine-linear functions s_i, $i = 1, \dots, 4$, see the description of the set S in (10.11). Furthermore, we have already obtained the corresponding dual gap function d_1^* in Example 10.11. Hence we calculate the gradient of this function

$$\nabla d_1^*(y) = \begin{cases} -1, & \text{if } y \in]-\infty, -2], \\ \frac{1}{2}y, & \text{if } y \in]-2, 0[, \\ 0, & \text{if } y \in [0, 1] \cup]4, +\infty[, \\ y - 1, & \text{if } y \in]1, 2[, \\ -\frac{1}{2}y + 2, & \text{if } y \in [2, 4], \end{cases}$$

which is illustrated in Figure 10.4. This gradient is piecewise affine-linear as expected from Corollary 10.20. ◇

Piecewise smooth functions are, in particular, semismooth in the sense of [112, 113], see, for examlpe, [54, Proposition 7.4.6]. In principle, this observation therefore allows the application of second-order Newton-type methods for the minimization of the dual gap function.

11. Numerical Results for Quasi-Variational Inequalities

In this chapter we present some numerical results for QVIs based on our approaches in Section 8.1 and Section 10.1. To this end, we confine us in Section 11.1 to QVIs with moving sets and generalized moving sets and apply the TOMLAB/SNOPT solver and the TOMLAB/KNITRO solver to the reformulation of such QVIs as the constrained optimization problem (7.9). Then in Section 11.2 we apply a global spectral gradient method and a conjugate gradient method to the reformulation of a QVI as the unconstrained optimization problem based on the dual gap function d_α^* defined in Theorem 10.10. The results from Sections 11.1 and 11.2 were published in [75] and [72], respectively.

11.1. Primal Gap Function Approach

Here we present numerical results for the solution of QVIs based on the optimization reformulation from (7.9):

$$\min \; g_\alpha(x) \quad \text{subject to} \quad x \in X,$$

where g_α denotes the regularized gap function and X is the feasible set of the QVI, cf. (7.6) and (7.3), respectively. In order to apply suitable standard software to this problem, we have to distinguish two cases: First we have a QVI with a generalized moving set in which case (7.9) represents a smooth (continuously differentiable) optimization problem. Second, if the constraints are not given by a generalized moving set, g_α is not necessarily continuously differentiable everywhere, although our analysis in Chapter 9 shows that, also in this case, except for some pathological situations, we can expect differentiability at all local minimizers.

Since, for the nondifferentiable case, numerical results are presented in Section 6.1 for the special case of generalized Nash equilibrium problems, we decided to concentrate on QVIs defined by generalized moving sets in this section. More precisely, we consider both QVIs with (standard) moving sets and QVIs with generalized moving sets as defined in Section 8.1.

To this end, we recall that the generalized gap function g_α is well defined for all $x \in \mathbb{R}^n$ in the moving and generalized moving set cases whenever $K \neq \emptyset$. This observation is important since this allows to apply software that might generate non-feasible iterates. In particular, this enables us to use the TOMLAB/SNOPT 7.2.9 solver as the working horse for problem (7.9), especially since this method does not use any second-order derivatives. However, we compare the results also with the TOMLAB/KNITRO 8.0.0 solver applied to (7.9) although, formally, this solver uses second-order information and, therefore, is not a feasible method in our case since

				SNOPT Solver		KNITRO Solver	
Example	n	m	x^0	k	g_α^{opt}	k	g_α^{opt}
MovSet1A	5	1	$(0,\dots,0)$	9	8.1190e-09	6	1.7720e-09
			$(10,\dots,10)$	14	8.1687e-09	8	3.6953e-11
MovSet1B	5	1	$(0,\dots,0)$	57	-1.4553e-09	7	5.9139e-10
			$(10,\dots,10)$	89	-4.1061e-08	16	5.8887e-10
MovSet2A	5	1	$(0,\dots,0)$	9	3.1279e-13	5	4.6895e-10
			$(10,\dots,10)$	18	-1.9631e-11	9	4.6971e-10
MovSet2B	5	1	$(0,\dots,0)$	35	3.1292e-09	9	-1.4995e-05
			$(10,\dots,10)$	–	failure	–	failure
MovSet3A1	1000	1	$(0,\dots,0)$	55	1.5426e-06	6	-1.5727e-09
			$(10,\dots,10)$	54	1.5427e-06	11	1.5038e-09
MovSet3B1	1000	1	$(0,\dots,0)$	57	5.2083e-08	7	4.8238e-10
			$(10,\dots,10)$	56	5.2119e-08	12	4.4169e-10
MovSet3A2	2000	1	$(0,\dots,0)$	64	4.3399e-11	7	1.3183e-11
			$(10,\dots,10)$	63	3.0436e-11	11	1.4201e-11
MovSet3B2	2000	1	$(0,\dots,0)$	63	1.0951e-07	7	1.6163e-11
			$(10,\dots,10)$	63	1.0954e-07	13	9.7019e-11
MovSet4A1	400	801	$(10,\dots,10)$	3	4.2168e-12	3	5.4949e-13
MovSet4B1	400	801	$(10,\dots,10)$	3	3.0465e-12	3	-1.7639e-13
MovSet4A2	800	1601	$(10,\dots,10)$	4	2.1394e-12	3	7.3646e-13
MovSet4B2	800	1601	$(10,\dots,10)$	4	-2.6190e-13	3	8.0765e-13

Table 11.1.: Table with numerical results for QVIs with moving sets from paper [51]

the regularized gap function g_α may not be twice continuously differentiable everywhere. For more information about TOMLAB/SNOPT solver and TOMLAB/KNITRO solver, we refer to their User's Guides on the web sites [1] and [2], respectively.

For both solvers, we provide the starting point x^0 as well as the function and gradient values (including the derivative of g_α from (8.3)) for each test problem. Moreover, for KNITRO, we use the active set Sequential Linear-Quadratic Programming (SLQP) optimizer by setting `Prob.KNITRO.options.ALG=3`. Apart from this, all standard options are taken for both methods. Our implementation uses the regularization parameter $\alpha = 1$ for all test problems.

We use two groups of test examples: The first group consists of all the QVIs with (standard) moving sets from the test problem collection [51] (called MovSet*). For the second group, we modify these test problems to QVIs with generalized moving sets (called GenMovSet*) defined by the diagonal matrix $Q(x) = \operatorname{diag}\left(\frac{1}{x_1^2+1}, \dots, \frac{1}{x_n^2+1}\right)$. The numerical results for the first group are presented in Table 11.1, whereas Table 11.2 contains the numerical results for the second group.

For each test example, Tables 11.1 and 11.2 contain the following data: the name of the example, the number of variables n, the number of constraints s_i, $i = 1, \dots, m$, the starting point x^0, and for both solvers the number of iterations k needed until convergence and the final value of the generalized gap function g_α in column g_α^{opt} (whenever a solution was found). Here, the starting points in Table 11.1 are those taken from the paper [51] and implemented in the corresponding M-file `startingPoints.m`. The same starting points are used for the generalized moving set examples. The results for examples MovSet4* and GenMovSet4* with the starting

				SNOPT Solver		KNITRO Solver	
Example	n	m	x^0	k	g_α^{opt}	k	g_α^{opt}
GenMovSet1A	5	1	$(0,\dots,0)$	10	-8.0488e-13	6	2.9963e-08
			$(10,\dots,10)$	18	4.0500e-12	13	2.9963e-08
GenMovSet1B	5	1	$(0,\dots,0)$	21	-1.2869e-02	11	7.6188e-06
			$(10,\dots,10)$	18	-1.8537e-04	15	7.8063e-06
GenMovSet2A	5	1	$(0,\dots,0)$	8	1.9762e-11	6	7.7652e-09
			$(10,\dots,10)$	18	-3.3309e-10	10	7.7636e-09
GenMovSet2B	5	1	$(0,\dots,0)$	28	1.3524e-09	14	1.9856e-06
			$(10,\dots,10)$	–	failure	–	failure
GenMovSet3A1	1000	1	$(0,\dots,0)$	29	5.9914e-10	8	9.8173e-12
			$(10,\dots,10)$	42	6.0085e-10	17	3.9912e-10
GenMovSet3B1	1000	1	$(0,\dots,0)$	31	3.1845e-11	8	2.0142e-10
			$(10,\dots,10)$	43	3.3889e-11	17	1.9064e-10
GenMovSet3A2	2000	1	$(0,\dots,0)$	34	1.2260e-09	9	4.9325e-10
			$(10,\dots,10)$	51	1.2214e-09	16	-3.3733e-08
GenMovSet3B2	2000	1	$(0,\dots,0)$	36	7.7424e-11	8	-5.4514e-11
			$(10,\dots,10)$	59	6.5349e-11	18	-6.1479e-10
GenMovSet4A1	400	801	$(10,\dots,10)$	12	5.6944e-03	10	1.3273e-08
GenMovSet4B1	400	801	$(10,\dots,10)$	12	4.7289e-03	10	1.3704e-08
GenMovSet4A2	800	1601	$(10,\dots,10)$	13	6.7424e-14	10	2.6527e-08
GenMovSet4B2	800	1601	$(10,\dots,10)$	12	1.0695e-02	10	2.8100e-08

Table 11.2.: Table with numerical results for QVIs with generalized moving sets

k	x^k	$g_\alpha(x^k)$	$\#g_\alpha$
0	(10, 10)	1.2744e+01	1
1	(9.84901583, 9.84901583)	3.8544e-01	3
2	(9.82327425, 9.82327425)	1.2972e-02	5
3	(9.81753271, 9.81753271)	1.1948e-06	6
4	(9.81747717, 9.81747717)	6.1412e-12	7
5	(9.81747704, 9.81747704)	-7.7879e-20	8

Table 11.3.: Table with numerical results for Example 8.4

point equal to the zero vector (as suggested in [51]) are not contained in Tables 11.1 and 11.2 since the zero vector turned out to be a solution of these test problems and was immediately identified as such from both solvers.

Tables 11.1 and 11.2 show that all test examples can be solved within a very reasonable number of iterations except for examples MovSet2B and GenMovSet2B with the second starting point. These tables also indicate that the number of iterations needed by KNITRO is sometimes significantly smaller than the corresponding numbers for SNOPT. A possible explanation might be the fact that KNITRO uses second-order information. We also believe that this fact is responsible for the higher accuracy that is sometimes obtained by the KNITRO solver. In fact, SNOPT terminates for three of the four test examples called GenMovSet4* with the function value of g_α being around $10^{-2} - 10^{-3}$, whereas KNITRO is able to get much closer to zero. Nevertheless, the termination by SNOPT was successful in the sense that the standard stopping criteria of this solver were reached.

Note also that, in some cases, upon termination we have a negative function value g_α^{opt} in the corresponding columns of Tables 11.1 and 11.2. These negative values arise for two reasons: First, if the final iterate x^k is slightly outside the feasible region, then g_α might be negative. Second, negative values may arise due to inexact function evaluations (recall that the evaluation of g_α at a point x requires the solution of an optimization problem which, fortunately, automatically also gives the gradient $\nabla g_\alpha(x)$).

Finally, in Table 11.3, we come back to our Example 8.4 and present the corresponding iteration history, with all calculations being done by SNOPT (using KNITRO we obtained a similar iteration history). More precisely, for each iteration k, Table 11.3 provides the iteration vector x^k, the value of g_α at x^k as well as the cumulated number $\#g_\alpha$ of evaluations of the mapping g_α. Table 11.3 illustrates that the calculation of a solution for the starting point $x^0 = (10, 10)$ finishes successfully and has a fast local convergence rate. We also tried a number of different starting points and were always able to find a solution up to the required accuracy. Note, however, that Example 8.4 has infinitely many solutions, hence the method finds different solutions when using different starting points.

11.2. Dual Gap Function Approach

In view of Theorem 10.10, a solution $\bar{y}$ of the the dual unconstrained minimization problem

$$\min_{y\in\mathbb{R}^n} d_\alpha^*(y)$$

from (10.14) implies a solution $\bar{x} = \nabla f_\alpha^*(\bar{y})$ of the corresponding QVI. In this section, we apply this theory to a class of examples from the QVILIB test problem collection [51] that satisfy Assumption 10.1.

For the solution of the unconstrained minimization problem (10.14), we use two different first-order methods: the global spectral gradient (GSG) method from [115] and a conjugate gradient (CG) method. The details of the GSG method are summarized in Section 6.1. A variant of a conjugate gradient (CG) method, which was first presented by Polak and Ribière in [111], is restated in the following algorithm.

Algorithm 11.1 *(Conjugate gradient method)*

(S.0) Choose $y^0 \in \mathbb{R}^n$, $0 < \sigma < \rho < \frac{1}{2}$, *set* $p^0 := -\nabla d_\alpha^*(y^0)$ *and* $k := 0$.

(S.1) If a suitable termination criterion holds: STOP.

(S.2) (Restart) If $\nabla d_\alpha^*(y^k)^T p^k > 0$, *set* $p^k := -\nabla d_\alpha^*(y^k)$.

(S.3) (The strong Wolfe-Powell conditions) Choose a step size $t_k > 0$ *with*

$$d_\alpha^*\left(y^k + t_k p^k\right) \le d_\alpha^*(y^k) + \sigma t_k \nabla d_\alpha^*(y^k)^T p^k$$

Example	n	y^0	k	$\#d^*_\alpha$	d^*_α	$\|\nabla d^*_\alpha\|$
Scrim11	2400	$(0,\dots,0)$	32	33	1.8044e-08	1.9517e-05
Scrim11	2400	$(10,\dots,10)$	36	37	5.4686e-08	6.2638e-05
Scrim12	4800	$(0,\dots,0)$	32	33	2.7707e-08	3.3777e-05
Scrim12	4800	$(10,\dots,10)$	36	37	9.0629e-08	5.6006e-05
Scrim21	2400	$(0,\dots,0)$	32	33	1.9558e-08	1.8173e-05
Scrim21	2400	$(10,\dots,10)$	36	37	7.3633e-08	7.2570e-05
Scrim22	4800	$(0,\dots,0)$	32	33	1.7462e-08	2.9717e-05
Scrim22	4800	$(10,\dots,10)$	36	37	5.6927e-08	6.2776e-05

Table 11.4.: Numerical results with the global spectral gradient method

and

$$\left|\nabla d^*_\alpha\left(y^k + t_k p^k\right)^T p^k\right| \le -\rho \nabla d^*_\alpha(y^k)^T p^k.$$

(S.4) Set $y^{k+1} := y^k + t_k p^k$,

$$\beta_k := \frac{\left(\nabla d^*_\alpha(y^{k+1}) - \nabla d^*_\alpha(y^k)\right)^T \nabla d^*_\alpha(y^{k+1})}{\left\|\nabla d^*_\alpha(y^k)\right\|^2},$$

$p^{k+1} := -\nabla d^*_\alpha(y^{k+1}) + \beta_k^{PR} p^k$, $k \leftarrow k+1$, *and go to (S.1).*

Note that for the examples in Table 11.5 the case $\nabla d^*_\alpha(y^k)^T p^k > 0$ never occurred. Furthermore, we compute the step length t_k satisfying the strong Wolfe-Powell conditions whose implementation is based on the suggestion outlined in [92] and is summarized in [66, Algorithm 6.5]. For the computation of this step size we use the parameter $\sigma = 10^{-4}$ and $\rho = 0.1$ and, at each iteration k, the initial guess

$$t_k = \frac{-2d^*_\alpha(y^k)}{\nabla d^*_\alpha(y^k)^T p^k}.$$

For both methods, the termination criteria are $\|\nabla d^*_\alpha(y^k)\| \le 10^{-5}$ or $d^*_\alpha(y^k) \le 10^{-6}$.

For the computation of the conjugate functions of f_α and h_α from Lemmata 10.7 and 10.8, respectively, we use the TOMLAB/KNITRO solver with the active set SLQP optimizer by setting `Prob.KNITRO.options.ALG=3` and `Prob.KNITRO.options.FEASTOL=`10^{-10}, see the TOMLAB/KNITRO User's Guide on the web site [2] for more information about TOMLAB/KNITRO solver. Our implementation uses the regularization parameter $\alpha = 5$ for all test runs.

The class of test problems that we use here are named `Scrim*` in the test problem library QVILIB from [51]. This class corresponds to a large-scale transportation problem formulated as QVIs. Tables 11.4 and 11.5 contain the following data: the name of the example, the number of variables n, the starting point y^0, the number of iterations k, the cumulated number of dual gap function evaluations $\#d^*_\alpha$ needed until convergence, the final value of the dual gap function d^*_α, and the final value of the gradient norm $\|\nabla d^*_\alpha\|$.

Example	n	y^0	k	$\#d_\alpha^*$	d_α^*	$\|\nabla d_\alpha^*\|$
Scrim11	2400	$(0,\dots,0)$	15	37	3.1869e-07	1.5863e-04
Scrim11	2400	$(10,\dots,10)$	20	47	7.6852e-07	3.5654e-04
Scrim12	4800	$(0,\dots,0)$	15	37	8.8592e-07	2.4624e-04
Scrim12	4800	$(10,\dots,10)$	23	50	9.3831e-07	1.5281e-04
Scrim21	2400	$(0,\dots,0)$	15	37	3.2617e-07	1.5942e-04
Scrim21	2400	$(10,\dots,10)$	20	47	7.5009e-07	3.5802e-04
Scrim22	4800	$(0,\dots,0)$	15	37	8.6840e-07	2.3947e-04
Scrim22	4800	$(10,\dots,10)$	23	50	8.8691e-07	1.5345e-04

Table 11.5.: Numerical results with the conjugate gradient method

In view of the large number of variables in each example, the evaluation of the dual gap function is more expensive than the computations in the outer iterations for both methods. In Tables 11.4 and 11.5, we observe that the total number of function evaluations in the GSG method is less than in the CG method. Therefore, in spite of higher number of iterations, the total time until convergence in the GSG method is less than in the CG method. Furthermore, we achieve higher accuracy of the results with the GSG method although the termination criteria for both methods are the same. In any case, both methods were able to find a solution for all instances of this class of QVIs, that is, they never stopped at a local but not global minimum of d_α^*.

Final Remarks

To conclude this thesis, we summarize the main results, discuss open questions, and give some suggestions on future research topics.

In Section 2.2, for a class of convex parametric optimization problems, where the objective function is strongly convex and the feasible set is independent of the parameter and nonempty, we proved the piecewise smoothness of the solution mapping exploiting mainly a suitable CRCQ condition. Note that this result might be known in the literature, but we could not find an explicit reference. We applied this result to the dual gap functions established in this thesis for some classes of GNEPs and QVIs and obtained the piecewise smoothness of the gradient of these dual gap functions under certain assumptions. Additionally, one can show that with the solution mapping also the gradient of the regularized gap function for QVIs with generalized moving sets is piecewise smooth.

In Chapters 4 and 9, respectively, we investigated some structural properties of a constrained optimization reformulation of player convex GNEPs or QVIs whose objective function is the well-known (primal) regularized gap function given in Section 3.2 or 7.2. In particular, we proved that, apart from some exceptional cases, the objective functions are differentiable at every minimizer of the corresponding optimization problem. Hence the optimization problems are essentially differentiable and therefore allow the application of suitable algorithms for smooth optimization problems. On the other hand, for jointly convex GNEPs, one can characterize certain (normalized) solutions as the minima of a smooth optimization problem, cf. [78], whose objective function is once but not twice continuously differentiable. We believe that a similar analysis can be carried out in order to verify twice continuous differentiability of this function under convenient assumptions. The details are left as a future research topic.

In Chapters 5 and 10 we showed that a class of GNEPs and QVIs, respectively, can be reformulated as an unconstrained and smooth optimization problem using the respective dual gap function, which was developed via variational and convex analysis techniques. There are a couple of questions still open for the future research.

- First, when is a stationary point of our unconstrained optimization problems already a global minimizer and, therefore, provides a solution of the GNEP?

- Second, can we develop a second-order method with fast local convergence by employing the fact that the gradient is still piecewise smooth? Since the objective function is continuously differentiable with a semismooth gradient (under appropriate assumptions), a natural candidate would be the semismooth Newton method from [112, 113], however, the computation of the corresponding generalized Jacobians (or Hessians, in our case) might be rather expensive. Therefore, we believe that another Newton-type method based on the idea of the computable generalized Jacobian from [126] (see also [80] for an application within the framework of generalized Nash equilibrium problems) might be the better choice.

- Third, what happens in the case where the function evaluations are done only inexactly? This point is quite interesting from a practical point of view, since the evaluation of our unconstrained objective functions requires the solution of two optimization problems, which are strongly convex, but which might be difficult to compute exactly at least in the non-quadratic case for GNEPs.
- Finally, is it possible to adapt the dual gap function approach to some of the existing generalizations of QVI problems (see, e.g., [102])? Note, however, that this requires that we have convenient gap functions also for these generalized QVIs, and that still some linearity and convexity assumptions will be needed not only for the function F and the set-valued mapping S arising in the standard QVI, but also for the additional functions that occur within these generalized QVIs.

Bibliography

[1] *TOMLAB/SNOPT User's Guide.* http://tomopt.com/tomlab/products/snopt/.

[2] *TOMLAB/KNITRO User's Guide.* http://tomopt.com/tomlab/products/knitro/.

[3] L. Altangerel and G. Battur: *Perturbation approach to generalized Nash equilibrium problems with shared constraints.* Optimization Letters 6, 2012, pp. 1379–1391.

[4] H. Amann and Joachim Escher: *Analysis II.* Birkhäuser, Basel, 1999.

[5] A. S. Antipin, N. Mijajlovic, and M. Jacimovic: *A second-order iterative method for solving quasi-variational inequalities.* Computational Mathematics and Mathematical Physics 53, 2013, pp. 258–264.

[6] K.J. Arrow and G. Debreu: *Existence of an equilibrium for a competitive economy.* Econometrica 22, 1954, pp. 265–290.

[7] G. Auchmuty: *Variational principles for variational inequalities.* Numerical Functional Analysis and Optimization 10, 1989, pp. 863–874.

[8] A. Auslender: *Optimisation: Méthodes Numériques.* A Series of Maîtrise de mathématiques et applications fondamentales, Masson, Paris, New York, Barcelone, 1976.

[9] D. Aussel, R. Correa, and M. Marechal: *Gap functions for quasivariational inequalities and generalized Nash equilibrium problems.* Journal of Optimization Theory and Applications 151, 2011, pp. 474–488.

[10] C. Baiocchi and A. Capelo: *Variational and Quasivariational Inequalities: Applications to Free Boundary Problems.* Wiley, New York, 1984.

[11] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer: *Non-Linear Parametric Optimization.* Birkhäuser, Basel, 1983.

[12] J. Barzilai and J. Borwein: *Two point stepsize gradient methods.* IMA Journal of Numerical Analysis 8, 1988, pp. 141–148.

[13] A. Bensoussan: *Points de Nash dans le cas de fonctionnelles quadratiques et jeux differentiels lineaires a N personnes.* SIAM Journal on Control 12, 1974, pp. 460–499.

[14] A. Bensoussan, M. Goursat, and J.-L. Lions: *Contrôle impulsionnel et inéquations quasi-variationnelles stationnaires.* Comptes Rendus de l'Académie des Sciences de Paris 276, Série A, 1973, pp. 1279–1284.

[15] A. Bensoussan and J.-L. Lions: *Nouvelle formulation de problèmes de contrôle impulsionnel et applications.* Comptes Rendus de l'Académie des Sciences de Paris 276, Série A, 1973, pp. 1189–1192.

[16] A. Bensoussan and J.-L. Lions: *Nouvelles méthodes en contrôle impulsionnel.* Applied Mathematics and Optimization 1, 1975, pp. 289–312.

[17] C. Berge: *Espaces topologiques et fonctions multivoques.* Dunod, Paris, 1959.

[18] C. Berge: *Topological Spaces: Including a Treatment of Multi-valued Functions, Vector Spaces and Convexity.* Dover, Mineola, New York, 1997. Originally published in French: C. Berge: *Espaces topologiques, fonctions multivoques.* Dunod, Paris, 1962. Reprint of the English translation by E. M. Patterson, originally published by Oliver and Boyd, Edinburgh and London, 1963.

[19] D.P. Bertsekas: *Nonlinear Programming.* Athena Scientific, Belmont, Massachusetts, second edition 1999.

[20] J.R. Birge, L. Qi, and Z. Wei: *A variant of the Topkins-Veinott method for solving inequality constrained optimization problems.* Journal of Applied Mathematics and Optimization 41, 2000, pp. 309–330.

[21] J.F. Bonnans and A. Shapiro: *Perturbation Analysis of Optimization Problems.* Springer, New York, 2000.

[22] J.V. Burke, A.S. Lewis, and M.L. Overton: *A robust gradient sampling algorithm for nonsmooth, nonconvex optimization.* SIAM Journal on Optimization 15, 2005, pp. 751–779.

[23] D. Chan and J.-S. Pang: *The generalized quasi-variational inequality problem.* Mathematics of Operations Research 7, 1982, pp. 211–222.

[24] R. Cominetti, F. Facchinei, and J.-B. Lasserre: *Modern Optimization Modelling Techniques.* Birkhäuser, Basel, 2012.

[25] A.A. Cournot: *Recherches sur les principes mathématiques de la théorie des richesses.* Hachette, Paris, 1838.

[26] G. Debreu: *A social equilibrium existence theorem.* Proceedings of the National Academy of Sciences 38, 1952, pp. 886–893.

[27] S. Dempe: *Foundations of Bilevel Programming.* A Series of Nonconvex Optimization and Its Applications, Vol. 61, Kluwer Academic Publishers, Dordrecht, 2002.

[28] S. Dempe and D. Pallaschke: *Quasidifferentiability of optimal solutions in parametric nonlinear optimization.* Optimization 40, 1997, pp. 1–24.

[29] J.E. Dennis (Jr) and R.B. Schnabel: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, NJ, 1983.

[30] H. Dietrich: *A smooth dual gap function solution to a class of quasivariational inequalities.* Journal of Mathematical Analysis and Applications 235, 1999, pp. 380–393.

[31] H. Dietrich: *Optimal control problems for certain quasivariational inequalities.* Optimization 49, 2001, pp. 67–93.

[32] D. Dorsch, H. T. Jongen, and V. Shikhman: *On intrinsic complexity of Nash equilibrium problems and bilevel optimization.* Journal of Optimization Theory and Applications 159, 2013, pp. 606–634.

[33] D. Dorsch, H. T. Jongen, and V. Shikhman: *On structure and computation of generalized Nash equilibria.* SIAM Journal on Optimization 23, 2013, pp. 452–474.

[34] A. Dreves: *Globally Convergent Algorithms for the Solution of Generalized Nash Equilibrium Problems.* Dissertation, University of Würzburg, 2012, http://opus.bibliothek.uni-wuerzburg.de/volltexte/ 2012/6982.

[35] A. Dreves, F. Facchinei, A. Fischer, and M. Herrich: *A new error bound result for Generalized Nash Equilibrium Problems and its algorithmic application.* Computational Optimization and Applications 59, 2014, pp. 63–84.

[36] A. Dreves, F. Facchinei, C. Kanzow, and S. Sagratella: *On the Solution of the KKT Conditions of Generalized Nash Equilibrium Problems.* SIAM Journal on Optimization 21, 2011, pp. 1082–1108.

[37] A. Dreves, A. von Heusinger, C. Kanzow, and M. Fukushima: *A Globalized Newton Method for the Computation of Normalized Nash Equilibria.* Journal of Global Optimization 56, 2013, pp. 327–340.

[38] A. Dreves and C. Kanzow: *Nonsmooth optimization reformulations characterizing all solutions of jointly convex generalized Nash equilibrium problems.* Computational Optimization and Applications 50, 2011, pp. 23–48.

[39] A. Dreves, C. Kanzow, and O. Stein: *Nonsmooth optimization reformulations of player convex generalized Nash equilibrium problems.* Journal of Global Optimization 53, 2012, pp. 587–614.

[40] C. Dutang: *Existence theorems for generalized Nash equilibrium problems.* Journal of Nonlinear Analysis and Optimization: Theory and Applications 4, pp. 115–126.

[41] C. Dutang: *A survey of GNE computation methods : theory and algorithms.* Available online at http://hal.archives-ouvertes.fr/docs/00/81/35/31/PDF/meth-comp-GNE-dutangc-noformat.pdf.

[42] A.F. Izmailov and M.V. Solodov: *On error bounds and Newton-type methods for generalized Nash equilibrium problems.* Computational Optimization and Applications, published online, 2013, DOI: 10.1007/s10589-013-9595-y.

[43] F. Facchinei, A. Fischer, and M. Herrich: *A family of Newton methods for nonsmooth constrained systems with nonisolated solutions.* Mathemamatical Methods of Operations Resaerch 77, 2013, pp. 433–443.

[44] F. Facchinei, A. Fischer, and M. Herrich: *An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions.* Mathematical Programming 146, 2014, pp. 1–36.

[45] F. Facchinei, A. Fischer, and V. Piccialli: *On generalized Nash games and variational inequalities.* Operations Research Letters 35, 2007, pp. 159–164.

[46] F. Facchinei, A. Fischer, and V. Piccialli: *Generalized Nash Equilibrium Problems and Newton Methods.* Mathematical Programming 117, 2009, pp. 163-194.

[47] F. Facchinei and C. Kanzow: *Generalized Nash equilibrium problems.* Annals of Operations Research 175, 2010, pp. 177–211.

[48] F. Facchinei and C. Kanzow: *Penalty methods for the solution of generalized Nash equilibrium problems (with complete test problems).* Technical Report, Institute of Mathematics, University of Würzburg, Würzburg, Germany, February 2009.

[49] F. Facchinei and C. Kanzow: *Penalty methods for the solution of generalized Nash equilibrium problems.* SIAM Journal on Optimization 20, 2010, pp. 2228–2253.

[50] F. Facchinei, C. Kanzow, S. Karl, and S. Sagratella: *The semismooth Newton method for the solution of quasi-variational inequalities.* Computational Optimization and Applications, published online, 2014, DOI: 10.1007/s10589-014-9686-4.

[51] F. Facchinei, C. Kanzow, and S. Sagratella: *QVILIB: A library of quasi-variational inequality test problems.* Pacific Journal of Optimization 9, 2013, pp. 225–250.

[52] F. Facchinei, C. Kanzow, and S. Sagratella: *Solving quasi-variational inequalities via their KKT-conditions.* Mathematical Programming 144, 2014, pp. 369–412.

[53] F. Facchinei and L. Lampariello: *Partial penalization for the solution of generalized Nash equilibrium problems.* Journal of Global Optimization 50, 2011, pp. 39–57.

[54] F. Facchinei and J.-S. Pang: *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volumes I and II.* Springer, New York, 2003.

[55] F. Facchinei and J.-S. Pang: *Exact penalty functions for generalized Nash problems.* In G. Di Pillo and M. Roma (eds.): *Large-Scale Nonlinear Optimization.* Series Nonconvex Optimization and Its Applications 83, Springer, New York, 2006, pp. 115–126.

[56] F. Facchinei and J.-S. Pang: *Nash equilibria: the variational approach.* In D.P. Palomar and Y.C. Eldar (eds.): *Convex Optimization in Signal Processing and Communications.* Cambridge University Press, 2009, pp. 443–493.

[57] F. Facchinei, V. Piccialli, and M. Sciandrone: *Decomposition algorithms for generalized potential games.* Computational Optimization and Applications 50, 2011, pp. 237–262.

[58] A.V. Fiacco: *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming.* Academic Press, New York, 1983.

[59] A. Fischer: *Local behavior of an iterative framework for generalized equations with non-isolated solutions.* Mathematical Programming 94, 2002, pp. 91–124.

[60] A. Fischer, M. Herrich, and K. Schönefeld: *Generalized Nash Equilibrium Problems - Recent Advances and Challenges.* Pesquisa Operacional, to appear.

[61] A. Fischer, P.K. Shukla, and M. Wang: *On the inexactness level of robust Levenberg-Marquardt methods.* Optimization 59, 2010, pp. 273–287.

[62] M. Fukushima: *Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems.* Mathematical Programming 53, 1992, pp. 99–110.

[63] M. Fukushima: *A class of gap functions for quasi-variational inequality problems.* Journal of Industrial and Management Optimization 3, 2007, pp. 165–171.

[64] M. Fukushima: *Restricted generalized Nash equilibria and controlled penalty algorithm.* Computational Management Science 8, 2011, pp. 201–218.

[65] J. Gauvin: *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex optimization.* Mathematical Programming 12, 1977, pp. 136–138.

[66] C. Geiger and C. Kanzow: *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben.* Springer, Berlin, Heidelberg, 1999.

[67] F. Giannessi: *Separation of sets and gap functions for quasi-variational inequalities.* In F. Giannessi and A. Maugheri (eds.): *Variational Inequality and Network Equilibrium Problems.* Plenum Press, New York, 1995, pp. 101–121.

[68] P.E. Gill, W. Murray, M.A. Saunders, G.W. Stewart, and M.H. Wright: *Properties of a representation of a basis for the null space.* Mathematical Programming 33, 1985, pp. 172–186.

[69] E.G. Gol'stein: *Theory of Convex Programming.* Translations of Mathematical Monographs 36, American Mathematical Society, Providence, Rhode Island, 1972.

[70] G. Gürkan and J.-S. Pang, *Approximations of Nash equilibria.* Mathematical Programming 117, 2009, pp. 223–253.

[71] P.T. Harker: *Generalized Nash games and quasi-variational inequalities.* European Journal of Operational Research 54, 1991, pp. 81–94.

[72] N. Harms, T. Hoheisel, and C. Kanzow: *On a smooth dual gap function for a class of quasi-variational inequalities.* Journal of Optimization Theory and Applications 163, 2014, pp. 413–438.

[73] N. Harms, T. Hoheisel, and C. Kanzow: *On a smooth dual gap function for a class of player convex generalized Nash equilibrium problems.* Journal of Optimization Theory and Applications, published online, 2014, DOI: 10.1007/s10957-014-0631-6.

[74] N. Harms, C. Kanzow, and O. Stein: *On differentiability properties of player convex generalized Nash equilibrium problems.* Optimization, published online, 2013, DOI: 10.1080/02331934.2012.752822.

[75] N. Harms, C. Kanzow, and O. Stein: *Smoothness properties of a regularized gap function for quasi-variational inequalities.* Optimization Methods and Software 29, 2014, pp. 720–750.

[76] D.W. Hearn: *The gap function of a convex program.* Operations Research Letters 1, 1982, pp. 67–71.

[77] A. von Heusinger and C. Kanzow: *SC^1-optimization reformulations of the generalized Nash equilibrium problem.* Optimization Methods and Software 23, 2008, pp. 953–973.

[78] A. von Heusinger and C. Kanzow: *Optimization reformulations of the generalized Nash equilibrium problem using Nikaido-Isoda-type functions.* Computational Optimization and Applications 43, 2009, pp. 353–377.

[79] A. von Heusinger and C. Kanzow: *Relaxation methods for generalized Nash equilibrium problems with inexact line search.* Journal of Optimization Theory and Applications 143, 2009, pp. 159–183.

[80] A. von Heusinger, C. Kanzow, and M. Fukushima: *Newton's method for computing a normalized equilibrium in the generalized Nash game through fixed point formulation.* Mathematical Programming 132, 2012, pp. 99–123.

[81] J.-B. Hiriart-Urruty and C. Lemaréchal: *Fundamentals of Convex Analysis.* Springer, Berlin, Heidelberg, 2001.

[82] W.W. Hogan: *Point-to-set maps in mathematical programming.* SIAM Review 15, 1973, pp. 591–603.

[83] W.W. Hogan: *Directional derivatives for extremal-value functions with applications to the completely convex case.* Operations Research 21, 1973, pp. 188–209.

[84] T. Horst and N.V. Thoai: *DC programming: overview.* Journal of Optimization Theory and Applications 103, 1999, pp. 1–43.

[85] R. Janin: *Directional derivative of the marginal function in nonlinear programming.* Mathematical Programming Study 21, 1984, pp. 110–126.

[86] J.B. Krawczyk and S. Uryasev: *Relaxation algorithms to find Nash equilibria with economic applications.* Environmental Modeling and Assessment 5, 2000, pp. 63–73.

[87] K. Kubota and M. Fukushima: *Gap function approach to the generalized Nash equilibrium problem.* Journal of Optimization Theory and Applications 144, 2010, pp. 511–531.

[88] J. Kyparisis: *On uniqueness of Kuhn-Tucker multipliers in nonlinear programming.* Mathematical Programming 32, 1985, pp. 242–246.

[89] F. Lenzen, F. Becker, J. Lellmann, S. Petra, and C. Schnörr: *A class of quasi-variational inequalities for adaptive image denoising and decomposition.* Computational Optimization and Applications 54, 2013, pp. 371–398.

[90] E.S. Levitin: *Perturbation Theory in Mathematical Programming and its Applications.* John Wiley & Sons, Chichester, New York, 1994.

[91] P. Marcotte and J.P. Dussault: *A sequential linear programming algorithm for solving monotone variational inequalities.* SIAM Journal on Control and Optimization 27, 1989, pp. 1260–1278.

[92] J.J. Moré and D.C. Sorensen: *Newton's method.* In: G.H. Golub (ed): *Studies in Numerical Analysis.* The Mathematical Association of America, Washington, D.C., 1984, pp. 29–82.

[93] U. Mosco: *Implicit variational problems and quasi variational inequalities.* In: J. Gossez, E. Lami Dozo, J. Mawhin, and L. Waelbroeck (eds.): *Nonlinear Operators and the Calculus of Variations.* Lecture Notes in Mathematics Vol. 543, Springer, 1976, pp. 83–156.

[94] J.F. Nash: *Equilibrium points in n-person games.* Proceedings of the National Academy of Sciences of the USA 36, 1950, pp. 48–49.

[95] J.F. Nash: *Non-cooperative games.* Annals of Mathematics 54, 1951, pp. 286–295.

[96] Y. Nesterov and L. Scrimali: *Solving strongly monotone variational and quasi-variational inequalities.* CORE Discussion Paper 2006/107, Catholic University of Louvain, Center for Operations Research and Econometrics, 2006.

[97] H. Nikaido and K. Isoda: *Note on noncooperative convex games.* Pacific Journal of Mathematics 5, 1955, pp. 807–815.

[98] M.A. Noor: *An iterative scheme for a class of quasi variational inequalities.* Journal of Mathematical Analysis and Applications 110, 1985, pp. 463–468.

[99] M.A. Noor: *On merit functions for quasivariational inequalities.* Journal of Mathematical Inequalities 1, 2007, pp. 259–268.

[100] M.A. Noor: *On general quasi-variational inequalities.* J. King Saud University 24, 2012, pp. 81–88.

[101] M.A. Noor and K.I. Noor: *Some new classes of quasi split feasibility problems.* Applied Mathematics and Information Sciences 7, 2013, pp. 1547–1552.

[102] M.A. Noor, K.I. Noor, and A.G. Khan: *Some iterative schemes for solving extended general quasi variational inequalities.* Applied Mathematics and Information Sciences 7, 2013, pp. 917–925.

[103] J.V. Outrata and M. Kocvara: *On a class of quasi-variational inequalities.* Optimization Methods and Software 5, 1995, pp. 275–295.

[104] J.V. Outrata, M. Kocvara, and J. Zowe: *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results.* Kluwer Academic Publishers, Dordrecht and Boston, 1998.

[105] J.V. Outrata and J. Zowe: *A Newton method for a class of quasi-variational inequalities.* Computational Optimization and Applications 4, 1995, pp. 5–21.

[106] J.-S. Pang and M. Fukushima: *Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games.* Computational Management Science 2, 2005, pp. 21–56 (Erratum: ibid 6, 2009, pp. 373–375).

[107] J.S. Pang and D. Ralph: *Piecewise smoothness, local invertibility, and parametric analysis of normal maps.* Mathematics of Operations Research 21, 1996, pp. 401–426.

[108] J.-S. Pang and G. Scutari: *Nonconvex games with side constraints.* SIAM Journal on Optimization 21, 2011, pp. 1491–1522.

[109] J.-S. Pang and G. Scutari: *Joint sensing and power allocation in nonconvex cognitive radio games: Quasi-Nash equilibria.* IEEE Transactions on Signal Processing 61, 2013, pp. 2366–2382.

[110] J.-S. Pang, G. Scutari, F. Facchinei, and C. Wang: *Distributed power allocation with rate constraints in Gaussian parallel interference channels.* IEEE Transactions on Information Theory 54, 2008, pp. 3471–3489.

[111] E. Polak and G. Ribière: *Note sur la convergence de méthodes de directions conjuguées.* Revue Francaise d'Informatique et de Recherche Opérationnelle 16, 1969, pp. 35–43.

[112] L. Qi: *Convergence analysis of some algorithms for solving nonsmooth equations.* Mathematics of Operations Research 18, 1993, pp. 227–244.

[113] L. Qi and J. Sun: *A nonsmooth version of Newton's method.* Mathematical Programming 58, 1993, pp. 353–367.

[114] D. Ralph and S. Dempe: *Directional derivatives of the solution of a parametric nonlinear program.* Mathematical Programming 70, 1995, pp. 159–172.

[115] M. Raydan: *The Barzilai and Borwein gradient method for the large unconstrained minimization problem.* SIAM Journal on Optimization 7, 1997, pp. 26–33.

[116] R.T. Rockafellar: *Directional differentiability of the optimal value function in a nonlinear programming problem.* Mathematical Programming Study 21, 1984, pp. 213–226.

[117] R.T. Rockafellar and R.J.-B. Wets: *Variational Analysis.* A Series of Comprehensive Studies in Mathematics, Vol. 317, Springer, Berlin, Heidelberg, 1998.

[118] I.P. Ryazantseva: *First-order methods for certain quasi-variational inequalities in Hilbert space.* Computational Mathematics and Mathematical Physics 47, 2007, pp. 183–190.

[119] S. Scholtes: *Introduction to Piecewise Differentiable Equations.* Springer, New York, 2012.

[120] A. Schrijver: *Theory of Linear and Integer Programming.* Wiley, Interscience Series in Discrete Mathematics and Optimization, 1994.

[121] L. Scrimali: *Mixed behavior network equilibria and quasi-variational inequalities.* Journal of Industrial and Managament Optimization 5, 2009, pp. 363–379.

[122] A.H. Siddiqi and Q.H. Ansari: *Strongly nonlinear quasivariational inequalities.* Journal of Mathematical Analysis and Applications 149, 1990, pp. 444–450.

[123] I. Singer: *A Fenchel-Rockafellar type duality theorem for maximization.* Bulletin of the Australian Mathematical Society 20, 1979, pp. 193–198.

[124] O. Stein: *Bi-level Strategies in Semi-infinite Programming.* Kluwer, Boston, 2003.

[125] O. Stein: *On constraint qualifications in non-smooth optimization.* Journal of Optimization Theory and Applications 121, 2004, pp. 647–671.

[126] D. Sun, M. Fukushima, and L. Qi: *A computable generalized Hessian of the D-gap function and Newton-type methods for variational inequality problems.* Complementarity and Variational Problems: State of the Art, M.C. Ferris and J.-S. Pang (eds.), SIAM, Philadelphia, 1997, pp. 452–472.

[127] K. Taji: *On gap functions for quasi-variational inequalities.* Abstract and Applied Analysis, 2008, Article ID 531361.

[128] K. Taji and M. Fukushima: *A new merit function and a successive quadratic programming algorithm for variational inequality problems.* SIAM Journal on Optimization 6, 1996, pp. 704–713.

[129] K. Taji, M. Fukushima, and T. Ibaraki: *A globally convergent Newton method for solving strongly monotone variational inequalities.* Mathematical Programming 58, 1993, pp. 369–383.

[130] J.F. Toland: *Duality in non-convex optimization.* Journal of Mathematical Analysis and Applications 66, 1978, pp. 399–415.

[131] J.F. Toland: *A duality principle for non-convex optimisation and the calculus of variations.* Archive for Rational Mechanics and Analysis 71, 1979, pp. 41–61.

[132] D.M. Topkins and A.F. Veinott: *On the convergence of some feasible direction algorithms for nonlinear programming.* SIAM Journal on Control 5, 1967, pp. 268–279.

[133] S. Uryasev and R.Y. Rubinstein: *On relaxation algorithms in computation of noncooperative equilibria.* IEEE Transactions on Automatic Control 39, 1994, pp. 1263–1267.

[134] G. Wachsmuth: *On LICQ and the uniqueness of Lagrange multipliers.* Operations Research Letters 41, 2013, pp. 78–80.

[135] G. Zoutendijk: *Methods of Feasible Directions.* Elsevier, Amsterdam, 1960.