Contents lists available at ScienceDirect

Nonlinear Analysis
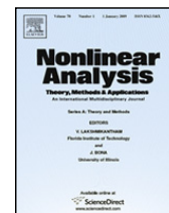
Exact penalty results for mathematical programs with vanishing constraints☆

Tim Hoheisel [a], Christian Kanzow [a,*], Jiří V. Outrata [b]

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

[b] Academy of Sciences, Institute of Information Theory and Automation, 18208 Prague, Czech Republic

ARTICLE INFO

ABSTRACT

A mathematical program with vanishing constraints (MPVC) is a constrained optimization problem arising in certain engineering applications. The feasible set has a complicated structure so that the most familiar constraint qualifications are usually violated. This, in turn, implies that standard penalty functions are typically non-exact for MPVCs. We therefore develop a new MPVC-tailored penalty function which is shown to be exact under reasonable assumptions. This new penalty function can then be used to derive (or recover) suitable optimality conditions for MPVCs.

1. Introduction

We consider a constrained optimization problem of the form

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

that we call a *Mathematical Program with Vanishing Constraints*, or MPVC for short, where all functions $f, H_i, G_i : \mathbb{R}^n \to \mathbb{R}$ are assumed to be continuously differentiable.

More generally, an MPVC is a mathematical program of the form

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

with some additional functions $g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ which represent some standard equality and inequality constraints. In order to keep our notation as simple as possible, we skip these standard constraints from this program and consider, from the

very beginning, the formulation (1) where only the difficult constraints are kept. Generalizations of our subsequent results to the more general setting are straightforward.

MPVCs are closely related to the class of mathematical programs with equilibrium (or complementarity) constraints, MPECs for short; see [1,2] for more details. In fact, it was already pointed out in [3] that it is possible to reformulate an MPVC as an MPEC so that the entire machinery from MPECs can, in principle, be applied to MPVCs. However, as noted in [3], the MPEC arising from MPVCs looses many desirable properties like, for example, uniqueness of solutions. Another reformulation of an MPVC as an MPEC was motivated by a comment of a referee and uses a continuously differentiable function $\delta : \mathbb{R} \to \mathbb{R}$ such that

$$\delta(t) \begin{cases} > 0, & \text{if } t > 0, \\ = 0, & \text{if } t \leq 0. \end{cases}$$

For example, one may take $\delta(t) := \max^2\{0, t\}$. Then it is easy to see that

$$b \geq 0, ab \leq 0 \iff b \geq 0, \delta(a) \geq 0, b\delta(a) = 0,$$

hence the MPVC constraints can be rewritten as

$$H_i(x) \geq 0, \qquad \tilde{G}_i(x) := \delta\big(G_i(x)\big) \geq 0, \qquad H_i(x)\tilde{G}_i(x) = 0 \quad \forall i = 1, \ldots, l$$

which is the typical setting of an MPEC. While this reformulation is certainly interesting, it has some disadvantages at least from a theoretical point of view. In particular, many MPEC-constraint qualifications are violated since $\nabla \tilde{G}_i(x^*) = 0$ whenever x^* is feasible for the original MPVC such that $G_i(x^*) < 0$. Moreover, the nonlinearity of δ can prevent a direct verification of the calmness of the respective perturbation map, provided the functions G_i, H_i are affine (cf. program (20) in Section 4). All these properties (constraint qualifications, calmness), however, will play an essential role in this paper.

This, in particular, motivates to consider the MPVC as an optimization problem for itself. In fact, there have already been published a couple of papers on MPVCs which investigate applicational, theoretical and numerical aspects: In [3], the first in the field of MPVCs, it is shown that this class of problems can be used as a unified framework for several problems from truss topology optimization. The papers [4–6] are mainly concerned with constraint qualifications and optimality conditions for MPVCs. Some numerical approaches are investigated in [7,8], where the first one is based on smoothing and regularization ideas and the latter employs a pure relaxation method. In [8] there is also presented some stability analysis, whereas [7] provides broad numerical results.

In this paper, however, we are interested in exact penalty results for MPVCs. To this end, we first recall some basic definitions and preliminary results in Section 2. We then state an exact penalty result in Section 3 within the framework of a rather general mathematical program. This result is then specialized to the MPVC-setting in Section 4, where we derive an MPVC-tailored penalty function and show that this new penalty function is exact under suitable assumptions. This exact penalty result is then used in Section 5 in order to give an alternative proof for the existence of suitable multipliers such that certain optimality conditions (called M-stationarity) hold at a local minimum of the MPVC. Section 6 then considers the exactness of the classical l_1-penalty function for MPVC; however, the conditions which guarantee exactness of the new penalty function considered in Section 4 are not sufficient for the exactness of the l_1-penalty function. We close with some final remarks in Section 7.

Notation: $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}_+ := [0, +\infty)$ is the set of nonnegative real numbers, and $\mathbb{R}_- := (-\infty, 0]$ are the nonpositive numbers. In addition to that we put $\bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$. Given a ($n$ index) set I, we write $\mathcal{P}(I)$ for the set of all partitions of I into two disjoint subsets of I, i.e. $(\beta_1, \beta_2) \in \mathcal{P}(I)$ if and only if $\beta_1 \cup \beta_2 = I$ and $\beta_1 \cap \beta_2 = \emptyset$. For a given set $S \subseteq \mathbb{R}^n$, we denote its convex hull by $\text{conv}(S)$. Moreover, for a nonempty closed (not necessarily convex) set $S \in \mathbb{R}^n$, the distance function $d_S : \mathbb{R}^n \to \mathbb{R}$ is given by

$$d_S(x) := \inf_{s \in S} \|x - s\|,$$

where $\| \cdot \|$ denotes an arbitrary l_p-norm in $\mathbb{R}^n$ for $p \in [1, \infty]$. Given a sequence $\{x^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ and a (not necessarily continuous) function $f : \mathbb{R}^n \to \mathbb{R}$, we write $x^k \underset{f}{\to} x$ if and only if $x^k \to x$ and $f(x^k) \to f(x)$. We further use the notation $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ for a multifunction or set-valued map, i.e., $\Phi(x)$ is a subset of $\mathbb{R}^n$. Its graph is defined as $\text{gph}\Phi := \{(x, y) \mid y \in \Phi(x)\}$. Finally, consider a mathematical program of the form

$$\min f(x) \quad \text{s.t.} \quad x \in X \tag{2}$$

for a given function $f : \mathbb{R}^n \to \mathbb{R}$ and a nonempty and closed feasible set $X \subseteq \mathbb{R}^n$. Any function of the form

$$P(x; \alpha) := f(x) + \alpha p(x)$$

with a (penalty) parameter $\alpha > 0$ will be called a *penalty function* of (2) provided that $p(x) \geq 0$ for all $x \in \mathbb{R}^n$ and $p(x) = 0$ if and only if $x \in X$. We say that this penalty function is *exact* at a local minimum x^* of (2) if there exists a finite penalty parameter $\bar{\alpha} > 0$ such that x^* is also a local minimum of the penalty function $P(x; \alpha)$ for all $\alpha \geq \bar{\alpha}$.

2. Preliminaries

Let X denote the feasible set of (1), and let $x^* \in X$ be an arbitrary feasible point. Then we define the index sets

$$
\begin{aligned}
I_+ &:= \big\{ i \mid H_i(x^*) > 0 \big\}, \\
I_0 &:= \big\{ i \mid H_i(x^*) = 0 \big\}.
\end{aligned}
\tag{3}
$$

Furthermore, we decompose the index set I_+ into the following subsets:

$$
\begin{aligned}
I_{+0} &:= \big\{ i \mid H_i(x^*) > 0, G_i(x^*) = 0 \big\}, \\
I_{+-} &:= \big\{ i \mid H_i(x^*) > 0, G_i(x^*) < 0 \big\}.
\end{aligned}
\tag{4}
$$

Similarly, we partition the set I_0 in the following way:

$$
\begin{aligned}
I_{0+} &:= \big\{ i \mid H_i(x^*) = 0, G_i(x^*) > 0 \big\}, \\
I_{00} &:= \big\{ i \mid H_i(x^*) = 0, G_i(x^*) = 0 \big\}, \\
I_{0-} &:= \big\{ i \mid H_i(x^*) = 0, G_i(x^*) < 0 \big\}.
\end{aligned}
\tag{5}
$$

Note that the first subscript indicates the sign of $H_i(x^*)$, whereas the second subscript stands for the sign of $G_i(x^*)$.

By means of these index sets, we are now in a position to state the two most prominent stationarity concepts for MPVCs.

Definition 2.1. Let x^* be feasible for (1).

(a) Then x^* is called *M-stationary* if there exist multipliers (η^G, η^H) such that

$$
0 = \nabla f(x^*) + \sum_{i=1}^{l} \eta_i^G \nabla G_i(x^*) - \sum_{i=1}^{l} \eta_i^H \nabla H_i(x^*)
\tag{6}
$$

and

$$
\begin{aligned}
&\eta_i^G = 0 \quad (i \in I_{+-} \cup I_{0-} \cup I_{0+}), \quad \eta_i^G \geq 0 \quad (i \in I_{00} \cup I_{+0}), \\
&\eta_i^H = 0 \quad (i \in I_+), \quad \eta_i^H \geq 0 \quad (i \in I_{0-}), \\
&\eta_i^G \eta_i^H = 0 \quad (i \in I_{00}).
\end{aligned}
\tag{7}
$$

(b) The point x^* is called *strongly stationary* if it is M-stationary and, in addition,

$$
\eta_i^G = 0, \quad \eta_i^H \geq 0 \quad (i \in I_{00}).
\tag{8}
$$

Apparently, strong stationarity implies M-stationarity and both concepts coincide as soon as the critical index set I_{00} is empty. Moreover, in [3], strong stationarity was shown to be equivalent to the standard *Karush–Kuhn–Tucker conditions* of (1) and hence, strong stationarity is a first order optimality condition in the presence of standard constraint qualifications, like the *Guignard constraint qualification (GCQ)* or the *Mangasarian–Fromovitz constraint qualification (MFCQ)*; see [9], for example. In turn, according to [5], M-stationarity even holds under some weaker and more specific assumptions like the MPVC-GCQ or the MPVC-MFCQ, which occur in Section 4.

The notion of the polar cone of a set is needed to establish several normal cones which will be employed in particular in Section 3.

Definition 2.2. Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a nonempty set. Then

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

is the *polar cone* of $\mathcal{C}$.

The prominent *tangent cone* is a standard tool in optimization and variational analysis. For a closed set $\emptyset \neq C \subseteq \mathbb{R}^n$ and $x^* \in C$, it is defined by

$$
\mathcal{T}_C(x^*) := \left\{ d \in \mathbb{R}^n \ \Big| \ \exists \{x^k\} \subseteq C, t_k \downarrow 0 : x^k \to x^* \text{ and } \frac{x^k - x^*}{t_k} \to d \right\}.
$$

An important device for our analysis is the so-called *limiting normal cone*.

Definition 2.3. Let $C \subseteq \mathbb{R}^n$ be a nonempty, closed set. Then

(a) the *Fréchet normal cone* to C at $a \in C$ is defined by $\hat{N}(a, C) := (\mathcal{T}_C(a))^\circ$, i.e., the Fréchet normal cone is the polar of the tangent cone.

(b) the *limiting normal cone* to C at $a \in C$ is defined by

$$N(a, C) := \left\{ \lim_{k \to \infty} w^k \;\middle|\; \exists \{a^k\} \subseteq C : a^k \to a, \; w^k \in \hat{N}(a^k, C) \; \forall k \in \mathbb{N} \right\}. \tag{9}$$

The Fréchet normal cone is sometimes also called the *regular normal cone*, most notably in [10], whereas the limiting normal cone comes with a number of different names, including *normal cone*, *basic normal cone*, and *Mordukhovich normal cone* due to the many contributions of Mordukhovich in this area, see, in particular, [11,12] for an extensive treatment and many applications of this cone. In the case of a convex set C, both the Fréchet normal cone and the limiting normal cone coincide with the standard normal cone from convex analysis; cf. [13].

Closely linked are the notions of the *Fréchet* and the *limiting subdifferential*, which may also be found in [10]. Note that we write *lsc* as an abbreviation of *lower semicontinuous*.

Definition 2.4. Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be lsc and $f(x)$ finite.

(a) The set

$$\partial_F f(x) := \left\{ s \in \mathbb{R}^n \;\middle|\; \liminf_{y \to x} \frac{f(y) - f(x) - s^T(y - x)}{\|y - x\|} \geq 0 \right\}$$

is called the *Fréchet subdifferential* of f at x.

(b) The set

$$\partial f(x) := \left\{ \lim_{k \to \infty} s^k \;\middle|\; \exists x^k \underset{f}{\to} x, \; s^k \in \partial_F f(x^k) \right\}$$

is called the *limiting subdifferential* of f at x.

3. A generalized mathematical program

In this section, we consider a general mathematical program of the form

$$\min f(x) \quad \text{s.t.} \quad F(x) \in \Lambda, \tag{10}$$

with locally Lipschitz functions $f : \mathbb{R}^n \to \mathbb{R}$, $F : \mathbb{R}^n \to \mathbb{R}^m$ and a nonempty closed set $\Lambda \subseteq \mathbb{R}^m$. This type of problem was already fruitfully employed in the MPEC-field in [14] (MPEC = mathematical programs with equilibrium constraints).

As soon as one tries to investigate exact penalty results for a class of optimization problems, the very closely linked concept of *calmness* of the respective problem, cf. [15–17], arises naturally for reasons explained below.

In order to define calmness for our general optimization problem (10), consider the associated family of perturbed problems

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

for some parameter $p \in \mathbb{R}^m$. Note that, obviously, it holds that (10) and $\Pi(0)$ are the same problems. The following definition of calmness is due to Burke; see [15, Def. 1.1].

Definition 3.1. Let x^* be feasible for $\Pi(0)$. Then the problem is called *calm* at x^* if there exist constants $\bar{\alpha} > 0$ and $\varepsilon > 0$ such that for all $(x, p) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying $x \in B_\varepsilon(x^*)$ and $F(x) + p \in \Lambda$, one has

$$f(x) + \bar{\alpha} \|p\| \geq f(x^*).$$

In this context $\bar{\alpha}$ and ε are called the *modulus* and the *radius* of calmness for $\Pi(0)$ at x^*. Note that the original definition by Clarke; see [17, Def. 6.4.1], also involves that $p \in B_\varepsilon(0)$. Actually, these definitions coincide as soon as the function F is continuous, as was coined in [15, Prop. 2.1], which is in particular fulfilled in our setup.

When Clarke established the notion of calmness as a tool for sensitivity analysis of parametrized optimization problems, he already was aware of its close connection to the concept of exact penalization. He showed that calmness is a sufficient condition for exact penalization. The full relation, however, is due to Burke see [15, Th. 1.1], and restated in the following result.

Proposition 3.2. *Let x^* be feasible for $\Pi(0)$. Then $\Pi(0)$ is calm at x^* with modulus $\bar{\alpha}$ and radius ε if and only if x^* is a minimum of*

$$P(x; \alpha) := f(x) + \alpha d_\Lambda(F(x)) \tag{11}$$

over $B_\varepsilon(x^*)$ for all $\alpha \geq \bar{\alpha}$.

Proof. See [15, Th. 1.1]. $\quad \square$

In the course of rising popularity of the calculus of multifunctions and their applications to optimization problems, another calmness concept has been established and successfully employed in the context of mathematical programming. The following definition of calmness of a multifunction can be found, e.g., in [10].

Definition 3.3. Let $\Phi : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ be a multifunction with a closed graph and $(u, v) \in \text{gph}\Phi$. Then we say that Φ is *calm* at (u, v) if there exist neighbourhoods U of u, V of v and a modulus $L \geq 0$ such that

$$\Phi(u') \cap V \subseteq \Phi(u) + L\|u - u'\|\mathbb{B} \quad \forall u' \in U. \tag{12}$$

The application to our mathematical programming setup from (10) and $\Pi(p)$ follows by virtue of the following multifunction $M : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, often named *perturbation map*, which is defined by

$$M(p) := \{x \in \mathbb{R}^n \mid F(x) + p \in \Lambda\}. \tag{13}$$

By means of the perturbation map, the feasible set of $\Pi(p)$ is then given by $M(p)$, in particular, one has $F^{-1}(\Lambda) = M(0)$.

Part of the gain from the notion of calmness of multifunctions for optimization is revealed by the following two results. In the first result, we see that calmness of the perturbation map at a particular point is in fact equivalent to the existence of local error bounds; see [18].

Proposition 3.4. *Let $x^* \in M(0)$ be feasible for (10). Then the following statements are equivalent.*

(1) *M is calm at $(0, x^*)$.*
(2) *There exists a neighbourhood U of x^* and a constant $\rho > 0$ such that*

$$d_{F^{-1}(\Lambda)}(x) \leq \rho d_\Lambda(F(x)) \quad \forall x \in U. \tag{14}$$

Proof. See [19, Corollary 1]. □

The second result shows that, roughly speaking, calmness of the perturbation map (Definition 3.3) yields calmness of the unperturbed problem $\Pi(0)$ (Definition 3.1).

Proposition 3.5. *Let $x^* \in M(0)$ be a local minimizer of (10) such that M is calm at $(0, x^*)$. Then $\Pi(0)$ is calm at x^*.*

Proof. By assumption, M is calm at $(0, x^*)$ and hence, due to Proposition 3.4, there exist constants $\tilde{\varepsilon}, \rho > 0$ such that

$$d_{F^{-1}(\Lambda)}(x) \leq \rho d_\Lambda(F(x)) \quad \forall x \in B_{\tilde{\varepsilon}}(x^*).$$

Now, choose $\hat{\varepsilon} \in (0, \tilde{\varepsilon}]$ such that f attains a minimum over $B_{\hat{\varepsilon}}(x^*) \cap F^{-1}(\Lambda)$ at x^*. Then put $\varepsilon := \frac{\hat{\varepsilon}}{2}$ and choose $x \in B_\varepsilon(x^*)$ arbitrarily. Moreover, let

$$x_0 \in \text{Proj}_{F^{-1}(\Lambda)}(x).$$

In particular, this implies $x_0 \in B_{\hat{\varepsilon}}(x^*)$. Together, one obtains

$$\begin{aligned} f(x^*) &\leq f(x_0) \\ &\leq f(x) + L\|x - x_0\| \\ &= f(x) + Ld_{F^{-1}(\Lambda)}(x) \\ &\leq f(x) + \rho L d_\Lambda(F(x)), \end{aligned} \tag{15}$$

where $L > 0$ denotes the local Lipschitz constant of f around x^*. If, now, we put $\bar{\alpha} := \rho L$ and note that, for $p \in \mathbb{R}^m$, we have $d_\Lambda(F(x)) \leq \|p\|$ whenever $F(x) + p \in \Lambda$, we apparently get the desired calmness of $\Pi(0)$. □

An immediate consequence is the following corollary.

Corollary 3.6. *Let $x^* \in M(0)$ be such that M is calm at $(0, x^*)$. Then the penalty function from (11) is exact at x^*.*

Proof. The proof follows immediately from Propositions 3.5 and 3.2. □

In the rest of this section, we will provide sufficient conditions for the calmness of the multifunction M at $(0, x^*)$ for some $x^* \in M(0)$. Thus, we automatically obtain sufficient conditions for the function $P(x; \alpha) = f(x) + \alpha d_\Lambda(F(x))$ to be exact at x^*. From now on we will assume the functions f and F to be continuously differentiable. Then we can define the following generalization of the Mangasarian–Fromovitz constraint qualification; see [14].

Definition 3.7. Let x^* be feasible for (10). We say that the *generalized Mangasarian–Fromovitz constraint qualification* (GMFCQ) holds at x^* if the following implication holds:

$$\left. \begin{array}{l} F'(x^*)^T \lambda = 0 \\ \lambda \in N_\Lambda(F(x^*)) \end{array} \right\} \implies \lambda = 0. \tag{16}$$

Note that, if $\Lambda = \mathbb{R}_-^m$, (16) reduces to standard MFCQ.

The notion of *GMFCQ* leads to the following result.

Proposition 3.8. *Let $x^* \in M(0)$ be feasible for* (10) *such that GMFCQ is satisfied. Then the perturbation map M is calm at* $(0, x^*)$.

Proof. See the proof of [14, Corollary 2.4]. □

The following corollary follows immediately.

Corollary 3.9. *Let $x^* \in M(0)$ be feasible for* (10) *such that GMFCQ is satisfied. Then the penalty function from* (11) *is exact at x^*.*

4. Deriving an exact penalty function for MPVCs

In order to derive an exact penalty function for the MPVC (1), we are guided by the results from Section 3, in particular Corollary 3.9. The path that we follow starts with a reformulation of the MPVC in the fashion of (10). Afterwards we will provide sufficient conditions for the GMFCQ to hold for the rewritten MPVC, which eventually provides an exact penalty function. Note, however, that the question whether GMFCQ holds or not, substantially depends on the chosen representation of the feasible set.

For the sake of reformulating the MPVC, consider the characteristic set

$$C := \{(a, b) \in \mathbb{R}^2 \mid b \geq 0, \ ab \leq 0\}, \tag{17}$$

and put

$$\Lambda_{VC} := \mathop{\times}_{i=1}^{l} C. \tag{18}$$

Furthermore, define the map $F : \mathbb{R}^n \to \mathbb{R}^{2l}$ by

$$F^{VC}(x) := (F_i^{VC}(x))_{i=1,\dots,l} := \begin{pmatrix} G_i(x) \\ H_i(x) \end{pmatrix}_{i=1,\dots,l}. \tag{19}$$

By means of these definitions, we are able to write the MPVC (1) as the following program

$$\min f(x) \quad \text{s.t.} \ F^{VC}(x) \in \Lambda_{VC}. \tag{20}$$

The perturbation map for (20) is consequently given by

$$M^{VC}(p) := \{x \in \mathbb{R}^n \mid F^{VC}(x) + p \in \Lambda_{VC}\}.$$

In order to find conditions to yield GMFCQ for (20), we need the following auxiliary result, which is concerned with calculating the limiting normal cone of the characteristic set C from (17).

Lemma 4.1. *Let $(a, b) \in C$. Then it holds that*

$$N_C((a, b)) = \begin{cases} \{0\} \times \{0\} & \text{if } b > 0, \ a < 0, \\ \mathbb{R}_+ \times \{0\} & \text{if } b > 0, \ a = 0, \\ \{0\} \times \mathbb{R}_- & \text{if } b = 0, \ a < 0, \\ \{0\} \times \mathbb{R} & \text{if } b = 0, \ a > 0, \\ \{(u, v) \mid u \geq 0, \ uv = 0\} & \text{if } a = b = 0. \end{cases} \tag{21}$$

Proof. See the proof of [5, Lemma 3.2]. □

With the aid of the above Lemma, we are now able to prove a first sufficiency result for GMFCQ in the MPVC setup.

Theorem 4.2. *Let $x^* \in M(0)$ be feasible for* (1) *and assume that for all $(\beta_1, \beta_2) \in \mathcal{P}(I_{00})$ the following two conditions are satisfied:*

(i) *There exists a vector $d \in \mathbb{R}^n$ such that*

$$\begin{aligned} \nabla G_i(x^*)^T d &> 0 \quad (i \in I_{+0} \cup \beta_2), \\ \nabla H_i(x^*)^T d &< 0 \quad (i \in I_{0-}), \\ \nabla H_i(x^*)^T d &= 0 \quad (i \in I_{0+} \cup \beta_1). \end{aligned} \tag{22}$$

(ii) *The gradients $\nabla H_i(x^*)$ $(i \in I_{0+} \cup \beta_1)$ are linearly independent.*

Then GMFCQ holds for (20).

Proof. Observe first that with $(a_i, b_i)^T := (F_i^{VC}(x^*)) = (G_i(x^*), H_i(x^*))^T$ we have

$$N_{\Lambda_{VC}}(F^{VC}(x^*)) = \mathop{\times}_{i=1}^{l} N_C((a_i, b_i)),$$

cf. [10, Proposition 6.41]. By means of Lemma 4.1 it follows that GMFCQ amounts to the condition

$$\left.\begin{array}{l}
0 = \sum_{i=1}^{l} \lambda_i^G \nabla G_i(x^*) + \sum_{i=1}^{l} \lambda_i^H \nabla H_i(x^*) \\
\lambda_i^G = 0 \ (i \in I_{+-} \cup I_{0+} \cup I_{0-}), \quad \lambda_i^G \geq 0 \ (i \in I_{+0} \cup I_{00}), \\
\lambda_i^H = 0 \ (i \in I_+), \quad \lambda_i^H \leq 0 \ (i \in I_{0-}), \\
\lambda_i^G \lambda_i^H = 0 \ (i \in I_{00}),
\end{array}\right\} \implies \lambda^G = \lambda^H = 0.$$

This is equivalent to

$$\left.\begin{array}{l}
0 = \sum_{i \in I_{+0} \cup I_{00}} \lambda_i^G \nabla G_i(x^*) + \sum_{i \in I_0} \lambda_i^H \nabla H_i(x^*) \\
\lambda_i^G \geq 0 \ (i \in I_{+0} \cup I_{00}), \\
\lambda_i^H \leq 0 \ (i \in I_{0-}), \\
\lambda_i^G \lambda_i^H = 0 \ (i \in I_{00}),
\end{array}\right\} \implies \begin{array}{l} \lambda_i^G = 0 \ (i \in I_{+0} \cup I_{00}), \\ \lambda_i^H = 0 \ (i \in I_0). \end{array}$$

This, eventually, is equivalent to the following condition: For all partitions $(\beta_1, \beta_2) \in \mathcal{P}(I_{00})$, the implication

$$\left.\begin{array}{l}
0 = \sum_{i \in I_{+0} \cup \beta_2} \lambda_i^G \nabla G_i(x^*) + \sum_{i \in I_{0-} \cup I_{0+} \cup \beta_1} \lambda_i^H \nabla H_i(x^*) \\
\lambda_i^G \geq 0 \ (i \in I_{+0} \cup \beta_2), \\
\lambda_i^H \leq 0 \ (i \in I_{0-}),
\end{array}\right\} \implies \begin{array}{l} \lambda_i^G = 0 \ (i \in I_{+0} \cup \beta_2) \\ \lambda_i^H = 0 \ (i \in I_{0-} \cup I_{0+} \cup \beta_1) \end{array} \tag{23}$$

holds. Invoking Motzkin's theorem of the alternative, cf. [20], for example, we see that the implication (23) is, in case that $I_{0-} \cup I_{+0} \cup \beta_2 \neq \emptyset$, equivalent to condition (i). In turn, if $I_{0-} \cup I_{+0} \cup \beta_2 = \emptyset$, (23) reduces to the linear independence of the gradients $\nabla H_i(x^*)$ $(i \in I_{0+} \cup \beta_1)$, which is condition (ii). $\square$

In the MPVC-field, the following variant of the (standard) Mangasarian–Fromovitz and linear independence constraint qualifications have shown to be useful tools.

Definition 4.3. Let x^* be feasible for (1). Then we say that

(a) MPVC-MFCQ is satisfied at x^* if the gradients

$$\nabla H_i(x^*) \ (i \in I_{0+} \cup I_{00}) \tag{24}$$

are linearly independent, and there exists a vector d such that

$$\begin{array}{ll}
\nabla H_i(x^*)^T d > 0 & \forall i \in I_{0-}, \\
\nabla G_i(x^*)^T d < 0 & \forall i \in I_{+0} \cup I_{00}, \\
\nabla H_i(x^*)^T d = 0 & \forall i \in I_{0+} \cup I_{00}.
\end{array} \tag{25}$$

(b) MPVC-LICQ is satisfied at x^* if the gradients

$$\nabla H_i(x^*) \ (i \in I_0) \quad \text{and} \quad \nabla G_i(x^*) \ (i \in I_{00} \cup I_{+0})$$

are linearly independent.

These constraint qualifications were formally introduced in [5].

The following result, which is an immediate consequence of Theorem 4.2, will state that *MPVC-MFCQ* is a sufficient condition for calmness of the perturbation map M^{VC}.

Corollary 4.4. *Let x^* be feasible for (1) such that MPVC-MFCQ holds at x^*. Then M^{VC} is calm at $(0, x^*)$.*

Proof. MPVC-MFCQ obviously implies conditions (i) and (ii) from Theorem 4.2 and hence, GMFCQ holds. Due to Proposition 3.8, GMFCQ implies calmness of M^{VC} at $(0, x^*)$. $\square$

Putting all pieces of information together, we can state a satisfactory exact penalty result for the MPVC.

Theorem 4.5. *Let x^* be feasible for (1) such that MPVC-MFCQ holds at x^*. Then the function*

$$P^{VC}(x, \alpha) := f(x) + \alpha d_{\Lambda_{VC}}(F^{VC}(x)) \tag{26}$$

is exact at x^*.

Our goal is now to find an explicit representation for the penalty function from (26). To this end, the following elementary result is crucial.

Lemma 4.6. *Let C be given by* (17)*. Then for* $(a, b) \in C$ *we have*

$$d_C(a, b) = \max\{0, -b, \min\{a, b\}\} = \begin{cases} \min\{a, b\}, & \text{if } a, b \geq 0, \\ 0, & \text{if } a \leq 0, \ b \geq 0, \\ -b, & \text{if } b \leq 0. \end{cases}$$

Note that the previous result holds for an arbitrary l_p-norm to induce the distance function.

Corollary 4.7. *Let* $x \in \mathbb{R}^n$*. Then we have*

$$d_{A_{VC}}(F^{VC}(x)) = \|\left(d_C(F_i^{VC}(x))\right)_{i=1,\dots,l}\| = \|\left(\max\{0, -H_i(x), \min\{G_i(x), H_i(x)\}\}\right)_{i=1,\dots,l}\|.$$

5. An alternative proof for *M*-stationarity

We consider again the penalty function P^{VC} from (26). Under certain assumptions (like MPVC-MFCQ, cf. Theorem 4.5), this penalty function is exact, hence a local minimum of the MPVC is also a local minimizer of $P^{VC}(\cdot, \alpha)$ for some $\alpha > 0$. This implies that $0 \in \partial_x P(x^*, \alpha)$, and this condition can be used in order to derive optimality conditions for the MPVC itself. The question now is what type of optimality condition we can expect to get from this condition. Since, on the one hand, MPVC-MFCQ gives exactness of the penalty function P^{VC}, but, on the other hand, is not enough in order to guarantee that strong stationarity holds at a local minimizer x^* of MPVC, it is not possible to derive strong stationarity from the condition $0 \in \partial_x P(x^*, \alpha)$. The best we can expect to get is therefore M-stationarity, and this is precisely the aim of this section.

Hence, suppose that x^* is a local minimizer of $P^{VC}(\cdot, \alpha)$ for some $\alpha > 0$, so that $0 \in \partial_x P(x^*, \alpha)$. In view of the definition of P^{VC} in (26) we are, for obvious reasons, particularly interested in the limiting subdifferential of the distance function d_C from Lemma 4.6. To this end, we define $\phi : \mathbb{R}^2 \to \mathbb{R}$ by

$$\phi(a, b) := d_C(a, b). \tag{27}$$

Then the limiting subdifferential of ϕ at points from the set C is given in the below lemma.

Lemma 5.1. *Let* $\phi : \mathbb{R}^2 \to \mathbb{R}$ *be defined by* (27) *and let* $(a, b) \in C$*. Then we have*

$$\partial\phi(a, b) = \begin{cases} \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} & \text{if } b > 0, a < 0, \\ \text{conv}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} & \text{if } b > 0, a = 0, \\ \text{conv}\left\{ \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} & \text{if } b = 0, a > 0, \\ \text{conv}\left\{ \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} & \text{if } b = 0, a < 0, \\ \text{conv}\left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\} \cup \text{conv}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} & \text{if } a = b = 0. \end{cases}$$

Proof. Due to the fact that $\phi(a, b) = d_C(a, b)$ for all $(a, b) \in \mathbb{R}^2$, where d_C can be induced by any l_p-norm in $\mathbb{R}^2$, especially by the Euclidean norm, we may invoke [10, Example 8.53], which yields that

$$\partial\phi(a, b) = N((a, b), C) \cap \mathbb{B} \quad \forall (a, b) \in C, \tag{28}$$

where $\mathbb{B}$ denotes the closed Euclidean unit ball in $\mathbb{R}^2$ around the origin. The representation of the limiting normal cone from Lemma 4.1 together with (28) eventually gives the desired result. $\square$

The following main result of this section reveals that exactness of the penalty function P^{VC} from (11) at a local minimizer of the MPVC yields M-stationarity as an optimality condition.

Theorem 5.2. *Let* x^* *be a local minimizer of the MPVC* (1) *such that* P^{VC} *is exact at* x^**. Then M-stationarity holds at* x^**.*

Proof. Due to the fact that P^{VC} is exact at the local minimizer x^* of (1), there exists a penalty parameter $\alpha > 0$ such that x^* is also a local minimizer of $P^{VC}(\cdot, \alpha)$. In particular, we thus have $0 \in \partial_x P^{VC}(x^*, \alpha)$. Now, recall that by Corollary 4.7 we have

$$P^{VC}(x, \alpha) = f(x) + \alpha \|(\phi(G_i(x^*), H_i(x^*)))_{i=1,\dots,l}\|,$$

for an arbitrary l_p-norm $\| \cdot \|$. Due to the fact that P^{VC} is exact for an arbitrary l_p-norm if and only if it is exact when using the l_1-norm, we restrict ourselves to this case, since we may apply well-known sum rules for the limiting subdifferential then. Thus, consider the case that

$$P^{VC}(x, \alpha) = f(x) + \alpha \sum_{i=1}^{l} \phi(G_i(x^*), H_i(x^*)).$$

Invoking [10, Exercise 10.10] we hence obtain

$$0 \in \partial_x P^{VC}(x^*, \alpha) \subseteq \{\nabla f(x^*)\} + \alpha \sum_{i=1}^{l} \partial(\phi(G_i(x^*), H_i(x^*))),$$

and therefore, due to [21, p. 151], there exist vectors $(\rho_i, \nu_i) \in \partial\phi(G_i(x^*), H_i(x^*))$ for $i = 1, \ldots, l$ such that

$$0 = \nabla f(x^*) + \alpha \sum_{i=1}^{l} (\rho_i \nabla G_i(x^*) + \nu_i \nabla H_i(x^*)). \tag{29}$$

Now, put

$$\eta_i^G := \alpha \rho_i, \quad \eta_i^H := -\alpha \nu_i \quad \forall i = 1, \ldots, l.$$

Then (29) and Lemma 5.1 imply that (x^*, η^G, η^H) is an M-stationary point of (1). $\quad\square$

Combining the previous result with the sufficiency condition for the exactness of P^{VC} from Section 4, we can immediately show that MPVC-MFCQ yields M-stationarity at a local minimizer of (1), which is already well known; cf. [5].

Corollary 5.3. *Let x^* be a local minimizer of* (1) *such that MPVC-MFCQ holds. Then x^* is an M-stationary point.*

Proof. The proof follows immediately from Theorems 4.5 and 5.2. $\quad\square$

6. Exactness of the l_1-penalty function for MPVCs

The previous sections contain an MPVC-tailored penalty function that was shown to be exact under reasonable assumptions. On the other hand, one may view the MPVC as a standard constrained optimization problem and then consider the corresponding well-known l_1-penalty function as a natural candidate for an exact penalty function. Recall that this l_1-penalty function for (1) is given by

$$P(x, \alpha) := f(x) + \alpha \psi(x) := f(x) + \alpha \sum_{i=1}^{l} \max\{-H_i(x), 0\} + \alpha \sum_{i=1}^{l} \max\{G_i(x)H_i(x), 0\}. \tag{30}$$

Using the function

$$\varphi(a, b) := \max\{ab, 0\} - \min\{b, 0\} \tag{31}$$

(which was already used in [7] as the basis of an algorithm for the numerical solution of MPVCs), we can rewrite the l_1-penalty function as

$$P(x, \alpha) = f(x) + \alpha \sum_{i=1}^{l} \varphi(G_i(x), H_i(x)). \tag{32}$$

In what follows, we are now concerned with finding sufficient conditions for the exactness of the l_1-penalty function P from (30), (32).

It is commonly known that the l_1-penalty function of a nonlinear program is exact at a feasible point provided that MFCQ holds at this point; cf. [22]. In the context of MPVCs, however, this assumption is not reasonable, since it is too often violated; see [3].

Moreover, opposite to the penalty function from the previous section, MPVC-MFCQ cannot be a sufficient condition for exactness. This is due to the fact that exactness of the l_1-penalty function yields KKT conditions at a local minimizer, but MPVC-MFCQ does not necessarily guarantee this. The following example also shows that the l_1-penalty function is not exact in a number of rather standard situations.

Example 6.1. Consider the MPVC

$$\begin{aligned}
\min \quad & f(x) := -(x_1 + x_2) \\
\text{s.t.} \quad & H_1(x) := x_1 + x_2 \geq 0, \\
& G_1(x)H_1(x) := (x_1 + x_2)(x_1 + x_2) \leq 0.
\end{aligned} \tag{33}$$

Clearly, $x^* := (0, 0)$ is a local (in fact, global) minimizer and, for instance $x^k := (\frac{1}{k}, \frac{1}{k})$ is a sequence converging to x^*. However, to each penalty parameter $\alpha > 0$, we can find an index $k_\alpha \in \mathbb{N}$ such that

$$P(x^k; \alpha) = -\frac{2}{k} + \alpha \frac{4}{k^2} < 0 = P(x^*; \alpha) \qquad \text{for all } k \geq k_\alpha,$$

i.e., the penalty function $P(x; \alpha)$ is not exact. Note, however, that the perturbation map (13), associated with (33), satisfies GMFCQ at x^* in view of Theorem 4.2.

To derive a sufficient condition for the exactness of the l_1-penalty function, we employ a new notion from variational analysis. It is the so-called outer subdifferential introduced in [23].

Definition 6.2. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be lsc and $f(x)$ finite. Then the set

$$\partial^> f(x) := \left\{ \lim_{k \to \infty} s^k \mid \exists x^k \underset{f}{\to} x, f(x^k) > f(x), s^k \in \partial f(x^k) \right\}$$

is called the *outer subdifferential* of f at x.

The concept of the outer subdifferential is closely linked to exact penalization. In order to present the precise relationship, consider the optimization problem

$$\min f(x) \quad \text{s.t.} \quad x \in C \tag{34}$$

for a set $C \subseteq \mathbb{R}^n$ and a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$. Then the following result holds true.

Proposition 6.3. *Let $\Psi : \mathbb{R}^n \to \mathbb{R}_+$ be a penalty term in the sense that $\Psi(x) = 0$ if and only if $x \in C$. Moreover, let $x \in C$ such that $0 \notin \partial^> \Psi(x)$. Then, for all $\alpha > 0$ sufficiently large, the function $f + \alpha \Psi$ is an exact penalty function for (34) at x.*

Proof. The assumption that $0 \notin \partial^> \Psi(x)$ and the fact that the outer subdifferential is closed, yields a constant $\bar{\gamma} > 0$ such that

$$\|s\| \geq \bar{\gamma} \quad \forall s \in \partial^> \Psi(x).$$

This, by [23, Theorem 2.1], yields constants $c > 0$ (choose for example $c := \gamma^{-1}$ for $\gamma \in (0, \bar{\gamma})$) and $\delta > 0$ such that

$$d_C(y) \leq c\Psi(y) \quad \forall y \in B_\delta(x).$$

This, invoking [18, Theorem 3], implies that if x is a (local) minimizer of (34), then it is also a (local) minimizer of $f + \alpha \Psi$ for $\alpha > 0$ sufficiently large. $\quad\square$

Coming back to our MPVC setup, the above result tells us that a sufficient condition for the exactness of the function $P(\cdot, \cdot)$ from (30) at a local minimizer $x^* \in X$ is the condition $0 \notin M$ for a set $M \supseteq \partial^> \psi(x^*)$. Hence it is of great interest to find some handy upper estimate of $\partial^> \psi(x^*)$. To this end, define the function

$$\vartheta : \mathbb{R}^{2l} \to \mathbb{R}_+, \quad \vartheta(y) := \sum_{i=1}^{l} \varphi(y_i), \tag{35}$$

where $y = (y_i)_{i=1}^{l}$ and $y_i \in \mathbb{R}^2$ for $i = 1 \ldots, l$, and φ denotes the function from (31).

Then with the function F^{VC} from (19), we have $\psi = \vartheta \circ F^{VC}$. Now, what we obviously need is some kind of chain and sum rule for the outer subdifferential. For these purposes, consider the following propositions.

Proposition 6.4. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable, let $\vartheta : \mathbb{R}^m \to \mathbb{R}_+$ be Lipschitz, and put $f := \vartheta \circ F$. Moreover, let $C \subseteq \mathbb{R}^m$ be closed such that*

$$\vartheta(y) = 0 \iff y \in C.$$

Then, for $F(x^*) \in C$, one has

$$\partial^> f(x^*) \subseteq F'(x^*)^T \partial^> \vartheta(F(x^*)).$$

Proof. By the definition of the outer subdifferential and taking into account that $f(x^*) = 0$, we have

$$\partial^> f(x^*) = \left\{ \lim_{k \to \infty} s^k \;\middle|\; \exists x^k \to x^*, f(x^k) > 0, s^k \in \partial f(x^k) \right\}$$

$$= \left\{ \lim_{k \to \infty} s^k \;\middle|\; \exists x^k \to x^*, F(x^k) \notin C, s^k \in \partial f(x^k) \right\}$$

$$\subseteq \left\{ \lim_{k \to \infty} s^k \;\middle|\; \exists x^k \to x^*, F(x^k) \notin C, s^k \in F'(x^k)^T \partial \vartheta(F(x^k)) \right\}, \tag{36}$$

where the inclusion is due to [21, p. 151].

Now, we claim that

$$\left\{ \lim_{k\to\infty} s^k \;\middle|\; \exists x^k \to x^*, F(x^k) \notin C, s^k \in F'(x^k)^T \partial\vartheta(F(x^k)) \right\}$$

$$= F'(x^*)^T \left\{ \lim_{k\to\infty} \xi^k \;\middle|\; \exists x^k \to x^*, F(x^k) \notin C, \xi^k \in \partial\vartheta(F(x^k)) \right\}. \tag{37}$$

In fact, the inclusion '$\supseteq$' follows immediately from the fact that, for $\{x^k\} \to x^*$, we have $F'(x^k) \to F'(x^*)$ due to the continuity of F'. On the other hand, the reverse inclusion is a consequence of the uniform boundedness of the limiting subdifferential which guarantees that any sequence $\{b_k\} \subseteq \partial\vartheta(x^k)$ is bounded for $\{x^k\} \to x^*$.

Now, since $\vartheta(y) > 0$ whenever $y \notin C$, we have

$$\left\{ \lim_{k\to\infty} \xi^k \;\middle|\; \exists x^k \to x^*, F(x^k) \notin C, \xi^k \in \partial\vartheta(F(x^k)) \right\} \subseteq \partial^>\vartheta(F(x^*)),$$

and hence the assertion follows from (36) and (37). $\quad\square$

Proposition 6.5. *Consider a Lipschitz function $\Phi : \mathbb{R}^n \to \mathbb{R}_+$ and a closed set $C \subseteq \mathbb{R}^n$ such that*

$$\Phi(v) = 0 \iff v \in C.$$

Define the function $f : \mathbb{R}^{nl} \to \mathbb{R}_+$ by

$$f(y) := \sum_{i=1}^l \Phi(y_i),$$

where $y = (y_i)_{i=1}^l$ and $y_i \in \mathbb{R}^n$. Then for $y^* \in \times_{i=1}^l C$ it holds that

$$\partial^> f(y^*) \subseteq \bigcup_{j=1}^l \partial\Phi(y_1^*) \times \ldots \times \partial^>\Phi(y_j^*) \times \cdots \times \partial\Phi(y_l^*).$$

Proof. Note that we have $\partial^> f(y^*) \subseteq \partial f(y^*) = \times_{j=1}^l \partial\Phi(y_j^*)$, where the equality is due to [10, Proposition 10.5].

Now, take $\xi \in \partial^> f(y^*)$. By definition, there exist sequences $\{y^k\} \to y^*$ such that $f(y^k) > 0$ and $\psi^k \in \partial f(y^k)$ for all $k \in \mathbb{N}$ with $\xi = \lim_{k\to\infty} \psi^k$. Due to the fact that the index set $\{1, \ldots, l\}$ is finite, there exists an index $j \in \{1, \ldots, l\}$ and a subsequence $\{y^k\}_{k\in K}$ such that (without relabelling) $y_j^k \notin C$ for all k. This proves the result, since we have $\psi_j \in \partial^>\Phi(y_j^*)$ then. $\quad\square$

Due to the previous two results, we can infer that for $x^* \in X$ we have (recall that $\psi = \vartheta \circ F^{VC}$)

$$\partial^>\psi(x^*) \overset{\text{Proposition 6.4}}{\subseteq} (F^{VC})'(x^*)^T \partial^>\vartheta(F^{VC}(x^*))$$

$$\overset{\text{Proposition 6.5}}{\subseteq} (F^{VC})'(x^*)^T \bigcup_{j=1}^l \partial\varphi(F_1^{VC}(x^*)) \times \cdots \times \partial^>\varphi(F_j^{VC}(x^*)) \times \cdots, \times \partial\varphi(F_l^{VC}(x^*))$$

$$= \bigcup_{j=1}^l \left\{ \sum_{i=1}^l \alpha_i \nabla G_i(x^*) + \beta_i \nabla H_i(x^*) \;\middle|\; \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \in \partial^>\varphi(F_j^{VC}(x^*)), \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \in \partial\varphi(F_i^{VC}(x^*))\ (i \neq j) \right\}. \tag{38}$$

It is hence of particular interest to know the outer and the limiting subdifferentials of φ at points from the set C. This information is provided by the following two lemmas.

Lemma 6.6. *Consider the function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ from (31), and the set $C \subseteq \mathbb{R}^2$ from (17) and let $(a, b) \in C$. Then it holds that*

$$\partial^>\varphi(a, b) = \begin{cases} \emptyset & \text{if } b > 0, a < 0, \\ \begin{pmatrix} b \\ 0 \end{pmatrix} & \text{if } b > 0, a = 0, \\ \begin{pmatrix} 0 \\ a-1 \end{pmatrix} & \text{if } b = 0, a < 0, \\ \left\{ \begin{pmatrix} 0 \\ a \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\} & \text{if } b = 0, a > 0, \\ \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\} & \text{if } a = b = 0. \end{cases} \tag{39}$$

Proof. In view of Definition 6.2, we are interested in the limiting subdifferential of φ at points $(a, b) \notin C$. For these purposes, we claim that

$$\partial\varphi(a, b) = \begin{cases} \begin{pmatrix} b \\ a \end{pmatrix} & \text{if } a, b > 0, \\ \begin{pmatrix} 0 \\ -1 \end{pmatrix} & \text{if } a > 0, b < 0, \\ \begin{pmatrix} b \\ a - 1 \end{pmatrix} & \text{if } a < 0, b < 0, \\ \left\{ \begin{pmatrix} v \\ -1 \end{pmatrix} \middle| v \in [b, 0] \right\} & \text{if } a = 0, b < 0. \end{cases} \tag{40}$$

The first three cases of this formula are easily seen due to the fact that φ is smooth in a neighbourhood of (a, b) and thus, one has $\partial\varphi(a, b) = \{\nabla\varphi(a, b)\}$. The case $a = 0, b < 0$ can be verified as follows: The function φ is regular in the sense of [17, Def. 2.3.4], as was already noted in [7, Lem. 3.3]. Therefore, the limiting subdifferential of this function coincides with the Clarke subdifferential, for which the corresponding formulas are also given in [7, Lem. 3.3].

We are now in a position to prove the formula for the outer subdifferential. For these purposes, consider the five relevant cases separately and recall that $\varphi(a, b) > 0$ if and only if $(a, b) \notin C$.

(i) $b > 0, a < 0$: In this case, there exists no sequence $(a_k, b_k) \to (a, b)$ with $(a_k, b_k) \notin C$ and so $\partial^>\varphi(a, b) = \emptyset$.

(ii) $b > 0, a = 0$: For a sequence $(a_k, b_k) \to (a, b)$ with $(a_k, b_k) \notin C$ one has $a_k, b_k > 0$. Hence it follows that if $s^k \in \partial\varphi(a_k, b_k)$, we see from (40) that $s^k = \begin{pmatrix} b_k \\ a_k \end{pmatrix}$ and hence $\lim_{k\to\infty} s^k = \begin{pmatrix} b \\ 0 \end{pmatrix}$.

(iii) $b = 0, a < 0$: If $(a_k, b_k) \to (a, b)$ and $(a_k, b_k) \notin C$, there remains the case $a_k, b_k < 0$. Hence, we have $s_k = \begin{pmatrix} b_k \\ a_k - 1 \end{pmatrix}$ for $s^k \in \partial\varphi(a_k, b_k)$. This yields $\lim_{k\to\infty} s^k = \begin{pmatrix} 0 \\ a - 1 \end{pmatrix}$.

(iv) $b = 0, a > 0$: Here, if $(a_k, b_k) \to (a, b)$ and $(a_k, b_k) \notin C$, we have $a_k, b_k > 0$ or $a_k > 0, b_k < 0$ and hence $s^k \in \left\{ \begin{pmatrix} b_k \\ a_k \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}$ which implies $\lim_{k\to\infty} s^k \in \left\{ \begin{pmatrix} 0 \\ a \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}$.

(v) $a = b = 0$: In this case, for a sequence $(a_k, b_k) \to (a, b)$ there may occur all cases of $(a_k, b_k) \in \mathbb{R}^2 \setminus C$. Hence if $s^k \in \partial\varphi(a_k, b_k)$ one has $s^k \in \left\{ \begin{pmatrix} b_k \\ a_k \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} b_k \\ a_k - 1 \end{pmatrix}, \begin{pmatrix} v_k \\ -1 \end{pmatrix} \right\}$ for $v_k \in [b_k, 0]$. Hence one obtains $\lim_{k\to\infty} s^k \in \left\{ \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$.

Altogether, this completes the proof. $\quad\square$

Lemma 6.7. *Consider the function* $\varphi : \mathbb{R}^2 \to \mathbb{R}$ *from* (31)*, the set* $C \subseteq \mathbb{R}^2$ *from* (17)*, and let* $(a, b) \in C$*. Then we have*

$$\partial\varphi(a, b) = \begin{cases} \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} & \text{if } b > 0, a < 0, \\ \operatorname{conv}\left\{ \begin{pmatrix} b \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} & \text{if } b > 0, a = 0, \\ \operatorname{conv}\left\{ \begin{pmatrix} 0 \\ a \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\} & \text{if } b = 0, a > 0, \\ \operatorname{conv}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ a - 1 \end{pmatrix} \right\} & \text{if } b = 0, a < 0, \\ \operatorname{conv}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\} & \text{if } a = b = 0. \end{cases}$$

Proof. Similar to the proof of the previous result, we recall an observation from [7] that the mapping φ is regular in the sense of Clarke, hence the limiting subdifferential is identical with the generalized gradient by Clarke, for which the corresponding representations can be found in [7]. $\quad\square$

At least, the foregoing results allow us to state some kind of sufficient condition for exactness of the l_1-penalty function.

Corollary 6.8. *Let* x^* *be feasible for* (1) *such that* $I_{00} = \emptyset$ *and MPVC-LICQ holds at* x^**. Then the penalty function from* (30) *is exact at* x^**.*

Proof. Due to Proposition 6.3, it suffices to show that, under the above assumptions, we have $0 \notin \partial^>\psi(x^*)$. For these purposes, suppose that $0 \in \partial^>\psi(x^*)$, then by (38) there exists $j \in \{1, \ldots, l\}$ such that

$$0 = \sum_{i=1}^{l} \alpha_i \nabla G_i(x^*) + \beta_i \nabla H_i(x^*),$$

where

$$(\alpha_i, \beta_i)^T \in \partial\varphi(F_i^{VC}(x^*)) \quad (i \neq j), \qquad (\alpha_j, \beta_j)^T \in \partial^>\varphi(F_j^{VC}(x^*)).$$

Due to the MPVC-LICQ assumption and Lemma 6.7, we obtain $(\alpha_i, \beta_i) = (0, 0)$ for all $i = 1, \ldots, l$. In particular, we have $(\alpha_j, \beta_j) = (0, 0)$ which contradicts the fact that $(\alpha_j, \beta_j)^T \in \partial^>\varphi(F_j^{VC}(x^*))$ as we have $j \notin I_{00}$; cf. Lemma 6.6. $\square$

It is currently not known whether the previous result holds without the assumption $I_{00} = \emptyset$. The current technique of proof does not allow us to verify this statement since the outer subdifferential of φ for indices $i \in I_{00}$ contains the zero vector, whereas all other outer subdifferentials are either empty or consist of nonzero elements; cf. Lemma 6.7. In order to avoid this problem, one needs a smaller estimate for the outer subdifferential of ψ than the one derived in (38). This, however, is a nontrivial task, because it requires a more refined analysis of the configuration of $\text{Im}(F)$ and Λ (in the notation of (10)). In any case, we know that MPVC-MFCQ cannot be a sufficient condition for the l_1-penalty function to be exact (see the corresponding discussion at the beginning of this section), in contrast to the more specialized (MPVC-tailored) penalty function considered in Section 4.

7. Final remarks

This paper gives exact penalty results for mathematical programs with vanishing constraints (MPVCs). In particular, it shows exactness for a new, MPVC-tailored penalty function under suitable conditions which, on the other hand, do not guarantee exactness of the well-known l_1 (or l_p with $p \in [1, \infty]$) penalty function. In fact, it is currently an open question under which assumptions the l_1-penalty function is exact in the MPVC-context if we do not want to assume that the bi-active index set I_{00} is empty at a local minimum.

We believe that our new penalty function can be used not only as a theoretical tool (like the derivation of optimality conditions, as shown in this paper), but also from a practical point of view, especially for the globalization of suitable (locally convergent) algorithms.

Acknowledgement

References

[1] Z.-Q. Luo, J.-S. Pang, D. Ralph, Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, UK, 1996.
[2] J.V. Outrata, M. Kočvara, J. Zowe, Nonsmooth Approach to Optimization Problems with Equilibrium Constraints, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
[3] W. Achtziger, C. Kanzow, Mathematical programs with vanishing constraints: Optimality conditions and constraint qualifications, Mathematical Programming 114 (2008) 69–99.
[4] T. Hoheisel, C. Kanzow, On the Abadie and Guignard constraint qualification for mathematical progammes with vanishing constraints, Optimization 58 (2009) 431–448.
[5] T. Hoheisel, C. Kanzow, Stationary conditions for mathematical programs with vanishing constraints using weak constraint qualifications, Journal of Mathematical Analysis and Applications 337 (2008) 292–310.
[6] T. Hoheisel, C. Kanzow, First- and second-order optimality conditions for mathematical programs with vanishing constraints, Applications of Mathematics 52 (2007) 495–514. Special issue dedicated to J.V. Outrata's 60th birthday.
[7] W. Achtziger, T. Hoheisel, C. Kanzow, A smoothing-regularization approach to mathematical programs with vanishing constraints, Preprint 284, Institute of Mathematics, University of Würzburg, Würzburg, November 2008.
[8] A.F. Izmailov, M.V. Solodov, Mathematical programs with vanishing constraints: Optimality conditions, sensitivity and a relaxation method, Journal of Optimization Theory and Applications 142 (2009) 501–532.
[9] D.W. Peterson, A review of constraint qualifications in finite-dimensional spaces, SIAM Review 15 (1973) 639–654.
[10] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, in: A Series of Comprehensive Studies in Mathematics, vol. 317, Springer, Berlin, Heidelberg, 1998.
[11] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation I. Basic Theory, in: A Series of Comprehensive Studies in Mathematics, vol. 330, Springer, Berlin, Heidelberg, 2006.
[12] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation II. Applications, in: A Series of Comprehensive Studies in Mathematics, vol. 331, Springer, Berlin, Heidelberg, 2006.
[13] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[14] M.L. Flegel, C. Kanzow, J.V. Outrata, Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints, Set-Valued Analysis 15 (2007) 139–162.
[15] J.V. Burke, Calmness and exact penalization, SIAM Journal on Control and Optimization 29 (1991) 493–497.
[16] J.V. Burke, An exact penalization viewpoint of constrained optimization, SIAM Journal on Control and Optimization 29 (1991) 968–998.
[17] F.H. Clarke, Optimization and Nonsmooth Analysis, John Wiley & Sons, New York, 1983, Reprinted by SIAM, Philadelphia, PA, 1990.
[18] J.-S. Pang, Error bounds in mathematical programming, Mathematical Programming 79 (1997) 299–332.
[19] R. Henrion, J.V. Outrata, Calmness of constraint systems with applications, Mathematical Programming 104 (2005) 437–464.
[20] O.L. Mangasarian, Nonlinear Programming, McGraw Hill, New York, 1969, Reprinted by SIAM, Philadelphia, PA, 1994.
[21] J.M. Borwein, A.S. Lewis, Convex analysis and nonlinear optimization, in: Theory and Examples, in: CMS Books in Mathematics, Springer-Verlag, New York, 2000.
[22] S.-P. Han, O.L. Mangasarian, Exact penalty functions in nonlinear programming, Mathematical Programming 17 (1979) 251–269.
[23] A.D. Ioffe, J.V. Outrata, On metric and calmness qualification conditions in subdifferential calculus, Set-Valued Analysis 16 (2008) 199–227.