Constraint Qualifications and Stationarity Concepts for Mathematical Programs with Equilibrium Constraints

Michael L. Flegel

Dissertation

Institute of Applied Mathematics and Statistics
University of Würzburg

Constraint Qualifications and Stationarity Concepts for Mathematical Programs with Equilibrium Constraints

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

vorgelegt von

Michael L. Flegel

aus

Bad Kreuznach

Würzburg, 2005

Eingereicht am: 9. Dezember 2004
bei der Fakultät für Mathematik und Informatik

1. Gutachter: Prof. Dr. Christian Kanzow, Universität Würzburg
2. Gutachter: Prof. Dr. Jiří Outrata, Czech Academy of Sciences

Tag der mündlichen Prüfung: 11. März 2005

To my parents.

ای کاش که جای آرمیدن بودی
تا این ره دور را رسیدن بودی
کاش از پی صد هزار سال از دل خاک
مانند چمن امید بر دمیدن بودی
— رباعیات عمر خیام

"Would but the Desert of Fountain yield
one glimpse—if dimly, yet indeed, reveal'd,
Toward which the fainting Traveler might spring,
As springs the trampled herbage of the field!"
— Omar Khayyam, The Rubaiyat

Preface

The doctoral thesis you hold in your hands is the result of research conducted during my time as a post-graduate student at the Institute for Applied Mathematics and Statistics at the University of Würzburg. Its creation was a gradual process, intermediate goals manifesting in various preprints and papers. These can be seen as snapshots in time of the evolution of the material, the field, and myself that lead to the results found in the following pages.

It was in large part the insights and ideas of my supervisor, Christian Kanzow, that lead to the creation of our publications, and thereby inevitably to the creation of this thesis. At several points during its inception, it could have swayed onto a different path than was ultimately followed. The reason for its current form may be found in the guidance and, at times, purposeful restraint thereof, by Christian Kanzow. I am greatly indebted to him, not only professionally, but also personally, for providing me with this opportunity and for always being there when I needed his help. He has my deepest gratitude.

It was also him that introduced me to the international community of researchers in my field. Of these, I am particularly grateful to Jiří Outrata for agreeing to be the co-referee of this thesis, and also for the many interesting discussions on MPECs and more general topics that lead to some joint work with him and Christian Kanzow. Also, he pointed out an error in an early draft of Section 5.3, suggesting a fix, which I gratefully incorporated in the form of Lemma 5.26.

Another researcher who was of great help is Jane Ye. I thank her, in particular, for the in-depth discussions we had about the material covered in Section 5.2. Though I would like to take credit for it, this material is solely due to her.

As much as the content of this thesis is based on rigorous mathematical reasoning, it would not have been possible without the support I had at home, from my parents, my brother Sven, Oma, and my greater family. It was them that helped me through frustrating and trying times, as well as being there to share in the joyous moments. Thank you.

This thesis went through several drafts before reaching its current form. My thanks go out to Christian Nagel for proof-reading parts of the draft. Also, once more, Christian Kanzow proof-read large parts if not all of the material reproduced here (and in some cases penning them himself), simply due to his involvement in the publications that lead up to this thesis.

A few other people deserve special mention: Alexis Iglauer, for being a good friend; my colleagues (too many to all be listed; you know who you are), for providing a pleasant work-

ing environment; Carl Geiger, for igniting my love for mathematics; Manfred Dobrowolski, for his help in creating Figure 2.1; my students, for giving me something to look forward to every week; and finally, Niloufar Hoevels, for her help in selecting the quotation found on the previous page.

Michael L. Flegel Würzburg, March 2005

Contents

III

Abbreviations

ACQ	Abadie constraint qualification
AMPL	a mathematical programming language
BT	bundle trust region
CQ	constraint qualification
GCQ	Guignard constraint qualification
LICQ	linear independence constraint qualification
MFCQ	Mangasarian-Fromovitz constraint qualification
MPAEC	mathematical program with affine equilibrium constraints
MPCC	mathematical program with complementarity constraints
MPEC	mathematical program with equilibrium constraints
NCP	nonlinear complementarity problem
NLP	nonlinear program
OPVIC	optimization problem with variational inequality constraints
SCQ	Slater constraint qualification
SMFCQ	strict Mangasarian-Fromovitz constraint qualification
SQP	sequential quadratic programming
TNLP	tightened nonlinear program
WSCQ	weak Slater constraint qualification

Notation

Spaces and Orthants

$\mathbb{R}$	the real numbers
$\mathbb{R}_-$	the left half line
$\mathbb{R}_+$	the right half line
$\mathbb{R}^n$	the n-dimensional real vector space
$\mathbb{R}^n_-$	the nonpositive orthant in $\mathbb{R}^n$
$\mathbb{R}^n_+$	the nonnegative orthant in $\mathbb{R}^n$

Sets

$\{x\}$	the set consisting of the vector x
$\operatorname{conv} \mathcal{S}$	convex hull of the set $\mathcal{S}$
$\operatorname{cl}(\mathcal{S})$	closure of the set $\mathcal{S}$
$\mathcal{S}_1 \subseteq \mathcal{S}_2$	$\mathcal{S}_1$ is a subset of $\mathcal{S}_2$
$\mathcal{S}_1 \subset \mathcal{S}_2$	$\mathcal{S}_1$ is a proper subset of $\mathcal{S}_2$
$\mathcal{S}_1 \backslash \mathcal{S}_2$	the set of elements contained in $\mathcal{S}_1$ but not in $\mathcal{S}_2$
$B_\varepsilon(z)$	open ball of radius ε around z
(a, b)	an open interval in $\mathbb{R}$
$[a, b]$	a closed interval in $\mathbb{R}$
$\lvert\delta\rvert$	cardinality of the set δ
α	$\{i \mid G_i(z^*) = 0, H_i(z^*) > 0\}$
β	$\{i \mid G_i(z^*) = 0, H_i(z^*) = 0\}$, degenerate or biactive set
γ	$\{i \mid G_i(z^*) > 0, H_i(z^*) = 0\}$
$\mathcal{P}(\beta)$	set of partitions of β
$\mathcal{Z}$	the feasible region of the MPEC

Vectors

$x \in \mathbb{R}^n$	column vector in $\mathbb{R}^n$
(x, y)	column vector $(x^T, y^T)^T$
x_i	i-th component of x
x_δ	vector in $R^{\lvert\delta\rvert}$ consisting of components x_i, $i \in \delta$
$x \geq y$	componentwise comparison $x_i \geq y_i$, $i = 1, \dots, n$
$x > y$	strict componentwise comparison $x_i > y_i$, $i = 1, \dots, n$
$\min\{x, y\}$	the vector whose i-th component is $\min\{x_i, y_i\}$
$\max\{x, y\}$	the vector whose i-th component is $\max\{x_i, y_i\}$
$\lVert x\rVert$	Euclidean norm of x
$\lVert x\rVert_q$	q-norm of x
$e^i \in \mathbb{R}^n$	the i-th vector of the canonical basis of $\mathbb{R}^n$

Cones

$\mathcal{T}(z^*)$	tangent cone to the feasible set $\mathcal{Z}$ of the MPEC at z^*
$\mathcal{T}(z^*, \mathcal{D})$	tangent cone to an arbitrary set $\mathcal{D}$ at z^*
$\mathcal{T}^{lin}(z^*)$	linearized tangent cone of the MPEC at z^*
$\mathcal{T}^{lin}_{\text{MPEC}}(z^*)$	MPEC-linearized tangent cone of the MPEC at z^*
$\hat{N}(v, \Omega)$	Fréchet normal cone to Ω at v
$N^{\pi}(v, \Omega)$	proximal normal cone to Ω at v
$N(v, \Omega)$	limiting normal cone to Ω at v
$N^{\text{Cl}}(v, \Omega)$	Clarke normal cone to Ω at v
$N^{\text{conv}}(v, \Omega)$	convex normal cone to convex Ω at v
$\mathcal{C}^*$	dual cone of $\mathcal{C}$

Functions

$f : \mathbb{R}^n \to \mathbb{R}^m$	function that maps $\mathbb{R}^n$ to $\mathbb{R}^m$
$f_i : \mathbb{R}^n \to \mathbb{R}$	i-th component function of f
$\operatorname{gph} f$	graph of the function f
$\operatorname{epi} f$	epigraph of the function f
$\Phi : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$	a multifunction that maps $\mathbb{R}^p$ to subsets of $\mathbb{R}^q$
$\operatorname{gph} \Phi$	graph of the multifunction Φ
$\nabla f(z)$	gradient of the function $f : \mathbb{R}^n \to \mathbb{R}$ at z, column vector
$\nabla_x f(x, y)$	gradient of f with respect to the variable x
$f'(z)$	Jacobian in $\mathbb{R}^{m \times n}$ of $f : \mathbb{R}^n \to \mathbb{R}^m$ at z
$\hat{\partial} f(z)$	Fréchet subgradient of a function $f : \mathbb{R}^n \to \mathbb{R}$
$\partial^{\pi} f(z)$	proximal subgradient of a function $f : \mathbb{R}^n \to \mathbb{R}$
$\partial f(z)$	limiting subgradient of a function $f : \mathbb{R}^n \to \mathbb{R}$
$\partial^{\text{Cl}} f(z)$	Clarke subgradient of a function $f : \mathbb{R}^n \to \mathbb{R}$
$\partial^{\text{conv}} f(z)$	convex subgradient of a convex function $f : \mathbb{R}^n \to \mathbb{R}$
$\Pi_X(x)$	a (not necessarily unique) projection of x onto a closed set X
$\operatorname{dist}(x, X)$	Euclidean distance between x and closed X
$P_q(\cdot; \rho)$	exact l_q MPEC-penalty function

Sequences

$\{a^k\} \subseteq \mathbb{R}^n$	a sequence in $\mathbb{R}^n$
$a^k \to a$	a convergent sequence with limit a
$a^k \searrow a$	a convergent sequence in $\mathbb{R}$ with limit a and $a^k > a$ for all $k = 1, 2, \ldots$
$\lim_{k \to \infty} a^k$	limit of the convergent series $\{a^k\}$

Chapter 1

Introduction

In the 1940s, the term 'programming' was used by large organizations to describe planning and scheduling activities. It quickly became apparent that the amount of each activity could be represented by a *variable*, and that natural *constraints*, imposed on these variables, could be described mathematically by a set of equations and inequalities.

It is in the nature of such constraints that they in general allow more than a single set of variables to satisfy them. Thus, a solution could be chosen that was optimal in some sense. Maximizing profit or minimizing cost are typical examples of criteria to choose a solution of the constraints. This could also be formulated mathematically in form of an *objective* function.

This system, the objective function to be optimized subject to some constraints, was referred to as *mathematical programming*. The simplest form of such programs is the *linear program*. When computing machines became available in the late 1940s, the simplex method to solve linear programs was introduced, and implemented on these early ancestors to the modern computer. This is, incidentally, where the modern meaning of the term 'programming' originates from.

Since the advent of mathematical programming, as computing power increased, programs of higher and higher dimension and complexity could be solved. Over the last half century, the *nonlinear program*, the most natural generalization of the linear program, has been investigated in depth, both in terms of theory and numerical solution.

Ever more detailed distinctions concerning the constraint structure have been made, spawning whole research areas within mathematical programming. In this thesis, we will concentrate on one such special class of nonlinear program.

Consider, therefore, the constrained optimization problem

$$\begin{array}{ll} \min & f(z) \\ \text{s.t.} & g(z) \leq 0, \quad h(z) = 0, \\ & G(z) \geq 0, \quad H(z) \geq 0, \quad G(z)^T H(z) = 0, \end{array} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$, $G : \mathbb{R}^n \to \mathbb{R}^l$, and $H : \mathbb{R}^n \to \mathbb{R}^l$ are continuously differentiable functions. Due to the presence of the complementarity constraint $G(z)^T H(z) = 0$, such programs are known as *mathematical programs with complementarity* *constraints*, or *MPCCs*. Originally, programs of type (1.1) occurred in equilibrium problems, also giving rise to the name *mathematical programs with equilibrium constraints*, or *MPECs*. Since the latter yields the more pronounceable moniker, we shall refer to (1.1) as an MPEC in the following.

The continuous differentiability of the data in (1.1) may be relaxed in favor of Lipschitz or lower semicontinuity, yielding a more general program, and with it more general results. Since this does not provide any additional insight into the problem, however, we will assume continuous differentiability of our data. In those rare cases where it should become necessary for our discussion to drop the continuous differentiability, we will explicitly alert the reader to this fact.

An important question is whether problems of type (1.1) actually appear in real-world applications. At this point it should be noted that MPECs are often also stated as *optimization problems with variational inequality constraints*, or *OPVICs*. The KKT formulations of such OPVICs are then programs of type (1.1). See [42] for a detailed discussion of this.

MPECs occur, for instance, in economic analysis. Extending the Nash game to incorporate a leader who anticipates the moves of his competitors yields a leader-follower, or Stackelberg game, an MPEC, see [42, 54, 12]. Other applications include network design, facility location and production problems [42], as well as optimal chemical equilibria and environmental economics [12].

Many engineering problems also yield MPECs. One example is minimizing the surface area of a membrane subject to the constraint that it must touch a given obstacle, rigid or compliant, at a predetermined contact area [54]. Such problems occur, for example, in chip design. Alternate interpretations of the same problem also yield applications in lubrication problems and filtration of liquids in porous media.

Another engineering application is the shape design of elastic-plastic structures. Materials such as metals can be modeled to behave elastically up to certain deflection, after which they behave plastically. The point at which elastic behavior switches to plastic behavior can be described by a complementarity problem. Optimizing the shape of the structure in some way then yields an MPEC. This type of problem occurs in metal part design (such as load bearing beams) and can also be applied to optimizing masonry structures [54].

Coulomb friction describes the contact friction between two solid objects. If the two objects are in relative motion to each other, the magnitude of the friction force is proportional to the normal force between the two objects. If the two objects are resting, this force is nil. Switching from resting to sliding (or reversing direction) introduces a jump discontinuity into the system. Discretizing the resulting differential equation in a certain fashion yields a complementarity problem [69], see also [68]. Optimizing some aspect of the system then yields an MPEC.

One interesting application of this is the so-called Michael Schumacher problem [67]. It describes the object of navigating a race car, whose tires are modeled using Coulomb friction, around a race track in the shortest time possible.

For a more thorough treatment of possible applications of MPECs, the interested reader is referred to the extensive discussion in [54], as well as the monographs [10], [42], and [12].

The ultimate aim, of course, is to find a global optimizer of the MPEC (1.1), or any nonlinear program, for that matter. However, this is a rather high-handed goal. Instead, focus is most often placed on finding a local minimizer.

A major problem that one is faced with is identification of a local minimizer. For this purpose, various *stationarity concepts* have been introduced over the years. They are in general easier to verify than the optimality of any given point. However, the constraints of the nonlinear program must satisfy a *constraint qualification*, or *CQ*, in order for a local minimizer of this program to satisfy any stationarity concept.

This thesis is dedicated to the thorough investigation of MPECs, stationarity concepts and constraint qualifications that are tailored to MPECs, and the connection that exists between the various stationarity concepts and constraint qualifications. Only first order necessary conditions are investigated. Sufficient conditions, i.e. conditions that ensure that any stationarity concept implies the local optimality of a point, are for the most part not considered. This remains the object of future research.

Nonetheless, it is our opinion that, at the time of print, this thesis contains an exhaustive discussion of the state of the art of the first order theory of MPECs of the form (1.1). We are, however, by no means the first to attempt an investigation of MPECs. A barrage of publications exist in this field, many of which will be referred to throughout this thesis.

To gather a glimpse as to why MPECs need special treatment and cannot simply be viewed as standard nonlinear programs, we single out the complementarity constraint $G(z)^T H(z) = 0$. As will become apparent fairly quickly in our treatment of MPECs, this complementarity constraint is the source of all our worries.

In standard mathematical programming, the linear program is the simplest program that we can think of. If we imagine all the functions, f, g, h, G, and H, to be affine, the complementarity term $G(z)^T H(z)$ would still be quadratic, leaving us with a constraint that is difficult to handle. Of course, if either G_i or H_i is constant for all $i = 1, \ldots, l$, the complementarity term $G(z)^T H(z)$ would be affine. However, if the feasible set of the MPEC (1.1) is nonempty, constant constraints can simply be dropped. If the feasible set is empty, the problem of finding a solution is moot. We can, therefore, assume without loss of generality that none of the constraint functions is constant.

Having motivated why the complementarity constraint $G(z)^T H(z) = 0$ is problematic, we investigate it in a little more detail in the following. Together with the constraints $G(z) \geq 0$ and $H(z) \geq 0$, it quickly becomes apparent that for all $i = 1, \ldots, l$, either $G_i(z) = 0$ or $H_i(z) = 0$, *or both*, must vanish. The separate treatment of these three cases will become of prime importance. We therefore introduce sets, α, β, γ, in the following fashion: For any given feasible point z^* of the MPEC (1.1), let

$$\begin{aligned} \alpha &:= \alpha(z^*) := \{i \mid G_i(z^*) = 0,\ H_i(z^*) > 0\}, \\ \beta &:= \beta(z^*) := \{i \mid G_i(z^*) = 0,\ H_i(z^*) = 0\}, \\ \gamma &:= \gamma(z^*) := \{i \mid G_i(z^*) > 0,\ H_i(z^*) = 0\}. \end{aligned} \tag{1.2}$$

Note that, although not apparent in their notation, these sets depend on the point z^*.

The set β is called the *degenerate* or *biactive* set. If it is empty, we say that z^* satisfies *strict complementarity*. In this case, all stationarity concepts tailored to MPECs collapse into a single one, and we are left with few results of interest.

In fact, many of the early attempts at coping with MPECs assumed strict complementarity to hold. It was quickly discovered, however, that in general, strict complementarity cannot be expected to hold. Perusing any of the numerous examples found in the literature and this thesis demonstrate that even the simplest MPECs do not satisfy strict complementarity. We therefore assume throughout this monograph that $\beta \neq \varnothing$.

The MPEC (1.1), as well as the index sets (1.2), will be constant companions for the remainder of this thesis.

We now discuss the structure of this thesis, which, in essence, is divided into two parts. Part I contains a discussion on constraint qualifications. Chapter 2 contains a recapitulation of constraint qualifications known from standard nonlinear programming and how they apply to MPECs. In this application it will become clear that common constraint qualifications are, in general, too strong for MPECs, necessitating the introduction of constraint qualifications tailored to MPECs. This is done in Chapter 3.

We then proceed, in Part II, to discuss stationarity concepts. Chapter 4 introduces and discusses the various stationarity concepts that have been the center of MPEC research. One of the stronger stationarity concepts, M-stationarity, is of particular importance to MPECs and is discussed in great detail in Chapter 5.

Chapter 5 can, in a way, be seen as the pièce de résistance of this thesis. A considerable amount of time was invested in the investigation of M-stationarity. The result is a short and concise proof of one of the more important results in the theoretical discussion of MPECs.

Rounding up the discussion of MPECs is a brief foray into their numerical treatment in Chapter 6.

Chapters 2 through 5, however, are the centerpiece of this thesis. As such, they should be read as a whole. Results stated in one chapter are readily referred to and used in another. Nevertheless, a certain amount of repetition was engaged in to abet the lucidity of the text.

To a certain degree, the original progression of how results were obtained was preserved. This gives some insight into the thought processes that led to the results.

Parts of the original work contained in this thesis have appeared in various publications [17, 20, 21] and preprints [15, 19, 16, 18, 22] which were created by the author, his advisor Christian Kanzow and, in the case of [22], a collaborator, Jiří Outrata. Where appropriate and necessary, other sources have been used. This is documented in each instance throughout this thesis. Nonetheless, a few publications deserve special mention at this stage.

The monographs [42, 54] introduced MPECs as a subject worthy of attention and as such played a large part in popularizing MPECs. The paper by Scheel and Scholtes [62] was among the first to introduce the format (1.1) of MPECs, along with introducing MPEC variants of standard constraint qualifications and the stationarity concepts they implied.

Pang and Fukushima [55] first considered Guignard CQ (although not calling it that) and investigated it in connection with MPEC stationarity concepts. Outrata [53] introduced M-stationarity, ushering in a large body of work, culminating in the obsoletion of all weaker stationarity concepts. This was first accomplished by Ye [78].

Notation

The notation we use in this thesis has, to a certain degree, become standard within the MPEC community. For a quick overview, we refer the reader to the preamble of this monograph.

For completeness' sake, we briefly summarize the notation used, going into some detail where necessary to avoid ambiguity.

The space of the real numbers is denoted by $\mathbb{R}$, the n-dimensional real vector space by $\mathbb{R}^n$.

Vectors $x \in \mathbb{R}^n$ are always understood to be column vectors, its transpose is given by x^T. Given two vectors x and y, we use $(x, y) := (x^T, y^T)^T$ for ease of notation.

Given a vector $x \in \mathbb{R}^n$, x_i denotes its i-th component, while for a set $\delta \subseteq \{1, \dots, n\}$ we denote by $x_\delta \in \mathbb{R}^{|\delta|}$ that vector which consists of those components of x which correspond to the indices in δ.

Functions such as max and min are understood componentwise when applied to vectors. Similarly, comparisons such as $\leq$ and $\geq$ are also understood componentwise.

By $\mathbb{R}^n_+ := \{x \in \mathbb{R}^n \mid x \geq 0\}$ we describe the nonnegative orthant of $\mathbb{R}^n$. Similarly, $\mathbb{R}^n_- := \{x \in \mathbb{R}^n \mid x \leq 0\}$ denotes the nonpositive orthant of $\mathbb{R}^n$.

By $B_\varepsilon(z)$ we denote the open ball around z with diameter ε with respect to the Euclidian norm; the dimension is given by the dimension of z. The q-norm of a vector $z \in \mathbb{R}^n$ is given by

$$\|z\|_q = \begin{cases} \left(\sum_{i=1}^n |z_i|^q\right)^{\frac{1}{q}} & : \ q \in [1, \infty), \\ \max_{i=1,\dots,n}\{|z_i|\} & : \ q = \infty. \end{cases}$$

If no index is given, $\|z\|$ denotes the Euclidean (2-)norm of the vector z.

The Euclidian distance between a point x and a closed set X is denoted by $\operatorname{dist}(x, X)$, while a (not necessarily unique) projection of x onto X is denoted by $\Pi_X(x)$.

For any set $\mathcal{S} \subseteq \mathbb{R}^n$, we denote by $\operatorname{cl}(\mathcal{S})$ and $\operatorname{conv}(\mathcal{S})$ its closure and convex hull, respectively.

Given a finite set $\beta \subset \mathbb{N}$, we define the set of its partitions as follows:

$$\mathcal{P}(\beta) := \{(\beta_1, \beta_2) \mid \beta_1 \cup \beta_2 = \beta, \beta_1 \cap \beta_2 = \varnothing\}.$$

Differential operators such as ∂ and ∇ are always applied to all arguments of the function following it and yield a column vector. If a differential operator is applied to only part of the arguments, this is denoted by an appropriate subscript, i.e. we have $\partial f(x, y) = (\partial_x f(x, y), \partial_y f(x, y))$. The Jacobian of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ at a point z is denoted by $f'(z) \in \mathbb{R}^{m \times n}$.

Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$, its graph is defined by

$$\operatorname{gph} f := \{(x, y) \in \mathbb{R}^{n+m} \mid y = f(x)\}.$$

If $m = 1$, we can additionally define its epigraph, given by

$$\operatorname{epi} f := \{(z, \zeta) \in \mathbb{R}^{n+1} \mid \zeta \geq f(z)\}.$$

Expanding the notion of a function, we can define a *multifunction*, or *set-valued function*. We denote this by $\Phi : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$, which states that elements of $\mathbb{R}^p$ are mapped to subsets of $\mathbb{R}^q$.

Similar to the case for normal functions, the graph of a multifunction is given by

$$\operatorname{gph} \Phi := \{(u, v) \in \mathbb{R}^{p+q} \mid v \in \Phi(u)\}.$$

Sequences in $\mathbb{R}^n$ are denoted by $\{a^k\} \subseteq \mathbb{R}^n$. To denote convergence, we write $a^k \to a$. It then holds that $\lim_{k\to\infty} a^k = a$. For a convergent sequence $\{a^k\} \subseteq \mathbb{R}$ with $a^k > a$ for all $k = 1, 2, \ldots$, we write $a^k \searrow a$.

Part I

Constraint Qualifications

Chapter 2

Standard Constraint Qualifications

In this chapter we will investigate constraint qualifications known from standard nonlinear programming and how they pertain to MPECs. To do this, we first recall the most common constraint qualifications in the context of standard nonlinear programming as well as some results. We then proceed to discuss these CQs as they are applied to MPECs.

2.1 The Standard Nonlinear Program

In order to facilitate the introduction and discussion of standard CQs, we will, for the time being, concentrate on a more general standard nonlinear program, which we state in the following manner:

$$\begin{array}{rl} \min & f(z) \\ \text{s.t.} & g(z) \leq 0, \\ & h(z) = 0, \end{array} \tag{2.1}$$

with continuously differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^p$. Note that our MPEC (1.1) is obviously a special case of this standard nonlinear program (2.1).

Undoubtedly among the better known CQs are the linear independence and Mangasarian-Fromovitz constraint qualifications. We will therefore introduce them in the following definition, and refer the interested reader to the literature [5, 4, 56, 27] for a more detailed discussion.

Definition 2.1 *Let z^* be a feasible point of the program* (2.1)*. We then say that*

(a) the linear independence constraint qualification*, or* LICQ*, holds at z^* if the gradient vectors*

$$\begin{array}{ll} \nabla g_i(z^*), & \forall i \in \mathcal{I}_g, \\ \nabla h_i(z^*), & \forall i = 1, \ldots, p \end{array} \tag{2.2}$$

are linearly independent, where

$$\mathcal{I}_g := \{i \mid g_i(z^*) = 0\} \tag{2.3}$$

is the set of the active inequalities of g in z^*;

(b) *the* Mangasarian-Fromovitz constraint qualification*, or* MFCQ *[45], holds at z^* if the gradient vectors*

$$\nabla h_i(z^*) \quad \forall i = 1, \ldots, p$$

are linearly independent and there exists a vector $d \in \mathbb{R}^n$ such that

$$\begin{aligned} \nabla g_i(z^*)^T d < 0, &\quad \forall i \in \mathcal{I}_g, \\ \nabla h_i(z^*)^T d = 0, &\quad \forall i = 1, \ldots, p. \end{aligned} \tag{2.4}$$

Note that, though not apparent from the notation, $\mathcal{I}_g$ depends on the vector z^*.

If it is clear from context, we will in the following tacitly omit "at z^*" when referring to constraint qualifications. This holds for all CQs we shall introduce in this chapter and the next.

It is well known that if MFCQ holds at a local minimizer z^* of the program (2.1), there exist Lagrange multipliers λ^g and λ^h such that the Karush-Kuhn-Tucker (KKT) conditions are satisfied (see Section 4.1 for details). We can now use such a Lagrange multiplier to define a strict variant of MFCQ, which will become important when we investigate first order optimality conditions for MPECs.

Definition 2.2 *Let z^* be a local minimizer of the program* (2.1)*. We then say that the* strict Mangasarian-Fromovitz constraint qualification*, or* SMFCQ *[24], holds if the gradient vectors*

$$\begin{aligned} \nabla g_i(z^*), &\quad \forall i \in \mathcal{J}_g, \\ \nabla h_i(z^*), &\quad \forall i = 1, \ldots, p, \end{aligned} \tag{2.5}$$

are linearly independent, and there exists a vector $d \in \mathbb{R}^n$ such that

$$\begin{aligned} \nabla g_i(z^*)^T d < 0, &\quad \forall i \in \mathcal{K}_g, \\ \nabla g_i(z^*)^T d = 0, &\quad \forall i \in \mathcal{J}_g, \\ \nabla h_i(z^*)^T d = 0, &\quad \forall i = 1, \ldots, p, \end{aligned} \tag{2.6}$$

where the sets

$$\mathcal{J}_g := \{i \in \mathcal{I}_g \mid \lambda_i^g > 0\}, \tag{2.7}$$
$$\mathcal{K}_g := \{i \in \mathcal{I}_g \mid \lambda_i^g = 0\} \tag{2.8}$$

depend on the Lagrange multiplier λ^g.

Note that, as in the case of $\mathcal{I}_g$, $\mathcal{J}_g$ and $\mathcal{K}_g$ depend on the vector z^*, though this is not explicitly made apparent in the notation.

The definition of SMFCQ requires the existence of a Lagrange multiplier. As we shall see in Proposition 2.4, SMFCQ implies MFCQ, which in turn guarantees the existence of such a Lagrange multiplier, provided z^* is a local minimizer. That is why we confine ourselves to local minimizers z^* of (2.1) in the definition of SMFCQ. This, however, is no restriction, since we are interested in constraint qualifications in order to obtain necessary optimality conditions (again, see Section 4.1). Therefore, the only points of interest to us are the local minimizers of the program (2.1).

It is well known that LICQ implies MFCQ, see, e.g., [27, Satz 2.41], [56, Figure 4] or [5, Theorem 6.2.3 iii]. The proof is elementary and its technique may also be applied to prove that SMFCQ is implied by LICQ, and that it implies MFCQ. This is made precise in Proposition 2.4. We wish to give a proof since we will use the same technique for the proof of Lemma 3.12. In order to do this, we require the following result.

Lemma 2.3 *Let $a_i \in \mathbb{R}^n$, $i = 1, \dots, m$ be given, a_i, $i = 1, \dots, s$, $s \leq m$ be linearly independent, and let there exist a vector $d \in \mathbb{R}^n$ such that*

$$\begin{aligned} a_i^T d &= 0, \qquad && \forall i = 1, \dots, s, \\ a_i^T d &< 0, \qquad && \forall i = s+1, \dots, m \end{aligned} \tag{2.9}$$

holds. Then, for every $r \leq s$, there exists a vector $b \in \mathbb{R}^n$ such that

$$\begin{aligned} a_i^T b &= 0, \qquad && \forall i = 1, \dots, r, \\ a_i^T b &< 0, \qquad && \forall i = r+1, \dots, m \end{aligned}$$

holds.

Proof. Let $\hat{d} \in \mathbb{R}^n$ satisfy the following set on linear equations:

$$\begin{aligned} a_i^T \hat{d} &= 0, \qquad && \forall i = 1, \dots, r, \\ a_i^T \hat{d} &= -1, \qquad && \forall i = r+1, \dots, s. \end{aligned}$$

We know that such a $\hat{d}$ exists since the a_i, $i = 1, \dots, s$ are linearly independent.

We now set

$$d(\delta) := d + \delta \hat{d},$$

with any vector d satisfying the conditions (2.9). It is easy to see that $d(\delta)$ satisfies the following conditions for all $\delta > 0$:

$$\begin{aligned} a_i^T d(\delta) &= 0, \qquad && \forall i = 1, \dots, r, \\ a_i^T d(\delta) &< 0, \qquad && \forall i = r+1, \dots, s. \end{aligned}$$

Since the inequality in (2.9) is strict, there exists an $\varepsilon > 0$ such that

$$a_i^T d(\delta) < 0, \qquad \forall i = s+1, \dots, m$$

for all $\delta \in (0, \varepsilon)$. Setting $b := d(\delta)$ for an arbitrary $\delta \in (0, \varepsilon)$ concludes the proof. $\square$

We can now use this lemma to prove the following proposition.

Proposition 2.4 *Let z^* be a feasible point of the program* (2.1)*. Then the following chain of implications holds at z^*:*

$$\text{LICQ } (\Longrightarrow \text{ SMFCQ}) \Longrightarrow \text{ MFCQ}.$$

The implication in brackets only holds when z^* is a local minimizer of (2.1)*, since only then is SMFCQ defined.*

Proof. We only prove the statement that LICQ implies SMFCQ since the remaining implication reduces to an identical application of Lemma 2.3.

Since $\mathcal{J}_g \subseteq \mathcal{I}_g$, the linear independence of the gradients in (2.2) obviously implies the linear independence of the gradients in (2.5). We then observe that the assumptions of Lemma 2.3 are satisfied with a_i equated to the gradients in (2.2) by setting $d = 0$ and noting that $s = m$. Applying Lemma 2.3 for an appropriate $r \leq m$ then yields that the conditions (2.6) are satisfied. This completes the proof. □

2.1.1 The Tangent and Linearized Tangent Cone

We now turn our attention to a different class of constraint qualifications. This class includes the well-known Abadie CQ. It is defined using the tangent and linearized tangent cones, which we introduce in the following definition.

Definition 2.5 *Let $\mathcal{Z}$ denote the feasible region of the program* (2.1)*. Then, for any $z^* \in \mathcal{Z}$, we call*

$$\mathcal{T}(z^*) := \left\{ d \in \mathbb{R}^n \;\middle|\; \exists \{z^k\} \subseteq \mathcal{Z}, \exists t_k \searrow 0 \,:\, z^k \to z^* \text{ and } \frac{z^k - z^*}{t_k} \to d \right\} \tag{2.10}$$

the tangent cone *of $\mathcal{Z}$ at z^*, and*

$$\begin{aligned} \mathcal{T}^{lin}(z^*) = \{ d \in \mathbb{R}^n \mid\ & \nabla g_i(z^*)^T d \leq 0, \quad \forall i \in \mathcal{I}_g, \\ & \nabla h_i(z^*)^T d = 0, \quad \forall i = 1, \ldots, p \ \} \end{aligned} \tag{2.11}$$

the linearized tangent cone *of $\mathcal{Z}$ at z^*.*

Obviously, the tangent cone depends on the feasible set $\mathcal{Z}$. We did not indicate this dependence for ease of notation. If the tangent cone to a set other than the feasible set $\mathcal{Z}$ is required, we indicate this by a second argument, $\mathcal{T}(z^*) = \mathcal{T}(z^*, \mathcal{Z})$.

Note that the tangent cone is also referred to as the *Bouligand* or *contingent* cone.

Further note that the inclusion

$$\mathcal{T}(z^*) \subseteq \mathcal{T}^{lin}(z^*) \tag{2.12}$$

(see, e.g., [27, Lemma 2.32]) always holds and that $\mathcal{T}(z^*)$ is closed (see, e.g., [27, Lemma 2.29]) but not necessarily convex, while $\mathcal{T}^{lin}(z^*)$ is polyhedral and hence both closed and convex.

Remark. We call $\mathcal{T}(z^*)$ and $\mathcal{T}^{lin}(z^*)$ *cones.* The mathematical definition of a cone $\mathcal{C}$ is that if $z \in \mathcal{C}$, then so is μz for all $\mu \geq 0$. Note that we specifically include the origin in the cone. This is handled differently across the literature. Sometimes the origin is included, and sometimes it is not.

Additionally, the linearized tangent cone $\mathcal{T}^{lin}(z^*)$ is a *polyhedral convex cone*, i.e. there exists a matrix A such that $\mathcal{T}^{lin}(z^*) = \{d \mid Ad \leq 0\}$, see [5, Theorem 3.2.4].

Before continuing, we will introduce the concept of the dual cone, which will crop up again and again throughout this thesis.

Definition 2.6 *Given an arbitrary cone $\mathcal{C}$, its* dual cone *$\mathcal{C}^*$ is defined as*

$$\mathcal{C}^* := \{v \in \mathbb{R}^n \mid v^T d \geq 0, \ \forall d \in \mathcal{C}\}. \tag{2.13}$$

Obviously, the dual cone is a cone, justifying its name.

Since we only need the dual of a cone, we have defined it this way, although, of course, the dual of an arbitrary set may be defined.

Cones and their duals have been the subject of extensive research in the past (see, in particular, [5, 60]). Note that in the literature, commonly the *polar cone* is considered (see, e.g., [61]), which is simply the negative of the dual cone.

We are now in a position to define two constraint qualifications using the tangent and linearized tangent cones.

Definition 2.7 *Let $z^* \in \mathcal{Z}$ be feasible for the program* (2.1)*. We say that the* Abadie constraint qualification*, or* ACQ*, holds at z^* if*

$$\mathcal{T}(z^*) = \mathcal{T}^{lin}(z^*), \tag{2.14}$$

and that the Guignard constraint qualification*, or* GCQ*, holds at z^* if*

$$\mathcal{T}(z^*)^* = \mathcal{T}^{lin}(z^*)^*. \tag{2.15}$$

For more information on the Abadie and Guignard CQs, we refer the reader to [5] and the overview article [56].

Abadie CQ obviously implies Guignard CQ. Indeed, Guignard CQ can be shown to be weaker by counter example (see Example 2.18 or [56]). In standard nonlinear programming, Abadie CQ is commonly the weakest CQ that is considered. This is because it usually is weak enough. As will become apparent in the next section, however, Abadie CQ is too strong in the context of MPECs. That is why we consider the lesser known, but weaker, Guignard CQ.

Additionally, Guignard CQ is, in a certain sense, the weakest possible constraint qualification for nonlinear programming, see Theorem 4.15 and the discussion following it.

We have already stated that Guignard CQ is implied by Abadie CQ, and we know how LICQ, SMFCQ, and MFCQ stand in relation to one another, see Proposition 2.4. The question how these two sets of constraint qualifications relate to each other is answered by the following proposition.

Proposition 2.8 *Let z^* be a feasible point of the program* (2.1)*. Then the following chain of implications holds:*

$$\text{LICQ } (\Longrightarrow \text{ SMFCQ}) \Longrightarrow \text{ MFCQ} \Longrightarrow \text{ ACQ} \Longrightarrow \text{ GCQ}.$$

The implication in brackets only holds when z^* is a local minimizer of (2.1)*, since only then is SMFCQ defined.*

Proof. The first two implications were proven in Proposition 2.4, while the last one follows directly from Definition 2.7. The remaining statement that MFCQ implies ACQ may be found in [27, Satz 2.39]. □

When Monique Guignard originally introduced Guignard CQ [30], she did so in an infinite-dimensional setting in what we will refer to as the *primal form*. In the case of programs of the form (2.1), however, the primal definition is equivalent to the dual formulation (2.15), as we will show in Corollary 2.11. The dual formulation is also commonly used in the finite-dimensional literature (see, e.g., [5, 66, 29]) to define Guignard CQ.

In order to introduce this primal formulation of Guignard CQ (see (2.18)), we collect some useful information about the dual cone in the following lemma.

Lemma 2.9 *Let $\mathcal{C}$ and $\tilde{\mathcal{C}}$ be arbitrary nonempty cones. Then the following hold:*

(i) $\mathcal{C}^*$ *is a closed convex cone;*

(ii) $\mathcal{C} \subseteq \tilde{\mathcal{C}}$ *implies* $\tilde{\mathcal{C}}^* \subseteq \mathcal{C}^*$*;*

(iii) $\mathcal{C} \subseteq \mathcal{C}^{**}$*;*

(iv) if $\mathcal{C}$ *is convex, then* $\mathcal{C}^{**} = \mathrm{cl}(C)$*;*

(v) $\mathcal{C}^{**} = \mathrm{cl}(\mathrm{conv}(\mathcal{C}))$*;*

(vi) $(\mathcal{C} \cup \tilde{\mathcal{C}})^* = \mathcal{C}^* \cap \tilde{\mathcal{C}}^*$.

Proof. Properties (i) and (ii) are easily verified (see also [5, Lemma 3.1.8]).

Considering the bidual cone of $\mathcal{C}$,

$$\begin{aligned}\mathcal{C}^{**} &= \{d \in \mathbb{R}^n \mid v^T d \geq 0 \quad \forall v \in \mathcal{C}^*\}\\ &= \{d \in \mathbb{R}^n \mid v^T d \geq 0 \quad \forall v : v^T \tilde{d} \geq 0 \quad \forall \tilde{d} \in \mathcal{C}\},\end{aligned}$$

property (iii) follows immediately.

Property (iv) has been proven in [5, Theorem 3.1.11], using separation theorems.

We know that $\mathcal{C}^{**}$ is closed and convex (property (i)). Therefore, since $\mathcal{C} \subseteq \mathcal{C}^{**}$ (property (iii)), we have $\mathrm{cl}(\mathrm{conv}(\mathcal{C})) \subseteq \mathcal{C}^{**}$. To prove the reverse inclusion, note that $\mathrm{cl}(\mathrm{conv}(\mathcal{C})) = \mathrm{cl}(\mathrm{conv}(\mathcal{C}))^{**}$ (property (iv)). Since $\mathcal{C} \subseteq \mathrm{cl}(\mathrm{conv}(\mathcal{C}))$, it follows that $\mathrm{cl}(\mathrm{conv}(\mathcal{C}))^* \subseteq \mathcal{C}^*$ by dualizing. Dualizing again yields $\mathcal{C}^{**} \subseteq \mathrm{cl}(\mathrm{conv}(\mathcal{C}))^{**} = \mathrm{cl}(\mathrm{conv}(\mathcal{C}))$. This proves property (v).

Property (vi) is the statement of [5, Theorem 3.1.9]. □

After first stating the following lemma, we will use the dual cone to deduce an equivalent formulation of GCQ in Corollary 2.11.

Lemma 2.10 *The following equality holds:*

$$\mathcal{T}(z^*)^* = \mathcal{T}^{\mathrm{G}}(z^*)^*, \tag{2.16}$$

with

$$\mathcal{T}^{\mathrm{G}}(z^*) := \mathrm{cl}(\mathrm{conv}(\mathcal{T}(z^*))). \tag{2.17}$$

Proof. Consider

$$\mathcal{T}^{\mathrm{G}}(z^*) = \mathrm{cl}(\mathrm{conv}(\mathcal{T}(z^*))) = \mathcal{T}(z^*)^{**},$$

where we used Lemma 2.9 (v) for the second equality. We dualize this to start off the following string of equations. Roman numerals indicate which point of Lemma 2.9 was used for that particular equality:

$$\mathcal{T}^{\mathrm{G}}(z^*)^* = \mathcal{T}(z^*)^{***} \overset{\text{(i),(iv)}}{=} \mathcal{T}(z^*)^*.$$

This completes the proof. □

Corollary 2.11 *Let z^* be a feasible point of the program* (2.1)*. Then GCQ holds in z^* if and only if*

$$\mathcal{T}^{\mathrm{G}}(z^*) = \mathcal{T}^{lin}(z^*) \tag{2.18}$$

holds, where $\mathcal{T}^{\mathrm{G}}(z^*)$ is defined in (2.17)*. We call* (2.18) *the* primal formulation *of GCQ.*

Proof. Dualizing (2.18) and using Lemma 2.10 yields

$$\mathcal{T}(z^*)^* = \mathcal{T}^{\mathrm{G}}(z^*)^* = \mathcal{T}^{lin}(z^*)^*,$$

i.e. GCQ in the form (2.18) implies (2.15).

To prove the converse implication, the following string of equalities again uses Roman numerals to indicate which point of Lemma 2.9 is used. In addition, the equality marked with ($*$) is acquired by dualizing (2.15).

$$\mathcal{T}^{lin}(z^*) \overset{\text{(iv)}}{=} \mathcal{T}^{lin}(z^*)^{**} \overset{(*)}{=} \mathcal{T}(z^*)^{**} \overset{\text{(v)}}{=} \mathrm{cl}(\mathrm{conv}(\mathcal{T}(z^*))) = \mathcal{T}^{\mathrm{G}}(z^*).$$

For the first equality note that $\mathcal{T}^{lin}(z^*)$ is polyhedral and as such closed and convex. □

Similar to calling (2.18) the primal formulation of Guignard CQ, we will, should the distinction become necessary, refer to (2.15) as the *dual formulation* of Guignard CQ.

In the context of Abadie CQ, we have the inclusion $\mathcal{T}(z^*) \subseteq \mathcal{T}^{lin}(z^*)$ (see (2.12)). We have similar results in the context of Guignard CQ, stated in the following Lemma.

Lemma 2.12 *Let z^* be a feasible point of the program* (2.1). *Then the inclusions*

$$\mathcal{T}^{lin}(z^*)^* \subseteq \mathcal{T}(z^*)^* \tag{2.19}$$

and

$$\mathcal{T}^{\mathrm{G}}(z^*) \subseteq \mathcal{T}^{lin}(z^*) \tag{2.20}$$

hold.

Proof. The first inclusion follows directly from the fact that $\mathcal{T}(z^*) \subseteq \mathcal{T}^{lin}(z^*)$ (see (2.12)) and Lemma 2.9 (ii). The second follows immediately from (2.12), (2.17) and the fact that $\mathcal{T}^{lin}(z^*)$ is both convex and closed. □

Results such as (2.12) and Lemma 2.12 are useful when we wish to demonstrate that GCQ or ACQ, respectively, hold. In that case, we only need to show that the remaining inclusion holds.

We now have two equivalent formulations of Guignard CQ. The dual formulation (2.15) will play the more important role in the following chapter, and indeed in the remainder of this thesis. The primal formulation (2.18) will be used to demonstrate an interesting fact about the Guignard CQ in the context of MPECs (see Lemma 2.20).

To close off this section, we will briefly discuss a special case of the program (2.1). Consider

$$\begin{array}{rl} \min & f(z) \\ \text{s.t.} & g(z) \leq 0, \\ & Bz + b = 0, \end{array} \tag{2.21}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable, the component functions $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are convex, $B \in \mathbb{R}^{p \times n}$, and $b \in \mathbb{R}^p$. Such a program is called a *convexly constrained nonlinear program.* Note that the feasible region is, in fact, convex.

For this type of program there exists the Slater CQ (see, e.g., [44, Slater's constraint qualification 5.4.3] or [27, Definition 2.44]), of which we will now define two variants.

Definition 2.13 *We say that the convexly constrained program* (2.21) *satisfies the* weak Slater constraint qualification*, or* WSCQ*, at z^* if there exists a vector $\hat{z} \in \mathbb{R}^n$ such that*

$$g_i(\hat{z}) < 0 \quad \forall i \in \mathcal{I}_g \quad \textit{and} \quad B\hat{z} + b = 0. \tag{2.22}$$

We say that it satisfies the Slater constraint qualification*, or* SCQ*, if there exists a vector $\hat{z} \in \mathbb{R}^n$ such that*

$$g(\hat{z}) < 0 \qquad \textit{and} \qquad B\hat{z} + b = 0. \tag{2.23}$$

Note that WSCQ depends on z^* through the set $\mathcal{I}_g$ of active inequality constraints in z^*. Clearly, SCQ implies WSCQ. The advantage of SCQ is that it can be stated independently of any point z^* that might be of interest to us.

We now state the following proposition, clarifying the relationship between Slater CQ and Abadie CQ. A proof is not given, since the technique is the same as we will use in the proof of Theorem 3.17. Also, proofs may be found in the literature (see, e.g., [27, Satz 2.45]).

Proposition 2.14 *Let z^* be feasible for the convexly constrained program* (2.21)*. If it satisfies WSCQ, it also satisfies ACQ.*

We have now introduced those CQs from standard nonlinear programming that we wish to investigate. In the next section we will discuss these CQs in the context of MPECs.

2.2 Application to MPECs

In this section we will apply the constraint qualifications introduced in the previous section to our MPEC (1.1). We skip LICQ and SMFCQ and immediately discuss MFCQ. As it turns out, MFCQ is violated at every feasible point of the MPEC (1.1) (see, e.g., [7]). Since both LICQ and SMFCQ imply MFCQ, they, too, are violated at every feasible point of the MPEC (1.1). This is made precise in the following proposition.

Proposition 2.15 *Let z^* be an arbitrary feasible point of the MPEC* (1.1)*. Then MFCQ is violated at z^*. In particular, LICQ and SMFCQ are also violated at z^*.*

Proof. It suffices to show that MFCQ is violated at z^*. To this end, let

$$\theta(z) := G(z)^T H(z) \tag{2.24}$$

denote the complementarity term of (1.1). Utilizing the function θ as well as the index sets α, β, and γ from (1.2), we apply the conditions (2.4) of MFCQ to obtain that the gradient vectors

$$\begin{aligned} &\nabla h_i(z^*), \quad \forall i = 1, \ldots, p, \\ &\nabla \theta(z^*) \end{aligned} \tag{2.25}$$

are linearly independent, and that there exists a vector $d \in \mathbb{R}^n$ such that

$$\begin{aligned} \nabla g_i(z^*)^T d &< 0, \quad \forall i \in \mathcal{I}_g, \\ \nabla h_i(z^*)^T d &= 0, \quad \forall i = 1, \ldots, p, \\ \nabla G_i(z^*)^T d &> 0, \quad \forall i \in \alpha \cup \beta, \\ \nabla H_i(z^*)^T d &> 0, \quad \forall i \in \gamma \cup \beta, \\ \nabla \theta(z^*)^T d &= 0. \end{aligned} \tag{2.26}$$

Now, consider the case that $\alpha = \gamma = \emptyset$. Then the gradient of θ is given by

$$\nabla \theta(z^*) = \sum_{i=1}^{l} \left[G_i(z^*) \nabla H_i(z^*) + H_i(z^*) \nabla G_i(z^*) \right] = 0.$$

This holds since $\beta = \{1, \ldots, l\}$ and therefore $G_i(z^*) = H_i(z^*) = 0$ for all $i = 1, \ldots, l$. The condition (2.25) for linear independence is therefore violated.

Assume therefore, that α or γ is not empty. It then holds that

$$\nabla \theta(z^*)^T d \;=\; \sum_{i=1}^{l} \left[G_i(z^*) \nabla H_i(z^*) + H_i(z^*) \nabla G_i(z^*) \right]^T d$$

$$= \sum_{i\in\gamma} \underbrace{G_i(z^*)}_{>0} \underbrace{\nabla H_i(z^*)^T d}_{>0} + \sum_{i\in\alpha} \underbrace{H_i(z^*)}_{>0} \underbrace{\nabla G_i(z^*)^T d}_{>0} \neq 0,$$

clearly violating the last line of (2.26). Note that only the partial sums are necessary since all other terms vanish due to the nature of the sets α, β, and γ.

Since z^* was chosen arbitrarily, MFCQ and hence SMFCQ and LICQ are violated at every feasible point of the MPEC (1.1). □

Next, let us consider the Slater CQs. Clearly, the constraints of the MPEC (1.1) can only be affine when either $G_i(\cdot)$ or $H_i(\cdot)$ is constant for every $i = 1, \ldots, l$. In this case, however, we can state the MPEC without the complementarity term and we are left with a standard nonlinear program of the form (2.1). Therefore, we can assume, without loss of generality, that the MPEC (1.1) is never a convexly constrained program, rendering the application of either Slater CQ useless.

Additionally, Slater CQ can be easily verified never to hold for any feasible point of (1.1): There exists no $\hat{z} \in \mathbb{R}^n$ such that $G(z^*) > 0$, $H(\hat{z}) > 0$, and $G(\hat{z})^T H(\hat{z}) = 0$.

The weak Slater CQ has a chance of holding in the nondegenerate case, i.e. if $\beta = \emptyset$. In this case, WSCQ requires the existence of a $\hat{z} \in \mathbb{R}^n$ such that

$$\begin{aligned}
&g_i(\hat{z}) < 0, && \forall i \in \mathcal{I}_g, \\
&h_i(\hat{z}) = 0, && \forall i = 1, \ldots, p, \\
&G_i(\hat{z}) > 0, && \forall i \in \alpha \cup \beta = \alpha, \\
&H_i(\hat{z}) > 0, && \forall i \in \gamma \cup \beta = \gamma, \\
&G(\hat{z})^T H(\hat{z}) = 0.
\end{aligned}$$

However, as mentioned in the introduction, the nondegenerate case is of little interest, since then all results that we will present collapse into the corresponding result from nonlinear programming.

Again, though, we wish to stress that the MPEC (1.1) is not convexly constrained in general, rendering the application of WSCQ of little use.

2.2.1 The Abadie Constraint Qualification

In this section, we turn our attention to the Abadie CQ. To this end, we need to determine the tangent and linearized tangent cone for a feasible point z^* of the MPEC (1.1). From here on out, we denote the feasible region of the MPEC by $\mathcal{Z}$. The tangent cone $\mathcal{T}(z^*)$ is then simply given by (2.10).

If we apply the rules of (2.11) to determine the linearized tangent cone $\mathcal{T}^{lin}(z^*)$ of the

MPEC (1.1) in z^*, we obtain

$$\begin{aligned}\mathcal{T}^{lin}(z^*) = \{d \in \mathbb{R}^n \mid\ & \nabla g_i(z^*)^T d \leq 0, && \forall i \in \mathcal{I}_g, \\ & \nabla h_i(z^*)^T d = 0, && \forall i = 1, \ldots, p, \\ & \nabla G_i(z^*)^T d \geq 0, && \forall i \in \alpha \cup \beta, \\ & \nabla H_i(z^*)^T d \geq 0, && \forall i \in \gamma \cup \beta, \\ & \nabla \theta(z^*)^T d = 0\,\},\end{aligned} \tag{2.27}$$

with $\theta(z) = G(z)^T H(z)$ given as in (2.24). An easy calculation shows that $\mathcal{T}^{lin}(z^*)$ may alternatively be expressed as

$$\begin{aligned}\mathcal{T}^{lin}(z^*) = \{d \in \mathbb{R}^n \mid\ & \nabla g_i(z^*)^T d \leq 0, && \forall i \in \mathcal{I}_g, \\ & \nabla h_i(z^*)^T d = 0, && \forall i = 1, \ldots, p, \\ & \nabla G_i(z^*)^T d = 0, && \forall i \in \alpha, \\ & \nabla H_i(z^*)^T d = 0, && \forall i \in \gamma, \\ & \nabla G_i(z^*)^T d \geq 0, && \forall i \in \beta, \\ & \nabla H_i(z^*)^T d \geq 0, && \forall i \in \beta\,\}.\end{aligned} \tag{2.28}$$

This representation of the linearized tangent cone $\mathcal{T}^{lin}(z^*)$ will become relevant when we investigate Guignard CQ.

Before discussing the Abadie CQ in a more rigorous fashion, we would like to point out a fundamental problem it has. Since the linearized tangent cone $\mathcal{T}^{lin}(z^*)$ is a polyhedral convex cone, it is, in particular, convex. However, the tangent cone $\mathcal{T}(z^*)$ is not convex in general. This is true for arbitrary nonlinear programs, of course, but it poses a particular problem for MPECs: MPECs, by their very nature, have nonconvex feasible regions, and as such are prone to nonconvex tangent cones. Example 2.18 demonstrates this behavior.

In order to shed some light on this qualitative discussion, we first need to introduce a program derived from the MPEC (1.1) for an arbitrary feasible point z^*: Given a partition $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$, we call the the nonlinear program

$$\begin{array}{ll} \min & f(z) \\ \text{s.t.} & g(z) \leq 0, \qquad\quad h(z) = 0, \\ & G_{\alpha\cup\beta_1}(z) = 0, \quad H_{\alpha\cup\beta_1}(z) \geq 0, \\ & G_{\gamma\cup\beta_2}(z) \geq 0, \quad H_{\gamma\cup\beta_2}(z) = 0 \end{array} \tag{2.29}$$

the (β_1, β_2)-restricted nonlinear program [55] and denote it by $\mathrm{NLP}_*(\beta_1, \beta_2)$. Note that the program $\mathrm{NLP}_*(\beta_1, \beta_2)$ depends on the vector z^* through the index sets α, β, and γ.

We will require the tangent and linearized tangent cones of $\mathrm{NLP}_*(\beta_1, \beta_2)$ in the following. Therefore, we denote the tangent cone of $\mathrm{NLP}_*(\beta_1, \beta_2)$ at z^* by

$$\mathcal{T}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*). \tag{2.30}$$

The linearized tangent cone at z^* takes on the following form:

$$
\begin{aligned}
\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*) = \{d \in \mathbb{R}^n \mid \nabla g_i(z^*)^T d \le 0, &\quad \forall i \in \mathcal{I}_g,\\
\nabla h_i(z^*)^T d = 0, &\quad \forall i = 1,\dots,p,\\
\nabla G_i(z^*)^T d = 0, &\quad \forall i \in \alpha \cup \beta_1,\\
\nabla H_i(z^*)^T d = 0, &\quad \forall i \in \gamma \cup \beta_2,\\
\nabla G_i(z^*)^T d \ge 0, &\quad \forall i \in \beta_2,\\
\nabla H_i(z^*)^T d \ge 0, &\quad \forall i \in \beta_1 \}.
\end{aligned}
\tag{2.31}
$$

The auxiliary program (2.29) will play an important role throughout this monograph. We therefore establish an important relationship between its tangent cone and the tangent cone of the MPEC (1.1) in the following lemma.

Lemma 2.16 *Let z^* be a feasible point of the MPEC* (1.1)*. Then the tangent cone of the MPEC* (1.1) *at z^* may be expressed by*

$$
\mathcal{T}(z^*) = \bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*). \tag{2.32}
$$

Proof. The idea of this proof is due to [55]. Let $\mathcal{Z}$ denote the feasible region of the MPEC (1.1) and let $\mathcal{Z}_{(\beta_1,\beta_2)}$ denote the feasible region of $\mathrm{NLP}_*(\beta_1,\beta_2)$ (see (2.29)).

We first show that there exists a neighborhood $\mathcal{N} \subseteq \mathbb{R}^n$ of z^* such that

$$
\mathcal{Z} \cap \mathcal{N} = \bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} (\mathcal{Z}_{(\beta_1,\beta_2)} \cap \mathcal{N}). \tag{2.33}
$$

To this end, note that the constraints $g(z) \le 0$ and $h(z) = 0$ occur in both $\mathcal{Z}$ and $\mathcal{Z}_{(\beta_1,\beta_2)}$, and can thus be ignored in our analysis.

The reverse inclusion '$\supseteq$' in (2.33) holds for arbitrary $\mathcal{N} \subseteq \mathbb{R}^n$ since obviously $\mathcal{Z}_{(\beta_1,\beta_2)} \subseteq \mathcal{Z}$ for all $(\beta_1,\beta_2) \in \mathcal{P}(\beta)$.

All that remains is to show that the forward inclusion '$\subseteq$' in (2.33) holds for some $\mathcal{N}$. For this purpose, consider an arbitrary $i \in \alpha$. Due to the nature of the set α, it holds that $G_i(z^*) = 0$ and $H_i(z^*) > 0$. Since H_i is continuous, there exists an $\varepsilon_i > 0$ such that $H_i(z) > 0$ for all $z \in B_{\varepsilon_i}(z^*)$. Hence, due to the complementarity term in $\mathcal{Z}$, it holds that $G_i(z) = 0$ for all $z \in B_{\varepsilon_i}(z^*) \cap \mathcal{Z}$.

Analogously, for every $i \in \gamma$, there exist a $\varepsilon_i > 0$ such that $G_i(z) > 0$ for all $z \in B_{\varepsilon_i}(z^*)$ and hence $H_i(z) = 0$ for all $z \in B_{\varepsilon_i}(z^*) \cap \mathcal{Z}$.

We now set $\varepsilon := \min_{i\in\alpha\cup\gamma}\{\varepsilon_i\}$ and consider an arbitrary $z \in B_\varepsilon(z^*) \cap \mathcal{Z}$. Due to the complementarity term in $\mathcal{Z}$, there exists a partition $(\beta_1,\beta_2) \in \mathcal{P}(\beta)$, dependent on z, such that

$$
\begin{aligned}
G_{\beta_1}(z) = 0, &\quad H_{\beta_1}(z) \ge 0,\\
G_{\beta_2}(z) \ge 0, &\quad H_{\beta_2}(z) = 0.
\end{aligned}
$$

Setting $\mathcal{N} := B_\varepsilon(z^*)$, we have shown that for an arbitrary $z \in \mathcal{Z} \cap \mathcal{N}$ there exists a partition $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$ such that $z \in \mathcal{Z}_{(\beta_1,\beta_2)} \cap \mathcal{N}$. This demonstrates that (2.33) does indeed hold.

Finally, consider the following string of equalities:

$$\begin{aligned}
\mathcal{T}(z^*, \mathcal{Z}) &= \mathcal{T}(z^*, \mathcal{Z} \cap \mathcal{N}) \\
&\overset{(2.33)}{=} \mathcal{T}(z^*, \bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} (\mathcal{Z}_{(\beta_1,\beta_2)} \cap \mathcal{N})) \\
&\overset{(*)}{=} \bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}(z^*, \mathcal{Z}_{(\beta_1,\beta_2)} \cap \mathcal{N}) \\
&= \bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}(z^*, \mathcal{Z}_{(\beta_1,\beta_2)})\,.
\end{aligned}$$

Here, the first and last equalities follow directly from the definition (2.10) of the tangent cone. The equation marked with $(*)$ immediately follows from the fact that (2.33) involves the union of only finitely many partitions (β_1, β_2).

Substituting the appropriate notation $\mathcal{T}(z^*)$ and $\mathcal{T}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$ concludes the proof.□

We are now able to state the following characterization of Abadie CQ for the MPEC (1.1), which in large parts is due to [55].

Proposition 2.17 *Let z^* be a feasible point of the MPEC* (1.1) *and assume that Abadie CQ holds for all* $\mathrm{NLP}_*(\beta_1, \beta_2)$*, i.e.*

$$\mathcal{T}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*) = \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*) \tag{2.34}$$

for all $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$. Then the following statements are equivalent:

(a) the Abadie constraint qualification holds in z^*;

(b) there exists a partition $(\hat\beta_1, \hat\beta_2) \in \mathcal{P}(\beta)$ such that $\mathcal{T}(z^*) = \mathcal{T}_{\mathrm{NLP}_*(\hat\beta_1,\hat\beta_2)}(z^*)$;

(c) there exists a partition $(\hat\beta_1, \hat\beta_2) \in \mathcal{P}(\beta)$ such that $\mathcal{T}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*) \subseteq \mathcal{T}_{\mathrm{NLP}_*(\hat\beta_1,\hat\beta_2)}(z^*)$ for all $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$.

Proof. (a)⇒(b) Since Abadie CQ holds, $\mathcal{T}(z^*) = \mathcal{T}^{lin}(z^*)$. Hence, $\mathcal{T}(z^*)$ is polyhedral and [55, Proposition 3] may be applied to yield the implication.

(b)⇒(a) Because (2.34) holds, we have $\mathcal{T}(z^*) = \mathcal{T}^{lin}_{\mathrm{NLP}_*(\hat\beta_1,\hat\beta_2)}(z^*)$, and hence $\mathcal{T}(z^*)$ is polyhedral. Consequently, $\mathcal{T}(z^*)$ is generated by linear constraints and is therefore equal to its linearization $\mathcal{T}^{lin}(z^*)$, i.e. (a) holds.

(b)⇔(c) The equivalence of (b) and (c) follows immediately from Lemma 2.16. Note that (2.34) is not needed for this equivalence. □

Geometrically, Proposition 2.17 can be interpreted as follows: While $\mathcal{T}(z^*)$ is equal to a finite union of (in the presence of (2.34)) polyhedral cones (see (2.32)), Abadie CQ holds if and only if there is at least one "big" tangent cone in the union (2.32) which contains all the other tangent cones. However, it is not difficult to find counterexamples, see Example 2.18.

The assumption in Proposition 2.17 that Abadie CQ holds for all $\mathrm{NLP}_*(\beta_1, \beta_2)$ is a very weak assumption. Since the $\mathrm{NLP}_*(\beta_1, \beta_2)$ do not have a complementarity constraint, the problems that have been discussed in this section don't apply. Therefore, assuming Abadie CQ for the $\mathrm{NLP}_*(\beta_1, \beta_2)$ is reasonable.

2.2.2 The Guignard Constraint Qualification

As we discussed in the previous section, a fundamental problem with Abadie CQ is that the tangent cone $\mathcal{T}(z^*)$ is nonconvex in general. If we look at the primal formulation (2.18) of Guignard CQ, we see that we consider the closure of the convex hull of $\mathcal{T}(z^*)$. Since the linearized tangent cone $\mathcal{T}^{lin}(z^*)$ is polyhedral and hence both closed and convex, Guignard CQ (in either form) has a better chance of holding than Abadie CQ.

The following example demonstrates that this problem with Abadie CQ, namely the nonconvexity of the tangent cone $\mathcal{T}(z^*)$, may (at least in this simple case) be sidestepped by Guignard CQ.

Example 2.18 Consider the program

$$\begin{array}{ll} \min & f(z^*) := z_1 + z_2 \\ \text{s.t.} & G(z) := z_1 \geq 0, \\ & H(z) := z_2 \geq 0, \\ & G(z)^T H(z) = z_1 z_2 = 0. \end{array}$$

The origin $z^* = 0 \in \mathbb{R}^2$ is the unique minimizer. The tangent cone $\mathcal{T}(z^*)$ is easily verified to be

$$\mathcal{T}(0) = \{d \in \mathbb{R}^2 \mid d_1 \geq 0, d_2 \geq 0, d_1 d_2 = 0\},$$

while the linearized tangent cone $\mathcal{T}^{lin}(0)$ is

$$\mathcal{T}^{lin}(0) = \{d \in \mathbb{R}^2 \mid d_1 \geq 0, d_2 \geq 0\} = \mathbb{R}^2_+.$$

Obviously, Abadie CQ does not hold since $\mathcal{T}(0) \neq \mathcal{T}^{lin}(0)$, whereas Guignard CQ does hold since $\mathcal{T}^{\mathrm{G}}(0) = \mathrm{cl}(\mathrm{conv}(\mathcal{T}(0)))$, like $\mathcal{T}^{lin}(0)$, is equal to the nonnegative orthant $\mathbb{R}^2_+$.

In the above example, the convex hull of $\mathcal{T}(0)$ is already closed, making the closure operation redundant. The question arises, whether this is so in general, and if it is not, when this might be the case. In order to answer this question, let us consider an arbitrary convex cone $\mathcal{C}$. Then the set $\{y \mid \mathcal{C} + y = \mathcal{C}\}$ is called the *lineality space* of $\mathcal{C}$. For a detailed discussion of the lineality space, see [60].

We now recall [5, Lemma 3.1.6] and [60, Corollary 9.1.3] in the following lemma.

Lemma 2.19 *Let $\mathcal{C}_1, \dots, \mathcal{C}_m$ be non-empty convex cones in $\mathbb{R}^n$. Then the following hold:*

(i)

$$\mathcal{C}_1 + \cdots + \mathcal{C}_m = \operatorname{conv}(\mathcal{C}_1 \cup \cdots \cup \mathcal{C}_m). \tag{2.35}$$

(ii) Additionally, let $\mathcal{C}_1, \dots, \mathcal{C}_m$ satisfy the following condition: if $d_i \in \operatorname{cl}(\mathcal{C}_i)$ for $i = 1, \dots, m$ and $d_1 + \cdots + d_m = 0$, then d_i is in the lineality space of $\operatorname{cl}(\mathcal{C}_i)$ for $i = 1, \dots, m$. Then

$$\operatorname{cl}(\mathcal{C}_1 + \cdots + \mathcal{C}_m) = \operatorname{cl}(\mathcal{C}_1) + \cdots + \operatorname{cl}(\mathcal{C}_m). \tag{2.36}$$

We will now use this lemma to prove the following result.

Lemma 2.20 *Given a feasible point z^* of the MPEC* (1.1)*, let Abadie CQ hold for all* $\mathrm{NLP}_*(\beta_1, \beta_2)$ *(see* (2.34)*). Then $\operatorname{conv}(\mathcal{T}(z^*))$ is closed, i.e. we have*

$$\mathcal{T}^{\mathrm{G}}(z^*) = \operatorname{cl}(\operatorname{conv}(\mathcal{T}(z^*))) = \operatorname{conv}(\mathcal{T}(z^*)).$$

Proof. Since, by assumption, Abadie CQ holds for all $\mathrm{NLP}_*(\beta_1, \beta_2)$ and $\mathcal{T}(z^*)$ can be written in the form (2.32), the following holds:

$$\mathcal{T}(z^*) = \bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*). \tag{2.37}$$

Next, we want to verify that Lemma 2.19 (ii) can be applied to the $\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$. To this end, we recall that $\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$ is polyhedral and hence a closed convex cone. Now let $d_{(\beta_1,\beta_2)} \in \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$ for each $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$ such that

$$\sum_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} d_{(\beta_1,\beta_2)} = 0. \tag{2.38}$$

We will now show that $d_{(\beta_1,\beta_2)}$ is in the lineality space of $\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$ for all $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$.

Multiplying (2.38) by $\nabla G_i(z^*)^T$ yields

$$\sum_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \nabla G_i(z^*)^T d_{(\beta_1,\beta_2)} = 0 \qquad \forall i = 1, \dots, l. \tag{2.39}$$

Since $d_{(\beta_1,\beta_2)} \in \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$, it holds, in particular, that $\nabla G_i(z^*)^T d_{(\beta_1,\beta_2)} \geq 0$ for $i \in \alpha \cup \beta$. Hence (2.39) implies $\nabla G_i(z^*)^T d_{(\beta_1,\beta_2)} = 0$ for all $i \in \alpha \cup \beta$ and $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$.

For an arbitrary partition $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$, we now want to verify that $\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*) + d_{(\beta_1,\beta_2)} = \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$, i.e. that $d_{(\beta_1,\beta_2)}$ is in the lineality space of $\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$. For the reverse inclusion, let d be arbitrarily given such that $d + d_{(\beta_1,\beta_2)} \in \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$. It then follows that

$$\nabla G_i(z^*)^T d = \nabla G_i(z^*)^T d + \underbrace{\nabla G_i(z^*)^T d_{(\beta_1,\beta_2)}}_{=0} = \nabla G_i(z^*)^T (d + d_{(\beta_1,\beta_2)}), \tag{2.40}$$

satisfies the appropriate conditions in the representation (2.31) of $\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$ for all $i \in \alpha \cup \beta$.

Similar equalities can be shown to hold for $g'_{\mathcal{I}_g}(z^*)d$, $h'(z^*)d$, and $H'_{\gamma\cup\beta}(z^*)d$. This demonstrates that $d \in \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$.

Conversely, we choose an arbitrary $d \in \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$. Since $d_{(\beta_1,\beta_2)} \in \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$ and $\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$ is convex, it holds by standard properties of convex cones that $d + d_{(\beta_1,\beta_2)} \in \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$. Hence we have proven that $\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)+d_{(\beta_1,\beta_2)} = \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$.

Therefore, $d_{(\beta_1,\beta_2)}$ is in the lineality space of $\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$ and we may apply Lemma 2.19 (ii). Note that $\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$ is convex, allowing the application of Lemma 2.19.

Now consider the following:

$$
\begin{aligned}
\mathrm{cl}(\mathrm{conv}(\mathcal{T}(z^*))) &\overset{(2.37)}{=} \mathrm{cl}(\mathrm{conv}\Big(\bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)\Big)) \\
&\overset{(2.35)}{=} \mathrm{cl}(\sum_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)) \\
&\overset{(2.36)}{=} \sum_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathrm{cl}(\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)) \\
&= \sum_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*) \\
&\overset{(2.35)}{=} \mathrm{conv}\Big(\bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)\Big) \\
&\overset{(2.37)}{=} \mathrm{conv}(\mathcal{T}(z^*)).
\end{aligned}
$$

This concludes the proof. $\square$

We wish to point out that, as in the case of Proposition 2.17, assuming Abadie CQ for all $\mathrm{NLP}_*(\beta_1, \beta_2)$ is a very weak assumption.

Since the tangent cone $\mathcal{T}(z^*)$ is always closed, one might hope that the result of Lemma 2.20 held for the convex hull of arbitrary closed cones. The following example, communicated to us by Marco López [41], shows, however, that such a statement is not true in general. Consequently, the statement of Lemma 2.20 is a property of MPECs (under certain assumptions) and is violated in a more general setting.

Example 2.21 Consider the closed, nonconvex cone in $\mathbb{R}^3$ generated by the set

$$\mathcal{S} = \{(-1,0,0)^T\} \cup \{x \in \mathbb{R}^3 \mid \|x - (2,0,1)^T\| \leq 1\}.$$

The convex hull of cone($\mathcal{S}$) (see Figure 2.1) is

$$\mathrm{conv}(\mathrm{cone}(\mathcal{S})) = \{x \in \mathbb{R}^3 \mid x_3 > 0\} \cup \{x \in \mathbb{R}^3 \mid x_2 = x_3 = 0\},$$

which is not closed.

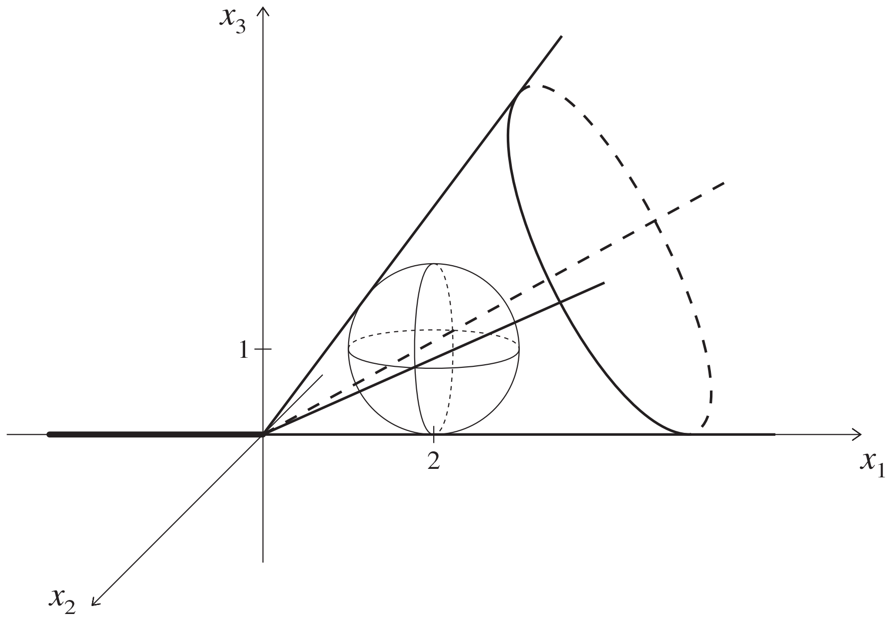

Figure 2.1: Illustration of cone($\mathcal{S}$) from Example 2.21.

The purpose of this chapter was to apply several constraint qualifications known from standard nonlinear programming to MPECs. We saw that most of the standard constraint qualifications are violated in every feasible point of the MPEC (1.1). Although Abadie CQ has a chance of being satisfied, Proposition 2.17 demonstrates that this is true only under very restrictive circumstances. One obvious drawback of the Abadie CQ was the nonconvexity of the tangent cone $\mathcal{T}(z^*)$. Guignard circumnavigates this issue by taking the convex hull of $\mathcal{T}(z^*)$. Thus, Guignard CQ seems to be the only constraint qualification from nonlinear programming that has any merit in the context of MPECs. This is also substantiated to a certain extent by the results of the following chapters.

Chapter 3

MPEC Constraint Qualifications

In this chapter we will investigate constraint qualifications that are tailored specifically to MPECs. The need for these arose from the fact that standard constraint qualifications were considered too strong, an aspect of MPECs we discussed in Chapter 2. Though we elucidated that Guignard is weak enough for MPECs, it has been largely overlooked in the MPEC community. This may be due to the fact that it is of little importance in standard nonlinear programming, since there Abadie CQ is commonly believed to be weak enough.

In this chapter, we first investigate MPEC constraint qualifications that have been introduced in past research. We then proceed to introduce a new type of tangent cone associated with MPECs and use it to define new constraint qualifications. We also discuss the relationship between these new CQs as well as their relationship to constraint qualifications introduced in Chapter 2. Just like linear and convexly constrained programs are commonly discussed as special cases of nonlinear programs, we will discuss special cases of MPECs and how they are connected to the MPEC constraint qualifications.

3.1 The Tightened Nonlinear Program

We commence by recalling some constraint qualifications for MPECs. These have appeared before, most notably in [62]. In order to do this, we first need to introduce the following program, dependent on z^*, and called the *tightened nonlinear program* TNLP(z^*):

$$\begin{array}{lll} \min & f(z) & \\ \text{s.t.} & g(z) \leq 0, & h(z) = 0, \\ & G_{\alpha\cup\beta}(z) = 0, & G_\gamma(z) \geq 0, \\ & H_\alpha(z) \geq 0, & H_{\gamma\cup\beta}(z) = 0. \end{array} \tag{3.1}$$

The above nonlinear program is called *tightened* since the feasible region is a subset of the feasible region of the MPEC (1.1). This implies that if z^* is a local minimizer of the MPEC (1.1), then it is also a local minimizer of the corresponding tightened nonlinear program TNLP(z^*).

We now use this program to define suitable MPEC variants of the standard linear independence, Mangasarian-Fromovitz- and strict Mangasarian-Fromovitz constraint qualifications.

Definition 3.1 *The MPEC* (1.1) *is said to satisfy* MPEC-LICQ (MPEC-MFCQ, MPEC-SMFCQ) *in a suitable vector z^* if the corresponding* TNLP(z^*) *satisfies LICQ (MFCQ, SMFCQ) in that vector z^*.*

Note that, as for standard SMFCQ, MPEC-SMFCQ is only defined for local minimizers of the MPEC (1.1). As discussed above, a local minimizer of the MPEC (1.1) is a local minimizer of the associated program TNLP(z^*). Since SMFCQ is only defined for local minimizers of TNLP(z^*), MPEC-SMFCQ is only defined for local minimizers of TNLP(z^*) and hence of the MPEC (1.1).

Since the MPEC variants of LICQ, SMFCQ, and MFCQ are defined using the nonlinear program TNLP(z^*), they inherit all properties that are known for TNLP(z^*). In particular, this includes the following corollary to Proposition 2.4.

Corollary 3.2 *Let z^* be a feasible point of the MPEC* (1.1)*. Then the following chain of implications holds at this point:*

$$\textit{MPEC-LICQ} \ (\Longrightarrow \textit{MPEC-SMFCQ}) \Longrightarrow \textit{MPEC-MFCQ}.$$

The implication in brackets only holds when z^* is a local minimizer of (1.1)*, since only then is MPEC-SMFCQ defined.*

Since we will need them in in the following, we shall explicitly write down the constraint qualifications from Definition 3.1. The MPEC-LICQ expands to the condition that the gradient vectors

$$\begin{aligned} &\nabla g_i(z^*), && \forall i \in \mathcal{I}_g, \\ &\nabla h_i(z^*), && \forall i = 1, \dots, p, \\ &\nabla G_i(z^*), && \forall i \in \alpha \cup \beta, \\ &\nabla H_i(z^*), && \forall i \in \gamma \cup \beta \end{aligned} \tag{3.2}$$

must be linearly independent. Here, $\mathcal{I}_g = \{i \mid g_i(z^*) = 0\}$ is defined as in (2.3). MPEC-LICQ can also be defined using the so-called *relaxed* nonlinear program, which we shall not elaborate upon here. The resulting definition, however, is the same, see, e.g., [55].

Similarly, MPEC-MFCQ expands to the following set of conditions: The gradient vectors

$$\begin{aligned} &\nabla h_i(z^*), && \forall i = 1, \dots, p, \\ &\nabla G_i(z^*), && \forall i \in \alpha \cup \beta, \\ &\nabla H_i(z^*), && \forall i \in \gamma \cup \beta \end{aligned} \tag{3.3}$$

are linearly independent, and there exists a vector $d \in \mathbb{R}^n$ such that

$$
\begin{aligned}
\nabla g_i(z^*)^T d < 0, &\quad \forall i \in \mathcal{I}_g, \\
\nabla h_i(z^*)^T d = 0, &\quad \forall i = 1, \dots, p, \\
\nabla G_i(z^*)^T d = 0, &\quad \forall i \in \alpha \cup \beta, \\
\nabla H_i(z^*)^T d = 0, &\quad \forall i \in \gamma \cup \beta.
\end{aligned}
\tag{3.4}
$$

Recall that a local minimizer z^* of the MPEC (1.1) is a local minimizer of the corresponding TNLP(z^*). Furthermore, MPEC-MFCQ at z^* is defined as standard MFCQ of TNLP(z^*). Together with the local optimality of z^*, this implies the existence of a Lagrange multiplier λ^* such that (z^*, λ^*) satisfies the KKT conditions of TNLP(z^*) (see Propositions 4.4 and 2.8). Therefore, if we assume that MPEC-MFCQ holds for a local minimizer z^* of the MPEC (1.1), we can use any Lagrange multiplier λ^* (which we now know exists) to define the MPEC-SMFCQ, i.e., taking (z^*, λ^*), we require the following to hold: The gradient vectors

$$
\begin{aligned}
\nabla g_i(z^*), &\quad \forall i \in \mathcal{J}_g, \\
\nabla h_i(z^*), &\quad \forall i = 1, \dots, p, \\
\nabla G_i(z^*), &\quad \forall i \in \alpha \cup \beta, \\
\nabla H_i(z^*), &\quad \forall i \in \gamma \cup \beta
\end{aligned}
\tag{3.5}
$$

are linearly independent, and there exists a vector $d \in \mathbb{R}^n$ such that

$$
\begin{aligned}
\nabla g_i(z^*)^T d = 0, &\quad \forall i \in \mathcal{J}_g, \\
\nabla g_i(z^*)^T d < 0, &\quad \forall i \in \mathcal{K}_g, \\
\nabla h_i(z^*)^T d = 0, &\quad \forall i = 1, \dots, p, \\
\nabla G_i(z^*)^T d = 0, &\quad \forall i \in \alpha \cup \beta, \\
\nabla H_i(z^*)^T d = 0, &\quad \forall i \in \gamma \cup \beta.
\end{aligned}
\tag{3.6}
$$

Here, $\mathcal{J}_g$ and $\mathcal{K}_g$ are defined as in (2.7) and (2.8), respectively. Note that, as in the case of standard nonlinear programming, the above assumption that MPEC-MFCQ holds in order to define MPEC-SMFCQ is no restriction since MPEC-SMFCQ implies MPEC-MFCQ (see Corollary 3.2).

3.1.1 Relationship to Standard CQs

We dedicate the remainder of this section to discussing the relationship of the MPEC constraint qualifications introduced in Definition 3.1 to the constraint qualifications known from standard nonlinear programming.

A quick comparison yields that MPEC-LICQ differs from standard LICQ for the MPEC (1.1) by the absence of the gradient of the complementarity term $\nabla\theta(z^*)$.

Far more interesting is the relationship of the various MPEC constraint qualifications to the Guignard CQ. First, we will show that MPEC-LICQ implies Guignard CQ. Before we can do so, however, we need to state the following lemma, which will facilitate the proof of Theorem 3.4.

Lemma 3.3 *Let the cones*

$$\begin{aligned} \mathcal{K}_1 := \{d \in \mathbb{R}^n \mid\ & a_i^T d \geq 0, \qquad \forall i = 1, \ldots, k, \\ & b_j^T d = 0, \qquad \forall j = 1, \ldots, l\} \end{aligned} \tag{3.7}$$

and

$$\begin{aligned} \mathcal{K}_2 = \{v \in \mathbb{R}^n \mid\ & v = \sum_{i=1}^{k} \alpha_i a_i + \sum_{j=1}^{l} \beta_j b_j \\ & \alpha_i \geq 0, \quad \forall i = 1, \ldots, k\}, \end{aligned} \tag{3.8}$$

be given. Then $\mathcal{K}_1 = \mathcal{K}_2^*$ *and* $\mathcal{K}_1^* = \mathcal{K}_2$.

Proof. (See also [5, Theorem 3.2.2].)

$$\begin{aligned} \mathcal{K}_2^* &= \{d \in \mathbb{R}^n \mid v^T d \geq 0 \quad \forall v \in \mathcal{K}_2\} \\ &= \{d \in \mathbb{R}^n \mid \sum_{i=1}^{k} \alpha_i a_i^T d + \sum_{j=1}^{l} \beta_i b_j^T d \geq 0 \quad \forall \alpha_i \geq 0,\ i = 1, \ldots, k\} \\ &\overset{(*)}{=} \{d \in \mathbb{R}^n \mid a_i^T d \geq 0\ \forall i = 1, \ldots, k, \quad b_j^T d = 0\ \forall j = 1, \ldots, l\} \\ &= \mathcal{K}_1. \end{aligned}$$

The reverse direction of the equation marked with $(*)$ is immediately obvious. Let us now consider the forward direction. For every $i_0 \in \{1, \ldots, k\}$, set $\alpha_{i_0} := 1$, $\alpha_i := 0$ for every $i \neq i_0$, and $\beta_j = 0$ for every $j = 1, \ldots, l$. Then $a_{i_0}^T d \geq 0$. Since $i_0 \in \{1, \ldots, k\}$ was chosen arbitrarily, it follows that $a_i^T d \geq 0$ for all $i = 1, \ldots, k$. Similarly, it can be proven that $b_j^T d = 0$ for all $j = 1, \ldots, l$.

Since $\mathcal{K}_2$ is a polyhedral cone, we can invoke Lemma 2.9 (v) to infer that $\mathcal{K}_1^* = \mathcal{K}_2$, proving the second equality. □

We now use Lemma 3.3 to prove the following theorem.

Theorem 3.4 *If a feasible point* z^* *of the MPEC* (1.1) *satisfies MPEC-LICQ, it also satisfies Guignard CQ.*

Proof. Since $\mathcal{T}^{lin}(z^*)^* \subseteq \mathcal{T}(z^*)^*$ (see Lemma 2.12), it suffices to show that

$$\mathcal{T}(z^*)^* \subseteq \mathcal{T}^{lin}(z^*)^* \tag{3.9}$$

holds. By virtue of Lemma 2.16, it holds that

$$\mathcal{T}(z^*) = \bigcup_{(\beta_1, \beta_2) \in \mathcal{P}(\beta)} \mathcal{T}_{\mathrm{NLP}_*(\beta_1, \beta_2)}(z^*).$$

Dualizing this yields

$$\mathcal{T}(z^*)^* = \bigcap_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)^* \tag{3.10}$$

(see Lemma 2.9 (vi)).

We now investigate the conditions (3.2), reproduced here for ease of reference. The gradient vectors

$$\begin{aligned}
&\nabla g_i(z^*), \quad && \forall i \in \mathcal{I}_g, \\
&\nabla h_i(z^*), \quad && \forall i = 1, \dots, p, \\
&\nabla G_i(z^*), \quad && \forall i \in \alpha \cup \beta, \\
&\nabla H_i(z^*), \quad && \forall i \in \gamma \cup \beta
\end{aligned}$$

are required to be linearly independent. This condition, however, is identical to the condition that standard LICQ holds for $\mathrm{NLP}_*(\beta_1, \beta_2)$ at z^* for every $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$. Hence, since MPEC-LICQ holds at z^*, LICQ holds for all $\mathrm{NLP}_*(\beta_1, \beta_2)$ at z^*. Therefore, Abadie CQ holds for each $\mathrm{NLP}_*(\beta_1, \beta_2)$ (see Proposition 2.8), i.e. we have

$$\mathcal{T}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*) = \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*) \tag{3.11}$$

for every $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$. Thus, we can apply Lemma 3.3 to the representation (2.31) of $\mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$, yielding the dual of $\mathcal{T}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)$ as follows:

$$\begin{aligned}
\mathcal{T}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)^* &= \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)^* \\
&= \{v \in \mathbb{R}^n \mid v = -\sum_{i\in\mathcal{I}_g} u_i^g \nabla g_i(z^*) - \sum_{i=1}^{p} u_i^h \nabla h_i(z^*) \\
&\qquad + \sum_{i\in\alpha\cup\beta} u_i^G \nabla G_i(z^*) + \sum_{i\in\gamma\cup\beta} u_i^H \nabla H_i(z^*), \\
&\qquad u^g_{\mathcal{I}_g} \geq 0, \quad u^G_{\beta_2} \geq 0, \quad u^H_{\beta_1} \geq 0\}.
\end{aligned} \tag{3.12}$$

Taking $v \in \mathcal{T}(z^*)^*$ arbitrarily, (3.10) yields that

$$v \in \mathcal{T}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)^* \quad \text{and} \quad v \in \mathcal{T}_{\mathrm{NLP}_*(\beta_2,\beta_1)}(z^*)^*$$

for an arbitrary partition $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$ and its “complement” $(\beta_2, \beta_1) \in \mathcal{P}(\beta)$.

Since all gradient vectors in (3.12) are linearly independent (MPEC-LICQ holds), $u^g_{\mathcal{I}_g}$, u^h, $u^G_{\alpha\cup\beta}$, and $u^H_{\gamma\cup\beta}$ are uniquely defined. Hence it follows from the fact that v is in both $\mathcal{T}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)^*$ and $\mathcal{T}_{\mathrm{NLP}_*(\beta_2,\beta_1)}(z^*)^*$ that $u^G_\beta \geq 0$ and $u^H_\beta \geq 0$. Therefore,

$$\begin{aligned}
v \in \{v \in \mathbb{R}^n \mid v = &-\sum_{i\in\mathcal{I}_g} u_i^g \nabla g_i(z^*) - \sum_{i=1}^{p} u_i^h \nabla h_i(z^*) \\
&+ \sum_{i\in\alpha\cup\beta} u_i^G \nabla G_i(z^*) + \sum_{i\in\gamma\cup\beta} u_i^H \nabla H_i(z^*), \\
&u^g_{\mathcal{I}_g} \geq 0, \quad u^G_\beta \geq 0, \quad u^H_\beta \geq 0\}
\end{aligned}$$

$$= \mathcal{T}^{lin}(z^*)^*,$$

which proves (3.9). Note that the above representation of $\mathcal{T}^{lin}(z^*)^*$ can be gleaned by applying Lemma 3.3 to the representation (2.28) of $\mathcal{T}^{lin}(z^*)$. □

Another constraint qualification that implies Guignard CQ is MPEC-SMFCQ, as we will state in the following theorem. We have, unfortunately, been unsuccessful in finding a direct proof such as in the case of MPEC-LICQ. Instead, the proof relies on results that we will present in Chapter 4. We therefore defer the proof until we have the tools we need (see page 68). Thematically the statement should occur here, however, which is why it is stated here.

Theorem 3.5 *If a local minimizer z^* of the MPEC* (1.1) *satisfies MPEC-SMFCQ, it also satisfies Guignard CQ.*

3.2 The MPEC-Linearized Tangent Cone

In Definition 3.1 we introduced MPEC variants of some common constraint qualifications. The question arises whether suitable variants of the Abadie and Guignard constraint qualifications can also be found.

Example 2.18 demonstrated that the problem with Abadie CQ was that the tangent cone $\mathcal{T}(z^*)$ was nonconvex in general, while the linearized tangent cone was convex by its very nature. Guignard CQ circumnavigated this problem by, in a sense, convexifying the tangent cone $\mathcal{T}(z^*)$ (see Corollary 2.11).

Another way to sidestep this problem of Abadie CQ is to "unconvexify" the linearized tangent cone. This suggests the definition of the following set, which we call the *MPEC-linearized tangent cone* for the MPEC (1.1) at z^*:

$$\begin{aligned}
\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*) := \{d \in \mathbb{R}^n \mid \; & \nabla g_i(z^*)^T d \le 0, \quad \forall i \in \mathcal{I}_g, \\
& \nabla h_i(z^*)^T d = 0, \quad \forall i = 1, \dots, p, \\
& \nabla G_i(z^*)^T d = 0, \quad \forall i \in \alpha, \\
& \nabla H_i(z^*)^T d = 0, \quad \forall i \in \gamma, \\
& \nabla G_i(z^*)^T d \ge 0, \quad \forall i \in \beta, \\
& \nabla H_i(z^*)^T d \ge 0, \quad \forall i \in \beta, \\
& (\nabla G_i(z^*)^T d) \cdot (\nabla H_i(z^*)^T d) = 0, \; \forall i \in \beta \}.
\end{aligned} \tag{3.13}$$

This set has appeared in [62, 55] before, but was not investigated further in either paper.

Note that although the cone $\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)$ is called the MPEC-*linearized* tangent cone, it is not a polyhedral cone. Linearizing the factors of the complementarity term individually bestows $\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)$ with a quadratic term. In particular, $\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)$ is not convex in general.

The idea of $\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)$ is to "linearize under the complementarity." This is a theme that will recur again and again in the remainder of this monograph.

Obviously, we have

$$\mathcal{T}^{lin}_{\text{MPEC}}(z^*) \subseteq \mathcal{T}^{lin}(z^*), \tag{3.14}$$

which can be seen by comparing $\mathcal{T}^{lin}_{\text{MPEC}}(z^*)$ with the representation (2.28) of $\mathcal{T}^{lin}(z^*)$. However, the relationship between $\mathcal{T}(z^*)$ and $\mathcal{T}^{lin}_{\text{MPEC}}(z^*)$ is not that apparent. In order to answer this question, we first express $\mathcal{T}^{lin}_{\text{MPEC}}(z^*)$ in terms of the programs $\text{NLP}_*(\beta_1, \beta_2)$ in the following counterpart to Lemma 2.16

Lemma 3.6 *Let z^* be a feasible vector of the MPEC* (1.1). *Then it holds that*

$$\mathcal{T}^{lin}_{\text{MPEC}}(z^*) = \bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}^{lin}_{\text{NLP}_*(\beta_1,\beta_2)}(z^*)\,. \tag{3.15}$$

Proof. This is easily verified by comparing (2.31) to (3.13). Also, it has previously been stated in [55]. □

Using this lemma, we can now answer the question of the relationship between $\mathcal{T}(z^*)$ and $\mathcal{T}^{lin}_{\text{MPEC}}(z^*)$. We do this in the following lemma.

Lemma 3.7 *Let z^* be a feasible point of the MPEC* (1.1). *Then the inclusion*

$$\mathcal{T}(z^*) \subseteq \mathcal{T}^{lin}_{\text{MPEC}}(z^*) \tag{3.16}$$

holds.

Proof. It is known from nonlinear programming that

$$\mathcal{T}_{\text{NLP}_*(\beta_1,\beta_2)}(z^*) \subseteq \mathcal{T}^{lin}_{\text{NLP}_*(\beta_1,\beta_2)}(z^*)$$

(see (2.12)). It follows immediately that

$$\bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}_{\text{NLP}_*(\beta_1,\beta_2)}(z^*) \subseteq \bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}^{lin}_{\text{NLP}_*(\beta_1,\beta_2)}(z^*). \tag{3.17}$$

Together with Lemmas 2.16 and 3.6, this yields

$$\mathcal{T}(z^*) = \bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}_{\text{NLP}_*(\beta_1,\beta_2)}(z^*) \subseteq \bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}^{lin}_{\text{NLP}_*(\beta_1,\beta_2)}(z^*) = \mathcal{T}^{lin}_{\text{MPEC}}(z^*),$$

which proves the result. □

In view of Lemma 3.7, the inclusion (3.14) is supplemented to yield the following:

$$\mathcal{T}(z^*) \subseteq \mathcal{T}^{lin}_{\text{MPEC}}(z^*) \subseteq \mathcal{T}^{lin}(z^*). \tag{3.18}$$

Having collected useful information about the MPEC-linearized tangent cone $\mathcal{T}^{lin}_{\text{MPEC}}(z^*)$, we now use it to define MPEC variants of the classical Abadie and Guignard constraint qualification in a natural fashion.

Definition 3.8 *Let z^* be feasible for the MPEC* (1.1). *We say that the* MPEC-Abadie constraint qualification*, or* MPEC-ACQ*, holds at z^* if*

$$\mathcal{T}(z^*) = \mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*), \tag{3.19}$$

and that the MPEC-Guignard constraint qualification*, or* MPEC-GCQ*, holds at z^* if*

$$\mathcal{T}(z^*)^* = \mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)^*. \tag{3.20}$$

Obviously, MPEC-ACQ implies MPEC-GCQ, as can be seen by dualizing (3.19). The converse is not true, as the following example demonstrates.

Example 3.9 Consider the following MPEC, the feasible set of which is inspired by [56]:

$$\begin{array}{ll} \min & f(z) := z_1^2 + z_2^2 \\ \text{s.t.} & G(z) := z_1^2 \geq 0, \\ & H(z) := z_2^2 \geq 0, \\ & G(z)^T H(z) = z_1^2 z_2^2 = 0. \end{array}$$

The origin $z^* = (0, 0)$ is the unique minimizer. It is easily verified that the tangent and MPEC-linearized tangent cones at the origin reduce to

$$\mathcal{T}(0) = \{d \in \mathbb{R}^2 \mid d_1 d_2 = 0\}$$

and

$$\mathcal{T}^{lin}_{\mathrm{MPEC}}(0) = \mathbb{R}^2,$$

respectively. Clearly, MPEC-ACQ does not hold. However, it is easily verified that

$$\mathcal{T}(0)^* = \mathcal{T}^{lin}_{\mathrm{MPEC}}(0)^* = \{0\}.$$

Hence, MPEC-GCQ does hold and we have demonstrated that MPEC-GCQ is indeed weaker than MPEC-ACQ.

As in the case of the standard Guignard CQ, we can state MPEC-GCQ in primal form, which we do in the following Lemma.

Proposition 3.10 *Let z^* be a feasible point of the MPEC* (1.1). *Then MPEC-GCQ holds in z^* if and only if*

$$\mathrm{cl}(\mathrm{conv}(\mathcal{T}(z^*))) = \mathrm{conv}(\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)) \tag{3.21}$$

holds. We call (3.21) *the* primal formulation *of MPEC-GCQ.*

Proof. To show that (3.20) implies (3.21), consider the following string of equalities, the Roman numerals indicating which point of Lemma 2.9 is used:

$$\mathrm{cl}(\mathrm{conv}(\mathcal{T}(z^*))) \overset{(\mathrm{v})}{=} \mathcal{T}(z^*)^{**} \overset{(*)}{=} \mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)^{**} \overset{(\mathrm{v})}{=} \mathrm{cl}(\mathrm{conv}(\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*))).$$

Note that we obtained the equality marked with $(*)$ by dualizing (3.20). It remains to be shown that $\text{conv}(\mathcal{T}^{lin}_{\text{MPEC}}(z^*))$ is closed. But $\mathcal{T}^{lin}_{\text{MPEC}}(z^*)$ has the representation (3.15) from Lemma 3.6. We can therefore simply copy the proof of Lemma 2.20 following equation (2.37) word by word, yielding that $\text{conv}(\mathcal{T}^{lin}_{\text{MPEC}}(z^*))$ is in fact closed.

For the converse implication, consider the following string of equalities, where Roman numerals again indicate which point of Lemma 2.9 is used. Additionally, Lemma 2.10 is applied to yield the first equality.

$$\begin{aligned}
\mathcal{T}(z^*)^* &= \text{cl}(\text{conv}(\mathcal{T}(z^*)))^* \\
&\overset{(3.21)}{=} \text{conv}(\mathcal{T}^{lin}_{\text{MPEC}}(z^*))^* \\
&\overset{(\circ)}{=} \text{cl}(\text{conv}(\mathcal{T}^{lin}_{\text{MPEC}}(z^*)))^* \\
&\overset{(\text{v})}{=} \mathcal{T}^{lin}_{\text{MPEC}}(z^*)^{***} \\
&\overset{(\text{i}),(\text{iv})}{=} \text{cl}(\mathcal{T}^{lin}_{\text{MPEC}}(z^*)^*) \\
&\overset{(\text{i})}{=} \mathcal{T}^{lin}_{\text{MPEC}}(z^*)^*.
\end{aligned}$$

Note that $(\circ)$ holds because $\text{conv}(\mathcal{T}^{lin}_{\text{MPEC}}(z^*))$ is closed by the above arguments. This completes the proof. □

Having introduced the primal formulation of MPEC-GCQ, we also call (3.20) the *dual formulation* of MPEC-GCQ, to distinguish the two.

For standard Guignard CQ, we proved Lemma 2.20, which stated that $\text{conv}(\mathcal{T}(z^*))$ was already closed if Abadie CQ held for all $\text{NLP}_*(\beta_1, \beta_2)$. This, of course, may also be done in the case of MPEC-GCQ. However, here it makes little sense, since, as we will show, Abadie CQ for all $\text{NLP}_*(\beta_1, \beta_2)$ implies MPEC-ACQ (see Corollary 3.11), and hence MPEC-GCQ.

3.2.1 Sufficient Conditions for MPEC-ACQ

We dedicate this section to bridging the gap between MPEC-ACQ and MPEC-GCQ, and those constraint qualifications defined utilizing $\text{TNLP}(z^*)$ in Definition 3.1. In particular, this involves finding sufficient conditions for MPEC-ACQ.

The following corollary to Lemmas 2.16 and 3.6 gives a first insight into this question.

Corollary 3.11 *If, for every partition $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$, the Abadie constraint qualification holds for* $\text{NLP}_*(\beta_1, \beta_2)$*, i.e.*

$$\mathcal{T}_{\text{NLP}_*(\beta_1,\beta_2)}(z^*) = \mathcal{T}^{lin}_{\text{NLP}_*(\beta_1,\beta_2)}(z^*) \qquad \forall (\beta_1, \beta_2) \in \mathcal{P}(\beta),$$

then

$$\mathcal{T}(z^*) = \mathcal{T}^{lin}_{\text{MPEC}}(z^*),$$

i.e. MPEC-ACQ holds.

Before we can prove our next result, which will clarify the relationship between MPEC-MFCQ and MPEC-ACQ, we need the following lemma.

Lemma 3.12 *If a feasible point z^* of the MPEC* (1.1) *satisfies MPEC-MFCQ, then classic MFCQ is satisfied in z^* by the corresponding nonlinear program* $\mathrm{NLP}_*(\beta_1, \beta_2)$ (2.29) *for any partition $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$.*

Proof. The gradient vectors

$$\begin{aligned} &\nabla h_i(z^*), && \forall i = 1, \ldots, p, \\ &\nabla G_i(z^*), && \forall i \in \alpha \cup \beta_1, \\ &\nabla H_i(z^*), && \forall i \in \gamma \cup \beta_2 \end{aligned}$$

are linearly independent since they are a subset of the linearly independent gradient vectors (3.3) in MPEC-MFCQ.

A simple application of Lemma 2.3 yields the existence of a $d \in \mathbb{R}^n$ such that

$$\begin{aligned} &\nabla g_i(z^*)^T d < 0, && \forall i \in \mathcal{I}_g, \\ &\nabla h_i(z^*)^T d = 0, && \forall i = 1, \ldots, p, \\ &\nabla G_i(z^*)^T d = 0, && \forall i \in \alpha \cup \beta_1, \\ &\nabla H_i(z^*)^T d = 0, && \forall i \in \gamma \cup \beta_2, \\ &\nabla G_i(z^*)^T d > 0, && \forall i \in \beta_2, \\ &\nabla H_i(z^*)^T d > 0, && \forall i \in \beta_1. \end{aligned}$$

This completes the proof. $\square$

We are now able to prove the following theorem which shows that, analogous to standard nonlinear programming, MPEC-MFCQ implies MPEC-ACQ.

Theorem 3.13 *If a feasible point z^* of the MPEC* (1.1) *satisfies MPEC-MFCQ, it also satisfies MPEC-ACQ.*

Proof. Since, by Lemma 3.12, MPEC-MFCQ implies MFCQ for every $\mathrm{NLP}_*(\beta_1, \beta_2)$ with $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$, which in turn implies that Abadie CQ holds for every such $\mathrm{NLP}_*(\beta_1, \beta_2)$ (see Proposition 2.8), we have, by Corollary 3.11, that MPEC-ACQ holds. $\square$

The following example, taken from [55], demonstrates that the reverse of Theorem 3.13 does not hold.

Example 3.14 Consider the following MPEC:

$$\begin{array}{rl} \min & f(z) := z_1 - z_2 \\ \text{s.t.} & g(z) := z_2 \leq 0, \\ & G(z) := z_1 \geq 0, \\ & H(z) := z_1 + z_2 \geq 0, \\ & G(z)^T H(z) = z_1(z_1 + z_2) = 0. \end{array}$$

The origin $z^* = (0, 0)$ is the unique minimizer. It is easily verified that

$$\mathcal{T}(0) = \mathcal{T}^{lin}_{\mathrm{MPEC}}(0) = \{(d_1, d_2) \mid d_2 \le 0, d_1 + d_2 = 0\},$$

showing that MPEC-ACQ holds in z^*. However, there fails to exist a $d \in \mathbb{R}^2$ such that

$$\nabla g(0)^T d = d_2 < 0, \quad \nabla G(0)^T d = d_1 = 0, \quad \nabla H(0)^T d = d_1 + d_2 = 0,$$

showing that MPEC-MFCQ does not hold in z^*.

3.2.2 Specific MPECs

This section may be seen as a continuation of the last, in the sense that we will introduce programs, which, under certain assumptions, imply MPEC-ACQ.

Mathematical Programs with Affine Equilibrium Constraints

The first program that we will introduce is the *mathematical program with affine equilibrium constraints*, or *MPAEC*. This is somewhat of a misnomer since we also require the non-complementarity constraints to be affine. Specifically, an MPAEC has the form

$$\begin{array}{ll} \min & f(z) \\ \text{s.t.} & Az + a \le 0, \quad Bz + b = 0, \\ & Cz + c \ge 0, \quad Dz + d \ge 0, \quad (Cz + c)^T(Dz + d) = 0, \end{array} \tag{3.22}$$

with $A \in \mathbb{R}^{m \times n}$, $a \in \mathbb{R}^m$, $B \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$, $C, D \in \mathbb{R}^{l \times n}$, and $c, d \in \mathbb{R}^l$.

The MPAEC (3.22) clearly reiterates the idea of "linearizing under the complementarity" we discussed in the context of $\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)$, see page 32.

Analogous to standard nonlinear programming, an MPAEC needs no constraint qualification (other than the linearity of its data) to guarantee that MPEC-ACQ holds. We show this in the following theorem.

Theorem 3.15 *Let z^* be a feasible point of the MPAEC* (3.22). *Then MPEC-ACQ holds in z^*.*

Proof. By assumption, the functions making up the constraints of each $\mathrm{NLP}_*(\beta_1, \beta_2)$ with $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$ are affine linear. Hence Abadie CQ holds for all $\mathrm{NLP}_*(\beta_1, \beta_2)$ (see, e.g., [4, Lemma 5.1.4] or [27, Satz 2.42]). By Corollary 3.11, MPEC-ACQ holds. $\square$

MPEC-Convex Constraints

As we discussed in Section 2.2, the MPEC (1.1) we consider never has convex constraints. We can, however, transfer the ideas from standard convex programming to MPECs. To this end, we again linearize the factors of the complementarity term individually, obtaining the following program:

$$\begin{array}{ll} \min & f(z) \\ \text{s.t.} & g(z) \leq 0, \quad h(z) = 0, \\ & G(z) \geq 0, \quad H(z) \geq 0, \quad G(z)^T H(z) = 0, \end{array} \tag{3.23a}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable, the component functions $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are convex, and the remaining functions are affine linear:

$$\begin{aligned} h(z) &:= Vz + \eta, \\ G(z) &:= Wz + \chi, \\ H(z) &:= Xz + \xi, \end{aligned} \tag{3.23b}$$

with $V \in \mathbb{R}^{p \times n}$, $\eta \in \mathbb{R}^p$, $W \in \mathbb{R}^{l \times n}$, $\chi \in \mathbb{R}^l$, $X \in \mathbb{R}^{l \times n}$, and $\xi \in \mathbb{R}^l$. We denote the rows of V, W, and X by v_i^T, w_i^T, and x_i^T, respectively. We apologize for the discrepancy in notation compared to the MPAEC (3.22). However, the proof of Theorem 3.17 is simplified greatly thereby.

Borrowing from standard nonlinear programming, we call a program of this type *MPEC-convexly constrained*. Note, however, that the feasible reason of (3.23) is *not* convex. Also note that, as in the case of convexly constrained programs, the objective function f is not in general convex.

We shall now define appropriate modifications of the Slater and weak Slater constraint qualifications, cf. Definition 2.13.

Definition 3.16 *The program* (3.23) *is said to satisfy the* weak MPEC-Slater constraint qualification*, or* MPEC-WSCQ*, in a feasible vector z^* if there exists a vector $\hat{z}$ such that*

$$\begin{aligned} g_i(\hat{z}) &< 0, & \forall i &\in \mathcal{I}_g, \\ h_i(\hat{z}) &= 0, & \forall i &= 1, \ldots, p, \\ G_i(\hat{z}) &= 0, & \forall i &\in \alpha \cup \beta, \\ H_i(\hat{z}) &= 0, & \forall i &\in \gamma \cup \beta. \end{aligned} \tag{3.24}$$

It is said to satisfy the MPEC-Slater constraint qualification*, or* MPEC-SCQ*, if there exists a vector $\hat{z}$ such that*

$$\begin{aligned} g_i(\hat{z}) &< 0, & \forall i &= 1, \ldots, m, \\ h_i(\hat{z}) &= 0, & \forall i &= 1, \ldots, p, \\ G_i(\hat{z}) &= 0, & \forall i &= 1, \ldots, l, \\ H_i(\hat{z}) &= 0, & \forall i &= 1, \ldots, l. \end{aligned} \tag{3.25}$$

Note that, just as in standard nonlinear programming, the weak MPEC-Slater CQ is characterized by the fact that it depends on the vector z^* (through the sets $\mathcal{I}_g$, α, β, and γ). The appeal of the MPEC-Slater CQ is that it is stated independently of z^*. However, MPEC-SCQ is potentially much more restrictive than MPEC-WSCQ, since we require $G_\gamma(\hat{z}) = 0$ and $H_\alpha(\hat{z}) = 0$. Note that MPEC-SCQ implies MPEC-WSCQ, analogous to standard nonlinear programming.

We now show the weak MPEC-Slater CQ to imply MPEC-ACQ. Unfortunately it is not possible to reduce the proof of the following theorem to results from nonlinear programming and Corollary 3.11 as was the case for Theorems 3.13 and 3.15. This is because, although the $\mathrm{NLP}_*(\beta_1, \beta_2)$ are convexly constrained programs, neither MPEC-WSCQ nor MPEC-SCQ implies that the Slater constraint qualification holds for any $\mathrm{NLP}_*(\beta_1, \beta_2)$, as can be seen by considering that WSCQ of $\mathrm{NLP}_*(\beta_1, \beta_2)$ is satisfied if there exists a $\hat{z} \in \mathbb{R}^n$ such that

$$\begin{aligned} &g_{\mathcal{I}_g}(\hat{z}) < 0, && h(\hat{z}) = 0, \\ &G_{\alpha\cup\beta_1}(\hat{z}) = 0, && G_{\beta_2}(\hat{z}) > 0, \\ &H_{\gamma\cup\beta_2}(\hat{z}) = 0, && H_{\beta_1}(\hat{z}) > 0. \end{aligned} \tag{3.26}$$

Clearly, neither MPEC-SCQ nor MPEC-WSCQ implies that there exists a $\hat{z}$ such that $G_{\beta_2}(\hat{z}) > 0$ and $H_{\beta_1}(\hat{z}) > 0$ (except in the nondegenerate case when $\beta = \varnothing$, which we have excluded from the start). We therefore have to fall back on a more elementary proof.

Theorem 3.17 *Let z^* be a feasible point of the program* (3.23). *If it satisfies MPEC-WSCQ, it also satisfies MPEC-ACQ.*

Proof. By virtue of Lemma 3.7 we know that $\mathcal{T}(z^*) \subseteq \mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)$. Therefore, all that remains to be shown is

$$\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*) \subseteq \mathcal{T}(z^*). \tag{3.27}$$

We take the same path here as was taken in [27, Satz 2.45] for standard nonlinear programming. To this end, we define the following cone:

$$\begin{aligned} \mathcal{T}^{strict}_{\mathrm{MPEC}}(z^*) := \{ d \in \mathbb{R}^n \mid\ & \nabla g_i(z^*)^T d < 0, && \forall i \in \mathcal{I}_g, \\ & \nabla h_i(z^*)^T d = 0, && \forall i = 1, \ldots, p, \\ & \nabla G_i(z^*)^T d = 0, && \forall i \in \alpha, \\ & \nabla H_i(z^*)^T d = 0, && \forall i \in \gamma, \\ & \nabla G_i(z^*)^T d \geq 0, && \forall i \in \beta, \\ & \nabla H_i(z^*)^T d \geq 0, && \forall i \in \beta, \\ & (\nabla G_i(z^*)^T d) \cdot (\nabla H_i(z^*)^T d) = 0, && \forall i \in \beta \}. \end{aligned} \tag{3.28}$$

Note that the difference between $\mathcal{T}^{strict}_{\mathrm{MPEC}}(z^*)$ and $\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)$ (see (3.13)) lies in the strict inequality for g.

Now, to prove (3.27), we show the following two inclusions:

$$\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*) \subseteq \mathrm{cl}(\mathcal{T}^{strict}_{\mathrm{MPEC}}(z^*)) \subseteq \mathcal{T}(z^*). \tag{3.29}$$

To prove the first inclusion of (3.29), we take a vector $\hat{z}$ satisfying the Slater conditions (3.24) and set

$$\hat{d} := \hat{z} - z^*.$$

Now the following holds for all $i \in \mathcal{I}_g$ because g is convex by assumption:

$$\nabla g_i(z^*)^T \hat{d} \le \underbrace{g_i(\hat{z})}_{<0} - \underbrace{g_i(z^*)}_{=0} < 0, \qquad \forall i \in \mathcal{I}_g. \tag{3.30}$$

Similarly, the following hold for the linear functions h, G, and H:

$$\begin{aligned}
\nabla h_i(z^*)^T \hat{d} &= v_i^T \hat{z} - v_i^T z^* = \underbrace{h_i(\hat{z})}_{=0} - \underbrace{h_i(z^*)}_{=0} = 0, && \forall i = 1, \dots, p, \\
\nabla G_i(z^*)^T \hat{d} &= w_i^T \hat{z} - w_i^T z^* = \underbrace{G_i(\hat{z})}_{=0} - \underbrace{G_i(z^*)}_{=0} = 0, && \forall i \in \alpha \cup \beta, \\
\nabla H_i(z^*)^T \hat{d} &= x_i^T \hat{z} - x_i^T z^* = \underbrace{H_i(\hat{z})}_{=0} - \underbrace{H_i(z^*)}_{=0} = 0, && \forall i \in \gamma \cup \beta.
\end{aligned} \tag{3.31}$$

We now use the vector $\hat{d}$, which we know has the properties (3.30) and (3.31), to define the following function:

$$d(\delta) := d + \delta \hat{d},$$

where $d \in \mathcal{T}_{\text{MPEC}}^{lin}(z^*)$ is chosen arbitrarily.

We want to show that $d(\delta) \in \mathcal{T}_{\text{MPEC}}^{strict}(z^*)$ for all $\delta > 0$. To this end, let $\delta > 0$ be fixed for the time being. Then the following holds:

$$\begin{aligned}
\nabla g_i(z^*)^T d(\delta) &= \underbrace{\nabla g_i(z^*)^T d}_{\le 0} + \delta \underbrace{\nabla g_i(z^*)^T \hat{d}}_{<0} < 0, \qquad \forall i \in \mathcal{I}_g, \\
\nabla h_i(z^*)^T d(\delta) &= \nabla h_i(z^*)^T d + \delta \nabla h_i(z^*)^T \hat{d} = 0, \qquad \forall i = 1, \dots, p, \\
\nabla G_i(z^*)^T d(\delta) &= \begin{cases} \underbrace{\nabla G_i(z^*)^T d}_{=0} + \delta \underbrace{\nabla G_i(z^*)^T \hat{d}}_{=0} = 0, & \forall i \in \alpha, \\ \underbrace{\nabla G_i(z^*)^T d}_{\ge 0} + \delta \underbrace{\nabla G_i(z^*)^T \hat{d}}_{=0} \ge 0, & \forall i \in \beta, \end{cases} \\
\nabla H_i(z^*)^T d(\delta) &= \begin{cases} \underbrace{\nabla H_i(z^*)^T d}_{=0} + \delta \underbrace{\nabla H_i(z^*)^T \hat{d}}_{=0} = 0, & \forall i \in \gamma, \\ \underbrace{\nabla H_i(z^*)^T d}_{\ge 0} + \delta \underbrace{\nabla H_i(z^*)^T \hat{d}}_{=0} \ge 0, & \forall i \in \beta. \end{cases}
\end{aligned}$$

Furthermore, for all $i \in \beta$ it holds that

$$\begin{aligned}(\nabla G_i(z^*)^T d(\delta)) \cdot (\nabla H_i(z^*)^T d(\delta)) = & \underbrace{(\nabla G_i(z^*)^T d) \cdot (\nabla H_i(z^*)^T d)}_{=0} \\ & + \delta(\nabla G_i(z^*)^T d) \cdot \underbrace{(\nabla H_i(z^*)^T \hat{d})}_{=0} \\ & + \delta\, \underbrace{(\nabla G_i(z^*)^T \hat{d})}_{=0} \cdot (\nabla H_i(z^*)^T d) \\ & + \delta^2\, \underbrace{(\nabla G_i(z^*)^T \hat{d})}_{=0} \cdot \underbrace{(\nabla H_i(z^*)^T \hat{d})}_{=0} \\ = {} & 0.\end{aligned}$$

Comparing the properties of $d(\delta)$ to $\mathcal{T}_{\text{MPEC}}^{strict}(z^*)$ (see (3.28)), we see that $d(\delta) \in \mathcal{T}_{\text{MPEC}}^{strict}(z^*)$ for all $\delta > 0$. Since $\text{cl}(\mathcal{T}_{\text{MPEC}}^{strict}(z^*))$ is closed by definition, it holds that $d = \lim_{\delta \searrow 0} d(\delta) \in \text{cl}(\mathcal{T}_{\text{MPEC}}^{strict}(z^*))$. Hence, the inclusion

$$\mathcal{T}_{\text{MPEC}}^{lin}(z^*) \subseteq \text{cl}(\mathcal{T}_{\text{MPEC}}^{strict}(z^*)) \tag{3.32}$$

holds.

To prove the second inclusion of (3.29), let $d \in \mathcal{T}_{\text{MPEC}}^{strict}(z^*)$ and $\{t_k\}$ be a sequence with $t_k \searrow 0$. Setting

$$z^k := z^* + t_k d,$$

we have

$$z^k \to z^* \quad \text{and} \quad \frac{z^k - z^*}{t_k} \to d.$$

If we can prove that $\{z^k\} \subseteq \mathcal{Z}$, it would follow that $d \in \mathcal{T}(z^*)$ (see (2.10)) and we would have shown $\mathcal{T}_{\text{MPEC}}^{strict}(z^*) \subseteq \mathcal{T}(z^*)$, and, since $\mathcal{T}(z^*)$ is closed, also that $\text{cl}(\mathcal{T}_{\text{MPEC}}^{strict}(z^*)) \subseteq \mathcal{T}(z^*)$. By the mean value theorem, it holds that there exists a vector ζ on the connecting line between z^k and z^* such that the following holds:

$$\begin{aligned} g_i(z^k) &= g_i(z^*) + \nabla g_i(\zeta)^T (z^k - z^*) \\ &= g_i(z^*) + t_k \nabla g_i(\zeta)^T d \\ &= \left\{ \begin{array}{ll} \underbrace{g_i(z^*)}_{=0} + t_k \underbrace{\nabla g_i(\zeta)^T d}_{<0,\ \forall k > k_0}, & \forall i \in \mathcal{I}_g \\ \underbrace{g_i(z^*)}_{<0} + t_k \nabla g_i(\zeta)^T d, & \forall i \notin \mathcal{I}_g \end{array} \right\} \leq 0, \qquad \forall k > k_1, \end{aligned}$$

where $k_0 \geq 0$ and $k_1 \geq k_0$ are sufficiently large integers. Note that the convexity of g does not enter here.

The linear functions are handled similarly:

$$h_i(z^k) = v_i^T z^k + \eta_i = \underbrace{v_i^T z^* + \eta_i}_{=0} + t_k \underbrace{v_i^T d}_{=0} = 0, \qquad \forall i = 1, \dots, p,$$

$$G_i(z^k) = w_i^T z^k + \chi_i = \left\{ \begin{array}{ll} \underbrace{w_i^T z^* + \chi_i}_{=0} + t_k \underbrace{w_i^T d}_{=0}, & \forall i \in \alpha \\ \underbrace{w_i^T z^* + \chi_i}_{=0} + t_k \underbrace{w_i^T d}_{\geq 0}, & \forall i \in \beta \\ \underbrace{w_i^T z^* + \chi_i}_{>0} + t_k w_i^T d, & \forall i \in \gamma \end{array} \right\} \geq 0, \qquad \forall k > k_2,$$

$$H_i(z^k) = x_i^T z^k + \xi_i = \left\{ \begin{array}{ll} \underbrace{x_i^T z^* + \xi_i}_{>0} + t_k x_i^T d, & \forall i \in \alpha \\ \underbrace{x_i^T z^* + \xi_i}_{=0} + t_k \underbrace{x_i^T d}_{\geq 0}, & \forall i \in \beta \\ \underbrace{x_i^T z^* + \xi_i}_{=0} + t_k \underbrace{x_i^T d}_{=0}, & \forall i \in \gamma \end{array} \right\} \geq 0, \qquad \forall k > k_3,$$

with $k_2 \geq 0$ and $k_3 \geq 0$ sufficiently large integers.

Finally, taking into consideration the definition of $\mathcal{T}_{\text{MPEC}}^{strict}(z^*)$ (see (3.28)), the following holds for the product $G_i(z^k) \cdot H_i(z^k)$:

$$G_i(z^k) \cdot H_i(z^k) = \left\{ \begin{array}{ll} (\underbrace{w_i^T z^* + \chi_i}_{=0} + t_k \underbrace{w_i^T d}_{=0})(x_i^T z^* + \xi_i + t_k x_i^T d), & \forall i \in \alpha \\ (\underbrace{w_i^T z^* + \chi_i}_{=0} + t_k w_i^T d)(\underbrace{x_i^T z^* + \xi_i}_{=0} + t_k x_i^T d) = t_k^2 \underbrace{(w_i^T d)(x_i^T d)}_{=0}, & \forall i \in \beta \\ (w_i^T z^* + \chi_i + t_k w_i^T d)(\underbrace{x_i^T z^* + \xi_i}_{=0} + t_k \underbrace{x_i^T d}_{=0}), & \forall i \in \gamma \end{array} \right\} = 0.$$

The above results demonstrate that $z^k \in \mathcal{Z}$ for all $k \geq \max\{k_0, k_1, k_2, k_3\}$ and hence $d \in \mathcal{T}(z^*)$. As mentioned in the discussion following Definition 2.5, $\mathcal{T}(z^*)$ is closed, yielding that $\text{cl}(\mathcal{T}_{\text{MPEC}}^{strict}(z^*)) \subseteq \mathcal{T}(z^*)$.

Together with (3.32), we therefore have $\mathcal{T}_{\text{MPEC}}^{lin}(z^*) \subseteq \mathcal{T}(z^*)$ and hence MPEC-ACQ is satisfied, concluding this proof. □

Note that the technique of the proof of Theorem 3.17 was adapted from [27, Satz 2.45], and may also be used to prove Proposition 2.14.

The following corollary follows immediately from Theorem 3.17 and Definition 3.16.

Corollary 3.18 *Let z^* be a feasible point of the program* (3.23). *If it satisfies MPEC-SCQ, it also satisfies MPEC-ACQ.*

In addition to the sufficient conditions for MPEC-ACQ discussed in the previous two sections, Ye [78] has introduced MPEC variants of several CQs not discussed here, and has shown them to imply MPEC-ACQ. The interested reader is referred to [78] for more details.

3.2.3 Comparison between the basic CQ [42] and MPEC-ACQ

The question was raised by Danny Ralph [57] whether MPEC-ACQ may be identical to the *basic CQ* used in [42]. In this section we try to shed some light on this. It may be noted in advance that if the constraints consist only of complementarity constraints that take on a certain form, then the basic CQ and MPEC-ACQ do indeed coincide. To see this, however, some effort is required. The remainder of this section is dedicated to that question. To facilitate a direct comparison with [42], we adopt their notation in the following.

Let us therefore consider an MPEC in the notation used in [42]:

$$\begin{array}{rl} \min & f(x,y) \\ \text{s.t.} & (x,y) \in Z \subseteq \mathbb{R}^{n+m}, \\ & y \in \mathcal{S} := \mathrm{SOL}(F(x,\cdot), C(x)) \end{array} \tag{3.33}$$

with

$$C(x) := \{y \in \mathbb{R}^m \mid g(x,y) \le 0\}. \tag{3.34}$$

Here $y \in \mathrm{SOL}(F(x,\cdot), C(x))$ denotes a solution of the variational inequality

$$\begin{array}{l} y \in C(x), \\ (v-y)^T F(x,y) \ge 0, \qquad \forall v \in C(x). \end{array} \tag{3.35}$$

Since we are interested in mathematical programs with nonlinear complementarity problems as constraints rather than variational inequalities, we consider the case where $g(x,y) := -y$. Then the program (3.33) reduces to the following program:

$$\begin{array}{rl} \min & f(x,y) \\ \text{s.t.} & (x,y) \in Z, \\ & y \ge 0, \\ & (v-y)^T F(x,y) \ge 0, \qquad \forall v \ge 0, \end{array}$$

which obviously is equivalent to the following program:

$$\begin{array}{rl} \min & f(x,y) \\ \text{s.t.} & (x,y) \in Z, \\ & y \ge 0, \quad F(x,y) \ge 0, \quad y^T F(x,y) = 0. \end{array} \tag{3.36}$$

Now, the *basic CQ* from [42] is said to hold in (x^*, y^*) if there exists a nonempty set M' of Lagrange multipliers of the variational inequality (3.35) such that

$$\mathcal{T}((x^*,y^*);\mathcal{F}) = \mathcal{T}((x^*,y^*);Z) \cap \left(\bigcup_{\lambda \in M'} \mathrm{Gr}(\mathcal{LS}_{(x^*,y^*,\lambda)}) \right), \tag{3.37}$$

where $\mathcal{T}(d;\mathcal{D}) = \mathcal{T}(d,\mathcal{D})$ denotes the tangent cone to the set $\mathcal{D}$ in the point d (see Definition 2.5), $\mathcal{F}$ denotes the feasible region of (3.36), $\mathrm{Gr}(\cdot)$ denotes the graph of a multifunction, and

$$\mathcal{LS}_{(x^*,y^*,\lambda)}(dx) = \mathrm{SOL}(\nabla_x L(x^*,y^*,\lambda)dx, \nabla_y L(x^*,y^*,\lambda), \mathcal{K}(x^*,y^*,\lambda;dx)) \tag{3.38}$$

is the solution set of the affine variational inequality

$$\begin{aligned} &dy \in \mathcal{K}(x^*,y^*,\lambda;dx), \\ &(v-dy)^T(\nabla_x L(x^*,y^*,\lambda)dx + \nabla_y L(x^*,y^*,\lambda)dy) \geq 0, \quad \forall v \in \mathcal{K}(x^*,y^*,\lambda;dx), \end{aligned} \tag{3.39}$$

where

$$L(x,y,\lambda) := F(x,y) + \sum_{i=1}^{l} \lambda_i \nabla_y g_i(x,y) = F(x,y) - \lambda \tag{3.40}$$

is the Lagrangian,

$$\mathcal{K}(x^*,y^*,\lambda;dx) := \{dy \in \mathbb{R}^m \mid (dx,dy) \in \mathcal{K}(x^*,y^*,\lambda)\} \tag{3.41}$$

is the *directional critical set*, and

$$\begin{aligned} \mathcal{K}(x^*,y^*,\lambda) := \big\{&(dx,dy) \in \mathbb{R}^{n+m} \mid \\ &\nabla_x g_i(x^*,y^*)^T dx + \nabla_y g_i(x^*,y^*)^T dy \leq 0, \forall i \in \{i \mid g_i(x^*,y^*) = 0 \wedge \lambda_i = 0\}, \\ &\nabla_x g_i(x^*,y^*)^T dx + \nabla_y g_i(x^*,y^*)^T dy = 0, \forall i \in \{i \mid g_i(x^*,y^*) = 0 \wedge \lambda_i > 0\}\big\} \\ = \big\{&(dx,dy) \in \mathbb{R}^{n+m} \mid \\ &e_i^T dy \geq 0, \quad \forall i \in \{i \mid y_i^* = 0 \wedge \lambda_i = 0\}, \\ &e_i^T dy = 0, \quad \forall i \in \{i \mid \lambda_i > 0\}\big\} \end{aligned} \tag{3.42}$$

is the *lifted critical cone* ($e_i \in \mathbb{R}^m$ is the i-th unit vector).

If, in (3.37), M' is assumed to be the whole set of Lagrange multipliers, the *full CQ* is said to hold.

The Lagrange multipliers λ of the variational inequality (3.35) must satisfy the following conditions:

$$\begin{aligned} &L(x,y,\lambda) = F(x,y) - \lambda = 0, \\ &\lambda \geq 0, \quad y \geq 0, \quad \lambda^T y = 0. \end{aligned} \tag{3.43}$$

Hence $\lambda^* := F(x^*,y^*)$ is the only Lagrange multiplier and $M' := \{\lambda^*\}$ reduces to a singleton, so that the basic and full CQ coincide. Note also that, because of the complementarity conditions in (3.43), $\lambda_i > 0$ implies $g_i(x^*,y^*) = -y_i^* = 0$ in (3.42).

Let us again consider the basic CQ (3.37). Note that it is somewhat different from our MPEC-ACQ since we also linearize the non-complementarity constraints. We void this difference in our subsequent discussion by setting $Z := \mathbb{R}^{n+m}$. Combining this with the fact that M' reduces to a singleton for our program (3.36) results in the basic (or full) CQ reducing to

$$\mathcal{T}((x^*, y^*); \mathcal{F}) = \mathrm{Gr}(\mathcal{LS}_{(x^*,y^*,\lambda^*)}). \tag{3.44}$$

Now, by virtue of [42, (1.3.14)], $dy \in \mathcal{LS}_{(x^*,y^*,\lambda^*)}(dx)$ holds if and only if there exists a multiplier μ such that

$$\begin{aligned} \nabla_x L(x^*, y^*, \lambda^*)dx + \nabla_y L(x^*, y^*, \lambda^*)dy - A^T\mu = 0,\\ Ady - b \geq 0,\\ \mu \geq 0,\\ \mu^T(Ady - b) = 0, \end{aligned} \tag{3.45}$$

with

$$A := \begin{pmatrix} e_i^T & i \in \{i \mid y_i^* = 0 \wedge \lambda_i^* = 0\} \\ -e_i^T & \\ e_i^T & i \in \{i \mid \lambda_i^* > 0\} \end{pmatrix}, \qquad b := 0.$$

Combining $z := (x, y)$ and $d = (dx, dy)$, and remembering that $\lambda^* = F(x^*, y^*)$, the set $\mathrm{Gr}(\mathcal{LS}_{(x^*,y^*,\lambda^*)})$ can be written as follows:

$$\begin{aligned} \mathrm{Gr}(\mathcal{LS}_{(z^*,\lambda^*)}) = \{d \in \mathbb{R}^{n+m} \mid \exists\mu \ : \ & \nabla F(z^*)^T d - A^T\mu = 0, && (3.46a)\\ & \mu \geq 0, \quad \mu^T A dy = 0, && (3.46b)\\ & e_i^T dy \geq 0, \quad i \in \beta && (3.46c)\\ & e_i^T dy = 0, \quad i \in \gamma\}, && (3.46d) \end{aligned}$$

where β and γ (and α, see below) are defined as in (1.2), for the program (3.36).

The conditions (3.46c) and (3.46d) reduce to $d_{n+i} \geq 0$ for $i \in \beta$ and $d_{n+i} = 0$ for $i \in \gamma$ respectively. From (3.46a) it follows that $\nabla F_i(z^*)^T d = 0$ for $i \in \alpha$, and $\nabla F_i(z^*)^T d = \mu_{j(i)} \geq 0$ for $i \in \beta$, where the index j depends on i.

Let us now consider (3.46b):

$$\begin{aligned} \mu^T A dy &= \mu^T \begin{pmatrix} \overbrace{\begin{matrix} d_{n+i} \end{matrix}}^{=0} & i \in \beta \\ \begin{matrix} -d_{n+i} \\ \underbrace{d_{n+i}}_{=0} \end{matrix} & i \in \gamma \end{pmatrix} \\ &= \sum_{i\in\beta} \mu_{j(i)} d_{n+i} \\ &= \sum_{i\in\beta} \underbrace{(\nabla F_i(z^*)^T d)}_{\geq 0} \underbrace{d_{n+i}}_{\geq 0} = 0. \end{aligned}$$

From this, it follows that

$$(\nabla F_i(z^*)^T d) \cdot d_{n+i} = 0, \qquad \forall i \in \beta.$$

Collecting everything we have gathered about $\mathrm{Gr}(\mathcal{LS}_{(z^*,\lambda^*)})$, we arrive at the following result:

$$\begin{aligned}\mathrm{Gr}(\mathcal{LS}_{(z^*,\lambda^*)}) = \{d \in \mathbb{R}^{n+m} \mid\ & \nabla F_i(z^*)^T d = 0, && \forall i \in \alpha,\\ & d_{n+i} = 0, && \forall i \in \gamma,\\ & \nabla F_i(z^*)^T d \geq 0, && \forall i \in \beta,\\ & d_{n+i} \geq 0, && \forall i \in \beta,\\ & (\nabla F_i(z^*)^T d) \cdot d_{n+i} = 0, && \forall i \in \beta\},\end{aligned}$$

which is equal to $\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)$ of the MPEC (3.36). Hence the basic CQ [42] and MPEC-ACQ coincide in this case. Note that they do not coincide in the presence of non-complementarity constraints. However, by replacing $\mathcal{T}((x^*, y^*); Z)$ with its linearization in (3.37), we can define the *extended basic CQ*,

$$\mathcal{T}((x^*, y^*); \mathcal{F}) = \mathcal{T}^{lin}((x^*, y^*); Z) \cap \left(\bigcup_{\lambda \in M'} \mathrm{Gr}(\mathcal{LS}_{(x^*,y^*,\lambda)}) \right),$$

which is equivalent to MPEC-ACQ in the context discussed here, i.e. where the MPEC takes the form (3.36).

3.3 Revisiting the Guignard CQ

To conclude this chapter on MPEC constraint qualifications, we wish to turn our attention toward the Guignard CQ once more. We have already discussed that both MPEC-LICQ and MPEC-SMFCQ imply standard GCQ. In this section, we will use the definition of MPEC-GCQ to learn more about Guignard CQ itself.

A direct comparison between MPEC-GCQ and GCQ reveals that the condition that is missing to close the gap between the two is the so-called intersection property, which we will formally introduce in the following definition.

Definition 3.19 *Let z^* be a feasible point of the MPEC* (1.1). *We say that the* intersection property *holds at z^* if*

$$\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)^* = \mathcal{T}^{lin}(z^*)^*. \tag{I}$$

The name "intersection property" stems from a more general setting, where (I) is expressed using the intersection of certain normal cones. We will not go into this here, but will nevertheless keep the name for consistency's sake. The interested reader is referred to [22] for details.

A trivial corollary to the definitions of GCQ, MPEC-GCQ, and the intersection property (see Definitions 2.7, 3.8, and 3.19, respectively) is the following.

Corollary 3.20 *Let z^* be a feasible point of the MPEC* (1.1). *Then GCQ holds if and only if MPEC-GCQ as well as the intersection property* (I) *hold.*

Though it provides an elegant result, the intersection property (I) is not very tangible. We therefore dedicate the remainder of the section to finding sufficient conditions for (I). The spadework for this has been done by Pang and Fukushima [55], albeit with a different purpose in mind. We will therefore extensively fall back on their results in the following discussion.

To this end, we must first introduce the concept of nonsingularity, as used in [55, 70].

Definition 3.21 *Given the linear system*

$$Ax \leq b, \qquad Cx = d, \tag{3.47}$$

an inequality $A_i x \leq b_i$ is said to be nonsingular *if there exists a feasible solution of the system* (3.47) *which satisfies this inequality strictly. Here A_i denotes the i-th row of the matrix A.*

We will now apply nonsingularity to the linearized tangent cone $\mathcal{T}^{lin}(z^*)$ (see (2.28)). To facilitate this, we introduce two new sets: Let β^G denote the subset of β consisting of all indices $i \in \beta$ such that the inequality $\nabla G_i(z^*)^T d \geq 0$ is nonsingular in the system defining $\mathcal{T}^{lin}(z^*)$. Similarly, we denote by β^H the nonsingular set pertaining to the inequalities $\nabla H_i(z^*)^T d \geq 0$. Note that β^G and β^H depend on z^*.

Using the sets β^G and β^H renders the following representation of $\mathcal{T}^{lin}(z^*)$ (cf. (2.28)):

$$\begin{aligned} \mathcal{T}^{lin}(z^*) = \{ d \in \mathbb{R}^n \mid \; & \nabla g_i(z^*)^T d \leq 0, && \forall i \in \mathcal{I}_g, \\ & \nabla h_i(z^*)^T d = 0, && \forall i = 1, \ldots, p, \\ & \nabla G_i(z^*)^T d = 0, && \forall i \in \alpha \cup \beta \backslash \beta^G, \\ & \nabla H_i(z^*)^T d = 0, && \forall i \in \gamma \cup \beta \backslash \beta^H, \\ & \nabla G_i(z^*)^T d \geq 0, && \forall i \in \beta^G, \\ & \nabla H_i(z^*)^T d \geq 0, && \forall i \in \beta^H \; \}. \end{aligned} \tag{3.48}$$

We will now also use the sets β^G and β^H to define the following assumption (A). Note that (A) is equivalent to [55, (A2)] by Lemma 1 of the same reference.

(A) Given the feasible point z^*, there exists a partition $(\beta_1^{GH}, \beta_2^{GH}) \in \mathcal{P}(\beta^G \cap \beta^H)$ such that

$$\sum_{i \in \mathcal{I}_g} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(z^*) - \sum_{i \in \alpha \cup \beta} \lambda_i^G \nabla G_i(z^*) - \sum_{i \in \gamma \cup \beta} \lambda_i^H \nabla H_i(z^*) = 0$$
$$\Longrightarrow \begin{cases} \lambda_i^G = 0, & \forall i \in \beta_1^{GH} \\ \lambda_i^H = 0, & \forall i \in \beta_2^{GH}. \end{cases}$$

We are now able to prove that assumption (A) implies the intersection property (I).

Lemma 3.22 *If a feasible point z^* of the MPEC* (1.1) *satisfies assumption (A), it also satisfies the intersection property* (I).

Proof. We know that

$$\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*) \subseteq \mathcal{T}^{lin}(z^*)$$

(see (3.14)), and hence

$$\mathcal{T}^{lin}(z^*)^* \subseteq \mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)^*$$

by Lemma 2.9 (ii). Therefore, all that remains to be shown is that

$$\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)^* \subseteq \mathcal{T}^{lin}(z^*)^*. \tag{3.49}$$

By Lemmas 3.6 and 2.9 (vi), we obtain the following representation of $\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)^*$:

$$\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)^* = \bigcap_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)^*. \tag{3.50}$$

Now, to prove (3.49), we take an arbitrary $v \in \mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)^*$. By virtue of (3.50), we have

$$v \in \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)^* \qquad \forall(\beta_1,\beta_2) \in \mathcal{P}(\beta). \tag{3.51}$$

Consider the specific partition of β given by

$$\hat{\beta}_1 := \beta^H \backslash \beta_2^{GH}, \quad \hat{\beta}_2 := \beta \backslash \hat{\beta}_1. \tag{3.52}$$

Here $(\beta_1^{GH}, \beta_2^{GH}) \in \mathcal{P}(\beta^G \cap \beta^H)$ is a partition of $\beta^G \cap \beta^H$ that satisfies assumption (A). Note that $\beta^G \backslash \beta_1^{GH} \subseteq \hat{\beta}_2$.

Since $v \in \mathcal{T}^{lin}_{\mathrm{NLP}_*(\hat{\beta}_1,\hat{\beta}_2)}(z^*)^*$ as well as $v \in \mathcal{T}^{lin}_{\mathrm{NLP}_*(\hat{\beta}_2,\hat{\beta}_1)}(z^*)^*$, we can apply Lemma 3.3 to both of these cones, yielding the existence of vectors $u = (u^g, u^h, u^G, u^H)$ and $w = (w^g, w^h, w^G, w^H)$ with

$$\begin{array}{llll} u_i^g \geq 0 & \forall i \in \mathcal{I}_g, & w_i^g \geq 0 & \forall i \in \mathcal{I}_g, \\ u_i^G \geq 0 & \forall i \in \hat{\beta}_2, & w_i^G \geq 0 & \forall i \in \hat{\beta}_1, \\ u_i^H \geq 0 & \forall i \in \hat{\beta}_1, & w_i^H \geq 0 & \forall i \in \hat{\beta}_2 \end{array}$$

such that

$$\begin{aligned} v &= -\sum_{i\in\mathcal{I}_g} u_i^g \nabla g_i(z^*) - \sum_{i=1}^{p} u_i^h \nabla h_i(z^*) + \sum_{i\in\alpha\cup\beta} u_i^G \nabla G_i(z^*) + \sum_{i\in\gamma\cup\beta} u_i^H \nabla H_i(z^*) \\ &= -\sum_{i\in\mathcal{I}_g} w_i^g \nabla g_i(z^*) - \sum_{i=1}^{p} w_i^h \nabla h_i(z^*) + \sum_{i\in\alpha\cup\beta} w_i^G \nabla G_i(z^*) + \sum_{i\in\gamma\cup\beta} w_i^H \nabla H_i(z^*). \end{aligned} \tag{3.53}$$

The choice of the sets $\hat{\beta}_1$ and $\hat{\beta}_2$ guarantee, in particular, that

$$\begin{array}{llll} u_i^G \geq 0 & \forall i \in \beta^G \backslash \beta_1^{GH}, & w_i^G \geq 0 & \forall i \in \beta_1^{GH}, \\ u_i^H \geq 0 & \forall i \in \beta^H \backslash \beta_2^{GH}, & w_i^H \geq 0 & \forall i \in \beta_2^{GH}. \end{array} \tag{3.54}$$

Taking the difference of the two representations of v in (3.53) and applying assumption (A) yields that

$$\begin{aligned} u_i^G - w_i^G &= 0 \qquad \forall i \in \beta_1^{GH} \quad \text{and} \\ u_i^H - w_i^H &= 0 \qquad \forall i \in \beta_2^{GH}. \end{aligned}$$

Together with (3.54), this yields that

$$u_i^G \geq 0 \quad \forall i \in \beta^G \quad \text{and} \quad u_i^H \geq 0 \quad \forall i \in \beta^H.$$

Finally, applying Lemma 3.3 to the representation (3.48) of $\mathcal{T}^{lin}(z^*)$ yields that $v \in \mathcal{T}^{lin}(z^*)^*$. This concludes the proof. $\square$

The purpose of introducing assumption (A) was to offer a more tangible and more easily verifiable property than the intersection property (I). Determining the sets β^G and β^H requires checking whether an inequality is nonsingular in the sense of Definition 3.21. To avoid having to do this, we recall the following concept, stronger than assumption (A), see, e.g., [78].

Definition 3.23 *Let a feasible point z^* of the MPEC* (1.1) *be given. The* partial MPEC-LICQ *is said to hold at z^* if the implication*

$$\sum_{i \in \mathcal{I}_g} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(z^*) - \sum_{i \in \alpha \cup \beta} \lambda_i^G \nabla G_i(z^*) - \sum_{i \in \gamma \cup \beta} \lambda_i^H \nabla H_i(z^*) = 0$$
$$\Longrightarrow \begin{cases} \lambda_\beta^G = 0 \\ \lambda_\beta^H = 0 \end{cases}$$

holds.

Note that partial MPEC-LICQ obviously implies assumption (A) since $\beta_1^{GH} \subseteq \beta$ and $\beta_2^{GH} \subseteq \beta$. This, together with Corollary 3.20 and Lemma 3.22 gives the following corollary.

Corollary 3.24 *Let z^* be a feasible point of the MPEC* (1.1) *satisfying MPEC-GCQ. Furthermore, let z^* satisfy any one of the following conditions*

(a) intersection property (I);

(b) assumption (A);

(c) partial MPEC-LICQ.

Then GCQ holds at z^**.*

A few notes on Corollary 3.24 are in order. If we replace MPEC-GCQ with the stronger MPEC-ACQ, statements (b) and (c) have been stated in [19]. If MPEC-GCQ is replaced by the still stronger assumption that all the associated nonlinear programs $\text{NLP}_*(\beta_1, \beta_2)$ satisfy the standard Abadie CQ, statement (b) has, in essence, been shown in [55].

Summary

In closing, we wish to collect the most important results from this chapter. Figure 3.1 shows the relationship of all the constraint qualifications discussed in this chapter. These relationships hold for all feasible z^*, with the exception of MPEC-SMFCQ, which is only defined for local minimizers of the MPEC (1.1).

$$\begin{tikzcd}
 & \text{MPEC-LICQ} \arrow[d, Rightarrow] \arrow[r, Rightarrow] & \text{Guignard CQ} \arrow[dr, Leftrightarrow] & \\
 & \text{MPEC-SMFCQ} \arrow[d, Rightarrow] \arrow[ur, Rightarrow] & & \text{MPEC-GCQ + (I)} \\
\text{MPEC-Slater CQs} \arrow[dr, Rightarrow] & \text{MPEC-MFCQ} \arrow[d, Rightarrow] & \text{MPAEC} \arrow[dl, Rightarrow] & \\
 & \text{MPEC-ACQ} \arrow[d, Rightarrow] & & \\
 & \text{MPEC-GCQ} & &
\end{tikzcd}$$

Figure 3.1: Relationships of constraint qualifications.

We intentionally included the implication arrow pointing from MPEC-LICQ to Guignard CQ, because MPEC-SMFCQ is defined only for local minimizers of the MPEC (1.1), while MPEC-LICQ and Guignard CQ are defined for all feasible points of the MPEC.

Part II

Stationarity Concepts

Chapter 4

Necessary Optimality Conditions for MPECs

In the previous chapters we have focused on constraint qualifications, general and ones tailored specifically to MPECs. Their purpose, of course, is to provide the means to obtain necessary optimality conditions. This is the center of focus for this chapter and the next. In the current chapter, we will reobtain known first order results for MPECs using a very simple and elegant approach. As a by-product, we will obtain a new stationarity concept, A-stationarity. This is of passing interest only, however, since this result will be supplanted by M-stationarity in Chapter 5.

4.1 A Standard Nonlinear Programming Approach

Stationarity concepts can be divided into two types, one involving descent directions of the objective, the other centered around the Lagrange function. We will first introduce the former type, before going on to discuss the Lagrange function.

Definition 4.1 *Let a feasible point z^* of the MPEC* (1.1) *be given. We then call z^** B-stationary*, if*

$$\nabla f(z^*) \in \mathcal{T}(z^*)^*, \tag{4.1}$$

we call it MPEC-linearized B-stationary*, if*

$$\nabla f(z^*) \in \mathcal{T}^{lin}_{\text{MPEC}}(z^*)^*, \tag{4.2}$$

and linearized B-stationary*, if*

$$\nabla f(z^*) \in \mathcal{T}^{lin}(z^*)^*. \tag{4.3}$$

In view of the inclusions

$$\mathcal{T}^{lin}(z^*)^* \subseteq \mathcal{T}^{lin}_{\text{MPEC}}(z^*)^* \subseteq \mathcal{T}(z^*)^*$$

(obtained by applying Lemma 2.9 (vi) to (3.18)), we have

$$\text{linearized B-stationarity} \Longrightarrow \text{MPEC-linearized B-stationarity} \Longrightarrow \text{B-stationarity.}$$

The 'B' in "B-stationarity," incidentally, refers to Bouligand.

There has been some disagreement within the MPEC community as to the names of the variants of B-stationarity. In particular, Scheel and Scholtes [62] refer to MPEC-linearized B-stationarity (4.2) as B-stationarity. However, we feel that this only adds to the confusion, since in standard nonlinear programming, (4.1) is referred to as B-stationarity, and an MPEC is nothing more than a specialization of a nonlinear program.

To differentiate the two concepts, we use the names B-stationarity and MPEC-linearized B-stationarity, and have added linearized B-stationarity for completeness' sake.

Before we discuss stationarity concepts based on the Lagrangian in the context of MPECs, we shall discuss them using the more general nonlinear program (2.1), which we reproduce here, for ease of reference:

$$\begin{array}{rl} \min & f(z) \\ \text{s.t.} & g(z) \leq 0, \\ & h(z) = 0, \end{array} \tag{4.4}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable.

The *Lagrange function* associated with the program (4.4) is given by

$$L(z, \lambda) := f(z) + \sum_{i=1}^{m} \lambda_i^g g_i(z) + \sum_{i=1}^{p} \lambda_i^h h_i(z). \tag{4.5}$$

The variable $\lambda = (\lambda^g, \lambda^h)$ in (4.4) is called the *Lagrange multiplier*. With the Lagrange function, we can define the following *Karush-Kuhn-Tucker*, or *KKT*, conditions of the program (4.4):

$$\begin{aligned} &\nabla_z L(z, \lambda) = \nabla f(z) + \sum_{i=1}^{m} \lambda_i^g \nabla g(z) + \sum_{i=1}^{p} \lambda_i^h \nabla h(z) = 0, \\ &h(z) = 0, \\ &g(z) \leq 0, \qquad \lambda^g \geq 0, \qquad g(z)^T \lambda^g = 0. \end{aligned} \tag{4.6}$$

Note that any point z satisfying the KKT conditions is, in particular, feasible for the program (4.4). Any vector (z^*, λ^*) satisfying (4.6) is called a *KKT point* of the program (4.4). For a detailed discussion of the Lagrange function and KKT conditions, we refer the interested reader to [27, 5, 4, 44].

Various of the proofs to follow require the well-known Farkas' theorem of the alternative. We therefore recall it at this point and refer to, e.g., [44, Theorem 2.4.6] for a proof.

Lemma 4.2 *(Farkas)*
Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^n$ be given. Then the following are equivalent:

(a) it holds that $b^T d \leq 0$ for all $d \in \mathbb{R}^n$ with $Ad \leq 0$;

(b) there exists a vector $y \in \mathbb{R}^m$ such that the system $A^T y = b$, $y \geq 0$ has a solution.

Before we can show that a KKT point, under certain circumstances, is a necessary optimality condition for the program (4.4), we recall the following lemma, a proof of which may be found in [27, Lemma 2.30].

Lemma 4.3 *Let z^* be a local minimizer of the program* (4.4)*. Then z^* is B-stationary.*

The following proposition is a well-known first order condition, known from nonlinear programming. Although a proof may be found, e.g., in [5, Theorem 6.2.4], we give a proof here, since we will use the same technique for a later result, Theorem 4.16.

Proposition 4.4 *Let z^* be a local minimizer of the program* (4.4) *and let Guignard CQ hold at z^*. Then there exists a Lagrange multiplier λ^* such that (z^*, λ^*) satisfies the KKT conditions* (4.6)*, i.e. (z^*, λ^*) is a KKT point.*

Proof. Since z^* is a local minimizer, we may apply Lemma 4.3. Together with the fact that Guignard CQ holds at z^*, we obtain that

$$\nabla f(z^*) \in \mathcal{T}(z^*)^* \stackrel{\text{GCQ}}{=} \mathcal{T}^{lin}(z^*)^*.$$

Here, $\mathcal{T}^{lin}(z^*)$ denotes the linearized tangent cone of the program (4.4) given by (2.11). Plugging in the definition of the dual cone (see Definition 2.6), we obtain that $-\nabla f(z^*)^T d \leq 0$ for all $d \in \mathbb{R}^n$ with $Ad \leq 0$, where A is defined by

$$A := \begin{pmatrix} g'_{\mathcal{I}_g}(z^*) \\ h'(z^*) \\ -h'(z^*) \end{pmatrix}.$$

Farkas' theorem of the alternative, Lemma 4.2, then yields the existence of a $y \geq 0$ such that $A^T y = -\nabla f(z^*)$. We then denote the components of y by $\lambda^g_{\mathcal{I}_g}$, λ^{h+}, and λ^{h-}, in that order. Setting $\lambda^g_i := 0$ for all $i \notin \mathcal{I}_g$ as well as $\lambda^h := \lambda^{h+} - \lambda^{h-}$ then immediately yields that $(z^*, \lambda^*) := (z^*, \lambda^g, \lambda^h)$ is a KKT point of the program (4.4). This concludes the proof. $\square$

Since Guignard CQ is implied by all the standard constraint qualifications we introduced in Chapter 2 (see Proposition 2.8), the KKT conditions (4.6) are, of course, a first order optimality condition under any of these CQs.

4.2 A Direct Application to MPECs

In this section we will investigate the results we obtain when we apply Proposition 4.4 to our MPEC (1.1).

Consider, therefore, a KKT point $(z^*, \hat\lambda)$ of the MPEC (1.1) with $\hat\lambda = (\hat\lambda^g, \hat\lambda^h, \hat\lambda^G, \hat\lambda^H, \hat\lambda^\theta)$. Since z^* is, in particular, feasible, we shall neglect conditions which pertain only to feasibility. Setting $\theta(z) := G(z)^T H(z)$ yields the following representation of the essential conditions for a KKT point (cf. (4.6)):

$$
\begin{aligned}
&0 = \nabla f(z^*) + \sum_{i=1}^m \hat\lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^p \hat\lambda_i^h \nabla h_i(z^*) - \sum_{i=1}^l \big[\hat\lambda_i^G \nabla G_i(z^*) + \hat\lambda_i^H \nabla H_i(z^*)\big] + \hat\lambda^\theta \nabla\theta(z^*),\\
&\theta(z^*) = 0,\\
&G(z^*) \geq 0, \qquad \hat\lambda^G \geq 0, \qquad (\hat\lambda^G)^T G(z^*) = 0,\\
&H(z^*) \geq 0, \qquad \hat\lambda^H \geq 0, \qquad (\hat\lambda^H)^T H(z^*) = 0,\\
&g(z^*) \leq 0, \qquad \hat\lambda^g \geq 0, \qquad (\hat\lambda^g)^T g(z^*) = 0.
\end{aligned}
\tag{4.7}
$$

Keeping in mind that

$$
\nabla\theta(z^*) = \sum_{i=1}^l \big[G_i(z^*)\nabla H_i(z^*) + H_i(z^*)\nabla G_i(z^*)\big],
$$

we order the sums in the first line of (4.7) by gradient. Setting $\lambda^g := \hat\lambda^g$, $\lambda^h := \hat\lambda^h$, $\lambda_\alpha^G := \hat\lambda_\alpha^G - \hat\lambda^\theta H_\alpha(z^*)$, $\lambda_{\gamma\cup\beta}^G := \hat\lambda_{\gamma\cup\beta}^G$, $\lambda_{\alpha\cup\beta}^H := \hat\lambda_{\alpha\cup\beta}^H$, and $\lambda_\gamma^H := \hat\lambda_\gamma^H - \hat\lambda^\theta G_\gamma(z^*)$ then yields the following representation of (4.7):

$$
\begin{aligned}
&0 = \nabla f(z^*) + \sum_{i=1}^m \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^p \lambda_i^h \nabla h_i(z^*) - \sum_{i=1}^l \big[\lambda_i^G \nabla G_i(z^*) + \lambda_i^H \nabla H_i(z^*)\big],\\
&\lambda_\alpha^G \ \text{ free}, \qquad \lambda_\beta^G \geq 0, \qquad \lambda_\gamma^G = 0,\\
&\lambda_\gamma^H \ \text{ free}, \qquad \lambda_\beta^H \geq 0, \qquad \lambda_\alpha^H = 0,\\
&g(z^*) \leq 0, \qquad \lambda^g \geq 0, \qquad (\lambda^g)^T g(z^*) = 0.
\end{aligned}
\tag{4.8}
$$

Note that we exploit the complementarity terms in (4.7) and the nature of the sets α and γ to get $\lambda_\gamma^G = 0$ and $\lambda_\alpha^H = 0$.

Based on the representation (4.8), a point z^* is called *strongly stationary* [62] or *primal-dual stationary* [55] if z^* is feasible for the MPEC (1.1) and there exists a Lagrange multiplier $\lambda^* = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ such that (z^*, λ^*) satisfies the conditions (4.8).

In the context of MPECs, it is common to refer to the concept of strong stationarity and the representation (4.8) that goes with it, rather than to a KKT point.

We have shown KKT conditions of an MPEC to imply strong stationarity. The converse is also true as we demonstrate in the following proposition.

Proposition 4.5 *Let z^* be feasible for the MPEC* (1.1). *Then there exists a Lagrange multiplier $\hat\lambda$ such that $(z^*, \hat\lambda)$ satisfies the KKT conditions* (4.7) *if and only if there exists a Lagrange multiplier λ^* such that (z^*, λ^*) satisfies the conditions* (4.8) *for strong stationarity.*

Proof. The forward direction has been demonstrated in the discussion leading up to this proposition.

For the reverse direction, consider a point (z^*, λ^*) with $\lambda^* = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ that satisfies (4.8). We then set $\hat{\lambda}^g := \lambda^g$, $\hat{\lambda}^h := \lambda^h$, $\hat{\lambda}^G_{\gamma\cup\beta} := \lambda^G_{\gamma\cup\beta}$, and $\hat{\lambda}^H_{\alpha\cup\beta} := \lambda^H_{\alpha\cup\beta}$. Furthermore, we set

$$\hat{\lambda}^\theta := \max\Big\{ \max_{i\in\alpha}\Big\{\frac{-\lambda_i^G}{H_i(z^*)}\Big\}, \max_{i\in\gamma}\Big\{\frac{-\lambda_i^H}{G_i(z^*)}\Big\}\Big\}.$$

Note that $\hat{\lambda}^\theta$ is well defined, since, by definition of the sets α and γ, it holds that $H_\alpha(z^*) > 0$ and $G_\gamma(z^*) > 0$ (see (1.2)). This choice of $\hat{\lambda}^\theta$ guarantees that $\hat{\lambda}^G_\alpha := \lambda^G_\alpha + \hat{\lambda}^\theta H_\alpha(z^*) \geq 0$ and $\hat{\lambda}^H_\gamma := \lambda^H_\gamma + \hat{\lambda}^\theta G_\gamma(z^*) \geq 0$ are nonnegative. Thus, given the conditions (4.8). we can construct a point $(z^*, \hat{\lambda})$ satisfying the conditions (4.7). This concludes the proof. □

Since a KKT point is equivalent to a strongly stationary point, it follows immediately from Proposition 4.4 that strong stationarity is a necessary optimality condition under Guignard CQ. This is stated in the following theorem.

Theorem 4.6 *Let z^* be a local minimizer of the MPEC* (1.1)*. If Guignard CQ holds in z^*, then there exists a Lagrange multiplier λ^* such that (z^*, λ^*) satisfies the conditions* (4.8)*, i.e. z^* is strongly stationary.*

At first glance it is easy to reason that since both MPEC-LICQ and MPEC-SMFCQ imply Guignard CQ, strong stationarity is a necessary first order condition under them. This reasoning is indeed valid for MPEC-LICQ. However, the proof of Theorem 3.5 (see page 68), which states that MPEC-SMFCQ implies Guignard CQ, requires that we already know that strong stationarity is a first order condition under MPEC-SMFCQ. This, therefore, has to be obtained using different techniques, which we do in Theorem 4.14.

A question of some interest is whether a KKT point, or equivalently, a strongly stationary point, is a sufficient optimality condition, perhaps under some additional assumptions. In standard nonlinear programming, a KKT point is a sufficient optimality condition if the program in question is convex.

We have already stated that the MPEC (1.1) is never a convex program. However, we introduced the MPEC-convexly constrained program, see (3.23), and were able to adapt the Slater CQs to this program. In a similar fashion, we are now able to show that strong stationarity is a sufficient optimality condition if the MPEC is MPEC-convexly constrained and has a convex objective function. This is stated in the following theorem.

Theorem 4.7 *Let z^* be a strongly stationary point of the MPEC-convexly constrained program* (3.23) *with convex objective function $f : \mathbb{R}^n \to \mathbb{R}$. Then z^* is a local minimizer of* (3.23)*.*

Proof. Since f and g_i for all $i = 1, \ldots, m$ are convex, we may write

$$f(z) - f(z^*) \;\geq\; \nabla f(z^*)^T(z - z^*), \tag{4.9a}$$

$$g_i(z) - g_i(z^*) \;\geq\; \nabla g_i(z^*)^T(z - z^*), \qquad \forall i = 1, \ldots, m \tag{4.9b}$$

for all $z \in \mathbb{R}^n$.

Now, let z be an arbitrary feasible point of the MPEC (3.23). Using (4.9a) and expressing $\nabla f(z^*)$ with the help of the gradient of the Lagrangian (see (4.8)), we obtain the following:

$$\begin{aligned} f(z) &\geq f(z^*) + \nabla f(z^*)^T (z - z^*) \\ &= f(z^*) - \sum_{i \in \mathcal{I}_g} \underbrace{\lambda_i^g}_{\geq 0} \underbrace{\nabla g_i(z^*)^T (z - z^*)}_{\leq \underbrace{g_i(z)}_{\leq 0} - \underbrace{g_i(z^*)}_{=0} \, (4.9\mathrm{b})} \\ &\quad - \sum_{i=1}^{p} \lambda_i^h \nabla h_i(z^*)^T (z - z^*) \\ &\quad + \sum_{i \in \alpha \cup \beta} \lambda_i^G \nabla G_i(z^*)^T (z - z^*) \\ &\quad + \sum_{i \in \gamma \cup \beta} \lambda_i^H \nabla H_i(z^*)^T (z - z^*) \\ &\geq f(z^*) - \sum_{i=1}^{p} \lambda_i^h \big[\underbrace{v_i^T z + \eta_i}_{=h_i(z)=0} - \underbrace{(v_i^T z^* + \eta_i)}_{=h_i(z^*)=0} \big] \\ &\quad + \sum_{i \in \alpha \cup \beta} \underbrace{\lambda_i^G}_{\geq 0 \text{ for } i \in \beta} \big[\underbrace{w_i^T z + \chi_i}_{=G_i(z) \geq 0} - \underbrace{(w_i^T z^* + \chi_i)}_{=G_i(z^*)=0} \big] \\ &\quad + \sum_{i \in \gamma \cup \beta} \underbrace{\lambda_i^H}_{\geq 0 \text{ for } i \in \beta} \big[\underbrace{x_i^T z + \xi_i}_{=H_i(z) \geq 0} - \underbrace{(x_i^T z^* + \xi_i)}_{=H_i(z^*)=0} \big] \\ &\geq f(z^*) + \sum_{i \in \alpha} \lambda_i^G G_i(z) + \sum_{i \in \gamma} \lambda_i^H H_i(z). \end{aligned} \tag{4.10}$$

Given the nature of the sets α and γ and the fact that G and H are continuous, there exists a neighborhood $\mathcal{N}$ of z^* such that

$$G_\gamma(z) > 0, \quad \forall z \in \mathcal{Z} \cap \mathcal{N},$$
$$H_\alpha(z) > 0, \quad \forall z \in \mathcal{Z} \cap \mathcal{N}.$$

Due to the complementarity term, this yields

$$G_\alpha(z) = 0 \quad \text{and} \quad H_\gamma(z) = 0 \quad \forall z \in \mathcal{Z} \cap \mathcal{N}.$$

Hence it follows from (4.10) that

$$f(z) \geq f(z^*), \qquad \forall z \in \mathcal{Z} \cap \mathcal{N}.$$

This concludes the proof. □

Note that, unlike in standard nonlinear programming, strong stationarity at z^* only implies *local* optimality in the setting of Theorem 4.7. In standard nonlinear programming, a KKT point of a convex program implies that this point is a *global* minimizer of the program.

KKT-type Stationarity Concepts

Although strong stationarity is a first order condition under MPEC-LICQ, Guignard CQ, and, as we will show later, under MPEC-SMFCQ, it is not a necessary optimality condition under any of the weaker constraint qualifications, MPEC-MFCQ, MPEC-ACQ, or MPEC-GCQ (Example 4.13 demonstrates this). The investigation of these constraint qualifications gave rise to a whole cacophony of stationarity concepts tailored to MPECs, which we will introduce in the following.

To this end, we define the following conditions, differing from strong stationarity (4.8) only in the restrictions imposed upon λ_β^G and λ_β^H:

$$
\begin{aligned}
&0 = \nabla f(z^*) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(z^*) - \sum_{i=1}^{l} \left[\lambda_i^G \nabla G_i(z^*) + \lambda_i^H \nabla H_i(z^*)\right],\\
&\lambda_\alpha^G \ \text{free}, \qquad \lambda_\beta^G \ \text{free}, \qquad \lambda_\gamma^G = 0,\\
&\lambda_\gamma^H \ \text{free}, \qquad \lambda_\beta^H \ \text{free}, \qquad \lambda_\alpha^H = 0,\\
&g(z^*) \le 0, \qquad \lambda^g \ge 0, \qquad (\lambda^g)^T g(z^*) = 0.
\end{aligned} \tag{4.11}
$$

With this, we can define the following stationarity concepts.

Definition 4.8 *Let z^* be feasible for the MPEC* (1.1). *We then call z^**

(a) weakly stationary *[55], if there exists a Lagrange multiplier $\lambda^* = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ such that (z^*, λ^*) satisfies the conditions* (4.11)*;*

(b) A-stationary *[19], if there exists a Lagrange multiplier $\lambda^* = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ such that (z^*, λ^*) satisfies the conditions* (4.11) *and it additionally holds that $\lambda_i^G \ge 0$ or $\lambda_i^H \ge 0$ for all $i \in \beta$;*

(c) C-stationary *[19], if there exists a Lagrange multiplier $\lambda^* = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ such that (z^*, λ^*) satisfies the conditions* (4.11) *and it additionally holds that $\lambda_i^G \lambda_i^H \ge 0$ for all $i \in \beta$;*

(d) M-stationary *[19], if there exists a Lagrange multiplier $\lambda^* = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ such that (z^*, λ^*) satisfies the conditions* (4.11) *and it additionally holds that either $\lambda_i^G > 0$ and $\lambda_i^H > 0$, or $\lambda_i^G \lambda_i^H = 0$ for all $i \in \beta$;*

(e) strongly stationary *[19], if there exists a Lagrange multiplier $\lambda^* = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ such that (z^*, λ^*) satisfies the conditions* (4.11) *and it additionally holds that $\lambda_i^G \ge 0$ and $\lambda_i^H \ge 0$ for all $i \in \beta$.*

A word on the choice of names is in order. The ‘A’ stands for “alternative,” since either Lagrange multiplier must be nonnegative. Also, A-stationarity was first obtained when we investigated the MPEC-Abadie CQ, so the ‘A’ originally also stood for “Abadie.” Finally, ‘C’ and ‘M’ come from the fact that they are obtained when the Clarke and Mordukhovich calculi are applied (see Theorem 4.12 and Chapter 5).

This multitude of stationarity concepts has spawned the joking reference “alphabet soup of stationarity concepts,” which is particularly fitting if, as is common, weak and strong stationarity are referred to as W- and S-stationarity.

Note that in the nondegenerate case, i.e. when $\beta = \varnothing$, all the stationarity concepts defined in Definition 4.8 coincide.

The following string of implications follows directly from Definition 4.8:

$$\text{strongly stationary} \Longrightarrow \text{M-stationary} \begin{array}{c} \nearrow \quad \text{A-stationary} \quad \searrow \\ \searrow \quad \text{C-stationary} \quad \nearrow \end{array} \text{weakly stationary}$$

Furthermore, clearly M-stationarity is the intersection of A- and C-stationarity.

4.3 An Indirect Application to MPECs

In this section we will focus on obtaining first order conditions under the constraint qualifications we introduced in Chapter 3. In the previous section, we already found strong stationarity to be a first order condition under Guignard CQ and hence under MPEC-LICQ. For this result we will give an alternate proof since it provides us with a simple and elegant approach to first order conditions. We will also use this approach to acquire first order conditions under both MPEC-SMFCQ and MPEC-MFCQ.

We commence by proving, using an alternate technique, that strong stationarity is a necessary optimality condition under MPEC-LICQ. This result has been proved before, using still different techniques, see [62, Theorem 2, (2)] as well as [55, Theorem 3], where strong stationarity is referred to as *primal-dual stationarity.*

Theorem 4.9 *Let z^* be a local minimizer of the MPEC* (1.1)*. If MPEC-LICQ holds in z^*, then there exists a unique Lagrange multiplier λ^* such that the conditions* (4.8) *are satisfied. In particular, z^* is strongly stationary.*

Proof. We shall consider two programs (the reason for this will become clear as the proof unfolds): $\text{NLP}_*(\beta_1, \beta_2)$ and its (in a sense) complementary program $\text{NLP}_*(\beta_2, \beta_1)$ (note the inverted positions of β_1 and β_2). The vector z^* is a local minimizer of these two programs since in both cases it is feasible and the feasible region of the corresponding program is a subset of the feasible region of the original MPEC (1.1).

As in the proof of Theorem 3.4, the condition for MPEC-LICQ at z^* is identical to the condition for standard LICQ of $\text{NLP}_*(\beta_1, \beta_2)$ at z^*. Hence LICQ holds for all $\text{NLP}_*(\beta_1, \beta_2)$ at z^*.

Since LICQ at z^* implies Guignard CQ at z^* (see Proposition 2.8), we can apply Proposition 4.4 to obtain the KKT conditions (4.6) of $\mathrm{NLP}_*(\beta_1, \beta_2)$. Hence, there exists a Lagrange multiplier $\tilde{\lambda}$ such that

$$
\begin{aligned}
&0 = \nabla f(z^*) + \sum_{i=1}^{m} \tilde{\lambda}_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \tilde{\lambda}_i^h \nabla h_i(z^*) - \sum_{i=1}^{l} \left[\tilde{\lambda}_i^G \nabla G_i(z^*) + \tilde{\lambda}_i^H H_i(z^*)\right],\\
&G_{\alpha\cup\beta_1}(z^*) = 0, \quad G_{\gamma\cup\beta_2}(z^*) \geq 0, \quad \tilde{\lambda}^G_{\gamma\cup\beta_2} \geq 0, \quad (\tilde{\lambda}^G_{\gamma\cup\beta_2})^T G_{\gamma\cup\beta_2}(z^*) = 0,\\
&H_{\gamma\cup\beta_2}(z^*) = 0, \quad H_{\alpha\cup\beta_1}(z^*) \geq 0, \quad \tilde{\lambda}^H_{\alpha\cup\beta_1} \geq 0, \quad (\tilde{\lambda}^H_{\alpha\cup\beta_1})^T H_{\alpha\cup\beta_1}(z^*) = 0,\\
&h(z^*) = 0, \quad g(z^*) \leq 0, \quad \tilde{\lambda}^g \geq 0, \quad (\tilde{\lambda}^g)^T g(z^*) = 0.
\end{aligned}
\tag{4.12}
$$

Similarly, taking the complementary program $\mathrm{NLP}_*(\beta_2, \beta_1)$, there exists a Lagrange multiplier $\hat{\lambda}$ such that the following KKT conditions are satisfied:

$$
\begin{aligned}
&0 = \nabla f(z^*) + \sum_{i=1}^{m} \hat{\lambda}_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \hat{\lambda}_i^h \nabla h_i(z^*) - \sum_{i=1}^{l} \left[\hat{\lambda}_i^G \nabla G_i(z^*) + \hat{\lambda}_i^H H_i(z^*)\right],\\
&G_{\alpha\cup\beta_2}(z^*) = 0, \quad G_{\gamma\cup\beta_1}(z^*) \geq 0, \quad \hat{\lambda}^G_{\gamma\cup\beta_1} \geq 0, \quad (\hat{\lambda}^G_{\gamma\cup\beta_1})^T G_{\gamma\cup\beta_1}(z^*) = 0,\\
&H_{\gamma\cup\beta_1}(z^*) = 0, \quad H_{\alpha\cup\beta_2}(z^*) \geq 0, \quad \hat{\lambda}^H_{\alpha\cup\beta_2} \geq 0, \quad (\hat{\lambda}^H_{\alpha\cup\beta_2})^T H_{\alpha\cup\beta_2}(z^*) = 0,\\
&h(z^*) = 0, \quad g(z^*) \leq 0, \quad \hat{\lambda}^g \geq 0, \quad (\hat{\lambda}^g)^T g(z^*) = 0.
\end{aligned}
\tag{4.13}
$$

Note that since $G_\gamma(z^*) > 0$ and $H_\alpha(z^*) > 0$ due to the definition of the sets α and γ, the complementarity terms in (4.12) and (4.13) yield that $\tilde{\lambda}^G_\gamma = \hat{\lambda}^G_\gamma = 0$ and $\tilde{\lambda}^H_\alpha = \hat{\lambda}^H_\alpha = 0$. Similarly, $\tilde{\lambda}^g_i = \hat{\lambda}^g_i = 0$ for all $i \notin \mathcal{I}_g$.

We use this information in the following reasoning where we equate the first equations of (4.12) and (4.13):

$$
\begin{aligned}
0 &= \sum_{i=1}^{m} (\hat{\lambda}_i^g - \tilde{\lambda}_i^g) \nabla g_i(z^*) + \sum_{i=1}^{p} (\hat{\lambda}_i^h - \tilde{\lambda}_i^h) \nabla h_i(z^*)\\
&\quad - \sum_{i=1}^{l} \left[(\hat{\lambda}_i^G - \tilde{\lambda}_i^G) \nabla G_i(z^*) + (\hat{\lambda}_i^H - \tilde{\lambda}_i^H) \nabla H_i(z^*)\right]\\
&= \sum_{i\in\mathcal{I}_g} (\hat{\lambda}_i^g - \tilde{\lambda}_i^g) \nabla g_i(z^*) + \sum_{i=1}^{p} (\hat{\lambda}_i^h - \tilde{\lambda}_i^h) \nabla h_i(z^*)\\
&\quad - \sum_{i\in\alpha\cup\beta} (\hat{\lambda}_i^G - \tilde{\lambda}_i^G) \nabla G_i(z^*) - \sum_{i\in\gamma\cup\beta} (\hat{\lambda}_i^H - \tilde{\lambda}_i^H) \nabla H_i(z^*).
\end{aligned}
$$

Since all terms involved are linearly independent due to MPEC-LICQ, we have $\tilde{\lambda} = \hat{\lambda}$. By setting $\lambda^* := \tilde{\lambda} = \hat{\lambda}$, we see that λ^* has the combined characteristics of the Lagrange multipliers in both (4.12) and (4.13), in particular

$$
(\lambda^G_\beta)^* \geq 0, \qquad (\lambda^G_\gamma)^* = 0,
$$

$$(\lambda_\beta^H)^* \geq 0, \qquad (\lambda_\alpha^H)^* = 0,$$

satisfying the conditions (4.8) for strong stationarity.

Uniqueness of the Lagrange multiplier follows immediately from the fact that the gradients occurring in the gradient of the Lagrangian in (4.8) are linearly independent due to MPEC-LICQ. This concludes the proof. □

We now use the method of proof of Theorem 4.9 to investigate MPEC-MFCQ.

Theorem 4.10 *Let z^* be a local minimizer of the MPEC* (1.1). *If MPEC-MFCQ holds in z^* then there exists a Lagrange multiplier $\lambda^* = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ such that*

$$\begin{aligned}
&0 = \nabla f(z^*) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(z^*) - \sum_{i=1}^{l} \left[\lambda_i^G \nabla G_i(z^*) + \lambda_i^H \nabla H_i(z^*)\right],\\
&\begin{array}{lll} \lambda_\alpha^G \ \ \textit{free}, & & \lambda_\gamma^G = 0,\\ & \lambda_i^G \geq 0 \vee \lambda_i^H \geq 0 \quad \forall i \in \beta & \\ \lambda_\gamma^H \ \ \textit{free}, & & \lambda_\alpha^H = 0, \end{array}\\
&g(z^*) \leq 0, \qquad \lambda^g \geq 0, \qquad g(z^*)^T \lambda^g = 0,
\end{aligned} \tag{4.14}$$

i.e. z^* is A-stationary.

Proof. Since MPEC-MFCQ holds at z^*, we may apply Lemma 3.12 to acquire that standard MFCQ holds for every $\mathrm{NLP}_*(\beta_1, \beta_2)$. Since MFCQ implies Guignard CQ, we once more apply Proposition 4.4 to obtain the existence of a Lagrange multiplier $\tilde{\lambda}$ such that the KKT conditions (4.12) are satisfied.

By the same reasoning as in the proof of Theorem 4.9, we have that $\tilde{\lambda}_\gamma^G = 0$, $\tilde{\lambda}_\alpha^H = 0$, and $\tilde{\lambda}_i^g = 0$ for all $i \notin \mathcal{I}_g$.

Additionally, we are able to conclude that

$$\tilde{\lambda}_{\beta_2}^G \geq 0, \qquad \tilde{\lambda}_{\beta_1}^H \geq 0.$$

Setting $\lambda^* := \tilde{\lambda}$ proves the result. □

Note that the proof of Theorem 4.10 holds for an arbitrary partition (β_1, β_2) of the index set β. Hence we can choose, a priori, such a partition and obtain corresponding Lagrange multipliers $(\lambda^G)^*$ and $(\lambda^H)^*$ such that $(\lambda_i^G)^* \geq 0$ for all $i \in \beta_1$ and $(\lambda_i^H)^* \geq 0$ for all $i \in \beta_2$. This is the statement of the following corollary.

Corollary 4.11 *Let $z^* \in \mathbb{R}^n$ be a local minimizer of the MPEC* (1.1). *If MPEC-MFCQ holds in z^* then for every partition $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$ there exists a Lagrange multiplier $\lambda^* = \lambda^*(\beta_1, \beta_2)$, dependent on the partition (β_1, β_2), such that z^* satisfies the conditions* (4.14).

The first order condition in Corollary 4.11 is called a *primal-dual first-order condition* in [42] and this particular one can be found in Theorem 3.3.6 of the same reference.

A different result for MPEC-MFCQ is stated in [62, Theorem 2 (1)], where C-stationarity is shown to be a necessary optimality condition under MPEC-MFCQ. We reproduce the proof here to demonstrate how our approaches differ.

Theorem 4.12 *Let $z^* \in \mathbb{R}^n$ be a local minimizer of the MPEC* (1.1)*. If MPEC-MFCQ holds in z^*, then there exists a Lagrange multiplier $\lambda^* = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ such that*

$$
\begin{aligned}
&0 = \nabla f(z^*) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(z^*) - \sum_{i=1}^{l} \left[\lambda_i^G \nabla G_i(z^*) + \lambda_i^H \nabla H_i(z^*) \right],\\
&\begin{array}{ll} \lambda_\alpha^G & \textit{free}, \\ \lambda_\gamma^H & \textit{free}, \end{array} \quad \lambda_i^G \lambda_i^H \geq 0 \quad \forall i \in \beta, \quad \begin{array}{l} \lambda_\gamma^G = 0, \\ \lambda_\alpha^H = 0, \end{array} \\
&g(z^*) \leq 0, \qquad \lambda^g \geq 0, \qquad g(z^*)^T \lambda^g = 0,
\end{aligned}
\tag{4.15}
$$

i.e. z^* is C-stationary.

Proof. Clearly, the MPEC (1.1) may equivalently be written as

$$
\begin{array}{rl}
\min & f(z) \\
\text{s.t.} & g(z) \leq 0, \quad h(z) = 0, \\
& \min\{G_i(z), H_i(z)\}, \quad i = 1, \ldots, l.
\end{array}
\tag{4.16}
$$

Application of [8, Theorem 6.1.1] yields the existence of $r \geq 0$, $\lambda^g \geq 0$, λ^h, and $\lambda^{\min}$, not all zero, such that

$$
\begin{aligned}
&0 = r \nabla f(z^*) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \lambda^h \nabla h_i(z^*) + \sum_{i=1}^{l} \lambda^{\min} c_i, \\
&g(z^*) \leq 0, \quad \lambda^g \geq 0, \quad g(z^*)^T \lambda^g = 0,
\end{aligned}
\tag{4.17}
$$

with

$$
c_i \in \partial^{\mathrm{Cl}} \min\{G_i(z^*), H_i(z^*)\} = \begin{cases} \{\nabla G_i(z^*)\} & : \ i \in \alpha, \\ \operatorname{conv}\{\nabla G_i(z^*), \nabla H_i(z^*)\} & : \ i \in \beta, \\ \{\nabla H_i(z^*)\} & : \ i \in \gamma. \end{cases}
\tag{4.18}
$$

Here, ∂^{Cl} denotes the Clarke subgradient, see [8]. Also, Section 5.1 contains some information on it. The values of c_i follow immediately from [8, Proposition 2.3.12, Proposition 2.3.6 (a), and Proposition 2.3.1].

Now, for every $i \in \beta$ there exists a $\mu_i \in [0, 1]$ such that

$$
c_i = \mu_i \nabla G_i(z^*) + (1 - \mu_i) \nabla H_i(z^*).
$$

If we set $\lambda_i^G := \lambda_i^{\min}\mu_i$ and $\lambda_i^H := \lambda_i^{\min}(1-\mu_i)$ for all $i \in \beta$, we obtain that $\lambda_i^G\lambda_i^H = (\lambda_i^{\min})^2\mu_i(1-\mu_i) \geq 0$ for all $i \in \beta$ due to the fact that $\mu_i \in [0,1]$.

Furthermore, noting the values of c_α and c_γ, we set $\lambda_\alpha^H := 0$ and $\lambda_\gamma^G := 0$ to acquire the following representation of (4.17):

$$
\begin{gathered}
0 = r\nabla f(z^*) + \sum_{i=1}^m \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^p \lambda_i^h \nabla h_i(z^*) - \sum_{i=1}^l \left[\lambda_i^G \nabla G_i(z^*) + \lambda_i^H \nabla H_i(z^*)\right], \\
\begin{array}{lll} \lambda_\alpha^G \quad \text{free}, & & \lambda_\gamma^G = 0, \\ & \lambda_i^G\lambda_i^H \geq 0 \quad \forall i \in \beta, & \\ \lambda_\gamma^H \quad \text{free}, & & \lambda_\alpha^H = 0, \end{array} \\
g(z^*) \leq 0, \qquad \lambda^g \geq 0, \qquad g(z^*)^T\lambda^g = 0.
\end{gathered}
\tag{4.19}
$$

These conditions may be interpreted as Fritz John-type C-stationarity conditions. They differ from C-stationarity only in the presence of the nonnegative multiplier r (see (4.15)). Note that this result holds without any assumptions. In particular, it holds without MPEC-MFCQ.

All that remains to be shown is that $r \neq 0$. To this end, assume that $r = 0$. Taking the fact into account that some of the multipliers vanish, the first line of (4.19) may then be written as

$$
0 = \sum_{i \in \mathcal{I}_g} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^p \lambda_i^h \nabla h_i(z^*) + \sum_{i \in \alpha\cup\beta} \lambda_i^G \nabla G_i(z^*) + \sum_{i \in \gamma\cup\beta} \lambda_i^H \nabla H_i(z^*). \tag{4.20}
$$

Next, we take a $d \in \mathbb{R}^n$ satisfying the conditions (3.4) of MPEC-MFCQ, and multiply it with the transpose of (4.20):

$$
0 = \sum_{i \in \mathcal{I}_g} \underbrace{\lambda_i^g}_{\geq 0} \underbrace{\nabla g_i(z^*)^T d}_{<0}.
$$

This immediately yields $\lambda^g = 0$.

The remaining gradients in (4.20) are linearly independent since MPEC-MFCQ holds (see (3.3)). Therefore, $\lambda^h = 0$, $\lambda^G = 0$ and $\lambda^H = 0$.

This, however, is a contradiction to the statement that not all of r, λ^g, λ^h, λ^G, and λ^H are zero (see [8, Theorem 6.1.1]). Therefore, $r \neq 0$. Scaling yields $r = 1$. This concludes the proof. □

Note that Theorems 4.10 and 4.12 merely show the existence of Lagrange multipliers with certain characteristics, but not the exclusion of other Lagrange multipliers. In fact, as the following example (taken from [62]) demonstrates, the respective conditions are not, in general, satisfied by the same set of Lagrange multipliers.

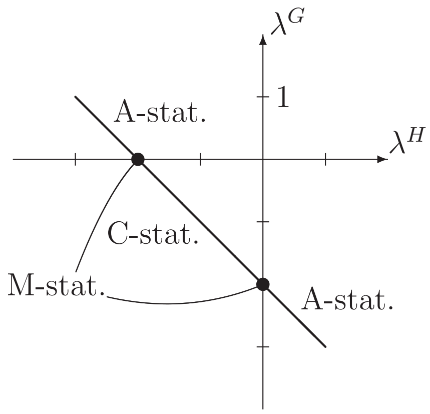

Figure 4.1: The Lagrange multipliers of Example 4.13.

Example 4.13 Consider the following MPAEC:

$$\begin{array}{ll} \min & f(z) := z_1 + z_2 - z_3 \\ \text{s.t.} & g(z) := \begin{pmatrix} -4z_1 + z_3 \\ -4z_2 + z_3 \end{pmatrix} \leq 0, \\ & G(z) := z_1 \geq 0, \\ & H(z) := z_2 \geq 0, \\ & G(z)^T H(z) = z_1 z_2 = 0. \end{array}$$

The origin is the unique solution of this program, and it satisfies MPEC-MFCQ. Obviously, $\alpha = \gamma = \varnothing$ and $\beta = \{1\}$. It is easily verified that the corresponding Lagrange multipliers $\lambda^G := (\lambda_1^G)^*$ and $\lambda^H := (\lambda_1^H)^*$ are subject to the following restrictions:

$$\begin{aligned} &\lambda^H \in [-3, 1], \\ &\lambda^G = -\lambda^H - 2 \end{aligned}$$

(see Figure 4.1).

The Lagrange multipliers satisfying A-stationarity are $\{(\lambda^G, \lambda^H) \mid \lambda^H \in [-3, -2] \cup [0, 1],\ \lambda^G = -\lambda^H - 2\}$, while the conditions for C-stationarity are satisfied by the multipliers $\{(\lambda^G, \lambda^H) \mid \lambda^H \in [-2, 0],\ \lambda^G = -\lambda^H - 2\}$.

Only for $(\lambda^G, \lambda^H) = (0, -2)$ and $(\lambda^G, \lambda^H) = (-2, 0)$ are the conditions for A- and C-stationarity satisfied simultaneously. Since M-stationarity is the intersection of A- and C-stationarity, it too is satisfied for these multipliers.

Note that Example 4.13 also yields that strong stationarity is not a first order condition under MPEC-MFCQ. The example was originally used in [62] for this purpose.

Of considerable interest is the fact that M-stationarity is a necessary optimality condition under MPEC-MFCQ in the case of Example 4.13. The suspicion arose that this might be so in general.

To try to verify this, we first note that the conditions (4.11) for weak stationarity are convex in the Lagrange multipliers λ^g, λ^g, λ^G, and λ^H. Now the question is whether a convex combination of the $2^{|\beta|}$ different A-stationary points (see Corollary 4.11) and the C-stationary (see Theorem 4.12) point can be found to yield an M-stationary point. In fact, the above example demonstrates (and it is easily verified in the general case) that for $|\beta| = 1$ there exists a convex combination of two A-stationary points which yields an M-stationary point. We did not require knowledge of the C-stationary point.

We have run computer simulations to test this in the scenario of $|\beta| = 2$. To this end, we generated all possible partitions of β and for each of these partitions we allowed (arbitrary) values equal to, greater than, and (in the case of free sign) less than 0. So, for each of the 4 A-stationary points, we had $3 \cdot 3 \cdot 2 \cdot 2 = 36$ "sign distributions." We then took all 36^4 possible combinations and checked whether an M-stationary point could be generated by convex combination. This did indeed work. Note that the C-stationary point was not needed here either.

Not needing the C-stationary point seemed very important to us, since we had, at the time, also shown that A-stationarity was a necessary optimality condition under MPEC-ACQ (see Theorem 4.16), not however C-stationarity.

For $|\beta| \geq 3$ computer simulations became infeasible. In the case of $|\beta| = 3$ we already have 8 different partitions and $(3 \cdot 2)^3 = 216$ different "sign distributions." We would therefore have to test $216^8 \approx 4.7 \cdot 10^{18}$ different configurations. Even if we (hopelessly optimistically) assume that we can test one configuration within one clock tick of a modern 3GHz processor, we would still need 50 years to test every configuration.

We tried to find a combinatorial approach to answer this question. Unfortunately, we failed to find a proof using this method. Instead, approaches using the Mordukhovich calculus were eventually able to yield the result. See Chapter 5 for elaborations on this approach.

To come back to the original objective of this section, we finally turn our attention to MPEC-SMFCQ, and use our technique to prove that strong stationarity is a necessary optimality condition under it. This result has appeared before in [62, Theorem 2 (2)], where it was derived using different techniques.

Theorem 4.14 *Let z^* be a local minimizer of the MPEC* (1.1). *If MPEC-SMFCQ holds in z^* then there exists a unique Lagrange multiplier λ^* such that the conditions* (4.8) *hold. In particular, z^* is strongly stationary.*

Proof. First note that MPEC-SMFCQ is defined since z^* is a local minimizer. Since MPEC-SMFCQ implies MPEC-MFCQ (see Corollary 3.2), the proof of Theorem 4.10 yields the existence of a Lagrange multiplier $\tilde{\lambda}$ which satisfies the KKT conditions (4.12). By the same arguments we have the existence of a Lagrange multiplier $\hat{\lambda}$ which satisfies the KKT conditions (4.13).

It is easily verified (taking into account that $G_\beta(z^*) = 0$ and $H_\beta(z^*) = 0$ by definition of the index set β) that both Lagrange multipliers $\tilde{\lambda}$ and $\hat{\lambda}$ also satisfy the KKT conditions of the TNLP(z^*) (3.1), here stated for a multiplier λ:

$$\begin{aligned}
&0 = \nabla f(z^*) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(z^*) - \sum_{i=1}^{l} \left[\lambda_i^G \nabla G_i(z^*) + \lambda_i^H H_i(z^*)\right],\\
&G_{\alpha\cup\beta}(z^*) = 0, \quad G_\gamma(z^*) \geq 0, \quad \lambda_\gamma^G \geq 0, \quad (\lambda_\gamma^G)^T G_\gamma(z^*) = 0,\\
&H_{\gamma\cup\beta}(z^*) = 0, \quad H_\alpha(z^*) \geq 0, \quad \lambda_\alpha^H \geq 0, \quad (\lambda_\alpha^H)^T H_\alpha(z^*) = 0,\\
&h(z^*) = 0, \quad g(z^*) \leq 0, \quad \lambda^g \geq 0, \quad (\lambda^g)^T g(z^*) = 0.
\end{aligned} \tag{4.21}$$

Since SMFCQ holds for the TNLP(z^*) (3.1) by Definition 3.1, the Lagrange multiplier is unique (cf. [37]) and hence $\tilde{\lambda} = \hat{\lambda}$. By the same arguments as in the proof of Theorem 4.9, we set $\lambda^* := \tilde{\lambda} = \hat{\lambda}$, which satisfies the combined characteristics of (4.12) and (4.13). This implies strong stationarity.

Since a Lagrange multiplier satisfying (4.8) also satisfies the KKT conditions (4.21) of the TNLP(z^*) (3.1), and SMFCQ holds for this TNLP(z^*), the Lagrange multiplier is unique (cf., again, [37]). This concludes the proof. □

The result of Theorem 4.14 strongly suggests that MPEC-SMFCQ should imply standard Guignard CQ, since Theorem 4.14 is a stronger result than Theorem 4.6: Though both MPEC-SMFCQ and Guignard CQ imply strong stationarity to be a necessary optimality condition, MPEC-SMFCQ additionally implies the uniqueness of the Lagrange multiplier.

We already stated this implication in Theorem 3.5. However, as we observed in the discussion preceding Theorem 3.5, we have been unsuccessful at finding a direct proof.

With the aid of Theorem 4.14, however, we are now in a position to give a proof of Theorem 3.5, albeit using intricate arguments.

With this purpose in mind, we state the following theorem, originally due to Gould and Tolle [29].

Theorem 4.15 *Let z^* be a feasible point of the MPEC* (1.1)*. Further suppose that for every continuously differentiable objective function f which assumes a local minimizer at z^* under the constraints of the MPEC* (1.1)*, there exists a Lagrange multiplier $\lambda = \lambda(f)$, dependent on f, such that the conditions* (4.8) *for strong stationarity hold. Then Guignard CQ holds at z^*.*

Proof. Recalling that the standard KKT conditions (4.7) at z^* of the MPEC (1.1) are equivalent to z^* being strongly stationary (see Proposition 4.5), this result immediately follows from [5, Theorem 6.3.2]. □

We are now in a position to give a proof of Theorem 3.5 which states that if a feasible point z^* of the MPEC (1.1) satisfies MPEC-SMFCQ, it also satisfies Guignard CQ.

Proof of Theorem 3.5. Theorem 4.14 states that strong stationarity is a necessary optimality condition under MPEC-SMFCQ. This holds for arbitrary MPECs of type (1.1). In particular, this stays true if we take our MPEC and vary the objective function f arbitrarily, but in such way as for the MPEC to assume a local minimizer at z^*. Therefore, by Theorem 4.15, Guignard CQ holds at z^*. □

As was already mentioned, we were unable to find a direct proof for Theorem 3.5. The fact that the proof of Theorem 4.15 relies on the shrewd construction of a nonintuitive continuous function suggests that a direct proof would require a similar construction. The proofs of this result in [29] and [5] both use this approach, though using different functions.

Another interesting facet of Theorem 4.15 is that Guignard CQ obviously is the weakest possible constraint qualification that will guarantee strong stationarity to hold for a local minimizer of the MPEC (1.1). See also the discussion in [5, Section 6.3] in this context. Therefore, by Corollary 3.20, this also demonstrates that the intersection property (I) is the minimum requirement in addition to MPEC-GCQ for strong stationarity to be a first order optimality condition.

4.4 First Order Optimality Conditions under MPEC-GCQ

Up until now, we have discussed those MPEC constraint qualifications that we obtained with the help of TNLP(z^*) (see Definition 3.1). The next question is, of course, which first order optimality conditions can be obtained under MPEC-ACQ and MPEC-GCQ. We cannot expect strong stationarity since MPEC-ACQ is implied by MPEC-MFCQ which, in turn, is not sufficient for strong stationarity (see Example 4.13). However, our next result shows that MPEC-GCQ and hence MPEC-ACQ does imply the weaker A-stationarity.

Theorem 4.16 *Let $z^* \in \mathbb{R}^n$ be a local minimizer of the MPEC* (1.1). *If MPEC-GCQ holds in z^*, then there exists a Lagrange multiplier λ^* such that the conditions* (4.14) *hold, i.e. z^* is A-stationary.*

Proof. Since z^* is a local minimizer, it holds that z^* is B-stationary, i.e.

$$\nabla f(z^*) \in \mathcal{T}(z^*)^*$$

(see Lemma 4.3). Since MPEC-GCQ holds, this is equivalent to

$$\nabla f(z^*) \in \mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)^* = \bigcap_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)^*, \tag{4.22}$$

where the equality follows directly from Lemmas 3.6 and 2.9 (vi).

Choosing a partition $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$ arbitrarily, it follows from (4.22) that

$$\nabla f(z^*) \in \mathcal{T}^{lin}_{\mathrm{NLP}_*(\beta_1,\beta_2)}(z^*)^*. \tag{4.23}$$

We now follow standard arguments from nonlinear programming. The condition (4.23) can also be written as

$$-\nabla f(z^*)^T d \leq 0 \qquad \forall d \in \mathbb{R}^n \text{ with } Ad \leq 0,$$

where the matrix $A \in \mathbb{R}^{(|\mathcal{I}_g|+2p+2|\alpha|+2|\gamma|+3|\beta|)\times n}$ is given by

$$A := \begin{pmatrix} g'_{\mathcal{I}_g}(z^*) \\ h'(z^*) \\ -h'(z^*) \\ G'_{\alpha\cup\beta_1}(z^*) \\ -G'_{\alpha\cup\beta_1}(z^*) \\ H'_{\gamma\cup\beta_2}(z^*) \\ -H'_{\gamma\cup\beta_2}(z^*) \\ -G'_{\beta_2}(z^*) \\ -H'_{\beta_1}(z^*) \end{pmatrix}.$$

Farkas' theorem of the alternative, Lemma 4.2, yields that

$$A^T y = -\nabla f(z^*), \qquad y \geq 0$$

has a solution. Now let us denote the components of y by $\lambda^g_{\mathcal{I}_g}$, λ^{h+}, λ^{h-} ($\lambda^{h+}, \lambda^{h-} \in \mathbb{R}^p$), $\lambda^{G+}_{\alpha\cup\beta_1}$, $\lambda^{G-}_{\alpha\cup\beta_1}$, $\lambda^{H+}_{\gamma\cup\beta_2}$, $\lambda^{H-}_{\gamma\cup\beta_2}$, $\lambda^G_{\beta_2}$, and $\lambda^H_{\beta_1}$, in that order. We then set $\lambda^h := \lambda^{h+} - \lambda^{h-}$, $\lambda^G_{\alpha\cup\beta_1} := \lambda^{G+}_{\alpha\cup\beta_1} - \lambda^{G-}_{\alpha\cup\beta_1}$, and $\lambda^H_{\gamma\cup\beta_2} := \lambda^{H+}_{\gamma\cup\beta_2} - \lambda^{H-}_{\gamma\cup\beta_2}$. Additionally, we set $\lambda^g_i := 0$ ($i \notin \mathcal{I}_g$), $\lambda^G_\gamma := 0$, and $\lambda^H_\alpha := 0$. The resulting vector $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ satisfies the conditions (4.14) for A-stationarity. Setting $\lambda^* := \lambda$ completes the proof. $\square$

Note that, as in the case of Corollary 4.11, we may choose the partition $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$ for which the conditions for A-stationarity should hold, a priori, see (4.23). This immediately yields the following corollary.

Corollary 4.17 *Let $z^* \in \mathbb{R}^n$ be a local minimizer of the MPEC* (1.1)*. If MPEC-GCQ holds in z^* then for every partition $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$ there exists a Lagrange multiplier $\lambda^* = \lambda^*(\beta_1, \beta_2)$, dependent on the partition (β_1, β_2), such the conditions* (4.14) *hold.*

This corollary was, in part, the basis of our hope that M-stationarity might be a first order condition under MPEC-ACQ (we had not thought of MPEC-GCQ at the time), see the discussion following Example 4.13. We in fact prove M-stationarity to be a first order condition under MPEC-GCQ in Chapter 5.

Summary

To give an overview of the important results concerning the various stationarity concepts from this chapter, we collect them in Figure 4.2.

Note that the proof of Proposition 4.4 yields that, in fact, B-stationarity implies the existence of a KKT point in the presence of GCQ. However, a KKT point of the MPEC

$$\begin{tikzcd}
& \text{local minimizer} \arrow[dl, Rightarrow, "\text{MPEC-MFCQ}"'] \arrow[d, Rightarrow] & \\
\text{C-stationary} & \text{B-stationary} \arrow[r, Rightarrow, "\text{GCQ}"] \arrow[d, Rightarrow, shift left, "\text{MPEC-GCQ}"] & \text{strongly stationary} \arrow[d, Leftrightarrow, "(*)"] \\
& \text{MPEC-lin. B-stat.} \arrow[u, Rightarrow, shift left] \arrow[d, Rightarrow] & \text{linearized B-stat.} \arrow[l, Rightarrow] \\
& \text{A-stationary} &
\end{tikzcd}$$

Figure 4.2: Relationships of stationarity concepts.

(1.1) is equivalent to strong stationarity (see Proposition 4.5). Additionally, the proof of Theorem 4.16 yields that MPEC-linearized B-stationarity implies A-stationarity.

The equivalence marked with $(*)$,

$$\text{strongly stationary} \iff \text{linearized B-stationary},$$

follows by the following arguments: For the forward implication, multiply the first line of (4.8) by $d \in \mathcal{T}^{lin}(z^*)$:

$$\begin{aligned}
\nabla f(z^*)^T d \;=\; & -\sum_{i \in \mathcal{I}_g} \underbrace{\lambda_i^g}_{\geq 0} \underbrace{\nabla g_i(z^*)^T d}_{\leq 0} - \sum_{i=1}^{p} \lambda_i^h \underbrace{\nabla h_i(z^*)^T d}_{=0} \\
& + \sum_{i \in \alpha} \lambda_i^G \underbrace{\nabla G_i(z^*)^T d}_{=0} + \sum_{i \in \gamma} \lambda_i^H \underbrace{\nabla H_i(z^*)^T d}_{=0} \\
& + \sum_{i \in \beta} \big[\underbrace{\lambda_i^G}_{\geq 0} \underbrace{\nabla G_i(z^*)^T d}_{\geq 0} + \underbrace{\lambda_i^H}_{\geq 0} \underbrace{\nabla H_i(z^*)^T d}_{\geq 0} \big] \;\geq\; 0.
\end{aligned}$$

The reverse implication follows directly from the proof of Proposition 4.4 and the fact that a KKT point of the MPEC (1.1) may equivalently be expressed as a strongly stationary point.

Chapter 5

M-Stationarity

We dedicate this chapter to showing that M-stationarity is a necessary first order condition under the weakest constraint qualification we have introduced in this thesis, MPEC-GCQ. This result is of some significance since M-stationarity is second only to strong stationarity, while MPEC-GCQ is the weakest constraint qualification for MPECs that we know of.

We present two proofs of this result. The first proof, presented in Section 5.2, is an approach by Jane Ye translated to an MPEC setting. Originally, Ye had formulated the approach for *optimization problems with variational inequality constraints*, or *OPVICs*. Unfortunately, her chain of reasoning contained an erroneous proof. We patch this gap with an alternate result, which is particularly well-suited to dealing with MPECs.

Ye's approach brings with it the introduction of several new concepts for MPECs. Though these are very interesting and expand the knowledge we have of MPECs, we lose track of our original goal, M-stationarity under MPEC-GCQ, since the approach is very cumbersome and bloated. We therefore go on, in Section 5.3, to present a very compact and elegant proof of the same result.

5.1 Preliminaries

Some background is needed in order to investigate the question of M-stationarity further. Central to this is the calculus introduced by Boris Mordukhovich [46, 40, 47]. This is, incidentally, where M-stationarity gets its name from. In this section, we will introduce those elements of the Mordukhovich calculus that we require for our analysis. Additionally, we will recall some properties of multifunctions that we will also need.

Normal Cones

To start off, we will now introduce some normal cones. For more detail on the normal cones we use here, see [47, 40, 61, 8].

Definition 5.1 *Let $\Omega \subseteq \mathbb{R}^l$ be nonempty and closed, and $v \in \Omega$ be given. We call*

$$\hat{N}(v,\Omega) := \{w \in \mathbb{R}^l \mid \limsup_{\substack{v^k \to v \\ \{v^k\} \subseteq \Omega \setminus \{v\}}} w^T(v^k - v)/\left\|v^k - v\right\| \leq 0\} \tag{5.1}$$

the Fréchet normal cone *to Ω at v,*

$$N^\pi(v,\Omega) := \{w \in \mathbb{R}^l \mid \exists \mu > 0 : w^T(u-v) \leq \mu \left\|u - v\right\|^2 \ \forall u \in \Omega\} \tag{5.2}$$

the proximal normal cone *to Ω at v,*

$$N(v,\Omega) := \{\lim_{k\to\infty} w^k \mid \exists \{v^k\} \subseteq \Omega : \lim_{k\to\infty} v^k = v, \ w^k \in N^\pi(v^k,\Omega)\} \tag{5.3}$$

the limiting normal cone *to Ω at v, and*

$$N^{\mathrm{Cl}}(v,\Omega) := \operatorname{cl}\operatorname{conv} N(v,\Omega) \tag{5.4}$$

the Clarke normal cone *to Ω at v. If Ω additionally is convex, we call*

$$N^{\mathrm{conv}}(v,\Omega) := \{w \in \mathbb{R}^l \mid w^T(u-v) \leq 0 \ \forall u \in \Omega\} \tag{5.5}$$

the standard normal cone *to Ω at v.*

By convention, we set $\hat{N}(v,\Omega) = N^\pi(v,\Omega) = N(v,\Omega) = N^{\mathrm{Cl}}(v,\Omega) = N^{\mathrm{conv}}(v,\Omega) := \emptyset$ if $v \notin \Omega$. By $N_\Omega^\times : \mathbb{R}^l \rightrightarrows \mathbb{R}^l$ we denote the multifunction that maps $v \mapsto N^\times(v,\Omega)$, where $\times$ is a placeholder for one of the normal cones defined above.

Since the limiting normal cone is the most important one in our subsequent analysis, we did not furnish it with an index to simplify notation.

Often the normal cones from Definition 5.1 are defined using any nonempty set Ω, i.e. without the assumption that Ω is closed. However, we only require normals to closed sets in our analysis. Assuming that Ω is closed by default facilitates the notation.

Note that if v is in the interior of Ω, any of the above normal cones reduces to $\{0\}$.

The Fréchet normal cone is also referred to as the *regular normal cone*, most notably in [61]. The limiting normal cone is also called the *Mordukhovich normal cone* [73]. Finally, the Clarke normal cone defined in (5.4) is in fact the well-known Clarke normal cone [8].

In the following proposition, we recall the interesting result that, in the definition of the limiting normal cone $N(v,\Omega)$, $N^\pi(v^k,\Omega)$ may be replaced by $\hat{N}(v^k,\Omega)$. For a proof, see [47, Proposition 2.2].

Proposition 5.2 *Let $\Omega \subseteq \mathbb{R}^l$ be nonempty and convex, and $v \in \Omega$ be given. Then it holds that*

$$\begin{aligned} N(v,\Omega) &= \{\lim_{k\to\infty} w^k \mid \exists \{v^k\} \subseteq \Omega : \lim_{k\to\infty} v^k = v, \ w^k \in N^\pi(v^k,\Omega)\} \\ &= \{\lim_{k\to\infty} w^k \mid \exists \{v^k\} \subseteq \Omega : \lim_{k\to\infty} v^k = v, \ w^k \in \hat{N}(v^k,\Omega)\}. \end{aligned} \tag{5.6}$$

In particular, it holds that $\hat{N}(v,\Omega) \subseteq N(v,\Omega)$.

We will now show that the normal cones defined in Definition 5.1 coincide if Ω is convex.

Proposition 5.3 *Let $\Omega \subseteq \mathbb{R}^l$ be nonempty, closed, and convex. Then the normal cones defined in Definition 5.1 coincide for every $v \in \Omega$, i.e. we have*

$$\hat{N}(v,\Omega) = N^{\pi}(v,\Omega) = N(v,\Omega) = N^{\mathrm{Cl}}(v,\Omega) = N^{\mathrm{conv}}(v,\Omega) \tag{5.7}$$

for all $v \in \Omega$.

Proof. For an arbitrary $v \in \Omega$ it obviously holds that

$$N^{\mathrm{conv}}(v,\Omega) \subseteq N^{\pi}(v,\Omega) \subseteq N(v,\Omega) \subseteq N^{\mathrm{Cl}}(v,\Omega). \tag{5.8}$$

By [8, Proposition 2.4.4], it further holds that

$$N^{\mathrm{Cl}}(v,\Omega) = N^{\mathrm{conv}}(v,\Omega)$$

if Ω is convex. Together with (5.8) this yields that

$$N^{\pi}(v,\Omega) = N(v,\Omega) = N^{\mathrm{Cl}}(v,\Omega) = N^{\mathrm{conv}}(v,\Omega).$$

Similarly, it holds that

$$N^{\mathrm{conv}}(v,\Omega) \subseteq \hat{N}(v,\Omega) \subseteq N(v,\Omega) \subseteq N^{\mathrm{Cl}}(v,\Omega), \tag{5.9}$$

where the middle inclusion is due to Proposition 5.2. By the same arguments as above, we obtain the statement of the proposition. □

We now state a Cartesian product rule for the limiting normal cone. This will become important throughout Sections 5.2 and 5.3. A proof may be found in [61, Proposition 6.41].

Proposition 5.4 *Let $\Omega_1 \subseteq \mathbb{R}^{l_1}$ and $\Omega_2 \subseteq \mathbb{R}^{l_2}$ be nonempty and closed. Then, for any $v_1 \in \Omega_1$ and $v_2 \in \Omega_2$ it holds that*

$$N((v_1,v_2),\Omega_1 \times \Omega_2) = N(v_1,\Omega_1) \times N(v_2,\Omega_2). \tag{5.10}$$

To cope with the complementarity term in the constraints of the MPEC (1.1), we require a result which investigates the limiting normal cone to a complementarity set. This result was originally stated in a slightly different format by Outrata in [53, Lemma 2.2], see also [76, Proposition 3.7]. We wish to reproduce a proof here since this result is essential to our analysis. To this end, we will need several auxiliary results. The main result is then presented in Proposition 5.8.

We will use the fact that the limiting normal cone may be expressed as the limit of the Fréchet normal cone (see Proposition 5.2). Therefore, we state, in the following two lemmas, two intuitively clear results for the Fréchet normal cone.

The first lemma states, in essence, that the Fréchet normal cone is local in nature.

Lemma 5.5 *Let $\Omega_1 \subseteq \mathbb{R}^l$ be nonempty and closed, let $\Omega_2 \subseteq \mathbb{R}^l$ be closed, and let $v \in \Omega_1 \backslash \Omega_2$ be given. Then it holds that*

$$\hat{N}(v, \Omega_1 \cup \Omega_2) = \hat{N}(v, \Omega_1). \tag{5.11}$$

Proof. The inclusion $\hat{N}(v, \Omega_1 \cup \Omega_2) \supseteq \hat{N}(v, \Omega_1)$ holds trivially.

For the remaining inclusion, consider the following. Since $v \notin \Omega_2$ and Ω_2 is closed, there exists a neighborhood $\mathcal{V}$ of v such that $\mathcal{V} \nsubseteq \Omega_2$. Then,

$$\hat{N}(v, \Omega_1 \cup \Omega_2) \subseteq \hat{N}(v, \Omega_1),$$

since we can, without loss of generality, pass to a subsequence such that $\{v^k\} \subseteq ((\Omega_1 \cup \Omega_2) \cap \mathcal{V}) \backslash \{v\} \subseteq (\Omega_1 \cap \mathcal{V}) \backslash \{v\}$ in the definition of the Fréchet normal cone (see (5.1)). $\square$

The following lemma investigates the Fréchet normal cone to the union of two sets at a point that lies in the intersection of the two sets.

Lemma 5.6 *Let $\Omega_1, \Omega_2 \subseteq \mathbb{R}^l$ be nonempty and closed, let $\Omega_1 \cap \Omega_2$ be nonempty, and let $v \in \Omega_1 \cap \Omega_2$ be given. Then it holds that*

$$\hat{N}(v, \Omega_1 \cup \Omega_2) = \hat{N}(v, \Omega_1) \cap \hat{N}(v, \Omega_2). \tag{5.12}$$

Proof. Let $w \in \hat{N}(v, \Omega_1 \cup \Omega_2)$ be arbitrarily given. Then, for every sequence $\{v^k\} \subseteq \Omega_1 \backslash \{v\} \subseteq (\Omega_1 \cup \Omega_2) \backslash \{v\}$ with $v^k \to v$ it holds that

$$\limsup_{k\to\infty} w^T(v^k - v)/\left\|v^k - v\right\| \leq 0.$$

This obviously implies that $w \in \hat{N}(v, \Omega_1)$. Similarly, it can be shown that $w \in \hat{N}(v, \Omega_2)$. Together we have

$$w \in \hat{N}(v, \Omega_1) \cap \hat{N}(v, \Omega_2).$$

Conversely, let $w \in \hat{N}(v, \Omega_1) \cap \hat{N}(v, \Omega_2)$ be given. Suppose that there exists a sequence $\{v^k\} \subseteq (\Omega_1 \cup \Omega_2) \backslash \{v\}$ with $v^k \to v$ such that

$$\limsup_{k\to\infty} w^T(v^k - v)/\left\|v^k - v\right\| > 0. \tag{5.13}$$

Now we pass, if necessary, to a subsequence such that either $\{v^k\} \subseteq \Omega_1$ or $\{v^k\} \subseteq \Omega_2$. Since $w \in \hat{N}(v, \Omega_1) \cap \hat{N}(v, \Omega_2)$, however, either $w \in \hat{N}(v, \Omega_1)$ or $w \in \hat{N}(v, \Omega_2)$ results in a contradiction to (5.13). Hence $w \in \hat{N}(v, \Omega_1 \cup \Omega_2)$, completing the proof. $\square$

With the help of the previous two lemmas, we are now able to prove the following lemma, which explicitly determines the Fréchet normal cone to the univariate complementarity set for every point within this set.

Lemma 5.7 *Let the set*

$$\mathcal{C}_i := \{(a_i, b_i) \in \mathbb{R}^2 \mid a_i \geq 0, b_i \geq 0, a_i b_i = 0\} \tag{5.14}$$

be given. Then, for an arbitrary $(a_i, b_i) \in \mathcal{C}_i$, the Fréchet normal cone to $\mathcal{C}_i$ in (a_i, b_i) is given by

$$\hat{N}((a_i, b_i), \mathcal{C}_i) = \begin{cases} \mathbb{R} \times \{0\} & : \ a_i = 0, b_i > 0, \\ \{0\} \times \mathbb{R} & : \ a_i > 0, b_i = 0, \\ \mathbb{R}^2_- & : \ a_i = 0, b_i = 0. \end{cases} \tag{5.15}$$

Proof. It is immediately obvious that $\mathcal{C}_i$ can be rewritten as $\mathcal{C}_i = \mathcal{C}_i^1 \cup \mathcal{C}_i^2$ with

$$\mathcal{C}_i^1 := \{0\} \times \mathbb{R}_+ \qquad \text{and} \qquad \mathcal{C}_i^2 := \mathbb{R}_+ \times \{0\}.$$

We now consider three cases.

$a_i = 0, b_i > 0$. Obviously, it holds that $(a_i, b_i) \in \mathcal{C}_i^1 \backslash \mathcal{C}_i^2$. Invoking Lemma 5.5 ($\mathcal{C}_i^1$ and $\mathcal{C}_i^2$ are obviously closed), we obtain that

$$\begin{aligned} \hat{N}((a_i, b_i), \mathcal{C}_i) &= \hat{N}((a_i, b_i), \mathcal{C}_i^1) \\ &= \{(w_1, w_2) \in \mathbb{R}^2 \mid \limsup_{\substack{b_k \to b_i \\ b_k \in \mathbb{R}_+ \backslash \{b_i\}}} \frac{w_2 (b_k - b_i)}{|b_k - b_i|} \leq 0\} \\ &= \{(w_1, w_2) \in \mathbb{R}^2 \mid w_2 = 0\} \\ &= \mathbb{R} \times \{0\}. \end{aligned}$$

The penultimate equality can be deduced as follows: Assume $w_2 > 0$. Choosing the sequence $b_k := b_i + \frac{1}{k} > b_i > 0$ for all $k = 1, 2, \ldots$ then yields $w_2 > 0$ as the limit. Similarly, if $w_2 < 0$, choosing $b_k := b_i - \frac{1}{k + k_0}$ for all $k = 1, 2, \ldots$ with $k_0 > \frac{1}{b_i}$ yields $-w_2 > 0$ as the limit. These two statements contradict that the limit superior be nonpositive. Therefore w_2 must vanish (the limit in this case is 0).

$a_i > 0, b_i = 0$. Similar arguments as for the case above yield that

$$\hat{N}((a_i, b_i), \mathcal{C}_i) = \{0\} \times \mathbb{R}.$$

$a_i = 0, b_i = 0$. In this final case, it holds that $(a_i, b_i) \in \mathcal{C}_i^1 \cup \mathcal{C}_i^2$. Invoking Lemma 5.6 yields that

$$\begin{aligned} \hat{N}((a_i, b_i), \mathcal{C}_i) &= \hat{N}((a_i, b_i), \mathcal{C}_i^1) \cap \hat{N}((a_i, b_i), \mathcal{C}_i^2) \\ &= \{(w_1, w_2) \in \mathbb{R}^2 \mid \limsup_{b_k \searrow 0} w_2 \frac{b_k}{|b_k|} \leq 0\} \cap \\ &\qquad \{(w_1, w_2) \in \mathbb{R}^2 \mid \limsup_{a_k \searrow 0} w_1 \frac{a_k}{|a_k|} \leq 0\} \end{aligned}$$

$$
\begin{aligned}
&= \{(w_1,w_2) \in \mathbb{R}^2 \mid w_2 \le 0\} \cap \{(w_1,w_2) \in \mathbb{R}^2 \mid w_1 \le 0\} \\
&= (\mathbb{R} \times \mathbb{R}_-) \cap (\mathbb{R}_- \times \mathbb{R}) \\
&= \mathbb{R}^2_-.
\end{aligned}
$$

This completes the proof. □

We are finally in a position to state and prove the following proposition, which explicitly determines the limiting normal cone to a multivariate complementarity set. This result will become of vital importance in Sections 5.2 and 5.3.

Proposition 5.8 *Let the set*

$$\mathcal{C} := \{(a,b) \in \mathbb{R}^{2l} \mid a \ge 0,\ b \ge 0\ a^T b = 0\} \tag{5.16}$$

be given. Then, for an arbitrary $(a,b) \in \mathcal{C}$, *define*

$$\mathcal{I}_a = \{i \mid a_i = 0, b_i > 0\}, \quad \mathcal{I}_b = \{i \mid a_i > 0, b_i = 0\}, \quad \mathcal{I}_{ab} = \{i \mid a_i = 0, b_i = 0\}.$$

Then the limiting normal cone to $\mathcal{C}$ *in* (a,b) *is given by*

$$N((a,b),\mathcal{C}) = \{(x,y) \in \mathbb{R}^{2l} \mid x_{\mathcal{I}_b} = 0, y_{\mathcal{I}_a} = 0, (x_i < 0 \wedge y_i < 0) \vee x_i y_i = 0\ \forall i \in \mathcal{I}_{ab}\}. \tag{5.17}$$

Proof. We can map the set $\mathcal{C}$ isomorphically to the Cartesian product

$$\tilde{\mathcal{C}} := \mathcal{C}_1 \times \cdots \times \mathcal{C}_l$$

with

$$\mathcal{C}_i := \{(a_i,b_i) \in \mathbb{R}^2 \mid a_i \ge 0, b_i \ge 0, a_i b_i = 0\}$$

for $i = 1,\dots,l$, by simply rearranging the components of $\mathcal{C}$ in an appropriate fashion.

We now consider a single $\mathcal{C}_i$ for a fixed $i \in \{1,\dots,l\}$. As in the proof of Lemma 5.7, $\mathcal{C}_i$ can be rewritten as $\mathcal{C}_i = \mathcal{C}_i^1 \cup \mathcal{C}_i^2$ with

$$\mathcal{C}_i^1 := \{0\} \times \mathbb{R}_+ \qquad \text{and} \qquad \mathcal{C}_i^2 := \mathbb{R}_+ \times \{0\}.$$

Again, we consider three cases.

$i \in \mathcal{I}_a$**.** Since $b_i > 0$, we may replace $\mathcal{C}_i$ with $\mathcal{C}_i^1$ in the definition of the limiting normal cone in the following manner:

$$
\begin{aligned}
N((a_i,b_i),\mathcal{C}_i) &= \{\lim_{k\to\infty} w^k \mid \exists\{v^k\} \subseteq \mathcal{C}_i : \lim_{k\to\infty} v^k = (a_i,b_i), w^k \in \hat{N}(v^k,\mathcal{C}_i)\} \\
&= \{\lim_{k\to\infty} w^k \mid \exists\{v^k\} \subseteq \mathcal{C}_i^1 : \lim_{k\to\infty} v^k = (a_i,b_i), w^k \in \hat{N}(v^k,\mathcal{C}_i)\} \\
&= \{\lim_{k\to\infty} w^k \mid \exists\{b_k\}, b_k > 0 : \lim_{k\to\infty} b_k = b_i, w^k \in \hat{N}((0,b_k),\mathcal{C}_i)\}.
\end{aligned}
$$

Since $b_k > 0$ for all $k = 1,2,\dots$, invoking Lemma 5.7 yields that $\hat{N}((0,b_k),\mathcal{C}_i) = \mathbb{R} \times \{0\}$ for all $k = 1,2,\dots$. Hence, we obtain that

$$N((a_i,b_i),\mathcal{C}_i) = \mathbb{R} \times \{0\}. \tag{5.18}$$

$i \in \mathcal{I}_b$**.** Similar to the case $i \in \mathcal{I}_a$, it can be shown that

$$N((a_i, b_i), \mathcal{C}_i) = \{0\} \times \mathbb{R}. \tag{5.19}$$

$i \in \mathcal{I}_{ab}$**.** As evident from Lemma 5.7, it holds that

$$N((0,0), \mathcal{C}_i) \subseteq (\mathbb{R} \times \{0\}) \cup (\{0\} \times \mathbb{R}) \cup \mathbb{R}^2_-, \tag{5.20}$$

since the Fréchet normal cone in the expression for the limiting normal cone (see (5.6)) cannot assume any other values. We will show that equality holds in (5.20) by constructing appropriate sequences.

Setting $v^k := (0, \frac{1}{k}) \in \mathcal{C}_i$ for all $k = 1, 2, \ldots$, we obtain that $\lim_{k\to\infty} \hat{N}((0, \frac{1}{k}), \mathcal{C}_i) = \mathbb{R} \times \{0\} \subseteq N((0,0), \mathcal{C}_i)$. Similarly, choosing $v^k := (\frac{1}{k}, 0) \in \mathcal{C}_i$ for all $k = 1, 2, \ldots$ yields that $\{0\} \times \mathbb{R} \subseteq N((0,0), \mathcal{C}_i)$. Finally, setting $\{v^k\}$ equal to the nullsequence, we obtain $\mathbb{R}^2_- \subseteq N((0,0), \mathcal{C}_i)$. With this, we have shown that equality holds in (5.20):

$$N((0,0), \mathcal{C}_i) = (\mathbb{R} \times \{0\}) \cup (\{0\} \times \mathbb{R}) \cup \mathbb{R}^2_-. \tag{5.21}$$

Now, since the sets $\mathcal{C}_i$ are obviously closed, we can apply Proposition 5.4 to obtain

$$N((a_1, b_1, \ldots, a_l, b_l), \tilde{\mathcal{C}}) = N((a_1, b_1), \mathcal{C}_1) \times \cdots \times N((a_l, b_l), \mathcal{C}_l).$$

Finally, we acquire (5.17) from (5.18), (5.19), and (5.21) by applying the appropriate inverse mapping to acquire $N((a, b), \mathcal{C})$ from $N((a_1, b_1, \ldots, a_l, b_l), \tilde{\mathcal{C}})$. This completes the proof. □

The following result is a first indication of the possibility of bridging the gap between the Mordukhovich calculus and MPEC-GCQ, since it puts a normal cone and the tangent cone in relation.

Proposition 5.9 *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ denote the feasible set of the MPEC* (1.1)*, let $\mathcal{Z}$ be non-empty, and let $z^* \in \mathcal{Z}$ be given. Then it holds that*

$$\hat{N}(z^*, \mathcal{Z}) = -\mathcal{T}(z^*)^*. \tag{5.22}$$

Proof. We first observe that $\mathcal{Z}$ is closed, and hence the Fréchet normal $\hat{N}(z^*, \mathcal{Z})$ is well defined. The statement of the proposition may then be found in [61, Proposition 6.5]. □

For more information on normal cones, the interested reader is referred to the works of Mordukhovich [46], Clarke [8], and Rockafellar and Wets [61].

Subgradients

Just as the Clarke subgradient can be defined in terms of the Clarke normal cone [8, Corollary to Theorem 2.4.9], alternate subgradients can be defined using the normal cones introduced in Definition 5.1.

Definition 5.10 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous. Then the* Fréchet subgradient *of f in $z \in \mathbb{R}^n$ is given by*

$$\hat{\partial} f(z) := \{\xi \mid (\xi, -1) \in \hat{N}((z, f(z)), \operatorname{epi} f)\}, \tag{5.23}$$

the proximal subgradient *of f in $z \in \mathbb{R}^n$ by*

$$\partial^\pi f(z) := \{\xi \mid (\xi, -1) \in N^\pi((z, f(z)), \operatorname{epi} f)\}, \tag{5.24}$$

the limiting subgradient *of f in $z \in \mathbb{R}^n$ by*

$$\partial f(z) := \{\xi \mid (\xi, -1) \in N((z, f(z)), \operatorname{epi} f)\}, \tag{5.25}$$

and the Clarke subgradient *of f in $z \in \mathbb{R}^n$ by*

$$\partial^{\mathrm{Cl}} f(z) := \{\xi \mid (\xi, -1) \in N^{\mathrm{Cl}}((z, f(z)), \operatorname{epi} f)\}. \tag{5.26}$$

If f is convex, the convex subgradient *of f in $z \in \mathbb{R}^n$ is given by*

$$\partial^{\mathrm{conv}} f(z) := \{\xi \mid (\xi, -1) \in N^{\mathrm{conv}}((z, f(z)), \operatorname{epi} f)\}. \tag{5.27}$$

Here, epi *denotes the epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$:*

$$\operatorname{epi} f := \{(z, \zeta) \in \mathbb{R}^{n+1} \mid \zeta \geq f(z)\}.$$

Remark. It immediately follows from (5.8) and (5.9) and the definition of the various subgradients that the inclusions

$$\partial^\pi f(z) \subseteq \partial f(z) \subseteq \partial^{\mathrm{Cl}} f(z) \tag{5.28}$$

and

$$\hat{\partial} f(z) \subseteq \partial f(z) \tag{5.29}$$

hold for all $f : \mathbb{R}^n \to \mathbb{R}$ locally Lipschitz continuous in $z \in \mathbb{R}^n$. If f is convex (and, therefore, locally Lipschitz continuous) in $z \in \mathbb{R}^n$, it follows from Proposition 5.3 that

$$\hat{\partial} f(z) = \partial^\pi f(z) = \partial f(z) = \partial^{\mathrm{Cl}} f(z) = \partial^{\mathrm{conv}} f(z). \tag{5.30}$$

The following proposition collects some useful properties and calculus rules of the limiting subgradient.

Proposition 5.11 *(a) Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous in z and $\alpha \geq 0$ be given. Then it holds that*

$$\partial(\alpha f)(z) = \alpha \partial f(z). \tag{5.31}$$

(b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable in z. Then the limiting subgradient reduces to a singleton:

$$\partial f(z) = \{\nabla f(z)\}. \tag{5.32}$$

(c) Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous in z. Then the following sum rule holds:

$$\partial(f + g)(z) \subseteq \partial f(z) + \partial g(z). \tag{5.33}$$

(d) Let $F : \mathbb{R}^n \to \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ be locally Lipschitz continuous in z and $F(z)$, respectively. Then the following chain rule holds for $\Phi : \mathbb{R}^n \to \mathbb{R}$ defined by $\Phi(z) := (f \circ F)(z)$:

$$\partial\Phi(z) \subseteq \{\partial(\eta F)(z) \mid \eta \in \partial f(\zeta)\big|_{\zeta=F(z)}\}. \tag{5.34}$$

Proof. The nonnegative scalar multiplication property (a) is easily verified, while (b)–(d) may be found, e.g., in their respective order, in [40, Remark 4C.3], [40, Proposition 5A.4], and [40, Theorem 5A.8]. □

For more detail on the calculus of the limiting subgradient, the interested reader is referred to the works of Loewen [40] and Mordukhovich [47]. Also of interest in this context is a comparison to the closely related calculus for the Clarke subgradient, see [8, Section 2.3].

Lipschitz Properties of Multifunctions

We will require some properties and results pertaining to multifunctions in our analysis. We dedicate the remainder of this section to introducing these.

To begin, we need the concept of calmness of multifunctions and the closely related idea of upper Lipschitz continuity.

Definition 5.12 *Let $\Phi : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ be a multifunction with a closed graph, and let $(u, v) \in \operatorname{gph} \Phi$ be given. We say that Φ is* calm *at (u, v) if there exist neighborhoods $\mathcal{U}$ of u, $\mathcal{V}$ of v, and a modulus $L \geq 0$ such that*

$$\Phi(\tilde{u}) \cap \mathcal{V} \subseteq \Phi(u) + B_{L\|\tilde{u}-u\|}(0) \qquad \forall \tilde{u} \in \mathcal{U}. \tag{5.35}$$

If (5.35) *holds with $\mathcal{V} = \mathbb{R}^q$, Φ is said to be* locally upper Lipschitz continuous at u *[59].*

Note that calmness is sometimes also referred to as *pseudo upper Lipschitz continuity* (see, e.g., [79]) and that in [61] both calmness and local upper Lipschitz continuity are referred to as calmness.

A concept closely related to calmness and local upper Lipschitz continuity is the *Aubin property*, also called *pseudo Lipschitz continuity* [79]. Since we do not need it in our analysis, we refer the interested reader to [61].

If Φ is a single valued function, the notion of local upper Lipschitz continuity reduces to the condition that

$$\|\Phi(\tilde{u}) - \Phi(u)\| \leq L\|\tilde{u} - u\|$$

for all $\tilde{u} \in \mathcal{U}$. This is a one-sided Lipschitz continuity condition, explaining the name.

Some readers may be familiar with calmness in conjunction with nonlinear programming, see, e.g., [8, Definition 6.4.1]. There exists a connection between this calmness and calmness of a specific multifunction. Although we will not go into this relationship here, in Section 5.2 we will investigate the relationship between calmness of a specific multifunction (see (5.43)) and an MPEC variant of nonlinear programming calmness (cf. Definition 5.18). See, in particular, Proposition 5.22 and Lemma 5.23 in connection with this.

Next, we introduce a special class of multifunction, which will occur repeatedly in our analysis.

Definition 5.13 *We say that a multifunction $\Phi : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ is a* polyhedral multifunction *if its graph is the union of finitely many polyhedral convex sets.*

A convenient property of polyhedral multifunctions is that they are upper Lipschitz continuous everywhere. This result is due to Robinson [59, Proposition 1] and is stated in the following proposition.

Proposition 5.14 *Let $\Phi : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ be a polyhedral multifunction. Then Φ is locally upper Lipschitz continuous at every $u \in \mathbb{R}^p$.*

5.2 An Exact Penalization Approach

In this section, we present the path described by Jane Ye [76, 78] to obtain M-stationarity as a necessary first order condition under MPEC-GCQ. Originally, this approach was carried through for OPVICs rather than MPECs [76], while the result was stated under the stronger MPEC-ACQ [78].

Unfortunately, there is a gap in Ye's approach. The proof of [76, Theorem 3.2] is erroneous (as the author agrees [77]), which invalidates [78, Theorem 3.1] since it relies on that result.

Several alternate versions of [76, Theorem 3.2] have been suggested by Ye herself as well as the author. Among these are [73, Corollary 4.2], [81, Theorem 4.12] (note that there is a sign error in this reference), [79, Theorem 3.2 (c)], [80, Theorem 1.3], [46, Corollary 7.5.1], and [48, Theorem 4.4]. Any of these results may be suitably adapted to fill the gap in the original proof.

We chose to use the result by Treiman [73, Corollary 4.2] to replace [76, Theorem 3.2]. This has two predominant advantages. For one, the Treiman result is in a format more compatible with our MPEC (1.1), as opposed to an OPVIC. Secondly, we avoid the Mordukhovich coderivative. Both of these facts facilitate the notation of our proofs considerably, in particular that of Theorem 5.16. For some background on coderivatives in general, see [61, Chapter 8.G], and we refer to [47] for the Mordukhovich coderivative in particular.

In the course of investigating Ye's approach, we transfer several definitions and results introduced for OPVICs [76] to the MPEC setting. Among these are MPEC variants of calmness and a local error bound, under which we show M-stationarity to be a necessary first order condition. We then proceed to show that an MPAEC (see (3.22)) satisfies these constraint qualifications, finally using the idea of [78, Theorem 3.1] to prove our result.

We now state a theorem by Treiman [73, Corollary 4.2] which we will use to prove a Fritz John-type condition in Theorem 5.16. This is one of two instances (the other being Lemma 5.26) where it is important to know that the limiting normal cone may be expressed as the limit of the proximal normal cone, since [73] relies on that definition.

Theorem 5.15 *Let the program*

$$\begin{array}{rl} \min & f(z) \\ s.t. & g(z) \leq 0, \quad h(z) = 0, \\ & z \in U \end{array} \tag{5.36}$$

be given, where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^p$ are locally Lipschitz continuous functions and $U \subseteq \mathbb{R}^n$ is closed. Furthermore, let z^* be a local minimizer of (5.36). *Then there exist $r \geq 0$, $\lambda^g \geq 0$, and λ^h, not all zero, such that*

$$\begin{aligned} &0 \in r\partial f(z^*) + \sum_{i=1}^{m} \lambda_i^g \partial g(z^*) + \sum_{i=1}^{p} \partial(\lambda_i^h h_i)(z^*) + N(z^*, U), \\ &g(z^*)^T \lambda^g = 0. \end{aligned} \tag{5.37}$$

We will now apply Theorem 5.15 to our problem (1.1). Note, however, that we use a slightly weaker smoothness assumption on f since this result will later be applied to a specific MPEC whose objective function is locally Lipschitz continuous but not differentiable in general.

Theorem 5.16 *(Fritz John-type M-stationarity condition)*
Let z^* be a local minimizer of the MPEC (1.1), *where the objective function f is locally Lipschitz continuous and all other functions are continuously differentiable. Then there exists $r \geq 0$, $\lambda^g \in \mathbb{R}^m$, $\lambda^h \in \mathbb{R}^p$, $\lambda^G \in \mathbb{R}^l$, $\lambda^H \in \mathbb{R}^l$, not all zero, such that*

$$\begin{aligned} &0 \in r\partial f(z^*) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(z^*) - \sum_{i=1}^{l} \left[\lambda_i^G \nabla G_i(z^*) + \lambda_i^H \nabla H_i(z^*)\right], \\ &\begin{array}{lll} \lambda_\alpha^G \ \text{free}, & & \lambda_\gamma^G = 0, \\ & (\lambda_i^G > 0 \wedge \lambda_i^H > 0) \ \vee \ \lambda_i^G \lambda_i^H = 0 \quad \forall i \in \beta & \\ \lambda_\gamma^H \ \text{free}, & & \lambda_\alpha^H = 0, \end{array} \\ &g(z^*) \leq 0, \qquad \lambda^g \geq 0, \qquad g(z^*)^T \lambda^g = 0. \end{aligned} \tag{5.38}$$

Proof. We will prove this result by applying Theorem 5.15 to our MPEC (1.1). In order to facilitate the proof, we introduce slack variables, ξ and η, in the following equivalent reformulation of the MPEC (1.1):

$$\begin{array}{rl} \min\limits_{(z,\xi,\eta)} & f(z) \\ \text{s.t.} & g(z) \le 0, \quad h(z) = 0, \\ & \Gamma(z,\xi,\eta) := \begin{pmatrix} G(z) - \xi \\ H(z) - \eta \end{pmatrix} = 0, \\ & (z,\xi,\eta) \in \mathbb{R}^n \times \mathcal{C} \end{array} \tag{5.39}$$

with

$$\mathcal{C} := \{(\xi,\eta) \in \mathbb{R}^{2l} \mid \xi \ge 0,\, \eta \ge 0,\, \xi^T\eta = 0\}.$$

We now apply Theorem 5.15 to the program (5.39). This is possible because, in particular, all functions involved are locally Lipschitz continuous and $\mathbb{R}^n \times \mathcal{C}$ is obviously closed.

Setting $\xi^* := G(z^*)$ and $\eta^* := H(z^*)$, Theorem 5.15 states that there exist $r \ge 0$, $\lambda^g \ge 0$, λ^h, and $\lambda^\Gamma =: (-\lambda^G, -\lambda^H)$, not all zero, such that

$$\begin{aligned} 0 \in & \begin{pmatrix} r\partial f(z^*) \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^m \lambda_i^g \partial g_i(z^*) \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^p \partial(\lambda_i^h h_i)(z^*) \\ 0 \\ 0 \end{pmatrix} \\ & + \sum_{i=1}^{2l} \partial(\lambda_i^\Gamma \Gamma_i)(z^*,\xi^*,\eta^*) + N\big((z^*,\xi^*,\eta^*), \mathbb{R}^n \times \mathcal{C}\big) \\ = & \begin{pmatrix} r\partial f(z^*) \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^m \lambda_i^g \nabla g_i(z^*) \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^p \lambda_i^h \nabla h_i(z^*) \\ 0 \\ 0 \end{pmatrix} \\ & - \begin{pmatrix} \sum_{i=1}^l \big(\lambda_i^G \nabla G_i(z^*) + \lambda_i^H \nabla H_i(z^*)\big) \\ -\lambda^G \\ -\lambda^H \end{pmatrix} + \begin{pmatrix} 0 \\ N\big((\xi^*,\eta^*),\mathcal{C}\big) \end{pmatrix}, \end{aligned} \tag{5.40}$$

where we used that all functions but f are continuously differentiable, as well as Proposition 5.4.

We now take a closer look at those components pertaining to ξ and η in (5.40),

$$(-\lambda^G, -\lambda^H) \in N\big((\xi^*,\eta^*),\mathcal{C}\big). \tag{5.41}$$

Invoking Proposition 5.8, we obtain the following rules for the components of λ^G and λ^H:

$$(\lambda_i^G, \lambda_i^H) \in \begin{cases} \{(a,b) \mid a \text{ free}, b = 0\} & : \ \xi_i^* = 0, \eta_i^* > 0, \\ \{(a,b) \mid (a > 0 \wedge b > 0) \vee ab = 0\} & : \ \xi_i^* = 0, \eta_i^* = 0, \\ \{(a,b) \mid a = 0, b \text{ free}\} & : \ \xi_i^* > 0, \eta_i^* = 0. \end{cases} \tag{5.42}$$

The statement of this theorem is thus obtained if we take into account that $\xi_i^* = G_i(z^*)$, $\eta_i^* = H_i(z^*)$, the definitions of α, β, and γ (see (1.2)), (5.42), (5.40), the final condition in (5.37), and the fact that (z^*,ξ^*,η^*) is feasible for the reformulated MPEC (5.39). □

Note that a version of Theorem 5.16 employing stronger smoothness assumptions appears in [78, Theorem 2.1].

Once more, as in the proof of Theorem 4.12, we have Fritz John-type conditions, see (5.38). The difference is in the restrictions imposed upon the Lagrange multipliers λ_β^G and λ_β^H (compare (5.38) to (4.19)). As in the proof of Theorem 4.12, we may use MPEC-MFCQ to guarantee that $r = 1$ in (5.38). This is stated in the following Theorem.

Theorem 5.17 *Let $z^* \in \mathbb{R}^n$ be a local minimizer of the MPEC* (1.1). *If MPEC-MFCQ holds in z^*, then z^* is M-stationary.*

Proof. This follows immediately from Theorem 5.16 by the same arguments as employed in the proof of Theorem 4.12. Note that, since f is continuously differentiable, it holds that $\partial f(z^*) = \{\nabla f(z^*)\}$ by Proposition 5.11 (b). □

In standard nonlinear programming, *calmness* (not be confused with calmness of a multifunction) is often used as a constraint qualification in order to guarantee that $r > 0$ in a Fritz John-setting, yielding a KKT point as a necessary first order condition. Applying this reasoning to MPECs yields a KKT point (which is equivalent to a strongly stationary point, see Proposition 4.5) under calmness.

However, we know that any constraint qualification weaker than MPEC-MFCQ cannot imply strong stationarity to be a first order condition under it (see Example 4.13). Hence standard calmness is too strong for an investigation of MPEC-GCQ. Therefore we will define an MPEC variant of calmness.

To this end, we introduce a multifunction $\mathcal{Z} : \mathbb{R}^{m+p+2l} \rightrightarrows \mathbb{R}^n$ to facilitate the statements and proofs of the results to follow:

$$\begin{aligned}\mathcal{Z}(p,q,r,s) := \{z \in \mathbb{R}^n \mid\ & g(z) + p \le 0,\ h(z) + q = 0,\\ & G(z) + r \ge 0,\ H(z) + s \ge 0,\\ & (G(z)+r)^T(H(z)+s) = 0\}.\end{aligned} \tag{5.43}$$

Note that besides denoting the multifunction by $\mathcal{Z}$, we also denote the feasible set of the MPEC (1.1) by $\mathcal{Z} = \mathcal{Z}(0,0,0,0)$. This ambiguity in the notation was chosen deliberately because it provides an elegant connection between the multifunction $\mathcal{Z}$ and the feasible set of the MPEC.

We now use this set to define an MPEC variant of calmness, inspired by the definition of calmness for OPVICs, as proposed by Ye [76]. Note the close resemblance to standard calmness (see, e.g., [8, Definition 6.4.1]).

Definition 5.18 *The MPEC* (1.1) *is said to satisfy* MPEC-calmness in z^* *(or, alternatively, the MPEC is* MPEC-calm at z^**) if there exist $\varepsilon > 0$ and $\mu > 0$ such that for all $(p,q,r,s) \in B_\varepsilon(0) \subset \mathbb{R}^{m+p+2l}$ and all $z \in \mathcal{Z}(p,q,r,s) \cap B_\varepsilon(z^*)$ it holds that*

$$f(z^*) \le f(z) + \mu \left\|(p,q,r,s)\right\|. \tag{5.44}$$

Since all norms are equivalent in a finite-dimensional setting, MPEC-calmness (just as standard calmness) is independent of the norm used.

Note that if MPEC-calmness holds at z^*, the point z^* is a local minimizer of the MPEC (1.1) as can be seen by considering (5.44) with $(p, q, r, s) = (0, 0, 0, 0)$.

Further note that a straightforward application of OPVIC-calmness [76] yields a slightly degenerated version of the above, where s (or r) is omitted.

Also of interest is that, once more, we treated the factors of the complementarity term independently, not the complementarity term as a whole (as would be the case for standard calmness, see also Section 6.2). This is reminiscent of the way the MPEC-linearized tangent cone $\mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)$ is defined (see the discussion in Section 3.2).

Just as in standard nonlinear programming, MPEC-calmness is closely linked to exact penalization (see, e.g., [6]), which we will examine in the following proposition. See also Proposition 6.1 in this context.

Proposition 5.19 *Let z^* be a local minimizer of the MPEC* (1.1). *Then the following are equivalent:*

(a) the MPEC is MPEC-calm at z^*;

(b) there exists a $\rho_0 > 0$ such that for all $\rho \geq \rho_0$ the vector $(z^*, 0, 0) \in \mathbb{R}^{n+2l}$ is a local minimizer of

$$\begin{array}{rl} \min\limits_{(z,r,s)} & f(z) + \rho(\|\max\{0, g(z)\}\| + \|h(z)\| + \|(r, s)\|) \\ s.t. & G(z) + r \geq 0, \quad H(z) + s \geq 0, \\ & (G(z) + r)^T (H(z) + s) = 0. \end{array} \tag{5.45}$$

Proof. (a) $\Rightarrow$ (b). Let the conditions for MPEC-calmness at z^* from Definition 5.18 be satisfied. Choose $\delta \leq \frac{\varepsilon}{3}$ such that

$$\max_{z \in B_\delta(z^*)} \|\max\{0, g(z)\}\| \leq \frac{\varepsilon}{3},$$
$$\max_{z \in B_\delta(z^*)} \|h(z)\| \leq \frac{\varepsilon}{3}.$$

This exists due to the continuity of h and g. Note that $\delta > 0$.

Furthermore, let an arbitrary $(z, r, s) \in B_\delta(z^*, 0, 0)$ such that (z, r, s) is feasible for the program (5.45) be given. Then set

$$p_i := -\max\{0, g_i(z)\}, \ \forall i = 1, \dots, m \qquad \text{and} \qquad q := h(z).$$

With this choice of p and q, it holds that

$$\|(p, q, r, s)\| \leq \|p\| + \|q\| + \|(r, s)\| \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

and hence $(p, q, r, s) \in B_\varepsilon(0)$, as well as $z \in \mathcal{Z}(p, q, r, s) \cap B_\varepsilon(z^*)$.

Since z^* satisfies MPEC-calmness, condition (5.44) holds and we obtain

$$\begin{aligned} f(z^*) &\leq f(z) + \mu \, \|(p,q,r,s)\| \\ &\leq f(z) + \mu\big(\, \|p\| + \|q\| + \|(r,s)\| \, \big) \\ &= f(z) + \mu(\|\max\{0, g(z)\}\| + \|h(z)\| + \|(r,s)\|) \\ &\leq f(z) + \rho(\|\max\{0, g(z)\}\| + \|h(z)\| + \|(r,s)\|), \end{aligned} \tag{5.46}$$

for all $\rho \geq \mu > 0$. Together with the fact that (z, r, s) was chosen arbitrarily from a neighborhood of $(z^*, 0, 0)$ and satisfying the constraints of (5.45), this implies that $(z^*, 0, 0)$ is a local minimizer of the program (5.45) for all $\rho \geq \rho_0 := \mu$.

(b) $\Rightarrow$ (a). The following arguments apply to the Euclidian norm $\|\cdot\| = \|\cdot\|_2$. The extension to an arbitrary norm is trivial since all norms are equivalent in finite dimensions. Only $\tilde{\mu}$ and μ below would be adjusted.

Let ε be given such that $(z^*, 0, 0)$ is a global minimizer of (5.45) with $\rho = \rho_0$ in the ball $B_{2\varepsilon}(z^*, 0, 0)$ (note the radius is 2ε).

Now, let $((p, q, r, s), z) \in \operatorname{gph} \mathcal{Z}$ be chosen arbitrarily such that $z \in B_\varepsilon(z^*)$ as well as $(p, q, r, s) \in B_\varepsilon(0)$. In this case, the vector (z, r, s) is, in particular, feasible for (5.45) and within the region where $(z^*, 0, 0)$ is a global minimizer. Therefore, it holds for $\rho = \rho_0$ that

$$\begin{aligned} f(z^*) &\leq f(z) + \rho_0(\|\max\{0, g(z)\}\| + \|h(z)\| + \|(r,s)\|) \\ &\leq f(z) + \tilde{\mu}(\|\max\{0, g(z)\}\|_1 + \|h(z)\|_1 + \|(r,s)\|_1) \\ &= f(z) + \tilde{\mu}(\|\max\{0, g(z) + p - p\}\|_1 + \|h(z) + q - q\|_1 + \|(r,s)\|_1) \\ &\leq f(z) + \tilde{\mu}(\| \max\{0, \underbrace{g(z) + p}_{\leq 0}\}\|_1 + \|p\|_1 + \| \underbrace{h(z) + q}_{=0} \|_1 + \|q\|_1 + \|(r,s)\|_1) \\ &= f(z) + \tilde{\mu}(\|p\|_1 + \|q\|_1 + \|(r,s)\|_1) \\ &= f(z) + \tilde{\mu} \, \|(p,q,r,s)\|_1 \\ &\leq f(z) + \mu \, \|(p,q,r,s)\| \, , \end{aligned}$$

for some $\tilde{\mu}, \mu > 0$, due to the equivalence of norms. This is exactly the definition of MPEC-calmness. $\square$

We can now use this proposition to prove that M-stationarity is a necessary first order condition under MPEC-calmness.

Theorem 5.20 *Let z^* be a local minimizer of the MPEC* (1.1) *at which the MPEC is MPEC-calm. Then there exists a Lagrange multiplier $\lambda^* := (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ such that*

$$\begin{aligned} &0 = \nabla f(z^*) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(z^*) - \sum_{i=1}^{l} \big[\lambda_i^G \nabla G_i(z^*) + \lambda_i^H \nabla H_i(z^*)\big], \\ &\begin{matrix} \lambda_\alpha^G \ \ \textit{free}, \\ \lambda_\gamma^H \ \ \textit{free}, \end{matrix} \qquad (\lambda_i^G > 0 \wedge \lambda_i^H > 0) \ \vee \ \lambda_i^G \lambda_i^H = 0 \quad \forall i \in \beta \qquad \begin{matrix} \lambda_\gamma^G = 0, \\ \lambda_\alpha^H = 0, \end{matrix} \\ &g(z^*) \leq 0, \qquad \lambda^g \geq 0, \qquad g(z^*)^T \lambda^g = 0, \end{aligned} \tag{5.47}$$

i.e. z^* *is M-stationary.*

Proof. The idea and proof of this theorem is based on the same principles as [76, Theorem 3.6].

Since the MPEC (1.1) satisfies MPEC-calmness at z^*, $(z^*, 0, 0)$ is a local minimizer of (5.45) according to Proposition 5.19 for $\rho = \rho_0$.

Interpreting (5.45) as an MPEC with complementarity constraints only, we apply Theorem 5.16 to acquire the existence of $(r, \lambda^G, \lambda^H) \neq (0, 0, 0)$ such that $r \geq 0$, λ^G, and λ^H satisfy the appropriate conditions in (5.38), and

$$0 \in r\partial\tilde{f}(z^*, 0, 0) - \sum_{i=1}^{l} \big[\lambda_i^G \begin{pmatrix} \nabla G_i(z^*) \\ e_i \\ 0 \end{pmatrix} + \lambda_i^H \begin{pmatrix} \nabla H_i(z^*) \\ 0 \\ e_i \end{pmatrix}\big], \tag{5.48}$$

where $\tilde{f}(z, r, s) := f(z) + \rho_0(\|\max\{0, g(z)\}\|_1 + \|h(z)\|_1 + \|(r, s)\|_1)$, and $e_i \in \mathbb{R}^l$ is the i-th unit vector. Note that we choose the 1-norm specifically. Also note that Theorem 5.16 may be applied since $\tilde{f}$ is locally Lipschitz continuous.

If we assume $r = 0$, it follows immediately from (5.48) that $\lambda^G = \lambda^H = 0$, a contradiction to $(r, \lambda^G, \lambda^H) \neq (0, 0, 0)$. Therefore, $r = 1$ can be assumed without loss of generality.

Now, let us consider $\partial_z \tilde{f}(z^*, 0, 0)$. Using the properties for limiting subgradients from Proposition 5.11 (note that $\rho_0 \geq 0$), we obtain

$$\begin{aligned} \partial_z \tilde{f}(z^*, 0, 0) &\subseteq \nabla f(z^*) + \rho_0 \big[\sum_{i=1}^{m} \partial \max\{0, g_i(z^*)\} + \sum_{i=1}^{p} \partial|h_i(z^*)|\big] \\ &\subseteq \nabla f(z^*) + \rho_0 \big[\sum_{i=1}^{m} \{\eta_i^g \nabla g(z^*) \mid \eta_i^g \in \partial \max\{0, \zeta\}\big|_{\zeta = g_i(z^*)}\} \\ &\qquad + \sum_{i=1}^{p} \{\eta_i^h \nabla h(z^*) \mid \eta_i^h \in \partial|\zeta|\big|_{\zeta = h_i(z^*)}\}\big]. \end{aligned} \tag{5.49}$$

Both $\max\{0, \cdot\}$ and $|\cdot|$ are convex functions, and in view of (5.30), all subgradients coincide, and their limiting subgradients are given by [8, Propositions 2.3.12, 2.3.1, & 2.3.6 (a)].

Since z^* is, in particular, feasible, it follows that $h(z^*) = 0$ and $g(z^*) \leq 0$. Hence we state the limiting subgradients of $\max\{0, \cdot\}$ and $|\cdot|$ only for the relevant part of their domain:

$$\partial \max\{0, x\} = \begin{cases} 0 & : x < 0, \\ [0, 1] & : x = 0, \end{cases} \qquad \partial|x| = [-1, 1] \text{ for } x{=}0.$$

Incorporating this into (5.49) and substituting $\partial_z \tilde{f}(z^*, 0, 0)$ into (5.48) with $r = 1$ yields the first part of the conditions of M-stationarity (see (5.47)).

Setting $\lambda_i^g := 0$ for all $i \notin \mathcal{I}_g$ yields the remainder of the conditions (5.47) for M-stationarity, thereby completing the proof. □

In Theorem 5.20 we showed M-stationarity to be a necessary optimality condition under MPEC-calmness. Therefore, MPEC-calmness plays the role of a constraint qualification in this context. It has the slight caveat of not being stated independently of the objective function f, however.

To remedy this, standard nonlinear programming employs local error bounds, which are stated independently of the objective function f, and are shown to imply standard calmness.

Following this example, we now define an MPEC-variant of a local error bound, which we will show to imply MPEC-calmness. Again, the following definition is inspired by Ye [76, Definition 4.1].

Definition 5.21 *The constraint system of the MPEC* (1.1)*,*

$$\begin{array}{l} g(z) \leq 0, \quad h(z) = 0, \\ G(z) \geq 0, \quad H(z) \geq 0, \quad G(z)^T H(z) = 0, \end{array} \tag{5.50}$$

is said to have a local MPEC-error bound *at z^* if there exist $\nu > 0$ and $\delta > 0$ such that*

$$\operatorname{dist}(z, \mathcal{Z}) \leq \nu \left\| (p, q, r, s) \right\| \tag{5.51}$$

for all $(p, q, r, s) \in B_\delta(0)$ and all $z \in \mathcal{Z}(p, q, r, s) \cap B_\delta(z^*)$.

In the above definition, a local MPEC-error bound is a specific property of the *constraint system* (5.50) of the MPEC. However, to facilitate notation, we will also say that the *MPEC has a local MPEC-error bound at z^**.

We now show that a local MPEC-error bound implies MPEC-calmness. This is also inspired by an analogous result for OPVICs by Ye [76, Proposition 4.2].

Proposition 5.22 *Let z^* be a local minimizer of the MPEC* (1.1)*, and let the constraint system* (5.50) *of the MPEC have a local MPEC-error bound at z^*. Then the MPEC is MPEC-calm at z^*.*

Proof. Let $\tilde{\delta}$ be chosen such that z^* is a global minimizer of f in $\mathcal{Z} \cap B_{2\tilde{\delta}}(z^*)$. Furthermore, let δ be the quantity from the definition of the local MPEC-error bound (see Definition 5.21). Setting $\varepsilon := \min\{\tilde{\delta}, \delta\}$, we have that $(p, q, r, s) \in B_\varepsilon(0)$ and $z \in \mathcal{Z}(p, q, r, s) \cap B_\varepsilon(z^*)$ satisfy (5.51). By our choice of z and the fact that $z^* \in \mathcal{Z}$ is feasible, it holds that

$$\|\Pi_{\mathcal{Z}}(z) - z^*\| \leq \|\Pi_{\mathcal{Z}}(z) - z\| + \|z - z^*\| \leq 2\varepsilon.$$

Using this, and the facts that $2\varepsilon \leq 2\tilde{\delta}$ and f is continuously differentiable and hence locally Lipschitz continuous, it follows that

$$\begin{aligned} f(z^*) &\leq f(\Pi_{\mathcal{Z}}(z)) \\ &= f(z) + (f(\Pi_{\mathcal{Z}}(z)) - f(z)) \\ &\leq f(z) + L \left\|\Pi_{\mathcal{Z}}(z) - z\right\| \\ &= f(z) + L\operatorname{dist}(z, \mathcal{Z}) \\ &\overset{(5.51)}{\leq} f(z) + L\nu \left\|(p, q, r, s)\right\| . \end{aligned} \tag{5.52}$$

Setting $\mu := L\nu$ yields the conditions for MPEC-calmness at z^*. This concludes the proof.□

Next, we will show that an MPAEC (see (3.22)) has a local MPEC-error bound at every feasible point. This result will play an important role when we prove M-stationarity to be a necessary optimality condition under MPEC-GCQ in Theorem 5.25.

In order to facilitate the proof, we first investigate the relationship between the local MPEC-error bound and the multifunction $\mathcal{Z}$ in the following lemma. Again, this relationship is analogous to a result for OPVICs, see [76].

Lemma 5.23 *Let z^* be feasible for the MPEC* (1.1)*. Then the MPEC has a local MPEC-error bound at z^* if and only if the multifunction $\mathcal{Z}$ is calm at* $(0, z^*)$*.*

Proof. Let the MPEC have a local MPEC-error bound at z^*. It then follows from (5.51) that

$$z \in \mathcal{Z} + B_{\nu\|(p,q,r,s)\|}(0)$$

for all $(p, q, r, s) \in B_\delta(0)$ and $z \in \mathcal{Z}(p, q, r, s) \cap B_\delta(z^*)$. By setting $\mathcal{U} := B_\delta(0)$, $\mathcal{V} := B_\delta(z^*)$, and $L := \nu$, we acquire

$$\mathcal{Z}(p, q, r, s) \cap \mathcal{V} \subseteq \mathcal{Z}(0, 0, 0, 0) + B_{L\|(p,q,r,s)\|}(0)$$

for all $(p, q, r, s) \in \mathcal{U}$. These are the conditions for calmness of the multifunction $\mathcal{Z}$ at $(0, z^*)$.

For the converse implication, let the multifunction $\mathcal{Z}$ be calm at $(0, z^*)$. We set $\nu := L$ and choose δ such that $B_\delta(0) \subseteq \mathcal{U}$ and $B_\delta(z^*) \subseteq \mathcal{V}$. Then, for arbitrary $(p, q, r, s) \in B_\delta(0)$ and $z \in \mathcal{Z}(p, q, r, s) \cap B_\delta(z^*)$ it follows from (5.35) that

$$z \in \mathcal{Z} + B_{\nu\|(p,q,r,s)\|}(0),$$

which, in turn, implies that

$$\operatorname{dist}(z, \mathcal{Z}) \leq \nu \|(p, q, r, s)\| .$$

Since $(p, q, r, s) \in B_\delta(0)$ and $z \in \mathcal{Z}(p, q, r, s) \cap B_\delta(z^*)$ were chosen arbitrarily, the MPEC has a local MPEC-error bound at z^*. □

Lemma 5.23, together with Proposition 5.22, bridges the gap between the concepts of calmness of a multifunction and MPEC-calmness. We hinted at this connection in our discussion following the definition of calmness of a multifunction, Definition 5.12. A similar connection exists between calmness of a multifunction and standard nonlinear programming calmness.

The following result states that an MPAEC (3.22) has a local MPEC-error bound. This, though important by itself, will be instrumental when we prove M-stationarity to be a first order optimality condition under MPEC-GCQ in Theorem 5.25.

Proposition 5.24 *Let z^* be an arbitrary feasible point of the MPAEC* (3.22)*. Then the constraint system of the MPAEC has a local MPEC-error bound at z^*.*

Proof. For proofs employing the method we shall use, see [59, 43, 76]. Alternatively, this result may be obtained by following the techniques of the proof for [43, Theorem 2.3].

We will show that the multifunction $\mathcal{Z}$ is calm at $(0, z^*)$, which, by virtue of Lemma 5.23, is equivalent to the MPAEC having a local MPEC-error bound at z^*.

To this end, consider the graph of $\mathcal{Z}$:

$$
\begin{aligned}
\operatorname{gph} \mathcal{Z} &= \{(p,q,r,s,z) \mid z \in \mathcal{Z}(p,q,r,s)\} \\
&= \{(p,q,r,s,z) \mid \ Az + a + p \leq 0, \ Bz + b + q = 0, \\
&\qquad\qquad\qquad\qquad\ Cz + c + r \geq 0, \ Dz + d + s \geq 0, \\
&\qquad\qquad\qquad\qquad\ (Cz + c + r)^T (Dz + d + s) = 0 \,\} \\
&= \bigcup_{(\nu_1,\nu_2) \in \mathcal{P}(\{1,\ldots,l\})} \{(p,q,r,s,z) \mid Az + a + p \leq 0, \ Bz + b + q = 0, \\
&\qquad\qquad\qquad\qquad\ C_{\nu_1} z + c_{\nu_1} + r_{\nu_1} = 0, \ C_{\nu_2} z + c_{\nu_2} + r_{\nu_2} \geq 0, \\
&\qquad\qquad\qquad\qquad\ D_{\nu_1} z + d_{\nu_1} + s_{\nu_1} \geq 0, \ D_{\nu_2} z + d_{\nu_2} + s_{\nu_2} = 0 \,\}.
\end{aligned}
$$

Here, C_{ν_1} denotes the matrix made up of those rows of C which correspond to the index set ν_1. The matrices C_{ν_2}, D_{ν_1}, and D_{ν_2} are defined similarly.

Obviously, $\operatorname{gph} \mathcal{Z}$ is the union of finitely many polyhedral convex sets. Therefore, $\mathcal{Z}$ is a polyhedral multifunction, see Definition 5.13. By virtue of Proposition 5.14, $\mathcal{Z}$ is locally upper Lipschitz continuous for all $(p,q,r,s) \in \mathbb{R}^{m+p+2l}$.

In particular, $\mathcal{Z}$ is locally upper Lipschitz continuous at the origin. This implies calmness of $\mathcal{Z}$ at $(0, z^*)$ for arbitrary $z^* \in \mathcal{Z}(0,0,0,0)$ (see Definition 5.12). Therefore, by virtue of Lemma 5.23, the MPAEC has a local MPEC-error bound at z^*, completing the proof. $\square$

We are finally able to present our main result, the proof of which is inspired by a similar result under MPEC-ACQ by Ye [78, Theorem 3.1].

Theorem 5.25 *Let z^* be a local minimizer of the MPEC* (1.1) *at which MPEC-GCQ holds. Then there exists a Lagrange multiplier $\lambda^* = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ such that*

$$
\begin{gathered}
0 = \nabla f(z^*) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(z^*) - \sum_{i=1}^{l} \left[\lambda_i^G \nabla G_i(z^*) + \lambda_i^H \nabla H_i(z^*) \right], \\
\begin{array}{lll}
\lambda_\alpha^G \ \text{free}, & & \lambda_\gamma^G = 0, \\
 & (\lambda_i^G > 0 \wedge \lambda_i^H > 0) \ \vee \ \lambda_i^G \lambda_i^H = 0 \quad \forall i \in \beta & \\
\lambda_\gamma^H \ \text{free}, & & \lambda_\alpha^H = 0,
\end{array} \\
g(z^*) \leq 0, \qquad \lambda^g \geq 0, \qquad g(z^*)^T \lambda^g = 0,
\end{gathered}
\tag{5.53}
$$

i.e. z^* is M-stationary.

Proof. Since z^* is a local minimizer of (1.1), B-stationarity holds, i.e.

$$\nabla f(z^*) \in \mathcal{T}(z^*)^*,$$

and since MPEC-GCQ holds, this is equivalent to

$$\nabla f(z^*) \in \mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*)^*.$$

Using the definition of the dual cone (see Definition 2.6), this yields

$$\nabla f(z^*)^T d \geq 0 \qquad \forall d \in \mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*). \tag{5.54}$$

This, in turn, is equivalent to $d^* = 0$ being a local minimizer of

$$\begin{array}{ll} \min\limits_{d} & \nabla f(z^*)^T d \\ \text{s.t.} & d \in \mathcal{T}^{lin}_{\mathrm{MPEC}}(z^*), \end{array} \tag{5.55}$$

which is an MPAEC. By Proposition 5.24, the constraint system of (5.55) has a local MPEC-error bound, which in turn implies that (5.55) is MPEC-calm at $d^* = 0$ by Proposition 5.22.

Keeping in mind that in (5.55), d is the variable, and that $d^* = 0$ solves (5.55), Theorem 5.20 then yields the existence of Lagrange multipliers $\lambda^g_{\mathcal{I}_g}$, λ^h, $\lambda^G_{\alpha\cup\beta}$, and $\lambda^H_{\gamma\cup\beta}$ such that, in particular

$$\begin{aligned} 0 = \nabla f(z^*) &+ \sum_{i\in\mathcal{I}_g} \lambda^g_i \nabla g_i(z^*) + \sum_{i=1}^{p} \lambda^h_i \nabla h_i(z^*) - \sum_{i\in\alpha} \lambda^G_i \nabla G_i(z^*) - \sum_{i\in\gamma} \lambda^H_i \nabla H_i(z^*) \\ &- \sum_{i\in\beta} \left[\lambda^G_i \nabla G_i(z^*) + \lambda^H_i \nabla H_i(z^*)\right], \\ &\lambda^G_\alpha \text{ free}, \qquad \lambda^H_\gamma \text{ free}, \qquad \lambda^h \text{ free}, \qquad \lambda^g_{\mathcal{I}_g} \geq 0, \\ &(\lambda^G_i > 0 \wedge \lambda^H_i > 0) \ \vee \ \lambda^G_i \lambda^H_i = 0 \quad \forall i \in \beta. \end{aligned} \tag{5.56}$$

The latter condition follows from the fact that since $d^* = 0$, it holds that $\nabla G_i(z^*)^T d^* = \nabla H_i(z^*)^T d^* = 0$ for all $i \in \beta$, and hence the whole set β is degenerate in the corresponding constraints of (5.55). Also note that, since λ^G_α and λ^H_γ have no sign restriction imposed upon them, their signs were chosen in such a manner as to facilitate the notation of the proof.

By setting $\lambda^g_i := 0$ for all $i \notin \mathcal{I}_g$, $\lambda^G_\gamma := 0$, and $\lambda^H_\alpha := 0$ we obtain the conditions (5.53) for M-stationarity with $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ from for M-stationarity from (5.56). This concludes the proof. □

It is interesting to note that we use the M-stationarity conditions (5.56) for the program (5.55) to acquire M-stationarity conditions of our source problem (1.1). This fundamental idea is due to Ye, see [78, Theorem 3.1].

Another point of interest is the fact that an MPAEC implies both MPEC-ACQ and MPEC-calmness (see Theorem 3.15 and Propositions 5.24 and 5.22, respectively), and that both these paths lead to M-stationarity being a necessary optimality condition (see Theorem 5.25 and Theorem 5.20, respectively). See also Figure 5.1.

5.3 A Direct Proof

We have shown M-stationarity to be a necessary optimality condition under MPEC-GCQ in Theorem 5.25. The proof used auxiliary concepts such as MPEC-calmness and a local MPEC-error bound to prove that $r = 1$ in a Fritz John-type M-stationarity setting. Although this approach offers considerable insight into MPECs and will lead to a strong exact penalization result (see Proposition 6.1), it is somewhat cumbersome and lengthy.

Therefore, we present an alternate, more elementary and direct proof of Theorem 5.25 in this section. This approach does not rely on a Fritz John-type result such as Theorem 5.15. Instead, it solely relies on the ability to determine the limiting normal cone to a complementarity set (cf. Proposition 5.8).

We start off, however, identical to the proof of Theorem 5.25, with Jane Ye's idea of reformulating our problem as an MPAEC.

In the following, we consider a local minimizer z^* of our MPEC (1.1) at which MPEC-GCQ is satisfied. Then, by the arguments of the proof of Theorem 5.25, we have that $d^* = 0$ is a minimizer of

$$\begin{array}{ll} \min\limits_{d} & \nabla f(z^*)^T d \\ \text{s.t.} & d \in \mathcal{T}^{lin}_{\text{MPEC}}(z^*). \end{array} \tag{5.57}$$

It is easily verified that $d^* = 0$ being a minimizer of (5.57) is equivalent to $(d^*, \xi^*, \eta^*) = (0, 0, 0)$ being a minimizer of

$$\begin{array}{ll} \min\limits_{(d,\xi,\eta)} & \nabla f(z^*)^T d \\ \text{s.t.} & (d, \xi, \eta) \in \mathcal{D} := \mathcal{D}_1 \cap \mathcal{D}_2 \end{array} \tag{5.58}$$

with

$$\begin{array}{lll} \mathcal{D}_1 := \{(d, \xi, \eta) \mid & \nabla g_i(z^*)^T d \leq 0, & \forall i \in \mathcal{I}_g, \\ & \nabla h_i(z^*)^T d = 0, & \forall i = 1, \dots, p, \\ & \nabla G_i(z^*)^T d = 0, & \forall i \in \alpha, \\ & \nabla H_i(z^*)^T d = 0, & \forall i \in \gamma, \\ & \nabla G_i(z^*)^T d - \xi_i = 0, & \forall i \in \beta, \\ & \nabla H_i(z^*)^T d - \eta_i = 0, & \forall i \in \beta\,\} \end{array} \tag{5.59}$$

and

$$\mathcal{D}_2 := \{(d, \xi, \eta) \mid \xi \geq 0,\ \eta \geq 0,\ \xi^T \eta = 0\}. \tag{5.60}$$

Since $(0, 0, 0)$ is a minimizer of (5.58), B-stationarity holds, i.e.

$$(\nabla f(z^*), 0, 0) \in \mathcal{T}((0, 0, 0), \mathcal{D})^*,$$

where $\mathcal{T}((0, 0, 0), \mathcal{D})$ denotes the tangent cone to the set $\mathcal{D}$ in the point $(0, 0, 0)$ (see Definition 2.5). By virtue of Proposition 5.9, this is equivalent to

$$(-\nabla f(z^*), 0, 0) \in \hat{N}((0, 0, 0), \mathcal{D}) \subseteq N((0, 0, 0), \mathcal{D}), \tag{5.61}$$

where the inclusion holds due to Proposition 5.2.

In order to calculate $N((0,0,0),\mathcal{D})$ in a fashion conducive to our goal, we need to consider the normal cones $\mathcal{D}_1$ and $\mathcal{D}_2$ separately. To be able to do this, we need some auxiliary results. We start off with stating, without proof, a result by Henrion, Jourani, and Outrata, see [31, Corollary 4.2].

Lemma 5.26 *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be arbitrary sets, and let $z \in \mathcal{C}_1 \cap \mathcal{C}_2$ be given. If the multifunction $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, given by*

$$\Phi(v) := \{w \in \mathcal{C}_1 \mid v + w \in \mathcal{C}_2\}, \tag{5.62}$$

is calm at $(0, z)$, it holds that

$$N(z, \mathcal{C}_1 \cap \mathcal{C}_2) \subseteq N(z, \mathcal{C}_1) + N(z, \mathcal{C}_2). \tag{5.63}$$

We now show that a function of type (5.62) defined using the program (5.58) is calm. This is done by showing that it is a polyhedral multifunction and hence locally upper Lipschitz continuous and hence calm.

Lemma 5.27 *Let the multifunction $\Phi : \mathbb{R}^{n+2|\beta|} \rightrightarrows \mathbb{R}^{n+2|\beta|}$ be given by*

$$\Phi(v) := \{w \in \mathcal{D}_1 \mid v + w \in \mathcal{D}_2\}. \tag{5.64}$$

Then Φ is a polyhedral multifunction.

Proof. Since the graph of Φ may be expressed as

$$\begin{aligned}
\operatorname{gph}\Phi = \{(d^v, \xi^v, \eta^v, d^w, \xi^w, \eta^w) \in \mathbb{R}^{2(n+2|\beta|)} \mid \\
&\nabla g_i(z^*)^T d^w \le 0, && \forall i \in \mathcal{I}_g, \\
&\nabla h_i(z^*)^T d^w = 0, && \forall i = 1, \dots, p, \\
&\nabla G_i(z^*)^T d^w = 0, && \forall i \in \alpha, \\
&\nabla H_i(z^*)^T d^w = 0, && \forall i \in \gamma, \\
&\nabla G_i(z^*)^T d^w - \xi_i^w = 0, && \forall i \in \beta, \\
&\nabla H_i(z^*)^T d^w - \eta_i^w = 0, && \forall i \in \beta, \\
&\xi^v + \xi^w \ge 0,\ \eta^v + \eta^w \ge 0,\ (\xi^v + \xi^w)^T(\eta^v + \eta^w) = 0\, \} \\
= \bigcup_{(\nu_1,\nu_2)\in\mathcal{P}(\{1,\dots,|\beta|\})} \{(d^v, \xi^v, \eta^v, d^w, \xi^w, \eta^w) \in \mathbb{R}^{2(n+2|\beta|)} \mid \\
&\nabla g_i(z^*)^T d^w \le 0, && \forall i \in \mathcal{I}_g, \\
&\nabla h_i(z^*)^T d^w = 0, && \forall i = 1, \dots, p, \\
&\nabla G_i(z^*)^T d^w = 0, && \forall i \in \alpha, \\
&\nabla H_i(z^*)^T d^w = 0, && \forall i \in \gamma, \\
&\nabla G_i(z^*)^T d^w - \xi_i^w = 0, && \forall i \in \beta, \\
&\nabla H_i(z^*)^T d^w - \eta_i^w = 0, && \forall i \in \beta,
\end{aligned}$$

$$\begin{aligned} \xi^v_{\nu_1} + \xi^w_{\nu_1} = 0, \quad & \xi^v_{\nu_2} + \xi^w_{\nu_2} \geq 0, \\ \eta^v_{\nu_1} + \eta^w_{\nu_1} \geq 0, \quad & \eta^v_{\nu_2} + \eta^w_{\nu_2} = 0 \}, \end{aligned}$$

it is obviously the union of finitely many polyhedral convex sets. By Definition 5.13, Φ is therefore a polyhedral multifunction. □

Since Φ defined in (5.64) is a polyhedral multifunction, Proposition 5.14 may be invoked to show that Φ is locally upper Lipschitz continuous at every point $v \in \mathbb{R}^{n+2|\beta|}$. It is therefore, in particular, calm at $(0,(0,0,0)) \in \operatorname{gph}\Phi$. By invoking Lemma 5.26, we see that (5.61) implies

$$(-\nabla f(z^*), 0, 0) \in N((0,0,0), \mathcal{D}_1) + N((0,0,0), \mathcal{D}_2). \tag{5.65}$$

Now, $\mathcal{D}_1$ is obviously a convex set. Hence the limiting normal cone to $\mathcal{D}_1$ coincides with the standard convex normal cone given by (5.5). A comparison with the definition of the dual cone (see Definition 2.6) yields that

$$N((0,0,0), \mathcal{D}_1) = -\mathcal{D}_1^*. \tag{5.66}$$

Given the fact that $\mathcal{D}_1$ is a polyhedral convex set, its dual, and by (5.66), its limiting normal cone at the origin is given by Lemma 3.3.

This yields the existence of λ^g, λ^h, λ^G and λ^H with $\lambda^g_{\mathcal{I}_g} \geq 0$ such that

$$\begin{aligned} \begin{pmatrix} -\nabla f(z^*) \\ 0 \\ 0 \end{pmatrix} \in & \sum_{i \in \mathcal{I}_g} \lambda^g_i \begin{pmatrix} \nabla g_i(z^*) \\ 0 \\ 0 \end{pmatrix} + \sum_{i=1}^p \lambda^h_i \begin{pmatrix} \nabla h_i(z^*) \\ 0 \\ 0 \end{pmatrix} \\ & - \sum_{i \in \alpha} \lambda^G_i \begin{pmatrix} \nabla G_i(z^*) \\ 0 \\ 0 \end{pmatrix} - \sum_{i \in \gamma} \lambda^H_i \begin{pmatrix} \nabla H_i(z^*) \\ 0 \\ 0 \end{pmatrix} \\ & - \sum_{i \in \beta} \Big[\lambda^G_i \begin{pmatrix} \nabla G_i(z^*) \\ -e^i \\ 0 \end{pmatrix} + \lambda^H_i \begin{pmatrix} \nabla H_i(z^*) \\ 0 \\ -e^i \end{pmatrix} \Big] \\ & + N((0,0,0), \mathcal{D}_2). \end{aligned} \tag{5.67}$$

where e^i denotes that unit vector in $\mathbb{R}^{|\beta|}$ which corresponds to the position of i in β. Note that since the signs in the second and third lines of (5.67) are arbitrary, they were chosen to facilitate the notation of the proof.

First, we take a look at the second and third components in (5.67). To this end, we rewrite the normal cone to $\mathcal{D}_2$ in the following fashion:

$$\begin{aligned} N((0,0,0), \mathcal{D}_2) &= N(0, \mathbb{R}^n) \times N((0,0), \{(\xi, \eta) \mid \xi \geq 0, \eta \geq 0, \xi^T \eta = 0\}) \\ &= \{0\} \times N((0,0), \{(\xi, \eta) \mid \xi \geq 0, \eta \geq 0, \xi^T \eta = 0\}). \end{aligned} \tag{5.68}$$

Here the first equality is due to the Cartesian product rule (see Proposition 5.4). The second equality uses that 0 is in the interior of $\mathbb{R}^n$, and hence any normal cone reduces to $\{0\}$.

Substituting (5.68) into (5.67) yields

$$(-\lambda_\beta^G, -\lambda_\beta^H) \in N((0,0), \{(\xi, \eta) \mid \xi \geq 0, \eta \geq 0, \xi^T \eta = 0\}).$$

Applying Proposition 5.8, we obtain that

$$(\lambda_i^G > 0 \ \wedge \ \lambda_i^H > 0) \quad \vee \quad \lambda_i^G \lambda_i^H = 0$$

for all $i \in \beta$. Note that since we need to determine the limiting normal cone in the point $(0,0)$, it holds that $\mathcal{I}_a = \mathcal{I}_b = \varnothing$ in Proposition 5.8.

Finally, we set $\lambda_\gamma^G := 0$, $\lambda_\alpha^H := 0$, and $\lambda_i^g := 0$ for all $i \notin \mathcal{I}_g$ and have thus acquired the conditions for M-stationarity (5.53) with $\lambda^* := (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$, completing our alternate proof of Theorem 5.25, which we restate here, for completeness' sake.

Theorem 5.28 *Let z^* be a local minimizer of the MPEC* (1.1) *at which MPEC-GCQ holds. Then there exists a Lagrange multiplier λ^* such that the conditions* (5.53) *hold, i.e. z^* is M-stationary.*

The central idea of the approach presented in this section was to separate the benign constraints from the complementarity constraints (divided here into $\mathcal{D}_1$ and $\mathcal{D}_2$) and consider the two types of constraints separately. Applying Lemma 5.26 yields that

$$N((0,0,0), \mathcal{D}) \subseteq N((0,0,0), \mathcal{D}_1) + N((0,0,0), \mathcal{D}_2), \tag{5.69}$$

enabling us to obtain (5.65) from (5.61). Note that this does not hold in general, but is a direct consequence of our MPAEC (5.57) having constraints characterized by affine functions (see Lemma 5.27).

If it were possible to separate the Fréchet normal cone to $\mathcal{D}_1 \cap \mathcal{D}_2$ in a fashion similar to (5.69), we would be able to obtain strong stationarity as a first order condition. However, the Fréchet normal does not admit this type of result. Also, we know that strong stationarity cannot be a necessary optimality condition under MPEC-GCQ (see Example 4.13).

Summary

To round up this chapter, we collect the most important results concerning constraint qualifications and stationarity concepts in Figure 5.1. Note that the results of Chapter 4 presented in Figure 4.2 are supplanted by the M-stationarity result Theorem 5.25 of the current chapter.

$$\begin{tikzcd}[arrows=Rightarrow]
 & \text{MPEC-LICQ} \arrow[d] & & \\
 & \text{MPEC-SMFCQ} \arrow[r] \arrow[dd] & \text{Guignard CQ} \arrow[dr] & \\
 & & & \text{strong stationarity} \\
 & \text{MPEC-MFCQ} \arrow[dd] & & \\
\text{MPEC-Slater CQs} \arrow[dr] & & \text{MPAEC} \arrow[dl] \arrow[dr] & \\
 & \text{MPEC-ACQ} \arrow[d] & & \text{local MPEC-error bound} \arrow[d] \\
 & \text{MPEC-GCQ} \arrow[dr] & & \text{MPEC-calmness} \arrow[dl] \\
 & & \text{M-stationarity} &
\end{tikzcd}$$

Figure 5.1: Necessary optimality conditions for a local minimizer z^* of the MPEC (1.1) under various constraint qualifications.

Chapter 6

Numerical Experiments

The centerpiece of this thesis has been the discussion of constraint qualifications and optimality conditions for MPECs. The previous chapters have focused on this theoretical aspect. Though this could stand by itself, we wish to dedicate this final chapter to embark on a short sojourn into the numerical solution of MPECs.

The time allotted to this stands in no comparison to the time spent on investigating the theory of MPECs, however. We therefore give a brief overview of existing approaches before generalizing the exact penalty result, Proposition 5.19, from Chapter 5. We discuss exact penalization in the context of MPEC algorithms and their globalization and then go on to gather some numerical experience with an exact penalization formulation of our MPEC (1.1). This should not be understood as a proposed algorithm, however, but rather an experiment to see whether our exact penalty result stands up to real-world problems.

6.1 Overview

For nonlinear programs, a multitude of very successful algorithms have been developed over the years. Among these are various versions of interior point methods and sequential quadratic programming, or SQP, methods. A large collection of freely and commercially available solvers exist. For a comprehensive list, the reader is referred to the NEOS server [51].

Unfortunately, the convergence analysis of such solvers often require that MFCQ or the stronger LICQ hold. Therefore, since MFCQ does not hold at any feasible point of an MPEC (see Proposition 2.15), we cannot expect any of these algorithms to perform well when applied to an MPEC. Substantiating this, early tests performed on MPECs yielded poor results, see, e.g., [42, 3].

Instead, several approaches have been suggested to solve MPECs numerically. We will touch upon some of them briefly in the following. The discussion presented here, however, is not intended to be a comprehensive survey of MPEC algorithms, but rather to convey some of the ideas used in the numerical treatment of MPECs. The reader is asked to follow up on those references that are given for more information.

The most broadly practiced idea is to try to solve an MPEC utilizing existing solvers with little or no alteration. This has the advantage that these implementations have been around for a while and have thus been thoroughly tested and perfected to a degree not possible if starting from scratch.

To be able to apply standard solvers to an MPEC, care must be taken to remove, or at least mitigate, the effect of the complementarity term in (1.1). In the following, we recall several ideas of how this might be accomplished.

Regularization techniques solve a sequence of nonlinear programs, given by

$$\begin{array}{ll} \min & f(z) \\ \text{s.t.} & g(z) \le 0, \quad h(z) = 0, \\ & G(z) \ge 0, \quad H(z) \ge 0, \quad G(z)^T H(z) \le \tau_k, \end{array} \tag{6.1}$$

where a sequence $\{\tau_k\}$ is chosen with $\tau_k \searrow 0$. Note that the original MPEC (1.1) is recovered for $\tau_k = 0$. See [63, 58] for discussions of this idea.

One drawback of this method is, of course, that a whole sequence of nonlinear programs must be solved. To counteract this, Demiguel et. al. [11], perform one step of an interior point method, and then update τ_k.

Another solution technique uses so-called NCP-functions $\phi : \mathbb{R}^2 \to \mathbb{R}$, defined by

$$\phi(a, b) = 0 \quad :\Longleftrightarrow \quad a \ge 0,\ b \ge 0,\ ab = 0.$$

The MPEC (1.1) is then equivalently expressed by

$$\begin{array}{ll} \min & f(z) \\ \text{s.t.} & g(z) \le 0, \quad h(z) = 0, \\ & \phi(G_i(z), H_i(z)) = 0, \quad \forall i = 1, \dots, l. \end{array} \tag{6.2}$$

A straightforward application of solvers to this nonlinear program will fail for the same reasons they fail for the original formulation (1.1).

Leyffer [39] has shown that, in an SQP setting, appropriate linearizations of the min-function, defined by $\phi(a, b) := \min\{a, b\}$ (see also (6.10)) remain feasible near a solution point of an MPEC. Furthermore, if an element of the Clarke subgradient is chosen at nondifferentiable points of ϕ, this has no adverse effect, under certain additional conditions, on the local convergence of an SQP algorithm. This is substantiated by good numerical results, see [39].

Rather than using the Clarke subgradient at nondifferentiable points of ϕ, it can be smoothed using a parameter τ, giving rise to *smoothing techniques.* The smoothing is chosen in such a fashion so that for $\tau = 0$, the (nonsmooth) NCP function ϕ is recovered. I.e., if ϕ_τ denotes the smoothed NCP function, it holds that $\phi_0 \equiv \phi$.

Again, a sequence of nonlinear programs, parameterized by $\tau_k \searrow 0$,

$$\begin{array}{ll} \min & f(z) \\ \text{s.t.} & g(z) \le 0, \quad h(z) = 0, \\ & \phi_{\tau_k}(G_i(z), H_i(z)) = 0, \quad \forall i = 1, \dots, l. \end{array} \tag{6.3}$$

may be solved. This was done, e.g., in [25].

Other ideas again include explicitly updating the parameter τ_k in an algorithm, or having a solver implicitly update it by treating τ as a variable and adding an additional constraint:

$$\begin{aligned} \min_{(z,\tau)} \quad & f(z) \\ \text{s.t.} \quad & g(z) \leq 0, \quad h(z) = 0, \\ & \phi_\tau(G_i(z), H_i(z)) = 0, \quad \forall i = 1, \ldots, l \\ & e^\tau - 1 = 0. \end{aligned} \tag{6.4}$$

The nature of the constraint $e^\tau - 1 = 0$ is such that starting an SQP algorithm with $\tau_0 > 0$ guarantees that $\tau_k > 0$ for all iterates of an SQP implementation, keeping the smoothed NCP function ϕ_τ smooth. See [34] for an exhaustive discussion of both the implicit and explicit approaches.

An intermediate approach to using smoothed NCP functions was described by Facchinei, Jiang and Qi [14]. In each iteration, given $\tau_k > 0$, they solve the smoothed program (6.3) up to a specified numerical accuracy ε_k. The parameters ε_k and τ_k are then updated appropriately and the next iteration is initiated.

Another technique, the *penalization technique*, completely removes the complementarity term from the constraints and adds it to the objective in the following manner:

$$\begin{aligned} \min \quad & f(z) + \rho G(z)^T H(z) \\ \text{s.t.} \quad & g(z) \leq 0, \quad h(z) = 0, \\ & G(z) \geq 0, \quad H(z) \geq 0. \end{aligned} \tag{6.5}$$

Here, ρ is a penalty parameter. Once more, a sequence of programs (6.5), with $\rho_k \to \infty$, may be solved. This approach is investigated, e.g., in [32]. See also [58, Section 5] for an investigation of exact penalization properties of the program (6.5).

Anitescu [3, 2] proposes a hybrid of the regularization and penalization techniques, where he relaxes the complementarity constraint as in (6.1) and penalizes the relaxation parameter τ in the objective, treating it as a variable. Additionally, he does this not only with the complementarity constraint, but with all constraints. This approach is inspired by the *elastic mode*, as implemented by some solvers, see, e.g., [28].

Any of the algorithmic approaches discussed above are shown to locally, and in some cases, globally, converge to a C-, M-, or even strongly stationary point. Different degrees of first order, second order, and dual conditions are required for these results. The reader is referred to the references given above for more information.

We now turn our attention to algorithmic ideas which are not based on the premise that we have at our disposal general solvers for nonlinear programming. Outrata [52], and Outrata, Kočvara and Zowe [54] propose to use a trust region bundle method, if the program has the special form

$$\begin{aligned} \min \quad & f(x, y) \\ \text{s.t.} \quad & y \in S(x), \\ & x \in \mathcal{X}. \end{aligned} \tag{6.6}$$

Here, S assigns a solution of some lower-level problem to each $x \in \mathcal{X}$. In particular, S may describe a variational inequality, which may then be expressed as an MPEC.

If we assume that S has a *unique* solution for each $x \in \mathcal{X}$, the program (6.6) may be rewritten as

$$\begin{array}{ll} \min & f(x, S(x)) \\ \text{s.t.} & x \in \mathcal{X}. \end{array} \tag{6.7}$$

Note that S is nonsmooth in general.

Then, if $\mathcal{X}$ is described by simple constraints, we are able to compute an element of the Clarke subgradient of S, and f is continuously differentiable, we may apply a bundle trust region algorithm to solve (6.7).

The applications described in [54] were solved using this method. For a description of the bundle method used, see [65, 54], as well as Section 6.3 below. Section 6.3 also contains a more mundane application of the bundle trust region algorithm to the MPEC (1.1).

Yet another idea to solve MPECs borrows the idea from sequential quadratic programming of linearizing the constraints of a nonlinear program, but treating the complementarity term separately. The same idea that lead to the introduction of the MPEC-linearized tangent cone (3.13), "linearizing under the complementarity," keeps the same information for the complementarity constraint that is kept for the other constraints. This yields a sequential MPAEC programming approach.

In his doctoral thesis, Stöhr [71] does just this, adding a trust region element for globalization. A serious drawback, however, is that the MPAEC to be solved in each iteration is not a quadratic program. Therefore, a considerable amount of work is required to solve these MPAECs.

Stöhr introduces a convex majorant of an exact penalty reformulation of the inner MPAEC and solves this using existing techniques. Unfortunately, it is not a priori clear how to choose this majorant. It is fairly easy to verify whether a C-stationary point of the MPAEC has been found. If, however, strong stationarity is desired, an exponentially large number of convex majorants may have to be solved.

To mitigate this, Stöhr only requires that the exact penalty reformulation of the MPAEC satisfy some descent criterion. For more details, the interested reader is referred to [71] as well as [64]. We will very briefly discuss a similar sequential MPAEC programming idea in the next section.

6.2 Exact Penalization

A property of any of the methods discussed in the previous section is that they are local in nature. Therefore, they need to be globalized by some means. Invariably, a penalty function is required for this. A line search method looks to decrease the value of the penalty function along a direction given by the solution of some inner program. A trust region approach modifies the trust region radius based on comparing the actual with the predicted improvement using some penalty function.

Another situation where penalization becomes useful to a certain degree, is when we wish to apply a solver for unconstrained optimization to a constrained optimization problem. The constraints are then added to the objective via a penalty term. Experience has shown that, given the option, other methods should always be preferred to such an approach. The simplicity of this idea, however, has some appeal, and accordingly, we gather some numerical data using it in Section 6.3.

Given the central role played by penalization in numerical algorithms, we now return to the exact penalty result presented in Proposition 5.19. In Chapter 5, this result was used to prove that M-stationarity is a necessary optimality condition under MPEC-calmness (see Theorem 5.20). Proposition 5.19 is, however, of little use for numerical applications since the program (5.45) only penalizes part of the original constraints, keeping some of the constraints. However, we can use the method of proof of Proposition 5.19 to prove a more useful penalty result, which we present in the following proposition.

Proposition 6.1 *Let z^* be a local minimizer of the MPEC* (1.1). *Then the following are equivalent:*

(a) the MPEC is MPEC-calm at z^*;

(b) there exists a $\rho_1 > 0$ such that for all $\rho \geq \rho_1$, the vector z^* is a local minimizer of

$$\min_z \; f(z) + \rho(\|\max\{0, g(z)\}\| + \|h(z)\| + \|\min\{G(z), H(z)\}\|). \tag{6.8}$$

Proof. As mentioned in the discussion leading up to this proposition, the proof of this result is carried out analogously to the proof of Proposition 5.19.
(a) $\Rightarrow$ (b). Choose $\delta \leq \varepsilon$ such that

$$\begin{aligned} \max_{z \in B_\delta(z^*)} \|\max\{0, g(z)\}\| &\leq \frac{\varepsilon}{4}, \\ \max_{z \in B_\delta(z^*)} \|h(z)\| &\leq \frac{\varepsilon}{4}, \\ \max_{z \in B_\delta(z^*)} \|\min\{G(z), H(z)\}\| &\leq \frac{\varepsilon}{4}. \end{aligned}$$

This exists due to the continuity of g, h, G, and H.

Furthermore, let $z \in B_\delta(z^*)$ be arbitrarily given. Then set

$$\begin{aligned} p_i &:= -\max\{0, g_i(z)\} && \forall i = 1, \ldots, m, \\ q &:= h(z), \\ r_i := s_i &:= -\min\{G_i(z), H_i(z)\} && \forall i = 1, \ldots, l. \end{aligned}$$

With this choice of p, q, r, and s it holds that

$$\|(p, q, r, s)\| \leq \|p\| + \|q\| + \|r\| + \|s\| \leq \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon$$

and hence $(p,q,r,s) \in B_\varepsilon(0)$ as well as $z \in \mathcal{Z}(p,q,r,s) \cap B_\varepsilon(z^*)$.

Since z^* satisfies MPEC-calmness, condition (5.44) holds and we obtain

$$\begin{aligned} f(z^*) &\le f(z) + \mu \,\|(p,q,r,s)\| \\ &\le f(z) + \mu(\|p\| + \|q\| + \|r\| + \|s\|) \\ &= f(z) + \mu(\|\max\{0, g(z)\}\| + \|h(z)\| + 2\,\|\min\{G(z), H(z)\}\|) \\ &\le f(z) + \rho(\|\max\{0, g(z)\}\| + \|h(z)\| + \|\min\{G(z), H(z)\}\|) \end{aligned} \tag{6.9}$$

for all $\rho \ge 2\mu > 0$. Since $z \in B_\delta(z^*)$ was chosen arbitrarily, (6.9) implies that z^* is a local minimizer of the program (6.8) for all $\rho \ge \rho_1 := 2\mu$.

(b) $\Rightarrow$ (a). Similar to the situation in the proof of Proposition 5.19, we prove this for the Euclidean norm only. An extension to an arbitrary norm is then trivial.

Let ε be given such that z^* is a global minimizer of (6.8) with $\rho = \rho_1$ in the ball $B_\varepsilon(z^*)$. Now, let $(p,q,r,s) \in B_\varepsilon(0)$ and $z \in \mathcal{Z}(p,q,r,s) \cap B_\varepsilon(z^*)$ be chosen arbitrarily. It then follows from (6.8) that

$$\begin{aligned} f(z^*) &\le f(z) + \rho_1(\|\max\{0, g(z)\}\| + \|h(z)\| + \|\min\{G(z), H(z)\}\|) \\ &\le f(z) + \tilde{\mu}(\|\max\{0, g(z)\}\|_1 + \|h(z)\|_1 + \|\min\{G(z), H(z)\}\|_1) \\ &= f(z) + \tilde{\mu}(\|\max\{0, g(z) + p - p\}\|_1 + \|h(z) + q - q\|_1 \\ &\qquad\qquad + \|\min\{G(z) + r - r, H(z) + s - s\}\|_1) \\ &\le f(z) + \tilde{\mu}(\|\max\{0, \underbrace{g(z) + p}_{\le 0}\}\|_1 + \|p\|_1 + \|\underbrace{h(z) + q}_{=0}\|_1 + \|q\|_1 \\ &\qquad\qquad + \|\underbrace{\min\{G(z) + r, H(z) + s\}}_{=0}\|_1 + \|r\|_1 + \|s\|_1) \\ &= f(z) + \tilde{\mu}\,\|(p,q,r,s)\|_1 \\ &\le f(z) + \mu\,\|(p,q,r,s)\| \end{aligned}$$

for some $\tilde{\mu}, \mu > 0$, due to the equivalence of norms. This concludes the proof. $\square$

Note that in [64, Theorem 3.11] an exact penalty result for MPECs is stated under much stronger conditions.

Remark. Proposition 6.1 states, in particular, that MPEC-calmness at z^* is equivalent to z^* being a local minimizer of the unconstrained program (6.8). It is known from standard nonlinear programming (see, e.g., [6, Theorem 1.1]) that exact penalization in the sense of (6.8) is equivalent to standard calmness at z^* of the program

$$\begin{array}{ll} \min & f(z) \\ \text{s.t.} & g(z) \le 0, \\ & h(z) = 0, \\ & \min\{G(z), H(z)\} = 0. \end{array} \tag{6.10}$$

Hence standard calmness at z^* of the program (6.10) is equivalent to MPEC-calmness at z^* of the MPEC (1.1). Note that the feasible regions of the MPEC (1.1) and the program (6.10) are identical, but that the min-term in (6.10) is not differentiable.

The advantage of the program (6.8) over (5.45) is that it is an unconstrained optimization problem. For this reason, the objective of the program (6.8), is often considered independently of any optimization problem. This is made precise in the following definition, where we explicitly restrict ourselves to the q-norms for $q \in [1, \infty]$.

Definition 6.2 *Given the MPEC* (1.1) *and an arbitrary* $q \in [1, \infty]$, *we call*

$$P_q(z; \rho) := f(z) + \rho(\|\max\{0, g(z)\}\|_q + \|h(z)\|_q + \|\min\{G(z), H(z)\}\|_q), \tag{6.11}$$

the exact l_q MPEC-penalty function *of the MPEC* (1.1).

Note that, as with MPEC-calmness and standard calmness, the exact l_q MPEC-penalty function P_q is identical to the standard exact l_q penalty function of the program (6.10). The standard l_q penalty function of the MPEC (1.1) takes on the form

$$\begin{aligned} \tilde{P}_q(z; \rho) :=& f(z) + \rho\big(\|\max\{0, g(z)\}\|_q + \|h(z)\|_q \\ &+ \|\max\{0, -G(z)\}\|_q + \|\max\{0, -H(z)\}\|_q + |G(z)^T H(z)|\big). \end{aligned} \tag{6.12}$$

At this point we wish to point out that P_q is, in fact, an *exact* penalty function (see, e.g., [27, Definition 5.8] for a formal definition). Proposition (6.1) states, in particular, that there exists a *finite* penalty parameter $\rho_1 > 0$ such that a local minimizer z^* of the MPEC (1.1) is also a local minimizer of the exact l_q penalty function $P_q(\cdot; \rho)$ for all $\rho \geq \rho_1$. Recall that MPEC-calmness at z^* implies that z^* is a local minimizer of the MPEC (1.1).

A Sequential MPAEC Programming Approach

As was mentioned at the beginning of this section, penalty functions are the basis of globalization techniques for sequential programming algorithms. We will now briefly investigate our exact l_q MPEC-penalty function in this context.

Given an MPEC (1.1) and a current iterate z^k, not necessarily feasible for the MPEC, we define the following inner MPAEC:

$$\begin{array}{ll} \min & \nabla f(z^k)^T d + \frac{1}{2} d^T H_k d \\ \text{s.t.} & g(z^k) + g'(z^k) d \leq 0, \\ & h(z^k) + h'(z^k) d = 0, \\ & G(z^k) + G'(z^k) d \geq 0, \\ & H(z^k) + H'(z^k) d \geq 0, \\ & \big(G(z^k) + G'(z^k) d\big)^T \big(H(z^k) + H'(z^k) d\big) = 0, \end{array} \tag{6.13}$$

where H_k is a positive definite approximation of the Hessian of the Lagrange function of the MPEC (cf. (4.5)).

Note that (6.13) is in fact an MPAEC with a convex quadratic objective function. Further note that in an SQP setting, the inner program would be a quadratic program where the final line of (6.13) would be replaced by a linearization of $\theta(z) = G(z)^T H(z)$ around $z = z^k$.

A solution d^k of the MPAEC (6.13) may then be used as a search direction for an outer iteration.

A similar algorithm was investigated in detail in the thesis of Stöhr [71]. His globalization uses a trust region approach by adding a constraint of the form $\|d\| \leq \Delta$ to the MPAEC (6.13). The algorithm then decides to alter the trust region Δ based on information gleaned from the penalty function P_q, see [71] for details.

Another idea to globalize the sequential MPAEC programming approach described above is to use a line search, which attempts to obtain a decrease in the exact l_q MPEC-penalty function along the search direction d^k provided by the MPAEC (6.13). This idea, however, fails, since it is easy to construct examples where a solution of the MPAEC (6.13) is not a descent direction of the exact l_q MPEC-penalty function, as is demonstrated by the following example.

Example 6.3 Let the MPEC

$$\begin{array}{rl} \min & -2z + \frac{1}{2}z^2 \\ \text{s.t.} & 1 + 2z \geq 0, \\ & 2 - 2z \geq 0, \\ & (1+2z)(2-2z) = 0, \end{array} \tag{6.14}$$

as well as an initial iterate $z^0 = 0$ be given. Note that z^0 is not feasible for the program (6.14) and hence not a solution. Then the inner MPAEC is given by

$$\begin{array}{rl} \min & -2d + \frac{1}{2}H_0 d^2 \\ \text{s.t.} & 1 + 2d \geq 0, \\ & 2 - 2d \geq 0, \\ & (1+2d)(2-2d) = 0, \end{array} \tag{6.15}$$

where we choose $H_0 := 1$ as the initial approximation to the Hessian of the Lagrangian. It is then easily verified that $d^0 = 1$ is the global minimizer of the MPAEC (6.15). Interestingly, it holds that $\beta = \varnothing$, i.e. strict complementarity is satisfied.

We now investigate the value of the exact l_q MPEC-penalty function P_q at the point $z^0 + \tau d^0 = \tau$ for $\tau > 0$. This will yield information about whether d^0 is a descent direction of P_q. Keeping in mind that for scalar arguments, it holds that $\|\cdot\|_q \equiv \|\cdot\|_1$ for all $q \in [1, \infty]$, we have that, for $\tau > 0$ sufficiently small,

$$\begin{aligned} P_q(z^0 + \tau d^0; \rho) - P_q(z^0; \rho) &= P_q(\tau; \rho) - P_q(0; \rho) \\ &= -2\tau + \frac{1}{2}\tau^2 + \rho|\min\{1 + 2\tau, 2 - 2\tau\}| - \rho \\ &= -2\tau + \frac{1}{2}\tau^2 + \rho(1 + 2\tau) - \rho \\ &= \frac{1}{2}\tau^2 + 2\tau(\rho - 1) > 0 \end{aligned}$$

for $\rho \geq 1$. This implies that $d^0 = 1$ is not a descent direction of P_q.

Incidentally, it may similarly be shown that d^0 is not a descent direction of the standard exact l_q penalty function $\tilde{P}_q$ (see (6.12)) either.

The above example demonstrates that it is not possible to use the exact l_q MPEC-penalty function in a line search setting.

Short of attempting to improve Stöhr's trust region approach by using the recently introduced filter methods (see, e.g., [9]), the sequential MPAEC programming approach does not instill much hope for improved numerical behavior. This is compounded by the fact that it is not clear how to obtain a solution of the inner MPAEC (6.13).

Since the potential gain of improving Stöhr's method stands in no comparison to the time we have available for a numerical investigation of MPECs, we now concentrate on a more naive use of exact penalization; applying unconstrained optimization solvers to the exact l_q MPEC-penalty function. This is the focus of the next section.

6.3 A Bundle Trust Region Implementation

In this section we will apply existing unconstrained solvers to an exact penalty reformulation of our MPEC (1.1). For simplicity's sake, we confine ourselves to the exact l_1 MPEC-penalty function,

$$P_1(z;\rho) = f(z) + \rho\Big(\sum_{i=1}^{m} \max\{0, g_i(z)\} + \sum_{i=1}^{p} |h_i(z)| + \sum_{i=1}^{l} |\min\{G_i(z), H_i(z)\}|\Big). \quad (6.16)$$

As was already mentioned, P_1 is not differentiable. Therefore, a tempting idea is to apply the simplex method of Nelder and Mead [50] to P_1. This method does not require any derivative information of the function to be minimized. It has the added advantage of being easy to implement. However, initial tests using the Nelder-Mead method yielded abysmal results. Though the algorithm converged, it was almost invariably to a point not feasible for the MPEC (1.1) to within a desired numerical accuracy. This necessitated several restarts until feasibility was finally obtained. For problems of dimension 5, as many as 20 restarts were necessary to acquire the desired accuracy in feasibility. Additionally, even if feasibility was finally obtained, it was usually not within an acceptable range of a local minimizer.

Although improvements of the classic Nelder-Mead method exist (see, e.g., [74, 49]), we chose not to pursue this path any further. We also did not look to other derivative free optimization solvers which build an increasingly accurate model of the objective function which is then minimized (see the overview articles [75, 36]).

Instead, we opted to use an algorithm which is able to use that derivative information of P_1 which is available. At the points at which P_1 is not derivable, we are still able to compute an element of its Clarke subgradient $\partial^{\mathrm{Cl}} P_1(\cdot;\rho)$. The most successful methods that use this additional information are the bundle or cutting plane methods, see the monograph [35] for an overview.

A very successful approach combining the trust region philosophy with a bundle method was proposed by Schramm and Zowe [65], see also [54]. It is their bundle trust region, or BT, code that we use in our implementation. The FORTRAN sources of the nonconvex BT code were kindly provided by Jiří Outrata.

The idea of the BT code is to build an approximation of an objective function $f : \mathbb{R}^n \to \mathbb{R}$ using the *bundle* information $(f(y_i), g_i)$ for $i \in J_k$, where $g_i \in \partial^{\mathrm{Cl}} f(y_i)$ is a member of the Clarke subgradient of f at y_i and $J_k \subseteq \{0, 1, \ldots, k\}$ is some set determining which part of the bundle information is used in the current iteration. The reader is referred to [65] for details.

In particular, the BT code requires that the user provide a function which, for a given y, computes $f(y)$ and an arbitrary element of the Clarke subgradient $\partial^{\mathrm{Cl}} f(y)$. Since the function we wish to minimize is the exact l_1 MPEC-penalty function P_1, we require an element of its Clarke subgradient. The following lemma gives an approximation of the Clarke subgradient of P_1.

Lemma 6.4 *Let $z \in \mathbb{R}^n$ be arbitrarily given. Then the Clarke subgradient of the exact l_1 MPEC-penalty function P_1 is given by*

$$
\begin{aligned}
&\partial^{\mathrm{Cl}} P_1(z;\rho) \subseteq \\
&\quad \nabla f(z) + \rho \sum_{i=1}^{m} \begin{cases} 0 & : \ g_i(z) < 0, \\ \operatorname{conv}\{0, \nabla g_i(z)\} & : \ g_i(z) = 0, \\ \nabla g_i(z) & : \ g_i(z) > 0 \end{cases} \\
&\quad + \rho \sum_{i=1}^{p} \begin{cases} -\nabla h_i(z) & : \ h_i(z) < 0, \\ \operatorname{conv}\{-\nabla h_i(z), \nabla h_i(z)\} & : \ h_i(z) = 0, \\ \nabla h_i(z) & : \ h_i(z) > 0 \end{cases} \\
&\quad + \rho \sum_{i=1}^{l} \begin{cases} -\nabla G_i(z) & : \ 0 > G_i(z) < H_i(z), \\ \operatorname{conv}\{-\nabla G_i(z), \nabla G_i(z)\} & : \ 0 = G_i(z) < H_i(z), \\ \nabla G_i(z) & : \ 0 < G_i(z) < H_i(z), \\ -\nabla H_i(z) & : \ 0 > H_i(z) < G_i(z), \\ \operatorname{conv}\{-\nabla H_i(z), \nabla H_i(z)\} & : \ 0 = H_i(z) < G_i(z), \\ \nabla H_i(z) & : \ 0 < H_i(z) < G_i(z), \\ \operatorname{conv}\{-\nabla G_i(z), -\nabla H_i(z)\} & : \ 0 > G_i(z) = H_i(z), \\ \operatorname{conv}\left\{\xi_i \operatorname{conv}\{-\nabla G_i(z), -\nabla H_i(z)\} \mid \xi_i \in [-1, 1]\right\} & : \ 0 = G_i(z) = H_i(z), \\ \operatorname{conv}\{\nabla G_i(z), \nabla H_i(z)\} & : \ 0 < G_i(z) = H_i(z). \end{cases}
\end{aligned}
\tag{6.17}
$$

Proof. We invoke [8, Proposition 2.3.3] to split the subgradient of P_1 into a sum over subgradients. For the terms concerning g and h, we note that $|h_i(z)| = \max\{-h_i(z), h_i(z)\}$, and invoke [8, Proposition 2.3.12]. Note that $\partial^{\mathrm{Cl}} \phi(z) = \{\nabla \phi(z)\}$ for any continuously differentiable $\phi : \mathbb{R}^n \to \mathbb{R}$, see [8, Proposition 2.2.4].

Now consider the term $\phi_i(z) := |\min\{G_i(z), H_i(z)\}|$ for any $i \in \{1, \ldots, l\}$. If $G_i(z) < H_i(z)$, it holds that $\phi_i(z) = |G_i(z)|$ and $\partial^{\mathrm{Cl}} \phi_i(z)$ is again given by [8, Proposition 2.3.12]. Similarly, $\partial^{\mathrm{Cl}} \phi_i(z)$ can be determined if $G_i(z) > H_i(z)$.

For the case when $G_i(z) = H_i(z)$ first consider the following:

$$\begin{aligned}\phi_i(z) &= |\min\{G_i(z), H_i(z)\}| \\ &= |-\max\{-G_i(z), -H_i(z)\}| \\ &= |\max\{-G_i(z), -H_i(z)\}| \\ &= (u \circ v_i)(z)\end{aligned}$$

where $u : \mathbb{R} \to \mathbb{R}$ and $v_i : \mathbb{R}^n \to \mathbb{R}$ are defined by

$$u(t) := |t| \quad \text{and} \quad v_i(z) := \max\{-G_i(z), -H_i(z)\},$$

respectively. By virtue of the chain rule [8, Theorem 2.6.6] we then have

$$\begin{aligned}\partial^{\text{Cl}}\phi_i(z) &\subseteq \text{conv}\left\{\partial^{\text{Cl}}u(t)\big|_{t=v_i(z)} \cdot \partial^{\text{Cl}}v_i(z)\right\} \\ &\subseteq \text{conv}\left\{\partial^{\text{Cl}}u(t)\big|_{t=v_i(z)} \cdot \text{conv}\{-\nabla G_i(z), -\nabla H_i(z)\}\right\} \\ &\subseteq \begin{cases}\text{conv}\{-\nabla G_i(z), -\nabla H_i(z)\} & : \; -G_i(z) = -H_i(z) > 0, \\ \text{conv}\left\{\xi_i \,\text{conv}\{-\nabla G_i(z), -\nabla H_i(z)\} \mid \xi_i \in [-1,1]\right\} & : \; -G_i(z) = -H_i(z) = 0, \\ -\text{conv}\{-\nabla G_i(z), -\nabla H_i(z)\} & : \; -G_i(z) = -H_i(z) < 0,\end{cases}\end{aligned}$$

where the last two inclusions are obtained by applying [8, Proposition 2.3.12] first to v_i and then to u. This completes the proof. □

All that remains now is to embed the BT code in an appropriate framework. One problem is that the penalty parameter ρ_1 for which P_1 becomes exact (see Proposition 6.1) is not known a priori. We therefore start with an initial guess ρ^0 of a penalty parameter and increase it if the BT code does not converge, or if it converges to a point that is infeasible for the MPEC (1.1).

We now state the framework of the algorithm we implemented and then proceed to explain each step in detail.

Algorithm 6.5 *(bundle trust region framework)*

(S.0) Choose an initial guess z^{init} as well as $\rho^0 > 0$, $\sigma > 1$ and $\rho^{\max} \geq \rho^0$, and set $z^0 := z^{\text{init}}$ and $k := 0$.

(S.1) Obtain a minimizer z^k_{BT} of $P_1(\cdot;\rho^k)$ using the BT code with starting point z^k.
If BT converged, go to (S.2). Otherwise, go to (S.3).

(S.2) ***BT converged.***
If z^k is feasible for the MPEC (1.1)*, set $z^* := z^k_{\text{BT}}$ and STOP (convergence).*
If z^k is infeasible for the MPEC (1.1)*, set $z^{k+1} := z^k_{\text{BT}}$, choose $\rho^{k+1} \in (\rho^k, \sigma\rho^k]$, and go to (S.4).*

(S.3) ***BT failed.*** *Restart Step.*
Choose $\rho^{k+1} := \sigma\rho^k$, set $z^{k+1} := z^{\text{init}}$, and go to (S.4).

(S.4) If $\rho^{k+1} > \rho^{\max}$: STOP (failure).

(S.5) Set $k \leftarrow k+1$ and go to (S.1).

We now motivate each step of Algorithm 6.5, also stating implementation details. It should be noted that Algorithm 6.5, though it is plausible, is a purely heuristical approach. Neither is a convergence theory given, nor is there any reason to expect Algorithm 6.5 to converge in general.

We start off with the initialization step (S.0). We choose an initial guess and store it in z^{init}. This is necessary because z^k is changed from iteration to iteration, while the initial guess z^{init} is required in the restart step (S.3) (see below).

In real-world applications, often an initial guess is provided, based on knowledge of the underlying system. If, however, no such guess is available, we choose z^{init} to be feasible for the box constraints imposed on the variables. If $l \in (\mathbb{R} \cup \{-\infty\})^n$ and $u \in (\mathbb{R} \cup \{\infty\})^n$ denote the lower and upper bounds on the variables, respectively, then z^{init} is chosen in the following fashion:

$$z_i^{\text{init}} := \begin{cases} \frac{1}{2}(l_i + u_i) & : \; l_i > -\infty, u_i < \infty, \\ l_i & : \; l_i > -\infty, u_i = \infty, \\ u_i & : \; l_i = -\infty, u_i < \infty, \\ 0 & : \; l_i = -\infty, u_i = \infty \end{cases} \tag{6.18}$$

for all components $i = 1, \ldots, n$.

The three parameters ρ^0, σ, and $\rho^{\max}$ pertain to the penalty parameter. Here, $\rho^0 > 0$ is the initial penalty parameter, which is increased by at most the factor $\sigma > 1$ in each iteration (see steps (S.2) and (S.3)). The parameter $\rho^{\max}$ denotes the maximal allowed value of the penalty parameter. Should ρ^k exceed this parameter, we abort the algorithm (see step (S.4)). In our specific implementation, we chose

$$\begin{aligned} \rho^0 &:= 5, \\ \sigma &:= 10, \\ \rho^{\max} &:= 10^6. \end{aligned}$$

The BT step (S.1) attempts to minimize the exact l_1 MPEC-penalty function $P_1(\cdot; \rho^k)$ with respect to the first argument for the current penalty parameter ρ^k. The BT code requires several parameters to be set. Among these are the maximum allowed number of iterations `maxit` and number of function evaluations `maxcom`. Additionally, a termination criterion ε^{BT} and the maximal number of stored subgradients $k_{\max}$ are specified by the user. The latter denotes the maximal number of elements in the set J_k which determines the bundle of the current iteration in the BT code. The reader is referred to [65] for details. We have found the following choice of parameters to work reasonably well:

$$\begin{aligned} \texttt{maxit} &:= 100n, \\ \texttt{maxcom} &:= 2 * \texttt{maxit}, \end{aligned}$$

$$\varepsilon^{\mathrm{BT}} := 10^{-6},$$
$$k_{\max} := \begin{cases} 25 & : \ n < 20, \\ n+5 & : \ \text{otherwise.} \end{cases}$$

Here n denotes the dimension of the problem, as usual.

Finally, BT requires an estimate of the minimum function value f_m^k. The only requirement is that $f_m^k < P_1(z^k;\rho^k)$, where z^k denotes the current iterate of Algorithm 6.5. In the absence of any information pertaining to the minimum function value, we have found the choice

$$f_m^k := \begin{cases} 2P_1(z^k;\rho^k) & : \ P_1(z^k;\rho^k) < -1, \\ (P_1(z^k;\rho^k)-1)/2 & : \ P_1(z^k;\rho^k) > 1, \\ P_1(z^k;\rho^k)-1 & : \ \text{otherwise} \end{cases}$$

to work quite well. The idea behind this choice of f_m^k is that we keep the order of $P_1(z^k;\rho^k)$, while bounding f_m^k away from it by a value of at least 1.

There is one special case when f_m^k can be chosen more intelligently. If the BT code converged in the previous iteration (this implies that $k \geq 1$), we use $f_m^k := P_1(z_{\mathrm{BT}}^{k-1};\rho^{k-1}$ as the estimate for the minimum function value. Since $\rho^k > \rho^{k-1}$ and the penalty term is also strictly positive (see the discussion of the termination criterion below), it holds that $f_m^k = P_1(z^k;\rho^{k-1}) < P_1(z^k;\rho^k)$. Note that $z^k = z_{\mathrm{BT}}^{k-1}$. Furthermore, if the previous solution z_{BT}^{k-1} is a good approximation of a solution of the MPEC (1.1), f_m^k is potentially a very good estimate.

As mentioned earlier, the BT code requires that the user provide a function which evaluates $P_1(z;\rho)$ and returns an arbitrary element of its Clarke subgradient for a given $z \in \mathbb{R}^n$. Motivated by Lemma 6.4, we chose

$$\begin{aligned} \eta = \nabla f(z) &+ \rho \sum_{i=1}^{m} \begin{cases} 0 & : \ g_i(z) <= 0, \\ \nabla g_i(z) & : \ g_i(z) > 0 \end{cases} \\ &+ \rho \sum_{i=1}^{p} \begin{cases} -\nabla h_i(z) & : \ h_i(z) < 0, \\ 0 & : \ h_i(z) = 0, \\ \nabla h_i(z) & : \ h_i(z) > 0 \end{cases} \\ &+ \rho \sum_{i=1}^{l} \begin{cases} -\nabla G_i(z) & : \ 0 > G_i(z) < H_i(z), \\ 0 & : \ 0 = G_i(z) < H_i(z), \\ \nabla G_i(z) & : \ 0 < G_i(z) < H_i(z), \\ -\nabla H_i(z) & : \ 0 > H_i(z) < G_i(z), \\ 0 & : \ 0 = H_i(z) < G_i(z), \\ \nabla H_i(z) & : \ 0 < H_i(z) < G_i(z), \\ (-\nabla G_i(z) - \nabla H_i(z))/2 & : \ 0 > G_i(z) = H_i(z), \\ 0 & : \ 0 = G_i(z) = H_i(z), \\ (\nabla G_i(z) + \nabla H_i(z))/2 & : \ 0 < G_i(z) = H_i(z), \end{cases} \end{aligned}$$

Note that the evaluation of η is implemented exactly as described above. Due to the nature of machine numbers, it is highly unlikely that any of the cases where equality between two floating point numbers is required, will ever occur. This was done deliberately, based on the experience that this choice usually performs well.

After having run the BT code, we check whether it has converged or whether it failed for some reason. Reasons for failure include that the maximal number of iterations or function evaluations has been reached without satisfying a termination criterion. This might occur, for example, if the function $P_1(\cdot;\rho^k)$ is unbounded. In this case it is reasonable to assume that the current penalty parameter ρ^k is too small. We therefore jump to step (S.3), where the penalty parameter is increased. It should be noted, however, that the maximum number of iterations or function evaluations may also be reached for other reasons, among which is ill-posedness of the problem. Increasing the penalty parameter might not mitigate this problem.

Failure of the BT code can also be for numerical reasons, which occur internally in the BT code. In this case we also increase the penalty parameter and reset the starting point in step (S.3), in the hope that this might improve the behavior of the BT code.

Step (S.2) of Algorithm 6.5 is reached if the BT code in step (S.1) converged successfully. We then assume the point z^k_{BT} returned by the BT code to be a local minimizer of $P_1(\cdot;\rho^k)$. If z^k_{BT} is feasible for the MPEC (1.1), we assume that we have found a solution of that MPEC. Numerically, feasibility cannot be expected to hold exactly. We therefore only require the *feasibility gap*

$$\Delta_k^{\mathrm{feas}} := \max\Big\{\max_{i=1,\dots,m}\{g_i(z^k_{\mathrm{BT}})\}, \max_{i=1,\dots,p}\{|h_i(z^k_{\mathrm{BT}})|\}, \max_{i=1,\dots,l}\{|\min\{G_i(z^k_{\mathrm{BT}}), H_i(z^k_{\mathrm{BT}})\}|\}\Big\}$$

to fall below a threshold:

$$\Delta_k^{\mathrm{feas}} \le n\varepsilon^{\mathrm{feas}}. \tag{6.19}$$

In our implementation, we have chosen

$$\varepsilon^{\mathrm{feas}} := 10^{-8}.$$

Hence, if (6.19) is satisfied, we set $z^* := z^k_{\mathrm{BT}}$ and exit the algorithm with a message signifying convergence.

If, on the other hand, (6.19) is not satisfied, we increase the penalty parameter and enter the next iteration. As the next iterate, we choose $z^{k+1} := z^k_{\mathrm{BT}}$. This is motivated by the fact that, since the BT code converged, we hope to be fairly close to a solution of the MPEC (1.1).

This also suggests that we should reinitialize $z^{\mathrm{init}} := z^k_{\mathrm{BT}}$, so that we start from this putatively better point in case of a restart step occurring in a future iteration. During our tests, this scenario occurred exactly once, and reinitializing z^{init} as described above caused slower convergence.

Since we assume z^k_{BT} to be close to a solution of the MPEC (1.1), we also do not

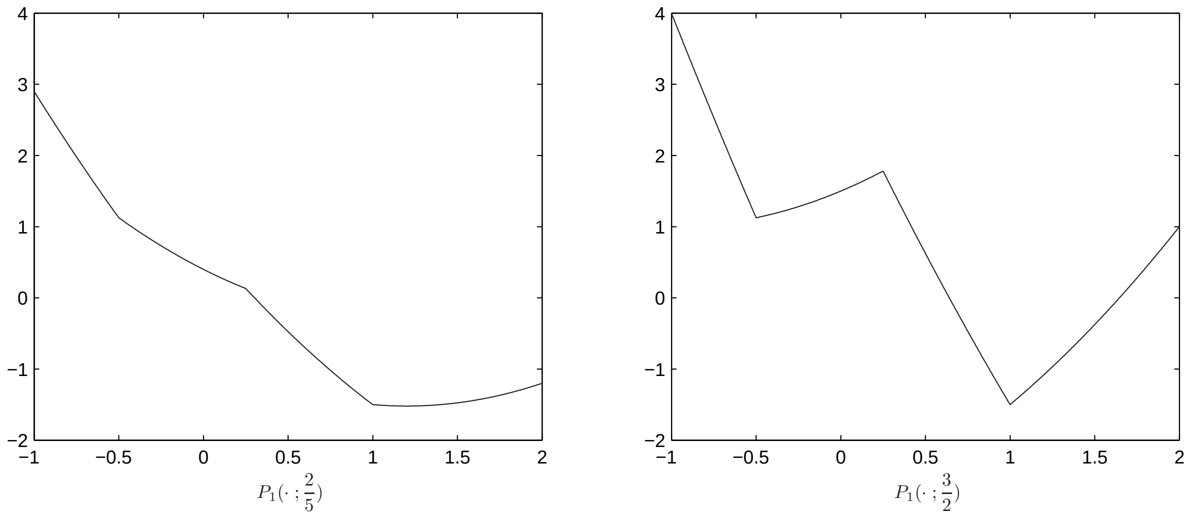

Figure 6.1: Exact l_1 MPEC-penalty function of (6.14) for $\rho = \frac{2}{5}$ and $\rho = \frac{3}{2}$.

multiply the penalty parameter by the full factor σ, but rather choose

$$\rho^{k+1} := \rho^k\big(1 + \frac{\sigma - 1}{\pi/2}\arctan(\frac{1}{10}\log_{10}\frac{\Delta_k^{\text{feas}}}{n\varepsilon^{\text{feas}}})\big).$$

This is a purely heuristical update rule, which we wish to motivate to a certain degree.

Since $\Delta_k^{\text{feas}}/n\varepsilon^{\text{feas}} > 1$ (see (6.19)), the logarithmic term is positive. The term $\log_{10}\frac{\Delta_k^{\text{feas}}}{n\varepsilon^{\text{feas}}}$ is a measure of the order by which the feasibility criterion is violated. The arc tangent was chosen because it grows reasonably fast for small values of its argument and has the constant asymptote $\pi/2$. The scaling is chosen such that $\rho^{k+1} \in (\rho^k, \sigma\rho^k]$. The factor $\frac{1}{10}$ in the argument of the arc tangent was added because it improved the behavior of the algorithm.

In summary, if the feasibility gap Δ_k^{feas} is small, i.e. on the order of $\varepsilon^{\text{feas}}$, ρ^k is only multiplied by a small factor, since we assume the penalty parameter only to be minutely too small. Additionally, the infeasibility of our iterate z_{BT}^k might only be due to numerical noise in some part of the algorithm and not because the penalty parameter ρ^k is too small. In this case, a small change in the penalty parameter might suffice to make the next iterate z_{BT}^{k+1} feasible for the MPEC (1.1).

If, however, Δ_k^{feas} is large, the penalty parameter is increased substantially. In fact, as Δ_k^{feas} tends to infinity, ρ^{k+1} tends to $\sigma\rho^k$.

To further illustrate our procedure in step (S.2), consider the penalty function of the program (6.14) from Example 6.3 for the penalty parameters $\rho = \frac{2}{5}$ and $\rho = \frac{3}{2}$ (see Figure 6.1). It is easily verified that $\mathcal{Z} = \{-\frac{1}{2}, 1\}$ is the feasible region of the MPEC (6.14) and that it attains its global minimizer at $z^* = 1$.

For $\rho = \frac{2}{5}$, however, the exact l_1 MPEC-penalty function $P_1(\cdot; \frac{2}{5})$ is minimized for $z = \frac{6}{5}$ (see graph on the left in Figure 6.1). However, this value of z is not feasible for the MPEC

(6.14). Moreover, our feasibility check (6.19) fails for $z = \frac{6}{5}$. Assuming that the BT code converged to $z^k_{\mathrm{BT}} = \frac{6}{5}$, we would increase the penalty parameter and enter the next iteration using this minimizer z^k_{BT} as the starting point for our next call to the BT code.

This starting point has the advantage that it is fairly close to the actual minimizer $z^* = 1$ of the MPEC (6.14). If we assume that the penalty parameter was increased to $\rho^{k+1} = \frac{3}{2}$ (see the graph on the right in Figure 6.1), it is reasonable to assume that the BT code will find the minimizer of $P_1(\cdot; \frac{3}{2})$ quicker starting from $z^{k+1} = \frac{6}{5}$, than it would starting from an arbitrarily chosen initial point z^{init}. This is compounded by the fact that the estimate for the minimizer is fairly accurate ($P_1(\frac{6}{5}; \frac{2}{5}) = -\frac{38}{25} = -1.52$, while $P_1(z^*; \rho^*) = -\frac{3}{2} = -1.5$ for ρ^* sufficiently large).

One might also imagine a situation where several intermediate penalty functions are minimized, taking several iterations to converge to a feasible point of the MPEC (1.1).

In step (S.3) we perform a *restart step*. Since BT failed for some reason (see the discussion above), we assume that the penalty parameter was chosen too small. We therefore increase it by the maximum allowed factor σ, reinitialize the starting point and then go on to enter the next iteration in steps (S.4) and (S.5).

The reasoning behind reinitializing the starting point is that the BT code exited in an undefined state (i.e. it did not converge). Therefore it is not clear what z^k_{BT} has to do with the problem. Imagine that $P_1(\cdot; \rho^k)$ is unbounded below. As the BT code tries to find a minimizer, its iterates diverge. Once the maximal number of iterations or function evaluations is reached, BT returns with its current iterate, which will potentially have a very large norm, and thus not contain any useful information.

Alternately to the penalty parameter being too small, the starting point might also have been chosen poorly. This might particularly then be the case, when z^{init} was chosen using (6.18). Since we count on the user having knowledge of the underlying problem, however, we did not account for this in our implementation. If one run of Algorithm 6.5 fails, one might alter z^{init} and then try again.

Finally, step (S.4) checks whether the maximal allowed penalty parameter $\rho^{\max}$ has been exceeded and exits with a failure message if this is the case. If not, step (S.5) increases the iteration counter and enters the next iteration.

6.3.1 Application to the MacMPEC Test Suite

In this section we apply Algorithm 6.5 described in the previous section to the MacMPEC test suite maintained by Leyffer [38]. This suite consists of a collection of 168 MPECs coded in AMPL. Since we only have access to the student version of AMPL (which may be downloaded free of charge [1]), we are restricted to problems of dimension 300 or less. Furthermore, we omitted those problems that contained mixed complementarity problems in their constraints. We are left with a collection of 114 problems, see Table 6.1.

For a detailed introduction into the AMPL language, see the AMPL book [23]. For information on hooking solvers to AMPL, see [26], as well as the AMPL sources and example wrappers.

The bundle trust region code was kindly provided to us by Jiří Outrata. The FORTRAN sources came in two different versions; the standard nonconvex BT code and a variant which explicitly handles box constraints. The wrapper Algorithm 6.5 was written in C++. Two programs, each using a version of the BT code, were compiled on a SuSE Linux 9.0 [72] installation using the Intel FORTRAN 8.1 and Intel C++ 8.1 compilers [33]. The Intel compilers were invoked with the optimization options

`-mp -O3`.

The former option enforces IEEE 754 compliance, while the latter turns on the greatest degree of optimization. If IEEE 754 compliance is not enforced, floating point operations may be performed by modern SSE or SSE2 instructions, which do not conform to the IEEE 754 standard. This leads to unpredictable behavior of our software.

Additionally, the FORTRAN and C++ compilers were invoked with the options

`-arch pn4` and `-mcpu=pentium4`,

respectively, causing the compilers to generate code optimized for Pentium 4 processors.

Table 6.1 collects the results obtained with Algorithm 6.5. The first column contains the name of the problem. An asterisk denotes that the problem is a maximization problem. In particular, this means that higher values of the solution are better. The second column states the dimension of the problem. Information such as number of constraints, split up into different types of constraints, is not contained in the table. The reader is referred to [38] for such information. The third column then gives the best function value obtained by Leyffer [38].

The next six columns contain information on the performance of Algorithm 6.5 using the nonconvex bundle trust region code, which does not explicitly handle box constraints. The version that exploits box constraint information on the variables did not fare very well in our situation. It is therefore not contained in Table 6.1. For a qualitative impression of the variant with box constraints, see the performance profiles in Figures 6.2 and 6.3, and their discussion.

The first of these columns ('solution') contains the function value of the objective function at z^*, the vector that Algorithm 6.5 converged to. The second column ('CPU') states the CPU time in seconds taken to obtain this solution. The value given is the average of three runs performed on a dual AMD Athlon MP 2800+ with 3.5GB of RAM running SuSE Linux 9.0 with the vanilla 2.4.21-243-smp4G kernel installed.[1] The time of a single run was taken from the `_solve_user_time` variable in AMPL.

The third column ('eval.') states the number of function evaluations required. No differentiation is made between function, constraint, or gradient evaluations, since the BT codes simultaneously require one function and one subgradient evaluation of $P_1(\cdot;\rho)$ and

[1] Although we used the Intel compilers, we found the (pre-compiled) AMPL routines to cause segmentation faults on Intel Pentium 4 systems for some of the MacMPEC problems. For this reason we opted to run our tests on an AMD system, resulting in the somewhat incongruous pairing of Intel optimized code and an AMD CPU.

this entails determining the values a whole complement of objective, constraint, and their gradients.

The fourth column (‘k’) indicates how many outer iterations of Algorithm 6.5 were necessary before converging. The fifth column (‘ρ^k’) contains the final value of the penalty parameter, while the sixth column (‘Δ_k^{feas}’) denotes the feasibility gap at the solution.

If the algorithm failed, this is indicated by a dash in all six columns. The reason for failure is given by the following table:

1	BT code converged, but solution infeasible.
2	BT terminated with maximal number of iterations.
3	BT terminated with maximal number of function evaluations.
4	BT terminated with numerical problems.

Note that these errors only state the reason for termination of the final iteration of Algorithm 6.5. No information why earlier iterations failed is given. Experience has shown, however, that in most cases, the reason does not change over iterations of Algorithm 6.5.

Table 6.1: Results obtained with Algorithm 6.5.

			nonconvex bundle trust region code (BT)					
Problem	dim.	solution	solution	CPU [s]	eval.	k	ρ^k	Δ_k^{feas}
bar-truss-3	29	10166.6	—	—	—	—	—	—[2]
bard1	5	17	17	0.013	36	1	5	7.24052e-09
bard2*	12	-6598	6598	0.498	2783	3	254.606	4.30619e-13
bard3	6	-12.6787	-12.6787	0.11567	454	7	84.537	8.66922e-09
bard1m	5	17	17	0.012667	50	1	5	1.48674e-11
bard3m	6	-12.6787	-12.6787	0.079333	330	6	35.0066	8.57793e-09
bilevel1	10	-60	5	0.010667	46	1	5	2.83575e-15
bilevel2	16	-6600	-6598	0.18967	896	4	603.435	9.13528e-15
bilevel2m	16	-6600	-6598	0.18867	896	4	603.435	9.13528e-15
bilevel3	12	-12.6787	-6.89251	0.688	3555	7	4085.27	9.96276e-09
bilin*	8	5.6	12.9449	0.026667	1697	2	50	2.56206e-10
dempe	4	31.25	30.2964	0.024	143	6	770.607	7.87376e-09
design-cent-1*	12	1.86065	1.86065	0.03	142	1	5	2.61489e-09
design-cent-2*	13	3.48382	—	—	—	—	—	—[3]
design-cent-21*	13	3.48382	—	—	—	—	—	—[3]
design-cent-3*	15	(I)	3.72337	0.14867	689	1	5	1.26289e-09
design-cent-31*	15	3.72337	3.72337	0.14567	672	1	5	1.33984e-09
design-cent-4*	22	3.0792	3.0792	0.062333	245	1	5	7.16348e-12
desilva	6	-1	-1	0.013333	59	1	5	9.56584e-11
df1	2	0	4.62951e-17	0.0063333	27	2	6.54782	1.69332e-16

Table 6.1: *continued.*

			nonconvex bundle trust region code (BT)					
Problem	dim.	solution	solution	CPU [s]	eval.	k	ρ^k	Δ_k^{feas}
ex9.1.1	13	-6	-13	0.026	146	1	5	5.46571e-16
ex9.1.2	10	-6.25	-6.25	0.014667	73	1	5	2.80393e-16
ex9.1.3	23	-6	-6	0.080333	2076	2	50	1.20676e-17
ex9.1.4	10	-37	-37	0.014667	86	1	5	1.77636e-15
ex9.1.5	13	-1	-1	0.014667	71	1	5	1.70804e-17
ex9.1.6	14	-15	-49	0.028333	160	1	5	5.07531e-16
ex9.1.7	17	-6	-6	0.050333	1782	2	50	1.11022e-16
ex9.1.8	14	-3.25	-3.25	0.0096667	43	1	5	4.44089e-16
ex9.1.9	12	3.11111	—	—	—	—	—	—[1]
ex9.1.10	14	-3.25	-3.25	0.01	43	1	5	4.44089e-16
ex9.2.1	10	25	17	0.06	300	2	23.7458	1.39957e-12
ex9.2.2	10	100	100	0.025333	117	2	24.1911	2.27417e-10
ex9.2.3	16	-55	5	0.017333	81	1	5	8.88178e-16
ex9.2.4	8	0.5	0.5	0.034	140	1	5	2.42729e-11
ex9.2.5	8	9	9.8	0.03	124	1	5	2.08327e-12
ex9.2.6	16	-1	-1	0.0083333	36	1	5	1.10424e-09
ex9.2.7	10	25	17	0.059	300	2	23.7458	1.39957e-12
ex9.2.8	6	1.5	1.5	0.010667	58	3	11.7719	7.98837e-11
ex9.2.9	9	2	2	0.015667	85	1	5	9.86865e-17
flp2	4	0	8.93985e-12	0.013667	39	1	5	1.07681e-12
flp4-1	80	0	-7.47133e-13	0.018333	38	1	5	3.70971e-16
flp4-2	110	0	-7.84755e-09	0.028	35	1	5	2.09941e-12
flp4-3	140	0	6.46329e-12	0.11933	61	1	5	4.63977e-17
flp4-4	200	0	-2.27209e-11	0.31733	71	1	5	1.64191e-15
gnash10	13	-230.823	-230.823	0.52633	2165	2	50	4.60458e-11
gnash11	13	-129.912	-129.912	0.458	1843	2	50	1.32817e-11
gnash12	13	-36.9331	-36.9331	0.43933	1738	2	50	5.23889e-13
gnash13	13	-7.06178	-7.06178	0.408	1628	2	50	2.96707e-11
gnash14	13	-0.179046	-0.179046	0.394	1630	2	50	1.89243e-12
gnash15	13	-354.699	-354.699	0.51667	2251	2	50	7.79001e-13
gnash16	13	-241.442	-241.442	0.45233	2263	2	50	7.51536e-16
gnash17	13	-90.7491	-90.7491	0.14767	1033	2	50	1.19076e-11
gnash18	13	-25.6982	-25.294	0.33567	1674	2	50	2.13177e-10
gnash19	13	-6.11671	-6.11671	0.402	1642	2	50	1.60829e-12
gauvin	3	20	20	0.011	35	1	5	1.62631e-10
hs044-i	20	15.6178	17.0901	0.095	446	3	66.3408	9.86517e-12
incid-set1-8	117	3.816e-17	—	—	—	—	—	—[3]

Table 6.1: *continued.*

			nonconvex bundle trust region code (BT)					
Problem	dim.	solution	solution	CPU [s]	eval.	k	ρ^k	Δ_k^{feas}
incid-set1c-8	117	3.816e-17	—	—	—	—	—	—[4]
incid-set2-8	117	0.004518	—	—	—	—	—	—[2]
incid-set2c-8	117	0.005471	—	—	—	—	—	—[4]
jr1	2	0.5	0.5	0.0086667	23	1	5	3.2338e-09
jr2	2	0.5	0.5	0.0066667	20	1	5	4.94481e-09
kth1	2	0	2.77556e-16	0.0043333	8	1	5	5.55112e-17
kth2	2	0	1	0.0033333	16	1	5	2.68061e-09
kth3	2	0.5	1	0.015667	40	1	5	3.26993e-09
liswet1-050	152	0.01399	—	—	—	—	—	—[2]
nash1	6	7.88861e-30	2.10267e-13	0.0093333	25	1	5	9.9689e-13
outrata31	5	3.2077	3.2077	0.029	136	3	24.9124	8.0566e-11
outrata32	5	3.4494	3.7077	0.063667	254	2	18.208	9.68983e-09
outrata33	5	4.60425	4.60426	0.049333	178	2	9.16322	7.16122e-09
outrata34	5	6.59268	6.59271	0.071667	318	6	76.491	2.05598e-09
pack-comp1-8	75	0.6	—	—	—	—	—	—[4]
pack-comp1c-8	75	0.6	—	—	—	—	—	—[3]
pack-comp1p-8	107	0.6	—	—	—	—	—	—[2]
pack-comp2-8	75	0.673117	—	—	—	—	—	—[4]
pack-comp2c-8	75	0.673458	—	—	—	—	—	—[3]
pack-comp2p-8	107	0.673117	—	—	—	—	—	—[2]
pack-rig1-8	49	0.787932	—	—	—	—	—	—[2]
pack-rig1c-8	49	0.7883	—	—	—	—	—	—[3]
pack-rig1p-8	58	0.787932	—	—	—	—	—	—[4]
pack-rig2-8	49	0.780404	—	—	—	—	—	—[4]
pack-rig2c-8	49	0.799306	—	—	—	—	—	—[2]
pack-rig2p-8	58	0.780404	—	—	—	—	—	—[4]
pack-rig3-8	49	0.735202	—	—	—	—	—	—[2]
pack-rig3c-8	49	0.753473	—	—	—	—	—	—[3]
portfl-i-1	87	1.502e-05	—	—	—	—	—	—[4]
portfl-i-2	87	1.457e-05	—	—	—	—	—	—[4]
portfl-i-3	87	6.265e-06	—	—	—	—	—	—[4]
portfl-i-4	87	2.177e-06	—	—	—	—	—	—[4]
portfl-i-6	87	2.361e-06	—	—	—	—	—	—[4]
qpec-100-1	105	0.0990028	0.403459	184.48	12907	2	50	2.03838e-16
qpec-100-2	110	-6.26049	-0.824514	207.75	14108	2	50	4.47634e-15
qpec-100-3	110	-5.48287	-4.52547	261.97	16317	2	50	8.6187e-16
qpec-100-4	120	-3.60073	-0.978443	804.31	41736	4	5000	3.75029e-14

Table 6.1: *continued.*

			nonconvex bundle trust region code (BT)					
Problem	dim.	solution	solution	CPU [s]	eval.	k	ρ^k	Δ_k^{feas}
qpec-200-1	210	-1.93483	-1.83465	2873	34890	2	50	1.6459e-16
qpec-200-2	220	-24.0299	-3.7045	4935.7	59693	3	500	3.13392e-18
qpec-200-3	220	-1.95341	—	—	—	—	—	—[2]
qpec-200-4	240	-6.19323	—	—	—	—	—	—[4]
qpec1	30	80	80	0.021333	36	1	5	2.15143e-14
qpec2	30	45	45	0.0096667	28	1	5	2.46833e-10
ralph1	2	0	0	0.004	6	1	5	0
ralph2	2	0	—	—	—	—	—	—[4]
ralphmod	104	-683.033	0	176.53	11047	3	500	0
scholtes1	3	2	2	0.014333	36	1	5	3.397e-11
scholtes2	3	15	15.0001	0.029333	100	3	95.0174	3.91997e-10
scholtes3	2	0.5	1	0.0043333	7	1	5	4.99975e-10
scholtes4	3	-3.07336e-07	-1.44329e-15	0.005	13	1	5	6.29126e-16
scholtes5	3	1	1	0.021667	62	1	5	3.56206e-13
sl1	8	0.0001	—	—	—	—	—	—[2]
stackelberg1	3	-3266.67	-3266.67	0.033667	420	2	50	5.65407e-11
tap-09	86	109.143	—	—	—	—	—	—[1]
tap-15	194	184.295	—	—	—	—	—	—[1]
water-net	52	929.169	—	—	—	—	—	—[4]
water-FL	169	3411.92	—	—	—	—	—	—[4]

A few words on Table 6.1 are in order. Discounting the cases where Algorithm 6.5 failed, it found a better solution than the known documented solution [38] in 7 cases, while converging to a point with lesser function value in 19 instances.

It should be noted at this point that it is not clear whether any given point is a local minimizer (either of $P_1(\cdot;\rho)$ or of the MPEC (1.1)). All that may be said is that it is feasible to within numerical accuracy and its objective has the given function value.

It is of interest that even though the feasibility check (6.19) only requires feasibility of the order of 10^{-8}, in most cases, far greater accuracy was obtained. This indicates that the termination criterion $\varepsilon^{\mathrm{BT}}$ of the BT code might be increased. However, this lead to worse performance of Algorithm 6.5.

As seen, in particular, in the cases of the 'qpec' and the unsolved 'pack' and 'portfl' problems, it seems that higher dimensions (> 100) pose difficulties for our approach and the bundle trust region algorithm.

In the case of 'bilin', Algorithm 6.5, using the standard BT code with $k_{\max} = 20$, converges to a point with objective function value 13, even better than the 12.9449 for

$k_{\max} = 25$ given in Table 6.1. Algorithm 6.5, using the BT code with box constraints, even converges to a point with objective 18.4 for this example.

Other cases where the BT code with box constraints performs better than the standard BT code are 'kth2' (1.10734e-08), 'kth3' (0.5), 'portfl-i-4' (2.17734e-06), and 'ralph2' (3.53133e-15). The numbers in brackets are the attained objective function values. In all cases, Algorithm 6.5, using the BT code with box constraints, matches the documented known solutions.

By and large, however, these few instances of the BT code with box constraints performing better than the BT code without, is far outweighed by the fact that the BT code with box constraints fails on many more problems. We suspect that the better performance of the standard BT code is due to a superior quality in implementation, rather than theoretical reasons.

To gain a better understanding of the performance of Algorithm 6.5 in its two variants, we avail ourselves of the performance profiles introduced by Dolan and Moré [13]. To this end, we define the set of solvers $\mathcal{S} := \{\mathrm{BT}, \mathrm{BT\ w/box\ constraints}\}$ and denote by $\mathcal{P}$ the set of all problems contained in Table 6.1. We then define the *performance ratio*

$$r_{p,s} := \begin{cases} \frac{t_{p,s}}{\min_{\sigma \in \mathcal{S}}\{t_{p,\sigma}\}} & : s \text{ converged}, \\ r_M & : s \text{ failed}, \end{cases}$$

for all $p \in \mathcal{P}$ and $s \in \mathcal{S}$, where $t_{p,s}$ denotes some measure of performance of the solver s on problem p. It must hold that $r_M \geq r_{p,s}$ for all $p \in \mathcal{P}$ and $s \in \mathcal{S}$ and that $r_M = r_{p,s}$ if and only if the solver s failed on problem p.

We then define the *performance profile*

$$\rho_s(\tau) := \frac{1}{|\mathcal{P}|} \big| \{ p \in \mathcal{P} \mid \log_2 r_{p,s} \leq \tau \} \big|$$

for all $s \in \mathcal{S}$.

In Figure 6.2, we plotted the performance profiles of the two variants of Algorithm 6.5, using CPU time as a measure of performance.

Two useful pieces of information may be gleaned from the performance profile. Firstly, the probability that a given solver $s \in \mathcal{S}$ will solve a problem is given by

$$\rho_s^* := \lim_{\tau \nearrow r_M} \rho_s(\tau).$$

In the graph (see Figure 6.2), this is the value at which the graph "evens out."

The second bit of information stems from the fact that performance profiles intrinsically compare solvers with each other. Plotted against the horizontal axis is the probability that the solver in question will solve a given problem with a performance of at most 2^τ times worse than the performance of the best solver. Simplistically speaking, the solver with the top-most graph performs the best.

As can be seen clearly in the two graphs, the BT code explicitly containing box constraints cannot compare to the standard nonconvex BT code, since it only solves 40% of

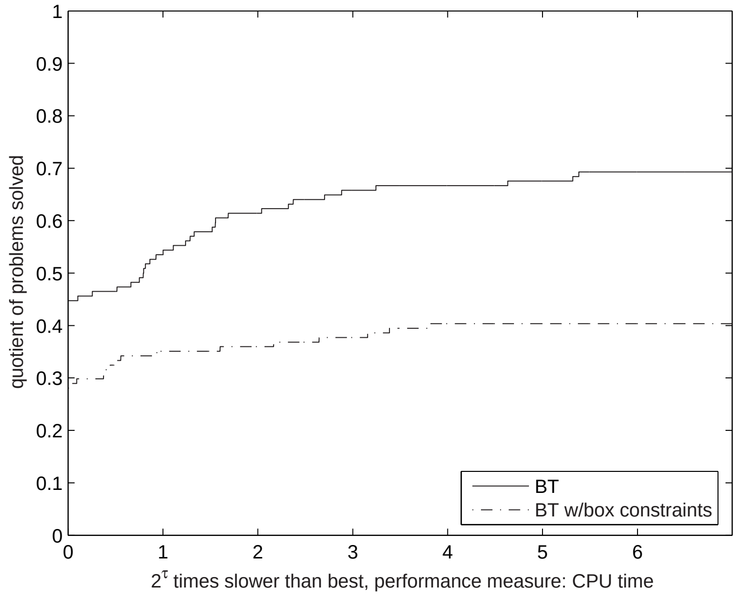

Figure 6.2: Performance profiles of Algorithm 6.5 for the MacMPEC test suite.

the problems, while the latter solves 69%. It should be noted, however, that 69% does not, in turn, compare to other methods. For instance, a method implemented by Leyffer using the filterSQP code [39] solves nearly 100% of the MacMPEC test suite.

As mentioned before, it seems that problems of higher dimension pose a bigger problem to our solvers. To verify this, we restrict the problem set $\mathcal{P}$ to those problems of dimension 30 or less and plot the performance profiles of this subset of problems, see Figure 6.3. In this case, Algorithm 6.5 with the standard BT code solves 92% of the problems, while the BT code with box constraints solves 55%.

We found this high quotient of lower dimensional problems solved somewhat surprising. The quotient, as well as the accuracy of the solutions, far surpassed our expectations of Algorithm 6.5. This is much more a reflection of the quality of the BT code, however, than it is of the merit of Algorithm 6.5. However, the results we obtained indicate that our exact penalty result, Proposition 6.1 stands up to our test suite. In particular, MPEC-calmness seems to be a reasonable assumption.

In closing, we would like to remind the reader once again that Algorithm 6.5 should not be viewed as a serious contender for solving MPECs. Rather, any one of the algorithmic approaches reviewed in Section 6.1 should be chosen.

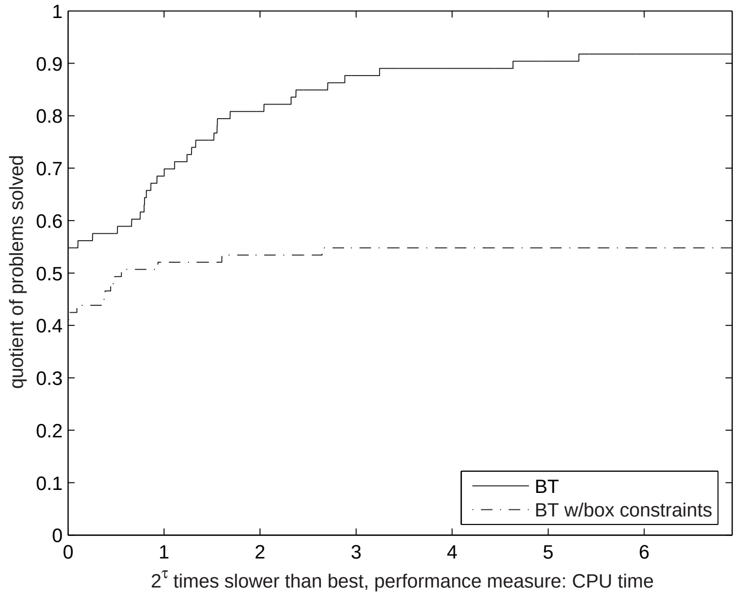

Figure 6.3: Performance profiles of Algorithm 6.5 for problems of dimension 30 or less.

Final Remarks

This thesis contains an exhaustive discussion of constraint qualifications and stationarity concepts for mathematical programs with equilibrium constraints. Great care was taken to deliver a concise account of the existing first order theory pertaining to these constraint qualifications and stationarity concepts.

Two chapters were dedicated to the discussion of constraint qualifications and how they apply to MPECs. First, standard constraint qualifications were investigated in the connection with MPECs. It was quickly found that all but the weakest constraint qualifications were too restrictive for MPECs. One exception was the Guignard CQ, elevating it to a special position for MPECs, whereas it has been largely overlooked in the standard nonlinear programming community.

Having accepted that standard constraint qualifications are not suited to the investigation of MPECs, we introduced several variants of standard constraint qualifications, specifically tailored to MPECs. Among these was the hitherto unknown MPEC-Guignard CQ, which we demonstrated to be the weakest of all known constraint qualifications for MPECs. It was found that relationships that existed between standard constraint qualifications also held for their MPEC counterparts.

Next, we investigated stationarity concepts and their relationship to the constraint qualifications we had introduced. Here too, it was quickly found that a standard KKT point, called a strongly stationary point by the MPEC community, was a fairly strong stationarity concept that was only a first order condition under the stronger constraint qualifications MPEC-LICQ, MPEC-SMFCQ, and Guignard CQ.

Following this realization, a whole set of KKT-type conditions tailored to MPECs were investigated. The same novel approach that yielded strong stationarity to be a necessary optimality condition under MPEC-LICQ also yielded that A-stationarity was a necessary optimality condition under MPEC-MFCQ. Using yet another simple technique, it could also be shown that A-stationarity was a necessary optimality condition under MPEC-ACQ and MPEC-GCQ.

Unfortunately, A-stationarity is among the very weakest stationarity concepts. Research by other people, using different techniques, had obtained C-stationarity to be a necessary first order condition under MPEC-MFCQ. C-stationarity, though qualitatively stronger than A-stationarity, is still very weak.

We therefore focused our attention on M-stationarity. Despite its strong nature and the very weak nature of MPEC-Guignard CQ, we were able to show that M-stationarity is a necessary first order conditions under MPEC-Guignard CQ.

To acquire this result, two paths were presented. The original, longer, path gives insight into the heart of the problem but loses track of its goal. It does, however, yield a strong exact penalty result as a by-product. The second proof, on the other hand, is short, elegant and concise. Both proofs, however, have in common that they require the calculus of Mordukhovich.

For theoretical purposes, this result has wide-spread ramifications: The investigation of weaker stationarity concepts, among which are A-, and C-stationarity, has become moot.

Bibliography

[1] *AMPL Student Version*. Internet Website.
`http://www.ampl.com/DOWNLOADS/details.html#StudentEd`.

[2] M. Anitescu, *Global convergence of an elastic mode approach for a class of mathematical programs with complementarity constraints*, tech. report, Argonne National Laboratory, 2004. Preprint ANL/MCS-P1143-0404.

[3] ——, *On solving mathematical programs with complementarity constraints as nonlinear programs*, tech. report, Argonne National Laboratory, 2004. Preprint ANL/MCS-P864-1200.

[4] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, New York, second ed., 1993.

[5] M. S. Bazaraa and C. M. Shetty, *Foundations of Optimization*, vol. 122 of Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, Heidelberg, New York, 1976.

[6] J. V. Burke, *Calmness and exact penalization*, SIAM Journal on Control and Optimization, 29 (1991), pp. 493–497.

[7] Y. Chen and M. Florian, *The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions*, Optimization, 32 (1995), pp. 193–209.

[8] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983.

[9] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust-Region Methods*, MPS/SIAM Series on Optimization, SIAM, Philadelphia, 2000.

[10] R. W. Cottle, J.-S. Pang, and R. E. Stone, *The Linear Complementarity Problem*, Computer Science and Scientific Computing, Academic Press, San Diego, 1992.

[11] A.-V. DeMiguel, M. P. Friedlander, F. J. Nogales, and S. Scholtes, *An interior-point method for MPECs based on strictly feasible relaxations.* Preprint ANL/MCS-P1150-0404, Mathematics and Computer Science Division, Argonne National Laboratory, April 2004.

[12] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht/Boston/London, 2002.

[13] E. D. Dolan and J. J. Moré, *Benchmarking optimization software with performance profiles*, Mathematical Programming, Series A, 91 (2002), pp. 201–213.

[14] F. Facchinei, H. Jiang, and L. Qi, *A smoothing method for mathematical programs with equilibrium constraints*, Mathematical Programming, 85 (1999), pp. 107–134.

[15] M. L. Flegel and C. Kanzow, *Optimality conditions for mathematical programs with equilibrium contraints: Fritz John and Abadie-type approaches.* Institute of Applied Mathematics and Statistics, University of Würzburg, Preprint 245, May 2002.

[16] ——, *An Abadie-type constraint qualification for mathematical programs with equilibrium constraints.* Institute of Applied Mathematics and Statistics, University of Würzburg, Preprint, November 2002.

[17] ——, *A Fritz John approach to first order optimality conditions for mathematical programs with equilibrium constraints*, Optimization, 52 (2003), pp. 277–286.

[18] ——, *A direct proof for M-stationarity under MPEC-GCQ for mathematical programs with equilibrium constraints.* Institute of Applied Mathematics and Statistics, University of Würzburg, Preprint, November 2004.

[19] ——, *On the Guignard constraint qualification for mathematical programs with equilibrium constraints.* Institute of Applied Mathematics and Statistics, University of Würzburg, Preprint, January 2004.

[20] ——, *Abadie-type constraint qualification for mathematical programs with equilibrium constraints*, Journal of Optimization Theory and Applications, 124 (2005).

[21] ——, *On M-stationarity for mathematical programs with equilibrium constraints.* Journal of Mathematical Analysis and Applications, to appear.

[22] M. L. Flegel, C. Kanzow, and J. Outrata, *Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints.* Institute of Applied Mathematics and Statistics, University of Würzburg, Preprint, October 2004.

[23] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL—A Modeling Language for Mathematical Programming*, Duxbury, Canada, 2nd ed., 2003.

[24] O. Fujiwara, S.-P. Han, and O. L. Mangasarian, *Local duality of nonlinear programs*, SIAM Journal on Control and Optimization, 22 (1984), pp. 162–169.

[25] M. Fukushima and J.-S. Pang, *Convergence of a smoothing continuation method for mathematical programs with complementarity constraints*, in Ill-Posed Variational Problems and Regularization Techniques, M. Thera and R. Tichatschke, eds., vol. 477 of Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin/Heidelberg, 1999, pp. 99–110.

[26] D. M. Gay, *Hooking your solver to AMPL.* Technical Report 97-4-06, Computing Sciences Research Center, Bell Laboratories, April 1997.

[27] C. Geiger and C. Kanzow, *Theorie und Numerik restringierter Optimierungsaufgaben*, Springer-Verlag, Berlin/Heidelberg, 2002.

[28] P. E. Gill, W. Murray, and M. A. Saunders, *User's guide for SNOPT 5.3: A Fortran apckage for large-scale nonlinear programming*, tech. report, Department of Mathematics, University of California, San Diego, 1997. Report NA 97-5.

[29] F. J. Gould and J. W. Tolle, *A necessary and sufficient qualification for constrained optimization*, SIAM Journal on Applied Mathematics, 20 (1971), pp. 164–172.

[30] M. Guignard, *Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space*, SIAM Journal on Control, 7 (1969), pp. 232–241.

[31] R. Henrion, A. Jourani, and J. Outrata, *On the calmness of a class of multifunctions*, SIAM Journal on Optimization, 13 (2002), pp. 603–618.

[32] X. M. Hu and D. Ralph, *Convergence of a penalty method for mathematical programming with complementarity constraints*, Journal of Optimization Theory and Applications, 123 (2004), pp. 365–390.

[33] *Intel Compilers*. Internet Website.
`http://www.intel.com/software/products/compilers/index.htm`.

[34] H. Jiang and D. Ralph, *Smooth SQP methods for mathematical programs with nonlinear complementarity constraints*, SIAM Journal of Optimization, 10 (2000), pp. 779–808.

[35] K. C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Springer-Verlag, Berlin, New York, 1985.

[36] T. G. Kolda, R. M. Lewis, and V. Torczon, *Optimization by direct search: New perspectives on some classical and modern methods*, SIAM Review, 45 (2003), pp. 385–482.

[37] J. Kyparisis, *On uniqueness of Kuhn-Tucker multipliers in nonlinear programming*, Mathematical Programming, 32 (1985), pp. 242–246.

[38] S. Leyffer, *MacMPEC Test Suite*. Internet Website. http://www-unix.mcs.anl.gov/~leyffer/MacMPEC/.

[39] ——, *Complementarity constraints as nonlinear equations: Theory and numerical experience*. Preprint ANL/MCS-P1054-0603, June 2003.

[40] P. D. Loewen, *Optimal Control via Nonsmooth Analysis*, vol. 2 of CRM Proceedings & Lecture Notes, American Mathematical Society, Providence, RI, 1993.

[41] M. A. López, *Personal communication*. French-German-Polish Conference on Optimization, Cottbus, Germany, September 2002.

[42] Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK, 1996.

[43] Z.-Q. Luo and P. Tseng, *Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem*, SIAM Journal on Optimization, 2 (1992), pp. 43–54.

[44] O. L. Mangasarian, *Nonlinear Programming*, Classics in Applied Mathematics, SIAM, Philadelphia, Classics ed., 1994. Identical to the 1969 book published by McGraw-Hill.

[45] O. L. Mangasarian and S. Fromovitz, *The Fritz John optimality necessary conditions in the presence of equality and inequality constraints*, Journal of Mathematical Analysis and Applications, 17 (1967), pp. 37–47.

[46] B. S. Mordukhovich, *Approximation Methods in Problems of Optimization and Control (in Russian)*, Nauka, Moscow, 1988.

[47] ——, *Generalized differential calculus for nonsmooth and set-valued mappings*, Journal of Mathematical Analysis and Applications, 183 (1994), pp. 250–288.

[48] ——, *Necessary conditions in nonsmooth minimization via lower and upper subgradients*, Set-Valued Analalysis, 12 (2004), pp. 163–193.

[49] L. Nazareth and P. Tseng, *Gilding the lily: A variant of the Nelder-Mead algorithm based on golden-section search*, Computational Optimization and Applications, 22 (2002), pp. 133–144.

[50] J. A. Nelder and R. Mead, *A simplex method for function minimization*, Computer Journal, 7 (1965), pp. 308–313.

[51] *NEOS Server for Optimization*. Internet Website. http://www-neos.mcs.anl.gov/.

[52] J. V. Outrata, *On optimization problems with variational inequality constraints*, SIAM Journal on Optimization, 4 (1994), pp. 340–357.

[53] ——, *Optimality conditions for a class of mathematical programs with equilibrium constraints*, Mathematics of Operations Research, 24 (1999), pp. 627–644.

[54] J. V. Outrata, M. Kočvara, and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, Nonconvex Optimization and its Applications, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[55] J.-S. Pang and M. Fukushima, *Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints*, Computational Optimization and Applications, 13 (1999), pp. 111–136.

[56] D. W. Peterson, *A review of constraint qualifications in finite-dimensional spaces*, SIAM Review, 15 (1973), pp. 639–654.

[57] D. Ralph, *Personal communication.* Third International Conference on Complementarity Problems, Cambridge, England, July 2002.

[58] D. Ralph and S. J. Wright, *Some properties of regularization and penalization schemes for MPECs.* Optimization Technical Report 03-04, Computer Sciences Department, University of Wisconsin-Madison, April 2004.

[59] S. M. Robinson, *Some continuity properties of polyhedral multifunctions*, Mathematical Programming Study, 14 (1981), pp. 206–214.

[60] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.

[61] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, vol. 317 of A Series of Comprehensive Studies in Mathematics, Springer, Berlin, Heidelberg, 1998.

[62] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity*, Mathematics of Operations Research, 25 (2000), pp. 1–22.

[63] S. Scholtes, *Convergence properties of a regularization scheme for mathematical programs with complementarity constraints*, SIAM Journal on Optimization, 11 (2001), pp. 918–936.

[64] S. Scholtes and M. Stöhr, *Exact penalization of mathematical programs with equilibrium constraints*, SIAM Journal on Control and Optimization, 37 (1999), pp. 617–652.

[65] H. Schramm and J. Zowe, *A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results*, SIAM Journal on Optimization, 2 (1992), pp. 121–152.

[66] P. Spellucci, *Numerische Verfahren der nichtlinearen Optimierung*, Internationale Schriftenreihe zur Numerischen Mathematik: Lehrbuch, Birkhäuser, Basel/Boston/Berlin, 1993.

[67] D. Stewart and M. Anitescu, *Computing optimal controls for friction problems.* Talk at the First MPS International Conference on Continuous Optimization, Troy, NY, August 2004.

[68] D. E. Stewart, *Rigid-body dynamics with friction and impact*, SIAM Review, 42 (2000), pp. 3–39.

[69] D. E. Stewart and J. C. Trinkle, *An implicit time-stepping scheme for rigid body dynamics with ineqlastic collisions and Coulomb friction*, International Journal for Numerical Methods in Engineering, 39 (1996), pp. 2673–2691.

[70] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions I*, vol. 163 of Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Springer-Verlag, Berlin/Heidelberg/New York, 1970.

[71] M. Stöhr, *Nonsmooth Trust Region Methods and Their Applications to Mathematical Programs with Equilibrium Constraints*, PhD thesis, Universität Fridericiana zu Karlsruhe, 1999.

[72] *SuSE Linux.* Internet Website. `http://www.suse.com`.

[73] J. S. Treiman, *Lagrange multipliers for nonconvex generalized gradients with equality, inequality, and set constraints*, SIAM Journal on Control and Optimization, 37 (1999), pp. 1313–1329.

[74] P. Tseng, *Fortified-descent simplicial search method: A general approach*, SIAM Journal on Optimization, 10 (1999), pp. 269–288.

[75] M. H. Wright, *Direct search methods: Once scorned, now respectable*, in Numerical Analysis 1995, D. F. Griffiths and G. A. Watson, eds., vol. 344 of Pitman Research Notes in Mathematics, Boca Raton, Florida, 1996, Dundee Biennial Conference in Numerical Analysis, CRC Press, pp. 191–208.

[76] J. J. Ye, *Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints*, SIAM Journal on Optimization, 10 (2000), pp. 943–962.

[77] ——, *Personal communication.* By email, October 2003.

[78] ——, *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints*, Journal on Mathematical Analysis and Applications, (2005).

[79] J. J. Ye and X. Y. Ye, *Necessary optimality conditions for optimization problems with variational inequality constraints*, Mathematics of Operations Research, 22 (1997), pp. 977–997.

[80] J. J. Ye and D. L. Zhu, *Multiobjective optimization problem with variational inequality constraints*, Mathematical Programming, 96 (2003), pp. 139–160.

[81] Q. J. Zhu, *Necessary conditions for constrained optimization problems in smooth banach spaces and applications*, SIAM Journal on Optimization, 12 (2002), pp. 1032–1047.