Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

On calmness conditions in convex bilevel programming

René Henrion[1], Thomas Surowiec[2]

submitted: April 20, 2010

[1] Weierstrass Institute
for Applied Analysis and Stochastics
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: rene.henrion@wias-berlin.de

[2] Humboldt University Berlin
Institute of Mathematics
Unter den Linden 6
10099 Berlin
Germany
E-Mail: surowiec@math.hu-berlin.de

No. 1508
Berlin 2010

2000 *Mathematics Subject Classification.* 90C30, 49J53.

Key words and phrases. Bilevel programming, partial calmness, M-stationarity, value function, calmness, discrete obstacle problem.

This work was supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

In this article we compare two different calmness conditions which are widely used in the literature on bilevel programming and on mathematical programs with equilibrium constraints. In order to do so, we consider convex bilevel programming as a kind of intersection between both research areas. The so-called partial calmness concept is based on the function value approach for describing the lower level solution set. Alternatively, calmness in the sense of multifunctions may be considered for perturbations of the generalized equation representing the same lower level solution set. Both concepts allow to derive first order necessary optimality conditions via tools of generalized differentiation introduced by Mordukhovich. They are very different, however, concerning their range of applicability and the form of optimality conditions obtained. The results of this paper seem to suggest that partial calmness is considerably more restrictive than calmness of the perturbed generalized equation. This fact is also illustrated by means of a dicretized obstacle control problem.

1 Introduction

Optimization problems of the abstract form

$$\min \{F(x,y) | y \in S(x),\, x \in \Omega\} \tag{1}$$

provide a fairly general framework for interesting subclasses like *bilevel programs* (BLPs) and *mathematical programs with equilibrium constraints* (MPECs). Here, $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is some objective function, $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a multifunction and $\Omega \subseteq \mathbb{R}^n$ is a fixed subset. More precisley, a BLP arises from (1) by specifying the multifunction S as the solution mapping associated with some parameter dependent optimization problem

$$S(x) := \arg\min_y \{f(x,y) | y \in C(x)\}, \tag{2}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is some objective function and $C : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ a feasible set mapping. The optimization problem defined in (2) is also called the *lower level problem* associated with the *upper level problem* (1) via the upper level decision variable x which acts as a parameter in (2). By contrast, an MPEC would be obtained from (1) by setting

$$S(x) := \left\{ y \in \mathbb{R}^m | 0 \in g(x,y) + N_{C(x)}(y) \right\}, \tag{3}$$

where $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, $C : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and 'N' represents some appropriate normal cone.

It is well-known that problem (1) with S either given by (2) or by (3) violates the standard constraint qualifications of nonlinear programming. Therefore, many attempts have been made to derive first order necessary optimality conditions for (1) under appropriate assumptions. A major tool employed for this purpose is the generalized differential calculus (subdifferential, normal cone, coderivative) of Mordukhovich, see [8]. Typically, the application of these calculus rules requires some kind of Lischitz like behavior (e.g., Aubin property or calmness) of certain multifunctions. Two basically different approaches have developed in the past: one is related with BLPs and uses the concept of *partial calmness* which is closely connected with the so called value function description of the solution set (2) (e.g., [10],[15],[17],[2]). The second approach relies on the MPEC structure and considers *calmness* of a perturbation mapping associated with the generalized equation inside (3) (e.g., [16],[11],[9],[4]). The aim of this paper is to compare both approaches in a setting where BLPs can be reformulated as MPECs.

The mapping (2) can be rephrased in the form (3) in case f is convex and continuously differentiable with respect to y and $C(x)$ is a closed convex set for all x. Indeed, in this convex setting, the solution set of the optimization problem (2) is equivalently described by the generalized equation

$$0 \in \nabla_y f(x, y) + N_{C(x)}(y),$$

where 'N' is the normal cone in the sense of convex analysis. For non-convex data, this last generalized equation would describe a set larger than the solutions of (2) and thus, BLPs can no longer be recast as MPECs. Conversely, MPECs cannot be formulated as BLPs either, since the mapping g in (3) may not be representable as the gradient of some function f.

In this paper, we consider bilevel programs with convex lower level problems of the type

$$\min \{F(x, y) | y \in S(x)\}, \tag{4}$$

where

$$\begin{aligned} S(x) \quad &: \quad = \arg\min_y \{f(x, y) | y \in C\}, \\ C \quad &: \quad = \{y \in \mathbb{R}^m | g_i(y) \le 0 \quad (i = 1, \dots, p)\}, \end{aligned}$$

$F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a continuously differentiable upper level objective function;
$f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a twice continuously differentiable lower level objective function;
$g : \mathbb{R}^m \to \mathbb{R}^p$ is a twice continuously differentiable lower level constraint mapping.
Moreover, we suppose that f is convex in y and the components g_i of g are also convex.

In contrast to the more general settings in (1), (2) and (3), we consider neither upper level constraints $x \in \Omega$ nor lower level constraints that depend on x. This relatively harmless simplification allows a considerable reduction in terms of notational effort, and therefore, allows us to focus the analysis on the essential points.

2 Basic Concepts and Notation

We commence by recalling some basic properties of a multifunction $\Phi : X \rightrightarrows Y$ between metric spaces X, Y. The graph of Φ is defined as

$$\text{gr } \Phi := \{(x, y) \in X \times Y | y \in \Phi(x)\}.$$

The inverse of Φ is a multifunction $\Phi^{-1} : Y \rightrightarrows X$ defined by

$$\Phi^{-1}(y) := \{x \in X | y \in \Phi(x)\}.$$

We shall make use of the following notions of Lipschitz continuity for multifunctions:

Definition 2.1. *Let $\Phi : X \rightrightarrows Z$ be a multifunction between metric spaces X, Z and consider a point, $(\bar{x}, \bar{z}) \in \text{gr } \Phi$. Then Φ is said to have the **Aubin property** at $(\bar{x}, \bar{z})$ provided there exists neighborhoods $\mathcal{U}$ of $\bar{x}$ and $\mathcal{V}$ of $\bar{z}$, and a constant $L > 0$ such that*

$$d(z, \Phi(x'')) \leq L||x' - x''|| \quad \forall z \in \mathcal{V} \cap \Phi(x') \; \forall x', x'' \in \mathcal{U}.$$

*Here, $d(z, \Phi(\bar{x})) = \inf_{z' \in \Phi(\bar{x})} d(z, z')$. A weaker Lipschitz property is obtained when fixing one of the x-arguments as $\bar{x}$. More precisely we say that Φ is **calm** at $(\bar{x}, \bar{z})$ provided there exists neighborhoods $\mathcal{U}$ of $\bar{x}$ and $\mathcal{V}$ of $\bar{z}$, and a constant $L > 0$ such that*

$$d(z, \Phi(\bar{x})) \leq L||x' - \bar{x}||, \; \forall z \in \mathcal{V} \cap \Phi(x'), \forall x' \in \mathcal{U}.$$

In the following, we collect some known results about the Aubin property and calmness of certain multifunctions we shall make use of in this paper. We start with

Theorem 2.2 (([6], Th. 3.6)). *Let $T_1 : X_1 \rightrightarrows X$ and $T_2 : X_2 \rightrightarrows X$ be multifunctions between metric spaces X_1, X_2, X. If T_1 is calm at $(x_1, x) \in \text{gr}\, T_1$, T_2 is calm at $(x_2, x) \in \text{gr}\, T_2$, T_2^{-1} has the Aubin property at (x, x_2) and $T_1(x_1) \cap T_2(\cdot)$ is calm at (x_2, x), then the multifunction $T_1 \cap T_2 : X_1 \times X_2 \rightrightarrows X$ defined by $(T_1 \cap T_2)(x_1, x_2) := T_1(x_1) \cap T_2(x_2)$ is calm at (x_1, x_2, x).*

Next, we consider a special multifunction $\Phi : \mathbb{R}^p \rightrightarrows \mathbb{R}^m$ defined by the parametric inequality system

$$\Phi(w) := \{y \in \mathbb{R}^m | h_i(y) \leq w_i \quad (i = 1, \ldots, p)\},$$

where $h : \mathbb{R}^m \to \mathbb{R}^p$ is a continuously differentiable mapping. The following fact is well-known:

Proposition 2.3. *If for some $(\bar{w}, \bar{y}) \in \text{gr}\, \Phi$ the Jacobian $Dh(\bar{y})$ is surjective, then Φ has the Aubin property at $(\bar{w}, \bar{y})$.*

Concerning the weaker calmness property of Φ, the following result holds true:

Proposition 2.4 (([3], Prop.1)). *If Φ is calm at $(\bar{w}, \bar{y}) \in \text{gr}\, \Phi$, then the contingent and linearized cones of $\Phi(\bar{w})$ at $\bar{y}$ coincide, i.e., $T_{\Phi(\bar{w})}(\bar{y}) = L_{\Phi(\bar{w})}(\bar{y})$, where*

$$\begin{aligned} T_{\Phi(\bar{w})}(\bar{y}) \; &: = \{d \in \mathbb{R}^m | \exists t_n \downarrow 0 \; \exists d_n \to d : \bar{y} + t_n d_n \in \Phi(\bar{w})\}, \\ L_{\Phi(\bar{w})}(\bar{y}) \; &: = \{d \in \mathbb{R}^m | \langle \nabla h_i(\bar{y}), d \rangle \leq 0 \quad \forall i \in I\}, \\ I \; &: = \{i \in \{1, \ldots, p\} | h_i(\bar{y}) = \bar{w} \quad \forall i \in I\}. \end{aligned}$$

In the context of bilevel programming the following calmness condition has been introduced in [15]:

Definition 2.5. *The bilevel program (4) is called* ***partially calm*** *at one of its local solutions* $(\bar{x}, \bar{y})$, *if there exist a neighborhood* $\mathcal{U}$ *of* $(0, \bar{x}, \bar{y})$ *and a constant* $L > 0$ *such that*

$$F(x, y) - F(\bar{x}, \bar{y}) + L|u| \geq 0 \quad \forall (u, x, y) \in \mathcal{U} : f(x, y) - \varphi(x) + u = 0, \, y \in C.$$

Here,

$$\varphi(x) := \inf \{f(x, y) | y \in C\} \tag{5}$$

denotes the optimal value function of the lower level problem in (4).

The following two definitions provide further important characterizations of the behaviour of multifunctions:

Definition 2.6. *A multifunction* $\Phi : X \rightrightarrows Y$ *between metric spaces* X, Y *is called* ***inner semicompact*** *at* $\bar{x} \in X$ *with* $\Phi(\bar{x}) \neq \emptyset$ *if for every sequence* $x_n \to \bar{x}$ *with* $\Phi(x_n) \neq \emptyset$ *there is a sequence* $y_n \in \Phi(x_n)$ *containing a bounded subsequence.*

Definition 2.7. *A multifunction* $\Phi : X \rightrightarrows Y$ *between metric spaces* X, Y *is called* ***inner semicontinuous*** *at* $(\bar{x}, \bar{y}) \in \operatorname{gr} \Phi$ *if for any sequence* $x_n \to \bar{x}$ *there is a sequence* $y_n \in \Phi(x_n)$ *with* $y_n \to \bar{y}$.

At the end of this section we recall some basic concepts from nonlinear programming. To this aim consider for an arbitrarily fixed $x \in \mathbb{R}^n$ the lower level problem

$$\min \{f(x, y) | g_i(y) \leq 0 \quad (i = 1, \ldots, p)\}. \tag{6}$$

associated with (4). The set of active indices at some $y \in C$ is defined as

$$I(y) := \{i \in \{1, \ldots, p\} \, | g_i(y) = 0\}. \tag{7}$$

A common constraint qualification for the set C is the following:

Definition 2.8. *The* ***Linear Independence Constraint Qualification*** *(LICQ) is said to hold at* $y \in C$ *if the set of gradients* $\{\nabla g_i(y)\}_{i \in I(y)}$ *is linearly independent.*

Note that in the case of convex functions g_i which we consider here, LICQ implies the existence of a Slater point for C, i.e., there is some $y^*$with $g_i(y^*) < 0$ for $i = 1, \ldots, p$. Sometimes a weaker constraint qualification than LICQ is sufficient to be required:

Definition 2.9. *The* ***Constant Rank Constraint Qualification*** *(CRCQ) is said to hold at* $y \in C$ *if there exists a neighbourhood* $\mathcal{U}$ *of* $\bar{y}$ *such that*

$$\operatorname{rank} \{\nabla g_i(y)\}_{i \in I} = \operatorname{rank} \{\nabla g_i(y')\}_{i \in I} \quad \forall y' \in \mathcal{U} \, \forall I \subseteq I(y).$$

Evidently, LICQ implies CRCQ, whereas the converse does not hold true. By

$$L(x, y, \lambda) := f(x, y) + \sum\nolimits_{i=1}^{p} \lambda_i g_i(y)$$

we denote the Lagrangean of problem (6). Now, fix any $(\bar{x}, \bar{y}) \in \operatorname{gr} S$. Then, $\bar{y}$ is a solution to (6) with $x = \bar{x}$, and hence, if LICQ is satisfied at $\bar{y}$, then there exists a unique multiplier $\lambda \in \mathbb{R}^p_+$ such that $\nabla_y L(\bar{x}, \bar{y}, \lambda) = 0$ and $\lambda_i g_i(\bar{y}) = 0$ for $i = 1, \ldots, p$.

Definition 2.10. *The **Strong Second-Order Sufficient Condition** (SSOSC) is said to hold at a solution $\bar{y}$ to (6) if for all $\lambda \geq 0$ satisfying $\nabla_y L(\bar{x}, \bar{y}, \lambda) = 0$ the Hessian $\nabla_y^2 L(\bar{x}, \bar{y}, \lambda)$ is positive definite on the subspace*

$$\bigcap\nolimits_{i=1}^{p} \ker \{\lambda_i \nabla g_i(\bar{y})\}.$$

Note that the multiplier λ in the previous definition is unique in case that LICQ holds at $\bar{y}$.

3 A comparison of partial calmness and calmness of the perturbation mapping

3.1 M-stationarity conditions

In this paper we want to compare two different calmness concepts leading to different necessary optimality conditions (stationarity in the sense of Mordukhovich or M-stationarity) for solutions of (4) in dual form. The following Theorem is a reduction of Theorems 3.1 and 5.1 in [2] to the setting considered here. It strongly relies on the concept of partial calmness as introduced in Def. 2.5.

Theorem 3.1. *Let $(\bar{x}, \bar{y})$ be a local solution to (4). Assume the following conditions:*

- *There exists some y^* with $g_i(y^*) < 0 \quad (i = 1, \ldots, p)$ (i.e. y^* is a Slater point for the lower level constraint set C).*
- *(4) is partially calm at $(\bar{x}, \bar{y})$.*
- *The solution set mapping S of the lower level problem in (4) is inner semicontinuous at $(\bar{x}, \bar{y})$ (see Def. 2.7) or inner semicompact at $\bar{x}$ (see Def. 2.6)*

Then, there are real numbers $\alpha > 0$, $\lambda_i \geq 0$, $\mu_i \geq 0$ $(i = 1, \ldots, p)$ such that

$$\begin{aligned} \nabla_x F(\bar{x}, \bar{y}) &= 0 \\ \nabla_y F(\bar{x}, \bar{y}) + \alpha \nabla_y f(\bar{x}, \bar{y}) + \sum\nolimits_{i=1}^{p} \mu_i \nabla g_i(\bar{y}) &= 0 \\ \nabla_y f(\bar{x}, \bar{y}) + \sum\nolimits_{i=1}^{p} \lambda_i \nabla g_i(\bar{y}) &= 0 \\ \lambda_i g_i(\bar{y}) = \mu_i g_i(\bar{y}) &= 0 \quad (i = 1, \ldots, p). \end{aligned}$$

It is easily seen that the last three relations can be equivalently aggregated in order to yield the following

Corollary 3.2. *Under the assumptions of Theorem 3.1 it holds that*

$$\nabla_x F(\bar{x}, \bar{y}) = 0 \quad \text{and} \quad \nabla_y F(\bar{x}, \bar{y}) \in \text{span}\, \{\nabla g_i(\bar{y})\}_{i \in I(\bar{y})},$$

where $I(\bar{y})$ is defined in (7).

In contrast to Theorem 3.1, a different system of necessary optimality conditions is derived in the following theorem via calmness of a perturbation mapping associated with the generalized equation $0 \in \nabla_y f(x,y) + N_C(y)$ which describes the lower level solution set mapping S of (4).

Theorem 3.3. *Let $(\bar{x}, \bar{y})$ be a local solution to (4). Assume the following conditions:*

- *There exists some y^* with $g_i(y^*) < 0 \quad (i = 1, \dots, p)$.*
- *The Constant Rank Constraint Qualification (CRCQ) is satisfied (see Def. 2.9).*
- *The perturbation mapping*
$$M(v) := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m | v \in \nabla_y f(x,y) + N_C(y)\} \tag{8}$$
is calm at $(0, \bar{x}, \bar{y})$ (see Def. 2.1).

Then, there are multipliers $\lambda \geq 0$, v^* and w^* such that

$$\begin{array}{rcl}
0 &=& \nabla_x F(\bar{x},\bar{y}) + \nabla_x^T \nabla_y f(\bar{x},\bar{y}) v^* \\
0 &=& \nabla_y F(\bar{x},\bar{y}) + \nabla_{yy} f(\bar{x},\bar{y}) v^* + \left(\sum_{i\in I(\bar{y})} \lambda_i \nabla^2 g_i(\bar{y})\right) v^* + \sum_{i\in I(\bar{y})} w_i^* \nabla g_i(\bar{y}) \\
0 &=& \langle \nabla g_i(\bar{y}), v^* \rangle \quad \forall i \in I(\bar{y}) : \lambda_i > 0 \\
0 &=& w_i^* \quad \forall i \in I(\bar{y}) : \lambda_i = 0, \ \langle \nabla g_i(\bar{y}), v^* \rangle < 0 \\
0 &\leq& w_i^* \quad \forall i \in I(\bar{y}) : \lambda_i = 0, \ \langle \nabla g_i(\bar{y}), v^* \rangle > 0 \\
0 &=& \nabla_y f(\bar{x},\bar{y}) + \sum_{i\in I(\bar{y})} \lambda_i \nabla g_i(\bar{y}) .
\end{array}$$

The theorem was proved in [14] (Cor. 4.1) under the stronger LICQ (see Def. 2.8) (see also [4], Th. 5.2 and 5.3 in a concrete setting). The possibility to weaken LICQ to CRCQ (in which case the multiplier λ is not necessarily unique) was shown in [14] (Cor. 6.2). We note that a very similar version of this theorem was earlier proven in [18], Th. 5.1, albeit under a stronger calmness condition not related to the mapping (8) but rather to an enhanced mapping including Lagrange multipliers.

It is evident that the stationarity conditions of Theorems 3.1 and 3.3 differ considerably and that the main reason for this is the different nature of the calmness conditions used (the other assumptions playing a subordinate technical role). The most striking difference concerning the results is the lack of second order terms in Theorem 3.1. This observation seems to suggest a quite restrictive role played by the partial calmness condition in Theorem 3.1. Consider for a moment a situation, where the lower level problem of (4) has no constraints (i.e., $C = \mathbb{R}^m$). Then, the inequality constraints $g_i(y) \leq 0$ are not present and the stationarity conditions of Theorem 3.1 reduce to

$$0 = \nabla F(\bar{x},\bar{y}), \ 0 = \nabla_y f(\bar{x},\bar{y}), \tag{9}$$

whereas Theorem 3.3 still yields the much richer information of the existence of v^* such that

$$\begin{array}{rcl}
0 &=& \nabla_x F(\bar{x},\bar{y}) + \nabla_x^T \nabla_y f(\bar{x},\bar{y}) v^* \\
0 &=& \nabla_y F(\bar{x},\bar{y}) + \nabla_{yy} f(\bar{x},\bar{y}) v^* \\
0 &=& \nabla_y f(\bar{x},\bar{y})
\end{array} \tag{10}$$

The following example illustrates the difference:

Example 3.4. *Consider the following specific instance of problem (4):*

$$\min\left\{x^2+cy|y\in S(x)\right\},\quad S(x):=\arg\min\left\{\frac{1}{2}y^2-xy\right\},$$

where $c\in\mathbb{R}$ is some parameter. The solution of this problem is easily found in a direct way as $(\bar{x},\bar{y})=\left(-\frac{c}{2};-\frac{c}{2}\right)$. Using (9), one would arrive at the stationary solution $(\bar{x},\bar{y})=(0,0)$ in case of $c=0$ and at no stationary solution at all if $c\neq 0$. Thus, (9) are not stationarity conditions in the typical case $c\neq 0$. In other words, the partial calmness condition of Theorems 3.1 is violated in the typical situation (whereas the other technical assumptions of the Theorem hold true). Applying (10) instead, the conditions excatly identify the solution of the problem for all $c\in\mathbb{R}$. Let us look more precisely at the calmness conditions themselves. If partial calmness was satisfied at a solution $(\bar{x},\bar{y})=\left(-\frac{c}{2};-\frac{c}{2}\right)$, then, according to Definition 2.5, there should exist a neighborhood $\mathcal{U}$ of $(0,\bar{x},\bar{y})$ and a constant $L>0$ such that

$$x^2+cy-(\frac{c^2}{4}-\frac{c^2}{2})+L|u|\geq 0\quad\forall\,(u,x,y)\in\mathcal{U}:\frac{1}{2}y^2-xy-\left(\frac{1}{2}x^2-x^2\right)+u=0.\quad(11)$$

Then, we may define a sequence $(u_n,x_n,y_n):=\left(-\frac{1}{2}n^{-2},-\frac{c}{2},-\frac{c}{2}-n^{-1}\right)\to(0,\bar{x},\bar{y})$ satisfying $\frac{1}{2}(x_n-y_n)+u_n=0$ but

$$x_n^2+cy_n+\frac{c^2}{4}+L|u_n|=-cn^{-1}+\frac{L}{2}n^{-2}<0\quad\text{for}\quad n>\frac{L}{2c}.$$

This is a contradiction with (11) and so partial calmness is violated for $c\neq 0$. As far as the perturbation mapping M (8) is concerned, it follows from $N_C(y)=\{0\}$ (due to absence of lower level constraints) that

$$M(v):=\{(x,y)|v\in y-x\}.$$

Hence, M is a continuous multifunction and in particular calm in the sense of Definition 2.1.

At the end of this section, we provide three independent conditions ensuring the calmness of the multifunction (8) needed in Theorem 3.3:

Theorem 3.5. *The multifunction (8) is calm at $(0,\bar{x},\bar{y})$, where $(\bar{x},\bar{y})$ is a local solution to (4), if one of the following three conditions is satisfied in problem (4):*

1. *The lower level constraint set C is a polyhedron and the lower level objective has the form $f(x,y)=y^T(Ay+Bx)$for matrices A,B of appropriate size.*

2. *The lower level problem satisfies LICQ and SSOSC (see Def. 2.8 and Def. 2.10).*

3. *The lower level objective satisfies that $\nabla^2_{xy}f(\bar{x},\bar{y})$ is surjective ('ample parameterization')*

Proof. Under the first condition, we may write the graph of M from (8) as

$$\text{gr } M = T^{-1}(\text{gr } N_C), \quad \text{where} \quad T(v,x,y) := \left(y, v - \frac{1}{2}Ay - Bx\right).$$

Since C is a polyhedron, gr N_C is a union of polyhedra and so is gr M as its preimage under a linear mapping. In other words, M is a polyhedral multifunction and as such it is calm as a consequence of a well-known theorem by Robinson [13]. Conditions 2. and 3. both imply actually the stronger Aubin property of M which entails calmness (see Def. 2.1). This was proved for condition 2. in [14] (Cor. 5.1, see also [4] (Prop. 5.1)) and follows for condition 3. from Mordukhovich's criterion for the Aubin property of multifunctions [8]. □

3.2 Uniformly weak sharp minima and value function constraint qualification

When comparing the partial calmness and calmness of the perturbation mapping in Theorems 3.1 and 3.3 it turns out that partial calmness is not a constraint qualification in the strict sense because it mixes the properties of the upper level objective function with properties of the constraint $y \in S(x)$. As a consequence, it is not sufficient to check that the constraints behave well in order to derive stationarity conditions for a whole class of objective functions. Rather, one has to check partial calmness in (2.5) for each instance of an objective function F anew. A way to formulate sufficient conditions for partial calmness independent of the concrete nature of the upper level objective is the use of uniformly weak sharp minima as proposed in [15]:

Definition 3.6. *The parametric lower level problem*

$$\min\{f(x,y)|y \in C\} \tag{12}$$

of (4) is said to have a uniformly weak sharp minimum around one of its solutions $\bar{y}$ at $x := \bar{x}$ if there exist $\varepsilon, \alpha > 0$ such that

$$f(x,y) - \varphi(x) \geq \alpha d(y, S(x)) \quad \forall x \in \mathbb{B}(\bar{x}, \varepsilon)\ \forall y \in C \cap \mathbb{B}(\bar{y}, \varepsilon).$$

Here, φ refers to the optimal value function (5) of the lower level problem and $\mathbb{B}(\bar{x}, \varepsilon)$ is the closed ball with radius ε around $\bar{x}$.

It can be shown ([15], Prop. 5.1) that for any locally Lipschitz function F the partial calmness condition in (2.5) is satisfied at a local solution $(\bar{x}, \bar{y})$ of (4) provided that the lower level problem has a uniformly weak sharp minimum around $(\bar{x}, \bar{y})$. Special conditions for lower level quadratic programs ensuring uniformly weak sharp minima are formulated in [15] (Prop. 5.2). Note, however, that these conditions are not satisfied in Example 3.4, because otherwise partial calmness could be guaranteed. In the following, we provide a strictly weaker constraint qualification than that of uniformly weak sharp minima, yet it is strong enough to still imply partial calmness for any locally Lipschitz objective F. Later we shall show that even this weaker constraint qualification is not satisfied (much less the existence of uniformly weak sharp minima) in typical lower level programs.

Definition 3.7. *With the notation of Definition 3.6, we say that the value function constraint qualification (VFCQ) is satisfied at* $(\bar{x}, \bar{y}) \in \text{gr } S$ *if the mapping*

$$K(u) := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m | y \in C, \ f(x, y) - \varphi(x) \leq u\}$$

is calm at $(0, \bar{x}, \bar{y})$ *in the sense of Definition 2.1.*

Proposition 3.8. *If the lower level problem (12) has a uniformly weak sharp minimum around one of its solutions* $\bar{y}$ *at* $x := \bar{x}$*, then (VFCQ) is satisfied at* $(\bar{x}, \bar{y})$*.*

Proof. Note first, that $(\bar{x}, \bar{y}) \in \text{gr } S$ by definition of S in (4) and that $K(0) = \text{gr } S$. Let $\varepsilon, \alpha > 0$ be the constants from Definition 3.6 and consider arbitrary $(u, x, y) \in \mathbb{B}((0, \bar{x}, \bar{y}), \varepsilon)$ such that $(x, y) \in K(u)$. Then,

$$d((x, y), K(0)) = d((x, y), \text{gr } S) \leq d(y, S(x)) \leq \alpha^{-1}(f(x, y) - \varphi(x)) \leq \alpha^{-1}u.$$

This, however, amounts to calmness of K at $(0, \bar{x}, \bar{y})$ in the sense of Definition 2.1. Consequently, (VFCQ) is satisfied. □

The following example shows that (VFCQ) is strictly weaker than the existence of uniformly weak sharp minima:

Example 3.9. *In (12) let* $C = [0, 2]$*,* $f(x, y) = xy$ *and* $(\bar{x}, \bar{y}) = (0, 1)$*. Then,*

$$S(x) = \begin{cases} \{2\} & \text{if } x < 0 \\ [0, 2] & \text{if } x = 0 \\ \{0\} & \text{if } x > 0 \end{cases}, \quad \varphi(x) = \begin{cases} 2x & \text{if } x < 0 \\ 0 & \text{if } x \geq 0 \end{cases}.$$

Now, choose any $(u, x, y) \in \mathbb{B}((0, \bar{x}, \bar{y}), 1/2)$ *such that* $(x, y) \in K(u)$*. A simple calculation shows that*

$$d((x, y), K(0)) = d((x, y), \text{gr } S) \leq 2u.$$

This implies that (VFCQ) is satisfied. On the other hand, the fact that, for instance, $d(\bar{y}, S(x)) = 1$ *holds true for all* $x < 0$ *entails the failure of the uniformly weak sharp minimum condition around* $(\bar{x}, \bar{y})$*.*

Though (VFCQ) is strictly weaker than the uniformly weak sharp minimum condition, it still guarantees the partial calmness condition:

Proposition 3.10. *Let* $(\bar{x}, \bar{y})$ *be a local solution of (4). If (VFCQ) is satisfied at* $(\bar{x}, \bar{y})$*, then (4) is partially calm at* $(\bar{x}, \bar{y})$*.*

Proof. According to Definition (3.7), let $\mathcal{U}$ be some neighbourhood of $(0, \bar{x}, \bar{y})$ and $L > 0$ be some constant such that

$$d((x, y), K(0)) \leq Lu \quad \forall (u, x, y) \in \mathcal{U} : (x, y) \in K(u).$$

Let $(x^*, y^*) \in K(0) = \text{gr } S$ be such that

$$d((x, y), K(0)) = \|(x, y) - (x^*, y^*)\|.$$

Since $(\bar{x}, \bar{y})$ is a local solution of (4), we may assume the neighbourhood $\mathcal{U}$ to be small enough to ensure that $F(\bar{x}, \bar{y}) \leq F(x^*, y^*)$. Finally, denote by $\varkappa > 0$ some Lipschitz constant of the upper level objective function F close to $(\bar{x}, \bar{y})$. Then,

$$\begin{aligned} F(\bar{x}, \bar{y}) - F(x, y) &\leq F(x^*, y^*) - F(x, y) \leq \varkappa \|(x, y) - (x^*, y^*)\| \\ &\leq \varkappa L u = \varkappa L |u| \qquad \forall (u, x, y) \in \mathcal{U} : (x, y) \in K(u). \end{aligned}$$

Here, the last equality relies on the fact that $u \geq 0$ whenever $(x, y) \in K(u)$ because $f(x, y) \geq \varphi(x)$ for all $y \in C$. By virtue of the definition of $K(u)$ this means that (4) is partially calm at $(\bar{x}, \bar{y})$. $\square$

Evidently, the last proposition applies to any locally Lipschitzian upper level objective F (despite our smoothness assumption) as this was the only property of F exploited in the proof. Propositions 3.8 and 3.10 show that the constraint qualification (VFCQ) is a stronger tool to guarantee the stationarity conditions of Theorem 3.1 than the condition of uniformly weak sharp minima. Nevertheless, we show in the following that (VFCQ) is much less efficient than the calmness condition on M in Theorem 3.3. First, recall that the mappings K and M from Definition 3.7 and Theorem 3.3 satisfy the relation $K(0) = M(0) = \mathrm{gr}\, S$.

Proposition 3.11. *If C is compact and (VFCQ) holds true at $(\bar{x}, \bar{y})$ then the perturbation mapping M from Theorem 3.3 is calm at $(0, \bar{x}, \bar{y})$. In particular, the same conclusion holds true if the lower level problem (12) has a uniformly weak sharp minimum around $(\bar{x}, \bar{y})$.*

Proof. According to Definition 3.7 and the mapping K defined there, there exist numbers $L, \varepsilon > 0$ such that

$$d((x, y), K(0)) \leq L|u| \quad \forall u \in [-\varepsilon, \varepsilon] \ \forall (x, y) \in K(u) \cap \mathbb{B}((\bar{x}, \bar{y}); \varepsilon). \tag{13}$$

Moreover, there is some $\Delta > 0$ such that $\|y^1 - y^2\| \leq \Delta$ for all $y^1, y^2 \in C$. Put $\tilde{\varepsilon} := \min\{\varepsilon, \varepsilon/\Delta\}$. Now, let $v \in \mathbb{B}(0; \tilde{\varepsilon})$ and $(x, y) \in M(v) \cap \mathbb{B}((\bar{x}, \bar{y}); \tilde{\varepsilon})$ be arbitrarily given. By definition of M, one has that

$$v \in \nabla_y f(x, y) + N_C(y),$$

which is equivalent to

$$(x, y) \in \arg\min_{y'} \{f(x, y') - \langle v, y' \rangle \, | y' \in C\}.$$

Therefore,

$$f(x, y) \leq f(x, y') - \langle v, y' - y \rangle \leq f(x, y') + \Delta \|v\| \quad \forall y' \in C.$$

Consequently, $f(x, y) \leq \varphi(x) + \Delta \|v\|$ and, thus, $(x, y) \in K(\Delta \|v\|)$. From $v \in \mathbb{B}(0; \tilde{\varepsilon})$, we derive that $\Delta \|v\| \leq \varepsilon$. Moreover, $(x, y) \in \mathbb{B}((\bar{x}, \bar{y}); \varepsilon)$. Summarizing, (13) may be invoked to yield

$$d((x, y), M(0)) = d((x, y), K(0)) \leq L\Delta \|v\|.$$

This implies that M is calm at $(0, \bar{x}, \bar{y})$. The second assertion follows now from Proposition 3.8. $\square$

The next theorem shows that in a typical setting for the lower level problem, the perturbation mapping M (8) is calm whereas (VFCQ) is violated. To this aim, we make the following assumptions on the lower level problem (6) at $(\bar{x},\bar{y}) \in \operatorname{gr} S$:

$$\text{LICQ and SSOSC hold at } \bar{y} \text{ (see Def. 2.8 and Def. 2.10).} \tag{14}$$

$$\#I(\bar{y}) < m \text{ (see(7)).} \tag{15}$$

$$\nabla_y f(\bar{x},\bar{y}) \neq 0. \tag{16}$$

These assumptions describe a somehow generic situation in nonlinear programming: (14) implies that the generalized equation describing the solution mapping S of (6) is *strongly regular* at $(\bar{x},\bar{y})$ in the sense of Robinson ([12], Th. 4.1). In particular, S is single-valued and Lipschitz continuous around $(\bar{x},\bar{y})$. (15) means that the solution $\bar{y}$ of (6) (at $x=\bar{x}$) is not attained in a corner of C but rather on a higher-dimensional face. Finally, (16) excludes a situation, where $\bar{y}$ is already an unconstrained solution of (6). In other words, the associated multiplier λ is nontrivial, such that the constraint $y \in C$ comes into play. Now, we have the following result:

Theorem 3.12. *For any $(\bar{x},\bar{y}) \in \operatorname{gr} S$ satisfying (14)-(16), M in (8) is calm (actually has the stronger Aubin property) at $(0,\bar{x},\bar{y})$, whereas (VFCQ) is violated at $(\bar{x},\bar{y})$. In particular, the lower level problem (12) has no uniformly weak sharp minimum around $(\bar{x},\bar{y})$.*

Proof. The statement about M follows from Theorem 3.5 (condition 2.), hence it remains to show that K fails to be calm at $(0,\bar{x},\bar{y})$ under the indicated conditions. This will entail that (VFCQ) is violated at $(\bar{x},\bar{y})$ and also via Proposition 3.8 that (12) has no uniformly weak sharp minimum around $(\bar{x},\bar{y})$. By strong regularity of S at $(\bar{x},\bar{y})$ (see discussion above), there exists some reals $\varkappa,\delta_1,\delta_2>0$ such that

$$\begin{aligned} \#\left\{S(x)\cap\mathbb{B}\left(\bar{y};\delta_2\right)\right\} &= 1 \quad \forall x\in\mathbb{B}\left(\bar{x};\delta_1\right) && (17)\\ d\left(S(x^1)\cap\mathbb{B}\left(\bar{y};\delta_2\right),S(x^2)\cap\mathbb{B}\left(\bar{y};\delta_2\right)\right) &\leq \varkappa\left\|x^1-x^2\right\| \quad \forall x^1,x^2\in\mathbb{B}\left(\bar{x};\delta_1\right) && (18)\end{aligned}$$

In particular, $S(\bar{x})\cap\mathbb{B}\left(\bar{y};\delta_2\right)=\{\bar{y}\}$. Assume for a moment that we have already shown that the restricted multifunction

$$\tilde{M}(u) := \{y\in\mathbb{R}^m | y\in C,\ f(\bar{x},y)-\varphi(\bar{x})\leq u\}$$

fails to be calm at $(0,\bar{y})$. Then, there are sequences $(u_n,y_n)\to(0,\bar{y})$ such that

$$y_n\in\tilde{M}(u_n) \text{ and } d(y_n,\tilde{M}(0))>n|u_n|.$$

Since $\tilde{M}(0)=S(\bar{x})$ and $y_n\to\bar{y}$ it follows that, for n large enough,

$$n|u_n| < d(y_n,\tilde{M}(0)) = d(y_n,S(\bar{x})) = d(y_n,S(\bar{x})\cap\mathbb{B}\left(\bar{y};\delta_2\right)) = \left\|y_n-\bar{y}\right\|. \tag{19}$$

If K were calm at $(0,\bar{x},\bar{y})$, then there would exist reals $L,\varepsilon>0$ such that (13) holds true. In particular, since $(\bar{x},y_n)\in K(u_n)$ (by $y_n\in\tilde{M}(u_n)$), it follows that

$$d\left(\left(\bar{x},y_n\right),K(0)\right)\leq L|u_n|.$$

Consequently, by $K(0)=\operatorname{gr} S$, there is a sequence $(x_n,v_n)\in\operatorname{gr} S$ such that

$$\|(\bar{x},y_n)-(x_n,v_n)\|\leq L|u_n|.$$

Considering without loss of generality the infinity-norm, we have that both

$$\|\bar{x} - x_n\|, \|y_n - v_n\| \leq L|u_n|.$$

In particular, $x_n \to \bar{x}$ and $v_n \to \bar{y}$. Then for large enough n, (17) entails that $S(x_n) \cap \mathbb{B}(\bar{y}; \delta_2) = \{v_n\}$ and (18) yields that

$$\|v_n - \bar{y}\| = d\left(S(x_n) \cap \mathbb{B}(\bar{y}; \delta_2), S(\bar{x}) \cap \mathbb{B}(\bar{y}; \delta_2)\right) \leq \varkappa \|x_n - \bar{x}\| \leq \varkappa L|u_n|.$$

This, finally leads to $\|y_n - \bar{y}\| \leq (\varkappa + 1) L|u_n|$, a contradiction with (19). Hence, K is not calm at $(0, \bar{x}, \bar{y})$ as was to be shown.

It remains to verify that $\tilde{M}$ fails to be calm at $(0, \bar{y})$. In order to show this, we want to invoke Theorem 2.2 by defining

$$\begin{aligned} T_1(v) \ &: = \{y \in \mathbb{R}^m | g_i(y) \leq v_i \quad (i = 1, \ldots, p)\} \\ T_2(u) \ &: = \{y \in \mathbb{R}^m | \, f(\bar{x}, y) - \varphi(\bar{x}) \leq u\}. \end{aligned}$$

Observe first that, due to Proposition 2.3, T_2 has the Aubin property at $(0, \bar{y})$ due to our assumption $\nabla_y f(\bar{x}, \bar{y}) \neq 0$. In particular, T_2 is calm at $(0, \bar{y})$. Similarly, our assumption, that LICQ holds at $\bar{y}$, implies again via Proposition 2.3 that T_1 has the Aubin property at $(0, \bar{y})$. So, T_1 is also calm at $(0, \bar{y})$. Finally, the inverse mapping

$$(T_2)^{-1}(y) = [f(\bar{x}, y) - \varphi(\bar{x}), \infty)$$

has trivially the Aubin property at $(\bar{y}, 0)$ by Lipschitz continuity of $f(\bar{x}, \cdot)$. Now, assume that $\tilde{M}$ was calm at $(0, \bar{y})$. Observing that $\tilde{M}(u) = T_1(0) \cap T_2(u)$, Theorem 2.2 would tell us that the mapping

$$M^*(u, v) := \{y \in \mathbb{R}^m | f(\bar{x}, y) - \varphi(\bar{x}) \leq u, \, g_i(y) \leq v_i \quad (i = 1, \ldots, p)\}$$

is calm at $(0, 0, \bar{y})$. Then, by Proposition 2.4, the contingent and linearized cones of $M^*(0, 0) = S(\bar{x})$ at $\bar{y}$ must coincide. Given that $S(\bar{x}) \cap \mathbb{B}(\bar{y}; \delta_2) = \{\bar{y}\}$, it follows for the contingent cone that $T_{M^*(0,0)}(\bar{y}) = \{0\}$. Let us calculate the linearized cone (recalling the definition of the index set $I(\bar{y})$ introduced above):

$$L_{M^*(0,0)}(\bar{y}) = \{h \in \mathbb{R}^m | \, \langle \nabla_y f(\bar{x}, \bar{y}), h \rangle \leq 0, \, \langle \nabla g_i(\bar{y}), h \rangle \leq 0 \quad (i \in I(\bar{y}))\}.$$

We claim that $L_{M^*(0,0)}(\bar{y})$ contains at least a 1-dimensional subspace. Indeed, by definition

$$L_{M^*(0,0)}(\bar{y}) \supseteq \{h \in \mathbb{R}^m | \, \langle \nabla_y f(\bar{x}, \bar{y}), h \rangle = 0, \, \langle \nabla g_i(\bar{y}), h \rangle = 0 \quad (i \in I(\bar{y}))\}.$$

Moreover, $(\bar{x}, \bar{y}) \in \operatorname{gr} S$ implies that $\bar{y}$ is a solution of the lower level problem at $x = \bar{x}$. Accordingly, the set of vectors

$$\{\nabla_y f(\bar{x}, \bar{y})\} \cup \{\nabla g_i(\bar{y})\}_{i \in I(\bar{y})}$$

is linearly dependent. Now, (15) entails that the rank of this set of vectors is at most $m - 1$. It follows that $L_{M^*(0,0)}(\bar{y})$ contains a linear subspace of at least dimension 1, which implies $L_{M^*(0,0)}(\bar{y}) \neq T_{M^*(0,0)}(\bar{y})$, showing that $\tilde{M}$ cannot be calm at $(0, \bar{y})$. This completes the proof. $\square$

4 A discrete obstacle problem

For the purpose of comparing the different calmness concepts discussed above as well as the resulting stationarity conditions according to Theorems 3.1 and 3.3, we consider the following instance of a bilevel problem (4) resulting from the control of a discrete obstacle problem [1]:

$$\min\left\{ \frac{1}{2}\|y-z\|^2 + \frac{c}{2}\|x\|^2 \,\middle|\, y \in S(x) \right\} \tag{20}$$

where

$$S(x) := \arg\min_y \left\{ \langle y, Ay\rangle - \langle Bx, y\rangle \,|\, y \in \mathbb{R}^m_+ \right\}.$$

Here, B is any matrix of appropriate size and A is a positive definite matrix, for instance the three-point Laplacian

$$A = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}. \tag{21}$$

In this problem, y refers to a state variable (e.g. shape of a 1-dimensional membrane), x is a control variable (force), z represents a given desired state and the zero level is an obstacle for the state variable (thus giving rise to the constraint $y \in \mathbb{R}^m_+$). The upper level objective

$$F(x,y) = \frac{1}{2}\|y-z\|^2 + \frac{c}{2}\|x\|^2$$

reflects a compromise between approaching the desired state and minimizing the cost of the control. The balance between both goals is determined by the coefficient $c > 0$. The minimization of the lower level objective

$$f(x,y) := \langle y, Ay\rangle - \langle Bx, y\rangle$$

determines the state as a function of the control (thereby respecting the lower level constraint $y \in C := \mathbb{R}^m_+$). The matrix B could be used to model a situation in which the control is defined only on a subinterval of the domain of the state. For instance, it might have the form

$$B = \begin{pmatrix} I \\ 0 \end{pmatrix}.$$

We next observe that all technical assumptions (apart from calmness) of Theorems 3.1 and 3.3 are satisfied. Indeed the lower constraint set C satisfies trivially (LICQ) at any of its feasible points. Much more it satisfies the weaker constraint qualification (CRCQ) and has a Slater point (e.g., $y^* = (1, \ldots, 1)$). Moreover, (SSOSC) is satisfied at any solution to the lower level problem given any value of the upper level variable x due to positive definiteness of A. Along with (LICQ) this implies (again via strong regularity as in the paragraph on top of Theorem 3.12) that the solution set mapping S of the lower level problem is locally unique and Lipschitz continuous around any point of its graph. In particular, S is inner semicontinuous at any point of its graph. Now we are in a position to state the following result on the failure of partial calmness in the obstacle problem:

Proposition 4.1. *If the desired state z contains at least one positive component, then partial calmness is violated at any local solution $(\bar{x}, \bar{y})$ of (20).*

Proof. Let $(\bar{x}, \bar{y})$ be a local solution of (20). Assume that partial calmness holds true at $(\bar{x}, \bar{y})$. Then, all assumptions of Theorem 3.1 are satisfied and we derive from the respective first stationarity condition that

$$0 = \nabla_x F(\bar{x}, \bar{y}) = c\bar{x},$$

which leads to $\bar{x} = 0$ due to $c > 0$. Now the associated lower level problem becomes

$$\min \{\langle y, Ay\rangle \,| y \geq 0\}$$

with its unique solution evidently given by $\bar{y} = 0$. Since also the resulting optimal value equals zero, we get that $\varphi(\bar{x}) = 0$ with φ from (5). Define z^+ by

$$z_i^+ := \max\{z_i, 0\} \quad (i = 1, \ldots, m).$$

By assumption, $z^+ \neq 0$. Introducing sequences

$$\begin{aligned} y_n &: = \frac{1}{\sqrt{n}\,\|z^+\|} z^+ \to 0 = \bar{y} \\ u_n &: = -\langle y_n, Ay_n\rangle \to 0, \end{aligned}$$

we observe that $y_n \in C = \mathbb{R}^m_+$ and $u_n = \varphi(\bar{x}) - f(\bar{x}, y_n)$. Consequently, we may apply the partial calmness definition (2.5) to the sequence $(u_n, \bar{x}, y_n)$ to derive the relation

$$\begin{aligned} 0 &\leq F(\bar{x}, y_n) - F(\bar{x}, \bar{y}) + L|u_n| = \frac{1}{2}\|y_n - z\|^2 - \frac{1}{2}\|z\|^2 + L\langle y_n, Ay_n\rangle \\ &= \frac{1}{2}\|y_n\|^2 - \langle y_n, z\rangle + L\langle y_n, Ay_n\rangle \\ &= \frac{1}{2n} - \frac{1}{\sqrt{n}\,\|z^+\|}\left\langle z^+, z\right\rangle + L\frac{1}{n\,\|z^+\|^2}\left\langle z^+, Az^+\right\rangle. \end{aligned}$$

Denoting the (positive) maximum eigenvalue of A by $\lambda^{\max}$ and observing that $\langle z^+, z\rangle = \|z^+\|^2$, one arrives at the contradiction

$$\sqrt{n}\,\|z^+\| \leq \frac{1}{2} + L\lambda^{\max}$$

for $n \to \infty$ because $\|z^+\| > 0$. Hence, our assumption that partial calmness holds true at $(\bar{x}, \bar{y})$ is false. □

The Proposition shows, that in the obstacle problem, by failure of partial calmness, the application of Theorem 3.1 is limited to settings, where $z \leq 0$. In this latter case, however, it is clear that the optimal solution of the problem is the trivial one: $\bar{x} = \bar{y} = 0$.

In contrast, calmness of the perturbed generalized equation (8) holds true in any case. Indeed, conditions 1. and 2. of Theorem 3.5 are satisfied simultaneously, where already one of them would have sufficed. Consequently, all assumptions of Theorem 3.3 are fulfilled. We recall, that the functions describing the set C in (4) are given by $g_i(y) := -y_i$ in the concrete setting of the obstacle problem, where $C = \mathbb{R}^m_+$. Consequently,

$$\nabla g_i(y) = -e_i, \quad \nabla^2 g_i(y) = 0,$$

with e_i denoting the standard unit vectors of $\mathbb{R}^n$. Writing down then the stationarity conditions of Theorem 3.3 in terms of the data of the obstacle problem, one arrives at

$$\begin{aligned}
0 &= c\bar{x} - B^T v^* && (22)\\
0 &= \bar{y} - z + 2Av^* - \sum\nolimits_{i\in I(\bar{y})} w_i^* e_i && (23)\\
0 &= v_i^* \quad \forall i \in I(\bar{y}) : \lambda_i > 0 && (24)\\
0 &= w_i^* \quad \forall i \in I(\bar{y}) : \lambda_i = 0,\ v_i^* > 0 && (25)\\
0 &\le w_i^* \quad \forall i \in I(\bar{y}) : \lambda_i = 0,\ v_i^* < 0 && (26)\\
0 &= 2A\bar{y} - B\bar{x} - \sum\nolimits_{i\in I(\bar{y})} \lambda_i e_i, && (27)
\end{aligned}$$

where $\lambda \geq 0$, v^* and w^* are certain multipliers whose existence is guaranteed. We may use these conditions to derive the following equivalent reformulation of problem (20) as a simple one-level optimization problem. To this aim, we make a specific assumption on the form of A which is fulfilled, for instance, for the three-point Laplacian (21) and we also suppose a certain structure of the matrix B, namely that

$$B = \begin{pmatrix} I_n \\ 0 \end{pmatrix} \tag{28}$$

for some $n \leq m$. This choice reflects a situation where the control x is supported by a (left) subinterval of the state's domain only (first m out of n discretization points). In accordance with (28), we partition the matrix A as

$$A = \begin{pmatrix} A^{(1)} \\ A^{(2)} \end{pmatrix}. \tag{29}$$

Theorem 4.2. *Assume that the positive definite matrix has the property*

$$A_{ij} \leq 0 \quad \forall i, j = 1, \ldots, m : i \neq j \tag{30}$$

(as it is the case for A in (21)). Assume further, that B is given as in (28). Then, the following holds true:

$(\bar{x}, \bar{y})$ is a solution to (20) if and only if $\bar{y}$ is a solution to

$$\min \left\{ \frac{1}{2}\|y - z\|^2 + 2c\left\|A^{(1)}y\right\|^2 \,\middle|\, y \in \mathbb{R}^m_+,\ A^{(2)}y = 0 \right\} \tag{31}$$

and

$$\bar{x} = 2A^{(1)}\bar{y}. \tag{32}$$

Proof. Let $(\bar{x}, \bar{y})$ be a solution to (20). We first show that $\bar{y}$ is a solution to (31) and that $\bar{x}$ satisfies (32). By Theorem 3.3 there exist multipliers $\lambda \geq 0$, v^* and w^* such that (22) - (27) are fulfilled. Assume that there is some $i \in I(\bar{y}) \cap \{1, \ldots, n\}$ such that $\lambda_i > 0$. We denote by A_i and B_i the ith rows of the matrices A and B, respectively. Since $B_j = e_j$ for $j \in \{1, \ldots, n\}$ (see (28)), we may derive from (27) that

$$0 = 2\langle A_i, \bar{y}\rangle - \langle B_i, \bar{x}\rangle - \lambda_i < 2\langle A_i, \bar{y}\rangle - \bar{x}_i. \tag{33}$$

Now, with $\bar{y} \geq 0$ and, given that $\bar{y}_i = 0$ due to $i \in I(\bar{y})$, it follows from (30) that

$$2\langle A_i, \bar{y}\rangle = 2\sum_{j \neq i} A_{ij}\bar{y}_j \leq 0. \tag{34}$$

Along with (33) this yields $\bar{x}_i < 0$. Now, with $B^T = (I_n|0)$, and recalling that $i \in \{1, \ldots, n\}$, we derive from (22) that

$$v_i^* = \left(B^T v^*\right)_i = c\bar{x}_i < 0.$$

This, however, contradicts (24). Thus, we have shown that $\lambda_i = 0$ for all $i \in I(\bar{y}) \cap \{1, \ldots, n\}$. We may now read the components of (27) as follows:

$$2\langle A_i, \bar{y}\rangle - \langle B_i, \bar{x}\rangle \begin{cases} = 0 & i \in I(\bar{y}) \cap \{1, \ldots, n\} \\ = 0 & i \notin I(\bar{y}) \\ \geq 0 & i \in \{n+1, \ldots, m\}. \end{cases} \tag{35}$$

In particular,

$$2\langle A_i, \bar{y}\rangle = \langle B_i, \bar{x}\rangle \quad \forall i \in \{1, \ldots, n\}.$$

Using the partition (29), we may compress these relations as $\bar{x} = 2A^{(1)}\bar{y}$ thus proving (32). Next, let $i \in I(\bar{y}) \cap \{n+1, \ldots, m\}$. Then, $\langle B_i, \bar{x}\rangle = 0$ and using again the argument leading to (34), we infer that $2\langle A_i, \bar{y}\rangle - \langle B_i, \bar{x}\rangle \leq 0$. Now, the second and third relations in (35) provide equality here: $2\langle A_i, \bar{y}\rangle - \langle B_i, \bar{x}\rangle = 0$ for all $i \in \{n+1, \ldots, m\}$. Referring again to the partition (29) and recalling that $\langle B_i, \bar{x}\rangle = 0$ for $i \in \{n+1, \ldots, m\}$, we arrive at $A^{(2)}\bar{y} = 0$. We now show that $\bar{y}$ is a solution to (31). Since $\bar{y} \in \mathbb{R}_+^m$, $\bar{y}$ is feasible in (31). Assume that $\bar{y}$ is not a solution to (31). Then there exists some $\hat{y} \in \mathbb{R}_+^m$ and satisfying $A^{(2)}\hat{y} = 0$ such that

$$\frac{1}{2}\|\hat{y} - z\|^2 + 2c\|A\hat{y}\|^2 < \frac{1}{2}\|\bar{y} - z\|^2 + 2c\|A\bar{y}\|^2. \tag{36}$$

Defining

$$\hat{x} := \begin{pmatrix} 2A^{(1)}\hat{y} \\ 0 \end{pmatrix},$$

we derive from $A^{(2)}\hat{y} = 0$ that

$$\|\hat{x}\|^2 = 4\left\|A^{(1)}\hat{y}\right\|^2 = 4\|A\hat{y}\|^2.$$

Similarly, the already proved relation (32) yields along with $A^{(2)}\bar{y} = 0$ that $\|\bar{x}\|^2 = 4\|A\bar{y}\|^2$. Consequently, (36) may be written as

$$\frac{1}{2}\|\hat{y} - z\|^2 + \frac{1}{2}c\|\hat{x}\|^2 < \frac{1}{2}\|\bar{y} - z\|^2 + \frac{1}{2}c\|\bar{x}\|^2. \tag{37}$$

On the other hand,

$$2A\hat{y} - B\hat{x} = \begin{pmatrix} 2A^{(1)}\hat{y} - \hat{x} \\ 2A^{(2)}\hat{y} - 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

It follows, that $\hat{y}$ satisfies the Kuhn-Tucker conditions for the lower level problem

$$\arg\min_y \left\{ \langle y, Ay \rangle - \langle B\hat{x}, y \rangle \, | y \in \mathbb{R}^m_+ \right\}.$$

As this is a convex optimisation problem (by positive definiteness of A), $\hat{y}$ must actually be a solution of this problem. In terms of the lower level solution mapping S, this means that $\hat{y} \in S(\hat{x})$. Now, (37) entails that $(\bar{x}, \bar{y})$ cannot be a solution to (20). This contradiction implies that $\bar{y}$ is a solution to (31). We have therefore proved one implication of the theorem. For the reverse direction, let now $(\bar{x}, \bar{y})$ be such that $\bar{y}$ solves (31) and $\bar{x}$ satisfies (32). Observe that the objective in (31) is strictly convex. Consequently, (31) $\bar{y}$ is the only solution to (31). Therefore, if (20) possesses a solution at all, this must coincide by the already proved implication with $(\bar{x}, \bar{y})$. Hence, we are done once we can show that (20) possesses a solution.

Though it can be verified that (20) does indeed have a solution via classical methods from the calculus of variations, see e.g., ([7]), we provide an elementary proof for the sake of completeness. To see this, we verify first that $S(x) \neq \emptyset$ for all $x \in \mathbb{R}^n$. Indeed, let $x \in \mathbb{R}^n$ be arbitrarily given. Then, $S(x)$ is the set of solutions to

$$\min_y \left\{ \langle y, Ay \rangle - \langle Bx, y \rangle \, | y \in \mathbb{R}^m_+ \right\}.$$

We put $H := \{ y \in \mathbb{R}^m | \, \langle y, Ay \rangle - \langle Bx, y \rangle \leq 1 \}$. Since $\emptyset \neq H \cap \mathbb{R}^m_+ \ni 0$, $S(x)$ is also the set of solutions to

$$\min_y \left\{ \langle y, Ay \rangle - \langle Bx, y \rangle \, | y \in H \cap \mathbb{R}^m_+ \right\}.$$

Denoting by $\lambda^{\min} > 0$ the smallest eigenvalue of A, we have that

$$\lambda^{\min} \|y\|^2 \leq \langle y, Ay \rangle \leq \langle Bx, y \rangle + 1 \leq \|y\| \left(\|Bx\| + 1 \right) \quad \forall y \in H : \|y\| \geq 1$$

This leads to

$$\|y\| \leq \left(\lambda^{\min} \right)^{-1} \|Bx\| + 1 \quad \forall y \in H : \|y\| \geq 1.$$

As a consequence, H is compact and the solution set $S(x)$ is nonempty. Hence, by the already employed strict convexity argument for the objective function of the lower level problem, $S(x)$ is a singleton for all $x \in \mathbb{R}^n$. On the other hand, referring back to the discussion in front of Proposition 4.1, S is a locally Lipschitz continuous mapping around all points of its graph. Summarizing, S is a continuos function (!) defined on all of $\mathbb{R}^n$. This allows us to equivalently rewrite (20) as the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|S(x) - z\|^2 + \frac{c}{2} \|x\|^2 \right\}. \tag{38}$$

Now, we use a similar argument as above, but this time for the upper level problem. Let

$$\bar{H} := \left\{ x \in \mathbb{R}^n \, \middle| \, \frac{1}{2} \|S(x) - z\|^2 + \frac{c}{2} \|x\|^2 \leq \|z\|^2 \right\}.$$

Since $S(0) = \{0\}$, one has that $\emptyset \neq \bar{H} \ni 0$, and so problem (38) is equivalent with

$$\min \left\{ \frac{1}{2} \|S(x) - z\|^2 + \frac{c}{2} \|x\|^2 \,\middle|\, x \in \bar{H} \right\}. \tag{39}$$

Since $\bar{H}$ is contained in a closed ball with radius $\|z\|\sqrt{2/c}$ it follows that $\bar{H}$ is compact. As the objective of (39) is continuous by continuity of S, problem (39) and, hence, the original problem (20) possesses a solution. This finishes the proof of the theorem. $\square$

Theorem 4.2 allows us to reduce the bilevel obstacle problem to a simple quadratic optimization problem the numerical solution of which is easily carried out by standard tools. Figure 1 provides an illustration for the solution of the obstacle problem under two different choices for the weighting coefficient c based on the solution of (31) along with the relation (32). The given interval is discretized into $m = 100$

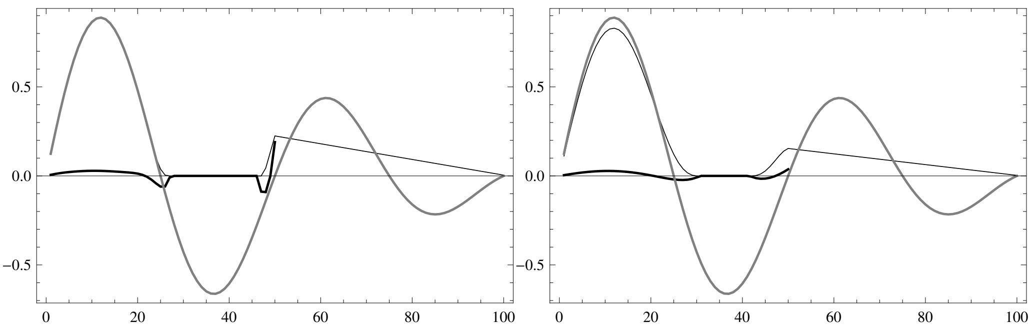

Figure 1: Illustration of solutions to the obstacle problem for coefficients $c = 1$ (left) and $c = 50$ (right). The plot shows the desired state z (thick, gray), the optimal state y (thin, black) and the optimal control x (thick, black).

points. Here, the control is defined on the first half of theinterval only (i.e., $n = 50$). For a small coefficient $c = 1$, the optimal state y, which is required to stay nonnegative, fits the desired state perfectly, whenever it too is positive in the first part of the interval. A substantially larger coefficient $c = 50$ forces the control to be smaller than in the first case. As a consequence, the discrepancy between optimal and desired state increases. Observe that the right-most value of the control determines the rate of linear decay of the optimal state on the second half of the entire interval.

5 Conclusions

The results of this paper (in particular, Prop. 3.11, Th. 3.12 and Prop. 4.1) seem to suggest that the concept of partial calmness via the value function approach is less efficient than calmness of the perturbation mapping (8) when used as a constraint qualification for M-stationarity in bilevel optimization problems. On the other hand, there are very recent developments where both approaches are combined (see [19]).

References

[1] M. Bergounioux and F. Mignot, *Optimal control of obstacle problems: Existence of Lagrange multipliers*, ESAIM Control Optim. Calc. Var. 5 (2000), pp. 45-70.

[2] S. Dempe, J. Dutta and B. Mordukhovich, *New necessary optimality conditions in optimistic bilevel programming*, Optimization 56 (2007), pp. 577-604.

[3] R. Henrion and J. Outrata, *Calmness of Constraint Systems with Applications*, Math.Program. 104 (2005), pp. 437-464.

[4] R. Henrion, J. Outrata and T. Surowiec, *Analysis of M-stationary points to an EPEC modeling oligopolistic competition in an electricity spot market*, Preprint No. 1433, Weierstrass Institute Berlin, Berlin, Germany, 2009 and submitted.

[5] R. Henrion, J. Outrata and T. Surowiec, *A note on the relation between strong and M-stationarity for a class of mathematical programs with equilibrium constraints*, to appear in Kybernetika.

[6] D. Klatte and B. Kummer, *Constrained Minima and Lipschitzian Penalties in Metric Spaces*, SIAM J. Optim. 13 (2002), pp. 619-633 .

[7] F. Mignot and J.-P. Puel, *Optimal Control in Some Variational Inequalities*, SIAM J. Con. Opt., 22 (1984), pp. 466-476.

[8] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation.*, Springer, Berlin, 2006.

[9] B.S. Mordukhovich and J.V. Outrata, *Coderivative analysis of quasi-variational inequalities with applications to stability and optimization*, SIAM J. Optim. 18 (2007), pp. 389-412.

[10] J.V. Outrata, *On the numerical solution of a class of Stackelberg problems*, Math. Methods Oper. Res. 34 (1990), pp. 255-277.

[11] J.V. Outrata, *A generalized mathematical program with equilibrium constraints*, SIAM J. Control Optim. 38 (2000), pp. 1623-1638.

[12] S. Robinson, *Strongly regular generalized equations*, Math. Oper. Res. 5 (1980), pp. 43-62.

[13] S. Robinson, *Some continuity properties of polyhedral multifunctions.* Math.Program. Study 14 (1981), pp. 206-214.

[14] T. Surowiec, *Explicit stationarity conditions and solution characterization for equilibrium problems with equilibrium constraints*, PhD thesis, Humboldt-University Berlin, Germany, 2010.

[15] J.J. Ye and D.L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization 33 (1995), pp. 9-27.

[16] J.J. Ye and X.Y. Ye, *Necessary optimality conditions for optimization problems with variational inequality constraints*, Math. Oper. Res. 22 (1997), pp. 977-997.

[17] J.J. Ye, D.L. Zhu and Q.J. Zhu, *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, SIAM J. Optim. 7 (1997), pp. 481-507.

[18] J.J. Ye, *Constraint Qualifications and Necessary Optimality Conditions for Optimization Problems with Variational Inequality Constraints*, SIAM J. Optim. 10 (2000), pp. 943-962.

[19] J.J. Ye, D. Zhu, *New Necessary Optimality Conditions for Bilevel Programs by Combining the MPEC and Value Function Approaches*, SIAM J. Optim. 20 (2010), pp. 1885-1905.