Mathematical Programs with Complementarity Constraints and Related Problems

Alexandra Schwartz
email: schwartz@gsc.tu-darmstadt.de

TECHNISCHE UNIVERSITÄT DARMSTADT

Graduate School CE

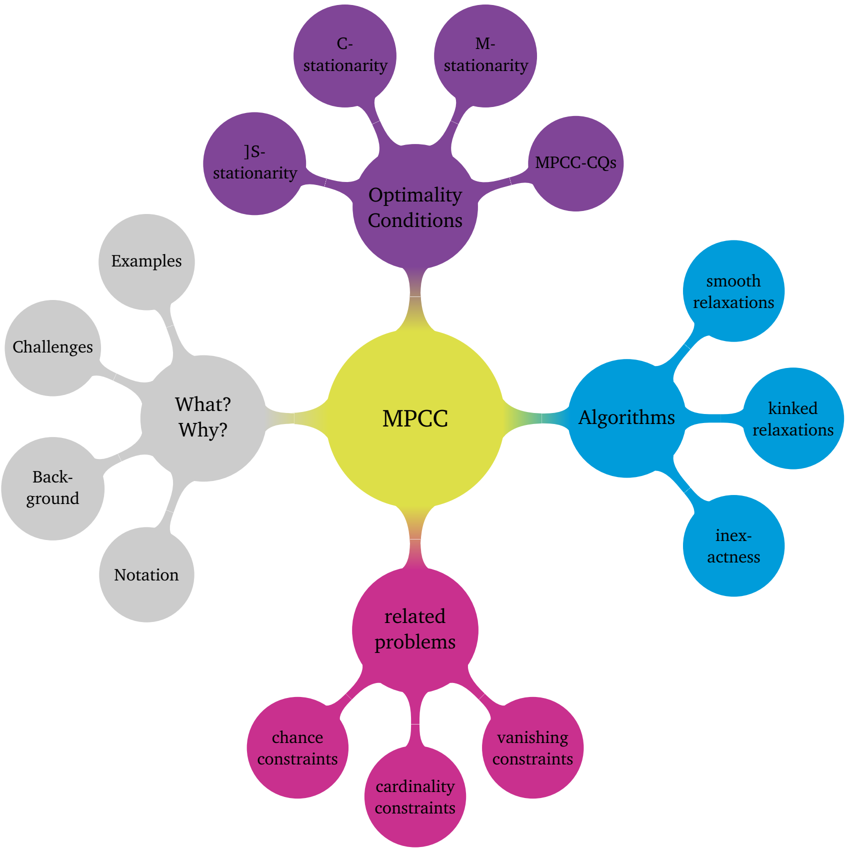
February 26, 2018

Abstract

These notes are meant to accompany a short course on mathematical programs with complementarity constraints (MPCC) and related problem classes given at the winter school *Modern Methods in Nonsmooth Optimization*. Since this course consists of only four lectures, we focus on only a few key aspects of the theoretical properties of this problem class and a few solution algorithms and provide references for alternative approaches. The contents covered in this lecture are admittedly partly chosen due to personal preference, but should also provide a good first step into the world of MPCCs.

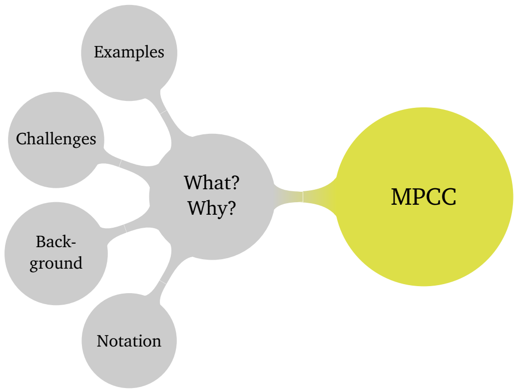
As general references for mathematical programs with complementarity constraints, I would like to mention the books [22] by Luo, Pang, and Ralph and [26] by Outrata, Kocvara, and Zowe. But of course there are also many other books, which at least partially cover MPCCs such as [9], which focuses on general bilevel programs, or [17], which provides Newton-type methods, which can also be applied to MPCCs.

For a more detailed background on the tools from variational analysis, I would like to mention the book [27] by Rockafellar and Wets, the books [23, 24] from Mordukhovich, [8] from Clarke and [29] by Schirotzek.



Contents

1 Introduction



Throughout this lecture we consider general *mathematical programs with complementarity constraints* (MPCC), which are of the form

$$\min_x f(x) \quad \text{s.t.} \quad \begin{aligned} & g(x) \leq 0, h(x) = 0, \\ & 0 \leq G(x) \perp H(x) \geq 0, \end{aligned} \tag{1.1}$$

where

- $x \in \mathbb{R}^n$ is the finite dimensional optimization variable,
- $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function,
- $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are the standard inequality and equality constraints
- and $G, H : \mathbb{R}^n \to \mathbb{R}^q$ are the *complementarity constraints*.

Here, the shorthand $0 \leq G(x) \perp H(x) \geq 0$ is an abbreviation for

$$G_i(x) \geq 0, \quad H_i(x) \geq 0, \quad G_i(x)H_i(x) = 0 \qquad \forall i = 1, \ldots, q.$$

Consequently, MPCCs can be seen standard *nonlinear programs* (NLP), but the complementarity constraints have a special structure, which prevents us from directly applying most of the classical results.

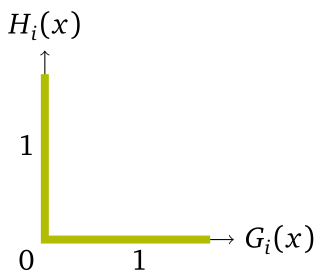


Figure 1.1: Structure of the feasible set of a complementarity constraint

The structure of the feasible set of one complementarity constraint is depicted in Figure 1.1. As one can see, at a feasible point x^* of the MPCC each complementarity constraint belongs to one of the following three cases:

$$\begin{aligned} I_{0+}(x^*) &:= \{i = 1, \dots, q \mid G_i(x^*) = 0, H_i(x^*) > 0\}, \\ I_{+0}(x^*) &:= \{i = 1, \dots, q \mid G_i(x^*) > 0, H_i(x^*) = 0\}, \\ I_{00}(x^*) &:= \{i = 1, \dots, q \mid G_i(x^*) = 0, H_i(x^*) = 0\}. \end{aligned}$$

In Section 1.1 we provide some examples for MPCCs, in Section 1.2 we discuss why MPCCs need special treatment, Section 1.3 provides some background knowledge needed later and Section 1.4 collects the notation used throughout these notes.

1.1 Examples and Applications

We begin by introducing some problem classes, which are closely related to MPCCs. Then we describe how these problems can be formulated as an MPCC. Finally, we provide an concrete application for MPCCs.

Class: Mathematical Programs with Equilibrium Constraints

MPCCs are closely related to the class of *mathematical programs with equilibrium constraints (MPEC)*. MPECs are optimization problems of the form

$$\min_{x,y} f(x,y) \quad \text{s.t.} \quad x \in X, y \in S(x),$$

where $X \subseteq \mathbb{R}^n$ is a nonempty feasible set and $f : \mathbb{R}^n \times \mathbb{R}^N \to \mathbb{R}$ is the objective function. The possibly set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^N$ describes the solution set of the *equilibrium constraint* and is often given only implicitly.

As we will see below, MPECs can often be reformulated as MPCCs. And since“ MPEC” is much easier to pronounce than “MPCC”, optimization problems with complementarity constraints are often also called MPECs in literature.

Class: Bilevel Optimization Problems

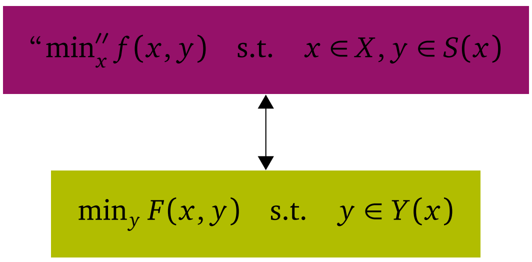


Figure 1.2: Bilevel Problem

One example for an MPEC are *bilevel optimization problems*, where

$$S(x) = \operatorname*{argmin}_y \{F(x,y) \mid y \in Y(x)\}$$

is the solution set of a lower level optimization problem

$$\min_y F(x,y) \quad \text{s.t.} \quad y \in Y(x)$$

with the variable y and the parameter x. Then the map $S(x)$ is usually not known explicitly.

In case the map $S(x)$ is not single-valued, i.e. y is not uniquely determined by x, we have to differentiate whether the variables x and y are controlled by the same entity. If yes, we can use the so called *optimistic formulation*

$$\min_{x,y} f(x,y) \quad \text{s.t.} \quad x \in X, y \in S(x).$$

If not, it can make more sense to consider the *pessimistic formulation*

$$\min_{x} \max_{y} f(y,x) \quad \text{s.t.} \quad x \in X, y \in S(x),$$

which is harder to solve.

Class: Stackelberg Problems

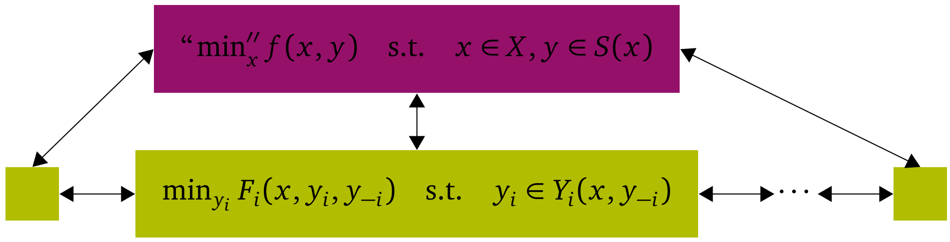


Figure 1.3: Stackelberg Problem

The bilevel optimization problem can be seen as a *single-leader-single-follower problem*, where the entity controlling x and minimizing f is the leader and the entity controlling y and minimizing F is the follower, who is reacting to the choice of x.

In case there is more than one follower and the followers do not only depend on the leader but also influence each other, we have to consider a game on the lower level, i.e. each follower i solves

$$\min_{y_i} F_i(x,y_i,y_{-i}) \quad \text{s.t.} \quad y_i \in Y_i(x,y_{-i}),$$

where y_i is the variable of player i and y_{-i} collects the variables of all other players on the lower level.

The set-valued map S could then be defined as the set of all *Nash equilibria* for a given value of x. Such *single-leader-multi-follower problems* are also called *Stackelberg problems*.

Reformulation: Lower Level to Complementarity Constraint

Assume that the feasible set $Y(x)$ of the lower level

$$\min_{y} F(x,y) \quad \text{s.t.} \quad y \in Y(x)$$

is given by some standard constraints, i.e.

$$Y(x) = \{y \mid G(x,y) \leq 0, H(x,y) = 0\}.$$

Then we know under constraint qualifications such as Abadie CQ or LICQ that an element $y \in S(x) = \text{argmin}_y\{F(x,y) \mid y \in Y(x)\}$ has to satisfy the *KKT conditions*

$$\begin{aligned} \nabla_y F(x,y) + \nabla_y G(x,y)\lambda + \nabla_y H(x,y)\mu &= 0, \\ 0 \leq \lambda \perp G(x,y) &\leq 0, \\ H(x,y) &= 0. \end{aligned}$$

Conversely, the KKT conditions are sufficient for an element of $S(x)$ if f, G_i are convex and H is affine linear in y. Thus, under suitable assumptions, we can reformulate a bilevel problem as the MPCC

$$\begin{aligned}\min_{x,y,\lambda,\mu} f(x,y) \quad \text{s.t.} \quad & x \in X,\\ & \nabla_y F(x,y) + \nabla_y G(x,y)\lambda + \nabla_y H(x,y)\mu = 0,\\ & 0 \le \lambda \perp -G(x,y) \ge 0,\\ & H(x,y) = 0.\end{aligned}$$

Application: Contest Design

Consider a lottery, where every player $i \in \{1,\dots,N\}$ has to decide on an input $y_i \ge 0$ and can win a price of value $v > 0$ with probability

$$\frac{w_i y_i}{\sum_{j=1}^N w_j y_j},$$

where $w_j > 0$ are some given weights. In case $w_j = 1$ for all players, the winning probability is exactly equal to the relative input of each player. Thus, every player i maximizes the objective function

$$F_i(w, y_i, y_{-i}) = v \frac{w_i y_i}{\sum_{j=1}^N w_j y_j} - c_i y_i,$$

where $c_i > 0$ describes the cost for the input of i.

Once can show that this game has exactly one Nash equilibrium for all given weights w. Thus, the map $w \mapsto S(w)$ collecting all Nash equilibria for a given w is single-valued.

Now consider the contest organizer, who can choose the weights w in order to optimize the objective

$$\max_w f(w,y) \quad \text{s.t.} \quad w \ge 0, y \in S(w).$$

One possible such objective would be to maximize the total input into the lottery, i.e.

$$f(w,y) = \sum_{j=1}^M y_j.$$

This is an example for a Stackelberg problem, where one can obtain an explicit formula for the solution of the lower level Nash game. But alternatively one can also reformulate the lower level using a variational inequality or KKT conditions.

1.2 Challenges

Let us come back to the general MPCC

$$\begin{aligned}\min_x f(x) \quad \text{s.t.} \quad & g(x) \le 0, h(x) = 0,\\ & 0 \le G(x) \perp H(x) \ge 0.\end{aligned}$$

Then we can obviously rewrite this equivalently as the following nonlinear program (MPCC-NLP)

$$\begin{aligned}\min_x f(x) \quad \text{s.t.} \quad & g(x) \le 0, h(x) = 0,\\ & -G(x) \le 0, -H(x) \le 0, G(x) \circ H(x) = 0.\end{aligned}$$

This raises the question: Why can we not solve MPCCs by just applying the standard theory from nonlinear optimization?

To answer this question, consider the following example from [28]:

Example 1.1. Consider the MPCC

$$\min_{x\in\mathbb{R}^3} x_1 + x_2 - x_3 \quad \text{s.t.} \quad \begin{aligned} &-4x_1 + x_3 \le 0, \\ &-4x_2 + x_3 \le 0, \\ &0 \le x_1 \perp x_2 \ge 0. \end{aligned}$$

This is an MPCC with the simplest possible complementarity constraint and two additional linear inequality constraints as well as a linear objective function. Since

$$x_3 \le 4\min\{x_1, x_2\} = 0,$$

the global minimum is $x^* = (0,0,0)$. Not let us check the KKT conditions at x^*. To do so, we have to examine if there exist multipliers $\lambda \in \mathbb{R}^4_+$, $\mu \in \mathbb{R}$ such that

$$\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + \lambda_1 \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ -4 \\ 1 \end{pmatrix} + \lambda_3 \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} + \lambda_4 \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

These three equations yield

$$\begin{aligned} \lambda_3 &= 1 - 4\lambda_1 \ge 0, \\ \lambda_4 &= 1 - 4\lambda_2 \ge 0, \\ \lambda_1 + \lambda_2 &= 1, \text{ where } \lambda_1 \ge 0, \lambda_2 \ge 0, \end{aligned}$$

which is obviously not possible. Thus, the global minimum of the MPCC is no KKT point.

If minima are no KKT points, then our standard first order optimality condition from nonlinear optimization is not necessary. Since solution algorithms for NLPs usually converge to KKT points, we thus might miss the solution of an MPCC if we apply a standard solution algorithm.

We know that a local solution of an NLP is a KKT point if a *constraint qualification* (CQ) holds there. Thus the special structure of the complementarity constraints seems to cause problems with constraint qualifications. We illustrate the cause of this problem on the following example:

Example 1.2. Consider the simplest possible MPCC

$$\min_{x\in\mathbb{R}^2} f(x) \quad \text{s.t.} \quad 0 \le x_1 \perp x_2 \ge 0.$$

and the corresponding MPCC-NLP

$$\min_{x\in\mathbb{R}^2} f(x) \quad \text{s.t.} \quad -x_1 \le 0, -x_2 \le 0, x_1 x_2 = 0.$$

Now we want to verify if LICQ holds at an arbitrary feasible point x. Let us look at the three possible cases separately:

If $x_1 = 0$ and $x_2 > 0$, the gradients of the active constraints are $(-1,0)^T$ and $(x_2,0)^T$ and thus linearly dependent. If $x_1 > 0$ and $x_2 = 0$, then the gradients of the active constraints are $(0,-1)$ and $(0,x_1)$ and also linearly dependent. If $x_1 = x_2 = 0$, all three constraints are active with the gradients $(-1,0)^T$, $(0,-1)^T$ and $(0,0)^T$ and also linearly dependent.

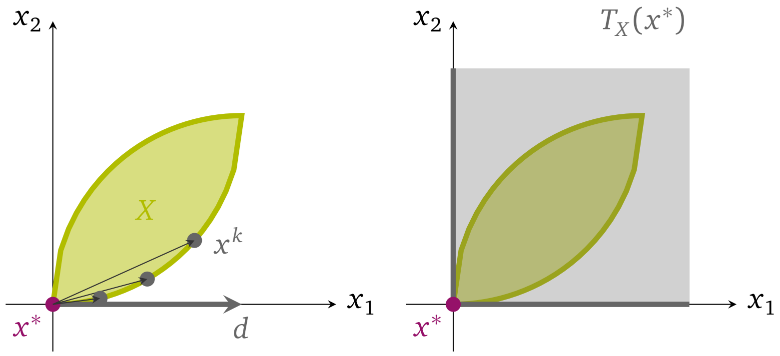


Figure 1.4: Illustration of the tangent cone

This problem is typical for MPCCs, in fact one can show that at every feasible point of an MPCC both LICQ and MFCQ are violated, see Exercise 1.2. While LICQ is a popular constraint qualification because it can easily be verified, it is one of the strongest constraint qualifications. So maybe weaker constraint qualifications have a better chance at being satisfied.

Recall that for $X \subseteq \mathbb{R}^n$ nonempty and $x^* \in X$ the *(Bouligand) tangent cone* is defined as

$$T_X(x^*) := \{d \in \mathbb{R}^n \mid \exists x^k \to_X x^*, t_k \geq 0 \ : \ t_k(x^k - x^*) \to d\},$$

see Figure 1.4 for an illustration.

In case

$$X = \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}$$

with $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ continuously differentiable, the *set of active inequalities* at $x^* \in X$ is defined as

$$I_g(x^*) := \{i = 1, \ldots, m \mid g_i(x^*) = 0\}$$

and the *linearized tangent cone* as

$$\begin{aligned} L_X(x^*) := \{d \in \mathbb{R}^n \quad | \quad & \nabla g_i(x^*)^T d \leq 0 \quad \forall i \in I_g(x^*), \\ & \nabla h(x^*)^T d = 0\}. \end{aligned}$$

Note that, although the notation does not indicate this, the linearized tangent cone to X depends on the description of X, i.e. on g and h.

Also recall that the *polar cone* to a nonempty set $C \subseteq \mathbb{R}^n$ is defined as

$$C^\circ := \{w \in \mathbb{R}^n \mid w^T d \leq 0 \quad \forall d \in C\}.$$

Then the weakest constraint qualifications usually used to ensure that local minima are KKT points are:

- *Abadie CQ*: $T_X(x^*) = L_X(x^*)$
- *Guignard CQ*: $T_X(x^*)^\circ = L_X(x^*)^\circ$

Let us see, if these weaker CQs are satisfied in Example 1.2.

Example 1.3. If $x_1 = 0$ and $x_2 > 0$, the gradients of the active constraints are $(-1,0)^T$ and $(x_2,0)^T$ and thus the linearized tangent cone is given by

$$L_X(x) = \{d \in \mathbb{R}^2 \mid d_1 = 0\}.$$

Since the tangent cone in these points is $\{0\} \times \mathbb{R}$, Abadie CQ and thus also Guignard CQ holds there.
If $x_1 > 0$ and $x_2 = 0$, then the gradients of the active constraints are $(0,-1)$ and $(0,x_1)$ and thus

$$L_X(x) = \{d \in \mathbb{R}^2 \mid d_2 = 0\} = T_X(x).$$

If $x_1 = x_2 = 0$, all three constraints are active with the gradients $(-1,0)^T$, $(0,-1)^T$ and $(0,0)^T$ and also linearly dependent. The resulting cones are

$$L_X(x) = \{d \in \mathbb{R}^2 \mid d_1 \geq 0, d_2 \geq 0\} \supsetneq T_X(x) = \{d \in \mathbb{R}^2 \mid 0 \leq d_1 \perp d_2 \geq 0\}.$$

Consequently Abadie CQ is violated in $x = (0,0)$. However, for the polar cones we obtain

$$L_X(x)^\circ = \{w \in \mathbb{R}^2 \mid w_1 \geq 0, w_2 \geq 0\} = T_X(x)^\circ,$$

i.e. Guignard CQ holds.

The behavior of this example is typical for MPCCs. If we have a biactive complementarity constraint, i.e. $G_i(x^*) = H_i(x^*)$, the feasible set and the tangent cone are usually nonconvex. On the other hand, the linearized tangent cone is polyhedral and thus always convex. Thus, Abadie CQ might not be satisfied. As we have seen in the previous example, even if Abadie CQ is violated, Guignard CQ still has a chance to hold. But as Example 1.2 shows, there are also very simple MPCCs, where Guignard CQ is violated at the solution.

If we cannot ensure that a constraint qualification holds at local minima, we cannot rely on the KKT conditions as necessary optimality conditions. But then we know that the Fritz-John conditions are necessary optimality conditions even without a constraint qualification. However, if we try to apply the standard Fritz-John conditions to the NLP formulation of the MPCC, we see that they are satisfied in every feasible point, see Exercise 1.3 and thus not a very useful optimality conditions.

Thus, if we want to solve MPCCs, we first have to derive suitable optimality conditions and then can develop tailored algorithms. We can also try to apply standard solvers directly, which sometimes works, but we should be aware of the potential problems.

1.3 Background

Let us recall some concepts, which will be used later.

1.3.1 Bouligand and Clarke Subdifferential

In this section, we consider *locally Lipschitz continuous* maps

$$F : \mathbb{R}^n \to \mathbb{R}^m.$$

To define a subdifferential for F we use the following, famous observation:

Theorem 1.4 (Rademacher's Theorem). *Let $U \subseteq \mathbb{R}^n$ be open and $F : U \to \mathbb{R}^m$ be locally Lipschitz continuous. Then F is differentiable almost everywhere, i.e. the set*

$$\{x \in U \mid F'(x) \text{ does not exist}\}$$

is a set of (Lebesgue) measure zero.

This motivates the following definitions:

Definition 1.5. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz continuous. Define the set*

$$D_F := \{x \in \mathbb{R}^n \mid F \text{ is differentiable in } x\}.$$

(a) Then for $x^* \in \mathbb{R}^n$ the set

$$\partial^B F(x^*) := \{M \in \mathbb{R}^{m\times n} \mid \exists (x^k)_k \to_{D_F} x^* \ : \ F'(x^k) \to M\}$$

is called the Bouligand subdifferential *of F.*

(b) For $x^* \in \mathbb{R}^n$ the set

$$\partial^C F(x^*) := \operatorname{conv} \partial^B F(x^*)$$

is called the Clarke subdifferential *of F. Its elements are called* (Clarke's) generalized Jacobians.

Beware: According to the previous definition, the elements of the Bouligand or Clarke subdifferential of a function $f : \mathbb{R}^n \to \mathbb{R}$ are row vectors, i.e. have dimension $1 \times n$. However, for functions mapping to $\mathbb{R}$ one sometimes also transposes the Bouligand and Clarke subdifferential and then speaks of (Clarke's) generalized gradient.

Before we collect useful properties of these subdifferentials, some examples:

- Consider the absolute value function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$. Then f is even globally Lipschitz continuous with $L = 1$ and $D_f = \mathbb{R} \setminus \{0\}$. Thus

  $$\partial^B f(x) = \begin{cases} \{-1\} & \text{if } x < 0, \\ \{-1, 1\} & \text{if } x = 0, \\ \{1\} & \text{if } x > 0. \end{cases}$$

  The Clarke subdifferential differs only in case $x = 0$ with

  $$\partial^C f(0) = [-1, 1].$$

- Consider the maximum function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \max\{x, 0\}$. Then f is even globally Lipschitz continuous with $L = 1$ and $D_f = \mathbb{R} \setminus \{0\}$. Thus

  $$\partial^B f(x) = \begin{cases} \{0\} & \text{if } x < 0, \\ \{0, 1\} & \text{if } x = 0, \\ \{1\} & \text{if } x > 0. \end{cases}$$

  The Clarke subdifferential differs only in case $x = 0$ with

  $$\partial^C f(0) = [0, 1].$$

- Consider the minimum function $f : \mathbb{R}^2 \to \mathbb{R}$, $f(a, b) = \min\{a, b\}$. Then f is even globally Lipschitz continuous with $L = 1$ and $D_f = \mathbb{R}^2 \setminus \{(a, a) \mid a \in \mathbb{R}\}$. Thus

  $$\partial^B f(a, b) = \begin{cases} \{(1, 0)\} & \text{if } a < b, \\ \{(1, 0), (0, 1)\} & \text{if } a = b, \\ \{(0, 1)\} & \text{if } a > b. \end{cases}$$

  The Clarke subdifferential differs only in case $a = b$ with

  $$\partial^C f(a, a) = \operatorname{conv}\{(1, 0), (0, 1)\}.$$

Some more examples for subdifferentials of commonly used vector norms are considered in Exercise 1.4.

As one might expect from the definition, the Bouligand and the Clarke subdifferential are closely related to the gradient, whenever it exists.

Proposition 1.6. *Let $F:\mathbb{R}^n\to\mathbb{R}^m$ be locally Lipschitz continuous and $x^*\in\mathbb{R}^n$. If F is continuously differentiable around x^*, then*

$$\partial^C F(x^*)=\partial^B F(x^*)=\{\nabla F(x^*)\}.$$

Proof. For all x^k close to x^* the Jacobian $F'(x^k)$ exists and converges to $F'(x^*)$. Thus $\partial^B F(x^*)=\{F'(x^*)\}$. Then the convex hull does not change anything, i.e. $\partial^C F(x^*)=\partial^B F(x^*)=\{\nabla F(x^*)\}$. □

The following example illustrates that differentiability of f alone is not enough to guarantee $\partial^C f(x^*)=\{\nabla f(x^*)\}$.

Example 1.7. Consider $f:\mathbb{R}\to\mathbb{R}$ with $f(x)=x^2\sin(\frac{1}{x})$. Then f is differentiable on $\mathbb{R}^n$ with

$$f'(x)=\begin{cases}2x\sin(\frac{1}{x})-\cos(\frac{1}{x}) & \text{if } x\neq 0,\\ 0 & \text{if } x=0.\end{cases}$$

Since f' is continuous on $\mathbb{R}\setminus\{0\}$, the function f is locally Lipschitz-continuous on $\mathbb{R}\setminus\{0\}$. One can also show that f is locally Lipschitz continuous around 0. For all $x\neq 0$ we know

$$\partial^C f(x)=\{\nabla f(x)\}=\{2x\sin(\frac{1}{x})-\cos(\frac{1}{x})\}.$$

However, at $x=0$, we have

$$\partial^C f(0)=\partial^B f(0)=[-1,1].$$

This can be verified directly using the definition of the subdifferentials.

Now let us collect some useful properties of the subdifferentials.

Lemma 1.8. *Let $F:\mathbb{R}^n\to\mathbb{R}^m$ be locally Lipschitz continuous and $x^*\in\mathbb{R}^n$ with local Lipschitz constant $L>0$. Then the following holds:*

(a) *$\partial^B F(x^*)$ is nonempty and compact with $\|M\|\le L$ for all $M\in\partial^B F(x^*)$.*

(b) *The map $x\mapsto\partial^B F(x)$ is* upper semicontinuous *at x^*, i.e. for all $\varepsilon>0$ exists $r>0$ such that*

$$\partial^B F(x)\subseteq\partial^B F(x^*)+B_\varepsilon(0)\qquad\forall x\in B_r(x^*).$$

This together with (a) implies that $x\mapsto\partial^B F(x)$ is locally bounded.

(c) *$\partial^C F(x^*)$ is nonempty, convex and compact with $\|M\|\le L$ for all $M\in\partial^C F(x^*)$.*

(d) *The map $x\mapsto\partial^C F(x)$ is* upper semicontinuous *and locally bounded.*

When it comes to computing the subdifferential of a given function, it is often useful to be able to use calculus rules. The Clarke subdifferential of a function $f:\mathbb{R}^n\to\mathbb{R}$ can be defined via a support function

$$f^\circ(x;d):=\limsup_{y\to x,t\downarrow 0}\frac{f(y+td)-f(y)}{t},$$

which allows to obtain fairly extensive collection of calculus rules. When we are interested in the case $F:\mathbb{R}^n\to\mathbb{R}^m$, we have to rely on the definition of the Clarke subdifferential via Rademacher's theorem, which complicates the proof of calculus rules. For this reason in the subsequent theorem we have to work with something similar to directional derivatives.

Theorem 1.9. *Let $H : \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^m \to \mathbb{R}^p$ be locally Lipschitz continuous and $x^* \in \mathbb{R}^n$. Then $F := G \circ H$ is locally Lipschitz continuous and for all $d \in \mathbb{R}^n$*

(a)
$$\partial^C F(x^*)d \subseteq \mathrm{conv}(\partial^C G(H(x^*))\partial^C H(x^*)d) = \mathrm{conv}(\partial^C G(H(x^*))\partial^C H(x^*))d;$$

(b) if G is continuously differentiable around $H(x^*)$
$$\partial^C F(x^*)d = G'(H(x^*))\partial^C H(x^*)d;$$

(c) if H is continuously differentiable around x^*
$$\partial^C F(x^*)d = \partial^C G(H(x^*))H'(x^*)d.$$

One can use more and better calculus rules for maps $f : \mathbb{R}^n \to \mathbb{R}$:

Theorem 1.10. *Let $f, f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^m \to \mathbb{R}^n$ be locally Lipschitz continuous and $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Then the following holds:*

(a) $\partial^C(cf)(x) = c\partial^C f(x)$ for all $c \in \mathbb{R}$.

(b) $\partial^C(f_1 + f_2)(x) \subseteq \partial^C f_1(x) + \partial^C f_2(x)$.

(c) $\partial^C(f_1 \cdot f_2)(x) \subseteq f_2(x)\partial^C f_1(x) + f_1(x)\partial^C f_2(x)$.

(d) $\partial^C(\frac{f_1}{f_2})(x) \subseteq \frac{f_2(x)\partial^C f_1(x) - f_1(x)\partial^C f_2(x)}{f_2(x)^2}$ if $f_2(x) \neq 0$.

(e) $\partial(f \circ G)(y) \subseteq \mathrm{conv}(\partial^C f(G(y)) \cdot \partial^C G(y))$ with equality if f is continuously differentiable or f convex and G continuously differentiable.

In general, the inclusions in the calculus rules can be strict. Consider for example $f_1(x) = |x|$, $f_2(x) = -|x|$ and $x^* = 0$. Then $(f_1 + f_2)(x) \equiv 0$ and thus
$$\partial^C(f_1 + f_2)(0) = \{0\}.$$
On the other hand
$$\partial^C f_1(0) + \partial^C f_2(0) = [-1, 1] + [-1, 1] = [-2, 2].$$

Remark 1.11 (Optimality Conditions). Note that the example $f(x) = |x|$ illustrates that the Clarke subdifferential has similar problems as the gradient, when it comes to characterizing local minima. While one can show that every local minimum x^* satisfies $0 \in \partial^C f(x^*)$, so does every local maximum. And analogously to the gradient of a nonconvex function, the reverse implication does not hold, i.e. $0 \in \partial^C f(x^*)$ does not imply that x^* is a local maximum or minimum.

Proposition 1.12. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz-continuous and $x^* \in \mathbb{R}^n$. Then the following holds:*

(a) $\partial^C(-f)(x^*) = -\partial^C f(x^*)$.

(b) If x^* is a local minimum or maximum of f, we have $0 \in \partial^C f(x^*)$.

For constrained optimization problems
$$\min_x f(x) \quad \text{s.t.} \quad g(x) \leq 0, h(x) = 0$$
one can prove an analogue the the Fritz-John conditions known from nonlinear optimization.

Theorem 1.13 (Fritz-John Conditions of Clarke). *Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ be locally Lipschitz-continuous . Let x^* be a local minimum of f on*
$$Z := \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}.$$
Then there exist multipliers $\alpha \geq 0$, $\lambda \in \mathbb{R}^m_+$, $\mu \in \mathbb{R}^p$ such that $(\alpha, \lambda, \mu) \neq (0, 0, 0)$, $\lambda_i = 0$ for all $i \notin I(x^*)$ and
$$0 \in \alpha\partial^C f(x^*) + \sum_{i=1}^m \lambda_i \partial^C g_i(x^*) + \sum_{i=1}^p \mu_i \partial^C h_i(x^*)$$

1.3.2 Fréchet and Limiting Normal Cones

Let us consider a nonempty feasible set $Z \subseteq \mathbb{R}^n$ and a point $x^* \in Z$. Then we have already seen that the *Bouligand tangent cone* at x^* is defined as

$$T_Z(x^*) := \{d \in \mathbb{R}^n \mid \exists (x^k)_k \to_Z x^*, (t_k)_k \geq 0 \; : \; t_k(x^k - x^*) \to d\}.$$

If we want to work with constraint qualifications such as Guignard CQ, we need to compute the corresponding polar cone

$$T_Z(x^*)^\circ := \{w \in \mathbb{R}^n \mid w^T d \leq 0 \quad \forall d \in T_Z(x^*)\}.$$

Definition 1.14. *Consider $Z \subseteq \mathbb{R}^n$ nonempty and $x^* \in Z$. Then the* Fréchet normal cone *of Z at x^* is defined as*

$$N_Z^F(x^*) := T_Z(x^*)^\circ = \{w \in \mathbb{R}^n \mid w^T d \leq 0 \quad \forall d \in T_Z(x^*)\}.$$

For $x^* \notin Z$ one defines $N_Z^F(x^*) := \emptyset$.

One can also provide a formula for the Fréchet normal cone, which is often used as its definition.

Proposition 1.15. *Consider $Z \subseteq \mathbb{R}^n$ nonempty and $x^* \in Z$. Then*

$$N_Z^F(x^*) := T_Z(x^*)^\circ = \{w \in \mathbb{R}^n \mid \limsup_{x \to x^*, x \in Z \setminus \{x^*\}} w^T \frac{x - x^*}{\|x - x^*\|} \leq 0\}.$$

Proof. Let us call the give set M. We begin by showing $M \subseteq N_Z^F(x^*) := T_Z(x^*)^\circ$. To this end consider an arbitrary $w \in M$ and $d \in T_Z(x^*)$. Then by definition of the tangent cone, we can find $(x^k)_k \subseteq Z$ and $t_k \geq 0$ such that $x^k \to x^*$ and $t_k(x^k - x^*) \to d$. Due to $w \in M$ we then know

$$0 \geq \limsup_{x \to x^*, x \in Z \setminus \{x^*\}} w^T \frac{x - x^*}{\|x - x^*\|} \geq \lim_{k \to \infty} w^T t_k(x^k - x^*) \cdot \frac{1}{t_k \|x^k - x^*\|} = w^T \frac{d}{\|d\|}.$$

Since this holds for all $d \in T_Z(x^*)$ we have shown $w \in N_Z^F(x^*)$.

To show the opposite inclusion $N_Z^F(x^*) \subseteq M$ assume one can find $w \in N_Z^F(x^*) \setminus M$. Since $w \notin M$ there exists $(x^k)_k \subseteq Z$ with $x^k \to x^*$ and

$$\lim_{k \to \infty} w^T \frac{x^k - x^*}{\|x^k - x^*\|} > 0.$$

Defining $t_k := \frac{1}{\|x^k - x^*\|}$ we obtain that

$$t_k(x^k - x^*) = \frac{x^k - x^*}{\|x^k - x^*\|} \to d^* \in T_Z(x^*)$$

at least on a subsequence. But then $w^T d^* > 0$, which is a contradiction to $w \in N_Z^F(x^*) = T_Z(x^*)^\circ$. □

Using the Frechet normal cone, we can rewrite *B-stationarity* as

$$\nabla f(x^*)^T d \geq 0 \quad \forall d \in T_X(x^*) \quad \Longleftrightarrow \quad -\nabla f(x^*) \in T_X(x^*)^\circ = N_X^F(x^*).$$

However, as the following example illustrates, the Fréchet normal cone is rather "unstable", which makes it hard to obtain good calculus rules.

Example 1.16. Consider the MPCC

$$\min_{x\in\mathbb{R}^2} f(x) \quad \text{s.t.} \quad 0 \le x_1 \perp x_2 \ge 0.$$

Then we have

$$N_X^F(x) = \begin{cases} \{0\}\times\mathbb{R} & \text{if } x_1 > 0, x_2 = 0,\\ \mathbb{R}\times\{0\} & \text{if } x_1 = 0, x_2 > 0,\\ (-\infty,0]^2 & \text{if } x_1 = 0, x_2 = 0.\end{cases}$$

Thus, if we approach $x^* = (0,0)$ e.g. with $x^k = (\frac{1}{k},0)$, then the Fréchet normal cone at x^k contains elements w with $w_2 > 0$ but at the limit these elements disappear.

This motivates the introduction of the following normal cone.

Definition 1.17. *Consider $Z \subseteq \mathbb{R}^n$ nonempty and* closed *and $x^* \in Z$. Then the* limiting or Mordukhovich normal cone *of Z at x^* is defined as*

$$N_Z^M(x^*) := \{w \in \mathbb{R}^n \mid \exists (x^k)_k \to x^*, w^k \in N_Z^F(x^k) \;:\; w^k \to w\}.$$

For $x^* \notin Z$ one defines $N_Z^M(x^*) := \emptyset$.

Since we defined $N_Z^F(x) = \emptyset$ for $x \notin Z$ it suffices to consider sequences $(x^k)_k \subseteq Z$ in the definition of the limiting normal cone.

As the name already indicates, the limiting normal cone is the limsup or closure of the Fréchet normal cone. Originally and also nowadays, the limiting normal cone is often defined based on the proximal normal cone instead of the Fréchet normal cone. However one can show that both constructions lead to the same limiting normal cone. Thus, we restrict ourselves to the Fréchet normal cone, which we already know from Guignard CQ.

Let us consider a few examples, which we need later on.

- For $Z = \{0\} \subset \mathbb{R}$ we know

  $$N_Z^F(x) = \begin{cases} \mathbb{R} & \text{if } x = 0,\\ \emptyset & \text{if } x \neq 0.\end{cases}$$

  Consequently, we also have

  $$N_Z^M(x) = \begin{cases} \mathbb{R} & \text{if } x = 0,\\ \emptyset & \text{if } x \neq 0.\end{cases}$$

- For $Z = (-\infty,0] \subset \mathbb{R}$ we know

  $$N_Z^F(x) = \begin{cases} \{0\} & \text{if } x < 0,\\ [0,\infty) & \text{if } x = 0,\\ \emptyset & \text{if } x > 0.\end{cases}$$

  Consequently, we also have

  $$N_Z^M(x) = \begin{cases} \{0\} & \text{if } x < 0,\\ [0,\infty) & \text{if } x = 0,\\ \emptyset & \text{if } x > 0.\end{cases}$$

- For $Z = \{x \in \mathbb{R}^2 \mid 0 \le x_1 \perp x_2\}$ we have

$$N_Z^F(x) = \begin{cases} \{0\} \times \mathbb{R} & \text{if } x_1 > 0, x_2 = 0, \\ \mathbb{R} \times \{0\} & \text{if } x_1 = 0, x_2 > 0, \\ (-\infty, 0] \times (-\infty, 0] & \text{if } x_1 = 0, x_2 = 0, \\ \emptyset & \text{otherwise.} \end{cases}$$

This leads to the following limiting normal cone

$$N_Z^M(x) = \begin{cases} \{0\} \times \mathbb{R} & \text{if } x_1 > 0, x_2 = 0, \\ \mathbb{R} \times \{0\} & \text{if } x_1 = 0, x_2 > 0, \\ ((-\infty, 0] \times (-\infty, 0]) \cup (\{0\} \times \mathbb{R}) \cup (\mathbb{R} \times \{0\}) & \text{if } x_1 = 0, x_2 = 0, \\ \emptyset & \text{otherwise.} \end{cases}$$

As we have seen in the examples, for the convex sets $\{0\}$ and $(-\infty, 0]$ the limiting normal cone coincides with the Fréchet normal cone and the convex normal cone. In Exercise 1.5 we prove that this is true for general convex sets. Note that this implies that the normal cones of a polyhedron

$$P = \{x \in \mathbb{R}^n \mid a_i^T x \le b_i \quad \forall i = 1, \dots, m\}$$

are given by

$$N_P^M(x^*) = N_P^F(x^*) = N_P(x^*) = \{w \in \mathbb{R}^n \mid w = \sum_{i \in I(x^*)} \lambda_i a_i, \lambda \ge 0\},$$

where $I(x^*) = \{i \mid a_i^T x = b_i\}$.

For the nonconvex set $\{x \in \mathbb{R}^2 \mid 0 \le x_1 \perp x_2\}$ however the limiting normal cone and the Fréchet normal cone differ at the "point of nonconvexity" $x = 0$.

Later, we kneed to compute the normal cone to an intersection of sets. A suitable result under quite weak assumptions can be found in [15, Corollary 4.2].

Theorem 1.18. *Consider two nonempty and closed sets $A, B \subseteq \mathbb{R}^n$ and $x^* \in A \cap B$. Assume that the set-valued function*

$$M(y) := \{x \in A \mid x + y \in B\}$$

is calm *in* $(0, x^*)$. *Then*

$$N_{A \cap B}^M(x^*) \subseteq N_A^M(x^*) + N_B^M(x^*).$$

The set-valued map M, that appears here, can be seen as a kind of perturbation map, i.e. which points x are feasible if we perturb the feasible set by the parameter y.

This is exactly exactly the kind of formula we need to proceed with our optimality conditions for MPCCs. Since the proof is rather involved and based on many concepts we have not introduced here, we skip the proof and instead take a look at the assumption of calmness. But first, in order to illustrate that the previous result is not true anymore, if we drop the assumption of calmness, consider the following example, which originates from a presentation given by Jiri Outrata.

Example 1.19. Consider the two sets

$$A = \{x \in \mathbb{R}^2 \mid (x_1+1)^2 + x_2^2 \leq 1\} \quad \text{and} \quad B = \{x \in \mathbb{R}^2 \mid x_1 \geq 0\}.$$

Since both sets are convex we do not have to differentiate between the various normal cones. The intersection is $A \cap B = \{0\}$ and thus $N^M_{A\cap B}(0) = \mathbb{R}^2$. On the other hand, we have $N^M_A(0) = \mathbb{R}_+ \times \{0\}$ and $N^M_B(0) = \mathbb{R}_- \times \{0\}$ and thus

$$N^M_A(0) + N^M_B(0) = \mathbb{R} \times \{0\} \not\supseteq \mathbb{R}^2 = N^M_{A\cap B}(0).$$

Consequently, the map

$$M(y) = \{x \in A \mid x + y \in B\} = \{x \in \mathbb{R}^2 \mid (x_1+1)^2 + x_2^2 \leq 1, x_1 \geq -y_1\}$$

cannot be calm at $(y^*, x^*) = (0,0)$.

Calmness is a Lipschitz-type property of set-valued maps. So let us introduce some related notions.

Definition 1.20. *Consider $Y \subseteq \mathbb{R}^n$ a set-valued map $M : Y \rightrightarrows \mathbb{R}^m$.*

(a) M has the Aubin property at $(y^*, x^*) \in \text{graph}(M)$, *if there exist $\delta > 0$, $\varepsilon > 0$, $L > 0$ such that*

$$M(y_1) \cap B_\varepsilon(x^*) \subseteq M(y_2) + L\|y_1 - y_2\| B_1(0) \quad \forall y_1, y_2 \in B_\delta(y^*).$$

(c) M is calm at $(y^*, x^*) \in \text{graph}(M)$ *if there exist $\delta > 0$, $\varepsilon > 0$, $L > 0$ such that*

$$M(y) \cap B_\varepsilon(x^*) \subseteq M(y^*) + L\|y - y^*\| B_1(0) \quad \forall y \in B_\delta(y^*).$$

The Aubin property is a Lipschitz-type property for set-valued maps, that results from the observation that one is often only interested in the behavior of $M(y)$ close to a point of interest $x^* \in M(y^*)$ and thus intersects the images $M(y)$ with $B_\varepsilon(x^*)$. Note that in the definition of the Aubin property we could also intersect $M(y_2)$ or the complete righthand side with $B_\varepsilon(x^*)$, but this would not change anything. However, the Aubin property is always violated if $M(y_2) = \emptyset$ for y_2 arbitrarily close to y^*. (Recall that $\emptyset + B = \emptyset$ for any set B.)

Calmness results from the Aubin property by fixing one of the two points y_1, y_2 to the point of interest y^*. It is thus a point-based property that does not transfer to a neighborhood, as e.g. Lipschitz continuity or the Aubin property do.

From single-valued functions we know that (affine) linear functions have nice properties such as global Lipschitz continuity. This motivates to consider the following class of set-valued maps.

Definition 1.21. *A set-valued map $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is called a* polyhedral set-valued map (polyhedral multifunction)*, if its graph*

$$\text{graph}(M) := \{(y, x) \mid x \in M(y)\} \subseteq \mathbb{R}^n \times \mathbb{R}^m$$

can be written as the union of finitely many polyhedral convex sets, i.e.

$$\text{graph}(M) = \bigcup_{i=1}^{p} P_i \quad \textit{where} \quad P_i = \{(y, x) \mid A_i(y, x) \leq b_i\}.$$

Our next step is to prove that polyhedral set-valued maps are calm.

Theorem 1.22. *Consider a polyhedral set-valued map $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. Then M is calm at every point $(y^*, x^*) \in \operatorname{graph}(M)$.*

Proof. Define the set-valued maps

$$M_i(y) := \{x \mid A_i(y, x) \leq b_i\}.$$

We first verify that each map M_i is Lipschitz continuous on its domain. To this end note

$$\begin{aligned} M_i(y) &:= \{x \mid A_i(y,x) \leq b_i\} \\ &= \{x \mid A_{ix}x \leq b_i - A_{iy}y\} \\ &= \{x \mid A_{ix}x \in b_i - A_{iy}y + \mathbb{R}^{l_i}_-\} \\ &= \{x \mid A_{ix}x \in (b_i - A_{iy}y + \mathbb{R}^{l_i}_-) \cap R\} \\ &= D[(b_i - A_{iy}y + \mathbb{R}^{p_i}_-) \cap R] + K \end{aligned}$$

where K and R are the kernel and the range of A_{ix} and D is its pseudoinverse. Thus define $L_i = \|D\|\|A_{iy}\|$. Then for all $y_1, y_2 \in \operatorname{dom}(M_i)$ we have

$$b_i - A_{iy}y_1 + \mathbb{R}^{l_i}_- \subseteq (b_i - A_{iy}y_2 + \mathbb{R}^{l_i}_-) + \|A_{iy}\|\|y_1 - y_2\|B_1(0)$$

and thus

$$M_i(y_1) \subseteq M_i(y_2) + \underbrace{\|D\|\|A_{iy}\|}_{=L_i} \|y_1 - y_2\|B_1(0).$$

This shows that all maps M_i are Lipschitz continuous with modulus L_i on their domain $\operatorname{dom}(M_i)$.

Now define $L := \max_{i=1}^p L_i$. Then all M_i are calm on $\operatorname{dom}(M_i)$ with modulus L. Now assume M was not calm at $(y^*, x^*) \in \operatorname{graph}(M)$ with modulus L. Then we could find $(y, x) \in \operatorname{graph}(M)$ arbitrarily close to (y^*, x^*) such that

$$x \notin M(y^*) + L\|y - y^*\|B_1(0).$$

Let $I \subseteq \{1, \ldots, p\}$ be the set of all indices i such that $(y^*, x^*) \in P_i$. Since all polyhedrons P_i are closed and (y, x) is arbitrarily close to (y^*, x^*), we have $(y, x) \in P_i$ for some $i \in I$. But then the calmness of M_i wit modulus L implies

$$x \in M_i(y^*) + l\|y - y^*\|B_1(0) \subseteq M(y^*) + L\|y - y^*\|B_1(0),$$

a contradiction. □

1.4 Notation

Let us close this section with some words on the notation used in this script.

- By $R_+ := [0, \infty)$ we denote the nonnegative real numbers. Analogously $R_- = (-\infty, 0]$.
- For two vectors $x, y \in \mathbb{R}^n$ we denote the componentwise (Hadamard) product by

$$x \circ y := (x_i y_i)_{i=1}^n.$$

- With $e_i \in \mathbb{R}^n$ we denote the i-th unit vector and with $e = (1, \ldots, 1)^T \in \mathbb{R}^n$ the sum of all unit vectors.

- If not specified otherwise for $x \in \mathbb{R}^n$ we denote an *arbitrary norm* on $\mathbb{R}^n$ by $\|x\|$. Most of the time, we use the *euclidean norm*

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^T x}.$$

  Based on the norm used at the moment, we denote an open ball centered at x with radius r by

$$B_r(x) := \{y \in \mathbb{R}^n \mid \|y - x\| < r\}.$$

- For a sequence of real numbers $(t_k)_k \subset \mathbb{R}$ with $t_k > 0$ converging to zero, we write $t_k \downarrow 0$.
- For a sequence $(x^k)_k \subseteq X$ with $x^k \to x^*$ we use the shorthand $x^k \to_X x^*$.
- Sometimes, we use the *Landau notation*

$$(a_k)_k = \mathrm{o}(b_k) \quad :\Longleftrightarrow \quad \frac{a_k}{b_k} \to 0 \quad \text{for } k \to \infty.$$

- For a function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ we call $x^* \in X$ a *local minimum of f on X*, if there is an $\varepsilon > 0$ such that

$$f(x) \geq f(x^*) \quad \forall x \in X \cap B_\varepsilon(x^*).$$

  If the above inequality holds even for all $x \in X$, we call x^* a *global minimum of f on X*.
- A function $F : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is called *globally Lipschitz continuous on X* if there is a constant $L > 0$ such that

$$\|F(x) - F(y)\| \leq L\|x - y\| \quad \forall x, y \in X.$$

  The function F is called *locally Lipschitz continuous at $x \in X$*, if there is a constant $L = L_x > 0$ and a radius $\varepsilon > 0$ such that

$$\|F(y) - F(z)\| \leq L\|y - z\| \quad \forall y, z \in X \cap B_\varepsilon(x).$$

  The function F is called *locally Lipschitz continuous on X* it it is locally Lipschitz continuous at all $x \in X$.

1.5 Exercises

Exercise 1.1 (Complementarity Conditions). For two vectors $x, y \in \mathbb{R}^n$ the condition

$$x_i \geq 0, y_i \geq 0, x_i y_i = 0 \qquad \forall i = 1, \ldots, n$$

is called *complementarity condition*. Prove, that it is equivalent to the following conditions:

(a) $x \geq 0, y \geq 0$ and $x_i = 0$ or $y_i = 0$ for all $i = 1, \ldots, n$.

(b) $x \geq 0, y \geq 0$ and $x^T y = 0$.

(c) $x \geq 0, y \geq 0$ and $x_i \cdot y_i \leq 0$ for all $i = 1, \ldots, n$.

(d) $\min\{x_i, y_i\} = 0$ for all $i = 1, \ldots, n$.

Exercise 1.2. Consider the NLP formulation

$$\min_x f(x) \quad \text{s.t.} \quad g(x) \leq 0, h(x) = 0,$$
$$G(x) \geq 0, H(x) \geq 0, G(x) \circ H(x) \leq 0.$$

and prove that MFCQ is violated in every feasible point.

Exercise 1.3. Consider the NLP formulation

$$\min_x f(x) \quad \text{s.t.} \quad g(x) \leq 0, h(x) = 0,$$
$$G(x) \geq 0, H(x) \geq 0, G(x) \circ H(x) \leq 0.$$

and prove that the Fritz-John conditions are satisfied in every feasible point.

Exercise 1.4. Compute the Bouligand subdifferential and the Clarke subdifferential of the following norms for all $x \in \mathbb{R}^n$:

(a) $f_2(x) = \|x\|_2$

(a) $f_1(x) = \|x\|_1$

(a) $f_\infty(x) = \|x\|_\infty$

Compare the results at $x = 0$.

Exercise 1.5. Consider a nonempty, closed and *convex* set $Z \subseteq \mathbb{R}^n$ and an arbitrary point $x^* \in Z$. Prove

$$N_Z^F(x^*) = N_Z^M(x^*) = N_Z(x^*).$$

Exercise 1.6. Consider nonempty and closed sets $Z_i \subseteq \mathbb{R}^{n_i}$ for $i = 1, \ldots, m$ and define $Z := Z_1 \times \ldots \times Z_m$. Consider an arbitrary point $x^* \in Z$.

(a) Prove

$$T_Z(x^*) \subseteq T_{Z_1}(x^*) \times \ldots \times T_{Z_m}(x^*)$$

and find an example, where the inclusion is strict. Prove that equality holds, if all sets Z_i are additionally convex.

Remark: In fact it suffices to require $N_{Z_i}^F(x^*) = N_{Z_i}^M(x^*)$ for all i to guarantee equality. This is always true for convex sets Z_i, see Exercise 1.5

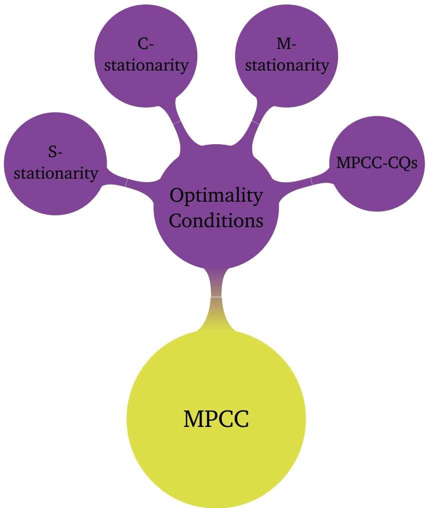
(b) Prove

$$N_Z^F(x^*) = N_{Z_1}^F(x^*) \times \ldots \times N_{Z_m}^F(x^*)$$

and

$$N_Z^M(x^*) = N_{Z_1}^M(x^*) \times \ldots \times N_{Z_m}^M(x^*)$$

2 Optimality Conditions



Throughout this section we consider mathematical programs with complementarity constraints (MPCC) of the form

$$\begin{aligned} \min_x f(x) \quad \text{s.t.} \quad & g(x) \le 0, h(x) = 0, \\ & 0 \le G(x) \perp H(x) \ge 0 \end{aligned}$$

with $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$, $G, H : \mathbb{R}^n \to \mathbb{R}^q$. We assume that all functions are at least once continuously differentiable.

We denote the feasible set by $X \subseteq \mathbb{R}^n$, i.e.

$$X := \{x \in \mathbb{R}^n \mid g(x) \le 0, h(x) = 0, 0 \le G(x) \perp H(x) \ge 0\}.$$

The feasible set of one complementarity constraint

$$G_i(x) \ge 0, H_i(x) \ge 0, G_i(x)H_i(x) = 0$$

is depicted in Figure 2.1.

This inspires the separation of the complementarity constraints into the following three index sets for a feasible point $x^* \in X$:

$$\begin{aligned} I_{0+}(x^*) &:= \{i \mid G_i(x^*) = 0, H_i(x^*) > 0\}, \\ I_{00}(x^*) &:= \{i \mid G_i(x^*) = 0, H_i(x^*) = 0\}, \\ I_{+0}(x^*) &:= \{i \mid G_i(x^*) > 0, H_i(x^*) = 0\}. \end{aligned}$$

Here, the first subscript indicates the status of G_i and the second stands for H_i. The set $I_{00}(x^*)$ is sometimes called the *biactive set*. Obviously these sets are pairwise disjoint and

$$\{1, \ldots, q\} = I_{0+}(x^*) \cup I_{00}(x^*) \cup I_{+0}(x^*).$$

Analogously to nonlinear optimization, for a feasible point $x^* \in X$ we denote the set of *active inequalities* by

$$I_g(x^*) = \{i \mid g_i(x^*) = 0\}.$$

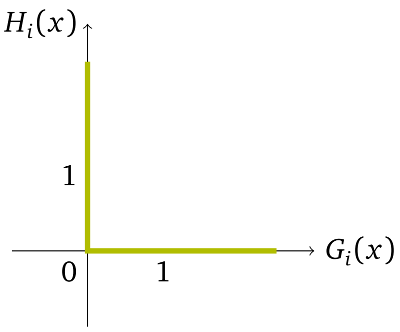


Figure 2.1: Feasible set of $G_i(x) \geq 0, H_i(x) \geq 0, G_i(x)H_i(x) = 0$

2.1 KKT based Optimality Conditions

Obviously we can interpret the MPCC as the standard nonlinear optimization problem (MPCC-NLP)

$$\begin{aligned} \min_x f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & G(x) \geq 0, \\ & H(x) \geq 0, \\ & G(x) \circ H(x) = 0, \end{aligned}$$

where $a \circ b$ denotes the componentwise (Hadamard) product of two vectors $a, b \in \mathbb{R}^q$.

And even though we have seen that local minima of an MPCC are not necessarily KKT points of MPCC-NLP, we also have seen that Guignard CQ has a chance to be satisfied. Thus, we can analyze the KKT conditions applied to MPCC-NLP and try to find suitable constraint qualifications.

2.1.1 Strong Stationarity

We know that, under constraint qualifications such as Abadie CQ or LICQ, every local solution of MPCC-NLP satisfies the KKT conditions. So let us begin by writing down these KKT conditions for MPCC-NLP:

At a KKT point x^*, there exist multipliers $\lambda^g \in \mathbb{R}^m_+$, $\mu^h \in \mathbb{R}^p$, $\lambda^G \in \mathbb{R}^q_+$, $\lambda^H \in \mathbb{R}^q_+$ and $\mu^{GH} \in \mathbb{R}^q$ such that

$$\begin{aligned} & \nabla f(x^*) + \nabla g(x^*)\lambda^g + \nabla h(x^*)\mu^h \\ + \sum_{i=1}^{q} \Big[-\lambda_i^G \nabla G_i(x^*) - \lambda^H \nabla H_i(x^*) + \mu_i^{GH}\big(H_i(x^*)\nabla G_i(x^*) + G_i(x^*)\nabla H_i(X^*)\big)\Big] & = 0 \\ , & \end{aligned}$$

and

$$\begin{aligned} 0 \geq g(x^*) \perp \lambda^g \geq 0, \\ h(x^*) = 0, \\ 0 \leq G(x^*) \perp \lambda^G \geq 0, \\ 0 \leq H(x^*) \perp \lambda^H \geq 0, \\ G(x^*) \circ H(x^*) = 0. \end{aligned}$$

Simplifying these conditions using $\gamma_i = \lambda_i^G - \mu^{GH} H_i(x^*)$ and $\nu_i = \lambda_i^H - \mu^{GH} G_i(x^*)$, we obtain

$$
\begin{aligned}
\nabla f(x^*) + \nabla g(x^*)\lambda^g + \nabla h(x^*)\mu^h - \nabla G(x^*)\gamma - \nabla H(x^*)\nu &= 0, \\
0 \geq g(x^*) \perp \lambda^g &\geq 0, \\
h(x^*) &= 0, \\
G(x^*) \geq 0, H(x^*) &\geq 0, \\
\gamma_i \in \mathbb{R},\ \nu_i &= 0 \qquad \forall i \in I_{0+}(x^*), \\
\gamma_i \geq 0,\ \nu_i &\geq 0 \qquad \forall i \in I_{0+}(x^*), \\
\gamma_i = 0,\ \nu_i &\in \mathbb{R} \qquad \forall i \in I_{+0}(x^*).
\end{aligned}
$$

This leads to our first optimality condition for MPCC, which was first introduced by Scheel and Scholtes in [28] and can also be found in [22].

Definition 2.1. *A feasible point $x^* \in X$ is called* strongly stationary (S-stationary) *if there exist multipliers $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ and $\gamma, \nu \in \mathbb{R}^q$ such that*

$$
\begin{aligned}
\nabla f(x^*) + \nabla g(x^*)\lambda + \nabla h(x^*)\mu - \nabla G(x^*)\gamma - \nabla H(x^*)\nu &= 0, \\
0 \geq g(x^*) \perp \lambda &\geq 0, \\
h(x^*) &= 0, \\
G(x^*) \geq 0, H(x^*) &\geq 0, \\
\nu_i &= 0 \qquad \forall i \in I_{0+}(x^*), \\
\gamma_i \geq 0,\ \nu_i &\geq 0 \qquad \forall i \in I_{00}(x^*), \\
\gamma_i &= 0 \qquad \forall i \in I_{+0}(x^*).
\end{aligned}
$$

From nonlinear optimization we know that the KKT conditions are necessary optimality conditions only under constraint qualifications, but we have already seen that CQs such as LICQ and MFCQ are violated at every feasible point of MPCC-NLP. Even Abadie CQ is often violated, because the tangent cone is typically nonconvex whereas the linearized tangent cone is polyhedral convex. More precisely, we have

$$
\begin{aligned}
L_X(x^*) = \{d \in \mathbb{R}^n \quad | \quad & \nabla g_i(x^*) \leq 0 \quad \forall i \in I_g(x^*), \\
& \nabla h(x^*)^T d = 0, \\
& \nabla G_i(x^*)^T d = 0 \quad \forall i \in I_{0+}(x^*), \\
& \nabla H_i(x^*)^T d = 0 \quad \forall i \in I_{+0}(x^*), \\
& \nabla G_i(x^*)^T d \geq 0, \nabla H_i(x^*)^T d \geq 0 \quad \forall i \in I_{00}(x^*)\}
\end{aligned}
$$

and

$$
\begin{aligned}
L_X(x^*)^\circ = \{w \in \mathbb{R}^n \quad | \quad & w = \nabla g(x^*)\lambda + \nabla h(x^*)\mu - \nabla G(x^*)\gamma - \nabla H(x^*)\nu, \\
& \lambda \geq 0 \text{ if } i \in I_g(x^*), \\
& \lambda_i = 0 \text{ if } i \notin I_g(x^*), \\
& \gamma_i = 0 \text{ if } i \in I_{+0}(x^*), \\
& \nu_i = 0 \text{ if } i \in I_{0+}(x^*), \\
& \gamma_i \geq 0,\ \nu_i \geq 0 \text{ if } i \in I_{00}(x^*)\}
\end{aligned}
$$

see Exercise 2.5 for details.

So let us take a closer look at the structure of the feasible set of MPCC around a point $x^* \in X$. Then for every $i \in I_{+0}(x^*)$ the continuity of G implies that $G_i(x) > 0$ and thus $H_i(x) = 0$ for all $x \in X$ close to x^*. Analogously we obtain $H_i(x) > 0$ and thus $G_i(x) = 0$ for all $x \in X$ close to x^* and all $i \in I_{0+}(x^*)$. For $i \in I_{00}(x^*)$ both $G_i(x) = 0, H_i(x^*) \geq 0$ and $G_i(x) \geq 0, H_i(x^*) = 0$ are possible in a neighborhood.

Thus, for an arbitrary set $I \subseteq I_{00}(x^*)$ we define the *tightened program* TNLP(x^*, I) as

$$\begin{aligned} \min_x f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & G_i(x) = 0, H_i(x) \geq 0 \quad \forall i \in I_{0+}(x^*) \cup I, \\ & G_i(x) \geq 0, H_i(x) = 0 \quad \forall i \in I_{+0}(x^*) \cup I^c, \end{aligned}$$

where $I^c := I_{00}(x^*) \setminus I$ denotes the complement of I with respect to the biactive set $I_{00}(x^*)$. Denote the feasible set of TNLP(x^*, I) by X_I. Then $x^* \in X_I$ for all $I \subseteq I_{00}(x^*)$ and there exists a radius $r > 0$ such that

$$X \cap B_r(x^*) = \bigcup_{I \subseteq I_{00}(x^*)} Z_I \cap B_r(x^*).$$

Since the tangent cone only depends on the feasible set in an arbitrarily small neighborhood and there are only finitely many $I \subseteq I_{00}(x^*)$, we obtain

$$T_X(x^*) = T_{X \cap B_r(x^*)}(x^*) = \bigcup_{I \subseteq I_{00}(x^*)} T_{X_I \cap B_r(x^*)}(x^*) = \bigcup_{I \subseteq I_{00}(x^*)} T_{X_I}(x^*).$$

Thus, passing to the polar cones, we obtain

$$T_X(x^*)^\circ = \bigcap_{I \subseteq I_{00}(x^*)} T_{X_I}(x^*)^\circ$$

For more details on calculus rules for cones and polar cones, see [27]. If some constraint qualification for X_I holds at x^*, we have

$$T_X(x^*)^\circ = \bigcap_{I \subseteq I_{00}(x^*)} T_{X_I}(x^*)^\circ = \bigcap_{I \subseteq I_{00}(x^*)} L_{X_I}(x^*)^\circ,$$

where

$$\begin{aligned} L_{X_I}(x^*) = \{d \in \mathbb{R}^n \quad | \quad & \nabla g_i(x^*) \leq 0 \quad \forall i \in I_g(x^*), \\ & \nabla h(x^*)^T d = 0, \\ & \nabla G_i(x^*)^T d = 0 \quad \forall i \in I_{0+}(x^*) \cup I, \\ & \nabla H_i(x^*)^T d = 0 \quad \forall i \in I_{+0}(x^*) \cup I^c, \\ & \nabla G_i(x^*)^T d \geq 0 \quad \forall i \in I^c, \\ & \nabla H_i(x^*)^T d \geq 0 \quad \forall i \in I\} \end{aligned}$$

and

$$\begin{aligned} L_{X_I}(x^*)^\circ = \{w \in \mathbb{R}^n \quad | \quad & w = \nabla g(x^*)\lambda + \nabla h(x^*)\mu - \nabla G(x^*)\gamma - \nabla H(x^*)\nu, \\ & \lambda \geq 0 \text{ if } i \in I_g(x^*), \\ & \lambda_i = 0 \text{ if } i \notin I_g(x^*), \\ & \gamma_i = 0 \text{ if } i \in I_{+0}(x^*), \\ & \nu_i = 0 \text{ if } i \in I_{0+}(x^*), \\ & \gamma_i \geq 0 \text{ if } i \in I^c, \\ & \nu_i \geq 0 \text{ if } i \in I\} \end{aligned}$$

For Guignard CQ to hold, the critical inclusion is

$$T_X(x^*)^\circ \subseteq L_X(x^*)^\circ.$$

So let us consider an arbitrary element $w \in T_X(x^*)^\circ$. If for $I \subseteq I_{00}(x^*)$ a CQ for TNLP(x^*, I) holds at x^*, we know $w \in L_{X_I}(x^*)^\circ$ and thus

$$w = \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \nabla h(x^*)\mu - \sum_{i \in I_{0+}(x^*) \cup I_{00}(x^*)} \gamma_i \nabla G_i(x^*) - \sum_{i \in I_{+0}(x^*) \cup I_{00}(x^*)} \nu_i \nabla H_i(x^*)$$

with $\lambda_i \geq 0$ for all $i \in I_g(x^*)$, $\gamma_i \geq 0$ for all $i \in I^c$ and $\nu_i \geq 0$ for all $i \in I$. If for $I^c \subseteq I_{00}(x^*)$ a CQ for TNLP(x^*, I^c) holds, too, then we also have $w \in L_{X_{I^c}}(x^*)^\circ$ and thus

$$w = \sum_{i \in I_g(x^*)} \lambda_i^c \nabla g_i(x^*) + \nabla h(x^*)\mu^c - \sum_{i \in I_{0+}(x^*) \cup I_{00}(x^*)} \gamma_i^c \nabla G_i(x^*) - \sum_{i \in I_{+0}(x^*) \cup I_{00}(x^*)} \nu_i^c \nabla H_i(x^*)$$

with $\lambda_i^c \geq 0$ for all $i \in I_g(x^*)$, $\gamma_i^c \geq 0$ for all $i \in I$ and $\nu_i^c \geq 0$ for all $i \in I^c$. If the gradients

$$\nabla g_i(x^*)(i \in I_g(x^*)),\ \nabla h_i(x^*)(i = 1, \ldots, p),\ \nabla G_i(x^*)(i \in I_{0+}(x^*) \cup I_{00}(x^*)),\ \nabla H_i(x^*)(i \in I_{+0}(x^*) \cup I_{00}(x^*))$$

are linearly independent, then w has a unique representation with respect to these gradients and thus $\lambda, \mu, \gamma, \nu$ and $\lambda^c, \mu^c, \gamma^c, \nu^c$ have to coincide. This implies

$$w = \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \nabla h(x^*)\mu - \sum_{i \in I_{0+}(x^*) \cup I_{00}(x^*)} \gamma_i \nabla G_i(x^*) - \sum_{i \in I_{+0}(x^*) \cup I_{00}(x^*)} \nu_i \nabla H_i(x^*)$$

with $\lambda_i \geq 0$ for all $i \in I_g(x^*)$, $\gamma_i \geq 0$ for all $i \in I_{00}(x^*)$ and $\nu_i \geq 0$ for all $i \in I_{00}(x^*)$. Thus we have shown $w \in L_X(x^*)^\circ$ and therefore $T_X(x^*)^\circ \subseteq L_X(x^*)^\circ$ under the above liner independence condition and the assumption that a CQ holds for the tightened programs TNLP(x^*, I). However, a closer look reveals that the used linear independence condition is exactly LICQ for all TNLP(x^*, I).

This motivates the following definition, see also [28]:

Definition 2.2. *At a feasible point $x^* \in X$ the* MPCC linear independence CQ (MPCC-LICQ) *holds, if*

$$\nabla g_i(x^*)(i \in I_g(x^*)),\ \nabla h_i(x^*)(i = 1, \ldots, p),\ \nabla G_i(x^*)(i \in I_{0+}(x^*) \cup I_{00}(x^*)),\ \nabla H_i(x^*)(i \in I_{+0}(x^*) \cup I_{00}(x^*))$$

are linearly independent.

Using this terminology, we have proven the following theorem, see also [13]:

Theorem 2.3. *Let $x^* \in X$ be feasible for the MPCC.*

(a) If MPCC-LICQ holds at x^* then standard Guignard CQ for the NLP formulation of the MPCC holds there, too.

(b) If x^* is a local minimum of MPCC and MPCC-LICQ holds there, then x^* is an S-stationary point.

For "convex" MPCCs one can also prove the opposite implication.

Theorem 2.4. *Consider MPCC, where f, g, h, G, H are continuously differentiable and f, g_i are convex and h, G, H are affine linear. Then every S-stationary point x^* is a local minimum.*

Proof. For all $I \subseteq I_{00}(x^*)$ we consider again the tightened programs TNLP(x^*, I)

$$\min_x f(x) \quad \text{s.t.} \quad \begin{aligned} & g(x) \leq 0, h(x) = 0, \\ & G_i(x) = 0, H_i(x) \geq 0 \quad \forall i \in I_{0+}(x^*) \cup I, \\ & G_i(x) \geq 0, H_i(x) = 0 \quad \forall i \in I_{+0}(x^*) \cup I^c, \end{aligned}$$

Then we know that $x^* \in X_I$ for all $I \subseteq I_{00}(x^*)$ and that there exists a radius $r > 0$ such that

$$X \cap B_r(x^*) = \bigcup_{I \subseteq I_{00}(x^*)} Z_I \cap B_r(x^*).$$

Thus, since there are only finitely many subsets $I \subseteq I_{00}(x^*)$ it suffices to show that x^* is a local minimum of TNLP(x^*, I) for all $I \subseteq I_{00}(x^*)$. However, the S-stationary point x^* is a KKT point of all TNLP(x^*, I) and all TNLP(x^*, I) are convex NLPs. Therefore, the statement follows from the corresponding standard result from nonlinear optimization.. □

In classical nonlinear optimization LICQ is the strongest CQ and we know that we can guarantee the KKT conditions under much weaker CQs. This raises the question whether weaker MPCC-CQs are sufficient for S-stationarity. To answer this question let us come back to Example 1.1.

Example 2.5. Consider again the MPCC

$$\min_{x \in \mathbb{R}^3} x_1 + x_2 - x_3 \quad \text{s.t.} \quad \begin{aligned} & -4x_1 + x_3 \leq 0, \\ & -4x_2 + x_3 \leq 0, \\ & 0 \leq x_1 \perp x_2 \geq 0, \end{aligned}$$

where we already know that the global minimum $x^* = (0, 0, 0)^T$ is not a KKT point and thus not S-stationary.

If we take a look at the gradients, we see that the four vectors

$$\begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -4 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

can of course not be linearly independent in $\mathbb{R}^3$. Consequently MPCC-LICQ is violated. But, as we will see later, MPCC-MFCQ holds at x^*.

Thus even in this very simple example, where all constraints (except for the complementarity) are linear, we cannot guarantee S-stationarity of minima as soon as MPCC-LICQ is violated.. Consequently, if we want to use weaker CQs than MPCC-LICQ, we must most likely also work with weaker optimality conditions.

2.1.2 Alternative and Weak Stationarity

Recall that we used the tightened programs TNLP(x^*, I)

$$\min_x f(x) \quad \text{s.t.} \quad \begin{aligned} & g(x) \leq 0, h(x) = 0, \\ & G_i(x) = 0, H_i(x) \geq 0 \quad \forall i \in I_{0+}(x^*) \cup I, \\ & G_i(x) \geq 0, H_i(x) = 0 \quad \forall i \in I_{+0}(x^*) \cup I^c, \end{aligned}$$

with $I \subseteq I_{00}(x^*)$ in order to prove that MPCC-LICQ guarantees S-stationarity of local minima of MPCC.

Motivated by these TNLPs, we define two more NLPs, whose feasible sets are locally around x^* a simplification of the feasible set of MPCC. The *relaxed program* RNLP(x^*) is defined as

$$\begin{aligned}\min_x f(x) \quad \text{s.t.} \quad & g(x) \le 0, h(x) = 0, \\ & G(x) = 0, H(x) \ge 0 \quad \forall i \in I_{0+}(x^*), \\ & G(x) \ge 0, H(x) = 0 \quad \forall i \in I_{+0}(x^*), \\ & G(x) \ge 0, H(x) \ge 0 \quad \forall i \in I_{00}(x^*)\end{aligned}$$

and the *tightened program* TNLP(x^*) as

$$\begin{aligned}\min_x f(x) \quad \text{s.t.} \quad & g(x) \le 0, h(x) = 0, \\ & G(x) = 0, H(x) \ge 0 \quad \forall i \in I_{0+}(x^*), \\ & G(x) \ge 0, H(x) = 0 \quad \forall i \in I_{+0}(x^*), \\ & G(x) = 0, H(x) = 0 \quad \forall i \in I_{00}(x^*).\end{aligned}$$

If we compare TNLP(x^*), RNLP(x^*) and TNLP(x^*, I), we see that they only differ in the constraints corresponding to the biactive index set $I_{00}(x^*)$. For $i \in I_{00}(x^*)$ we relax the complementarity constraint to $G_i(x) \ge 0, H_i(x) \ge 0$ in RNLP(x^*), we chose one of the two possibilities $G_i(x) = 0, H_i(x) \ge 0$ or $G_i(x) \ge 0, H_i(x) = 0$ in TNLP(x^*, I) and we tighten the constraint to $G_i(x) = 0, H_i(x) = 0$ in TNLP(x^*).

Since the feasible set of RNLP locally includes the feasible set of MPCC, which locally includes the feasible set of TNLP(x^*, I) for all $I \subseteq I_{00}(x^*)$ and for all $I \subseteq I_{00}(x^*)$ the feasible set of TNLP(x^*) is locally included in the feasible set of TNLP(x^*, I), we immediately obtain the following implications:

$$\begin{aligned} & x^* \text{is a local minimum of RNLP}(x^*) \\ \Longrightarrow \quad & x^* \text{is a local minimum of MPCC} \\ \Longrightarrow \quad & x^* \text{is a local minimum of TNLP}(x^*, I) \quad \forall I \subseteq I_{00}(x^*) \\ \Longrightarrow \quad & x^* \text{is a local minimum of TNLP}(x^*)\end{aligned}$$

Thus necessary optimality conditions for TNLP(x^*, I) and TNLP(x^*) are also local optimality conditions for MPCC.

Let us write down the KKT conditions for all auxiliary NLPs. Since they only differ in the constraints for $i \in I_{00}(x^*)$, we obtain the following for all problems:

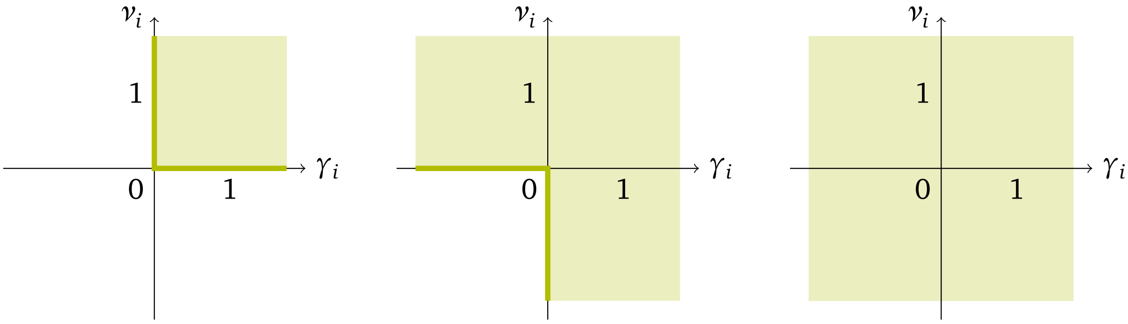
$$\begin{aligned}\nabla f(x^*) + \nabla g(x^*)\lambda + \nabla h(x^*)\mu - \nabla G(x^*)\gamma - \nabla H(x^*)\nu &= 0, \\ 0 \ge g(x^*) \perp \lambda^g &\ge 0, \\ h(x^*) &= 0, \\ G(x^*) \ge 0, H(x^*) &\ge 0, \\ \nu_i &= 0 \qquad \forall i \in I_{0+}(x^*), \\ \gamma_i &= 0 \qquad \forall i \in I_{+0}(x^*).\end{aligned}$$

For RNLP(x^*) we additionally have the condition

$$\gamma_i \ge 0, \, \nu_i \ge 0 \quad \forall i \in I_{00}(x^*),$$

for TNLP(x^*, I) we have the additional condition

$$\nu_i \ge 0 \quad \forall i \in I, \qquad \gamma_i \ge 0 \quad \forall i \in I_{00}(x^*) \setminus I,$$



(a) RNLP(x^*) **(b)** TNLP(x^*, I) **(c)** TNLP(x^*)

Figure 2.2: Sign constraints of the multipliers corresponding to $i \in I_{00}(x^*)$

and for TNLP(x^*) we do not have any sign constraints on γ_i, ν_i for $i \in I_{00}(x^*)$. These conditions are also displayed in Figure 2.2.

We see that the S-stationarity conditions are exactly the KKT conditions for RNLP(x^*), which provides another insight into why S-stationarity might be to strong as necessary optimality conditions for MPCCs. The KKT conditions for the tightened problems motivate the following optimality conditions for MPCCs, see [28, 12].

Definition 2.6. *A feasible point $x^* \in X$ is called* weakly stationary (W-stationary) *if there exist multipliers $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ and $\gamma, \nu \in \mathbb{R}^q$ such that*

$$\begin{aligned} \nabla f(x^*) + \nabla g(x^*)\lambda + \nabla h(x^*)\mu - \nabla G(x^*)\gamma - \nabla H(x^*)\nu &= 0, \\ 0 \geq g(x^*) \perp \lambda &\geq 0, \\ h(x^*) &= 0, \\ G(x^*) \geq 0, H(x^*) &\geq 0, \\ \nu_i &= 0 \qquad \forall i \in I_{0+}(x^*), \\ \gamma_i &= 0 \qquad \forall i \in I_{+0}(x^*). \end{aligned}$$

If additionally

$$\gamma_i \geq 0 \text{ or } \mu_i \geq 0 \quad \forall i \in I_{00}(x^*),$$

the point x^* is called A-stationary (A = alternative).

Since weak stationarity only considers optimality with respect to the smaller feasible set of the tightened program TNLP(x^*), using these conditions as necessary optimality conditions usually leads to too many candidates for minima of MPCC.

2.1.3 MPCC–tailored Constraint Qualifications

The tightened program can also be used to define other MPCC analogues of NLP CQs by saying that the MPCC CQ holds at $x^* \in X$ is the corresponding NLP CQ for TNLP(x^*) holds at x^*. Is is easy to see that MPCC-LICQ has this property.

To define some other useful MPCC CQs we first have to introduce the notion of positive linear dependence.

Definition 2.7. *A set of vectors $a_i \in \mathbb{R}^n$ for $i \in A$ and $b_i \in \mathbb{R}^n$ for $i \in B$ is called* positively linearly dependent *if there exist multipliers $\lambda_i \geq 0$ for $i \in A$ and $\mu_i \in \mathbb{R}$ for $i \in B$ such that $(\lambda, \mu) \neq (0, 0)$ and*

$$0 = \sum_{i \in A} \lambda_i a_i + \sum_{i \in B} \mu_i b_i.$$

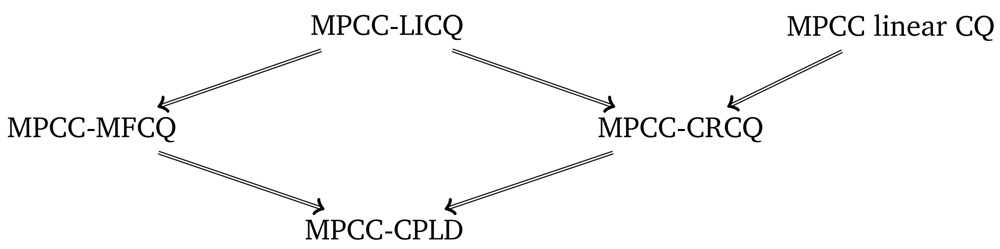


Figure 2.3: Relations between MPCC-CQs

Otherwise the vectors are called positively linearly independent.

Note that the sign restriction only applies to the first set of vectors. Obviously linear independence implies positive linear independence whereas positive linear dependence implies linear dependence.
Now we can define some more MPCC tailored CQs, see [28, 31, 16].

Definition 2.8. *Let $x^* \in X$ be feasible for MPCC.*

(a) MPCC Mangasarian-Fromovitz CQ (MPCC-MFCQ) *holds at x^* if the following gradients are positively linearly independent:*

$$\begin{aligned} &\nabla g_i(x^*)(i \in I_g(x^*)) \\ \textit{and} \quad &\nabla h_i(x^*)(i = 1, \dots, p), \nabla G_i(x^*)(i \in I_{0+}(x^*) \cup I_{00}(x^*)), \nabla H_i(x^*)(i \in I_{+0}(x^*) \cup I_{00}(x^*)). \end{aligned}$$

(b) MPCC constant rank CQ (MPCC-CRCQ) *holds at x^* if for all subsets $I_1 \subseteq I_g(x^*)$, $I_2 \subseteq \{1, \dots, p\}$, $I_3 \subseteq I_{0+}(x^*) \cup I_{00}(x^*)$, $I_4 \subseteq I_{+0}(x^*) \cup I_{00}(x^*)$ such that the gradients*

$$\nabla g_i(x)(i \in I_1), \nabla h_i(x)(i \in I_2), \nabla G_i(x)(i \in I_3), \nabla H_i(x)(i \in I_4).$$

are linearly dependent at x^*, they remain linearly dependent in a neighborhood of x^*.

(c) MPCC constant positive linear dependence CQ (MPCC-CPLD) *holds at x^* if for all subsets $I_1 \subseteq I_g(x^*)$, $I_2 \subseteq \{1, \dots, p\}$, $I_3 \subseteq I_{0+}(x^*) \cup I_{00}(x^*)$, $I_4 \subseteq I_{+0}(x^*) \cup I_{00}(x^*)$ such that the gradients*

$$\begin{aligned} &\nabla g_i(x)(i \in I_1) \\ \textit{and} \quad &\nabla h_i(x)(i \in I_2), \nabla G_i(x)(i \in I_3), \nabla H_i(x)(i \in I_4). \end{aligned}$$

are positively linearly dependent at x^*, they remain linearly dependent in a neighborhood of x^*.

(d) MPCC linear CQ*, if all constraints g, h, G, H are (affine) linear functions.*

Note that we define MPCC-MFCQ based on the alternative positive linear independence characterization (PLICQ) of MFCQ instead of the classical definition from nonlinear optimization, see Exercise 2.4 for details.

One immediately sees that MPCC-LICQ implies both MPCC-MFCQ and MPCC-CRCQ, whereas MPCC-MFCQ and MPCC-CRCQ each imply MPCC-CPLD. However, there is no relation between MPCC-MFCQ and MPCC-CRCQ. The MPCC linear CQ implies MPCC-CRCQ. See also Figure 2.3 for a visualization of the relations between the MPCC-CQs.

And due to the way these MPCC-CQs are defined, we immediately obtain the following theorem.

Theorem 2.9. *Let $x^* \in X$ be a local minimum of MPCC such that MPCC-CPLD or any stronger MPCC CQ holds. Then x^* is weakly stationary.*

If we look again at Example 2.5, all constraints are linear and thus MPCC linear CQ holds. One can also verify that MPCC-MFCQ holds. Thus, even under MPCC-MFCQ, which is only slightly weaker than MPCC-LICQ, S-stationarity is already not a necessary optimality condition anymore. Therefore, in the next sections we want to find necessary optimality conditions, which are are satisfied under these MPCC CQs but more selective than weak stationarity.

2.2 C-Stationarity

An alternative approach to obtain necessary optimality conditions for the MPCC is based on a reformulation of the complementarity conditions using a so called NCP function.

Definition 2.10. *A function* $\varphi : \mathbb{R}^2 \to \mathbb{R}$ *is called an* NCP function (nonlinear complementarity problem)*, if*

$$\varphi(a,b) = 0 \quad \Longleftrightarrow \quad 0 \leq a \perp b \geq 0.$$

The two most prominent examples for NCP functions are given below:

- *minimum function*: $\varphi(a,b) = \min\{a,b\}$
- *Fischer-Burmeister function*: $\varphi(a,b) = \sqrt{a^2+b^2} - (a+b)$

For more examples and useful properties of NCP functions see [32].

As the name already indicates, an NCP function φ can be used to reformulate complementarity constraints as equations

$$0 \leq G_i(x) \perp H_i(x) \geq 0 \quad \Longleftrightarrow \quad \varphi(G_i(x), H_i(x)) = 0.$$

Thus we can equivalently reformulate the MPCC

$$\begin{aligned} \min_x f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & 0 \leq G(x) \perp H(x) \geq 0 \end{aligned}$$

as

$$\begin{aligned} \min_x f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & \varphi(G_i(x), H_i(x)) = 0 \quad \forall i = 1, \dots, q. \end{aligned}$$

The only problem is that NPC functions usually are nondifferentiable at $(a,b) = (0,0)$. One can construct differentiable NCP functions, but these then necessarily have the property $\nabla\varphi(0,0) = (0,0)^T$. Using such a differentiable NCP function to reformulate the MPCC would lead to

$$\nabla\varphi(G_i(x), H_i(x)) = 0 \quad \forall i \in I_{00}(x^*)$$

and thus result in the same problems as the NLP formulation of the MPCC.

In the remainder of this section, we consider only the minimum function

$$\varphi(a,b) = \min\{a,b\}.$$

Thus function is differentiable for all (a,b) such that $a \neq b$.

Now consider an arbitrary feasible point $x^* \in X$ and rewrite the complementarity constraints as

$$\varphi(G_i(x), H_i(x)) = \min\{G_i(x), H_i(x)\} = 0.$$

For all $i \in I_{0+}(x^*)$ we locally have $H_i(x) > 0$ and thus the constraint locally reduces to the equation

$$\min\{G_i(x), H_i(x)\} = G_i(x) = 0.$$

For all $i \in I_{+0}(x^*)$ we locally have $G_i(x) > 0$ and thus the constraint reduces to

$$\min\{G_i(x), H_i(x)\} = H_i(x) = 0.$$

Consequently, the minimum function is differentiable for all $I \in I_{0+}(x^*) \cup I_{+0}(x^*)$ with

$$\nabla \min\{G_i(x^*), H_i(x^*)\} = \begin{cases} \nabla G_i(x^*) & \text{if } i \in I_{0+}(x^*), \\ \nabla H_i(x^*) & \text{if } i \in I_{+0}(x^*). \end{cases}$$

Unfortunately, in case $i \in I_{00}(x^*)$ we have

$$\min\{G_i(x^*), H_i(x^*)\} = G_i(x^*) = H_i(x^*) = 0$$

and thus the map $x \mapsto \min\{G_i(x), H_i(x)\}$ may be nondifferentiable at x^*. This is a problem if we want to obtain KKT-type optimality conditions. For this reason, we use the Clarke subdifferential as a substitute for the gradient.

Now we want to apply this approach to our nonsmooth reformulation of the MPCC as

$$\begin{aligned} \min_x f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & \min\{G_i(x), H_i(x)\} = 0 \quad \forall i = 1, \ldots, q. \end{aligned}$$

Since we assume that all functions f, g, h, G, H are continuously differentiable, they are locally Lipschitz continuous and so is $\min\{G_i(x), H_i(x)\}$. The Clarke subdifferential of f, g_i, h_i consists only of the respective gradients due to the *continuous* differentiability. For all $i \notin I_{00}(x^*)$ we have already seen that $\min\{G_i(x), H_i(x)\}$ is continuously differentiable and computed the gradient.

It remains to compute the Clarke subdifferential of $\min\{G_i(x), H_i(x)\}$ for all $i \in I_{00}(x^*)$. To do so, we can use the chain rule, which for $i \in I_{00}(x^*)$ provides the estimate

$$\partial^C \min\{G_i, H_i\}(x^*) \subseteq \operatorname{conv}\{\nabla G_i(x^*), \nabla H_i(x^*)\}.$$

Although the inner function is continuously differentiable, the outer function is nonconvex and thus the chain rule does not guarantee equality. However, one can show that

$$\partial^C \min\{G_i, H_i\}(x^*) = \operatorname{conv}\{\nabla G_i(x^*), \nabla H_i(x^*)\}.$$

If we assume e.g. $G_i(x) \geq H_i(x)$ in a neighborhood of x^*. Due to $H_i(x^*) = G_i(x^*) = 0$ the function $G_i(x) - H_i(x)$ then attains a local minimum at x^*, which implies $\nabla G_i(x^*) - \nabla H_i(x^*) = 0$ and thus

$$\partial^C \min\{G_i, H_i\}(x^*) = \{\nabla H_i(x^*)\} = \operatorname{conv}\{\nabla G_i(x^*), \nabla H_i(x^*)\}.$$

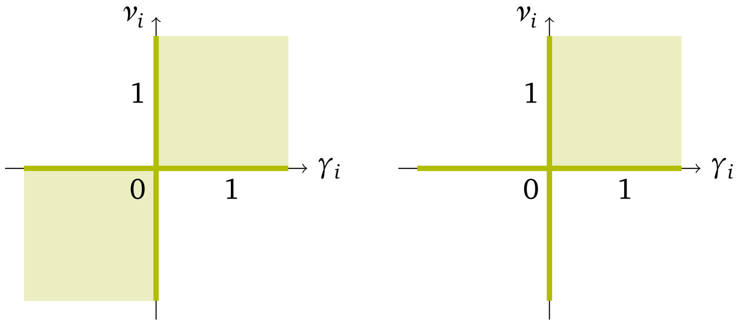
Thus the formula holds in all possible cases.

At a local minimum the Fritz-John conditions thus guarantee the existence of multipliers $\alpha \geq 0, \lambda \geq 0, \mu, \delta$ not all zero such that

$$\begin{aligned} 0 \quad \in \quad & \alpha \nabla f(x^*) + \sum_{i \in I_g} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{n} \mu_i \nabla h_i(x^*) \\ & + \sum_{i \in I_{0+}} \delta_i \nabla G_i(x^*) + \sum_{i \in I_{+0}} \delta_i \nabla H_i(x^*) + \sum_{i \in I_{00}} \delta_i \operatorname{conv}\{\nabla G_i(x^*), \nabla H_i(x^*)\}. \end{aligned}$$

Let $c_i \nabla G_i(x^*) + (1 - c_i) \nabla H_i(x^*)$ with $c_i \in [0, 1]$ be the needed elements of the convex hull. Then we can define

$$\gamma_i = \begin{cases} -\delta_i & \text{if } i \in I_{0+}(x^*), \\ -\delta_i c_i & \text{if } i \in I_{00}(x^*), \\ 0 & \text{if } i \in I_{+0}(x^*), \end{cases} \quad \text{and} \quad \nu_i = \begin{cases} -\delta_i & \text{if } i \in I_{+0}(x^*), \\ -\delta_i (1 - c_i) & \text{if } i \in I_{00}(x^*), \\ 0 & \text{if } i \in I_{0+}(x^*) \end{cases}$$



(a) C-stationarity **(b)** M-stationarity

Figure 2.4: Sign constraints on the multipliers corresponding to $i \in I_{00}(x^*)$

and obtain the condition

$$\alpha \nabla f(x^*) + \nabla g(x^*)\lambda + \nabla h(x^*)\mu - \nabla G(x^*)\gamma - \nabla H(x^*)\nu = 0,$$

where

$$\gamma_i \nu_i = c_i(1-c_i)\delta_i^2 \geq 0 \quad \forall i \in I_{00}(x^*).$$

While these Fritz-John conditions hold at every local minimum x^* without the need for a CQ, the multiplier α may be equal to zero. In this degenerate case the optimality condition is independent from the objective function. However, if $\alpha = 0$, we know that $\lambda, \mu, \nu, \gamma$ are not all equal to zero and

$$\nabla g(x^*)\lambda + \nabla h(x^*)\mu - \nabla G(x^*)\gamma - \nabla H(x^*)\nu = 0$$

Due to $\lambda \geq 0$ this cannot happen, if MPCC-MFCQ holds at x^*.

In case $\alpha > 0$ we can without loss of generality assume $\alpha = 1$ and thus obtain the following optimality conditions.

Definition 2.11. *A feasible point $x^* \in X$ is called* Clarke stationary (C-stationary) *if there exist multipliers $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ and $\gamma, \nu \in \mathbb{R}^q$ such that*

$$\begin{aligned} \nabla f(x^*) + \nabla g(x^*)\lambda + \nabla h(x^*)\mu - \nabla G(x^*)\gamma - \nabla H(x^*)\nu &= 0, \\ 0 \geq g(x^*) \perp \lambda &\geq 0, \\ h(x^*) &= 0, \\ G(x^*) \geq 0, H(x^*) &\geq 0, \\ \nu_i &= 0 \qquad \forall i \in I_{0+}(x^*), \\ \gamma_i &= 0 \qquad \forall i \in I_{+0}(x^*). \end{aligned}$$

and additionally

$$\gamma_i \cdot \mu_i \geq 0 \quad \forall i \in I_{00}(x^*).$$

A comparison of these optimality conditions with our previous ones yields that C-stationarity is weaker than S-stationarity, stronger than weak stationarity and has no immediate relation to A-stationarity.

Our observation concerning MPCC-MFCQ immediately yields the following result.

Theorem 2.12. *Let $x^* \in X$ be a local minimum of MPCC, where MPCC-MFQ holds. Then x^* is C-stationary.*

In fact, one can show that local minima are C-stationary under much weaker constraint qualifications such as an MPCC analogue of GCQ. We later obtain this as a corollary of a stronger result.

So far, we know that a local minimum x^* of MPCC satisfying MPCC-MFCQ must be both A- and C-stationary. If even MPCC-LICQ holds, the multipliers for both must be the same and thus satisfy both

$$\gamma_i \geq 0 \quad \text{or} \quad \nu_i \geq 0 \quad \forall i \in I_{00}(x^*)$$

and

$$\gamma_i \nu_i \geq 0 \quad \forall i \in I_{00}(x^*).$$

This results in the condition

$$\gamma_i \cdot \nu_i = 0 \quad \text{or} \quad \gamma_i \geq 0, \nu_i \geq 0 \quad \forall i \in I_{00}(x^*),$$

which illustrated in Figure 2.4. Feasible points of MPCC satisfying this stronger condition compared to A- or C-stationarity are later called M-stationary. Of course we already know that a local minimum satisfying MPCC-LICQ is S-stationary, which is even stronger than M-stationarity. However, we have already seen that it is not possible to weaken the assumption MPCC-LICQ significantly as local minima satisfying MPCC-MFCQ may already violate the S-stationarity conditions. In contrast to this, we will see that local minima are M-stationary even under MPCC-analogues of Guignard CQ.

2.3 M-Stationarity

We know that S-stationarity is a necessary optimality condition for the MPCC only under MPCC-LICQ and may already be violated under MPCC-FCQ. Thus, our goal is to derive an alternative stationarity condition, which is a strong as possible but holds under weaker constraint qualifications such as MPCC analogues of GUignard and Abadie CQ.

2.3.1 MPCC Analogues of Abadie and Guignard CQ

As we have seen already, classical Abadie CQ is most likely not satisfied for MPCCs whereas classical Guignard CQ has at least a chance to hold.

So far, we have introduces MPCC analogues of CQs using the tightened program TNLP(x^*)

$$\begin{aligned}
\min_x f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\
& G_i(x) \geq 0, H_i(x) = 0 \quad \forall i \in I_{+0}(x^*), \\
& G_i(x) = 0, H_i(x) \geq 0 \quad \forall i \in I_{0+}(x^*), \\
& G_i(x) = 0, H_i(x) = 0 \quad \forall i \in I_{00}(x^*).
\end{aligned}$$

To see that it does not make a lot of sense to define MPCC-ACQ and MPCC-GCQ the same way, consider the following example.

Example 2.13. Consider the simple MPCC

$$\min_{x \in \mathbb{R}^2} f(x) \quad \text{s.t.} \quad 0 \leq x_1 \perp x_2 \geq 0$$

and $x^* = 0$. Then the feasible set of TNLP(x^*) is equal to $\{x^*\}$ and thus the corresponding tangent cone and linearized tangent cone also consist only of $\{x^*\}$. However, the tangent cone of the original MPCC at x^* is the whole feasible set of MPCC. Thus, TLNP(x^*) satisfies ACQ and GCQ. (In fact this holds for general MPCCs under the previously defined MPCC-CQs.) But the considered tangent cones and linearized tangent cones do not reflect the local structure of the feasible set of MPCC.

The problem with ACQ was that the linearized tangent cone by definition is always convex, whereas the tangent cone of an MPCC usually is not. A possibility to adapt the definition of the linearized tangent cone better to the structure of MPCC is given next.

Definition 2.14. *Consider a point $x^* \in X$ feasible for MPCC. Then the* MPCC-linearized tangent cone *of MPCC at x^* is defined as*

$$\begin{aligned} L_X^{MPCC}(x^*) := \{d \in \mathbb{R}^n \quad | \quad & \nabla g_i(x^*)^T d \leq 0 \quad \forall i \in I_g(x^*), \\ & \nabla h_i(x^*)^T d = 0 \quad \forall i = 1, \ldots, p, \\ & \nabla G_i(x^*)^T d = 0 \quad \forall i \in I_{0+}(x^*), \\ & \nabla H_i(x^*)^T d = 0 \quad \forall i \in I_{+0}(x^*), \\ & \nabla G_i(x^*)^T d \geq 0 \quad \forall i \in I_{00}(x^*), \\ & \nabla H_i(x^*)^T d \geq 0 \quad \forall i \in I_{00}(x^*), \\ & (\nabla G_i(x^*)^T d)(\nabla H_i(x^*)^T d) = 0 \quad \forall i \in I_{00}(x^*)\}. \end{aligned}$$

This cone is related to the classical linearized tangent cone of X via

$$L_X^{\text{MPCC}}(x^*) = L_X(x^*) \cap \{d \in \mathbb{R}^n \mid (\nabla G_i(x^*)^T d)(\nabla H_i(x^*)^T d) = 0 \quad \forall i \in I_{00}(x^*)\}.$$

Consequently, we always have the inclusion

$$L_X^{\text{MPCC}}(x^*) \subseteq L_X(x^*)$$

and $L_X^{\text{MPCC}}(x^*)$ can be nonconvex. Moreover, one can show (see Exercise 2.7) that the inclusion

$$T_X(x^*) \subseteq L_X^{\text{MPCC}}(X^*)$$

also holds for all $x^* \in X$. This motivates the following definition.

Definition 2.15. *Consider a point $x^* \in X$ feasible for MPCC.*

(a) MPCC Abadie CQ (MPCC-ACQ) *holds at x^* if*

$$T_X(x^*) = L_X^{MPCC}(x^*).$$

(g) MPCC Guignard CQ (MPCC-GCQ) *holds at x^* if*

$$T_X(x^*)^\circ = L_X^{MPCC}(x^*)^\circ.$$

Since we defined MPCC-ACQ and MPCC-GCQ differently from the rest of the MPCC-CQs, one may ask how the relation between them is. So far, the weakest MPCC-CQ we have introduced, was MPCC-CPLD. And one can show that, same as in the standard NLP case, MPCC-CPLD implies MPCC-ACQ.

Theorem 2.16. *Consider a point $x^* \in X$ feasible for MPCC. If MPCC-CPLD holds at x^* then so does MPCC-ACQ.*

Proof. For the given point x^* consider again the tightened programs TNLP(x^*, I)

$$\begin{aligned} \min_x f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & G_i(x) \geq 0, H_i(x) = 0 \quad \forall i \in I_{+0}(x^*) \cup I, \\ & G_i(x) = 0, H_i(x) \geq 0 \quad \forall i \in I_{0+}(x^*) \cup I^c \end{aligned}$$

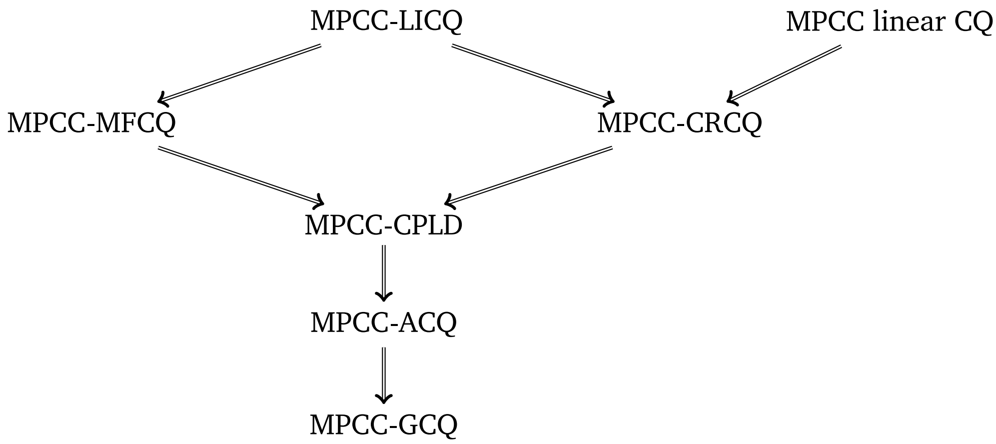


Figure 2.5: Relations between MPCC-CQs

for all subsets $I \subseteq I_{00}(x^*)$.

One can show that MPCC-CPLD implies standard CPLD for all TNLP(x^*, I), see Exercise 2.8. From nonlinear optimization we know that this implies that Abadie CQ for TNLP(x^*, I) also holds at x^*. Thus, we have

$$\begin{aligned} T_{X_I}(x^*) = L_{X_I}(x^*) = \{d \in \mathbb{R}^n \quad | \quad & \nabla g_i(x^*) \le 0 \quad \forall i \in I_g(x^*), \\ & \nabla h(x^*)^T d = 0, \\ & \nabla G_i(x^*)^T d = 0 \quad \forall i \in I_{0+}(x^*) \cup I, \\ & \nabla H_i(x^*)^T d = 0 \quad \forall i \in I_{+0}(x^*) \cup I^c, \\ & \nabla G_i(x^*)^T d \ge 0 \quad \forall i \in I^c, \\ & \nabla H_i(x^*)^T d \ge 0 \quad \forall i \in I\} \end{aligned}$$

for all $I \subseteq I_{00}(x^*)$. Since we already know that

$$T_X(x^*) = \bigcup_{I \subseteq I_{00}} T_{X_I}(x^*),$$

we thus obtain

$$\begin{aligned} T_X(x^*) = \bigcup_{I \subseteq I_{00}} \{d \in \mathbb{R}^n \quad | \quad & \nabla g_i(x^*) \le 0 \quad \forall i \in I_g(x^*), \\ & \nabla h(x^*)^T d = 0, \\ & \nabla G_i(x^*)^T d = 0 \quad \forall i \in I_{0+}(x^*), \\ & \nabla H_i(x^*)^T d = 0 \quad \forall i \in I_{+0}(x^*), \\ & \nabla G_i(x^*)^T d \ge 0, \nabla H_i(x^*)^T d = 0 \quad \forall i \in I^c, \\ & \nabla G_i(x^*)^T d = 0, \nabla H_i(x^*)^T d \ge 0 \quad \forall i \in I\} = L_X^{\text{MPCC}}(x^*). \end{aligned}$$

Thus, MPCC-ACQ holds at x^*. □

The relations between all MPCC-CQs introduced so far are collected in Figure 2.5.

2.3.2 M-Stationarity

Recall that we are still looking for a necessary optimality condition for MPCCs, which is stronger than A- and C-stationarity but holds under weaker MPCC-CQs than MPCC-LICQ. To derive the M-stationarity conditions, we begin with an argument similar to the KKT conditions.

Let $x^* \in X$ be a local minimum of MPCC. Then x^* is B-stationary, i.e.

$$\nabla f(x^*)^T d \geq 0 \quad \forall d \in T_X(x^*).$$

Contrary to the KKT conditions/S-stationarity, we now only assume that MPCC-GCQ holds at x^*. Then we obtain

$$-\nabla f(x^*) \in T_X(x^*)^\circ = L_X^{\text{MPCC}}(x^*)^\circ$$

and thus

$$\nabla f(x^*)^T \geq 0 \quad \forall d \in L_X^{\text{MPCC}}(x^*).$$

This implies that $d^* = 0$ is the (global) minimum of

$$\min_d \nabla f(x^*)^T d \quad \text{s.t.} \quad d \in L_X^{\text{MPCC}}(x^*).$$

Due to the definition of the MPCC linearized tangent cone, this is an MPCC with affine constraints. To simplify the feasible set even more, we define $D := D_1 \cap D_2$, where

$$\begin{aligned}
D_1 = \{(d,u,v) \in R^{n+2|I_{00}|} \quad | \quad & \nabla g_i(x^*)^T d \leq 0 \quad \forall i \in I_g(x^*), \\
& \nabla h_i(x^*)^T d = 0 \quad \forall i = 1,\dots,p, \\
& \nabla G_i(x^*)^T d = 0 \quad \forall i \in I_{0+}(x^*), \\
& \nabla H_i(x^*)^T d = 0 \quad \forall i \in I_{+0}(x^*), \\
& \nabla G_i(x^*)^T d - u_i = 0 \quad \forall i \in I_{00}(x^*), \\
& \nabla H_i(x^*)^T d - v_i = 0 \quad \forall i \in I_{00}(x^*)\}, \\
D_2 = \{(d,u,v) \in R^{n+2|I_{00}|} \quad | \quad & 0 \leq u \perp v \geq 0\}
\end{aligned}$$

So we separate the "simple" constraints from the "complicated" ones by introducing slack variables u, v and collecting the standard constraints in D_1 and the complementarity constraints with $i \in I_{00}(x^*)$ in D_2. Obviously d is feasible for the affine MPCC if and only if $(d, G_{I_{00}}(x^*), H_{I_{00}}(x^*)) \in D$. Consequently, $(0,0,0)^T$ is a solution of

$$\min_{(d,u,v)} \nabla f(x^*)^T d \quad \text{s.t.} \quad (d,u,v) \in D = D_1 \cap D_2.$$

Thus, $(0,0,0)^T$ is B-stationary and therefore

$$-\begin{pmatrix} \nabla f(x^*) \\ 0 \\ 0 \end{pmatrix} \in T_D(0,0,0)^\circ = T_{D_1 \cap D_2}(0,0,0)^\circ.$$

Since both D_1 and D_2 have a simple structure (D_1 is affine), we can compute the corresponding tangent and polar cones

$$\begin{aligned}
T_{D_1}(0,0,0) = L_{D_1}(0,0,0) = \{(s_1,s_2,s_2) \quad | \quad & \nabla g_i(x^*)^T s_1 \leq 0 \quad \forall i \in I_g(x^*), \\
& \nabla h_i(x^*)^T s_1 = 0 \quad \forall i = 1,\dots,p, \\
& \nabla G_i(x^*)^T s_1 = 0 \quad \forall i \in I_{0+}(x^*), \\
& \nabla H_i(x^*)^T s_1 = 0 \quad \forall i \in I_{0+}(x^*), \\
& \nabla G_i(x^*)^T s_1 - (s_2)_i = 0 \quad \forall i \in I_{00}(x^*), \\
& \nabla H_i(x^*)^T s_1 - (s_3)_i = 0 \quad \forall i \in I_{00}(x^*)\},
\end{aligned}$$

$$T_{D_1}(0,0,0)^\circ = \{w \quad | \quad w = \sum_{i\in I_g} \lambda_i \begin{pmatrix} \nabla g_i(x^*) \\ 0 \\ 0 \end{pmatrix} + \sum_{i=1}^{p} \mu_i \begin{pmatrix} \nabla h_i(x^*) \\ 0 \\ 0 \end{pmatrix} + \sum_{i\in I_{0+}} \gamma_i \begin{pmatrix} \nabla G_i(x^*) \\ 0 \\ 0 \end{pmatrix}$$
$$+ \sum_{i\in I_{+0}} \nu_i \begin{pmatrix} \nabla H_i(x^*) \\ 0 \\ 0 \end{pmatrix} + \sum_{i\in I_{00}} \gamma_i \begin{pmatrix} \nabla G_i(x^*) \\ -e_i \\ 0 \end{pmatrix} + \sum_{i\in I_{00}} \nu_i \begin{pmatrix} \nabla H_i(x^*) \\ 0 \\ -e_i \end{pmatrix},$$
$$\lambda \geq 0\}$$

and

$$\begin{aligned} T_{D_2}(0,0,0) &= \{(s_1, s_2, s_2) \mid 0 \leq s_2 \perp s_3 \geq 0\}, \\ T_{D_2}(0,0,0)^\circ &= \{(w_1, w_2, w_3) \mid w_2 \leq 0, w_3 \leq 0\}. \end{aligned}$$

Now to be able to use these formulae, we have to figure out how $T_{D_1}(0,0,0)^\circ$ and $T_{D_2}(0,0,0)^\circ$ are related to $T_{D_1\cap D_2}(0,0,0)^\circ$. Unfortunately, while we have the rules

$$T_{A\cup B}(x) = T_A(x) \cup T_B(x) \quad \text{and} \quad T_{A\cup B}(x)^\circ = T_A(x)^\circ \cap T_B(x)^\circ$$

for unions of sets, there are in general no such calculus rules for intersections of sets.

Example 2.17. Consider

$$A = \{x \in \mathbb{R} \mid x \sin \tfrac{1}{x} = 0\} \quad \text{and} \quad B = \{x \in \mathbb{R} \mid x \cos \tfrac{1}{x} = 0\}.$$

Then $A\cap B = \{0\}$ and thus $T_{A\cap B}(0) = \{0\}$, $T_{A\cap B}(0)^\circ = \mathbb{R}$. On the other hand, we obtain $T_A(0) = T_B(0) = \mathbb{R}$ and thus $T_A(0)^\circ = T_B(0)^\circ = \{0\}$. Consequently, for this example

$$T_{A\cap B}(0) = \{0\} \neq \mathbb{R} = T_A(0) \cap T_B(0).$$

Also for the polar cones, we see $T_{A\cap B}(0)^\circ = \mathbb{R}$ but all intuitive combinations in this example we obtain $(T_A(0) \cap T_B(0))^\circ = T_A(0)^\circ \cup T_B(0)^\circ = T_A(0)^\circ + T_B(0)^\circ = \{0\}$.

We will later show that for certain cones the intersection of sets can be realized by adding the respective cones. However, for the polar cone of the (Bouligand) tangent cone, we already know that in general

$$T_{D_1\cap D_2}(0,0,0)^\circ \not\subseteq T_{D_1}(0,0,0)^\circ + T_{D_2}(0,0,0)^\circ$$

because otherwise our formulae for the polar cones would imply that every local minimum of MPCC is S-stationary under MPCC-GCQ. But we have seen an example, where the local minimum is not S-stationary even under the stronger MPCC-MFCQ.

To solve this problem, we use the *limiting normal cone*, which has better calculus properties. To do so, recall that we know

$$-\begin{pmatrix} \nabla f(x^*) \\ 0 \\ 0 \end{pmatrix} \in N^F(0,0,0) = N^F_{D_1\cap D_2}(0,0,0) \subseteq N^M_{D_1\cap D_2}(0,0,0)$$

where

$$\begin{aligned} D_1 = \{(d,u,v) \in R^{n+2|I_{00}|} \quad | \quad & \nabla g_i(x^*)^T d \leq 0 \quad \forall i \in I_g(x^*), \\ & \nabla h_i(x^*)^T d = 0 \quad \forall i = 1, \dots, p, \\ & \nabla G_i(x^*)^T d = 0 \quad \forall i \in I_{0+}(x^*), \\ & \nabla H_i(x^*)^T d = 0 \quad \forall i \in I_{+0}(x^*), \\ & \nabla G_i(x^*)^T d - u_i = 0 \quad \forall i \in I_{00}(x^*), \\ & \nabla H_i(x^*)^T d - v_i = 0 \quad \forall i \in I_{00}(x^*)\}, \\ D_2 = \{(d,u,v) \in R^{n+2|I_{00}|} \quad | \quad & 0 \leq u \perp v \geq 0\}. \end{aligned}$$

We have to show that the set-valued map

$$M(y) := \{(d,u,v) \in D_1 \mid (d,u,v) + y \in D_2\}$$

is a polyhedral set-valued map. To see this, consider the graph

$$\begin{aligned}
\mathrm{graph}(M) &= \{(y,d,u,v) \mid (d,u,v) \in D_1, (d,u,v)+y \in D_2\} \\
&= \{(y,d,u,v) \mid \nabla g_i(x^*)^T d \le 0 \quad \forall i \in I_g(x^*), \\
&\qquad \nabla h_i(x^*)^T d = 0 \quad \forall i = 1,\dots,p, \\
&\qquad \nabla G_i(x^*)^T d = 0 \quad \forall i \in I_{0+}(x^*), \\
&\qquad \nabla H_i(x^*)^T d = 0 \quad \forall i \in I_{+0}(x^*), \\
&\qquad \nabla G_i(x^*)^T d - u_i = 0 \quad \forall i \in I_{00}(x^*), \\
&\qquad \nabla H_i(x^*)^T d - v_i = 0 \quad \forall i \in I_{00}(x^*), \\
&\qquad 0 \le u - y_u \perp v + y_v \ge 0\} \\
&= \bigcup_{I \subseteq I_{00}} \{(y,d,u,v) \mid \nabla g_i(x^*)^T d \le 0 \quad \forall i \in I_g(x^*), \\
&\qquad \nabla h_i(x^*)^T d = 0 \quad \forall i = 1,\dots,p, \\
&\qquad \nabla G_i(x^*)^T d = 0 \quad \forall i \in I_{0+}(x^*), \\
&\qquad \nabla H_i(x^*)^T d = 0 \quad \forall i \in I_{+0}(x^*), \\
&\qquad \nabla G_i(x^*)^T d - u_i = 0 \quad \forall i \in I_{00}(x^*), \\
&\qquad \nabla H_i(x^*)^T d - v_i = 0 \quad \forall i \in I_{00}(x^*), \\
&\qquad (u-y_u)_i = 0, (v+y_v)_i \ge 0 \quad \forall i \in I \\
&\qquad (u-y_u)_i \ge 0, (v+y_v)_i \ge 0 \quad \forall i \in I_{00}(x^*) \setminus I\},
\end{aligned}$$

which is obviously the union of finitely many polyhedra.

Thus, M is calm, which implies that

$$N^M_{D_1 \cap D_2}(0,0,0) \subseteq N^M_{D_1}(0,0,0) + N^M_{D_2}(0,0,0).$$

Since D_1 is polyhedral convex, we have

$$\begin{aligned}
N^M_{D_1}(0,0,0) &= N^F_{D_1}(0,0,0) = T_{D_1}(0,0,0)^\circ \\
&= \{w \mid w = \sum_{i \in I_g} \lambda_i \begin{pmatrix} \nabla g_i(x^*) \\ 0 \\ 0 \end{pmatrix} + \sum_{i=1}^{p} \mu_i \begin{pmatrix} \nabla h_i(x^*) \\ 0 \\ 0 \end{pmatrix} + \sum_{i \in I_{0+}} \gamma_i \begin{pmatrix} \nabla G_i(x^*) \\ 0 \\ 0 \end{pmatrix} \\
&\qquad + \sum_{i \in I_{+0}} \nu_i \begin{pmatrix} \nabla H_i(x^*) \\ 0 \\ 0 \end{pmatrix} + \sum_{i \in I_{00}} \gamma_i \begin{pmatrix} \nabla G_i(x^*) \\ -e_i \\ 0 \end{pmatrix} + \sum_{i \in I_{00}} \nu_i \begin{pmatrix} \nabla H_i(x^*) \\ 0 \\ -e_i \end{pmatrix}, \\
&\qquad \lambda \ge 0\}.
\end{aligned}$$

Using

$$C := \{(a,b) \in \mathbb{R}^2 \mid 0 \le a \perp b \ge 0\},$$

we can write (up to a slight reordering of the components of u and v)

$$D_2 = \mathbb{R}^n \times C^p$$

and thus obtain the limiting normal cone using the rules for cartesian products as

$$N^M_{D_2}(0,0,0) = \{w=(w_d,w_u,w_v) \mid w_d=0, \\ (w_u)_i(w_v)_i = 0 \text{ or } (w_u)_i \le 0, (w_v)_i \le 0 \quad \forall i \in I_{00}(x^*)\}.$$

Putting all pieces together, our optimality condition thus reads

$$-\begin{pmatrix} \nabla f(x^*) \\ 0 \\ 0 \end{pmatrix} \in N^F_{D_1\cap D_2}(0,0,0) \subseteq N^M_{D_1\cap D_2}(0,0,0) \subseteq N^M_{D_1}(0,0,0)+N^M_{D_2}(0,0,0)$$

and implies

$$\begin{aligned} -\nabla f(x^*) &= \sum_{i\in I_g} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) + \sum_{i\in I_{0+}} \gamma_i \nabla G_i(x^*) \\ &\quad + \sum_{i\in I_{+0}} \nu_i \nabla H_i(x^*) + \sum_{i\in I_{00}} \gamma_i \nabla G_i(x^*) + \sum_{i\in I_{00}} \nu_i \nabla H_i(x^*), \\ 0 &= -\gamma_i + (w_u)_i \quad \forall i \in I_{00}(x^*), \\ 0 &= -\nu_i + (w_v)_i \quad \forall i \in I_{00}(x^*), \end{aligned}$$

where $\lambda \ge 0$ and $(w_u)_i(w_v)_i = 0$ or $(w_u)_i \le 0, (w_v)_i \le 0$ for all $i \in I_{00}(x^*)$.
Bringing this into a form similar to our previous optimality conditions yields:

Definition 2.18. *A point $x^* \in X$ feasible for MPCC is called* M-stationary (Mordukhovich)*, if there exist multipliers $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ and $\gamma, \nu \in \mathbb{R}^q$ such that*

$$\begin{aligned} \nabla f(x^*) + \nabla g(x^*)\lambda + \nabla h(x^*)\mu - \nabla G(x^*)\gamma - \nabla H(x^*)\nu = 0, & \\ 0 \ge g(x^*) \perp \lambda \ge 0, & \\ h(x^*) = 0, & \\ G(x^*) \ge 0, H(x^*) \ge 0, & \\ \nu_i = 0 & \quad \forall i \in I_{0+}(x^*), \\ \gamma_i = 0 & \quad \forall i \in I_{+0}(x^*). \end{aligned}$$

and additionally

$$\gamma_i \cdot \mu_i = 0 \text{ or } \gamma_i \ge 0, \nu_i \ge 0 \quad \forall i \in I_{00}(x^*).$$

Due to our derivation of the M-stationarity conditions, we have already proven the following result.

Theorem 2.19. *Consider a local minimum $x^* \in X$ of MPCC, where MPCC-GCQ holds. Then x^* is M-stationary (and thus also C-, A- and W-stationary).*

Consequently M-stationarity is a necessary optimality condition, which is not much weaker than S-stationarity but, contrary to S-stationarity, holds under weak MPCC-CQs such as MPCC-GCQ.
To illustrate the advantage of M-stationarity over weaker stationarity concepts such as C-stationarity, consider the following example:

Example 2.20. Consider the MPCC

$$\min_{x\in\mathbb{R}^2} (x_1-1)^2 + (x_2-1)^2 \quad \text{s.t.} \quad 0 \le x_1 \perp x_2 \ge 0.$$

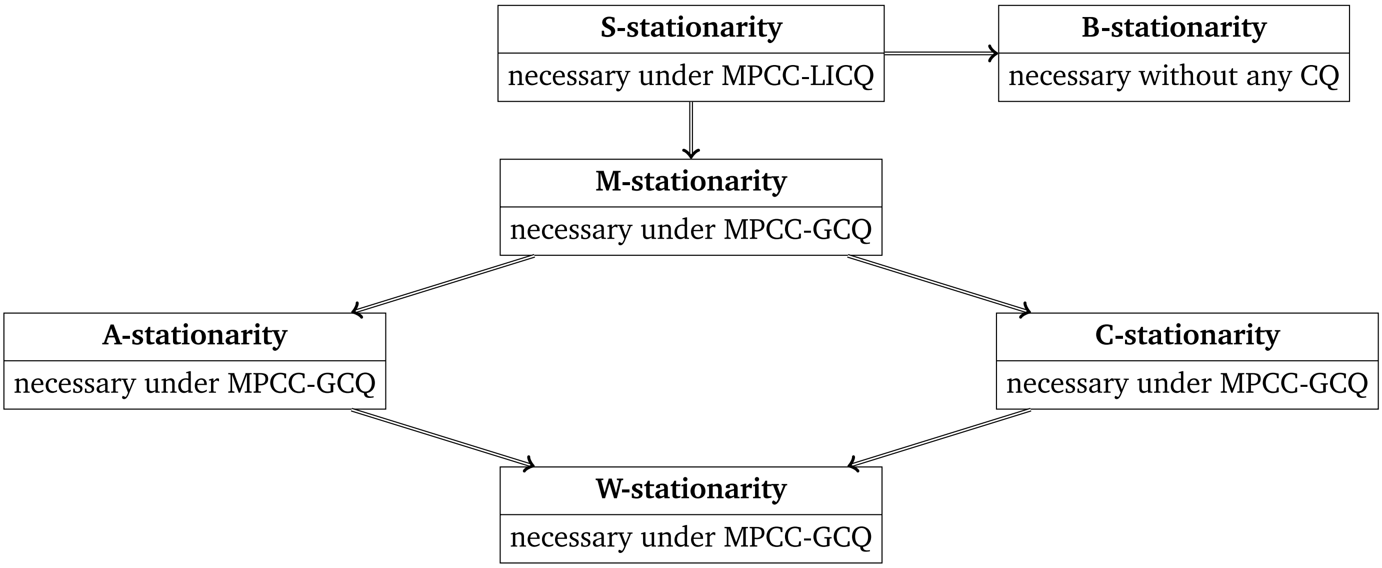


Figure 2.6: Relations between first order necessary conditions for MPCCs

Since we want to minimize the euclidean distance to the point $(1,1)$ but can only move inside the feasible set, this problem has two global minima $(1,0)$ and $(0,1)$ and a local maximum at $x^* = (0,0)$. Because MPCC-LICQ holds, we immediately know that the global minima are S-stationary.

Let us now consider the local maximum $x^* = 0$. There, the stationarity conditions

$$0 = \nabla f(x^*) - \gamma \nabla G(x^*) - \nu \nabla H(x^*) = \begin{pmatrix} -2 \\ -2 \end{pmatrix} - \gamma \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \nu \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

are satisfied for $\gamma = \nu = -\frac{1}{2}$ and thus the local maximum x^* is C-stationary but not M-stationary.

By now, we have seen a lot of different stationarity concepts. The relations between them are collected in Figure 2.6. All stationarity concepts differ only in the conditions on the multipliers γ_i, ν_i corresponding to a biactive complementarity constraint $i \in I_{00}(x^*)$. The corresponding “feasible sets” for these biactive multipliers are given in Figure 2.7.

2.4 Exercises

Exercise 2.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and $Z \subseteq \mathbb{R}^n$ nonempty. Prove that every local minimum $x^* \in Z$ of f on Z has the property

$$f'(x^*; d) = \nabla f(x^*)^T d \geq 0 \quad \forall d \in T_Z(x^*).$$

Since $T_Z(x^*)$ is the Bouligand tangent cone of Z at x^*, this condition is sometimes called Bouligand stationarity (B-stationarity).

Exercise 2.2. For $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ consider the feasible set

$$Z := \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}.$$

For a given point $x^* \in Z$ define the linearized feasible set as

$$Z^l := \{x \in \mathbb{R}^n \mid g(x^*) + \nabla g(x^*)^T (x - x^*) \leq 0, h(x^*) + \nabla h(x^*)^T (x - x^*) = 0\}$$

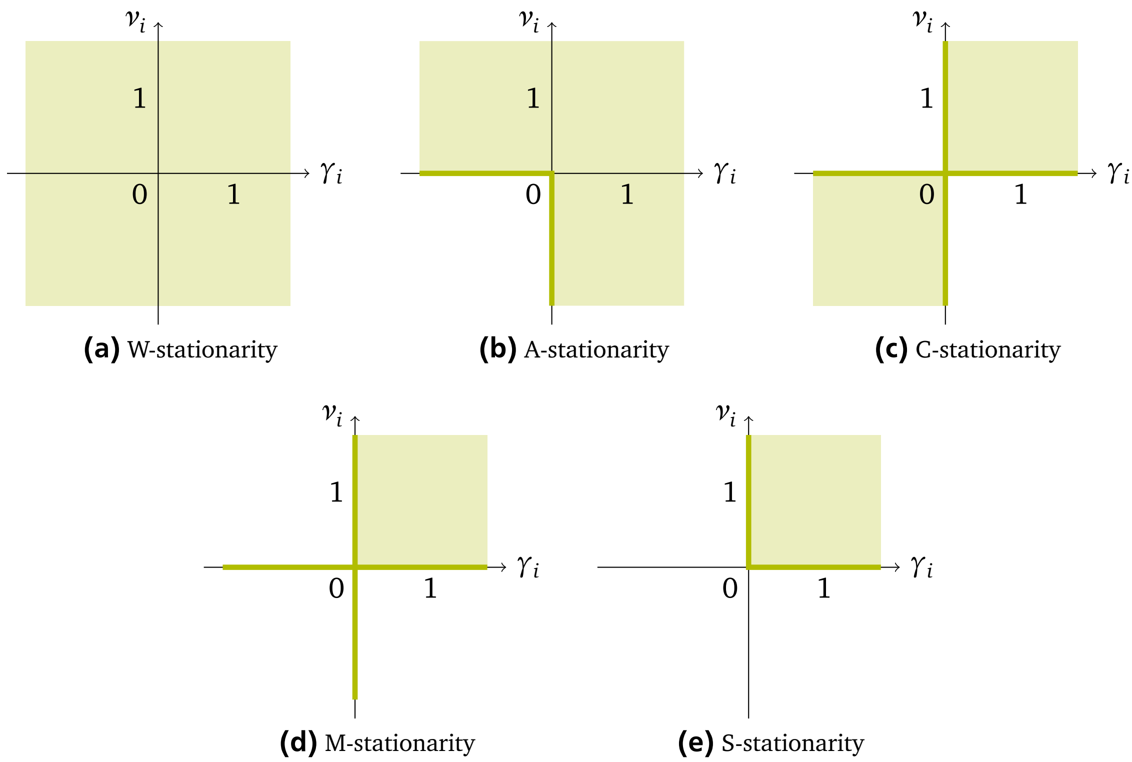


Figure 2.7: Sign constraints of the multipliers corresponding to $i \in I_{00}(x^*)$

(a) Prove
$$T_{Z^l}(x^*) = L_{Z^l}(x^*) = L_Z(x^*).$$

(b) Compare Z and Z^l for
$$Z = \{x \in \mathbb{R}^2 \mid x_1 \geq 0, \quad x_2 \geq 0, \quad x_1 x_2 = 0\}$$
and $x^* = (1,0)$ as well as $x^* = (0,0)$.

Exercise 2.3. Consider matrices $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times p}$ and $C \in \mathbb{R}^{m\times q}$. Then *Motzkin's transposition theorem* states that either the system

$$A^T y > 0, B^T y \geq 0, C^T y = 0$$

has a solution $y \in \mathbb{R}^n$ or the system

$$Ax_1 + Bx_2 + Cx_3 = 0,$$
$$x_1 \geq 0, x_1 \neq 0, x_2 \geq 0$$

has a solution $(x_1, x_2, x_3) \in \mathbb{R}^{m+p+q}$, but never both. Here, $A^T Y > 0$ means that all components of the vector $A^T y$ are positive.

(a) Prove the theorem.

(b) In the theorem the matrices B and C can be omitted but not A. Why is the theorem wrong if the matrix A is omitted?

Hint: The Farkas Lemma from linear or nonlinear optimization may be useful.

Exercise 2.4. Consider the standard NLP

$$\min_x f(x) \quad \text{s.t.} \quad g(x) \le 0, h(x) = 0$$

with $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ continuously differentiable and denote the feasible set by $Z \subseteq \mathbb{R}^n$. Then *positive linear independence CQ (PLICQ)* is said to hold at a feasible point $x^* \in Z$ if the gradients

$$\nabla g_i(x^*)(i \in I_g(x^*)) \text{ and } \nabla h_i(x^*)(i = 1, \ldots, p)$$

are positively linearly independent, i.e. if

$$\sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) = 0$$

and $\lambda \ge 0$ implies $(\lambda, \mu) = 0$.

Prove that PLICQ is equivalent to MFCQ.

Exercise 2.5. Consider the feasible set

$$X = \{x \in \mathbb{R}^n \mid g(x) \le 0, h(x) = 0, G(x) \ge 0, H(x) \ge 0, G(x) \circ H(x) = 0\}$$

for some continuously differentiable functions $g : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$, $G, H : \mathbb{R}^n \to \mathbb{R}^q$. (Here, $a \circ b$ denotes the componentwise product of two vectors.)

For an arbitrary feasible point $x^* \in X$ compute the linearized tangent cone $L_X(x^*)$ and the corresponding polar cone $L_X(x^*)^\circ$.

Hint: You can use that the polar cone of

$$K = \{d \mid A^T d \le 0, B^T d = 0\}$$

with some matrices A, B is given by

$$K^\circ = \{w = A\lambda + B\mu \mid \lambda \ge 0\}.$$

Exercise 2.6. Consider the feasible set

$$Z := \{x \in \mathbb{R}^n \mid g(x) \le 0,\ h(x) = 0,\ 0 \le G(x) \perp H(x) \ge 0\}$$

of our standard MPCC and for a given point $x^* \in X$ define the tightened feasible sets

$$\begin{aligned} Z_I := \{x \in \mathbb{R}^n \quad | \quad & g(x) \le 0,\ h(x) = 0, \\ & G_i(x) = 0, H_i(x) \ge 0 \quad \forall i \in I_{0+}(x^*) \cup I, \\ & G_i(x) \ge 0, H_i(x) = 0 \quad \forall i \in I_{+0}(x^*) \cup I^c\} \end{aligned}$$

for arbitrary subsets $I \subseteq I_{00}(x^*)$. Show the following:

(a) For all $I \subseteq I_{00}(x^*)$ we have $Z_I \subseteq Z$.

(b) There exists a radius $r > 0$ such that

$$Z \cap B_r(x^*) = \bigcup_{I \subseteq I_{00}(x^*)} Z_I \cap B_r(x^*).$$

Exercise 2.7. Consider the standard MPCC with the feasible set X. Show that for all $x^* \in X$ the following inclusion holds

$$T_X(x^*) \subseteq L_X^{\text{MPCC}}(x^*).$$

Exercise 2.8. Consider a feasible point $x^* \in X$ of MPCC, where MPCC-CPLD holds. Prove that this implies (standard) CPLD at x^* for the tightened problems TNLP(x^*, I)

$$\begin{aligned} \min_x f(x) \quad \text{s.t.} \quad & g(x) \le 0, h(x) = 0, \\ & G_i(x) \ge 0, H_i(x) = 0 \quad \forall i \in I_{+0}(x^*) \cup I, \\ & G_i(x) = 0, H_i(x) \ge 0 \quad \forall i \in I_{0+}(x^*) \cup I^c \end{aligned}$$

for all subsets $I \subseteq I_{00}(x^*)$.

Exercise 2.9. Consider a general feasible set of the form

$$Z = \{x \in \mathbb{R}^n \mid F(x) \in D\}$$

for some continuously differentiable map $F : \mathbb{R}^n \to \mathbb{R}^M$ and a nonempty set $D \subseteq \mathbb{R}^M$. Often (see below) the set D has a much simpler structure than the set Z. This motivates the definition of the *linearized tangent cone* as

$$L_Z^{\text{gen}}(x^*) := \{d \in \mathbb{R}^n \mid \nabla F(x^*)^T d \in T_D(F(x^*))\}.$$

(a) Write the feasible set of the standard NLP in the form $Z = \{x \in \mathbb{R}^n \mid F(x) \in D\}$ and show $L_Z(x^*) = L_Z^{\text{gen}}(x^*)$.

(b) Write the feasible set of the MPCC in the form $Z = \{x \in \mathbb{R}^n \mid F(x) \in D\}$ and show $L_X^{\text{MPCC}}(x^*) = L_Z^{\text{gen}}(x^*)$.

Exercise 2.10. Using normal cones, we can give shorter definitions of various stationarity conditions for MPCCs. To this end, rewrite the feasible set as $X = \{x \in \mathbb{R}^n \mid F(x) \in D\}$ as in Exercise 2.9.

(a) Show that $x^* \in X$ is S-stationary if and only if

$$0 \in \nabla f(x^*) + \nabla F(x^*)^T N_D^F(F(x^*)).$$

(b) Show that $x^* \in X$ is M-stationary if and only if

$$0 \in \nabla f(x^*) + \nabla F(x^*)^T N_D^M(F(x^*)).$$

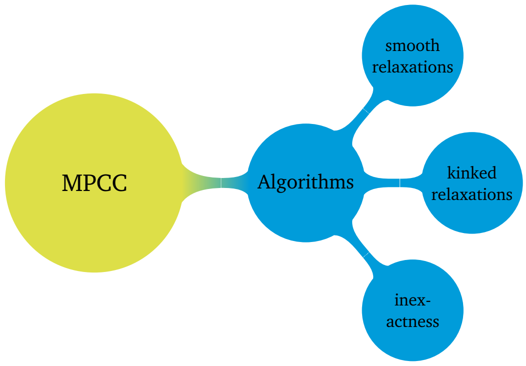
(c) The *Clarke normal cone* to a set Z at $x^* \in Z$ is defined as

$$N_Z^C(x^*) = \operatorname{conv} N_Z^M(x^*).$$

Which stationarity concept is equivalent to

$$0 \in \nabla f(x^*) + \nabla F(x^*)^T N_D^C(F(x^*))?$$

3 Algorithms



Here, we want to focus on one class of algorithms for MPCCs, the so called *relaxation algorithms*. The basic idea of all relaxation schemes is to get rid of the complicated complementarity constraints

$$G_i(x) \geq 0,\ H_i(x) \geq 0,\ G_i(x)H_i(x) = 0 \quad \forall i = 1, \ldots, q$$

by replacing these conditions in a suitable way such that the corresponding relaxed problem has nicer properties. The relaxed problem depends on a parameter $t > 0$ which has to be driven to zero in order to reobtain the underlying MPCC.

3.1 Smooth Relaxation Methods

Probably the first attempt to use a relaxation idea for solving MPCCs goes back to Scholtes [30]. The central idea of the relaxation scheme by Scholtes is to replace the MPCC by a sequence of parametrized NLPs of the form

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0, \quad \forall i = 1, \ldots, m, \\ & h_i(x) = 0 \quad \forall i = 1, \ldots, p, \\ & G_i(x) \geq 0 \quad \forall i = 1, \ldots, q, \\ & H_i(x) \geq 0 \quad \forall i = 1, \ldots, q, \\ & G_i(x)H_i(x) \leq t \quad \forall i = 1, \ldots, q. \end{array}$$

see Figure 3.1 for a geometric illustration.

We denote the relaxed problem by $R^S(t)$ and its feasible set by $X^S(t)$. Since, geometrically, this is a global relaxation of the complementarity conditions, we call this approach the *global relaxation method*.

For the convergence analysis, some index sets are needed:

$$\begin{aligned} I_g(x) &:= \{i \mid g_i(x) = 0\}, \\ I_G(x) &:= \{i \mid G_i(x) = 0\}, \\ I_H(x) &:= \{i \mid H_i(x) = 0\}, \\ I_{GH}(x;t) &:= \{i \mid H_i(x)G_i(x) = t\}. \end{aligned}$$

The following is the most basic convergence result for the global relaxation method.

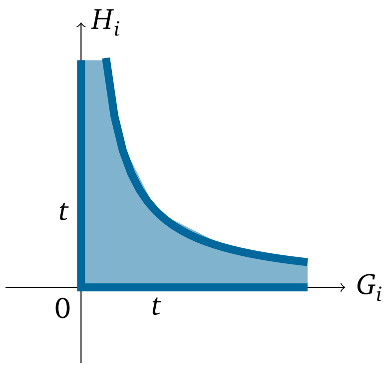


Figure 3.1: Geometric interpretation of the relaxation method by Scholtes

Theorem 3.1. *Let $(t_k)_k \downarrow 0$ and let x^k be a KKT point of $R^S(t_k)$ with $x^k \to x^*$ such that MPCC-MFCQ holds at x^*. Then x^* is a C-stationary point of* (1.1).

Proof. Since x^k is a KKT point of $R^S(t_k)$ there exist multipliers $(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^k)$ such that

$$\begin{aligned} 0 \;=\; & \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) - \sum_{i=1}^{q} \gamma_i^k \nabla G_i(x^k) \\ & - \sum_{i=1}^{q} \nu_i^k \nabla H_i(x^k) + \sum_{i=1}^{q} \delta_i^k [H_i(x^k)\nabla G_i(x^k) + G_i(x^k)\nabla H_i(x^k)] \end{aligned} \tag{3.1}$$

with

$$\begin{aligned} \lambda^k \geq 0 \quad & \text{and} \quad \operatorname{supp}(\lambda^k) \subseteq I_g(x^k), \\ \gamma^k \geq 0 \quad & \text{and} \quad \operatorname{supp}(\gamma^k) \subseteq I_G(x^k), \\ \nu^k \geq 0 \quad & \text{and} \quad \operatorname{supp}(\nu^k) \subseteq I_H(x^k), \\ \delta^k \geq 0 \quad & \text{and} \quad \operatorname{supp}(\delta^k) \subseteq I_{GH}(x^k; t_k) \end{aligned}$$

for all $k \in \mathbb{N}$. This implies

$$\operatorname{supp}(\gamma^k) \cap \operatorname{supp}(\delta^k) = \emptyset, \quad \operatorname{supp}(\nu^k) \cap \operatorname{supp}(\delta^k) = \emptyset \tag{3.2}$$

for all $k \in \mathbb{N}$. Moreover, for all $k \in \mathbb{N}$ sufficiently large, we have $I_g(x^k) \subseteq I_g(x^*)$, $I_G(x^k) \subseteq I_{00}(x^*) \cup I_{0+}(x^*)$, and $I_H(x^k) \subseteq I_{00}(x^*) \cup I_{+0}(x^*)$.

Our next step is to define suitable new multipliers

$$\tilde{\gamma}_i^k = \begin{cases} \gamma_i^k, & \text{if } i \in \operatorname{supp}(\gamma^k), \\ -\delta_i^k H_i(x^k), & \text{if } i \in \operatorname{supp}(\delta^k) \setminus I_{+0}(x^*), \\ 0, & \text{else,} \end{cases}$$

and

$$\tilde{\nu}_i^k = \begin{cases} \nu_i^k, & \text{if } i \in \operatorname{supp}(\nu^k), \\ -\delta_i^k G_i(x^k), & \text{if } i \in \operatorname{supp}(\delta^k) \setminus I_{0+}(x^*), \\ 0, & \text{else.} \end{cases}$$

With these multipliers, we can rewrite (3.1) as

$$
\begin{aligned}
0 \;=\; & \nabla f(x^k)+\sum_{i=1}^{m}\lambda_i^k\nabla g_i(x^k)+\sum_{i=1}^{p}\mu_i^k\nabla h_i(x^k)-\sum_{i=1}^{q}\tilde{\gamma}_i^k\nabla G_i(x^k)-\sum_{i=1}^{q}\tilde{\nu}_i^k\nabla H_i(x^k)\\
& +\sum_{i\in I_{+0}(x^*)}\delta_i^k H_i(x^k)\nabla G_i(x^k)+\sum_{i\in I_{0+}(x^*)}\delta_i^k G_i(x^k)\nabla H_i(x^k).
\end{aligned}
$$

If we assume that the sequence $\{(\lambda^k,\mu^k,\tilde{\gamma}^k,\tilde{\nu}^k,\delta^k_{I_{+0}(x^*)\cup I_{0+}(x^*)})\}$ was unbounded, then we can find a subsequence K such that the normed sequence converges:

$$
\frac{(\lambda^k,\mu^k,\tilde{\gamma}^k,\tilde{\nu}^k,\delta^k_{I_{+0}\cup I_{0+}})}{\|(\lambda^k,\mu^k,\tilde{\gamma}^k,\tilde{\nu}^k,\delta^k_{I_{+0}\cup I_{0+}})\|}\to_K(\lambda,\mu,\tilde{\gamma},\tilde{\nu},\delta_{I_{+0}\cup I_{0+}})\neq 0.
$$

The equation above then yields

$$
0 \;=\; \sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)+\sum_{i=1}^{p}\mu_i\nabla h_i(x^*)-\sum_{i=1}^{q}\tilde{\gamma}_i\nabla G_i(x^*)-\sum_{i=1}^{q}\tilde{\nu}_i\nabla H_i(x^*)
$$

where $\lambda\geq 0$ and for all $k\in K$ sufficiently large

$$
\begin{aligned}
\operatorname{supp}(\lambda) &\subseteq I_g(x^k)\subseteq I_g(x^*),\\
\operatorname{supp}(\tilde{\gamma}) &\subseteq I_G(x^k)\cup I_{GH}(x^k;t_k)\setminus I_{+0}(x^*)\subseteq I_{00}(x^*)\cup I_{0+}(x^*),\\
\operatorname{supp}(\tilde{\nu}) &\subseteq I_H(x^k)\cup I_{GH}(x^k;t_k)\setminus I_{0+}(x^*)\subseteq I_{00}(x^*)\cup I_{+0}(x^*).
\end{aligned}
$$

Additionally, $(\lambda,\mu,\tilde{\gamma},\tilde{\nu})\neq 0$ has to hold. Otherwise, $\delta_i>0$ would have to hold for at least one $i\in I_{+0}(x^*)\cup I_{0+}(x^*)$. Assume without loss of generality $\delta_i>0$ for an $i\in I_{+0}(x^*)$. This implies $\delta_i^k>0$ for all k sufficiently large and consequently $\tilde{\nu}_i^k=-\delta_i^k G_i(x^k)$ for those k. Because of $i\in I_{+0}(x^*)$, this yields $\tilde{\nu}_i=\lim_{k\in K}-\delta_i^k G_i(x^k)<0$, a contradiction to our assumption $\tilde{\nu}=0$.

However, $(\lambda,\mu,\tilde{\gamma},\tilde{\nu})\neq 0$ is a contradiction to the prerequisite that MPCC-MFCQ holds in x^*. Thus, we may assume without loss of generality that the sequence is convergent to some vector $(\lambda^*,\mu^*,\tilde{\gamma}^*,\tilde{\nu}^*,\delta^*_{I_{+0}(x^*)\cup I_{0+}(x^*)})$. It is easy to see that $\lambda^*\geq 0$ and $\operatorname{supp}(\lambda^*)\subseteq I_g(x^*)$. According to the definition of $\tilde{\gamma}^k$ and $\tilde{\nu}^k$, we have

$$
\operatorname{supp}(\tilde{\gamma}^*)\subseteq I_{00}(x^*)\cup I_{0+}(x^*),\quad \operatorname{supp}(\tilde{\nu}^*)\subseteq I_{00}(x^*)\cup I_{+0}(x^*).
$$

The continuous differentiability of f,g,h,G,H then implies

$$
0=\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)+\sum_{i=1}^{p}\mu_i^*\nabla h_i(x^*)-\sum_{i=1}^{q}\tilde{\gamma}_i^*\nabla G_i(x^*)-\sum_{i=1}^{q}\tilde{\nu}_i^*\nabla H_i(x^*).
$$

To prove the C-stationarity of x^*, it remains to show that $\tilde{\gamma}_i^*\tilde{\nu}_i^*\geq 0$ for all $i\in I_{00}(x^*)$. Assume that there is an $i\in I_{00}(x^*)$ with $\tilde{\gamma}_i^*<0$ and $\tilde{\nu}_i^*>0$ or with $\nu_i^*<0$ and $\tilde{\gamma}_i^*>0$. We consider only the first case, the second one can be treated similarly. Because of $\tilde{\gamma}_i^k\geq 0$, the condition $\gamma_i^*<0$ implies $i\in\operatorname{supp}(\delta^k)$ for all $k\in\mathbb{N}$ sufficiently large. This implies $i\notin\operatorname{supp}(\nu^k)$ in view of (3.2) and, therefore, $\tilde{\nu}_i^*\leq 0$ in contradiction to our assumption. □

Note that the corresponding result in [30] assumes MPCC-LICQ and shows that the whole sequence of multipliers corresponding to the KKT points x^k converges. Here we assume the weaker MPCC-MFCQ which, obviously, does not guarantee convergence of the corresponding sequence of multipliers. But the proof shows that one can extract a sequence of multipliers which stays bounded and is, therefore, convergent at least on a subsequence.

To see that this convergence results is sharp, i.e. without further assumptions one cannot expect more than C-stationarity of the limit, consider the following example.

Example 3.2. Consider the MPCC

$$\min_{x\in\mathbb{R}^2}(x_1-1)^2+(x_2-1)^2 \quad \text{s.t.} \quad 0\le x_1 \perp x_2 \ge 0.$$

Then the global minima are $(1,0)$ and $(0,1)$ and are S-stationary due to MPCC-LICQ. The point $x^*=(0,0)$ is a local maximum, but is C-stationary. And for $t>0$ sufficiently small, the point $(\sqrt{t},\sqrt{t})$ are KKT points of the relaxed problem.

The assumption of x^k being a KKT point of the relaxed problem $R^S(t_k)$ is based on the existence of multipliers. A priori, it is not clear that these multipliers really exist. The following result essentially guarantees the existence of these multipliers by showing that MPCC-MFCQ at a feasible point x^* of the original MPCC implies that standard MFCQ holds for the relaxed problems $R^S(t)$, at least locally around x^*.

Theorem 3.3. *Let x^* be feasible for* (1.1) *such that MPCC-MFCQ is satisfied at x^*. Then there exists a neighborhood $U(x^*)$ of x^* and $\bar{t}>0$ such that the following holds for all $t\in(0,\bar{t}]$: If $x\in U(x^*)$ is feasible for $R^S(t)$, then standard MFCQ for $R^S(t)$ holds in x.*

Proof. First note that, by continuity, for all $x\in X^S(t)$ sufficiently close to x^*, we have

$$\begin{array}{rcl} I_g(x) & \subseteq & I_g(x^*), \\ I_G(x) & \subseteq & I_{00}(x^*)\cup I_{0+}(x^*), \\ I_H(x) & \subseteq & I_{00}(x^*)\cup I_{+0}(x^*), \\ I_{GH}(x)\cap I_G(x) & = & \emptyset, \\ I_{GH}(x)\cap I_H(x) & = & \emptyset. \end{array} \tag{3.3}$$

Since MPCC-MFCQ holds, the gradients

$$\begin{array}{l}\{\nabla g_i(x^*)\mid i\in I_g(x^*)\}\cup \\ \big\{\{\nabla h_i(x^*)\mid i=1,\dots,p\}\cup\{\nabla G_i(x^*)\mid i\in I_{00}(x^*)\cup I_{0+}(x^*)\}\cup\{\nabla H_i(x^*)\mid i\in I_{00}(x^*)\cup I_{+0}(x^*)\}\big\}\end{array}$$

are positive-linearly independent. This implies that the set of gradients

$$\begin{array}{l}\{\nabla g_i(x)\mid i\in I_g(x^*)\}\cup \\ \big\{\{\nabla h_i(x)\mid i=1,\dots,p\}\cup\{\nabla G_i(x)\mid i\in I_{00}(x^*)\cup I_{0+}(x^*)\}\cup\{\nabla H_i(x)\mid i\in I_{00}(x^*)\cup I_{+0}(x^*)\}\big\}\end{array}$$

is also positive-linearly independent for all $x\in X^S(t)$ sufficiently close to x^*. Taking into account that

$$I_G(x)\cup\big(I_{GH}(x)\cap I_{0+}(x^*)\big)\cup\big(I_{GH}(x)\cap I_{00}(x^*)\big)\subseteq I_{00}(x^*)\cup I_{0+}(x^*)$$

and

$$I_H(x)\cup\big(I_{GH}(x)\cap I_{+0}(x^*)\big)\cup\big(I_{GH}(x)\cap I_{00}(x^*)\big)\subseteq I_{00}(x^*)\cup I_{+0}(x^*)$$

for all $x\in X^S(t)$ sufficiently close to x^* and using the fact that $G_i(x)>0, H_i(x)\approx 0$ for all $i\in I_{+0}(x^*)$ as well as $G_i(x)\approx 0, H_i(x)>0$ for all $i\in I_{0+}(x^*)$ whenever x is close to x^*, it follows that there is a neighborhood $U(x^*)$ such that the set of vectors

$$\begin{array}{rl} \nabla g_i(x) & (i\in I_g(x)), \\ \nabla h_i(x) & (i=1,\dots,p), \\ \nabla G_i(x) & (i\in I_G(x)), \\ \nabla H_i(x) & (i\in I_H(x)), \\ G_i(x)\nabla H_i(x)+H_i(x)\nabla G_i(x) & (i\in I_{GH}(x)\cap I_{0+}(x^*)), \\ G_i(x)\nabla H_i(x)+H_i(x)\nabla G_i(x) & (i\in I_{GH}(x)\cap I_{+0}(x^*)), \\ \nabla G_i(x) & (i\in I_{GH}(x)\cap I_{00}(x^*)), \\ \nabla H_i(x) & (i\in I_{GH}(x)\cap I_{00}(x^*)) \end{array} \tag{3.4}$$

is positive-linearly independent for all $x \in X^S(t) \cap U(x^*)$.

We now claim that standard MFCQ holds for the relaxed program $R^S(t)$ whenever $x \in X^S(t) \cap U(x^*)$. To this end, take an arbitrary $x \in X^S(t) \cap U(x^*)$. We have to show that

$$\begin{aligned} 0 &= \sum_{i \in I_g(x)} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \mu_i \nabla h_i(x) - \sum_{i \in I_G(x)} \alpha_i \nabla G_i(x) \\ &\quad - \sum_{i \in I_H(x)} \beta_i \nabla H_i(x) + \sum_{i \in I_{GH}(x)} \gamma_i (G_i(x) \nabla H_i(x) + H_i(x) \nabla G_i(x)) \end{aligned} \tag{3.5}$$

with $\mu \in \mathbb{R}^p$ and $\lambda, \alpha, \beta, \gamma \geq 0$ holds only for the null vector. To see this, we rewrite (3.5) as

$$\begin{aligned} 0 &= \sum_{i \in I_g(x)} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \mu_i \nabla h_i(x) - \sum_{i \in I_G(x)} \alpha_i \nabla G_i(x) - \sum_{i \in I_H(x)} \beta_i \nabla H_i(x) \\ &\quad + \sum_{i \in I_{GH}(x) \cap (I_{0+}(x^*) \cup I_{+0}(x^*))} \gamma_i (G_i(x) \nabla H_i(x) + H_i(x) \nabla G_i(x)) \\ &\quad + \sum_{i \in I_{00}(x^*) \cap I_{GH}(x)} (\gamma_i G_i(x)) \nabla H_i(x) + \sum_{i \in I_{00}(x^*) \cap I_{GH}(x)} (\gamma_i H_i(x)) \nabla G_i(x). \end{aligned} \tag{3.6}$$

Applying the positive-linear independence of the vectors from (3.4) to (3.6) and using (3.3), we immediately obtain that all coefficients from (3.5) are zero, and this completes the proof. □

An alternative but closely related smooth relaxation can be found in the paper [21] by Lin and Fukushima. They suggest to relax the MPCC to

$$\begin{aligned} \min_x f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & G_i(x) H_i(x) - t^2 \leq q0 \quad \forall i = 1, \ldots, q, \\ & (G_i(x) + t)(H_i(x) + t) - t^2 \geq 0 \quad \forall i = 1, \ldots, q, \end{aligned}$$

see Figure 3.2 for an illustration.

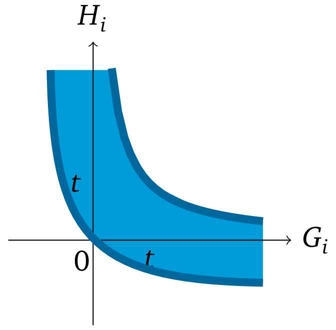


Figure 3.2: Geometric interpretation of the relaxation method by Lin and Fukushima

The relaxation method by Lin and Fukushima has exactly the same convergence properties ans the one by Scholtes, which we discussed above.

Another, only local relaxation goes back to the paper [31] by Steffensen and Ulbrich. Here, the idea is to relax the feasible set of MPCC only around the origin, i.e. around the critical biactive points. To do so, one uses regularization functions, which can be seen as a smoothing of the absolute value function around the origin.

Definition 3.4. $\theta : [-1, 1] \to \mathbb{R}$ *is called a* regularization function *if it satisfies the following conditions:*

(a) θ is twice continuously differentiable on $[-1,1]$;

(b) $\theta(-1) = \theta(1) = 1$;

(c) $\theta'(-1) = -1$ and $\theta'(1) = 1$;

(d) $\theta''(-1) = \theta''(1) = 0$;

(e) $\theta''(x) > 0$ for all $x \in (-1,1)$.

Note that condition (e) implies that θ is strictly convex on $[-1,1]$. The following result taken from [31, Lemma 3.1] reveals an immediate but crucial property of all regularization functions.

Lemma 3.5. *Let $\theta : [-1,1] \to \mathbb{R}$ be a regularization function. Then it holds that $\theta(x) > |x|$ for all $x \in (-1,1)$.*

Two simple examples of suitable regularization functions are

$$\theta(x) := \frac{2}{\pi}\sin\left(\frac{\pi}{2}x + \frac{3\pi}{2}\right) + 1 \quad \text{and} \quad \theta(x) := \frac{1}{8}\left(-x^4 + 6x^2 + 3\right),$$

cf. [31]. The second function is the Hermite interpolation polynomial satisfying the requirements from Definition 3.4.

This idea leads to the following relaxation from [31] (cf. Figure 3.3 for an illustration)

$$\begin{array}{ll}
\min & f(x) \\
\text{s.t.} & g_i(x) \le 0, \quad \forall i = 1, \ldots, m, \\
& h_j(x) = 0 \quad \forall j = 1, \ldots, p, \\
& G_i(x) \ge 0 \quad \forall i = 1, \ldots, q, \\
& H_i(x) \ge 0 \quad \forall i = 1, \ldots, q, \\
& \Phi^{SU}(x;t) \le 0 \quad \forall i = 1, \ldots, q
\end{array}$$

with

$$\Phi^{SU} : \mathbb{R}^n \to \mathbb{R}^q, \quad \Phi_i^{SU}(x;t) := G_i(x) + H_i(x) - \varphi(G_i(x) - H_i(x); t) \; \forall i = 1, \ldots, q$$

where

$$\varphi(\cdot;t) : \mathbb{R} \to \mathbb{R}, \quad \varphi(a;t) := \begin{cases} |a|, & \text{if } |a| \ge t, \\ t\theta(\frac{a}{t}), & \text{if } |a| < t, \end{cases} \tag{3.7}$$

and θ is a regularization function.

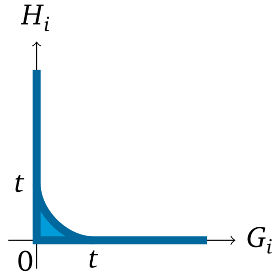


Figure 3.3: Geometric interpretation of the relaxation method by Steffensen and Ulbrich

This relaxation method still converges only to C-stationary points, but it suffices to assume MPCC-CPLD as opposed to MPCC-MFCQ.

3.2 Kinked Relaxation Methods

If one wants to ensure convergence of a relaxation method to more than C-stationary points, one either has to impose additional assumptions on the sequence of KKT points, use second order information or change the geometry of the feasible set of the relaxation.

Probably the first relaxation scheme, that can be guaranteed to converge to M-stationary points, can be found [18] by Kadrani et al. They propose the following relaxation, see also Figure 3.4:

$$\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \leq 0, \quad \forall i = 1, \ldots, m, \\
& h_j(x) = 0 \quad \forall j = 1, \ldots, p, \\
& G_i(x) \geq -t \quad \forall i = 1, \ldots, q, \\
& H_i(x) \geq -t \quad \forall i = 1, \ldots, q, \\
& (G_i(x) - t)(H_i(x) - t) \leq 0 \quad \forall i = 1, \ldots, q.
\end{aligned}$$

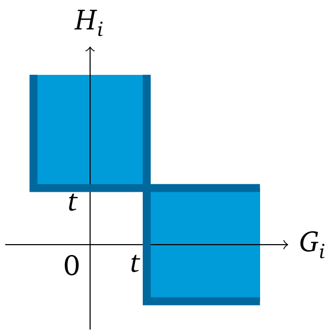


Figure 3.4: Geometric interpretation of the relaxation method by Kadrani et al.

To be precise, this method is not really a relaxation method, as the feasible sets of the relaxed problems do not include the whole feasible set of MPCC. But one can show that this method converges to M-stationary points under CC-CPLD. Furthermore, if MPCC-LICQ holds at $x^* \in X$, then standard Guignard CQ holds in all feasible points of the relaxed problem in a neighborhood. In fact, standard LICQ holds in all feasible points of the relaxed problem in a neighborhood except for those “on the kink”.

To overcome the problem that the feasible set of the MPCC is not completely included in the feasible set of the relaxed problems and that the feasible set of the relaxed problems is almost disconnected, Kanzow and Schwartz suggested another relaxation scheme in [19].

Our relaxation is based on the function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\varphi(a,b) = \begin{cases} ab, & \text{if } a+b \geq 0, \\ -\frac{1}{2}(a^2+b^2), & \text{if } a+b < 0. \end{cases}$$

This function has the following elementary properties.

Lemma 3.6. *(a) φ is an NCP-function, i.e. $\varphi(a,b) = 0$ if and only if $a \geq 0, b \geq 0, ab = 0$.*

(b) φ is continuously differentiable with gradient

$$\nabla \varphi(a,b) = \begin{cases} \begin{pmatrix} b \\ a \end{pmatrix}, & \textit{if } a+b \geq 0, \\ \begin{pmatrix} -a \\ -b \end{pmatrix}, & \textit{if } a+b < 0. \end{cases}$$

(c) φ has the property that

$$\varphi(a,b)\begin{cases} >0, & \text{if } a>0 \text{ and } b>0,\\ <0, & \text{if } a<0 \text{ or } b<0.\end{cases}$$

Based on this function, we define a continuously differentiable mapping $\Phi:\mathbb{R}^n\to\mathbb{R}^q$ componentwise by

$$\begin{aligned}\Phi_i(x;t) &:= \varphi\big(G_i(x)-t, H_i(x)-t\big)\\ &= \begin{cases}\big(G_i(x)-t\big)\big(H_i(x)-t\big), & \text{if } G_i(x)+H_i(x)\ge 2t,\\ -\frac{1}{2}\big((G_i(x)-t)^2+(H_i(x)-t)^2\big), & \text{if } G_i(x)+H_i(x)<2t,\end{cases}\end{aligned}$$

where $t\ge 0$ is an arbitrary parameter. With this function, we can formulate the *relaxed problem* $R^{KS}(t)$ for $t\ge 0$ as

$$\begin{aligned}\min f(x) \quad \text{s.t.} \quad & g_i(x)\le 0 \quad \forall i=1,\dots,m,\\ & h_i(x)=0 \quad \forall i=1,\dots,p,\\ & G_i(x)\ge 0 \quad \forall i=1,\dots,q,\\ & H_i(x)\ge 0 \quad \forall i=1,\dots,q,\\ & \Phi_i(x;t)\le 0 \quad \forall i=1,\dots,q.\end{aligned} \tag{3.8}$$

Hence, in our approach, we replace the complementarity conditions

$$G_i(x)\ge 0,\ H_i(x)\ge 0,\ G_i(x)H_i(x)=0 \quad \forall i=1,\dots,q$$

by the conditions

$$G_i(x)\ge 0,\ H_i(x)\ge 0,\ \Phi_i(x;t)\le 0 \quad \forall i=1,\dots,q$$

which, from a geometric point of view, gives a set of the form shown in Figure 3.5.

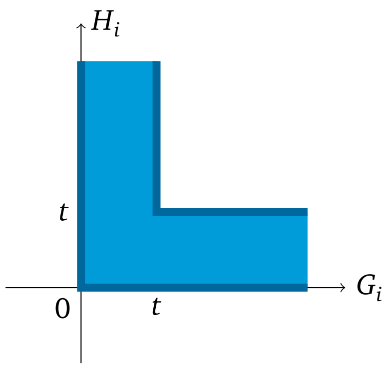


Figure 3.5: Geometric interpretation of the new regularization

Similar to the index sets used for MPCCs before, we define

$$\begin{aligned} I_g(x) &:= \{i \mid g_i(x)=0\},\\ I_G(x) &:= \{i \mid G_i(x)=0\},\\ I_H(x) &:= \{i \mid H_i(x)=0\},\\ I_\Phi(x;t) &:= \{i \mid \Phi_i(x;t)=0\}\end{aligned}$$

for $t \geq 0$ and x feasible for $R^{KS}(t)$. We also use a partition of the index set $I_\Phi(x;t)$ into the following three subsets:

$$\begin{aligned} I_\Phi^{00}(x;t) &:= \{i \in I_\Phi(x;t) \mid G_i(x) - t = 0, H_i(x) - t = 0\}, \\ I_\Phi^{0+}(x;t) &:= \{i \in I_\Phi(x;t) \mid G_i(x) - t = 0, H_i(x) - t > 0\}, \\ I_\Phi^{+0}(x;t) &:= \{i \in I_\Phi(x;t) \mid G_i(x) - t > 0, H_i(x) - t = 0\}. \end{aligned}$$

Note that these sets form a partition of $I_\Phi(x;t)$ since the definition of Φ implies that

$$\Phi_i(x;t) = 0 \quad \Longleftrightarrow \quad G_i(x) - t \geq 0,\ H_i(x) - t \geq 0,\ \big(G_i(x) - t\big)\big(H_i(x) - t\big) = 0.$$

In view of Lemma 3.6, the function Φ is continuously differentiable with its gradient given by

$$\nabla \Phi_i(x;t) = \begin{cases} (H_i(x) - t)\nabla G_i(x) + (G_i(x) - t)\nabla H_i(x), & \text{if } G_i(x) + H_i(x) \geq 2t, \\ -(G_i(x) - t)\nabla G_i(x) - (H_i(x) - t)\nabla H_i(x), & \text{if } G_i(x) + H_i(x) < 2t \end{cases} \tag{3.9}$$

for all $i = 1, \ldots, q$.

The following result summarizes some simple properties of the regularized program $R^{KS}(t)$.

Lemma 3.7. *For $t > 0$ let X and $X^{KS}(t)$ be the feasible sets of the MPCC* (1.1) *and $R^{KS}(t)$, respectively. Then the following three statements hold:*

(a) $X^{KS}(0) = X$.

(b) $X^{KS}(t_1) \subseteq X^{KS}(t_2)$ *for all* $0 \leq t_1 \leq t_2$.

(c) $\bigcap_{t \geq 0} X^{KS}(t) = X$.

The previous result shows, in particular, that the feasible set X of the original MPCC is always contained in the feasible set $X^{KS}(t)$ of the regularized program $R^{KS}(t)$ (in contrast to the approach by Kadrani et al., and that our relaxation exhibits the desired behavior $\lim_{t \downarrow 0} X^{KS}(t) = X$. Note also that, from a geometric point of view, our regularized problem has a much nicer feasible set than the one by Kadrani et al. which, we recall, consists of almost disconnected pieces.

If we solve $R^{KS}(t_k)$ for a sequence $\{t_k\} \downarrow 0$ and obtain KKT points $(x^k, \lambda^k, \mu^k, \gamma^k, \nu^k, \delta^k)$ of $R^{KS}(t_k)$, where $x^k \to x^*$, what kind of MPCC-stationarity can we expect in x^*? The next theorem gives an answer to this question.

Theorem 3.8. *Let $\{t_k\} \downarrow 0$ and $\{(x^k, \lambda^k, \mu^k, \gamma^k, \nu^k, \delta^k)\}$ be a sequence of KKT points of $R^{KS}(t_k)$ with $x^k \to x^*$. If MPCC-CPLD holds in x^*, then x^* is an M-stationary point of the MPCC* (1.1).

Proof. Obviously, x^* is feasible for the MPCC (1.1) and for all $k \in \mathbb{N}$ sufficiently large, we have

$$\begin{aligned} I_g(x^k) &\subseteq I_g(x^*), \\ I_G(x^k) \cup I_\Phi^{00}(x^k;t_k) \cup I_\Phi^{0+}(x^k;t_k) &\subseteq I_{00}(x^*) \cup I_{0+}(x^*), \\ I_H(x^k) \cup I_\Phi^{00}(x^k;t_k) \cup I_\Phi^{+0}(x^k;t_k) &\subseteq I_{00}(x^*) \cup I_{+0}(x^*). \end{aligned} \tag{3.10}$$

Since all $(x^k, \lambda^k, \mu^k, \gamma^k, \nu^k, \delta^k)$ are KKT points of $R^{KS}(t_k)$, we have

$$\begin{aligned} 0 &= \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) - \sum_{i=1}^q \gamma_i^k \nabla G_i(x^k) - \sum_{i=1}^q \nu_i^k \nabla H_i(x^k) \\ &\quad + \sum_{i=1}^q \delta_i^k \nabla \Phi_i(x^k;t_k) \end{aligned}$$

with

$$\begin{aligned}
\lambda_i^k = 0 \ \forall i \notin I_g(x^k) \quad &\text{and} \quad \lambda_i^k \geq 0 \ \forall i \in I_g(x^k),\\
\gamma_i^k = 0 \ \forall i \notin I_G(x^k) \quad &\text{and} \quad \gamma_i^k \geq 0 \ \forall i \in I_G(x^k),\\
\nu_i^k = 0 \ \forall i \notin I_H(x^k) \quad &\text{and} \quad \nu_i^k \geq 0 \ \forall i \in I_H(x^k),\\
\delta_i^k = 0 \ \forall i \notin I_\Phi(x^k; t) \quad &\text{and} \quad \delta_i^k \geq 0 \ \forall i \in I_\Phi(x^k; t_k).
\end{aligned}$$

Since the representation of $\nabla\Phi_i$ immediately gives $\nabla\Phi_i(x^k; t_k) = 0$ for all $i \in I_\Phi^{00}(x^k; t_k)$ and all $k \in \mathbb{N}$, we may also assume $\delta_i^k = 0$ for all $i \in I_\Phi^{00}(x^k; t_k)$ and all $k \in \mathbb{N}$. Thus, we can rewrite the equation above as

$$\begin{aligned}
0 \;=\; & \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) - \sum_{i=1}^q \gamma_i^k \nabla G_i(x^k) - \sum_{i=1}^q \nu_i^k \nabla H_i(x^k)\\
& + \sum_{i=1}^q \delta_i^{G,k} \nabla G_i(x^k) + \sum_{i=1}^q \delta_i^{H,k} \nabla H_i(x^k)
\end{aligned}$$

where

$$\begin{aligned}
\delta_i^{G,k} &= \begin{cases} \delta_i^k (H_i(x^k) - t_k), & \text{if } i \in I_\Phi^{0+}(x^k; t_k),\\ 0, & \text{else,} \end{cases}\\
\delta_i^{H,k} &= \begin{cases} \delta_i^k (G_i(x^k) - t_k), & \text{if } i \in I_\Phi^{+0}(x^k; t_k),\\ 0, & \text{else.} \end{cases}
\end{aligned}$$

Note that the multipliers $\delta^{G,k}$ and $\delta^{H,k}$ are nonnegative, too. Here, we may assume without loss of generality that the gradients corresponding to nonvanishing multipliers in this equation are linearly independent for all $k \in \mathbb{N}$ (note that this may change the multipliers, but a previously positive multiplier will stay at least nonnegative and a vanishing multiplier will remain zero).

Our next step is to prove that the sequence $\{(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^{G,k}, \delta^{H,k})\}$ is bounded. Assuming the contrary, we can find a subsequence K such that

$$\frac{(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^{G,k}, \delta^{H,k})}{\|(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^{G,k}, \delta^{H,k})\|} \to_K (\lambda, \mu, \gamma, \nu, \delta^G, \delta^H) \neq 0.$$

Dividing by $\|(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^{G,k}, \delta^{H,k})\|$ and passing to the limit in the equation above yields

$$\begin{aligned}
0 \;=\; & \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) - \sum_{i=1}^q \gamma_i \nabla G_i(x^*) - \sum_{i=1}^q \nu_i \nabla H_i(x^*)\\
& + \sum_{i=1}^q \delta_i^G \nabla G_i(x^*) + \sum_{i=1}^q \delta_i^H \nabla H_i(x^*),
\end{aligned}$$

i.e., the gradients

$$\begin{aligned}
&\{\nabla g_i(x^*) \mid i \in \operatorname{supp}(\lambda)\} \cup \big\{\{\nabla h_i(x^*) \mid i \in \operatorname{supp}(\mu)\}\\
&\cup \{\nabla G_i(x^*) \mid i \in \operatorname{supp}(\gamma) \cup \operatorname{supp}(\delta^G)\} \cup \{\nabla H_i(x^*) \mid i \in \operatorname{supp}(\nu) \cup \operatorname{supp}(\delta^H)\}\big\}
\end{aligned} \tag{3.11}$$

are positive-linearly dependent. MPCC-CPLD guarantees that they remain linearly dependent in a whole neighborhood. This, however, is a contradiction to the linear independence of these gradients in x^k. Here, we use

$$\operatorname{supp}(\lambda, \mu, \gamma, \nu, \delta^G, \delta^H) \subseteq \operatorname{supp}(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^{G,k}, \delta^{H,k})$$

for all k sufficiently large and (3.10).

Consequently, our assumption was wrong and thus the sequence $\{(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^{G,k}, \delta^{H,k})\}$ is bounded. Therefore, we can assume without loss of generality that the whole sequence is convergent to some limit $(\lambda^*, \mu^*, \gamma^*, \nu^*, \delta^{G,*}, \delta^{H,*})$. Since $I_G(x^k) \cap I_\Phi^{0+}(x^k; t_k) = \emptyset$ and $I_H(x^k) \cap I_\Phi^{+0}(x^k; t_k) = \emptyset$ for all $k \in \mathbb{N}$, it is easy to see that the multipliers

$$\hat{\gamma}_i = \begin{cases} \gamma_i^* & \text{if } i \in \operatorname{supp}(\gamma^*), \\ -\delta_i^{G,*} & \text{if } i \in \operatorname{supp}(\delta^{G,*}), \\ 0 & \text{else} \end{cases} \quad \text{and} \quad \hat{\nu}_i = \begin{cases} \nu_i^* & \text{if } i \in \operatorname{supp}(\nu^*), \\ -\delta_i^{H,*} & \text{if } i \in \operatorname{supp}(\delta^{H,*}), \\ 0 & \text{else} \end{cases}$$

are well-defined, and we obtain

$$0 \;=\; \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^p \mu_i^* \nabla h_i(x^*) - \sum_{i=1}^q \hat{\gamma}_i \nabla G_i(x^*) - \sum_{i=1}^q \hat{\nu}_i \nabla H_i(x^*).$$

Here, $\lambda^* \geq 0$ and

$$\begin{aligned} \operatorname{supp}(\lambda^*) \;&\subseteq\; I_g(x^k) \subseteq I_g(x^*), \\ \operatorname{supp}(\hat{\gamma}) \;&=\; \operatorname{supp}(\gamma^*) \cup \operatorname{supp}(\delta^{G,*}) \;\subseteq\; I_G(x^k) \cup I_\Phi^{0+}(x^k; t_k) \;\subseteq\; I_{00}(x^*) \cup I_{0+}(x^*), \\ \operatorname{supp}(\hat{\nu}) \;&=\; \operatorname{supp}(\nu^*) \cup \operatorname{supp}(\delta^{H,*}) \;\subseteq\; I_H(x^k) \cup I_\Phi^{+0}(x^k; t_k) \;\subseteq\; I_{00}(x^*) \cup I_{+0}(x^*) \end{aligned}$$

for all k sufficiently large. Consequently, we have $\hat{\gamma}_i = 0$ for all $i \in I_{+0}(x^*)$ and $\hat{\nu}_i = 0$ for all $i \in I_{0+}(x^*)$, i.e., $(x^*, \lambda^*, \mu^*, \hat{\gamma}, \hat{\nu})$ is at least a weakly stationary point of the MPCC (1.1). To prove M-stationarity, assume that there is an $i \in I_{00}(x^*)$ with $\hat{\gamma}_i < 0$ and $\hat{\nu}_i \neq 0$ (the case $\hat{\gamma}_i \neq 0$ and $\hat{\nu}_i < 0$ can be treated in a symmetric way). The condition $\hat{\gamma}_i < 0$ implies $i \in \operatorname{supp}(\delta^{G,*}) \subseteq I_\Phi^{0+}(x^k; t_k)$ for all k sufficiently large. Because of

$$I_\Phi^{0+}(x^k; t_k) \cap \big(I_H(x^k) \cup I_\Phi^{+0}(x^k; t_k)\big) = \emptyset$$

for all $k \in \mathbb{N}$, this yields $\hat{\nu}_i = 0$ in contradiction to our assumption. □

Analogously to the relaxation method by Kardani et al. one can show that if MPCC-LICQ holds at $x^* \in X$, then the relaxed problems locally satisfy LICQ almost everywhere and Guignard CQ everywhere, see [19] for the details.

3.3 The Effects of Inexactness

In the previous convergence results we assumed that we are able to compute KKT point sof the relaxed problems. However, in practice we usually end up with *inexact KKT points*:

Definition 3.9. *Let $x^* \in \mathbb{R}^n$ and $\varepsilon > 0$ be given. If there exist vectors $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p$ such that*

$$\begin{aligned} &\left\| \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) \right\|_\infty \leq \varepsilon, \\ &g_i(x^*) \leq \varepsilon,\ \lambda_i \geq -\varepsilon,\ |g_i(x^*)\lambda_i| \leq \varepsilon \qquad \forall i = 1, \dots, m, \\ &|h_i(x^*)| \leq \varepsilon \qquad \forall\, i = 1, \dots, p, \end{aligned} \tag{3.12}$$

*x^* is called an ε-*stationary point *of the NLP*

$$\min_x f(x) \quad s.t. \quad g(x) \leq 0, h(x) = 0.$$

.

Thus, it es interesting to study which effect inexact KKT points have on the convergence of the afore mentioned relaxation methods, see [20] for more details. We begin by discussing the effects on the Scholtes relaxation.

Theorem 3.10. *Let $\{t_k\} \downarrow 0$, $\varepsilon_k = O(t_k)$, $\{x^k\}$ be a sequence of ε_k-stationary points of $NLP^S(t_k)$ with multipliers $(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^k)$, and assume that $x^k \to x^*$. If MPEC-MFCQ holds in x^*, then x^* is a C-stationary point of the MPEC.*

Proof. Since all x^k are ε_k-stationary points of NLP(t_k), we have

$$\Bigg\| \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) - \sum_{i=1}^{q} \gamma_i^k \nabla G_i(x^k) - \sum_{i=1}^{q} \nu_i^k \nabla H_i(x^k) + \sum_{i=1}^{q} \delta_i^k \nabla \Phi_i^S(x^k; t_k) \Bigg\|_\infty \leq \varepsilon_k$$

with

$$\begin{array}{llll} g_i(x^k) \leq \varepsilon_k, & \lambda_i^k \geq -\varepsilon_k, & |\lambda_i^k g_i(x^k)| \leq \varepsilon_k & \forall i = 1, \dots, m, \\ |h_i(x^k)| \leq \varepsilon_k & & & \forall i = 1, \dots, p, \\ G_i(x^k) \geq -\varepsilon_k, & \gamma_i^k \geq -\varepsilon_k, & |\gamma_i^k G_i(x^k)| \leq \varepsilon_k & \forall i = 1, \dots, q, \\ H_i(x^k) \geq -\varepsilon_k, & \nu_i^k \geq -\varepsilon_k, & |\nu_i^k H_i(x^k)| \leq \varepsilon_k & \forall i = 1, \dots, q, \\ \Phi_i^S(x^k; t_k) \leq \varepsilon_k, & \delta_i^k \geq -\varepsilon_k, & |\delta_i^k \Phi_i^S(x^k; t_k)| \leq \varepsilon_k & \forall i = 1, \dots, q, \end{array}$$

where $\nabla \Phi_i^S(x^k; t_k) = H_i(x^k) \nabla G_i(x^k) + G_i(x^k) \nabla H_i(x^k)$. Obviously, the limit x^* is feasible for the MPEC (1.1). We define the multipliers

$$\delta_i^{G,k} := \begin{cases} \delta_i^k H_i(x^k) & \text{if } i \in I_{00}(x^*) \cup I_{0+}(x^*), \\ 0 & \text{if } i \in I_{+0}(x^*), \end{cases}$$
$$\delta_i^{H,k} := \begin{cases} \delta_i^k G_i(x^k) & \text{if } i \in I_{00}(x^*) \cup I_{+0}(x^*), \\ 0 & \text{if } i \in I_{0+}(x^*). \end{cases}$$

Then we have

$$\Bigg\| \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) - \sum_{i=1}^{q} \gamma_i^k \nabla G_i(x^k) + \sum_{i=1}^{q} \delta_i^{G,k} \nabla G_i(x^k) + \sum_{i \in I_{+0}} \delta_i^k H_i(x^k) \nabla G_i(x^k) - \sum_{i=1}^{q} \nu_i^k \nabla H_i(x^k) + \sum_{i=1}^{q} \delta_i^{H,k} \nabla H_i(x^k) + \sum_{i \in I_{0+}} \delta_i^k G_i(x^k) \nabla H_i(x^k) \Bigg\|_\infty \leq \varepsilon_k.$$

We claim that the multipliers $(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^{G,k}, \delta^{H,k}, \delta^k_{I_{+0} \cup I_{0+}})$ are bounded. If the sequence were unbounded, we could assume without loss of generality convergence of the sequence

$$\frac{(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^{G,k}, \delta^{H,k}, \delta^k_{I_{+0} \cup I_{0+}})}{\|(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^{G,k}, \delta^{H,k}, \delta^k_{I_{+0} \cup I_{0+}})\|} \to (\bar{\lambda}, \bar{\mu}, \bar{\gamma}, \bar{\nu}, \bar{\delta}^G, \bar{\delta}^H, \bar{\delta}_{I_{+0} \cup I_{0+}}) \neq 0.$$

Then the ε_k-stationarity of x^k yields

$$\sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(x^*) + \sum_{i=1}^{p} \bar{\mu}_i \nabla h_i(x^*) - \sum_{i=1}^{q} \bar{\gamma}_i \nabla G_i(x^*) - \sum_{i=1}^{q} \bar{\nu}_i \nabla H_i(x^*) + \sum_{i=1}^{q} \bar{\delta}_i^G \nabla G_i(x^*) + \sum_{i=1}^{q} \bar{\delta}_i^H \nabla H_i(x^*) = 0, \tag{3.13}$$

where we took into account that $H_i(x^k) \to 0$ ($i \in I_{+0}(x^*)$) and $G_i(x^k) \to 0$ ($i \in I_{0+}(x^*)$). For all $i = 1, \dots, m$, the ε_k-stationarity implies $\bar{\lambda}_i \geq 0$. If $\bar{\lambda}_i > 0$, we have $\lambda_i^k > c$ for some constant $c > 0$ and all k sufficiently large. This yields

$$0 \leq |g_i(x^k)| \leq \frac{\varepsilon_k}{|\lambda_i^k|} \leq \frac{\varepsilon_k}{c} \to 0$$

due to $\varepsilon_k \downarrow 0$ and thus $i \in I_g(x^*)$. Analogously, we have $\bar{\gamma}_i \geq 0$ for all $i = 1, \dots, q$ and $\bar{\gamma}_i > 0$ implies $G_i(x^k) \to 0$ and thus $i \in I_{00}(x^*) \cup I_{0+}(x^*)$, and also $\bar{\nu}_i \geq 0$ for all $i = 1, \dots, q$ with $\bar{\nu}_i > 0$ implying $H_i(x^k) \to 0$ and thus $i \in I_{00}(x^*) \cup I_{+0}(x^*)$.

By definition $\bar{\delta}_i^G \neq 0$ implies $i \in I_{00}(x^*) \cup I_{0+}(x^*)$ and $\bar{\delta}_i^H \neq 0$ is only possible if $i \in I_{00}(x^*) \cup I_{+0}(x^*)$. Thus, we know $\operatorname{supp}(\bar{\lambda}) \subseteq I_g(x^*)$ and

$$\operatorname{supp}(\bar{\gamma}) \cup \operatorname{supp}(\bar{\delta}^G) \subseteq I_{00}(x^*) \cup I_{0+}(x^*), \quad \operatorname{supp}(\bar{\nu}) \cup \operatorname{supp}(\bar{\delta}^H) \subseteq I_{00}(x^*) \cup I_{+0}(x^*).$$

Our next step is to show

$$\operatorname{supp}(\bar{\gamma}) \cap \operatorname{supp}(\bar{\delta}^G) = \emptyset \quad \text{and} \quad \operatorname{supp}(\bar{\nu}) \cap \operatorname{supp}(\bar{\delta}^H) = \emptyset.$$

Without loss of generality, we consider only the case $i \in \operatorname{supp}(\bar{\gamma}) \cap \operatorname{supp}(\bar{\delta}^G)$ and show that this is impossible. By the definition of $\bar{\gamma}$ and $\bar{\delta}^G$, this would imply $|\gamma_i^k| \to \infty$ and $|\delta_i^{G,k}| = |\delta_i^k H_i(x^k)| \to \infty$ for $k \to \infty$. Due to $H_i(x^k) \to H_i(x^*)$, the latter implies $|\delta_i^k| \to \infty$. The ε_k-stationarity then yields

$$0 \leq \frac{|G_i(x^k)|}{\varepsilon_k} \leq \frac{1}{|\gamma_i^k|} \to 0 \quad \text{for } k \to \infty.$$

Analogously, we obtain

$$0 \leq \frac{|G_i(x^k)H_i(x^k) - t_k|}{\varepsilon_k} \leq \frac{1}{|\delta_i^k|} \to 0 \quad \text{for } k \to \infty.$$

Together, these two limits imply $\frac{t_k}{\varepsilon_k} \to 0$, a contradiction to $\varepsilon_k = O(t_k)$.

Hence, if $(\bar{\lambda}, \bar{\mu}, \bar{\gamma}, \bar{\nu}, \bar{\delta}^G, \bar{\delta}^H) \neq 0$, (3.13) yields a contradiction to MPEC-MFCQ. If, on the other hand, $(\bar{\lambda}, \bar{\mu}, \bar{\gamma}, \bar{\nu}, \bar{\delta}^G, \bar{\delta}^H) = 0$, there has to be an $i \in I_{0+}(x^*) \cup I_{+0}(x^*)$ with $\bar{\delta}_i \neq 0$. First consider the case $i \in I_{0+}(x^*)$. Then by definition

$$\bar{\delta}_i^G = \lim_{k \to \infty} \frac{\delta_i^k H_i(x^k)}{\|(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^{G,k}, \delta^{H,k}, \delta^k_{I_{+0} \cup I_{0+}})\|} = \bar{\delta}_i \underbrace{H_i(x^*)}_{>0} \neq 0,$$

a contradiction to the assumption $\bar{\delta}_i^G = 0$. In an analogous way, we obtain a contradiction in the case $i \in I_{+0}(x^*)$.

Consequently, the sequence $\{(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^{G,k}, \delta^{H,k}, \delta^k_{I_{+0} \cup I_{0+}})\}$ is bounded and therefore converges to some limit $(\lambda^*, \mu^*, \tilde{\gamma}, \tilde{\nu}, \tilde{\delta}^G, \tilde{\delta}^H, \tilde{\delta}_{I_{+0} \cup I_{0+}})$ at least on a subsequence. By passing to this subsequence, we can assume convergence on the whole sequence. Using the same arguments as before, it is easy to see that $\lambda^* \geq 0$, $\tilde{\gamma} \geq 0$, $\tilde{\nu} \geq 0$, and $\operatorname{supp}(\lambda^*) \subseteq I_g(x^*)$,

$$\operatorname{supp}(\tilde{\gamma}) \cup \operatorname{supp}(\tilde{\delta}^G) \subseteq I_{00}(x^*) \cup I_{0+}(x^*), \quad \operatorname{supp}(\tilde{\nu}) \cup \operatorname{supp}(\tilde{\delta}^H) \subseteq I_{00}(x^*) \cup I_{+0}(x^*).$$

Let us define the multipliers $\gamma^* := \tilde{\gamma} - \tilde{\delta}^G$ and $\nu^* := \tilde{\nu} - \tilde{\delta}^H$. Then x^* together with the multipliers $(\lambda^*, \mu^*, \gamma^*, \nu^*)$ is a weakly stationary point of the MPEC (1.1).

In order to prove C-stationarity of x^*, assume that there were an $i \in I_{00}(x^*)$ such that $\gamma_i^* \nu_i^* < 0$. We consider without loss of generality only the case $\gamma_i^* < 0$, $\nu_i^* > 0$, the other one can be treated the same way. Since we know $\tilde{\gamma}_i \geq 0$, the assumption $\gamma_i^* < 0$ implies $\tilde{\delta}_i^G > 0$. Due to $i \in I_{00}(x^*)$ and $\tilde{\delta}_i^G = \lim_{k\to\infty} \delta_i^{G,k} = \lim_{k\to\infty} \delta_i^k H_i(x^k)$, we obtain $|\delta_i^k| \to \infty$ for $k \to \infty$ and consequently we can again conclude

$$0 \leq \frac{|G_i(x^k)H_i(x^k) - t_k|}{\varepsilon_k} \leq \frac{1}{|\delta_i^k|} \to 0 \quad \text{for } k \to \infty. \tag{3.14}$$

If $\tilde{\nu}_i > 0$, the ε_k-stationarity would imply

$$0 \leq \frac{|H_i(x^k)|}{\varepsilon_k} \leq \frac{1}{\nu_i^k} \to \frac{1}{\tilde{\nu}_i} \quad \text{for } k \to \infty,$$

i.e. the quotient $\frac{H_i(x^k)}{\varepsilon_k}$ would remain bounded. Due to $i \in I_{00}(x^k)$, this together with (3.14) would yield $\frac{t_k}{\varepsilon_k} \to 0$, a contradiction to $\varepsilon_k = O(t_k)$.

Consequently, $\tilde{\nu}_i = 0$ and, therefore, $\tilde{\delta}_i^H < 0$. Together with $\tilde{\delta}_i^G > 0$, the definition of $\tilde{\delta}_i^H, \tilde{\delta}_i^G$ and the ε_k-stationarity then implies $\delta_i^k \to +\infty$ for $k \to \infty$ and $G_i(x^k) < 0$, $H_i(x^k) > 0$ for all k sufficiently large. Using (3.14), this would imply

$$0 < \frac{t_k}{\varepsilon_k} < \frac{|G_i(x^k)H_i(x^k) - t_k|}{\varepsilon_k} \to 0 \quad \text{for } k \to \infty,$$

once more a contradiction to $\varepsilon_k = O(t_k)$. Thus, x^* with the multipliers $(\lambda^*, \mu^*, \gamma^*, \nu^*)$ is a C-stationary point of the MPEC (1.1). □

The following example illustrates that we cannot guarantee C-stationarity of the limit x^* without the assumption $\varepsilon_k = O(t_k)$ used in Theorem 3.10.

Example 3.11. Consider the MPEC

$$\min x_2 - x_1 \quad \text{s.t.} \quad x_1 \geq 0, x_2 \geq 0, x_1 x_2 = 0.$$

Now, for an arbitrary $t > 0$ the conditions for ε_t-stationarity of $\mathrm{NLP}^S(t)$ read as follows:

$$\begin{aligned}
\left\| \begin{pmatrix} -1 \\ 1 \end{pmatrix} - \gamma^t \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \nu^t \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \delta^t \begin{pmatrix} x_2^t \\ x_1^t \end{pmatrix} \right\|_\infty &\leq \varepsilon_t, \\
x_1^t \geq -\varepsilon_t,\ \gamma^t \geq -\varepsilon_t,\ |\gamma^t x_1^t| &\leq \varepsilon_t, \\
x_2^t \geq -\varepsilon_t,\ \nu^t \geq -\varepsilon_t,\ |\nu^t x_2^t| &\leq \varepsilon_t, \\
x_1^t x_2^t - t \leq \varepsilon_t,\ \delta^t \geq -\varepsilon_t,\ |\delta^t(x_1^t x_2^t - t)| &\leq \varepsilon_t.
\end{aligned}$$

If we choose $\varepsilon_t = 2\sqrt{t}$, it is easy to verify that $x^t = (\sqrt{t}, \sqrt{t})$ together with the multipliers $\gamma^t = 0$, $\nu^t = 2$ and $\delta^t = \frac{1}{\sqrt{t}}$ is an ε_t-stationary point of $\mathrm{NLP}^S(t)$. However, for $t \downarrow 0$, the sequence x^t converges to the origin, which is a weakly but not C-stationary point of the MPEC.

So the convergence properties of the Scholtes relaxation method stay the same if we replace exact KKT points of the relaxed problems by inexact approximations as long as the accuracy ε_k decreases at least as fast as the relaxation parameter t_k. One can show that the relaxation method by Lin and Fukushima enjoys the same favorable property. However, for all other relaxation methods, if one replaces KKT points by inexact approximations, one in general only obtains weakly stationary limits. We illustrate this in the example of the relaxation from Kanzow and Schwartz.

Theorem 3.12. *Let $\{t_k\} \downarrow 0$, $\{\varepsilon_k\} \downarrow 0$, $\{x^k\}$ be a sequence of ε_k-stationary points of NLP$^{KS}(t_k)$, and assume that $x^k \to x^*$ with MPEC-MFCQ holding in x^*. Then x^* is a weakly stationary point of the MPEC.*

Proof. Since all x^k are ε_k-stationary points of NLP(t_k), we have

$$\Big\| \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) \\ - \sum_{i=1}^{q} \gamma_i^k \nabla G_i(x^k) - \sum_{i=1}^{q} \nu_i^k \nabla H_i(x^k) + \sum_{i=1}^{q} \delta_i^k \nabla \Phi_i^{KS}(x^k; t_k) \Big\|_\infty \le \varepsilon_k$$

with

$$\begin{array}{llll} g_i(x^k) \le \varepsilon_k, & \lambda_i^k \ge -\varepsilon_k, & |\lambda_i^k g_i(x^k)| \le \varepsilon_k & \forall i = 1, \dots, m, \\ |h_i(x^k)| \le \varepsilon_k & & & \forall i = 1, \dots, p, \\ G_i(x^k) \ge -\varepsilon_k, & \gamma_i^k \ge -\varepsilon_k, & |\gamma_i^k G_i(x^k)| \le \varepsilon_k & \forall i = 1, \dots, q, \\ H_i(x^k) \ge -\varepsilon_k, & \nu_i^k \ge -\varepsilon_k, & |\nu_i^k H_i(x^k)| \le \varepsilon_k & \forall i = 1, \dots, q, \\ \Phi_i^{KS}(x^k; t_k) \le \varepsilon_k, & \delta_i^k \ge -\varepsilon_k, & |\delta_i^k \Phi_i^{KS}(x^k; t_k)| \le \varepsilon_k & \forall i = 1, \dots, q, \end{array}$$

where $\Phi_i^{KS}(x^k; t_k)$ is defined as before with the gradient

$$\nabla \Phi_i^{KS}(x^k; t_k) = \begin{cases} (H_i(x^k) - t_k)\nabla G_i(x^k) + (G_i(x^k) - t_k)\nabla H_i(x^k) & \text{if } G_i(x^k) + H_i(x^k) \ge 2t_k, \\ -(G_i(x^k) - t_k)\nabla G_i(x^k) - (H_i(x^k) - t_k)\nabla H_i(x^k) & \text{else.} \end{cases}$$

Hence, the limit x^* is obviously feasible for the MPEC (1.1). We define the multipliers

$$\tilde{\gamma}_i^k := \begin{cases} \gamma_i^k - \delta_i^k (H_i(x^k) - t_k) & \text{if } G_i(x^k) + H_i(x^k) \ge 2t_k, \\ \gamma_i^k + \delta_i^k (G_i(x^k) - t_k) & \text{else,} \end{cases}$$
$$\tilde{\nu}_i^k := \begin{cases} \nu_i^k - \delta_i^k (G_i(x^k) - t_k) & \text{if } G_i(x^k) + H_i(x^k) \ge 2t_k, \\ \nu_i^k + \delta_i^k (H_i(x^k) - t_k) & \text{else.} \end{cases}$$

Then we have

$$\Big\| \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) - \sum_{i=1}^{q} \tilde{\gamma}_i^k \nabla G_i(x^k) - \sum_{i=1}^{q} \tilde{\nu}_i^k \nabla H_i(x^k) \Big\|_\infty \le \varepsilon_k.$$

We claim that the multipliers $(\lambda^k, \mu^k, \tilde{\gamma}^k, \tilde{\nu}^k)$ are bounded. If the sequence were unbounded, we could assume without loss of generality convergence of the sequence

$$\frac{(\lambda^k, \mu^k, \tilde{\gamma}^k, \tilde{\nu}^k)}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \tilde{\nu}^k\|} \to (\bar{\lambda}, \bar{\mu}, \bar{\gamma}, \bar{\nu}) \neq 0.$$

Then the ε_k-stationarity of x^k yields

$$\sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(x^*) + \sum_{i=1}^{p} \bar{\mu}_i \nabla h_i(x^*) - \sum_{i=1}^{q} \bar{\gamma}_i \nabla G_i(x^*) - \sum_{i=1}^{q} \bar{\nu}_i \nabla H_i(x^*) = 0.$$

Additionally, the ε_k-stationarity yields $\bar{\lambda}_i \ge 0$ for all $i = 1, \dots, m$, and $\bar{\lambda}_i > 0$ implies $g_i(x^*) = 0$, hence $\text{supp}(\bar{\lambda}) \subseteq I_g(x^*)$.

Now consider an $i \in I_{+0}(x^*)$. This implies $G_i(x^k) + H_i(x^k) \ge 2t_k$ for all k sufficiently large and thus $\tilde{\gamma}_i^k = \gamma_i^k - \delta_i^k (H_i(x^k) - t_k)$. The ε_k-stationarity yields $\gamma_i^k G_i(x^k) \to 0$, hence $\gamma_i^k \to 0$, and

$$\delta_i^k \Phi_i^{KS}(x^k; t_k) = \delta_i^k (H_i(x^k) - t_k)(G_i(x^k) - t_k) \to 0,$$

thus $\delta_i^k(H_i(x^k)-t_k)\to 0$. Consequently, we have $\bar{\gamma}_i = 0$. This shows that $\text{supp}(\bar{\gamma}) \subseteq I_{00}(x^*) \cup I_{0+}(x^*)$. By a symmetric argument, we obtain $\text{supp}(\bar{\nu}) \subseteq I_{00}(x^*) \cup I_{+0}(x^*)$. Thus, the equation above reduces to

$$\sum_{i\in I_g} \bar{\lambda}_i \nabla g_i(x^*) + \sum_{i=1}^{p} \bar{\mu}_i \nabla h_i(x^*) - \sum_{i\in I_{00}\cup I_{0+}} \bar{\gamma}_i \nabla G_i(x^*) - \sum_{i\in I_{00}\cup I_{+0}} \bar{\nu}_i \nabla H_i(x^*) = 0$$

with $\bar{\lambda}_i \geq 0$ for all $i \in I_g(x^*)$. Hence MPEC-MFCQ implies $\bar{\lambda}_i = 0\,(i \in I_g)$, $\bar{\mu}_i = 0\,(i = 1,\ldots,p)$, $\bar{\gamma}_i = 0\,(i \in I_{00} \cup I_{0+})$, and $\bar{\nu}_i = 0\,(i \in I_{00} \cup I_{+0})$. Altogether, we get a contradiction to the fact that $(\bar{\lambda}, \bar{\mu}, \bar{\gamma}, \bar{\nu}) \neq 0$.

Hence the sequence $\{(\lambda^k, \mu^k, \tilde{\gamma}^k, \tilde{\nu}^k)\}$ is bounded. Without loss of generality, we can assume that the entire sequence $\{(\lambda^k, \mu^k, \tilde{\gamma}^k, \tilde{\nu}^k)\}$ converges to a limit $(\lambda^*, \mu^*, \gamma^*, \nu^*)$. The limit is then weakly stationary since the multipliers $\lambda^*, \mu^*, \gamma^*, \nu^*$ have the same properties as $\bar{\lambda}, \bar{\mu}, \bar{\gamma}, \bar{\nu}$. □

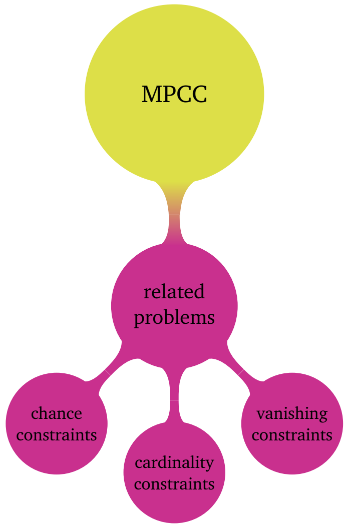
Under additional assumptions on the sequence, one can also ensure stronger stationarity properties of the limit, see [20] for details. But the following example illustrates, that additional assumptions are indeed necessary.

Example 3.13. Consider the two-dimensional MPEC

$$\min -x_1 - x_2 \quad \text{s.t.} \quad 0 \leq x_1 \perp x_2 \geq 0$$

and sequences $t \downarrow 0$, $\varepsilon_t = t^2$. Then it is easy to verify that the points $x^t = \big((1-t)t,\ (1-t)t\big)^T$ are ε_t-stationary points of $\text{NLP}^{KS}(t)$ with the multipliers $\gamma^t = 0, \nu^t = 0, \delta^t = \frac{1}{\varepsilon_t}$. On the other hand $x^t \to (0,0)^T$, which is a C-stationary point of the MPEC and satisfies even MPEC-LICQ, but is not an M-stationary point.

4 Related Problem Classes



Finally, we want to have a look at some classes opt optimization problems, which are related to MPCCs but need special treatment nonetheless.

4.1 Mathematical Programs with Vanishing Constraints

A *mathematical program with vanishing constraints* (MPVC) is of the form

$$\min_x f(x) \quad \text{s.t.} \quad \begin{aligned} & g(x) \leq 0, h(x) = 0, \\ & H(x) \geq 0, G(x) \circ H(x) \leq 0, \end{aligned}$$

where f, g, h, G, H have the same dimensions as in the MPCC. So contrary to an MPCC and MPVC lacks the constraint $G(x) \geq 0$. The feasible set of one vanishing constraint is depicted in Figure 4.1. We see, that in case $H_i(x) > 0$ we have the constraint $G_i(x) \leq 0$ but in case $H_i(x) = 0$ the constraint on G_i vanishes and all $G_i(x) \in \mathbb{R}$ are feasible. These constraints have applications e.g. in truss design, see for example [1, 2].

MPVCs suffer from similar problems as MPCCs when it comes to constraint qualifications.

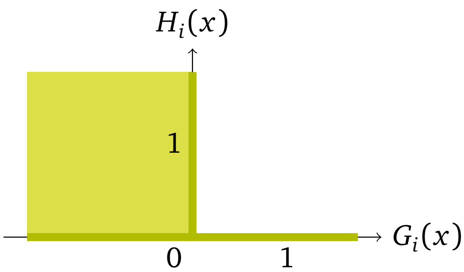


Figure 4.1: Illustration of the constraint $H_i(x) \geq 0, G_i(x)H_i(x) \leq 0$

Lemma 4.1. *Let x^* be feasible for MPVC and $H_i(x^*) = 0$ for at least one index $i \in \{1,\dots,q\}$. Then LICQ is violated in x^*.*

If there exists an index i such that $H_i(x^*) = 0$ and $G_i(x^*) \geq 0$, then MFCQ is violated in x^*.

In points, where there exists an index i such that $H_i(x^*) = 0$ and $G_i(x^*) = 0$, even Abadie CQ may be violated.

Unfortunately, the points, where LICQ or MFCQ fails are often of practical interest. I.e. in truss design they correspond to potential bars, which are not realized in the final truss.

One can reformulate an MPVC as an MPCC using slack variables as follows:

$$\begin{aligned} \min_{x,y} f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & G(x) - y \leq 0, \\ & 0 \leq H(x) \perp y \geq 0. \end{aligned}$$

Then x^* is a solution of the MPVC if and only if (x^*, y^*) with

$$y_i^* \begin{cases} = 0 & \text{if } H_i(x^*) > 0, \\ \geq \max\{0, G_i(x^*)\} & \text{if } H_i(x^*) = 0 \end{cases}$$

if a solution of the corresponding MPCC.

So in theory, it is possible to handle an MPVC using MPCC theory. However, this approach has some drawbacks: First, the slack variable y is not uniquely determined, which may cause numerical problems. Secondly, MPVCs have slightly better theoretical problems than MPCCs, e.g. LICQ and MFCQ are not violated in all feasible points and thus the KKT conditions have a better chance at being optimality conditions. Thus it makes sense to use the ideas from MPCC theory but to apply them to the MPVC structure directly.

4.2 Optimization Problems with Cardinality Constraints

Optimization problems with cardinality constraints are of the form

$$\begin{aligned} \min_{x} f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & \|x\|_0 \leq \kappa, \end{aligned} \tag{4.1}$$

where f, g, h have the same dimensions as before, $\kappa \in \{1,\dots,n-1\}$ and

$$\operatorname{supp}(x) = \{i \mid x_i \neq 0\} \quad \text{and} \quad \|x\|_0 = |\operatorname{supp}(x)|.$$

Thus, while $x \in \mathbb{R}^n$ only $\kappa < n$ components of x are allowed to be nonzero at the same time. The resulting feasible set is usually nonconvex and possibly even disconnected. And the function $x \mapsto \|x\|_0$ is discrete-valued and thus neither convex nor continuous. Also, although the notation may lead to a different impression, it is not a norm because it is not positively homogeneous.

One popular application of such constraints are portfolio optimization problems such as

$$\begin{aligned} \min_{x} x^T Q x \quad \text{s.t.} \quad & \mu^T x \geq \rho, \\ & e^T x = 1, \\ & \|x\|_0 \leq \kappa, \end{aligned}$$

see for example [4]. Here, one is given n possible assets and tries to choose κ out of them, which minimize the resulting risk $x^T Q x$ while at the same time providing expected returns of at least ρ.

Similar problems also appear in in compressed sensing, see e.g. [6], where they are often tackled by replacing $\|x\|_0$ with the l_1 norm. In the literature one can find a large number of different approaches to handle cardinality constraints such as mixed integer reformulations, DC approaches, exponential or convex approximations, greedy methods and heuristics, see for example [4, 14, 33, 25, 10, 11].

An alternative approach going back to [11, 5, 7] is based on a complementarity-type reformulation:

$$\begin{aligned}\min_{x,y} f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & 0 \leq y \leq e, e^T y \geq n - \kappa, \\ & x \circ y = 0.\end{aligned} \tag{4.2}$$

Here, y is an originally binary variable used to counting the zero components of x, of which there should be at least $n-\kappa$. Due to the special structure of the constraints, it is possible to relax the binary variable y to a continuous variable without destroying the relation between the two problems.

Lemma 4.2. *(a) x^* is feasible for* (4.1) *if and only if there exists y such that (x^*, y) is feasible for* (4.2).

(b) x^* is a global solution of (4.1) *if and only if (x^*, y) is a global solution of* (4.2) *for all y, for which it is feasible.*

(c) If x^* is a local solution of (4.1)*, then (x^*, y) is a local solution of* (4.2) *for all y, for which it is feasible.*

(d) If (x^*, y) is a local solution of (4.2) *with $\|x^*\|_0 = \kappa$, then x^* is a local solution of* (4.1).

If we consider a local solution (x^*, y) of (4.2) with $\|x^*\|_0 < \kappa$, it may not be a local solution of the original problem (4.1). This is illustrated by the following example. However, in numerical test, we have so far not encountered this case.

Example 4.3. Consider the 3-dimensional problem

$$\min_x \|x - (2,1,0)^T\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leq 1.$$

The global minimum is $x^* = (2,0,0)^T$ and the local minimum is $\hat{x} = (0,1,0)^T$. The point $\tilde{x} = (0,0,0)^T$ is not a local solution of the original problem, but e.g. together with $\tilde{y} = (1,1,0)$ it is a local solution of the reformulation.

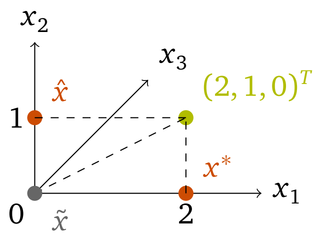


Figure 4.2: Illustration of Example 4.3

The complementarity-type constraint

$$y_i \geq 0, \quad x_i y_i = 0 \qquad \forall i = 1, \dots, n$$

is somewhere between a vanishing and a complementarity constraints. Compared to a complementarity constraint, we are lacking the constraint $x \geq 0$. However, in some applications, we know $x \geq 0$ or we can enforce this using a split $x = x^+ - x^-$ with $x^+, x^- \geq 0$. Compared to a vanishing constraint the difference is $x_i y_i = 0$ instead of $x_i y_i \leq 0$.

As one might expect, direct application of standard NLP theory to the reformulated problem (4.2) is problematic due to the fact that standard constraint qualifications are usually violated. So one might be tempted to try to apply MPCC theory instead, especially in cases, where one has the additional constraint $x \geq 0$. However, this also has its drawbacks: In all points where y is binary and $e^T y = n - \kappa$ (which is the case if $\|x\|_0 = \kappa$, MPCD-LICQ and MPCC-MFCQ are violated due to the gradients with respect to y being positively linearly dependent. Nonetheless, the reformulation has better theoretical properties than an arbitrary MPCC.

To see this, let us introduce the index set

$$I_0(x) := \{i = 1, \dots, n \mid x_i = 0\}$$

and define two optimality conditions.

Definition 4.4. *A feasible point* (x^*, y^*) *of the reformulation* (4.2) *is called*

- M-stationary*, if there are multipliers* λ, μ, γ *with*

$$\nabla f(x^*) + \sum_{i \in I_g} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) + \sum_{i \in I_0} \gamma_i e_i = 0$$
$$\lambda_i \geq 0 \quad \forall i \in I_g(x^*)$$

- S-stationary*, if additionally*

$$\gamma_i = 0 \quad \textit{for all} \quad i \in I_0(x^*) \ : \ y_i^* = 0.$$

The terminology is based on MPCC theory, i.e. S-stationarity is equivalent to a KKT point of the reformulation and M-stationarity is based on the limiting normal cone. One can also try to define analogues to C- and weak stationarity, but they coincide with M-stationarity here.

M-stationarity has the nice property that it independent from y and corresponds to the KKT conditions of the tightened problem

$$\min_x f(x) \quad \text{s.t.} \quad g(x) \leq 0, h(x) = 0, x_i = 0 \quad \forall i \in I_0(x^*).$$

This motivates the definition of tailored constraint qualifications as follows:

Definition 4.5. *Let* x^* *be feasible for* (4.1)*. Then CC-LICQ (CC-MFCQ, CC-CRCQ, CC-CPLD) holds there if LICQ (MFCQ; CRCQ, CPLD) for*

$$\min_x f(x) \quad \textit{s.t.} \quad g(x) \leq 0, h(x) = 0, x_i = 0 \quad \forall i \in I_0(x^*).$$

holds at x^*.

On can also define CC-analogues of Abadie and Guignard CQ similarly to the MPCC case. And contrary to MPCCs one can prove the following result:

Theorem 4.6. *Let* (x^*, y^*) *be a local minimum of the reformulation* (4.2) *and any CC constraint qualification such as CC-LICQ or CC-Guignard CQ hold there. Then* (x^*, y^*) *is S-stationary.*

For MPCCs we have seen that S-stationarity is a necessary optimality condition only under MPCC-LICQ but may already fail under MPCC-MFCQ.

To solve the continuous reformulation, one can employ similar relaxation algorithms as for MPCCs. And also here, one obtains better results.

Theorem 4.7. *Let $(t_k) \downarrow 0$ and (x^k, y^k) be KKT points of the Scholtes-type relaxed problems*

$$\begin{aligned} \min_{x,y} f(x) \quad s.t. \quad & g(x) \leq 0, h(x) = 0, \\ & 0 \leq y \leq e, e^T y \geq n - \kappa, \\ & -t_k e \leq x \circ y \leq t_k e. \end{aligned}$$

If $(x^k, y^k) \to (x^*, y^*)$ and CC-MFCQ holds there, then (x^*, y^*) is an S-stationary point.

For MPCCs we can in general only guarantee C-stationarity of the limit, which here would correspond to an M-stationary point.

4.3 Optimization Problems with Chance Constraints

Another class of problems, which can be reformulated using a complementarity-type constraint, are *optimization problems with chance constraints*, see [3]. These are of the form

$$\begin{aligned} \min_{x} f(x) \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & P(G(x, \xi) \leq 0) \geq 1 - \varepsilon, \end{aligned}$$

where $\varepsilon \in (0, 1)$ is a given parameter and ξ is a random vector with finitely many possible realization $\xi_1 \ldots, \xi_N$ and known probabilities $p_1, \ldots, p_N$. Then the problem can be rewritten equivalently as

$$\begin{aligned} \min_{x,y} \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & y \in \{0, 1\}^N, \\ & p^T y \geq 1 - \varepsilon, \\ & y_i G(x, \xi_i) \leq 0. \end{aligned}$$

If we again relax the binary variable y, we end up with the problem

$$\begin{aligned} \min_{x,y} \quad \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & 0 \leq y \leq e, \\ & p^T y \geq 1 - \varepsilon, \\ & y_i G(x, \xi_i) \leq 0 \quad \forall i = 1, \ldots, N, \end{aligned}$$

where one can show that feasible points still are equivalent to feasible points of the original problem. Obviously, the reformulated problem is very similar to the reformulation of the cardinality constrained problem. The differences are that the sum over y_i is now weighted by the probabilities p_i and that the constraint $y_i G(x, \xi_i) \leq 0$ is more of a vanishing type.

For these problems one can obtain a similar optimality as for the reformulation of cardinality constrained problem.

Bibliography

[1] Wolfgang Achtziger and Christian Kanzow. "Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications". In: *Mathematical Programming* 114.1 (2008), pp. 69–99.

[2] Wolfgang Achtziger, Christian Kanzow, and Tim Hoheisel. "On a relaxation method for mathematical programs with vanishing constraints". In: *GAMM-Mitteilungen* 35.2 (2012), pp. 110–130.

[3] Lukáš Adam and Martin Branda. "Nonlinear chance constrained problems: optimality conditions, regularization and solvers". In: *Journal of Optimization Theory and Applications* 170.2 (2016), pp. 419–436.

[4] Daniel Bienstock. "Computational study of a family of mixed-integer quadratic programming problems". In: *Mathematical programming* 74.2 (1996), pp. 121–140.

[5] Oleg P Burdakov, Christian Kanzow, and Alexandra Schwartz. "Mathematical programs with cardinality constraints: Reformulation by complementarity-type conditions and a regularization method". In: *SIAM Journal on Optimization* 26.1 (2016), pp. 397–425.

[6] Emmanuel J Candès and Michael B Wakin. "An introduction to compressive sampling". In: *IEEE signal processing magazine* 25.2 (2008), pp. 21–30.

[7] Michal Červinka, Christian Kanzow, and Alexandra Schwartz. "Constraint qualifications and optimality conditions for optimization problems with cardinality constraints". In: *Mathematical Programming* 160.1-2 (2016), pp. 353–377.

[8] Frank H Clarke. *Optimization and nonsmooth analysis*. Vol. 5. Siam, 1990.

[9] Stephan Dempe et al. *Bilevel programming problems*. Springer, 2015.

[10] David Di Lorenzo et al. "A concave optimization-based approach for sparse portfolio selection". In: *Optimization Methods and Software* 27.6 (2012), pp. 983–1000.

[11] Mingbin Feng et al. "Complementarity formulations of l0-norm optimization problems". In: (2013).

[12] Michael L Flegel and Christian Kanzow. "Abadie-type constraint qualification for mathematical programs with equilibrium constraints". In: *Journal of Optimization Theory and Applications* 124.3 (2005), pp. 595–614.

[13] Michael L Flegel and Christian Kanzow. "On the Guignard constraint qualification for mathematical programs with equilibrium constraints". In: *Optimization* 54.6 (2005), pp. 517–534.

[14] Antonio Frangioni and Claudio Gentile. "SDP diagonalizations and perspective cuts for a class of nonseparable MIQP". In: *Operations Research Letters* 35.2 (2007), pp. 181–185.

[15] René Henrion, Abderrahim Jourani, and Jiri Outrata. "On the calmness of a class of multifunctions". In: *SIAM Journal on Optimization* 13.2 (2002), pp. 603–618.

[16] Tim Hoheisel, Christian Kanzow, and Alexandra Schwartz. "Convergence of a local regularization approach for mathematical programmes with complementarity or vanishing constraints". In: *Optimization Methods and Software* 27.3 (2012), pp. 483–512.

[17] Alexey F Izmailov and Mikhail V Solodov. *Newton-type methods for optimization and variational problems*. Springer, 2014.

[18] Abdeslam Kadrani, Jean-Pierre Dussault, and Abdelhamid Benchakroun. “A new regularization scheme for mathematical programs with complementarity constraints”. In: *SIAM Journal on Optimization* 20.1 (2009), pp. 78–103.

[19] Christian Kanzow and Alexandra Schwartz. “A new regularization method for mathematical programs with complementarity constraints with strong convergence properties”. In: *SIAM Journal on Optimization* 23.2 (2013), pp. 770–798.

[20] Christian Kanzow and Alexandra Schwartz. “The price of inexactness: Convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited”. In: *Mathematics of Operations Research* 40.2 (2014), pp. 253–275.

[21] Gui-Hua Lin and Masao Fukushima. “A modified relaxation scheme for mathematical programs with complementarity constraints”. In: *Annals of Operations Research* 133.1-4 (2005), pp. 63–84.

[22] Zhi-Quan Luo, Jong-Shi Pang, and Daniel Ralph. *Mathematical programs with equilibrium constraints*. Cambridge University Press, 1996.

[23] Boris S Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory*. Vol. 330. Springer Science & Business Media, 2006.

[24] Boris S Mordukhovich. *Variational Analysis and Generalized Differentiation II: Applications*. Springer Science & Business Media, 2006.

[25] Walter Murray and Howard Shek. “A local relaxation method for the cardinality constrained portfolio optimization problem”. In: *Computational Optimization and Applications* 53.3 (2012), pp. 681–709.

[26] Jiri Outrata, Michal Kocvara, and Jochem Zowe. *Nonsmooth approach to optimization problems with equilibrium constraints: theory, applications and numerical results*. Vol. 28. Springer Science & Business Media, 2013.

[27] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*. Vol. 317. Springer Science & Business Media, 2009.

[28] Holger Scheel and Stefan Scholtes. “Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity”. In: *Mathematics of Operations Research* 25.1 (2000), pp. 1–22.

[29] Winfried Schirotzek. *Nonsmooth analysis*. Springer Science & Business Media, 2007.

[30] Stefan Scholtes. “Convergence properties of a regularization scheme for mathematical programs with complementarity constraints”. In: *SIAM Journal on Optimization* 11.4 (2001), pp. 918–936.

[31] Sonja Steffensen and Michael Ulbrich. “A new relaxation scheme for mathematical programs with equilibrium constraints”. In: *SIAM Journal on Optimization* 20.5 (2010), pp. 2504–2539.

[32] Defeng Sun and Liqun Qi. “On NCP-functions”. In: *Computational Optimization and Applications* 13.1-3 (1999), pp. 201–220.

[33] Xiaojin Zheng et al. “Successive convex approximations to cardinality-constrained convex programs: a piecewise-linear DC approach”. In: *Computational Optimization and Applications* 59.1-2 (2014), pp. 379–397.