MATHEMATICAL PROGRAMS WITH EQUILIBRIUM CONSTRAINTS: ENHANCED FRITZ JOHN-CONDITIONS, NEW CONSTRAINT QUALIFICATIONS AND IMPROVED EXACT PENALTY RESULTS

Christian Kanzow[1] and Alexandra Väth[1]

Preprint 292 October 2009

[1] University of Würzburg
Institute of Mathematics
Am Hubland, 97074 Würzburg, Germany

e-mail: kanzow@mathematik.uni-wuerzburg.de
alexandra.vaeth@mathematik.uni-wuerzburg.de

October 26, 2009

Abstract. Mathematical programs with equilibrium (or complementarity) constraints (MPECs for short) form a difficult class of optimization problems. The standard KKT conditions are not always necessary optimality conditions due to the fact that suitable constraint qualifications are often violated. Alternatively, one can therefore use the Fritz John-approach to derive necessary optimality conditions. While the usual Fritz John-conditions do not provide much information, we prove an enhanced version of the Fritz John-conditions. This version motivates the introduction of some new constraint qualifications (CQs) which can then be used in order to obtain, for the first time, a completely elementary proof of the fact that a local minimum is an M-stationary point under one of these CQs. We also show how these CQs can be used to obtain a suitable exact penalty result under weaker or different assumptions than those that can be found in the literature.

Key Words: Mathematical programs with equilibrium constraints, Fritz John conditions, Constraint qualification, M-stationarity, Exact penalization.

1 Introduction

We consider an optimization problem of the form

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

with continuously differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$, and $G, H : \mathbb{R}^n \to \mathbb{R}^q$ (in some of our results, however, f will only be assumed to be locally Lipschitz continuous). This problem is called a *mathematical program with equilibrium (or complementarity) constraints*, MPEC for short. The MPEC has its origin in bilevel programming (see [4]) which arises naturally in the various applications of the Stackelberg game in economic sciences. Further applications in engineering and natural sciences then led to the extension of bilevel programs to MPECs. The reader is referred to the monographs [12, 19] for further details.

MPECs are known to be difficult optimization problems due to the fact that many of the standard constraint qualifications (like the linear independence and the Mangasarian-Fromovitz constraint qualifications, LICQ and MFCQ for short) are violated at any feasible point. Hence, the usual KKT conditions are not always necessary optimality conditions, even in the case where all constraint functions g_i, h_i, G_i, H_i are linear. Special MPEC-tailored CQs can be used in order to show that a concept which is called *strong stationarity* gives first order optimality conditions, see [12, 20, 25, 22] for a corresponding discussion. However, this strong stationarity concept can be shown to be equivalent to the KKT conditions of an MPEC, see [7], hence also strong stationarity is not always a necessary optimality condition. A slightly weaker concept, introduced in [28, 17, 18, 26], is called *M-stationarity* and can be shown to provide first-order optimality conditions under fairly mild assumptions, cf. [6, 8, 27].

A natural idea to overcome the fact that standard constraint qualifications are typically violated is to use the Fritz John-approach since these conditions do not require any CQs. This was already done in [5], mainly in order to give a simple proof for strong stationarity to be a necessary optimality condition under a suitable CQ. However, it should also be noted that the standard Fritz John-conditions applied to MPECs do not give much information regarding the signs of the Lagrange multipliers. In [27] (see also [6]), the reader can find an improved version of the Fritz John-conditions which was used, based on the limiting subdifferential and the limiting co-derivatives by Mordukhovich, to obtain M-stationarity as a necessary optimality condition.

Here we follow an idea from [1] where enhanced Fritz John-conditions were obtained for standard nonlinear programs (with an additional abstract constraint). We exploit the special structure of the complementarity constraints within (1) in order to obtain further improved Fritz John-conditions which are even stronger than those obtained in [27, 6]. These enhanced Fritz John-conditions motivate the introduction of new constraint qualifications for MPECs which can then be used in order to obtain M-stationarity as a necessary optimality condition. While the latter is, in principle, a known result, we stress that we get, for the first time, a completely elementary proof of this fact, whereas all previous papers dealing with this topic make use of the limiting subdifferential and/or the limiting co-derivative by Mordukhovich, see [15, 16, 21] for an extensive treatment of these concepts. While these are very powerful tools from variational analysis, they are not known to the whole optimization community. Nevertheless, we will also use the limiting subdifferential to show how one of our new constraint qualification can be used to get an improved exact penalty result for MPECs.

The organization of the paper is as follows: We first give some background material in Section 2. We then prove our enhanced Fritz John-result in Section 3, where we also introduce the corresponding new CQs and state the M-stationary result. Section 4 contains our exact penalty result for MPECs. The precise relation between our new CQs and some existing ones is discussed in Section 5. We then close with some final remarks in Section 6.

Notation: $\|\cdot\|$ denotes an arbitrary norm in $\mathbb{R}^n$. Whenever it is important that a special l_p-norm is used, we will provide the corresponding index. Mostly, we will use the l_1-norm

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|,$$

but $\|\cdot\|_2$ and $\|\cdot\|_\infty$ will also appear. The open ball $B_\varepsilon(x) = \{y \in \mathbb{R}^n \mid \|y - x\| < \varepsilon\}$ is always defined using the norm in the respective context. Finally, given two vectors $x, y \in \mathbb{R}^n$, we write $0 \leq x \perp y \geq 0$ as a short-hand for $x \geq 0, y \geq 0$ and $x^T y = 0$.

2 Background Material

Here we introduce some useful notations and give precise definitions of the material that will be used in our subsequent sections.

First, let the feasible set of (1) be denoted by

$$X := \{x \in \mathbb{R}^n \mid g(x) \leq 0,\ h(x) = 0,\ G_i(x) \geq 0,\ H_i(x) \geq 0,\ G_i(x)H_i(x) = 0\ \forall i = 1, \ldots, q\}.$$

To facilitate the notation, we define the following index sets for an arbitrary $x^* \in X$:

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

Note that the first (second) subscript indicates whether $G_i(x^*)$ ($H_i(x^*)$) is zero or positive at the given point x^*. With this notation, we are now able to define a stationarity concept for MPECs, the so called M-stationarity.

Definition 2.1. *Let x^* be feasible for* (1). *Then x^* is called* M-stationary *if there are multipliers $(\lambda, \mu, \gamma, \nu)$ such that*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) - \sum_{i=1}^{q} \gamma_i \nabla G_i(x^*) - \sum_{i=1}^{q} \nu_i \nabla H_i(x^*) = 0$$

and $\lambda \geq 0$, $\lambda_i = 0$ for all $i \notin I_g(x^*)$, $\gamma_i = 0$ for all $i \in I_{+0}(x^*)$, $\nu_i = 0$ for all $i \in I_{0+}(x^*)$ and either $\gamma_i \nu_i = 0$ or $\gamma_i > 0, \nu_i > 0$ for all $i \in I_{00}(x^*)$.

The concept of *strong stationarity* requires the slightly sharper condition that $\gamma_i \geq 0$ and $\nu_i \geq 0$ for all $i \in I_{00}(x^*)$, whereas all other conditions from Defintion 2.1 remain unchanged. In general, however, strong stationarity is a necessary first order condition for MPECs only under relatively restrictive conditions. It might be violated even for the case where all functions g_i, h_i, G_i, H_i are linear. On the other hand, M-stationarity is satisfied under this assumption and in many other cases, cf. the corresponding discussion in [6, 8, 27]. The M-stationarity condition was independently introduced in [28, 17, 18, 26], and the letter M refers to Mordukhovich since the proof of M-stationarity being a necessary optimality condition relies on some concepts introduced by Mordukhovich. There also exist some weaker stationarity concepts like A- and C-stationarity, see [5, 22], however, they will not play any role in this paper and are typically viewed as being too weak.

We next define some cones that will appear later.

Definition 2.2. *(a) Let $C \subseteq \mathbb{R}^n$ be a nonempty set. The* polar cone *of C is defined as*

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

(b) Let $C \subseteq \mathbb{R}^n$ be a nonempty closed set and $x^* \in C$. The (Bouligand) tangent cone *(or* contingent cone*) of C in x^* is defined as*

$$\begin{aligned} T_C(x^*) &:= \{d \in \mathbb{R}^n \mid \exists \{x^k\} \to_C x^*, \{t_k\} \downarrow 0 : \frac{x^k - x^*}{t_k} \to d\} \\ &= \{d \in \mathbb{R}^n \mid \exists \{d^k\} \to d, \{t_k\} \downarrow 0 : \ x^* + t_k d^k \in C \ \ \forall k \in \mathbb{N}\}, \end{aligned}$$

where $\{x^k\} \to_C x^*$ denotes a sequence $\{x^k\}$ converging to x^* and satisfying $x^k \in C$ for all $k \in \mathbb{N}$.

(c) Let $C \subseteq \mathbb{R}^n$ be a nonempty closed set and $x^* \in C$. The Fréchet normal cone *of C in x^* is defined as*

$$N_C^F(x^*) := T_C(x^*)^\circ.$$

(d) Let $C \subseteq \mathbb{R}^n$ be a nonempty closed set and $x^* \in C$. The limiting normal cone *of C in x^* is defined as*

$$N_C(x^*) := \{d \in \mathbb{R}^n \mid \exists \{x^k\} \to_C x^*, d^k \in N_C^F(x^k) : \ d^k \to d\}.$$

The following subdifferentials stand in close relation to the cones defined above and will play an important role when it comes to the exactness of our penalty function.

Definition 2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous.*

(a) The Fréchet subdifferential *of f in x^* is defined as*

$$\partial^F f(x^*) := \{s \in \mathbb{R}^n \mid \liminf_{x \to x^*} \frac{f(x) - f(x^*) - s^T(x - x^*)}{\|x - x^*\|} \geq 0\}.$$

(b) The limiting subdifferential *of f in x^* is defined as*

$$\partial f(x^*) := \{s \in \mathbb{R}^n \mid \exists \{x^k\} \to x^*, s^k \in \partial^F f(x^k) : \ s^k \to s\}.$$

We next recall two constraint qualifications that will play a certain role in our discussion. In particular, these CQs will be used in order to show how our new constraint qualifications (to be introduced in the next section) are related to these more standard MPEC-CQs.

Definition 2.4. *Let x^* be feasible for* (1). *Then x^* is said to satisfy*

(a) MPEC-MFCQ *if the gradient vectors*

$$\begin{array}{ll} \nabla h_i(x^*) & \forall i = 1, \ldots, p, \\ \nabla G_i(x^*) & \forall i \in I_{0+}(x^*) \cup I_{00}(x^*), \\ \nabla H_i(x^*) & \forall i \in I_{+0}(x^*) \cup I_{00}(x^*) \end{array} \tag{2}$$

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

$$\begin{array}{rl} \nabla h_i(x^*)^T d = 0 & \forall i = 1, \ldots, p, \\ \nabla G_i(x^*)^T d = 0 & \forall i \in I_{0+}(x^*) \cup I_{00}(x^*), \\ \nabla H_i(x^*)^T d = 0 & \forall i \in I_{+0}(x^*) \cup I_{00}(x^*), \\ \nabla g_i(x^*)^T d < 0 & \forall i \in \mathcal{I}_g(x^*). \end{array} \tag{3}$$

(b) MPEC-ACQ*, if*

$$T_X(x^*) = L_{MPEC}(x^*),$$

where the MPEC-linearized tangent cone $L_{MPEC}(x^*)$ *is defined as*

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

Remark 2.5. Although the above definition of MPEC-MFCQ is the most common one, in our context a different formulation is more useful. Using Motzkin's theorem of the alternative, see, e.g., [14], it is not difficult to see that MPEC-MFCQ is equivalent to the following condition: There is no nonzero multiplier $(\lambda, \mu, \gamma, \nu) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q$ with

$$\lambda_i \geq 0 \ (i \in I_g(x^*)), \ \lambda_i = 0 \ (i \notin I_g(x^*)), \ \gamma_i = 0 \ (i \in I_{+0}(x^*)), \ \nu_i = 0 \ (i \in I_{0+}(x^*))$$

and

$$\sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) - \sum_{i \in I_{0+}(x^*) \cup I_{00}(x^*)} \gamma_i \nabla G_i(x^*) - \sum_{i \in I_{+0}(x^*) \cup I_{00}(x^*)} \nu_i \nabla H_i(x^*) = 0$$

3 A Fritz-John-type Result for MPECs

There exist different ways to obtain first-order optimality conditions for standard nonlinear programs. One is a geometric approach which requires that the tangent cone is equal to a suitable linearized cone (or, at least, that the polar cones of these two sets are identical). Another way is via an exact penalty function P, since the unconstrained first-order optimality condition $0 \in \partial P(x^*)$ for this penalty function can be used to obtain corresponding optimality conditions. A third way is via the Fritz John conditions which do not require any constraint qualifications, but have the disadvantage that they involve a multiplier also in front of the gradient of the objective function. However, under suitable standard CQs (like MFCQ), one can show that this multiplier is nonzero, and then one obtains the usual KKT optimality conditions. In fact, all three approaches end up at the same KKT conditions as first-order optimality conditions. The only difference is that other CQs are required in order to arrive at the KKT conditions.

The situation is different for MPECs. Different approaches lead to different optimality conditions (besides the fact that also different MPEC-tailored CQs are needed). Here, similar to [5, 27, 6], we take the Fritz John-approach. However, it is well-known that a direct application of the Fritz John-conditions to an MPEC is not very effective. Therefore, we are interested in an MPEC-tailored Fritz John condition. The following is the main result of this kind and motivated by similar ideas (for standard nonlinear programs) from [1].

Theorem 3.1. *Let x^* be a local minimum of the MPEC. Then, there are multipliers $\alpha, \lambda, \mu, \gamma, \nu$ such that:*

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

(ii) $\alpha \geq 0$, $\lambda_i \geq 0$ *for all* $i \in I_g(x^*)$, $\lambda_i = 0$ *for all* $i \notin I_g(x^*)$, $\gamma_i = 0$ *for all* $i \in I_{+0}(x^*)$, $\nu_i = 0$ *for all* $i \in I_{0+}(x^*)$ *and either* $\gamma_i > 0, \nu_i > 0$ *or* $\gamma_i\nu_i = 0$ *for all* $i \in I_{00}(x^*)$.

(iii) $\alpha, \lambda, \mu, \gamma, \nu$ *are not all equal to zero.*

(iv) If $\lambda, \mu, \gamma, \nu$ *are not all equal to zero, then there is a sequence* $\{x^k\} \to x^*$ *such that for all* $k \in \mathbb{N}$*:*
$f(x^k) < f(x^*)$,
if $\lambda_i > 0$ *(*$i \in \{1, \dots, m\}$*), then* $\lambda_i g_i(x^k) > 0$,
if $\mu_i \neq 0$ *(*$i \in \{1, \dots, p\}$*), then* $\mu_i h_i(x^k) > 0$,
if $\gamma_i \neq 0$ *(*$i \in \{1, \dots, q\}$*), then* $\gamma_i G_i(x^k) < 0$,
if $\nu_i \neq 0$ *(*$i \in \{1, \dots, q\}$*), then* $\nu_i H_i(x^k) < 0$.

Proof. We first formulate our MPEC (1) equivalently as

$$\begin{aligned} \min_{(x,y,z)} f(x) \quad \text{s.t.} \quad & g(x) \leq 0, \\ & h(x) = 0, \\ & y - G(x) = 0, \\ & z - H(x) = 0, \\ & (x, y, z) \in C, \end{aligned} \tag{4}$$

where the set

$$C := \{(x, y, z) \in \mathbb{R}^{n+q+q} \mid y_i \geq 0, z_i \geq 0, y_i z_i = 0 \text{ for all } i = 1, \dots, q\}$$

is nonempty and closed and we have a local minimum in (x^*, y^*, z^*) with $y^* = G(x^*), z^* = H(x^*)$. Now, we can apply the idea behind Proposition 2.1 from [1]: For each $k \in \mathbb{N}$, consider the penalized problem

$$\min_{x,y,z} F_k(x, y, z) \quad \text{s.t.} \quad (x, y, z) \in S \cap C,$$

where

$$
\begin{aligned}
F_k(x,y,z) &:= f(x) + \frac{k}{2}\sum_{i=1}^{m} \max\{0, g_i(x)\}^2 + \frac{k}{2}\sum_{i=1}^{p} h_i(x)^2 \\
&\quad + \frac{k}{2}\sum_{i=1}^{q} (y_i - G_i(x))^2 + \frac{k}{2}\sum_{i=1}^{q} (z_i - H_i(x))^2 + \frac{1}{2}\|(x,y,z) - (x^*,y^*,z^*)\|_2^2, \\
S &:= \{(x,y,z) \mid \|(x,y,z) - (x^*,y^*,z^*)\|_2 \le \varepsilon\}
\end{aligned}
$$

and $\varepsilon > 0$ is taken in such a way that $f(x) \ge f(x^*)$ for all $(x,y,z) \in S$ that are feasible for the reformulated MPEC (4). Because $S \cap C$ is compact and F_k is continuous, this problem has at least one solution (x^k, y^k, z^k) for all $k \in \mathbb{N}$. Our next step is to show that the sequence $\{(x^k, y^k, z^k)\}$ converges to (x^*, y^*, z^*). To this end, note that

$$
\begin{aligned}
& f(x^k) + \frac{k}{2}\sum_{i=1}^{m} \max\{0, g_i(x^k)\}^2 + \frac{k}{2}\sum_{i=1}^{p} h_i(x^k)^2 \\
& \quad + \frac{k}{2}\sum_{i=1}^{q} (y_i^k - G_i(x^k))^2 + \frac{k}{2}\sum_{i=1}^{q} (z_i^k - H_i(x^k))^2 + \frac{1}{2}\|(x^k,y^k,z^k) - (x^*,y^*,z^*)\|_2^2 \\
= \; & F_k(x^k,y^k,z^k) \;\le\; F_k(x^*,y^*,z^*) \;=\; f(x^*)
\end{aligned}
$$

for all $k \in \mathbb{N}$. Because $S \cap C$ is compact, the sequence $\{f(x^k)\}$ is bounded. This yields

$$
\begin{aligned}
& \lim_{k\to\infty} \max\{0, g_i(x^k)\} = 0 \quad \forall i = 1,\dots,m, \\
& \lim_{k\to\infty} h_i(x^k) = 0 \quad \forall i = 1,\dots,p, \\
& \lim_{k\to\infty} y_i^k - G_i(x^k) = 0 \quad \forall i = 1,\dots,q, \\
& \lim_{k\to\infty} z_i^k - H_i(x^k) = 0 \quad \forall i = 1,\dots,q
\end{aligned}
$$

because otherwise the left-hand side of the inequality above would become unbounded. Therefore, every accumulation point of $\{(x^k, y^k, z^k)\}$ is feasible for the reformulated MPEC (4). The compactness of $S \cap C$ ensures that there is at least one accumulation point. Let $(\bar{x}, \bar{y}, \bar{z})$ be an arbitrary accumulation point of the sequence. Then we know by continuity that

$$
f(\bar{x}) + \frac{1}{2}\|(\bar{x},\bar{y},\bar{z}) - (x^*,y^*,z^*)\|_2^2 \le f(x^*)
$$

and, on the other hand, by the feasibility of $(\bar{x}, \bar{y}, \bar{z})$

$$
f(x^*) \le f(\bar{x}).
$$

Together, this yields $\|(\bar{x},\bar{y},\bar{z}) - (x^*,y^*,z^*)\|_2 = 0$. Thus, the entire sequence $\{(x^k, y^k, z^k)\}$ converges to (x^*, y^*, z^*).

Consequently, we may assume without loss of generality that (x^k, y^k, z^k) is an interior point of S for all $k \in \mathbb{N}$. Then, the standard necessary optimality condition says that

$$
-\nabla F_k(x^k, y^k, z^k) \in N_C^F(x^k, y^k, z^k)
$$

for all $k \in \mathbb{N}$, where the gradient of F_k is given by

$$\begin{aligned}
&-\nabla F_k(x^k, y^k, z^k)\\
&= -\Bigg[\begin{pmatrix} \nabla f(x^k) \\ 0 \\ 0 \end{pmatrix} + \sum_{i=1}^{m} k \max\{0, g_i(x^k)\} \begin{pmatrix} \nabla g_i(x^k) \\ 0 \\ 0 \end{pmatrix} + \sum_{i=1}^{p} k h_i(x^k) \begin{pmatrix} \nabla h_i(x^k) \\ 0 \\ 0 \end{pmatrix}\\
&\quad - \sum_{i=1}^{q} k(y_i^k - G_i(x^k)) \begin{pmatrix} \nabla G_i(x^k) \\ -e_i \\ 0 \end{pmatrix} - \sum_{i=1}^{q} k(z_i^k - H_i(x^k)) \begin{pmatrix} \nabla H_i(x^k) \\ 0 \\ -e_i \end{pmatrix}\\
&\quad + \left(\begin{pmatrix} x^k \\ y^k \\ z^k \end{pmatrix} - \begin{pmatrix} x^* \\ y^* \\ z^* \end{pmatrix} \right) \Bigg]
\end{aligned}$$

and the Frechet normal cone of C in (x^k, y^k, z^k) is easily seen to be given by

$$N_C^F(x^k, y^k, z^k) = \left\{ \begin{pmatrix} 0 \\ \xi \\ \zeta \end{pmatrix} : \begin{array}{ll} \xi_i = 0, \zeta_i \in \mathbb{R} & \text{if } y_i^k > 0 \\ \zeta_i = 0, \xi_i \in \mathbb{R} & \text{if } z_i^k > 0 \\ \xi_i \leq 0, \zeta_i \leq 0 & \text{if } y_i^k = z_i^k = 0 \end{array} \right\}.$$

This yields

$$\begin{aligned}
0 \;=\; & \nabla f(x^k) + \sum_{i=1}^{m} k \max\{0, g_i(x^k)\} \nabla g_i(x^k) + \sum_{i=1}^{p} k h_i(x^k) \nabla h_i(x^k)\\
& - \sum_{i=1}^{q} k(y_i^k - G_i(x^k)) \nabla G_i(x^k) - \sum_{i=1}^{q} k(z_i^k - H_i(x^k)) \nabla H_i(x^k) + (x^k - x^*)
\end{aligned}$$

for all $k \in \mathbb{N}$ and also

$$\begin{array}{rcll}
k(y_i^k - G_i(x^k)) & = & -(y_i^k - y_i^*) & \text{if } \; y_i^k > 0, z_i^k = 0,\\
k(z_i^k - H_i(x^k)) & = & -(z_i^k - z_i^*) & \text{if } \; y_i^k = 0, z_i^k > 0,\\
\left.\begin{array}{r} k(y_i^k - G_i(x^k)) \\ k(z_i^k - H_i(x^k)) \end{array}\right. & \begin{array}{c} \geq \\ \geq \end{array} & \left.\begin{array}{l} -(y_i^k - y_i^*) \\ -(z_i^k - z_i^*) \end{array}\right\} & \text{if } \; y_i^k = z_i^k = 0.
\end{array}$$

Now define the multipliers

$$\delta_k := \sqrt{1 + \sum_{i=1}^{m} (k \max\{0, g_i(x^k)\})^2 + \sum_{i=1}^{p} (k h_i(x^k))^2 + \sum_{i=1}^{q} (k(y_i^k - G_i(x^k)))^2 + \sum_{i=1}^{q} (k(z_i^k - H_i(x^k)))^2}$$

and

$$
\begin{aligned}
\alpha_k &:= \frac{1}{\delta_k}, \\
\lambda_i^k &:= \frac{k \max\{0, g_i(x^k)\}}{\delta_k} \quad \forall i = 1, \dots, m, \\
\mu_i^k &:= \frac{k h_i(x^k)}{\delta_k} \quad \forall i = 1, \dots, p, \\
\gamma_i^k &:= \frac{k(y_i^k - G_i(x^k))}{\delta_k} \quad \forall i = 1, \dots, q, \\
\nu_i^k &:= \frac{k(z_i^k - H_i(x^k))}{\delta_k} \quad \forall i = 1, \dots, q.
\end{aligned}
$$

Because of $\|(\alpha_k, \lambda^k, \mu^k, \gamma^k, \nu^k)\|_2 = 1$ for all $k \in \mathbb{N}$, we may assume without loss of generality that the sequence of multipliers converges to some limit $(\alpha, \lambda, \mu, \gamma, \nu) \neq 0$. Now, we are interested in some properties of this limit. Because of the convergence of $\alpha_k \to \alpha$, we know that the sequence $\{\delta_k\}$ either diverges to $+\infty$ or converges to some positive value (greater or equal to one). We will use this fact later to obtain more information about the signs of γ and ν. By continuity and because of $x^k \to x^*$, we obtain

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

Furthermore, it is easy to see that $\alpha \geq 0$ and $\lambda \geq 0$. Now remember $(y^*, z^*) = (G(x^*), H(x^*))$ and $(x^k, y^k, z^k) \in C$ for all $k \in \mathbb{N}$. If $i \in I_{+0}(x^*)$, this yields $y_i^k > 0, z_i^k = 0$ for all k sufficiently large. Thus, we know

$$
\gamma_i = \lim_{k \to \infty} \frac{k(y_i^k - G_i(x^k))}{\delta_k} = \lim_{k \to \infty} \frac{-(y_i^k - y_i^*)}{\delta_k} = 0
$$

for all $i \in I_{+0}(x^*)$. Analogously, one can prove $\nu_i = 0$ for all $i \in I_{0+}(x^*)$. For $i \in I_{00}(x^*)$ at least one of the following three cases has to occur: If $y_i^k > 0, z_i^k = 0$ for infinitely many k, the same argumentation as above yields $\gamma_i = 0$. Analogously, if $y_i^k = 0, z_i^k > 0$ for infinitely many k, we obtain $\nu_i = 0$. If, however, $y_i^k = z_i^k = 0$ for infinitely many k, we obtain

$$
\begin{aligned}
\gamma_i &= \lim_{k \to \infty} \frac{k(y_i^k - G_i(x^k))}{\delta_k} \geq \lim_{k \to \infty} \frac{-(y_i^k - y_i^*)}{\delta_k} = 0, \\
\nu_i &= \lim_{k \to \infty} \frac{k(z_i^k - H_i(x^k))}{\delta_k} \geq \lim_{k \to \infty} \frac{-(z_i^k - z_i^*)}{\delta_k} = 0
\end{aligned}
$$

So for all $i \in I_{00}(x^*)$ we have either $\gamma_i > 0, \nu_i > 0$ or $\gamma_i \nu_i = 0$.

Finally, let us assume $(\lambda, \mu, \gamma, \nu) \neq 0$. Then $(\lambda^k, \mu^k, \gamma^k, \nu^k) \neq 0$ for all $k \in \mathbb{N}$ sufficiently large. Using the definition of these multipliers, it therefore follows that $(x^k, y^k, z^k) \neq (x^*, y^*, z^*)$ for all k sufficiently large. Consequently, we have

$$
f(x^k) < f(x^k) + \frac{1}{2}\|(x^k, y^k, z^k) - (x^*, y^*, z^*)\|_2^2 \leq f(x^*)
$$

for all $k \in \mathbb{N}$ sufficiently large. Furthermore, we have the following implication for all i and all k sufficiently large:

$$\begin{aligned}
&\lambda_i > 0 \quad \Longrightarrow \quad \lambda_i^k > 0 \quad \Longrightarrow \quad g_i(x^k) > 0 \quad \Longrightarrow \quad \lambda_i g_i(x^k) > 0, \\
&\mu_i \neq 0 \quad \Longrightarrow \quad \mu_i \mu_i^k > 0 \quad \Longrightarrow \quad \mu_i h_i(x^k) > 0.
\end{aligned}$$

Now let $i \in \{1, \ldots, q\}$ be an index with $\gamma_i \neq 0$. This implies $\gamma_i \gamma_i^k > 0$ or equivalently

$$\gamma_i(y_i^k - G_i(x^k)) > 0$$

for all k sufficiently large. We have seen above that if $y_i^k > 0$ for infinitely many k the multiplier γ_i has to be zero. Therefore, in our case $y_i^k = 0$ for all k sufficiently large and consequently

$$\gamma_i G_i(x^k) < 0$$

for all those k. One can prove the implication $\nu_i \neq 0 \Longrightarrow \nu_i H_i(x^k) < 0$ for all k sufficiently large analogously. □

Statements (i)–(iii) of Theorem 3.1 were shown previously in [27] (see also [6]) using a completely different technique of proof based on the limiting co-derivative. Here we improve the result from [27, 6] by showing that statement (iv) also holds. The idea for this proof (and the corresponding statements) is inspired by a corresponding result from [1]. We stress, however, that we did not simply apply the result from [1], but that we exploit the particular structure of the complementarity constraints within our MPEC in order to obtain suitable sign constraints on the multipliers.

Motivated by Theorem 3.1 and the related discussion in [1], we now define MPEC-analogues of some constraint qualifications introduced in [1] for standard nonlinear programs.

Definition 3.2. *A vector $x^* \in X$ is said to satisfy*

(a) MPEC generalized MFCQ*, if there is no multiplier* $(\lambda, \mu, \gamma, \nu) \neq (0, 0, 0, 0)$ *such that*

  (i) $\sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) - \sum_{i=1}^q \gamma_i \nabla G_i(x^*) - \sum_{i=1}^q \nu_i \nabla H_i(x^*) = 0,$

  (ii) $\lambda_i \geq 0$ *for all* $i \in I_g(x^*)$, $\lambda_i = 0$ *for all* $i \notin I_g(x^*)$, $\gamma_i = 0$ *for all* $i \in I_{+0}(x^*)$, $\nu_i = 0$ *for all* $i \in I_{0+}(x^*)$ *and either* $\gamma_i > 0, \nu_i > 0$ *or* $\gamma_i \nu_i = 0$ *for all* $i \in I_{00}(x^*)$,

(b) MPEC generalized pseudonormality*, if there is no multiplier* $(\lambda, \mu, \gamma, \nu)$ *such that*

  (i) $\sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) - \sum_{i=1}^q \gamma_i \nabla G_i(x^*) - \sum_{i=1}^q \nu_i \nabla H_i(x^*) = 0,$

  (ii) $\lambda_i \geq 0$ *for all* $i \in I_g(x^*)$, $\lambda_i = 0$ *for all* $i \notin I_g(x^*)$, $\gamma_i = 0$ *for all* $i \in I_{+0}(x^*)$, $\nu_i = 0$ *for all* $i \in I_{0+}(x^*)$ *and either* $\gamma_i > 0, \nu_i > 0$ *or* $\gamma_i \nu_i = 0$ *for all* $i \in I_{00}(x^*)$,

  (iii) there is a sequence $\{x^k\} \to x^*$ *such that the following is true for all* $k \in \mathbb{N}$*:*

$$\sum_{i=1}^m \lambda_i g_i(x^k) + \sum_{i=1}^p \mu_i h_i(x^k) - \sum_{i=1}^q \gamma_i G_i(x^k) - \sum_{i=1}^q \nu_i H_i(x^k) > 0.$$

(c) MPEC generalized quasinormality*, if there is no multiplier* $(\lambda, \mu, \gamma, \nu)$ *such that*

(i) $\sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) - \sum_{i=1}^q \gamma_i \nabla G_i(x^*) - \sum_{i=1}^q \nu_i \nabla H_i(x^*) = 0,$

(ii) $\lambda_i \geq 0$ *for all* $i \in I_g(x^*)$, $\lambda_i = 0$ *for all* $i \notin I_g(x^*)$, $\gamma_i = 0$ *for all* $i \in I_{+0}(x^*)$, $\nu_i = 0$ *for all* $i \in I_{0+}(x^*)$ *and either* $\gamma_i > 0, \nu_i > 0$ *or* $\gamma_i \nu_i = 0$ *for all* $i \in I_{00}(x^*)$,

(iii) $(\lambda, \mu, \gamma, \nu) \neq (0, 0, 0, 0)$,

(iv) there is a sequence $\{x^k\} \to x^*$ *such that the following is true for all* $k \in \mathbb{N}$*: For all* $\lambda_i > 0$ *we have* $\lambda_i g_i(x^k) > 0$*, for all* $\mu_i \neq 0$ *we have* $\mu_i h_i(x^k) > 0$*, for all* $\gamma_i \neq 0$ *we have* $-\gamma_i G_i(x^k) > 0$*, and for all* $\nu_i \neq 0$ *we have* $-\nu_i H_i(x^k) > 0$.

Obviously, the following implications hold:

$$\begin{aligned} \text{MPEC generalized MFCQ} \quad &\Longrightarrow \quad \text{MPEC generalized pseudonormality} \\ &\Longrightarrow \quad \text{MPEC generalized quasinormality.} \end{aligned}$$

A more indepth discussion of these new constraint qualifications will be given in Section 5 using a result from Section 4.

MPEC generalized MFCQ was already introduced in [27] under a different name, namely NNAMCQ (No Nonzero Abnormal Multiplier Constraint Qualification). The term MPEC GMFCQ, where G stands for generalized, also appears there and it is shown that, although MPEC GMFCQ is defined differently from our MPEC generalized MFCQ, both are equivalent. Furthermore, it is not difficult to see that it is weaker than the standard MPEC-MFCQ condition, see Remark 2.5. Note that MPEC generalized MFCQ is motivated by statements (i)–(iii) of Theorem 3.1 since this CQ guarantees that a local minimum is an M-stationary point for our MPEC, see also Theorem 3.3 below. On the other hand, MPEC generalized quasinormality is motivated by statements (i)–(iv). This CQ can be used to verify the following result.

Theorem 3.3. *Let* x^* *be a local minimum of* (1) *satisfying MPEC generalized quasinormality. Then* x^* *is an M-stationary point of* (1).

Proof. Suppose that x^* is a local minimum of our MPEC. Then Theorem 3.1 implies the existence of multipliers $\alpha, \lambda, \mu, \gamma, \nu$ such that statements (i)–(iv) of that result hold. Assume that $\alpha = 0$. Then the MPEC generalized quasinormality condition implies that $\lambda = \mu = \gamma = \nu = 0$, contradicting the fact that not all multipliers are zero. Hence $\alpha > 0$, and we may assume without loss of generality that $\alpha = 1$, showing that x^* is indeed an M-stationary point. □

Note that the proof of Theorem 3.3 is completely elementary and does not assume any knowledge of the limiting subdifferential or the limiting co-derivative by Mordukhovich, in contrast to all previous results known to the authors where it is also shown that a local minimum is an M-stationary point under suitable CQs.

In the moment, it is not clear why we also introduced the MPEC generalized pseudonormality condition since, so far, it is not really used anywhere. In the following section, however, we show that this CQ plays a fundamental role in order to prove an exact penalty result.

4 Exact Penalization of MPECs

Exact penalty results for MPECs are known in the literature, see [12, 13, 29, 23, 11] for example. In particular, it is known that MPEC-MFCQ implies exactness of a certain penalty function that will also appear in our context as a side-product. Some authors also use a partial penalization only, in fact, they sometimes penalize the standard constraints only, whereas the complementarity constraints are left as constraints (see [11]). We believe that this is not the right way to consider penalty functions for MPECs. In fact, the main idea of penalty methods is to get rid of possibly difficult constraints. So one should at least penalize the (complicated) complementarity constraints. For the sake of simplicity, we penalize all constraints in this section and obtain an exactness result for our penalty function under MPEC generalized pseudonormality which, we recall, is weaker than the usual MPEC-MFCQ condition.

In order to derive the exactness result, let us first rewrite our MPEC equivalently as

$$\min f(x) \qquad \text{s.t.} \qquad F(x) \in \Lambda, \tag{5}$$

where

$$F(x) := \begin{pmatrix} g_i(x)_{i=1,\dots,m} \\ h_i(x)_{i=1,\dots,p} \\ \begin{pmatrix} G_i(x) \\ H_i(x) \end{pmatrix}_{i=1,\dots,q} \end{pmatrix}$$

and

$$\Lambda := \begin{pmatrix} (-\infty, 0]^m \\ \{0\}^p \\ C^q \end{pmatrix}$$

with

$$C := \{(a, b) \in \mathbb{R}^2 \mid a \geq 0, b \geq 0, ab = 0\}.$$

The penalty function associated to (5) is

$$P_\alpha(x) := f(x) + \alpha \mathrm{dist}_\Lambda(F(x)). \tag{6}$$

Here, the distance function is defined by

$$\mathrm{dist}_\Lambda(F(x)) = \inf\{\|y - F(x)\| \mid y \in \Lambda\}, \tag{7}$$

where, in principle, the norm can be chosen arbitrarily. Our goal is to prove that the penalty function (6) is *exact* in a local minimum x^* satisfying a suitable constraint qualification, i.e. that there is a finite $\bar{\alpha} \geq 0$ such that x^* is an unconstrained local minimum of $P_\alpha(x)$ for all $\alpha \geq \bar{\alpha}$. It is well-known that exactness of this function using a specific norm implies exactness for all other norms as well. Therefore, we will restrict ourselves to the l_1-norm. In this case, $P_\alpha(x)$ is of the form

$$P_\alpha(x) = f(x) + \alpha \left[\sum_{i=1}^m \mathrm{dist}_{(-\infty,0]}(g_i(x)) + \sum_{i=1}^p \mathrm{dist}_{\{0\}}(h_i(x)) + \sum_{i=1}^q \mathrm{dist}_C((G_i(x), H_i(x))) \right] \tag{8}$$

and elementary calculations lead to the following explicit formulas for the corresponding distance functions.

Lemma 4.1. *Under the l_1-norm, the distance functions are given by the following expressions for $a, b \in \mathbb{R}$:*

$$
\begin{aligned}
dist_{(-\infty,0]}(a) &= \max\{a, 0\}, \\
dist_{\{0\}}(a) &= |a|, \\
dist_C((a,b)) &= \max\{-a, -b, -(a+b), \min\{a,b\}\} = \begin{cases} a \text{ or } b & a = b \geq 0, \\ b & a > b > 0, \\ -b & a > 0, b \leq 0, \\ -(a+b) & a \leq 0, b \leq 0, \\ -a & a \leq 0, b > 0, \\ a & b > a > 0. \end{cases}
\end{aligned}
$$

It follows that the penalty function we consider in this section is explicitly given by

$$
\begin{aligned}
P_\alpha(x) = f(x) + \alpha \Bigg[&\sum_{i=1}^{m} \max\{0, g_i(x)\} + \sum_{i=1}^{p} |h_i(x)| + \\
&\sum_{i=1}^{q} \max\{ -G_i(x), -H_i(x), -(G_i(x) + H_i(x)), \min\{G_i(x), H_i(x)\}\} \Bigg],
\end{aligned}
\tag{9}
$$

see also the discussion at the end of this section for the relation between this penalty function and another one which is used more frequently in the context of MPECs.

In order to prove the exactness result, we need to calculate the limiting subdifferentials of the distance functions stated in Lemma 4.1.

Lemma 4.2. *The limiting subdifferentials of the distance functions from Lemma 4.1 (recall that we use the l_1-norm here) are given by*

$$
\begin{aligned}
\partial dist_{(-\infty,0]}(a) &= \begin{cases} \{0\} & a < 0, \\ [0,1] & a = 0, \\ \{1\} & a > 0, \end{cases} \\
\partial dist_{\{0\}}(a) &= \begin{cases} \{-1\} & a < 0, \\ [-1,1] & a = 0, \\ \{1\} & a > 0, \end{cases}
\end{aligned}
$$

$$\partial dist_C((a,b)) = \begin{cases} \{(\xi,0)^T,(0,\zeta)^T \mid \xi,\zeta \in [0,1]\} \cup \{(\xi,\zeta)^T \mid \xi,\zeta \in [-1,0]\} & a=b=0, \\ \{(1,0)^T,(0,1)^T)\} & a=b>0, \\ \{(0,1)^T\} & a>b>0, \\ \{(0,\zeta)^T \mid \zeta \in [-1,1]\} & a>0, b=0, \\ \{(0,-1)^T\} & a>0, b<0, \\ \{(\xi,-1)^T \mid \xi \in [-1,0]\} & a=0, b<0, \\ \{(-1,-1)^T\} & a<0, b<0, \\ \{(-1,\zeta)^T \mid \zeta \in [-1,0]\} & a<0, b=0, \\ \{(-1,0)^T\} & a<0, b>0, \\ \{(\xi,0)^T \mid \xi \in [-1,1]\} & a=0, b>0, \\ \{(1,0)^T\} & b>a>0. \end{cases}$$

Proof. For convex functions, both the Fréchet and the limiting subdifferential coincide with the standard subdifferential from convex analysis, cf. [21]. This gives the expressions for the first two distance functions.

In order to get the expression for the limiting subdifferential of the third distance function $(a,b) \mapsto \mathrm{dist}_C(a,b)$, we also recall that the Fréchet and the limiting subdifferentials of a locally continuously differentiable function are equal to a single set, consisting of the gradient of that function, cf. [21, Example 8.8]. Together with the previous comment regarding (locally) convex functions, we obtain all statements except for the first two cases $a=b=0$ and $a=b>0$ which we will now treat separately.

First consider the case $a=b>0$. We claim that the Fréchet subdifferential is empty. To see this, assume there exists an element $s=(s_1,s_2) \in \partial^F \mathrm{dist}_C(a,b)$. Then consider the particular sequence $\{(a^k,b^k)\}$ with $(a^k,b^k) := (a+\frac{1}{k}, b)$. An elementary calculation then shows that

$$\frac{\mathrm{dist}_C(a^k,b^k) - \mathrm{dist}_C(a,b) - s^T \binom{a^k-a}{b^k-b}}{\left\| \binom{a^k-a}{b^k-b} \right\|} = \frac{b-b-\frac{1}{k}s_1}{\frac{1}{k}} = -s_1,$$

hence the limit inferior of this expression is nonnegative if and only if $s_1 \leq 0$. On the other hand, consider the particular sequence $\{(a^k,b^k)\}$ with $(a^k,b^k) := (a-\frac{1}{k}, b)$. Again, a simple calculation gives

$$\frac{\mathrm{dist}_C(a^k,b^k) - \mathrm{dist}_C(a,b) - s^T \binom{a^k-a}{b^k-b}}{\left\| \binom{a^k-a}{b^k-b} \right\|} = \frac{a-\frac{1}{k}-a+\frac{1}{k}s_1}{\frac{1}{k}} = -1+s_1,$$

and the limit inferior of this term is nonnegative if and only if $s_1 \geq 1$. This contradiction shows that s cannot belong to the Fréchet subdifferential. Hence, to obtain the elements s of the limiting subdifferential $\partial \mathrm{dist}_C(a,b)$, we only need to consider sequences $\{s^k\}$ converging to s with s^k being an element of the Fréchet subdifferential $\partial^F \mathrm{dist}_C(a^k,b^k)$ at points (a^k,b^k) satisfying $a^k,b^k > 0$ and $a^k \neq b^k$. The corresponding expressions were already calculated and show that the limiting subdifferential consists of the two vectors $(0,1)^T$ and $(1,0)^T$.

Finally, consider the case $a=b=0$. First, let us calculate the Fréchet subdifferential at this point. We claim that this Fréchet subdifferential is given by the rectangle $[-1,0] \times [-1,0]$. To see this, note that it is not difficult to see that the numerator

$$\mathrm{dist}_C(a^k, b^k) - \mathrm{dist}_C(0,0) - (s_1, s_2)\begin{pmatrix} a^k - 0 \\ b^k - 0 \end{pmatrix}$$

occuring in the definition of the Fréchet subdifferential is always nonnegative for all $(s_1, s_2) \in [-1,0] \times [-1,0]$, so these elements certainly belong to $\partial^F \mathrm{dist}_C(0,0)$. On the other hand, there cannot exist any other elements since, by taking the particular sequence $\{(a^k, b^k)\}$ with $(a^k, b^k) := (\frac{1}{k}, 0)$, we obtain

$$\frac{\mathrm{dist}_C(a^k, b^k) - \mathrm{dist}_C(0,0) - (s_1, s_2)\binom{a^k-0}{b^k-0}}{\left\|\binom{a^k-0}{b^k-0}\right\|} = \frac{0 - 0 - \frac{1}{k}s_1}{\frac{1}{k}} = -s_1,$$

whereas for the particular sequence $\{(a^k, b^k)\}$ with $(a^k, b^k) := (-\frac{1}{k}, 0)$, we get

$$\frac{\mathrm{dist}_C(a^k, b^k) - \mathrm{dist}_C(0,0) - (s_1, s_2)\binom{a^k-0}{b^k-0}}{\left\|\binom{a^k-0}{b^k-0}\right\|} = \frac{\frac{1}{k} - 0 + \frac{1}{k}s_1}{\frac{1}{k}} = 1 + s_1,$$

so that the definition of the Fréchet subdifferential shows that we necessarily have $s_1 \in [-1,0]$. A symmetric argument shows that also $s_2 \in [-1,0]$ is necessary for the vector $s = (s_1, s_2)$ belonging to $\partial^F \mathrm{dist}_C(0,0)$. Altogether, we therefore have $\partial^F \mathrm{dist}_C(0,0) = [-1,0] \times [-1,0]$.

Since the Fréchet subdifferential is a subset of the limiting subdifferential, it follows that $[-1,0] \times [-1,0] \in \partial \mathrm{dist}_C(0,0)$. The other elements $s \in \partial \mathrm{dist}_C(0,0)$ can be easily obtained by taking sequences $s^k \to s$ with $s^k \in \partial^F \mathrm{dist}_C(a^k, b^k)$ with $(a^k, b^k) \to (0,0)$ and $(a^k, b^k) \neq (0,0)$ together with the already known expressions for the Fréchet subdifferentials in these points. □

To prove the central result of this section, we will proceed in three steps. First, we need an auxiliary result, then we will prove that MPEC generalized pseudonormality implies the existence of local error bounds and, finally, we will use this fact to obtain exactness of our penalty function. Remember that we use the l_1-norm to measure distances.

Lemma 4.3. *Let x^* be feasible for* (1) *such that MPEC generalized pseudonormality holds in x^*. Then there are $\delta, c > 0$ such that for all $x \in B_\delta(x^*)$ with $x \notin X$ and all $\xi \in \partial dist_\Lambda F(x)$ the following estimate holds:*

$$\|\xi\|_1 \geq \frac{1}{c}.$$

Proof. Assume that the statement is wrong. Then, one can find a sequence $\{x^k\} \to x^*$ with $x^k \notin X$ and $\xi^k \in \partial \mathrm{dist}_\Lambda F(x^k)$ for all $k \in \mathbb{N}$ such that $\|\xi^k\|_1 \to 0$. To calculate $\partial \mathrm{dist}_\Lambda F(x^k)$ we may apply the sum rule from [2, Theorem 6.4.4] because the distance functions are Lipschitz continuous. Furthermore, we can use the chain rule stated in [2, p. 151], again because of the Lipschitz continuity of distance functions. This yields the existence of multipliers

$$\begin{aligned} \lambda_i^k &\in \partial \mathrm{dist}_{(-\infty,0]}(g_i(x^k)) \quad \forall\, i = 1,\ldots,m, \\ \mu_i^k &\in \partial \mathrm{dist}_{\{0\}}(h_i(x^k)) \quad \forall\, i = 1,\ldots,p, \\ (\gamma_i^k, \nu_i^k) &\in -\partial \mathrm{dist}_C((G_i(x^k), H_i(x^k))) \quad \forall\, i = 1,\ldots,q \end{aligned}$$

such that

$$\xi^k = \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) - \sum_{i=1}^{q} (\gamma_i^k \nabla G_i(x^k) + \nu_i^k \nabla H_i(x^k)) \tag{10}$$

for all $k \in \mathbb{N}$. Using Lemma 4.2, it is easy to see that the sequence $\{(\lambda^k, \mu^k, \gamma^k, \nu^k)\}$ is bounded. Hence, we may assume without loss of generality that it converges to some limit $(\lambda, \mu, \gamma, \nu)$. Taking the limit $k \to \infty$ in (10) and using the smoothness of g, h, G, H then yields

$$0 = \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) - \sum_{i=1}^{q} (\gamma_i \nabla G_i(x^*) + \nu_i \nabla H_i(x^*))$$

Furthermore, Lemma 4.2 yields

$$\begin{aligned} \lambda_i &\geq 0 & \forall i &= 1, \ldots, m, \\ \lambda_i &= 0 & \forall i &\notin I_g(x^*), \\ \gamma_i &= 0 & \forall i &\in I_{+0}(x^*), \\ \nu_i &= 0 & \forall i &\in I_{0+}(x^*), \\ \gamma_i \nu_i = 0 \text{ or } \gamma_i > 0, \nu_i &> 0 & \forall i &\in I_{00}(x^*). \end{aligned}$$

Additionally, it is easy to see that, for all $k \in \mathbb{N}$, we have

$$\begin{aligned} \lambda_i g_i(x^k) &\geq 0 & \forall i &= 1, \ldots, m, \\ \mu_i h_i(x^k) &\geq 0 & \forall i &= 1, \ldots, p, \\ -\gamma_i G_i(x^k) &\geq 0 & \forall i &= 1, \ldots, q, \\ -\nu_i H_i(x^k) &\geq 0 & \forall i &= 1, \ldots, q. \end{aligned}$$

Because of $x^k \notin X$ for all $k \in \mathbb{N}$, at least one constraint has to be violated infinitely many times. Using Lemma 4.2, it is easy to see that the corresponding product is strictly positive for all $k \in \mathbb{N}$ such that the constraint is violated, i.e. if the constraint $g_i(x^k) \leq 0$ is violated for infinitely many k we have $\lambda_i g_i(x^k) > 0$ for all those k, if the constraint $h_i(x^k) = 0$ is violated for infinitely many k we have $\mu_i h_i(x^k) > 0$ for all those k and finally, if the constraint $(G_i(x^k), H_i(x^k)) \in C$ is violated for infinitely many k we have $-(\gamma_i G_i(x^k) + \nu_i H_i(x^k)) > 0$ for all those k. This yields

$$\sum_{i=1}^{m} \lambda_i g_i(x^k) + \sum_{i=1}^{p} \mu_i h_i(x^k) - \sum_{i=1}^{q} (\gamma_i G_i(x^k) + \nu_i H_i(x^k)) > 0$$

at least on a subsequence $K \subseteq \mathbb{N}$. This, however, implies that MPEC generalized pseudonormality is violated in x^*, a contradiction. □

The following result about local error bounds is based on [24, Theorem 2.2], where a more general setting was considered. Hence, the proof for our special case is easier and therefore stated here. Again, distances are measured in the l_1-norm.

Lemma 4.4. *Let x^* be feasible for* (1) *and $\delta, c > 0$ such that $\|\xi\|_1 \geq \frac{1}{c}$ for all $x \in B_\delta(x^*)\backslash X$ and all $\xi \in \partial dist_\Lambda F(x)$. Then the following estimate holds for all $x \in B_{\frac{\delta}{2}}(x^*)$:*

$$dist_X(x) \leq nc \; dist_\Lambda F(x).$$

Proof. Assume that the statement is wrong. Then, there is an $\tilde{x} \in B_{\frac{\delta}{2}}(x^*)$ with

$$\mathrm{dist}_X(\tilde{x}) > nc\,\mathrm{dist}_\Lambda F(\tilde{x}).$$

Obviously, this implies $\tilde{x} \notin X$. Furthermore, we can choose a $t > 1$ such that

$$d := tnc\,\mathrm{dist}_\Lambda F(\tilde{x}) < \mathrm{dist}_X(\tilde{x}). \tag{11}$$

Because of $\tilde{x} \notin X$, we have $d > 0$. Furthermore, $\mathrm{dist}_\Lambda F(\tilde{x}) = \frac{d}{tnc}$ and thus

$$\mathrm{dist}_\Lambda F(\tilde{x}) \leq \inf_{x \in \mathbb{R}^n} \mathrm{dist}_\Lambda F(x) + \frac{d}{tnc}.$$

Application of Ekeland's variational principle [3, Theorem 7.5.1] to the continuous nonnegative function $x \mapsto \mathrm{dist}_\Lambda F(x)$ with $\varepsilon = \frac{d}{tnc}$ and $\lambda = d$ yields the existence of an $\bar{x}$ with the following properties:

$$\|\bar{x} - \tilde{x}\|_1 \leq d, \tag{12}$$

$$\mathrm{dist}_\Lambda F(\bar{x}) \leq \mathrm{dist}_\Lambda F(\tilde{x}), \tag{13}$$

$$\mathrm{dist}_\Lambda F(x) + \frac{1}{tnc}\|x - \bar{x}\|_1 > \mathrm{dist}_\Lambda F(\bar{x}) \quad \forall x \in \mathbb{R}^n, x \neq \bar{x}. \tag{14}$$

Equations (12) and (11) imply

$$\|\bar{x} - \tilde{x}\|_1 \leq d < \mathrm{dist}_X(\tilde{x}),$$

thus $\bar{x} \notin X$. According to (14), the function $x \mapsto \mathrm{dist}_\Lambda F(x) + \frac{1}{tnc}\|x - \bar{x}\|_1$ attains a global minimum in $\bar{x}$. Thus, by Fermat's rule, we have

$$0 \in \partial \left[\mathrm{dist}_\Lambda F(x) + \frac{1}{tnc}\|x - \bar{x}\|_1\right]_{x=\bar{x}}.$$

To calculate this subdifferential, we may invoke the sum rule from [2, Theorem 6.4.4] because distance functions are Lipschitz continuous. Furthermore, it is easy to see that the limiting subdifferential of the convex function $x \mapsto \frac{1}{tnc}\|x - \bar{x}\|_1$ in $\bar{x}$ is given by

$$\partial \left[\frac{1}{tnc}\|\cdot - \bar{x}\|_1\right](\bar{x}) = \frac{1}{tnc}\{\zeta \in \mathbb{R}^n \mid \|\zeta\|_\infty \leq 1\}.$$

Hence, we can find $\xi \in \partial \mathrm{dist}_\Lambda F(\bar{x})$ and ζ with $\|\zeta\|_\infty \leq 1$ such that

$$0 = \|\xi + \frac{1}{tnc}\zeta\|_1 \geq \|\xi\|_1 - \frac{1}{tnc}\|\zeta\|_1 \geq \|\xi\|_1 - \frac{1}{tnc}n,$$

consequently

$$\|\xi\|_1 \leq \frac{1}{tc} < \frac{1}{c}.$$

On the other hand, we have $\bar{x} \notin X$ and, using equations (12), (11) together with $x^* \in X$,

$$\|\bar{x} - x^*\|_1 \leq \|\bar{x} - \tilde{x}\|_1 + \|\tilde{x} - x^*\|_1 \leq d + \|\tilde{x} - x^*\|_1 < \mathrm{dist}_X(\tilde{x}) + \|\tilde{x} - x^*\|_1 \leq 2\|\tilde{x} - x^*\|_1 \leq 2\frac{\delta}{2} = \delta.$$

This, however, is a contradiction to our assumptions. □

Taking into account the previous two results, we can now follow [10] and get an exact penalty result for MPECs.

Theorem 4.5. *Let x^* be a local minimizer of* (1) *with f locally Lipschitz-continuous around x^* with modulus $L > 0$. If MPEC generalized pseudonormality holds in x^*, then the penalty function P_α defined in* (9) *is exact in x^*.*

Proof. According to Lemma 4.3 and Lemma 4.4, we can find constants $\delta, c > 0$ such that

$$\mathrm{dist}_X(x) \leq c\, \mathrm{dist}_\Lambda F(x)$$

for all $x \in B_\delta(x^*)$ (note that we redefined the constants δ, c from Lemma 4.4 to shorten the notation). Now choose $\varepsilon > 0$ such that $2\varepsilon < \delta$ and that f attains a global minimum in x^* on $B_{2\varepsilon}(x^*) \cap X$. Furthermore, we assume without loss of generality that L is the Lipschitz constant of f in $B_{2\varepsilon}(x^*)$. Then the following holds for every $x \in B_\varepsilon(x^*)$: Choose $x^p \in \mathrm{Proj}_X(x)$ arbitrary. This implies

$$\|x^p - x\|_1 \leq \|x^* - x\|_1 \leq \varepsilon \quad \Longrightarrow \quad \|x^p - x^*\|_1 \leq \|x^p - x\|_1 + \|x - x^*\|_1 \leq 2\varepsilon$$

and consequently we have

$$\begin{aligned} f(x^*) &\leq f(x^p) \\ &\leq f(x) + L\|x^p - x\|_1 \\ &= f(x) + L\,\mathrm{dist}_X(x) \\ &\leq f(x) + cL\,\mathrm{dist}_\Lambda F(x). \end{aligned}$$

Thus, the penalty function P_α is exact with $\bar{\alpha} = cL$. □

In the proof above we only used that f is locally Lipschitz-continuous around x^*, so it is not necessary to demand f to be smooth. On the other hand, mere continuity of f is not enough to guarantee exactness as the following example illustrates:

Example 4.6 Consider the following 2-dimensional example with linear constraints

$$\begin{aligned} \min f(x) \quad \text{s.t.} \quad & g(x) := x_2 \leq 0, \\ & G(x) := x_1 \geq 0, \\ & H(x) := x_1 + x_2 \geq 0, \\ & G(x)H(x) = x_1(x_1 + x_2) = 0, \end{aligned}$$

where f will be specified later. The feasible set is obviously given by

$$X = \{(t, -t) \in \mathbb{R}^2 \mid t \geq 0\},$$

hence $x^* := (0, 0)$ is feasible. Obviously, MPEC-MFCQ is violated in x^* as there are multipliers $(\lambda, \gamma, \nu) \neq (0, 0, 0)$ with $\lambda \geq 0$ such that

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

On the other hand, it is easy to see that MPEC generalized MFCQ is satisfied. Thus, we know that the penalty function P_α is exact for every locally Lipschitz continuous function f. Now let us consider the objective function

$$f(x) := -\sqrt{|x_1 + x_2|}$$

which is well defined and continuous in $\mathbb{R}^2$ but not locally Lipschitz continuous around x^*. Note that x^* is a local minimizer of f over X. However, x^* is not a local minimizer of P_α for any $\alpha > 0$. Evaluation of P_α at $x^k := (\frac{1}{k^2}, 0)$, $k \in \mathbb{N}$, for example yields

$$P_\alpha(x^k) = \frac{1}{k}\left(-1 + \alpha\frac{1}{k}\right),$$

which eventually becomes negative for all $\alpha > 0$. Hence, x^* with $P_\alpha(x^*) = 0$ is not a local minimizer of P_α for any $\alpha > 0$ or, equivalently, P_α is not exact in x^*. ◇

Theorem 4.5 is particularly interesting, because it also works for nonstrict local minima x^*. A similar exact penalty result based on pseudonormality can be found in [1, Proposition 4.2], however, that result requires x^* to be a strict local minimum, and it is stated in [1, Example 7.7] that this assumption might be crucial. In our case, we do not need a strict local minimum to guarantee exactness. We stress, however, that our technique of proof is also completely different from the one in [1], and it is not clear whether this technique can also be used to improve the result from [1].

We recall that (9) gives the explicit representation of the penalty function used within this section. Another popular penalty function that is typically taken by the authors in the MPEC-setting, see [23], takes into account the equivalence

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

so that it is a natural idea to add the absolute value of the min-function to our penalty term, resulting into the mapping

$$\tilde{P}_\alpha(x) := f(x) + \alpha\left[\sum_{i=1}^m \max\{0, g_i(x)\} + \sum_{i=1}^p |h_i(x)| + \sum_{i=1}^q \Big|\min\{G_i(x), H_i(x)\}\Big|\right]. \tag{15}$$

As a consequence of our previous results, we may also obtain an exact penalty result for $\tilde{P}_\alpha$. To this end, let us reconsider the original penalty function from (7). Using the l_1-norm, we then obtain the expression (8) for this distance-based penalty function. Using once again (for the sake of consistency) the l_1-norm to calculate the distances for all the terms that occur in (8) (cf. Lemma 4.1), we end up with the representation from (9). However, we could alternatively calculate the distances for each term using the l_∞-norm. Then it is not difficult to see that the last expression in Lemma 4.1 becomes

$$\mathrm{dist}_C(a, b) = \Big|\min\{a, b\}\Big|,$$

i.e., also this mapping may be viewed as a distance function. Taking into account that all norms are equivalent in finite dimensions, we immediately see that $\tilde{P}_\alpha$ is also an exact penalty function under the assumption of Theorem 4.5. This proves the following result.

Corollary 4.7. *Let x^* be a local minimum of* (1) *such that MPEC generalized pseudonormality holds in x^*. Then the penalty function $\tilde{P}_\alpha(x)$ from* (15) *is exact in x^*.*

5 Relations between Old and New CQs

In this section, we want to show how our new MPEC constraint qualifications are related to some existing ones. We already know that MPEC-MFCQ implies each of our new CQs. On the other hand, it is clear that our conditions cannot be weaker than MPEC-ACQ since the latter condition is not enough to imply exactness of penalty functions. In this way, one gets the idea that our conditions are somewhere between MPEC-MFCQ and MPEC-ACQ. In fact, we will show that the MPEC generalized pseudonormality condition is strictly between MPEC-MFCQ and MPEC-ACQ.

To this end, we consider once again the MPEC in the abstract formulation from the previous section, namely

$$\min\ f(x) \quad \text{s.t.} \quad F(x) \in \Lambda,$$

where f is locally Lipschitz-continuous and F is smooth.

The constraint qualification most commonly used in the context of exact penalty functions is MPEC-MFCQ. The reason for this is as follows: According to our proof of Theorem 4.5, the existence of local error bounds, i.e. the existence of constants $\delta, c > 0$ such that

$$\mathrm{dist}_{F^{-1}(\Lambda)}(x) \leq c\,\mathrm{dist}_{\Lambda} F(x) \quad \forall x \in B_\delta(x^*)$$

is a sufficient condition for exactness. It was shown in [9, Corollary 1] that the existence of local error bounds is equivalent to the calmness of the perturbation map

$$M(r) := \{x \in \mathbb{R}^n \mid F(x) + r \in \Lambda\}$$

at $(0, x^*)$, where calmness of a multifunction is defined as follows.

Definition 5.1. *Let $\Phi : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ be a multifunction with closed graph and $(u, v) \in gph(\Phi)$. Then we say that Φ is calm at (u, v) if there are neigbourhoods U of u, V of v, and a modulus $L \geq 0$ such that*

$$\phi(u') \cap V \subseteq \Phi(u) + L\|u' - u\|B_1(0) \quad \forall u' \in U.$$

Furthermore, it is well-known that the following condition

$$\left.\begin{array}{l} \nabla F(x^*)^T \lambda = 0, \\ \lambda \in N_\Lambda(F(x^*)) \end{array}\right\} \Longrightarrow \lambda = 0 \tag{16}$$

guarantees calmness of M in $(0, x^*)$, cf. for example [8, Proposition 3.8], and thus exactness of P_α. In the next result, we want to relate this condition to one of our constraint qualifications.

Lemma 5.2. *Condition* (16) *is equivalent to MPEC generalized MFCQ.*

Proof. Due to [21, Proposition 6.41], we may rewrite the limiting normal cone as

$$N_\Lambda(F(x^*)) = \bigtimes_{i=1}^{m} N_{(-\infty,0]}(g_i(x^*)) \times \bigtimes_{i=1}^{p} N_{\{0\}}(h_i(x^*)) \times \bigtimes_{i=1}^{q} N_C(G_i(x^*), H_i(x^*)).$$

Hence, condition (16) is equivalent to

$$\left.\begin{array}{l} \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) - \sum_{i=1}^q \left(\gamma_i \nabla G_i(x^*) + \nu_i \nabla H_i(x^*)\right) = 0, \\ \lambda_i \in N_{(-\infty,0]}(g_i(x^*)) \quad \forall i = 1, \dots, m, \\ \mu_i \in N_{\{0\}}(h_i(x^*)) \quad \forall i = 1, \dots, p, \\ (\gamma_i, \nu_i) \in -N_C(G_i(x^*), H_i(x^*)) \quad \forall i = 1, \dots, q, \end{array}\right\} \Longrightarrow (\lambda, \mu, \gamma, \nu) = 0,$$

which is exactly MPEC generalized MFCQ. $\square$

Because MPEC-MFCQ implies MPEC generalized MFCQ, MPEC-MFCQ also is a sufficient condition for exactness of P_α. However, recall from Example 4.6 that MPEC-MFCQ is strictly stronger than MPEC generalized MFCQ. Moreover, the following counterexample shows that also MPEC generalized MFCQ is strictly stronger than MPEC generalized pseudonormality. Thus, MPEC-MFCQ is a sufficient condition for exactness of P_α, but it is by far too restrictive.

Example 5.3 Consider the 2-dimensional minimization problem

$$\begin{aligned} \min f(x) \quad \text{s.t.} \quad & g(x) := x_1 + x_2 \leq 0, \\ & G(x) := x_1 \geq 0, \\ & H(x) := x_2 \geq 0, \\ & G(x)H(x) = x_1 x_2 = 0. \end{aligned}$$

The origin $x^* = (0,0)$ is feasible, and all constraints are active at x^*. To prove that MPEC generalized MFCQ is violated, we have to find $(\lambda, \gamma, \nu) \neq 0$ such that $\lambda \geq 0$, either $\gamma\nu = 0$ or $\gamma, \nu > 0$ and

$$\lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \gamma \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \nu \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Obviously, all vectors with this properties are of the form $(\lambda, \gamma, \nu) = c(1,1,1)$ with $c > 0$. Hence, MPEC generalized MFCQ is violated. MPEC generalized pseudonormality, on the other hand, is satisfied, because we have

$$\lambda g(x^k) - \gamma G(x^k) - \nu H(x^k) = c(x_1^k + x_2^k) - cx_1^k - cx_2^k = 0$$

for all sequences $x^k \to x^*$. $\diamond$

This example illustrates that MPEC generalized pseudonormality is indeed weaker than those constraint qualifications commonly used to guarantee exactness of P_α.

On the other hand, it is well-known that Abadie-CQ is not strong enough to guarantee exactness, cf. [1, Example 7.3]. Thus, it is not surprising that MPEC generalized pseudonormality is strictly stronger than MPEC-ACQ. To see this, we first need a technical result concerning the tangent cone. The proof of this result is rather straightforward, nevertheless, we stress that it is not a priori clear that this result holds since the set C is not regular in the sense of [21], see, in particular, Proposition 6.41 and the subsequent discussion in that reference.

Lemma 5.4. *Let x^* be feasible for* (1). *Then the tangent cone is given by*

$$T_\Lambda(F(x^*)) = \bigtimes_{i=1}^{m} T_{(-\infty,0]}(g_i(x^*)) \times \bigtimes_{i=1}^{p} T_{\{0\}}(h_i(x^*)) \times \bigtimes_{i=1}^{q} T_C(G_i(x^*), H_i(x^*)).$$

Proof. The inclusion "$\subseteq$" follows directly from [21, Proposition 6.41]. To prove the inclusion "$\supseteq$" consider arbitrary elements $d_{g_i} \in T_{(-\infty,0]}(g_i(x^*))$, $d_{h_i} \in T_{\{0\}}(h_i(x^*))$ and $(d_{G_i}, d_{H_i}) \in T_C(G_i(x^*), H_i(x^*))$, and define

$$d := (d_{g_i, i=1,\dots,m}, d_{h_i, i=1,\dots,p}, (d_{G_i}, d_{H_i})_{i=1,\dots,q}).$$

According to the definition of the tangent cone, there are sequences

$$\begin{aligned} d_{g_i}^k \to d_{g_i}, t_{g_i}^k \downarrow 0 \quad &\text{with} \quad g_i(x^*) + t_{g_i}^k d_{g_i}^k \leq 0, \\ d_{h_i}^k \to d_{h_i}, t_{h_i}^k \downarrow 0 \quad &\text{with} \quad h_i(x^*) + t_{h_i}^k d_{h_i}^k = 0, \\ (d_{G_i}^k, d_{H_i}^k) \to (d_{G_i}, d_{H_i}), t_{GH_i}^k \downarrow 0 \quad &\text{with} \quad 0 \leq G_i(x^*) + t_{GH_i}^k d_{G_i}^k \perp H_i(x^*) + t_{GH_i}^k d_{H_i}^k \geq 0 \end{aligned}$$

for all $k \in \mathbb{N}$. Consequently, we have

$$d^k := (d_{g_i,i=1,\dots,m}^k, d_{h_i,i=1,\dots,p}^k, (d_{G_i}^k, d_{H_i}^k)_{i=1,\dots,q}) \to d.$$

To prove $d \in T_\Lambda(F(x^*))$ it suffices to find a sequence $t^k \downarrow 0$ such that $F(x^*) + t^k d^k \in \Lambda$ for all $k \in \mathbb{N}$. Define

$$t^k := \min\{t_{g_i,i=1,\dots,m}^k, t_{GH_i,i=1,\dots,q}^k\}$$

for all $k \in \mathbb{N}$. Then we know $t^k \downarrow 0$ and it remains to show $F(x^*) + t^k d^k \in \Lambda$ for all $k \in \mathbb{N}$. Let $k \in \mathbb{N}$ be arbitrary but fixed and recall that x^* is feasible for (1). For all $i = 1, \dots, m$ two cases can occur: If $d_{g_i}^k < 0$, we have

$$g_i(x^*) + t^k d_{g_i}^k < g_i(x^*) \leq 0,$$

and if $d_{g_i}^k \geq 0$, we have

$$g_i(x^*) + t^k d_{g_i}^k \leq g_i(x^*) + t_{g_i}^k d_{g_i}^k \leq 0.$$

For all $i = 1, \dots, p$ we have $d_{h_i}^k = 0$ because of $h_i(x^*) = 0$ and $t_{h_i}^k > 0$. Consequently, we obtain

$$h_i(x^*) + t^k d_{g_i}^k = 0.$$

Now consider an $i \in I_{+0}(x^*)$. Because of $d_{G_i}^k \to d_{G_i}$ and $t_{GH_i}^k \downarrow 0$ this implies

$$G_i(x^*) + t_{GH_i}^k d_{G_i}^k > 0$$

for all $k \in \mathbb{N}$ sufficiently large and, consequently, $d_{H_i}^k = 0$ for these k. This implies

$$0 \leq G_i(x^*) + t^k d_{G_i}^k \perp H_i(x^*) + t^k d_{H_i}^k \geq 0$$

for all $k \in \mathbb{N}$ sufficiently large. By symmetry, we also obtain

$$0 \leq G_i(x^*) + t^k d_{G_i}^k \perp H_i(x^*) + t^k d_{H_i}^k \geq 0$$

for all $i \in I_{0+}(x^*)$. In remains to consider $i \in I_{00}(x^*)$. In this case, we know

$$0 \leq d_{G_i}^k \perp d_{H_i}^k \geq 0,$$

which directly implies

$$0 \leq G_i(x^*) + t^k d_{G_i}^k \perp H_i(x^*) + t^k d_{H_i}^k \geq 0$$

for all $k \in \mathbb{N}$. Together, this proves $F(x^*) + t^k d^k \in \Lambda$ for all $k \in \mathbb{N}$ sufficiently large. □

With this lemma, we can prove that MPEC generalized pseudonormality implies MPEC-ACQ.

Lemma 5.5. *Let x^* be feasible for* (1) *such that MPEC generalized pseudonormality holds in x^*. Then MPEC-ACQ also holds in x^*.*

Proof. As we have proven in Lemma 4.3 and Lemma 4.4, MPEC generalized pseudonormality implies the existence of local error bounds. According to [9, Corollary 1], the existence of local error bounds is equivalent to calmness of the perturbation map

$$M(r) := \{x \in \mathbb{R}^n \mid F(x) + r \in \Lambda\}$$

in $(0, x^*)$. Thus, we can apply Proposition 1 from the same paper and obtain $T_X(x^*) = L_X(x^*)$, where $L_X(x^*)$ is defined as

$$L_X(x^*) := \{d \in \mathbb{R}^n \mid \nabla F(x^*)^T d \in T_\Lambda(F(x^*))\}.$$

Because of Lemma 5.4, we may write $T_\Lambda(F(x^*))$ as

$$T_\Lambda(F(x^*)) = \bigtimes_{i=1}^{m} T_{(-\infty,0]}(g_i(x^*)) \times \bigtimes_{i=1}^{p} T_{\{0\}}(h_i(x^*)) \times \bigtimes_{i=1}^{q} T_C(G_i(x^*), H_i(x^*)).$$

Now, we can apply this knowledge to $L_X(x^*)$ and obtain

$$\begin{array}{rcll} L_X(x^*) & = & \{d \in \mathbb{R}^n & \mid \ \nabla g_i(x^*)^T d \in T_{(-\infty,0]}(g_i(x^*)) \ \forall i = 1, \dots, m, \\ & & & \quad \nabla h_i(x^*)^T d \in T_{\{0\}}(h_i(x^*)) \ \forall i = 1, \dots, p, \\ & & & \quad (\nabla G_i(x^*)^T d, \nabla H_i(x^*)^T d) \in T_C(G_i(x^*), H_i(x^*)) \ \forall i = 1, \dots, q\} \\ & = & \{d \in \mathbb{R}^n & \mid \ \nabla g_i(x^*)^T d \le 0 \ \forall i \in I_g(x^*), \\ & & & \quad \nabla h_i(x^*)^T d = 0 \ \forall i = 1, \dots, p, \\ & & & \quad \nabla G_i(x^*)^T d = 0 \ \forall i \in I_{0+}(x^*), \\ & & & \quad \nabla H_i(x^*)^T d = 0 \ \forall i \in I_{+0}(x^*), \\ & & & \quad \nabla G_i(x^*)^T d \ge 0 \ \forall i \in I_{00}(x^*), \\ & & & \quad \nabla H_i(x^*)^T d \ge 0 \ \forall i \in I_{00}(x^*), \\ & & & \quad (\nabla G_i(x^*)^T d)(\nabla H_i(x^*)^T d) = 0 \ \forall i \in I_{00}(x^*)\} \\ & = & L_{MPEC}(x^*), & \end{array}$$

where $L_{MPEC}(x^*)$ denotes the MPEC-linearized cone from Definition 2.4. Consequently, we have $T_X(x^*) = L_X(x^*) = L_{MPEC}(x^*)$, which is exactly MPEC-ACQ. □

Altogether, we have proven the following relations in this paper for a local minimum x^* of the MPEC (1):

$$\begin{array}{ccccc} & & \text{MPEC-MFCQ} & & \\ & & \Downarrow & & \\ & & \text{MPEC generalized MFCQ} & & \\ & & \Downarrow & & \\ & & \text{MPEC generalized pseudonormality} & \Rightarrow & \text{MPEC generalized quasinormality} \\ & & \Downarrow & & \Downarrow \\ \text{exactness of } P_\alpha & \Leftarrow & \text{calmness of } M \text{ in } (0, x^*) & & \text{M-stationarity} \\ & & \Downarrow & & \\ & & \text{MPEC-ACQ} & & \end{array}$$

6 Final Remarks

We presented an enhanced version of the Fritz John conditions for mathematical programs with equilibrium constraints (MPECs). This version was used, in particular, to obtain a simple and direct proof for a local minimum of the MPEC to be an M-stationary point under suitable assumptions. These assumptions are, in fact, certain (new) constraint qualifications introduced within this paper and motivated by the statements of the enhanced Fritz John conditions. One of these new constraint qualification is also used to prove a new exact penalty result for MPECs.

References

[1] D.P. Bertsekas and A.E. Ozdaglar: *Pseudonormality and a Lagrange multiplier theory for constrained optimization.* Journal of Optimization Theory and Applications 114, 2002, pp. 187-343.

[2] J.M. Borwein and A.S. Lewis: *Convex Analysis and Nonlinear Optimization.* Canadian Mathematical Society Books in Mathematics, Springer, New York, 2000.

[3] F.H. Clarke: *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York, 1983 (reprinted by SIAM, Philadelphia, PA, 1990).

[4] S. Dempe: *Foundations of Bilevel Programming.* Nonconvex Optimization and Its Applications, Vol. 61, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

[5] M.L. Flegel and C. Kanzow: *A Fritz John approach to first order optimality conditions for mathematical programs with equilibrium constraints.* Optimization 52, 2003, pp. 277-286.

[6] M.L. Flegel and C. Kanzow: *On M-stationary points for mathematical programs with equilibrium constraints.* Journal of Mathematical Analysis and Applications 310, 2005, pp. 286-302.

[7] M.L. Flegel and C. Kanzow: *On the Guignard constraint qualification for mathematical programs with equilibrium constraints.* Optimization 54, 2005, pp. 517-534.

[8] M.L. Flegel, C. Kanzow, and J.V. Outrata: *Optimality conditions for disjunctive programs with applications to mathematical programs with equilibrium constraints.* Set-Valued Analysis 15, 2007, pp. 139-162.

[9] R. Henrion and J.V. Outrata: *Calmness of constraint systems with applications.* Mathematical Programming 104, 2005, pp. 437-464.

[10] T. Hoheisel, C. Kanzow, and J.V. Outrata: *Exact penalty results for mathematical programs with vanishing constraints.* Preprint 289, Institute of Mathematics, University of Würzburg, Würzburg, May 2009 (revised October 2009).

[11] G. Liu, J.J. Ye, and J. Zhu: *Partial exact penalty for mathematical programs with equilibrium constraints.* Set-Valued Analysis 16, 2008, pp. 785-804.

[12] Z.-Q. Luo, J.-S. Pang, and D. Ralph: *Mathematical Programs with Equilibrium Constraints.* Cambridge University Press, Cambridge, UK, 1996.

[13] Z.-Q. Luo, J.-S. Pang, D. Ralph, and S.Q. Wu: *Exact penalization and stationarity conditions for mathematical programs with equilibrium constraints.* Mathematical Programming 75, 1996, pp. 19-76.

[14] O.L. Mangasarian: *Nonlinear Programming.* McGraw-Hill, New York, 1969 (reprinted by SIAM, Philadephia, PA, 1994).

[15] B.S. Mordukhovich: *Variational Analysis and Generalized Differentiation I. Basic Theory.* A Series of Comprehensive Studies in Mathematics, Vol. 330, Springer, Berlin, Heidelberg, 2006.

[16] B. S. Mordukhovich: *Variational Analysis and Generalized Differentiation II. Applications.* A Series of Comprehensive Studies in Mathematics, Vol. 331, Springer, Berlin, Heidelberg, 2006.

[17] J.V. Outrata: *Optimality conditions for a class of mathematical programs with equilibrium constraints.* Mathematics of Operations Research 24, 1999, pp. 627-644.

[18] J.V. Outrata: *A generalized mathematical program with equilibrium constraints.* SIAM Journal on Control and Optimization 38, 2000, pp. 1623-1638.

[19] J. V. Outrata, M. Kočvara, and J. Zowe: *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints.* Nonconvex Optimization and its Applications, Vol. 28, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[20] J.-S. Pang and M. Fukushima: *Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints.* Computational Optimization and Applications 13, 1999, pp. 111-136.

[21] R.T Rockafellar and R.J.-B. Wets: *Variational Analysis.* A Series of Comprehensive Studies in Mathematics, Vol. 317, Springer, Berlin, Heidelberg, 1998.

[22] H. Scheel and S. Scholtes: *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity.* Mathematics of Operations Research 25, 2000, pp. 1-22.

[23] S. Scholtes and M. Stöhr: *Exact penalization of mathematical programs with equilibrium constraints.* SIAM Journal on Control and Optimization 37, 1999, pp. 617–652.

[24] Z. Wu and J.J. Ye: *First-order and second-order conditions for error bounds.* SIAM Journal on Optimization 14, 2003, pp. 621–645.

[25] J.J. Ye: *Optimality conditions for optimization problems with complementarity constraints.* SIAM Journal on Optimization 9, 1999, pp. 374–387.

[26] J.J. Ye: *Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints.* SIAM Journal on Optimization 10, 2000, pp. 943–962.

[27] J.J. Ye: *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints.* Journal of Mathematical Analysis and Applications 307, 2005, pp. 350-369.

[28] J.J. Ye and X.Y. Ye: *Necessary optimality conditions for optimization problems with variational inequality constraints.* Mathematics of Operations Research 22, 1997, pp. 977-997.

[29] J.J. Ye, D.L. Zhu, and Q.J. Zhu: *Exact penalization and necessary optimality conditions for generalized bilevel programming problems.* SIAM Journal on Optimization 7, 1997, pp. 481-507.