NONCONVEX QUASI-VARIATIONAL INEQUALITIES: STABILITY ANALYSIS AND APPLICATION TO NUMERICAL OPTIMIZATION*

JOYDEEP DUTTA†, LAHOUSSINE LAFHIM‡, ALAIN ZEMKOHO§, AND SHENGLONG ZHOU¶

Abstract. We consider a parametric quasi-variational inequality (QVI) without any convexity assumption. Using the concept of *optimal value function*, we transform the problem into that of solving a nonsmooth system of inequalities. Based on this reformulation, new coderivative estimates as well as robust stability conditions for the optimal solution map of this QVI are developed. Also, for an optimization problem with QVI constraint, necessary optimality conditions are constructed and subsequently a tailored semismooth Newton-type method is designed, implemented, and tested on a wide range of optimization examples from the literature. In addition to the fact that our approach does not require convexity, its coderivative and stability analysis does not involve second order derivatives and subsequently, the proposed Newton scheme does not need third order derivatives.

Key words. Quasi-variational inequalities, stability analysis, optimal value function, optimization problems with quasi-variational inequality constraints, semismooth Newton method

AMS subject classifications. 90C26, 90C31, 90C33, 90C46, 90C55

1. Introduction.

In this paper, we consider the quasi-variational inequality (QVI) problem to find $y \in K(x,y)$, for a given parameter $x \in \mathbb{R}^n$, such that

$$\langle f_0(x,y),\ \varsigma - y\rangle \geq 0 \quad \forall \varsigma \in K(x,y) := \{\varsigma \in \mathbb{R}^m \mid g_0(x,y,\varsigma) \leq 0\}, \tag{1.1}$$

where $f_0 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is a continuously differentiable function and $K : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ denotes the feasibility set-valued map defined by a function $g_0 : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^q$, which is also assumed to be continuously differentiable. Quasi-variational inequalities have been widely studied in the literature, considering the large number of applications and the mathematical challenges involved in the process of solving them; see, e.g., [6, 36, 35] and references therein, for some numerical methods to solve different classes of the problem, as well a number of applications of QVIs. This paper has two main goals, with the first one being to study the stability of the solution maping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$

$$S(x) := \{y \in K(x,y) \mid \langle f_0(x,y),\ \varsigma - y\rangle \geq 0 \quad \forall \varsigma \in K(x,y)\} \tag{1.2}$$

associated to the QVI (1.1). More precisely, we aim to construct estimates of the coderivative of S (in the sense of Mordukhovich [31]) and sufficient conditions ensuring that this set-valued mapping is Lipschitz-like (in the sense of Aubin [3]). The second objective of the paper is to develop a numerical method to solve an optimization problem partly constrained by the QVI (1.1).

Some of these questions have been addressed in the literature (see, e.g., [33, 20, 19]), and as it is common in the broad literature on variational inequalities, the set $K(x,y)$ is assumed to be convex, for all (x,y). Hence, to study the stability/sensitivity of the parametric QVI (1.1), it is common to use the following generalized equation (GE) reformulation of S (1.2):

$$S(x) = \left\{y \in \mathbb{R}^m \,\middle|\, 0 \in f_0(x,y) + N_{K(x,y)}(y)\right\}. \tag{1.3}$$

Here, $N_{K(x,y)}(y)$ represents the normal cone, in the sense of convex analysis, to the set $K(x,y)$ at the point y; cf. next section for the definition of this concept. Obviously, the GE reformulation (1.3) of S (1.2) is only possible if K (1.1) is convex-valued. In the absence of this convexity assumption, this approach is not applicable. Secondly, the stability analysis of S (computation of the coderivative and construction of conditions ensuring that it is Lipschitz-like) conducted through (1.3) would require second order information, which is not available for various applications (see, e.g., [26]).

*Submitted to the editors on October 4, 2022

Funding: The work of AZ was supported by the EPSRC grant EP/V049038/1 and the Alan Turing Institute under the EPSRC grant EP/N510129/1.

†Department of Economic Sciences, Indian Institute of Technology Kanpur, India (jdutta@iitk.ac.in).

‡Department of Mathematics, Sidi Mohammed Ben Abdellah University, Morocco (lahoussine.lafhim@usmba.ac.ma).

§School of Mathematical Sciences, University of Southampton, UK (a.b.zemkoho@soton.ac.uk).

¶Department of EEE, Imperial College London, London, UK (shenglong.zhou@imperial.ac.uk).

Our analysis in this paper is based on the fact that the set-valued mapping S (1.2) can easily be rewritten, without any assumption, as

$$S(x) := \left\{y \in K(x,y) \,\middle|\, y^\top f_0(x,y) - \varphi(x,y) \leq 0\right\} \tag{1.4}$$

for all $x \in \mathbb{R}^n$. Here, φ denotes the optimal value function

$$\varphi(x,y) := \min_{\varsigma}\{\varsigma^\top f_0(x,y)| \ \varsigma \in K(x,y)\}. \tag{1.5}$$

The core of the analysis in this paper is based on this *value function reformulation* of the QVI (1.1). Clearly, unlike (1.3), transformation (1.4) does not require any convexity assumption. Hence, based on the value function reformulation, we develop completely new results for S (1.2) without any convexity assumption. Additionally, no second order derivatives is involved in our analysis, as it would be the case when using the GE reformulation in (1.3).

As for the second objective of this paper, we consider the following optimization problem with a quasi-variational inequality constraint (OPQVI):

$$\min_{x,y} F(x,y) \quad \text{s.t.} \quad G(x,y) \leq 0, \ \ y \in S(x). \tag{1.6}$$

Here, the functions $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$, and the set-valued mapping S is defined as in (1.2). Problem (1.6) has a two-level optimization structure, with F and G representing the upper-level objective and constraint functions, respectively. In the same vein, S corresponds to the optimal solution map of the lower-level problem, which, unlike in standard two-level/bilevel optimization (see, e.g., [9, 11]), is a QVI defined as in (1.1).

For problem (1.6), we consider the value function reformulation based on (1.4) and develop necessary optimality conditions for the problem under suitable *calmness* conditions. Subsequently, we develop a version of the semismooth Newton method tailored to these necessary optimality conditions and prove its convergence under a suitable framework. Extensive experiments on over 124 examples from the literature [33, 47] are then conducted to demonstrate the efficiency potential of our method.

Special classes of problem (1.6) have been studied in many papers in the literature; see, e.g., [1, 33, 25, 24, 20, 19]. However, the main focus has usually been on the derivation of necessary optimality conditions based on the GE reformulation (1.3), with a special attention on the construction of suitable qualification conditions. To the best of our knowledge, not much is available in the literature in terms of solution algorithms for problems of the form (1.6). Moreover, we are not aware of any work where the problem has been studied in the absence of convexity. Also, an attempt to develop a Newton-type method for the problem from the perspective of the GE reformulation (1.3) would require third order derivatives for the function g_0 describing K (1.1). This is another reason why it is attractive to develop such methods based on the value function reformulation (1.4), as we do in this paper. Additionally, the extend of the numerical experiments conducted in this paper could serve as based to accelerate work on the development of numerical methods for the problem.

For the reminder of the paper, in the next section, we first provide some preliminary tools from variational analysis that will be needed in the subsequent sections. In particular, focus in Section 2 is on generalized differentiation tools for nonsmooth functions and set-valued mappings. In Section 3, we develop stability results for the set-valued mapping (1.2); as our analysis is based on (1.4), the presence of the value function constraint motivates the introduction and study of a version of the *uniform weak sharp minimum* concept tailored to the QVI (1.1) and provide sufficient conditions ensuring that it holds. Finally, in Section 4, a *partial calmness* concept tailored to the value function reformulation of problem (1.6) is introduced to build necessary optimality conditions. Subsequently, suitable second order conditions are then introduced to establish the convergence of a semismooth Newton scheme introduced to solve the aforementioned necessary conditions.

2. Basic concepts and background material. We start this section with some basic notation to be mostly used in the following main sections of the paper. Considering a vector $x \in \mathbb{R}^n$ and a scalar $\epsilon > 0$, we denote by $\mathbb{U}_\epsilon(x) := \{y \in \mathbb{R}^n | \parallel y - x \parallel_\infty < \epsilon\}$ and $\mathbb{B}_\epsilon(x) := \{y \in \mathbb{R}^n | \parallel y - x \parallel_\infty \leq \epsilon\}$ the open and closed ϵ-balls around x, respectively. Whenever there is no confusion or for convenience at some places, we will use the notation $z := (x, y)$, $g(x, y) := g_0(x, y, y)$, and $f(x, y) := y^\top f_0(x, y)$. Subsequently, for the points $\bar{z} := (\bar{x}, \bar{y})$ and $(\bar{z}, \bar{\varsigma})$, the following notation will be used to collect index sets of active points for the constraints defined by the functions G, g, and g_0, respectively:

$$\begin{array}{lllll} I^1 & := & I^G(\bar{z}) & := & \{i \mid G_i(\bar{z}) = 0\}\,, \\ I^2 & := & I^g(\bar{z}) & := & \{j \mid g_j(\bar{z}) = 0\}\,, \\ I^3 & := & I^{g_0}(\bar{z}, \bar{\varsigma}) & := & \left\{j \mid g_{0_j}(\bar{z}, \bar{\varsigma}) = 0\right\}. \end{array} \tag{2.1}$$

If we associate to a point $\bar{z}$, feasible for the constraint defined by G (1.6), a Lagrange multiplier $\bar{u}$, we can proceed with the following standard partition of the indices:

$$\begin{array}{lllll} \eta^1 & := & \eta^G(\bar{z}, \bar{u}) & := & \{j|\ \bar{u}_j = 0,\ G_j(\bar{z}) < 0\}, \\ \theta^1 & := & \theta^G(\bar{z}, \bar{u}) & := & \{j|\ \bar{u}_j = 0,\ G_j(\bar{z}) = 0\}, \\ \nu^1 & := & \nu^G(\bar{z}, \bar{u}) & := & \{j|\ \bar{u}_j > 0,\ G_j(\bar{z}) = 0\}. \end{array} \tag{2.2}$$

Similarly, for a feasible point $\bar{z}$ (resp. $(\bar{z}, \bar{\varsigma})$) to g (resp. g_0) and a corresponding Lagrange multipliers $\bar{v}$ (resp. $\bar{w}$), we can analogously define the following partitions:

$$\begin{array}{lllllllll} \eta^2 & := & \eta^g(\bar{z}, \bar{v}), & \theta^2 & := & \theta^g(\bar{z}, \bar{v}), & \nu^2 & := & \nu^g(\bar{z}, \bar{v}), \\ \eta^3 & := & \eta^{g_0}(\bar{z}, \bar{\varsigma}, \bar{w}), & \theta^3 & := & \theta^{g_0}(\bar{z}, \bar{\varsigma}, \bar{w}), & \nu^3 & := & \nu^{g_0}(\bar{z}, \bar{\varsigma}, \bar{w}). \end{array} \tag{2.3}$$

2.1. Generalized derivatives for nonsmooth functions. Given that the optimal value function φ (1.5) is nonsmooth in general, we need generalized concepts of differentiability to deal with it. We start the discussion on this by recalling the *generalized directional derivative* in the sense of Clarke [7], which can be defined for a function $\psi : \mathbb{R}^n \to \mathbb{R}$ by

$$\psi^0(\bar{x}; d) := \limsup_{\substack{x \to \bar{x} \\ t \downarrow 0}} \frac{1}{t} \left[\psi(x + td) - \psi(x)\right].$$

This quantity exists if ψ is any function Lipschitz continuous around $\bar{x}$ [7, Proposition 2.1.1]. Utilizing this notion, the Clarke subdifferential can also be introduced:

$$\bar{\partial}\psi(\bar{x}) := \{\zeta \in \mathbb{R}^n|\ \psi^o(\bar{x}; d) \geq \langle \zeta, d\rangle,\ \forall d \in \mathbb{R}^n\}\,. \tag{2.4}$$

Note that $\partial\psi(\bar{x}) = \{\nabla\psi(\bar{x})\}$ if ψ if ψ is differentiable at $\bar{x}$, and for a convex function, this concept coincides with the subdifferential in sense of convex analysis. Furthermore, this concept can be extended to a vector-valued function $\psi : \mathbb{R}^n \to \mathbb{R}^m$ that is Lipschitz continuous around a point $\bar{x}$, as such a function is differentiable almost everywhere around this point. Hence, the *Clarke generalized Jacobian* of this function can be written as

$$\bar{\partial}\psi(\bar{x}) := \text{co}\ \{\lim\ \nabla\psi(x^n) :\ \ x^n \to \bar{x},\ \ x^n \in D_\psi\}\,, \tag{2.5}$$

where "co" stands for the convex hull and D_ψ represents the set of points where $\psi : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable. Note that the equality (2.5) coincides with the one in (2.4) for real-valued function. Following the expression in (2.5), the *B-subdifferential* (see, e.g., [37]) can be defined by

$$\partial_B\psi(\bar{x}) := \{\lim\ \nabla\psi(x^n) :\ \ x^n \to \bar{x},\ \ x^n \in D_\psi\}\,. \tag{2.6}$$

The concept of semismoothness will play an important role in convergence analysis of the Newton method that will be introduced in Section 4 of this paper. Let a function $\psi : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous around the point $\bar{x}$. Then, ψ will be said to be *semismooth* at $\bar{x}$ if the limit

$$\lim \{Vd' \mid V \in \partial\psi(\bar{x} + td'),\ d' \to d,\ t \downarrow 0\}$$

exists for all $d \in \mathbb{R}^n$. If, in addition,

$$Vd - \lim \{Vd' \mid V \in \partial\psi(\bar{x} + td'),\ d' \to d,\ t \downarrow 0\} = O(\|d\|^2)$$

holds for all $V \in \partial\psi(\bar{x} + d)$ with $d \to 0$, then ψ is said to be *strongly semismooth* at $\bar{x}$. The function ψ will be said to be ψ is SC^1 if it is continuously differentiable and $\nabla\psi$ is semismooth. Also, ψ is called *LC^2 function* if ψ is twice continuously differentiable and $\nabla^2\psi$ is locally Lipschitzian. Note that these concepts of semismoothness for a vector-valued function are extensions introduced in [38] from the original definitions for real-valued functions due to Mifflin [29].

2.2. Set-valued mappings and their generalized differentiation. Set-valued mappings and some related concepts will play a key role in the derivation of key results in this paper. At first, a set-valued map $\Psi\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *inner semicompact* at some point $\bar{x}$ with $\Psi(\bar{x}) \neq \emptyset$ if for every sequence $x_k \to \bar{x}$ with $\Psi(x_k) \neq \emptyset$ there is a sequence of $y_k \in \Psi(x_k)$ that contains a convergent subsequence as $k \to \infty$. It follows that the inner semicompactness holds in the finite-dimensional setting under consideration whenever Ψ is uniformly bounded around $\bar{x}$, i.e., there exists a neighborhood U of $\bar{x}$ and a bounded set $\Theta \subset \mathbb{R}^m$ such that $\Psi(x) \subset \Theta$ for all $x \in U$.

The mapping $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to be *inner semicontinuous* at $(\bar{x}, \bar{y}) \in \mathrm{gph}\Psi$ if for every sequence $x_k \to \bar{x}$ there is a sequence of $y_k \in \Psi(x_k)$ that converges to $\bar{y}$ as $k \to \infty$. For single-valued mappings $\Psi : \mathbb{R}^n \to \mathbb{R}^m$ this property obviously reduces to the continuity of Ψ at $\bar{x}$. Furthermore, if $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is inner semicompact at $\bar{x}$ with $\Psi(\bar{x}) = \{\bar{y}\}$, then Ψ is inner semicontinuous at $(\bar{x}, \bar{y})$. Obviously, the inner semicontinuity property is more restrictive than the inner semicompactness while bringing us to more precise results of the coderivative calculus, as it will be clear in the next section.

The *calmness* property of $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ holds at $(\bar{x}, \bar{y})$ with $\bar{y} \in \Psi(\bar{x})$ if there exist neighborhoods U and V of $\bar{x}$ and $\bar{y}$, respectively, and a constant $\ell > 0$ such that

$$d(y, \Psi(\bar{x})) \leq \ell \|x - \bar{x}\|, \quad \forall y \in V \cap \Psi(x), \quad \forall x \in U. \tag{2.7}$$

In the case where $V = \mathbb{R}^m$, this property goes back to Robinson [40] who called it the "upper Lipschitz property" of Ψ at $\bar{x}$. It is proved in [40] that the upper Lipschitz (and hence calmness) property holds at every point if the graph of Ψ is *piecewise polyhedral*, i.e., expressible as the union of finitely many polyhedral sets. Efficient conditions for the validity of the calmness property and its broad applications to variational analysis and optimization were strongly developed by Jiří Outrata and his collaborators; see, e.g., [22, 23, 33] among many other publications.

Considering some continuous functions $g_i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ for $i = 1, \ldots, p$ and $h_j : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ for $j = 1, \ldots, q$, we associate the set-valued mapping Ψ defined by

$$\Psi\left(x\right) := \left\{y \in \mathbb{R}^m \mid g_i\left(x, y\right) \leq 0,\ i = 1, \ldots, p,\ \ h_j\left(x, y\right) = 0,\ j = 1, \ldots, q\right\} \tag{2.8}$$

and the point $(\bar{x}, \bar{y}) \in \mathrm{gph}\Psi$ and set $\Omega \subset \mathbb{R}^n$, Ψ will be said to be *R-regular* at $(\bar{x}, \bar{y})$ w.r.t. $\Omega \subseteq \mathbb{R}^n$ if there are some positive numbers L, ϵ, and δ such that

$$d\left(y, \Psi\left(x\right)\right) \ \leq\ L \max\left\{0,\ \max\left\{g_i\left(x, y\right) \mid\ i = 1, \ldots, p\right\},\ \max\left\{\left|h_j\left(x, y\right)\right| \mid\ j = 1, \ldots, q\right\}\right\}. \tag{2.9}$$

for all $x \in \mathbb{U}_\delta\left(\bar{x}\right) \cap \Omega$ and $y \in \mathbb{U}_\epsilon\left(\bar{y}\right)$. For more details on R-regularity, see [28] and references therein.

To introduce generalized differentiation for a set-valued mapping, the concept of basic *normal cone* is used. For a closed subset C of $\mathbb{R}^n$, the *Mordukhovich* (also known as basic or limiting) normal cone to C at one of its points $\bar{x}$ is the set

$$N_C(\bar{x}) := \left\{ v \in \mathbb{R}^n \,\middle|\, \exists v_k \to v,\ x_k \to \bar{x}\,(x_k \in C) :\ v_k \in \widehat{N}_C(x_k) \right\}, \tag{2.10}$$

where $\widehat{N}_C$ denotes the dual of the contingent/Boulingand tangent cone to C:

$$\widehat{N}_C(\bar{x}) := \left\{v \in \mathbb{R}^n \middle|\ \langle v, u - \bar{x}\rangle \leq o(\|u - \bar{x}\|)\ \ \forall u \in C\right\}.$$

THEOREM 2.1 (Mordukhovich [30] and Rockafellar and Wets [41]). Let $C := \Omega \cap \psi^{-1}(\Xi)$, where $\Omega \subseteq \mathbb{R}^n$ and $\Xi \subseteq \mathbb{R}^m$ are closed sets and the function $\psi : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous function around $\bar{x}$. If there is no nonzero vector v such that $0 \in \partial\langle v, \psi\rangle(\bar{x}) + N_\Omega(\bar{x})$, $v \in N_\Xi(\psi(\bar{x}))$, then

$$N_C(\bar{x}) \subseteq \bigcup \left\{\partial\langle v, \psi\rangle(\bar{x}) + N_\Omega(\bar{x}) \mid v \in N_\Xi(\psi(\bar{x}))\right\}. \tag{2.11}$$

Equality holds in (2.11), provided that Ξ is normally regular at $\psi(\bar{x})$, i.e., $N_\Xi(\psi(\bar{x})) = \widehat{N}_\Xi(\psi(\bar{x}))$. This is obviously the case if Ξ is a convex set. Also note that in the theorem, the term $\partial\langle v, \psi\rangle(\bar{x})$

refers to the Mordukhovich subdifferential of the function $x \mapsto \sum_{i=1}^m v_i\psi_i(x)$ at $\bar{x}$. If $\psi : \mathbb{R}^n \to \mathbb{R}$, then the *Mordukhovich* (also known as basic) *subdifferential* of ψ at $\bar{x}$ can be defined by

$$\partial\psi(\bar{x}) := \{\xi \in \mathbb{R}^n |\, (\xi, -1) \in N_{\text{epi}\psi}(\bar{x}, \psi(\bar{x}))\},$$

where epiψ stands for the epigraph of ψ. If ψ is Lipschitz continuous around $\bar{x}$, then we can also define the *Clarke* (or convexified) *subdifferential* of ψ at $\bar{x}$:

$$\bar{\partial}\psi(\bar{x}) := \text{co}\,\partial\psi(\bar{x}).$$

Note that based on [22, Theorem 4.1], inclusion (2.11) also holds if the following set-valued mapping is calm at the point $(0, \bar{x})$:

$$\Psi(\varsigma) := \{x \in \Omega \,|\; \psi(x) + \varsigma \in \Xi\}. \tag{2.12}$$

Using the above concept of basic normal cone, we now introduce the notion of *coderivative* for a given set-valued map $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, at some point $(\bar{x}, \bar{y}) \in \text{gph}\,\Psi$, which corresponds to a homogeneous mapping $D^*\Psi(\bar{x}|\bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, defined by

$$D^*\Psi(\bar{x}|\bar{y})(y^*) := \{x^* \in \mathbb{R}^n |(x^*, -y^*) \in N_{\text{gph}\,\Psi}(\bar{x}, \bar{y})\}, \tag{2.13}$$

for all $y^* \in \mathbb{R}^m$. Here, $N_{\text{gph}\,\Psi}$ represents the basic normal cone (2.10) to gph Ψ.

We conclude this subsection with some further properties of set-valued mappings. Consider a set-valued mapping $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. It will be said to be *Lipschitz-like* around $(\bar{x}, \bar{y})$ if there exist neighbourhoods U of $\bar{x}$, V of $\bar{y}$, and a constant $l > 0$ such that

$$\Upsilon(x) \cap V \subseteq \Psi(u) + l \parallel u - x \parallel \mathbb{B} \quad \text{for all } x, u \in U. \tag{2.14}$$

with $\mathbb{B}$ denoting the unit ball in $\mathbb{R}^m$. If Ψ is a positively homogeneous mapping, its *outer norm* (resp. *inner norm*) is defined by

$$\parallel \Psi \parallel^+ = \sup_{x\in\mathbb{B}} \sup_{u\in\Psi(x)} \parallel u \parallel \quad (\text{resp. } \parallel \Psi \parallel^- = \sup_{x\in\mathbb{B}} \inf_{u\in\Psi(x)} \parallel u \parallel).$$

Furthermore, the infimum of all $l > 0$ for which (2.14) holds, also known as the *Lipschitz modulus* of Ψ at $(\bar{x}, \bar{y})$, is given via the outer norm of the coderivative of Ψ:

$$\text{lip}\Psi(\bar{x}, \bar{y}) = \inf\{l \in]0, +\infty[\; |\, (2.14) \text{ holds for some } U \text{ and } V\} = \parallel D^*\Upsilon(\bar{x}, \bar{y}) \parallel^+.$$

3. Coderivative and robust stability of solution maps. The plan here is to construct estimates for the coderivative of S (1.2) in terms of the derivatives of f_0 and g_0. In doing so, we provide new rules for the robust Lipschitz stability of S (1.2) based on the value function reformulation (1.4). To proceed, some further definitions and notations are in order. First, the Mangasarian-Fromowitz constraint qualification (MFCQ) for the constraint describing K (1.1), which will be said to hold at $(\bar{x}, \bar{y}, \hat{y})$ if there exists a vector $d \in \mathbb{R}^m$ such that

$$\nabla_3 g_{0j}(\bar{x}, \bar{y}, \hat{y})^\top d < 0 \quad \text{for} \quad j \in I^3 := I^3(\bar{x}, \bar{y}, \hat{y}) := \{i|\; g_{0_i}(\bar{x}, \bar{y}, \hat{y}) = 0\} \tag{3.1}$$

with $\nabla_3 g_{0j}(x, y, \varsigma)$ representing the Jacobian of the function $\varsigma \longmapsto g_{0j}(x, y, \varsigma)$ for a fixed vector $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$. Next, we associate to a vector $(\bar{x}, \bar{y}, \hat{y}, y^*)$ a set of Lagrange multipliers

$$\Lambda^0(\bar{x}, \bar{y}, \hat{y}, y^*) := \left\{ \begin{pmatrix} v \\ w \\ \lambda \end{pmatrix} \in \mathbb{R}^{2p+1} \;\middle|\; \begin{array}{l} \lambda \geq 0, \;\; w \in \Lambda(\bar{x}, \bar{y}, \hat{y}) \\ v \geq 0, \;\; g_0(\bar{x}, \bar{y}, \bar{y}) \leq 0, \;\; v^\top g_0(\bar{x}, \bar{y}, \bar{y}) = 0 \\ y^* + \lambda \left\{ \left[\sum_{k=1}^m \left(\delta_l f_k(\bar{x}, \bar{y}) + y_k \frac{\partial f_k}{\partial y_l}(\bar{x}, \bar{y}) \right) \right]_{l=1}^m \right. \\ \qquad \left. - \left[\sum_{\ell=1}^m \hat{y}_\ell \nabla_2 f_{0\ell}(\bar{x}, \bar{y}) + \nabla_2 g_0(\bar{x}, \bar{y}, \hat{y})^\top w \right] \right\} \\ \qquad\qquad + \nabla_2 g_0(\bar{x}, \bar{y}, \bar{y})^\top v = 0 \end{array} \right\}$$

related to the QVI (1.1), with $\delta_l := 1$ if $l = k$ and $\delta_l := 0$ if $l \neq k$, for $l = 1, \ldots, m$ and $k = 1, \ldots, m$. Recall that the other involved set of Lagrange multipliers $\Lambda(\bar{x}, \bar{y}, \hat{y})$ is defined by

$$\Lambda(z, \varsigma) := \left\{ w \in \mathbb{R}^q \,\middle|\, \nabla_2 \ell(z, \varsigma, w) = 0,\ \ w \geq 0,\ \ g_0(z, \varsigma) \leq 0,\ \ w^\top g_0(z, \varsigma) = 0 \right\}. \tag{3.2}$$

for $z := (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ and $\varsigma \in \mathbb{R}^m$. In (3.2), the function ℓ represents the Lagrangian function associated with the underlying parametric optimization problem (1.5):

$$\ell(z, \varsigma, w) := \varsigma^\top f_0(z) + w^\top g_0(z, \varsigma). \tag{3.3}$$

Furthermore, in (3.2), $\nabla_2 \ell(z, \varsigma, w) := f_0(z) + \sum_{j=1}^{q} w_j \nabla_2 g_{0j}(z, \varsigma)$ with $\nabla_2 g_{0j}(z, \varsigma)$ denoting the gradient of the function $\varsigma \to g_{0j}(z, \varsigma)$ when the vector $z := (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ is fixed.

Throughout, the linear independence constraint qualification (LICQ) holds at $(\bar{z}, \bar{\varsigma})$ if

$$\text{the family of vectors} \quad \left\{ \nabla_2 g_{0j}(\bar{z}, \bar{\varsigma}) \,\middle|\, j \in I^3 \right\} \quad \text{is linearly independent} \tag{3.4}$$

with I^3 given in (2.1) and $\nabla_2 g_0(z, \varsigma)$ stands for the gradient of g_0 w.r.t. the variable ς.

We can easily check that for a fixed parameter $x \in \mathbb{R}^n$, the set-valued mapping S (1.2) can be written as $S(x) = \{y \in \mathbb{R}^m | \, y \in S_0(x, y)\}$, where the mapping S_0 is defined from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^m$ by

$$S_0(x, y) := \arg\min_{\varsigma} \left\{ \varsigma^\top f_0(x, y) \,|\, \varsigma \in K(x, y) \right\}. \tag{3.5}$$

Theorem 3.1. *Consider a point $(\bar{x}, \bar{y})$ such that $\bar{y} \in S(\bar{x})$ and let $\hat{y} \in S_0(\bar{x}, \bar{y})$.*

(i) *Suppose that the MFCQ (3.1) holds at $(\bar{x}, \bar{y}, \bar{y})$ and $(\bar{x}, \bar{y}, \hat{y})$. Furthermore, assume that S_0 (3.5) is inner semicontinuous at $(\bar{x}, \bar{y}, \hat{y})$ and let the set-valued mapping*

$$\Psi(\theta) := \{(x, y) \in \Omega \,|\, f(x, y) - \varphi(x, y) \leq \theta \}, \tag{3.6}$$

with $\Omega := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \,|\, g(x, y) \leq 0 \}$, be calm at $(0, \bar{x}, \bar{y})$. Then, for all $y^* \in \mathbb{R}^m$,

$$\begin{aligned} D^* S(\bar{x}, \bar{y})(y^*) \subseteq & \bigcup_{(v, w, \lambda) \in \Lambda^0(\bar{x}, \bar{y}, \hat{y}, y^*)} \\ & \left\{ \nabla_1 g_0(\bar{x}, \bar{y}, \bar{y})^\top v + \lambda \left[\sum_{l=1}^{m} (\bar{y}_l - \hat{y}_l) \nabla_1 f_{0l}(\bar{x}, \bar{y}) - \nabla_1 g_0(\bar{x}, \bar{y}, \hat{y})^\top w \right] \right\}. \end{aligned} \tag{3.7}$$

(ii) *Suppose that in addition to the assumptions in* (i)*, the following qualification condition holds:*

$$\begin{aligned} & (v, w, \lambda) \in \Lambda^0(\bar{x}, \bar{y}, \hat{y}, 0) \implies \\ & \left[\nabla_1 g_0(\bar{x}, \bar{y}, \bar{y})^\top v + \lambda \left(\sum_{l=1}^{m} (\bar{y}_l - \hat{y}_l) \nabla_1 f_{0l}(\bar{x}, \bar{y}) - \nabla_1 g_0(\bar{x}, \bar{y}, \hat{y})^\top w \right) = 0 \right] \end{aligned} \tag{3.8}$$

Then the set-valued mapping S (1.2) is Lipschitz-like around $(\bar{x}, \bar{y})$ and its Lipschitz modulus at this point can be estimated by

$$\begin{aligned} \operatorname{lip} S(\bar{x}, \bar{y}) \leq \sup \Bigg\{ & \left\| \nabla_1 g_0(\bar{x}, \bar{y}, \bar{y})^\top v + \lambda \left[\sum_{l=1}^{m} (\bar{y}_l - \hat{y}_l) \nabla_1 f_{0l}(\bar{x}, \bar{y}) - \nabla_1 g_0(\bar{x}, \bar{y}, \hat{y})^\top w \right] \right\| : \\ & (v, w, \lambda) \in \Lambda^0(\bar{x}, \bar{y}, \hat{y}, y^*),\ \ \|y^*\| \leq 1 \Bigg\}. \end{aligned}$$

Proof. For (i), note that based on the assumptions made, we successively have the relationships

$$
\begin{aligned}
D^*S(\bar{x},\bar{y})(y^*) &\overset{(1)}{=} \left\{x^* \in \mathbb{R}^n \,\middle|\, (x^*,-y^*) \in N_{\text{gph}S}(\bar{x},\bar{y})\right\}\\
&\overset{(2)}{=} \left\{x^* \in \mathbb{R}^n \,\middle|\, (x^*,-y^*) \in N_{\Omega\cap\psi^{-1}(\mathbb{R}_-)}(\bar{x},\bar{y})\right\}\\
&\overset{(3)}{\subseteq} \left\{x^* \in \mathbb{R}^n \,|\, \exists\lambda \geq 0 : \ (x^*,-y^*) \in \partial\langle\lambda,\psi\rangle(\bar{x},\bar{y}) + N_\Omega(\bar{x},\bar{y})\right\}\\
&\overset{(4)}{\subseteq} \left\{x^* \in \mathbb{R}^n \,\middle|\, \begin{array}{r} \exists(v,\lambda): \ \ \lambda \geq 0, \ \ v \geq 0, \ \ g(\bar{x},\bar{y}) \leq 0, \ \ v^\top g(\bar{x},\bar{y}) = 0 \\ (x^*,-y^*) \in \nabla g(\bar{x},\bar{y})^\top v + \lambda\left(\nabla f(\bar{x},\bar{y}) - \bar{\partial}\varphi(\bar{x},\bar{y})\right) \end{array}\right\}\\
&\overset{(5)}{\subseteq} \bigcup_{(\lambda,v,w)\in\Lambda^0(\bar{x},\bar{y},\hat{y},y^*)} \left\{\nabla_1 g_0(\bar{x},\bar{y},\bar{y})^\top v \right.\\
&\qquad\qquad \left. + \lambda\left[\sum_{l=1}^m (\bar{y}_l - \hat{y}_l)\nabla_1 f_{0l}(\bar{x},\bar{y}) - \nabla_1 g_0(\bar{x},\bar{y},\hat{y})^\top w\right]\right\},
\end{aligned}
$$

where (1) corresponds to the definition of the coderivative (2.13) and (2) results from

$$\text{gph}\, S = \Omega \cap \psi^{-1}(\mathbb{R}_-) \ \text{ with } \ \Omega := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \,|\, g(x,y) \leq 0\}, \ \psi(x,y) := f(x,y) - \varphi(x,y),$$

where $g(x,y) := g_0(x,y,y)$. (3) is a consequence of the calmness of the mapping Ψ at $(0,\bar{x},\bar{y})$, as this guarantees, under the calmness framework (2.12), that

$$N_{\Omega\cap\psi^{-1}(\mathbb{R}_-)}(\bar{x},\bar{y}) \subseteq N_\Omega(\bar{x},\bar{y}) + \bigcup\left\{\partial\langle\lambda,\psi\rangle(\bar{x},\bar{y}) \,\middle|\, \lambda \in N_{\mathbb{R}_-}(\psi(\bar{x},\bar{y})) = \mathbb{R}_+\right\},$$

according to Theorem 2.1. As to (4), it follows from the fulfillment of the MFCQ (3.1) at $(\bar{z},\bar{y})$, as this ensures that we have the inclusion

$$N_\Omega(\bar{x},\bar{y}) \subseteq \left\{\nabla g(\bar{x},\bar{y})^\top v \,\middle|\, v \geq 0, \ \ g(\bar{x},\bar{y}) \leq 0, \ \ v^\top g(\bar{x},\bar{y}) = 0\right\}.$$

Finally, according to [32, Corollary 5.3], the combination of the inner semicontinuity of S_0 and MFCQ (3.1) both holding at the point $(\bar{x},\bar{y},\hat{y})$ leads to

$$\bar{\partial}\varphi(\bar{x},\bar{y}) \ \subseteq \ \left\{\sum_{\ell=1}^m \hat{y}_\ell \nabla f_{0\ell}(\bar{x},\bar{y}) + \nabla_{1,2} g_0(\bar{x},\bar{y},\hat{y})^\top w \,\middle|\, w \in \Lambda(\bar{x},\bar{y},\hat{y})\right\}. \tag{3.9}$$

It can easily be checked that (5) follows from this inclusion.

For assertion (ii), observe that from (3.7), we have the following inclusion:

$$
\begin{aligned}
D^*S(\bar{x},\bar{y})(0) \subseteq &\bigcup_{(v,w,\lambda)\in\Lambda^0(\bar{x},\bar{y},\hat{y},0)} \left\{\nabla_1 g_0(\bar{x},\bar{y},\bar{y})^\top v \right.\\
&\qquad \left. +\lambda\left[\sum_{l=1}^m (\bar{y}_l - \hat{y}_l)\nabla_1 f_{0l}(\bar{x},\bar{y}) - \nabla_1 g_0(\bar{x},\bar{y},\hat{y})^\top w\right]\right\}.
\end{aligned}
$$

Hence, it is clear that condition (3.8) is sufficient for $D^*S(\bar{x},\bar{y})(0) \subseteq \{0\}$. Thus, since the coderivative map is positively homogeneous, i.e., in particular $0 \in D^*S(\bar{x},\bar{y})(0)$, the coderivative/Mordukhovich criterion $D^*S(\bar{x},\bar{y})(0) = \{0\}$ is satisfied. This implies that S is Lipschitz-like around $(\bar{x},\bar{y})$. In this case, it is well-known (cf. [31]) that the exact Lipschitzian bound of S around $(\bar{x},\bar{y})$ is obtained as

$$\text{lip}\, S(\bar{x},\bar{y}) = \|D^*S(\bar{x},\bar{y})\| := \sup\{\|u\| \,|\, u \in D^*S(\bar{x},\bar{y})(y^*), \ \|y^*\| \leq 1\}.$$

Considering inclusion (3.7) one more time, we have the stated upper bound on $\text{lip}\, S(\bar{x},\bar{y})$. ∎

As it is clear from the proof of this theorem, the calmness of the set-valued mapping Ψ (3.6) is crucial. Hence, we are now going to provide some sufficient conditions to ensure that it holds. The first sufficient condition is based on the concept of *uniform weak sharp minimum condition*, which is well-known in parametric optimization; see, e.g., [12, 44] and references therein. Here, we introduced a version of the concept tailored to the parametric QVI (1.1).

Definition 3.2. *The QVI* (1.1) *will be said to satisfy a* local uniform weak sharp minimum *condition (LUWSMC) at* $(\bar{x}, \bar{y}) \in$ *gph* S*, if there exist some* $\alpha > 0$ *and* $\varepsilon > 0$ *such that*

$$\forall\, (x, y) \in \mathbb{U}_\epsilon\, (\bar{x}, \bar{y}) : \quad y \in K\,(x, y) \implies d(y,\, S(x)) \leq \alpha\, (f(x, y) - \varphi(x, y))\,. \tag{3.10}$$

Theorem 3.3. *If the LUWSMC holds at the point* $(\bar{x}, \bar{y}) \in$ *gph* S *then, the set-valued mapping* Ψ (3.6) *is calm at the point* $(0, \bar{x}, \bar{y})$.

Proof. Obviously, in this context, we have $\Psi(0) = \operatorname{gph} S$. For some numbers $\alpha > 0$ and $\varepsilon > 0$ such that (3.10) holds. Then, for any $(x, y, \varsigma) \in \mathbb{U}_\epsilon\, (\bar{x}, \bar{y}, 0)$ such that $(x, y) \in \Psi(\varsigma)$, we have

$$d\,((x, y), \Psi(0)) = d\,((x, y), \mathrm{gph} S) \leq d(y, S(x)) \leq \alpha\, (f(x, y) - \varphi(x, y)) \leq \alpha\varsigma = \alpha|\varsigma - 0|,$$

where the last inequality results from the fact that $(x, y) \in \Psi(\varsigma)$ implies that $\varsigma \geq 0$, as, by the definition of φ, it holds that $f(x, y) \geq \varphi(x, y)$ for all $y \in K\,(x, y)$. $\square$

Next, we provide another sufficient condition based on the R-regularity concept introduced in Subsection 2.2, but now for the fulfillment of the LUWSMC. To proceed, consider the expression of S in (1.4), and note that the R-regularity constraint qualification (RRCQ) will be said to hold at the point $(\bar{x}, \bar{y}) \in \operatorname{gph} S$ if S is R-regular (2.9) at $(\bar{x}, \bar{y})$ w.r.t. $\operatorname{dom} S$.

Proposition 3.4. *If the RRCQ holds at* $(\bar{x}, \bar{y})$ *and there is some neighborhood* $U \subset \mathbb{R}^n$ *of* $\bar{x}$ *and* $V \subset \mathbb{R}^n$ *of* $\bar{y}$ *such that dom* $K \cap (U \times V) = ($*dom* $S \cap U) \times V$*, then the LUWSMC is satisfied at* $(\bar{x}, \bar{y})$.

Proof. Fix $(\bar{x}, \bar{y}) \in$ gph S. Since, the mapping S is R-regular at $(\bar{x}, \bar{y})$ w.r.t. dom S, there exist $\sigma > 0$ and $\epsilon > 0$ such that for all $(x, y) \in \mathbb{U}_\epsilon\, (\bar{x}, \bar{y}) \cap (\text{dom } S \times V)$ we have the inequality

$$d\,(y, S\,(x)) \;\leq\; \sigma\; \max\{0, g\,(x, y)\,, f\,(x, y) - \varphi\,(x, y)\}.$$

From the definition of the optimal value function, for any $(x, y) \in \mathbb{U}_\epsilon\, (\bar{x}, \bar{y})$ with $g\,(x, y) \leq 0$, we have the inequality $f\,(x, y) - \varphi\,(x, y) \geq 0$. Hence, for all $(x, y) \in \mathbb{U}_\epsilon\, (\bar{x}, \bar{y}) \cap (\text{dom } S \cap U) \times V$, one gets

$$y \in K\,(x, y) \Longrightarrow d\,(y, S\,(x)) \;\leq\; \sigma\;\, (f\,(x, y) - \varphi\,(x, y))\,. \tag{3.11}$$

Since, dom $K \cap (U \times V) = (\text{dom } S \cap U) \times V$, we have for all $(x, y) \in \mathbb{U}_\epsilon\, (\bar{x}, \bar{y}) \cap (U \times V)$ that the condition (3.11) holds. Hence, the result. $\square$

Note that R-regularity for a mapping as stated in (2.8) is guaranteed under validity of the corresponding MFCQ. However, based on [12, 45], the MFCQ will automatically fail for the solution set-valued mappings S (1.4) and S_0 (3.5) with the value function reformulation. To overcome this failure, we next provide a more tractable framework for the fulfilment of the condition, based on the relaxed constant positive linear dependence constraint qualification (RCPLD), introduced in [2] and studied in various papers, including [5], which inspired thenext result. For the definition of the RCPLD and further studies around the topic, see the latter references, for example.

We now consider the specific case where we have a linear parametric QVI in the form (1.1) with

$$f_0(x, y) := Cy \quad \text{and} \quad K(x, y) := \{\varsigma \in \mathbb{R}^m |\; A\,(x, y)\,\varsigma \leq b\,(x, y)\}\,, \tag{3.12}$$

where $C \in \mathbb{R}^{m\times m}$ is a constant square matrix, $b : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ a twice continuously differentiable function, and $A : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{k\times m}$ being a matrix function made of twice continuously differentiable functions $a_{st}\,(x, y)$ for $s := 1, \ldots, k$ and $t := 1, \ldots, m$. Obviously, in this case, the corresponding solution set-valued mapping (1.4) can be rewritten as

$$S\,(x) = \{y \in K\,(x, y) \,|\; y^\top Cy - \varphi\,(x, y) \leq 0\}.$$

To state the corresponding sufficient condition for R-regularity, we need the dual problem

$$\max_{\sigma}\, \sigma^\top b\,(x, y) \quad \text{s.t.} \quad A\,(x, y)^\top \sigma = Cy, \;\, \sigma \geq 0$$

associated to (3.12). For each $s = 1, \dots, k$ define a single column matrix $D^s : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ with m rows, which are made of the elements

$$d_t^s(x,y) := \sum_{r=1}^{m} \frac{\partial a_{sr}(x,y)}{\partial y_t} y_r + a_{tt}(x,y) - \frac{\partial b_t(x,y)}{\partial y_t}, \ \ t = 1, \dots, m$$

and subsequently consider the single column matrix with mk rows obtained as

$$D(x,y) := \left[D^1(x,y), \ \ \dots, \ \ D^k(x,y)\right]^\top.$$

Proposition 3.5. *Fix $(\bar{x},\bar{y}) \in gph\ S$ and let $\bar{\varsigma} \in S_0(\bar{x},\bar{y})$ with corresponding dual optimal solution being $\bar{\sigma}_{I(\bar{x},\bar{y},\bar{\varsigma})}$, where $I(\bar{x},\bar{y},\bar{\varsigma}) := \{s = 1, \dots, m|\ \ a_s(x,y)^T \varsigma = b_s(x,y)\}$. Suppose that $\bar{\varsigma}$ is nondegenerate vertex and $\bar{\sigma}_s > 0$ for all $s \in I(\bar{x},\bar{y},\bar{\varsigma})$. Then, there exist a neighbourhood of $(\bar{x},\bar{y})$ in which φ is differentiable. Furthermore, for all (x,y), setting*

$$\mathcal{B}(x,y) := \left[\ 2Cy - \nabla\varphi(x,y) \qquad D(x,y)_{I(\bar{x},\bar{y},\bar{y})}\ \right]^\top,$$

if the set-valued mapping S_0 (3.5) is inner semicontinuous at $(\bar{x},\bar{y},\bar{\varsigma})$ and there is a neighbourhood U of $(\bar{x},\bar{y})$ such that for each index set $J \subset \{1, \cdots, |I(\bar{x},\bar{y},\bar{\varsigma})| + 1\}$, the matrix $\mathcal{B}_J(x,y)$ has constant row rank for all $(x,y) \in U$, then the mapping S (1.2) is R-regular at the point $(\bar{x},\bar{y})$.

Proof. Let $\bar{\varsigma} \in S_0(\bar{x},\bar{y})$ with corresponding dual optimal solution $\bar{\sigma}_{I(\bar{x},\bar{y},\bar{\varsigma})}$. Since $\bar{\varsigma}$ is nondegenerate vertex and $\bar{\sigma}_s > 0$ for $s \in I(\bar{x},\bar{y},\bar{\varsigma})$, one has from [42, Theorem 4.3] that the marginal function φ is differentiable around $(\bar{x},\bar{y})$. Moreover, recall that the solution set can be written as

$$S(x) = \{y \in \mathbb{R}^m | \bar{g}_0(x,y) = 0, \ \ \bar{g}_s(x,y) \leq 0, \ \forall s \in I\},$$

where $\bar{g}_0(x,y) = y^T C y - \varphi(x,y)$ and for each $s \in I$, $\bar{g}_s(x,y) := (x,y) = a_s(x,y)^\top y - b_s(x,y)$.

Now, observe that the assumptions of the the proposition guarantee that for each index set $J \subset \{1, \cdots, |I(\bar{x},\bar{y},\bar{y})| + 1\}$ the family possesses constant rank on U. Furthermore, S is inner semicontinuous at $(\bar{x},\bar{y})$ by inner semicontinuity of S_0 at $(\bar{x},\bar{y},\bar{\varsigma})$. We now can apply [5, Theorem 4.2] in order to obtain that S is R-regular at $(\bar{x},\bar{y})$. ∎

Example 3.6. Consider a QVI (1.1) of the form (3.12), defined with

$$C := \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad A(x,y) := \begin{pmatrix} -y_1 & 0 \\ 0 & -3y_2 \end{pmatrix}, \quad \text{and} \quad b(x,y) = \begin{pmatrix} x_1^2 \\ x_2^2 \end{pmatrix}. \tag{3.13}$$

The primal and dual optimization problems associated to QVI described in (3.13) can be written as

$$(\bar{P}_{x,y}) \ : \begin{cases} \min\limits_{\varsigma} \varsigma^T f_0(x,y) = \varsigma_1 y_1 + 2\varsigma_2 y_2 \\ \text{s.t.} \quad -\varsigma_1 y_1 + x_1^2 \leq 0, \\ \qquad\ \ -3\varsigma_2 y_2 + x_2^2 \leq 0, \end{cases} \quad \text{and} \quad (\bar{D}_{x,y}) \ : \begin{cases} \min\limits_{\sigma} \sigma x_1^2 + \sigma_2 x_2^2 \\ \text{s.t.} \quad \sigma_1 y_1 = y_1, \\ \qquad\ 3\sigma_2 y_2 = 2y_2 \leq 0, \end{cases}$$

respectively. Let $\frac{\sqrt{3}}{3} \geq a > 0$, $x \in \mathbb{R}_+^* \times \mathbb{R}_+^*$, and $y \in [a,x] \times [a,x]$. Then observe that

$$y = \left(x_1, \frac{\sqrt{3}}{3} x_2\right) \in K(x,y).$$

Furthermore, $y \in S(x)$. We can easily see that, if $y_1 > 0$ and $y_2 > 0$, the the points

$$\bar{\varsigma} = \left(\frac{x_1^2}{y_1}, \frac{x_2^2}{3y_2}\right) \quad \text{and} \quad \bar{\sigma} = \left(1, \frac{2}{3}\right)$$

are optimal for $\left(\bar{P}_{x,y}\right)$ and $\left(\bar{D}_{x,y}\right)$, respectively. Moreover, the LICQ holds for $\bar{\varsigma}$; additionally, as $\sigma_i > 0$ for $i = 1, 2$, the corresponding optimal value function $\varphi(x,y) = x_1^2 + \dfrac{2}{3}x_2^2$ is differentiable.

On the other hand, with $h_0(x,y) := y^T Cy - \varphi(x,y)$, the solution mapping S can be rewritten as

$$S(x) = \left\{ y \in \mathbb{R}^2 \,\middle|\, \begin{array}{lcl} h_0(x,y) & = & y_1^2 + 2y_2^2 - x_1^2 + \dfrac{2}{3}x_2^2 \leq 0 \\ h_1(x,y) & = & -\varsigma_1 y_1 + x_1^2 \leq 0 \\ h_2(x,y) & = & -3\varsigma_2 y_2 + x_2^2 \leq 0 \end{array} \right\}.$$

Let $\overline{y} \in S(\overline{x})$. Since,

$$\nabla_x h_0(x,y) = \begin{pmatrix} -2x_1 \\ \dfrac{4}{3}x_2 \end{pmatrix}, \quad \nabla_x h_1(x,y) = \begin{pmatrix} 2x_1 \\ 0 \end{pmatrix}, \quad \text{and} \quad \nabla_x h_2(x,y) = \begin{pmatrix} 0 \\ 2x_2 \end{pmatrix},$$

the matrix $\{\nabla_x h_0(x,y), \nabla_x h_1(x,y), \nabla_x h_2(x,y)\}$ has constant row rank for all (x,y) in a neighbourhood of $(\overline{x}, \overline{y})$ in which φ is differentiable. Therefore, from Proposition 3.5, the solution set-valued mapping S (1.2) corresponding to (3.13) is R-regular at the point $(\overline{x}, \overline{y})$.

Note that all the relationships discussed above can be summarized in the following diagram:

$$\text{LQVI}_\Omega \xRightarrow{\text{(Proposition 3.5)}} \text{RRCQ}_\Omega \xRightarrow{\text{(Proposition 3.4)}} \text{LUWSMC}_\Omega \xRightarrow{\text{(Theorem 3.3)}} \text{CALM}_\Omega$$

Fig. 1: *Note that LQVI represents the linear QVI considered in* (3.12). *For the definition of the RRCQ, see the discussion just before just before Proposition 3.4. See Definition 3.2 for the LUWSMC concept. As for CALM, it refers to the calmness of the set-valued mapping* Ψ (3.6) *assumed in Theorem 3.3. Note that the index* Ω *is used to differentiate* Ψ (3.6) *(based on* Ω*) from a slightly different mapping that will be introduced in the next section.*

As final point on the calmness requirement in Theorem 3.1, note that a parallel assumption can be made in the context of the generalized equation reformulation in (1.3); see [33, Theorem 3.1(ii)]. However, it was shown in [25] that in the context of the feasible set of a scenario of a bilevel optimization problem with unperturbed lower-level feasible set, which can have some similarity with a case of the QVI in (1.1) (see discussion in next section), the corresponding version of the calmness condition for the reformulation in (1.3) shows a behaviour better than the calmness of its version of the set-valued mapping Ψ (3.6). Nevertheless, our value function reformulation (1.4) of the QVI (1.1) has least two key advantages over its generalized equation model in (1.3): (1) it does not require convexity and (2) it does not require second order derivatives for the stability analysis in Theorem 3.1. We will see in the next section that these two points are really important in the design of a second order method for an optimization problem with QVI constraint.

To close this section, we are going to discuss the inner semicontinuity assumption on the set-valued mapping S_0 (3.5) also required in Theorem 3.1. First, note that this assumption can be weakened to the *inner semicompactness*. However, applying this weaker assumption will lead to a very loose upper bound of the coderivative of S in (3.7) due the convex hull operator that will appear on upper estimate of the subdifferential of φ (3.9); see, e.g., [10, 12]. This would subsequently lead to a stronger assumption in (3.8) as well as a looser upper bound for lip $S(\bar{x}, \bar{y})$ obtained in Theorem 3.1. Nevertheless, as discussed in Section 2, the inner semicompactness of S_0 is much weaker than its inner semicontinuity counterpart, and the latter holds, for example under the combination of the former and the uniqueness of S_0 at the corresponding point; for conditions ensuring the uniqueness of this mapping, see, e.g., [13] and references therein. The following adaptation of [28, Lemma 2.2] provides another scenario where S_0 (3.5) is inner semicontinuous without requiring its uniqueness.

PROPOSITION 3.7. *Let S_0 be R-regular at $(\bar{x},\bar{y},\bar{\varsigma}) \in gph\, S_0$ w.r.t. dom S_0. Moreover, let K be locally bounded at $(\bar{x},\bar{y}) \in dom\, S_0$ and inner semicontinuous at $(\bar{x},\bar{y},\bar{\varsigma})$ w.r.t. dom K, for some $\bar{\varsigma} \in K(\bar{x},\bar{y})$. Then S_0 is inner semicontinuous at $(\bar{x},\bar{y},\bar{\varsigma})$ w.r.t. dom S_0.*

Proof. With the notation $\overline{g}_0(x,y,\varsigma) := \varsigma^T f_0(x,y) - \varphi(x,y)$ and $\overline{g}_j(x,y,\varsigma) := g_j(x,y,\varsigma)$ for $j = 1,\dots,p$, the set-valued mapping S_0 can be rewritten as

$$S_0(x,y) = \{\varsigma \in \mathbb{R}^m |\ \overline{g}_{0_j}(x,y,\varsigma) \le 0,\ j \in J \cup \{0\}\}.$$

Let $(\bar{x},\bar{y},\bar{\varsigma}) \in \operatorname{gph} S_0$. Let $(x_k,y_k)_k \subset \operatorname{dom} S_0$ such that $x_k \to \bar{x}$ and $y_k \to \bar{y}$. First, we claim that φ is continuous at $(\bar{x},\bar{y})$. In deed, by definition, the functions $\overline{g}_1, \dots, \overline{g}_p$ are continuous (as they are continuously differentiable), then K is upper semicontinuous at $(\bar{x},\bar{y})$. Since, f_0 is continuous, it follows from [4, Theorem 4.2.1] that φ is lower semi continuous at $(\bar{x},\bar{y})$. Now, under the inner semicontinuity of K at $(\bar{x},\bar{y})$, we get from [4, Theorem 4.2.1], that φ is upper semicontinuous at $(\bar{x},\bar{y})$. Consequently, φ is continuous at $(\bar{x},\bar{y})$. On the other hand, the assumption of the proposition guarantee the existence of a constant $L > 0$, $\delta > 0$ and some $\epsilon > 0$ such that

$$d(\varsigma,\ S_0(x,y)) \ \le\ L \max\left\{0,\ \max\left\{\overline{g}_j(x,y,\varsigma) \mid\ j \in J \cup \{0\}\right\}\right\}.$$

for all $(x,y) \in \mathbb{U}_\epsilon(\bar{x},\bar{y})$ and all $\varsigma \in \mathbb{U}_\delta(\bar{\varsigma})$. Consequently,

$$d(\bar{\varsigma},\ S_0(x,y)) \ \le\ L \max\left\{0,\ \max\left\{\overline{g}_j(x,y,\bar{\varsigma}) \mid\ j \in J \cup \{0\}\right\}\right\}.$$

for all $(x,y) \in \mathbb{U}_\epsilon(\bar{x},\bar{y})$. Therefore, by continuity of $\overline{g}_0(x_k,y_k,\cdot), \overline{g}_1(x_k,y_k,\cdot), \cdots, \overline{g}_p(x_k,y_k,\cdot)$ and the choice of $(x_k,y_k)_k \subset \operatorname{dom} S_0$ the set $S_0(x_k,y_k)$ is nonempty and closed. Hence, there exist $\varsigma_k \in \Pi(\bar{\varsigma}, S_0(x_k,y_k))$ for sufficiently large k such that

$$\|\bar{\varsigma} - \varsigma_k\| \ \le\ L \max\left\{0,\ \max\left\{\overline{g}_j(x_k,y_k,\bar{\varsigma}) \mid\ j \in J \cup \{0\}\right\}\right\}.$$

Finally, the continuity of $\overline{g}_0, \overline{g}_1, \cdots, \overline{g}_p$ ensures that, $\lim\limits_{k\to+\infty} \|\bar{\varsigma} - \varsigma_k\| = 0$. Consequently S_0 is inner semicontinuous at $(\bar{x},\bar{y},\bar{\varsigma})$ w.r.t. dom S_0. ∎

4. Optimization problems with a QVI constraint. Recall that our primary goal in this section is to develop necessary optimality conditions and a solution algorithm for problem (1.6). Considering the value function reformulation (1.4) of the QVI (1.1), the problem can be rewritten as

$$\min\ F(z) \quad \text{s.t.} \quad G(z) \le 0,\ g(z) \le 0,\ f(z) - \varphi(z) \le 0, \tag{4.1}$$

taking into account the notation $z := (x,y)$, $g(x,y) := g_0(x,y,y)$, and $f(x,y) := y^\top f_0(x,y)$ introduced in Section 2, as well as the optimal value function φ defined in (1.5). In the next subsection, we start with the derivation of necessary optimality conditions for problem (4.1). Subsequently, a semismooth Newton method to solve the problem is proposed and tested on the BOLIB [47] library of bilevel optimization problems casted as OPQVI (1.6). Throughout this section, we assume that the functions F, G, f_0, and g_0 are twice continuously differentiable.

4.1. Necessary optimality conditions. We start here by recalling that a standard constraint qualification, such as the MFCQ, cannot hold for problem (4.1) [12, 45]. Hence, to derive necessary optimality conditions for the problem, we instead consider the partial penalization

$$\min_z\ F(z) + \lambda\left(f(z) - \varphi(z)\right) \quad \text{s.t.} \quad G(z) \le 0, \quad g(z) \le 0, \tag{4.2}$$

which moves the value function constraint $f(z) - \varphi(z) \le 0$, responsible for the failure of constraint qualifications, from the feasible set of problem (4.1) [12, 45]. In problem (4.2), $\lambda \in (0,\infty)$ corresponds to the penalization parameter. A close connection can be established between problems (4.1) and (4.2), based on the so-called *partial calmness* concept [45]. But before present the corresponding result, note that problem (4.1) will be said to be partially calm at one of its feasible points $\bar{z}$ if there is $\lambda \in (0, \infty)$ and a neighbourhood U of $(0, \bar{z})$ such that

$$F(z) - F(\bar{z}) + \lambda|\varsigma| \geq 0, \;\; \forall (\varsigma, z) \in U : G(z) \leq 0, \; g(z) \leq 0, \; f(z) - \varphi(z) + \varsigma = 0. \tag{4.3}$$

THEOREM 4.1. *Let $\bar{z}$ be locally optimal for problem (4.1). Then the problem is partially calm at $\bar{z}$ if and only if there exists $\lambda \in (0, \infty)$ such that this point is also locally optimal for problem (4.2).*

For a moment now, we are going to discuss sufficient conditions ensuring that problem (4.1) is partially calm. The first result of this series is a bit like a counterpart of Theorem 3.1.

THEOREM 4.2. *Let $\bar{z}$ be a local optimal solution of problem (4.1). The problem is partially calm at $\bar{z}$ if a set-valued mapping $\bar{\Psi}$, obtained by replacing Ω in (3.6) by the following set, is calm at $(0, \bar{z})$:*

$$\bar{\Omega} := \{ z := (x, y) \in \mathbb{R}^n \times \mathbb{R}^m \,|\, g(z) \leq 0, \;\; G(z) \leq 0 \} .$$

Proof. From the definition of the calmness of Ψ at $(0, \bar{z})$, as given in (2.7), there exists a neighborhood U of $(0, \bar{z})$ and a number $\ell > 0$ such that

$$d(z, \Psi(0)) \leq \ell|\varsigma| \;\; \forall (\varsigma, z) \in U, \;\; z \in \Psi(\varsigma). \tag{4.4}$$

As $\bar{z}$ is a local optimal solution of problem (4.1), $\Psi(0) \neq \emptyset$. Hence, consider a point $z^* \in \Psi(0)$ such that $d(z, \Psi(0)) = \|z - z^*\|$. Furthermore, also based on the fact that $\bar{z}$ is a local optimal solution of problem (4.1), assume without lost of generality that U is small enough such that $F(\bar{z}) \leq F(z^*)$, considering the fact that $\Psi(0)$ coincides with the feasible set of (4.1). Denoting by $\lambda > 0$ a Lipschtz constant of F near $\bar{z}$, it follows that for all $(\varsigma, z) \in U$ and $z \in \Psi(\varsigma)$,

$$F(\bar{z}) - F(z) \leq F(z^*) - F(z) \leq \lambda\|z - z^*\| \leq \ell\lambda|\varsigma|$$

by considering (4.4). The proof ends by comparing this condition with (4.3). ∎

Clearly, the framework of sufficient conditions ensuring the calmness of the set-valued mapping Ψ (3.6) can straightforwardly be extended to $\bar{\Psi}$ introduced in Theorem 4.2. Without unnecessarily repeating corresponding results from the previous section here, we can summarize them in the following extended counterpart of Figure 1:

$$\begin{tikzcd}
\mathrm{LQVI}_{\bar{\Omega}} \arrow[r, Rightarrow, "\overline{\text{(Proposition 3.5)}}"] & \mathrm{RRCQ}_{\bar{\Omega}} \arrow[r, Rightarrow, "\overline{\text{(Proposition 3.4)}}"] & \mathrm{LUWSMC}_{\bar{\Omega}} \arrow[d, Rightarrow, "\overline{\text{(Theorem 3.3)}}"] \\
\mathrm{PPEN} \arrow[r, Leftrightarrow, "\text{(Theorem 4.1)}"] & \mathrm{PACAL} & \mathrm{CALM}_{\bar{\Omega}} \arrow[l, Rightarrow, "\text{(Theorem 4.2)}"']
\end{tikzcd}$$

Fig. 2: *The index set $\bar{\Omega}$ is used here to distinguish the corresponding relationships with the ones in Figure 1, which are instead based on Ψ (3.6) defined with Ω. The lines over Proposition 3.5, Proposition 3.4, and Theorem 3.3 are used to specify that we are referring here to the versions of these results for the set-valued mapping $\bar{\Psi}$ introduced in Theorem 4.2. As for PACAL, it represents the partial calmness condition (4.3), while PPEN is used for the existence of $\lambda \in (0, \infty)$ such that a given point $\bar{z}$ is locally optimal for the partially penalized problem (4.2). Finally, note that for $LQVI_{\bar{\Omega}}$, $\bar{\Omega}$ is defined as in (3.12) with the additional constraint $G(x, y) := \bar{A}(x, y)y - \bar{b}(x, y) \leq 0$ associated to (1.6), with the functions $\bar{A}$ and $\bar{b}$ being of appropriate dimensions.*

Now, based on the partial penalization of problem (4.1), we are going to establish the necessary optimality conditions that will be at the center of the analysis in this section. To proceed, we need some constraint qualifications. The Mangasarian-Fromowitz constraint qualification (MFCQ) will be said to hold for G and g at a point $\bar{z}$ if there exists a vector $d \in \mathbb{R}^{n+m}$ such that we have

$$\nabla G_i(\bar{z})^\top d < 0, \;\; i \in I^1 \quad \text{and} \quad \nabla g_j(\bar{z})^\top d < 0, \;\; j \in I^2. \tag{4.5}$$

From now on, and as necessary, we will occasionally use the notation $g_0(z, \varsigma) := g_0(x, y, \varsigma)$ and obviously, $g_0(z, y) = g(z)$. Subsequently, $\nabla_1 g_0(z, \varsigma) := \nabla_{x,y} g_0(x, y, \varsigma)$ and $\nabla_2 g_0(z, \varsigma) := \nabla_3 g_0(x, y, \varsigma)$. However, we will write $\nabla g(z)$ for the derivative of the function $z := (x, y) \longmapsto g_0(x, y, y)$ at the point $z := (x, y)$. Based on these notation, the expression

$$\nabla_1 \ell(z, y, w) = \nabla f_0(z)^\top y + \nabla_1 g_0(z, y)^\top w$$

will be used in the sequel for the derivative w.r.t. $z := (x, y)$ of the Lagrangian of the underlying parametric optimization problem in (3.5).

THEOREM 4.3. *Let $z := (x, y)$ be a local optimal solution of problem (4.1). Suppose that S_0 is inner semicontinuous at the point $(z, y) \in gph\ S_0$, where the lower-level regularity condition (3.1) also holds. Furthermore, suppose that problem (4.1) is partially calm at z and upper-level regularity (4.5) is also satisfied at z. Then, there exist $\lambda \in]0, \infty[$, $u \in \mathbb{R}^p$, and $(v, w) \in \mathbb{R}^{2q}$ such that*

$$\nabla F(z) + \nabla G(z)^\top u + \nabla g(z)^\top v + \lambda \nabla f(z) - \lambda \nabla_1 \ell(z, y, w) = 0, \tag{4.6}$$

$$f_0(z) + \nabla_2 g_0(z, y)^\top w = 0, \tag{4.7}$$

$$u \geq 0, \;\; G(z) \leq 0, \;\; u^\top G(z) = 0, \tag{4.8}$$

$$v \geq 0, \;\; g(z) \leq 0, \;\; v^\top g(z) = 0, \tag{4.9}$$

$$w \geq 0, \;\; g_0(z, y) \leq 0, \;\; w^\top g_0(z, y) = 0. \tag{4.10}$$

Proof. The proof technique is well-known, see, e.g., [10, 12, 45], but we establish the result here for the sake of completeness. Start by observing that since S_0 (3.5) is inner semi-continuous at (z, y) and the MFCQ (3.1) holds at (z, y), then φ (1.5) is then locally Lipschitz continuous around z. Problem (4.1) is therefore a locally Lipschitz continuous optimization problem. Next, note that under the partial calmness condition, it follows from Theorem 4.1 that we can find a number $\lambda > 0$ such that z is a local optimal solution of problem (4.2). Applying the Lagrange multiplier rule for locally Lipschitz optimization on the latter problem, it follows that, as the upper-level regularity condition (4.5) holds at z, there exist $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^q$ such that (4.8) and (4.9) hold together with

$$\nabla F(z) + \nabla G(z)^\top u + \nabla g(z)^\top v + \lambda \nabla f(z) \in \lambda \bar{\partial} \varphi(z) \tag{4.11}$$

given that $\bar{\partial}(-\varphi)(z) = -\bar{\partial}\varphi(z)$, since φ is locally Lipschitz continuous around z. Furthermore, under the inner semicontinuity of S and fulfillment of the upper-level regularity condition (4.5) at z,

$$\bar{\partial}\varphi(z) \;\subseteq\; \left\{ \nabla_1 \ell(z, y, w) \;\middle|\; \begin{array}{c} \nabla_2 \ell(z, y, w) := f_0(z) + \nabla_2 g_0(z, y)^\top w = 0 \\ w \geq 0, \;\; g_0(z, y) \leq 0, \;\; w^\top g_0(z, y) = 0 \end{array} \right\}. \tag{4.12}$$

Obviously, combining (4.11) and (4.12), we get (4.6)–(4.7) and (4.10). ∎

A few comments on this result are in order. First, the lower- and upper-level regularity conditions as MFCQ can easily be verified. As for the inner semicontinuity assumption of the set-valued mapping S_0 (3.5), see relevant discussion at the end of the previous section, including Proposition 3.7. The only thing to add here with regards to relaxing the inner semicontinuity condition by using the inner semicompactness is that it would lead to a more complicated system of optimality conditions (see relevant discussion in [12]), which would create new challenges for the Newton scheme to be introduced in the next section. As for the partial calmness condition, see Figure 2 for a suitable framework ensuring that it holds. Finally, looking at the optimality conditions (4.6)–(4.10), it is clear, similarly to the discussion from the previous section, that they do not involve second order information as it is the case in [33], for example. Additionally, as shown in the series of papers [15, 17, 49, 43], optimality resulting from the value function approach have the potential to have a better numerical behaviour compared to the ones from GE-type reformulation. This was one of the motivations to study the version of the semismooth Newton method introduced in the next subsection.

4.2. Semismooth Newton-type method. From here on, we propose and study a method to solve the system (4.6)–(4.10). To proceed, the first thing to address is whether these conditions fully represent the optimality conditions of problem (1.6) based on reformulation (4.1). This comes as the underlying inner semicontinuity assumption in Theorem 4.3 presumes that $y \in S_0(z)$. It is important to point out that this inclusion is redundant if we impose the convexity of g_0 w.r.t. ζ in Theorem 4.3, as it will guaranty that the combination of (4.7) and (4.10) implies that $y \in S_0(z)$.

Secondly, we can easily check that the system (4.6)–(4.10) as a *nonsquare* system of equations. Using a trick introduced in [15], we can get a *square* system by adding the new variable $\xi \in \mathbb{R}^m$. To proceed, consider the Lagrangian-type function

$$L^\lambda(z,\xi,u,v,w) := \mathcal{L}^\lambda(z,u,v) - \lambda\ell(z,\xi,w),$$

where $\ell(z,\xi,w)$ is given in (3.3) and $\mathcal{L}^\lambda(z,u,v)$ is defined by

$$\mathcal{L}^\lambda(z,u,v) = F(z) + u^\top G(z) + v^\top g(z) + \lambda f(z)$$

Based on these tools, we can easily check that conditions (4.6)–(4.10) can be rewritten as

$$\Phi^\lambda(\zeta) := \begin{bmatrix} \nabla L^\lambda(z,\xi,u,v,w) \\ \sqrt{G(z)^2+u^2} + G(z) - u \\ \sqrt{g(z)^2+v^2} + g(z) - v \\ \sqrt{h(z,\xi)^2+w^2} + h(z,\xi) - w \end{bmatrix} = 0, \tag{4.13}$$

where the square root is understood vector-wise. Also, here, $h = g_0$, $\zeta := (z,\xi,u,v,w)$, and ∇L^λ represents the gradient of the function L^λ w.r.t. (z,ξ). This is actually a relaxation of the system (4.6)-(4.10), where we drop a constraint $y = \xi$. It is easy to show that if $S_0(z) = \{y\}$, then $\xi = y$ and this system reduces to (4.6)-(4.10). Most importantly, solving (4.13) is much easier because of its squareness and numerical experiments shows solving this system enables the proposed method to possess excellent performance in terms of finding local/global solutions.

Obviously, (4.13) is a system of $n + 2m + p + 2q$ equations with $n + 2m + p + 2q$ variables in ζ. This therefore allows for a natural extension of standard versions of the semismooth Newton method (see, e.g., [8, 14, 27, 39, 38]) to the bilevel optimization setting. In order to take full advantage of the structure of the function Φ^λ (4.13), we will use the following globalized version of the semismooth Newton method developed by De Luca et al. [8]. Recall that there are various other classes of functions generally known as NCP (nonlinear complementarity problem) functions that have been used in the literature to reformulate complementarity conditions into equations; see [18] and references therein for an extended list and related properties.

Note that the only difference between this algorithm and the original one in [8] is that in Step 0, we also have to provide the penalization parameter λ in (4.2). Also recall that in Step 2, $\partial_B\Phi^\lambda$ denotes the B-subdifferential (2.6). Obviously, equation $W^k d = -\Phi^\lambda(\zeta^k)$ has a solution if the matrix W^k is nonsingular. The latter holds in particular if the function Φ^λ is BD-regular. The function Φ^λ is said to be BD-regular at a point ζ if each element of $\partial_B\Phi^\lambda(\zeta)$ is nonsingular. Using this property, the convergence of Algorithm 4.1 can be established as follows [8]:

THEOREM 4.4. *Let problem* (1.6) *be* SC^1 *and* $\bar\zeta := (\bar z, \bar\xi, \bar u, \bar v, \bar w)$ *an accumulation point of a sequence generated by Algorithm 4.1 for some parameter* $\lambda > 0$. *Then* $\bar\zeta$ *is a stationary point of the problem of minimizing* Ψ^λ, *i.e.,* $\nabla\Psi^\lambda(\bar\zeta) = 0$. *If* $\bar\zeta$ *solves* $\Phi^\lambda(\zeta) = 0$ *and the function* Φ^λ *is BD-regular at* $\bar\zeta$, *then the algorithm converges to* $\bar\zeta$ *superlinearly and quadratically if problem* (1.6) *is* LC^2.

Note that problem (1.6) is SC^1 (resp. LC^2) if the functions F, G_i with $i = 1,\dots,p$, f_0, and $(g_0)_j$ with $j = 1,\dots,q$ are all SC^1 (resp. LC^2). Also note that problem (1.6) being SC^1 (resp. LC^2) guaranties that Φ^λ is semismooth (resp. strongly semismooth), cf. [27]. Results closely related to Theorem 4.4 are developed in [14, 27] and many other references therein. Observe that we have imposed BD-regularity in this theorem. We can replace it by the stronger CD-regularity, which refers to the non-singularity of all matrices in $\partial\Phi^\lambda(\bar\zeta)$. In the next section, we focus our attention on the derivation of conditions ensuring that CD-regularity holds.

Algorithm 4.1 Semi-smooth Newton Method for Bilevel Optimization

Step 0: Choose $\lambda, \epsilon, M > 0, \rho \in (0,1), \sigma \in (0,1/2), t > 2$, $\zeta^o := (z^o, \xi^o, u^o, v^o, w^o)$ and set $k := 0$.
Step 1: If $\|\Phi^\lambda(\zeta^k)\| < \epsilon$ or $k \leq M$, then stop.
Step 2: Choose $W^k \in \partial_B \Phi^\lambda(\zeta^k)$ and find the solution d^k of the system

$$W^k d = -\Phi^\lambda(\zeta^k).$$

If the above system is not solvable or if the condition

$$\nabla\Psi^\lambda(\zeta^k)^\top d^k \leq -\rho\|d^k\|^t \quad \text{with} \quad \Psi^\lambda(\zeta) := \frac{1}{2}\|\Phi^\lambda(\zeta)\|^2$$

is not satisfied, set $d^k = -\nabla\Psi^\lambda(\zeta^k)$.
Step 3: Find the smallest nonnegative integer s_k such that

$$\Psi^\lambda(\zeta^k + \rho^{s_k} d^k) \leq \Psi^\lambda(\zeta^k) + 2\sigma\rho^{s_k}\nabla\Psi^\lambda(\zeta^k)^\top d^k.$$

Then set $\alpha_k := \rho^{s_k}, \zeta^{k+1} := \zeta^k + \alpha_k d^k$, $k := k+1$ and go to **Step 1**.

4.2.1. CD-regularity. To provide sufficient conditions guaranteeing that CD-regularity holds for Φ^λ (4.13), we first construct an upper estimate of the generalized Jacobian of Φ^λ.

THEOREM 4.5. *Let the functions F, G, f_0, and g_0 be twice continuously differentiable at the point $\bar\zeta := (\bar z, \bar\xi, \bar u, \bar v, \bar w)$. If $\lambda > 0$, then Φ^λ is semismooth at $\bar\zeta$ and any $W^\lambda \in \partial\Phi^\lambda(\bar\zeta)$ can take the form*

$$W^\lambda = \begin{bmatrix} \nabla^2_{zz}L^\lambda(\bar\zeta) - \lambda\nabla^2_{zz}\ell(\bar\zeta) & -\lambda\nabla^2_{z\xi}\ell(\bar\zeta)^\top & \nabla G(\bar z)^\top & \nabla g(\bar z)^\top & -\lambda\nabla_1 h(\bar z,\bar\xi)^\top \\ -\lambda\nabla^2_{z\xi}\ell(\bar\zeta) & -\lambda\nabla^2_{\xi\xi}\ell(\bar\zeta) & O & O & -\lambda\nabla_1 h(\bar z,\bar\xi)^\top \\ \Lambda_1\nabla G(\bar z) & O & \Gamma_1 & O & O \\ \Lambda_2\nabla g(\bar z) & O & O & \Gamma_2 & O \\ \Lambda_3\nabla_1 h(\bar z,\bar\xi) & \Lambda_3\nabla_2 h(\bar z,\bar\xi) & O & O & \Gamma_3 \end{bmatrix} \tag{4.14}$$

where $\Lambda_i := \operatorname{diag}(a^i)$ and $\Gamma_i := \operatorname{diag}(b^i)$, $i = 1,2,3$, are such that

$$(a^i_j, b^i_j) \begin{cases} = & (0,-1) & \text{if } j \in \eta^i, \\ = & (1,0) & \text{if } j \in \nu^i, \\ \in & \{(\alpha,\beta) :\ (\alpha-1)^2 + (\beta+1)^2 \leq 1\} & \text{if } j \in \theta^i, \end{cases} \tag{4.15}$$

where η^i, ν^i, and θ^i with $i = 1,2,3$ are defined in (2.2) *and* (2.3)*.*

In the next result, we provide conditions ensuring that the function Φ^λ is CD-regular. To proceed, analogously to (3.4), the LICQ will be said to hold at $\bar z$ for problem 4.2 if the family of vectors

$$\left\{\nabla G_i(\bar z) :\ i \in I^1\right\} \cup \left\{\nabla g_j(\bar z) :\ j \in I^2\right\} \tag{4.16}$$

is linearly independent. Furthermore, let us introduce the cone of feasible directions

$$Q(\bar z, \bar\xi) := \left\{ (d^1; d^2) \left| \begin{array}{rll} \nabla G_j(\bar z)^\top d^1 & = 0, & j \in \nu^1 \\ \nabla g_j(\bar z)^\top d^1 & = 0, & j \in \nu^2 \\ \nabla h_j(\bar z,\bar\xi)^\top (d^1; d^2) & = 0, & j \in \nu^3 \end{array} \right. \right\},$$

where ν^i, $i = 1,2,3$ defined as in (2.2)–(2.3). The last condition is related the the strict complementarity condition (SCC). A point $(\bar z, \bar\xi, \bar w)$ will be said to satisfy the SCC if it holds that

$$\theta^3 := \theta^h(\bar z, \bar\xi, \bar w) = \emptyset. \tag{4.17}$$

For simplicity, we write $d^{12} := (d^1; d^2)$. By $\nabla^2\mathcal{L}^\lambda(\bar\zeta)$ and $\nabla^2\ell(\bar\zeta)$, we will denote the Hessian of the Lagrangian functions $\mathcal{L}^\lambda$ and ℓ w.r.t. z and (z,ξ), respectively.

THEOREM 4.6. *Assume that problem (1.6) is SC^1 and let the point $\bar{\zeta} := (\bar{z}, \bar{\xi}, \bar{u}, \bar{v}, \bar{w})$ satisfy the optimality conditions (4.6)–(4.10) for some $\lambda > 0$. Suppose that the LICQ (4.16) and (3.4) hold at $\bar{z}$ and $(\bar{z}, \bar{\xi})$, respectively. If additionally, for all $d^{12} \in Q(\bar{z}, \bar{\xi}) \setminus \{0\}$, we have*

$$(d^1)^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^1 > \lambda (d^{12})^\top \nabla^2 \ell(\bar{\zeta}) d^{12} \tag{4.18}$$

and the SCC (4.17) is also satisfied at $(\bar{z}, \bar{\xi}, \bar{w})$, then Φ^λ is CD-regular at $\bar{\zeta}$.

Proof. Let W^λ be any element from $\partial \Phi^\lambda(\bar{\zeta})$. Then, it can take the form described in Theorem 4.5, cf. (4.14)–(4.15). Hence, it follows that for any $d := (d^1, d^2, d^3, d^4, d^5)$ with $d^1 \in \mathbb{R}^{n+m}$, $d^2 \in \mathbb{R}^m$, $d^3 \in \mathbb{R}^p$, $d^4 \in \mathbb{R}^q$ and $d^5 \in \mathbb{R}^q$ such that $W^\lambda d = 0$, we have

$$\left[\nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) - \lambda \nabla^2_{zz} \ell(\bar{\zeta})\right] d^1 - \lambda \nabla^2_{z\xi} \ell(\bar{\zeta}) d^2 + \nabla G(\bar{z})^\top d^3 + \nabla g(\bar{z})^\top d^4 - \lambda \nabla_1 h(\bar{z}, \bar{\xi})^\top d^5 = 0, \tag{4.19}$$
$$- \lambda \nabla^2_{z\xi} \ell(\bar{\zeta}) d^1 - \lambda \nabla^2_{\xi\xi} \ell(\bar{\zeta}) d^2 - \lambda \nabla_2 h(\bar{z}, \bar{\xi})^\top d^5 = 0, \tag{4.20}$$
$$\forall j = 1, \dots, p,\ a^1_j \nabla G_j(\bar{z})^\top d^1 + b^1_j d^3_j = 0, \tag{4.21}$$
$$\forall j = 1, \dots, q,\ a^2_j \nabla g_j(\bar{z})^\top d^1 + b^2_j d^4_j = 0, \tag{4.22}$$
$$\forall j = 1, \dots, q,\ a^3_j \nabla h_j(\bar{z}, \bar{\xi})^\top d^{12} + b^3_j d^5_j = 0. \tag{4.23}$$

Adding $(d^1)^\top \times$(4.19) and $(d^2)^\top \times$(4.20) together derives

$$\begin{aligned} &(d^1)^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^1 - \lambda (d^{12})^\top \nabla^2 \ell(\bar{\zeta}) d^{12} \\ &\quad + (d^3)^\top \nabla G(\bar{z}) d^1 + (d^4)^\top \nabla g(\bar{z}) d^1 - \lambda (d^5)^\top \nabla h(\bar{z}, \bar{\xi}) d^{12} = 0. \end{aligned} \tag{4.24}$$

Recall that p and q represent the number of constraint functions G and g. For $i = 1, 2, 3$, let $p_1 := p$, $p_2 := q$ and $p_3 := q$. Then define 9 index sets by

$$P^i_1 := \{j \in \{1 \dots, p_i\} \mid a^i_j > 0,\ b^i_j < 0\}, \tag{4.25}$$
$$P^i_2 := \{j \in \{1 \dots, p_i\} \mid a^i_j = 0,\ b^i_j = -1\} \quad \supseteq \quad \eta^i, \tag{4.26}$$
$$P^i_3 := \{j \in \{1 \dots, p_i\} \mid a^i_j = 1,\ b^i_j = 0\} \quad \supseteq \quad \nu^i, \tag{4.27}$$

where the relations '$\supseteq$' can be verified easily from (4.15). For example, for any $j \in \eta^i$, we have $a^i_j = 0$, $b^i_j = -1$ from (4.15). Thus, $j \in P^i_2$. It follows from (4.21)–(4.23) that for $j \in P^1_2$, $j \in P^2_2$, and $j \in P^3_2$,

$$d^3_j = 0,\ d^4_j = 0,\ \text{ and } d^5_j = 0, \tag{4.28}$$

respectively due to $a^i_j = 0,\ b^i_j = -1$. As for $j \in P^1_3$, $j \in P^2_3$ and $j \in P^3_3$, we respectively get

$$\nabla G_j(\bar{z})^\top d^1 = 0,\ \nabla g_j(\bar{z})^\top d^1 = 0,\ \text{ and } \nabla h_j(\bar{z}, \bar{\xi})^\top d^{12} = 0. \tag{4.29}$$

due to $a^i_j = 1,\ b^i_j = 0$. Now observe that under the SCC (4.17), $\theta^3 = \emptyset$. Hence, from the corresponding counterpart of (4.15), it follows that $P^3_1 = \emptyset$. We can further check that for $j \in P^1_1$ and $j \in P^2_1$,

$$\nabla G_j(\bar{z})^\top d^1 = c^1_j d^3_j \ \text{ and } \ \nabla g_j(\bar{z})^\top d^1 = c^2_j d^4_j, \tag{4.30}$$

where $c^1_j := -b^1_j / a^1_j$ and $c^2_j := -b^2_j / a^2_j$, respectively. Since $P^3_1 = \emptyset$, $d^5_j = 0$ for $j \in P^3_2$ by (4.28) and $\nabla h_j(\bar{z}, \bar{\xi})^\top d^{12} = 0$ for $j \in P^3_3$ by (4.29), it follows that

$$(d^5)^\top \nabla h(\bar{z}, \bar{\xi}) d^{12} = \sum_{j \in P^3_2} d^5_j \nabla h_j(\bar{z}, \bar{\xi})^\top d^{12} + \sum_{j \in P^3_3} d^5_j \nabla h_j(\bar{z}, \bar{\xi})^\top d^{12} = 0 \tag{4.31}$$

Inserting (4.31) into (4.24) while taking into account (4.28)–(4.30), we get

$$0 = \underbrace{(d^1)^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^1 - \lambda (d^{12})^\top \nabla^2 \ell(\bar{\zeta}) d^{12}}_{=:\Delta} + \sum_{j \in P^1_1} c^1_j (d^3_j)^2 + \sum_{j \in P^2_1} c^2_j (d^4_j)^2 \geq \Delta,$$

where the inequality is owing to $c_j^1 > 0$ for $j \in P_1^1$ and $c_j^2 > 0$ for $j \in P_1^2$. Again, (4.29) and the fact that $\nu^i \subseteq P_3^i$ for $i = 1, 2, 3$ from (4.27) indicate that $d^{12} \in Q(\bar{z}, \bar{\xi})$, if $d^{12} \neq 0$ then $\Delta > 0$ from (4.18) which is contradicted with above inequality. Therefore, we have $d^{12} = 0$. This suffices to $d^1 = 0$, $d^2 = 0$, $d_j^3 = 0$ for $j \in P_1^1$ and $d_j^4 = 0$ for $j \in P_1^2$. Substituting these into (4.19)–(4.20), it follows from (4.28) and $P_1^3 = \emptyset$ that

$$\sum_{j \in P_3^1} d_j^3 \nabla G_j(\bar{z}) + \sum_{j \in P_3^2} d_j^4 \nabla g_j(\bar{z}) + \sum_{j \in P_3^3} (-\lambda d_j^5) \nabla_1 h_j(\bar{z}, \bar{\xi}) = 0, \tag{4.32}$$

$$\sum_{j \in P_3^3} d_j^5 \nabla_2 h_j(\bar{z}, \bar{\xi}) = 0. \tag{4.33}$$

Since the LICQ (3.4) is satisfied at $(\bar{z}, \bar{\xi})$ and $P_3^3 \subseteq I^3$ holds, it follows from (4.33) that $d_j^5 = 0$ for $j \in P_3^3$. Therefore $\sum_{j \in P_3^1} d_j^3 \nabla G_j(\bar{z}) + \sum_{j \in P_3^2} d_j^4 \nabla g_j(\bar{z}) = 0$. This together with the LICQ (4.16) at $\bar{z}$, and $P_3^i \subseteq I^i$ for $i = 1, 2$, we have $d_j^3 = 0$ for $j \in P_3^1$ and $d_j^4 = 0$ for $j \in P_3^2$. Hence, $d = 0$. ∎

For the next result, we consider one scenario to avoid imposing the SCC.

THEOREM 4.7. *Assume that problem (1.6) is SC^1 and let the point $\bar{\zeta} := (\bar{z}, \bar{\xi}, \bar{u}, \bar{v}, \bar{w})$ satisfy the optimality conditions (4.6)–(4.10) for some $\lambda > 0$. Suppose that the LICQ (4.16) and (3.4) hold at $\bar{z}$ and $(\bar{z}, \bar{\xi})$, respectively. If additionally, for all $(d^{12}, e) \in \left[Q(\bar{\zeta}) \times \mathbb{R}^{|P_1^3|}\right] \setminus \{0\}$,*

$$(d^1)^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^1 > \lambda \Big\{ (d^{12})^\top \nabla^2 \ell(\bar{\zeta}) d^{12} + \sum_{j \in P_1^3} c_j^3 (e_j)^2 \Big\}, \tag{4.34}$$

with $c_j^3 := -b_j^3 / a_j^3$ for $j \in P_1^3$ as (4.25). Then Φ^λ is CD-regular at $\bar{\zeta} := (\bar{z}, \bar{\xi}, \bar{u}, \bar{v}, \bar{w})$.

Proof. Also proceeding as in the proof of Theorem 4.6 while replacing (4.30) with

$$\nabla G_j(\bar{z}) d^1 = c_j^1 d_j^3,\ \nabla g_j(\bar{z}) d^{1,2} = c_j^2 d_j^4, \text{ and } \nabla h_j(\bar{z}, \bar{\xi}) d^{12} = c_j^3 d_j^5 \tag{4.35}$$

for $j \in P_1^1$, $j \in P_1^2$, and $j \in P_1^3$, respectively, we get equality

$$(d^1)^\top \nabla^2 \mathcal{L}^\lambda(\zeta) d^{1,2} - \lambda \left\{ (d^{12})^\top \nabla^2 \ell(\zeta) d^{12} + \sum_{j \in P_1^3} c_j^3 (d_j^5)^2 \right\} + \sum_{j \in P_1^1} c_j^1 (d_j^3)^2 + \sum_{j \in P_1^2} c_j^2 (d_j^4)^2 = 0$$

by inserting (4.28)–(4.29) and (4.35) in the counterpart of (4.24), as θ^3 is not necessarily empty. Hence, under assumption (4.34), we get $d^1 = 0$, $d^2 = 0$, $d^3 = 0$, $d_j^3 = 0$ for $j \in P_1^1$, $d_j^4 = 0$ for $j \in P_1^2$ and $d_j^5 = 0$ for $j \in P_1^3$. Similarly, the rest of the proof then follows as for Theorem 4.6. ∎

Considering the structure of the generalized second order subdifferential of φ (1.5) (see [46]), condition (4.34) can be seen as the most natural extension to our problem (4.2) of the strong second order sufficient condition used for example in [14, 27]. To see this, note that condition (4.34) can be replaced by the following condition, for all (d^{12}, e) in $\left[Q(\bar{\zeta}) \times \mathbb{R}^q\right] \setminus \{0\}$:

$$(d^1)^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^1 > \lambda \left\{ (d^{12})^\top \nabla^2 \ell(\bar{\zeta}) d^{12} + e^\top \nabla h(\bar{z}, \bar{\xi}) d^{12} \right\}.$$

4.2.2. Numerical experiments. Our aim in this section is to implement Algorithm 4.1, that we programmed in MATLAB (R2018a) on a Viglen desktop of 8GB memory and Inter(R) Core(TM) i5-4570 3.2Ghz CPU. To proceed with the presentation of the results, we first outline the implementation details. Recall that λ is a parameter for problem (4.2). Hence, one of the main difficulties in implementing Algorithm 4.1 is to be able to suitably select values of this parameter that can generate stationary points that could be optimal for problem (4.1). It therefore goes without saying that we cannot expect the same value of λ to be used for all the examples under consideration, as solutions obtained depend on this parameter. It turns out that we just need to take λ from a small set of constants. Precisely, we set $\lambda \in \{3^2, 3^1, 3^0, 3^{-1}, 3^{-2}\}$ for all the examples used in our experiments if there is no extra explanations.

For starting point, we choose $z^0 := [x^o; y^o]$ that is feasible for G and g_0 (or g), i.e., $G(z^o) \leq 0$ and $g(z^o) \leq 0$, respectively. If it is not easy to find such a point, we just set $x^o = 0$ and $y^o = 0$. After initializing (x^o, y^o), we choose $\xi^o = y^o$, $u^o = (|G_1(z^o)|, \ldots, |G_p(z^o)|)^\top$, $v^o = (|g_1(z^o)|, \ldots, |g_q(z^o)|)^\top$, and $w^o = v^o$. The choices of ξ^o, u^o, v^o, and w^o are not exclusive, as different initial points would render different solutions.

For the input parameters, we set $\epsilon = 10^{-6}$, $\beta = 10^{-8}$, $t = 2.1$, $\rho = 0.5$, $M = 1000$ and $\sigma = 10^{-4}$. We choose s^k from $\{-1, 0, 1, 2, \cdots\}$, which means that the starting step length of Newton direction is $\rho^{-1} = 2$. This contributes to make our algorithm render a desirable average performance over all examples. For the stop criteria, apart from setting $\|\Phi^\lambda(\zeta^k)\| < \epsilon$ or $k \leq M$, we also terminate the algorithm if $\|\Phi^\lambda(\zeta^k)\|$ does not change significantly within 100 steps, e.g., the variance of the vector $(\|\Phi^\lambda(\zeta^{k-100})\|, \|\Phi^\lambda(\zeta^{k-99})\|, \cdots, \|\Phi^\lambda(\zeta^k)\|)$ is less that 10^{-6}. The latter termination criterion was used for a very small portion of testing examples; resulting in a sequence $\{\zeta^k\}$ converging (to a point ζ^*), but which still did not satisfy $\|\Phi^\lambda(\zeta^*)\| < \epsilon$. A potential reason to explain such a phenomenon is that the corresponding selection of λ is not appropriate. For instance, for example `DempeDutta2012Ex24`, our method obtained the global optimal solution for all λ but with $\|\Phi^\lambda(\zeta^*)\| > \epsilon$ when $\lambda = 3^2$ and $\|\Phi^\lambda(\zeta^*)\| < \epsilon$ when $\lambda \in \{3^1, 3^0, 3^{-1}, 3^{-2}\}$. Therefore, to avoid such phenomenon which might result in unnecessary calculations, it is reasonable to terminate the algorithm if $\|\Phi^\lambda(\zeta^k)\|$ does not significantly vary within 100 steps.

Note that in order to pick elements from the generalized Jacobian of Φ^λ, cf. Step 2 of Algorithm 4.1, we extend the algorithm from [8] to our context.

Our main list of examples is based on bilevel optimization. But before moving to that, we first test the two examples considered in [33], one of the publications that motivated the work in this paper. The constraints in the first example describe the Nash game of two players [21] and in the second example describe the oligopolistic market equilibrium [34, 16].

Example 4.8. (`MordukhovichOutrata2007Ex63` [21, 33]) The problem is described as

$$F(x, y) := x - 3y_1 - 11y_2/3 + (y_1 - 9)^2/2$$

$$G(x, y) := \begin{bmatrix} -1 - x \\ -1 + x \end{bmatrix}$$

$$f_0(x, y) := \begin{bmatrix} -34 + 2y_1 + 8y_2/3 \\ -97/4 + 5y_1/4 + 2y_2 \end{bmatrix}$$

$$g_0(x, y, \xi) := \begin{bmatrix} y_1 + \xi_2 - 15 - x \\ y_2 + \xi_1 - 15 - x \end{bmatrix}$$

A local optimal solution to this problem is $\bar{x} = 0, \bar{y} = (9, 6)^\top$ with $F(\bar{x}, \bar{y}) = 49$.

Example 4.9. (`MordukhovichOutrata2007Ex64` [34, 16, 33]) The problem is described as

$$F(x, y) := (\max\{x - 135, 0\})^2 + 0.6(y_1 - 34)^2 + 0.6(y_2 - 16)^2$$

$$f_0(x, y) := \begin{bmatrix} -76 + 2y_1 + y_2 \\ -72 + y_1 + 2y_2 \end{bmatrix}$$

$$g_0(x, y, \xi) := \begin{bmatrix} -0.333x + \xi_1 + \xi_2 \\ -\xi_1 \\ -\xi_2 \end{bmatrix}$$

An optimal solution to this problem given in [33] is $\bar{x} = 135.15, \bar{y} = (30.95, 14.1)^\top$ with $F(\bar{x}, \bar{y}) = 7.77$.

Algorithm 4.1 is clearly related to the parameter λ. To see the impact of our method to λ, we solve Examples 4.8 and 4.9 by choosing $\lambda \in \{0.001, 0.002, \cdots, 0.5\}$. Results were reported in Figure 3, where let $\hat{x}, \hat{y}$ denote the solution obtained by Algorithm 4.1. For Example 4.8, Algorithm 4.1 yields the solutions such that $F(\hat{x}, \hat{y}) < F(\bar{x}, \bar{y})$ when $\lambda < 0.009$ due to $1 - F(\bar{x}, \bar{y})/F(\hat{x}, \hat{y}) > 0$ and $F(\hat{x}, \hat{y}) < 0$, and $F(\hat{x}, \hat{y}) = F(\bar{x}, \bar{y})$ when $\lambda \geq 0.009$ due to $1 - F(\bar{x}, \bar{y})/F(\hat{x}, \hat{y}) = 0$. Whilst for Example 4.9, $F(\bar{x}, \bar{y})$ increases and tends to be stable eventually with the rising of λ, with being smaller than $F(\hat{x}, \hat{y})$ when $\lambda < 0.018$.

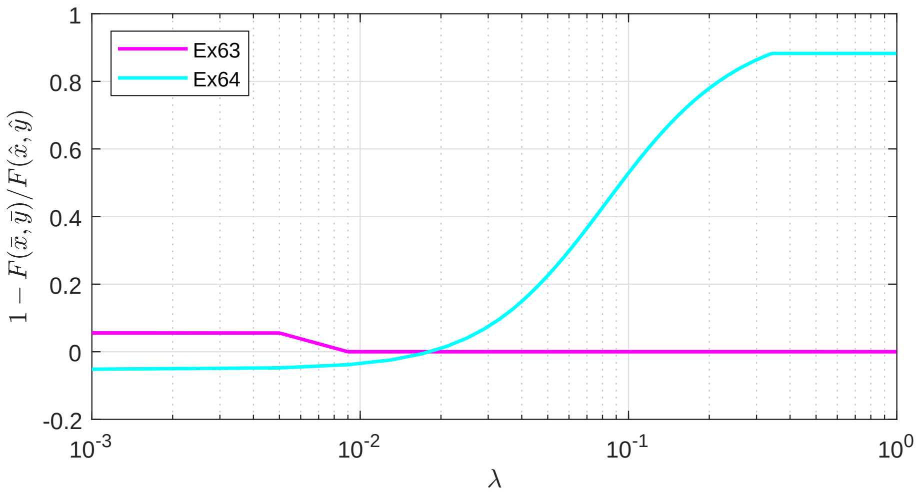

Fig. 3: The relative difference between the our and known objective function values, i.e., $1 - F(\bar{x}, \bar{y})/F(\hat{x}, \hat{y})$, under $\lambda \in [0.001, 1]$.

Next, we consider the test problems from the Bilevel Optimization LIBrary (BOLIB) [47]. Recall that a bilevel optimization can take the form

$$\min_{x,y} F(x, y) \quad \text{s.t.} \quad G(x, y) \le 0, \;\; y \in \arg\min_{y} \;\{f_1(x, y) : \;\; g_1(x, y) \le 0\}\,,$$

where $f_1 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $g_1 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$. This problem can obviously be cast as a QVI of the form (1.1) with $f_0(x, y) := \nabla_2 f_1(x, y)$, provided that the functions $f(x, .)$ and g_{1j}, $j = 1, \ldots, q$ are convex. We have tested Algorithm 4.1 on the 124 test problems contained in [47], which are all nonlinear bilevel optimization problems collected from the literature. Each example was solved using $\lambda \in \{3^2, 3^1, 3^0, 3^{-1}, 3^{-2}\}$ and two different starting points. Results of all examples were listed in Table 1, where $Sol.2$ represents our optimal value of the leader's objective function F and $Sol.1$ represents the known/true optimal optimal value of F. In addition, as we mentioned above, for each example we run Algorithm 4.1 under two starting points. However, to save space, results listed in Table 1 are generated under the first stating point. The the complete results and related technical details can be found in the supplementary material in [48].

No	**Examples**	$Sol.1$	$Sol.2$				
			$\lambda = 3^2$	$\lambda = 3^1$	$\lambda = 3^0$	$\lambda = 3^{-1}$	$\lambda = 3^{-2}$
1	`AiyoshiShimizu1984Ex2`	5.00	11.57	0.00	10.00	5.00	5.00
2	`AllendeStill2013`	-1.50	-6.25	-6.25	-7.14	-8.15	-8.49
3	`AnEtal2009`	2251.55	2251.55	2251.55	2251.55	2251.55	2251.55
4	`Bard1988Ex1`	17.00	48.42	30.59	17.00	7.68	2.00
5	`Bard1988Ex2`	-6600.00	-5772.97	-5852.78	-6600.00	-6600.00	-6600.00
6	`Bard1988Ex3`	-12.68	-10.36	-12.68	-12.79	-10.36	-10.36
7	`Bard1991Ex1`	2.00	2.00	2.00	2.00	2.00	2.00
8	`BardBook1998`	NaN	182.25	11.11	0.00	0.00	0.00
9	`Colson2002BIPA1`	0.00	0.00	0.00	0.00	0.00	0.00
10	`Colson2002BIPA2`	NaN	1.44	1.25	0.50	0.31	0.06
11	`Colson2002BIPA3`	0.31	1.40	0.50	0.50	0.31	0.06
12	`Colson2002BIPA4`	-29.20	-29.20	-13.00	-58.00	-58.00	-58.00

13	Colson2002BIPA5	5.00	5.00	7.43	5.00	1.70	0.36
14	CalamaiVicente1994a	250.00	250.00	250.00	250.00	250.00	250.00
15	CalamaiVicente1994b	17.00	25.00	30.59	17.00	17.00	2.00
16	CalamaiVicente1994c	2.00	2.00	2.00	2.00	2.00	2.00
17	CalveteGale1999P1	88.79	88.79	86.94	80.08	69.90	62.52
18	ClarkWesterberg1990a	2.75	2.75	2.75	2.96	0.89	0.20
19	Dempe1992a	NaN	0.00	0.00	-0.33	-0.81	-2.25
20	Dempe1992b	31.25	31.25	31.25	18.45	6.69	4.39
21	DempeDutta2012Ex24	0.00	14.05	172.12	251.19	0.00	0.00
22	DempeDutta2012Ex31	-1.00	0.00	-0.16	-0.46	-1.12	0.00
23	DempeFranke2011Ex41	-1.00	0.00	0.00	-1.00	-1.00	-1.00
24	DempeFranke2011Ex42	5.00	5.00	4.52	2.56	-0.22	-0.88
25	DempeFranke2014Ex38	2.13	3.13	2.13	0.61	-0.62	-0.93
26	DempeEtal2012	-1.00	-1.00	-1.00	-2.00	-3.00	-3.00
27	DempeLohse2011Ex31a	-6.00	0.00	-5.50	-5.50	-5.94	-5.99
28	DempeLohse2011Ex31b	-12.00	4.04	-12.00	-12.00	-12.00	-12.00
29	DeSilva1978	-1.00	-1.00	-1.00	-1.00	-1.28	-1.48
30	EdmundsBard1991	0.00	90.00	10.00	0.00	5.00	5.00
31	FalkLiu1995	-2.25	-2.25	-2.25	-2.25	-3.72	-3.98
32	FloudasZlobec1998	1.00	0.93	100.00	0.00	0.00	-1.00
33	GumusFloudas2001Ex1	2250.00	2250.00	2250.00	2018.86	1863.78	1652.20
34	GumusFloudas2001Ex3	-29.20	-58.00	-58.00	-58.00	-29.20	-58.00
35	GumusFloudas2001Ex4	9.00	9.00	9.00	9.00	4.84	1.00
36	GumusFloudas2001Ex5	0.19	0.19	0.19	0.19	0.19	0.19
37	HatzEtal2013	0.00	359.99	7414.39	0.00	0.00	0.00
38	HendersonQuandt1958	-3266.67	-3266.67	-3266.67	-3266.67	-3266.67	-3266.67
39	HenrionSurowiec2011	0.00	0.00	0.00	0.00	0.00	0.00
40	IshizukaAiyoshi1992a	0.00	0.00	0.00	0.00	0.00	0.00
41	KleniatiAdjiman2014Ex3	-1.00	-0.45	-1.00	-1.00	-1.00	-1.86
42	KleniatiAdjiman2014Ex4	-10.00	-3.18	-7.00	-9.00	-3.73	-5.04
43	LamparSagrat2017Ex23	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
44	LamparSagrat2017Ex31	1.00	1.00	1.00	1.00	1.00	1.00
45	LamparSagrat2017Ex32	0.50	0.50	0.50	0.50	0.50	0.50
46	LamparSagrat2017Ex33	0.50	0.50	0.50	0.50	0.50	0.50
47	LamparSagrat2017Ex35	0.80	1.25	0.80	0.80	0.03	0.00
48	LucchettiEtal1987	0.00	0.00	0.78	0.00	0.00	0.50
49	LuDebSinha2016a	1.94	1.11	1.11	1.11	1.11	1.11
50	LuDebSinha2016b	0.00	0.07	0.07	0.07	0.07	0.07
51	LuDebSinha2016c	1.12	1.12	1.97	1.12	1.12	1.12
52	LuDebSinha2016d	NaN	-155.04	-19.56	19.91	-16.27	-192.00
53	LuDebSinha2016e	NaN	9.75	7.85	28.65	5.44	8.35
54	LuDebSinha2016f	NaN	-160.00	-160.00	-160.00	-160.00	-18.54
55	MacalHurter1997	81.33	81.33	81.33	81.33	81.33	81.33
56	Mirrlees1999	0.01	0.01	0.01	0.01	0.01	0.01
57	MitsosBarton2006Ex38	0.00	0.00	0.00	0.00	0.00	0.00
58	MitsosBarton2006Ex39	-1.00	-0.02	-0.06	-0.33	-1.00	-1.00
59	MitsosBarton2006Ex310	0.50	0.50	-0.09	0.50	0.50	0.50
60	MitsosBarton2006Ex311	-0.80	-0.50	0.50	0.50	-0.80	-0.80
61	MitsosBarton2006Ex312	0.00	-1.00	-1.00	-1.00	-1.01	-1.02
62	MitsosBarton2006Ex313	-1.00	0.00	0.30	-2.00	-2.00	-2.00
63	MitsosBarton2006Ex314	0.25	0.06	0.06	0.06	0.03	0.00
64	MitsosBarton2006Ex315	0.00	0.00	-2.00	-2.00	-2.00	-2.00
65	MitsosBarton2006Ex316	-2.00	-0.54	-0.25	-2.00	-3.00	-3.00
66	MitsosBarton2006Ex317	0.19	0.19	0.19	0.00	0.00	0.00
67	MitsosBarton2006Ex318	-0.25	-1.00	-1.00	-1.00	-1.00	-1.00
68	MitsosBarton2006Ex319	-0.26	-0.26	-0.26	-0.26	-0.69	-1.20
69	MitsosBarton2006Ex320	0.31	0.05	0.04	0.02	0.00	0.00
70	MitsosBarton2006Ex321	0.21	0.21	0.21	0.21	0.00	0.00

71	MitsosBarton2006Ex322	0.21	0.21	0.21	0.21	0.01	0.01
72	MitsosBarton2006Ex323	0.18	0.18	0.18	0.05	0.18	0.18
73	MitsosBarton2006Ex324	-1.76	-1.75	-1.75	-1.75	-1.75	-1.76
74	MitsosBarton2006Ex325	-1.00	0.00	0.00	-1.00	0.00	0.00
75	MitsosBarton2006Ex326	-2.35	-2.00	-2.35	0.17	-1.19	-2.00
76	MitsosBarton2006Ex327	2.00	0.20	0.02	0.00	0.00	0.00
77	MitsosBarton2006Ex328	-10.00	-2.61	-4.24	-8.68	-7.94	-10.00
78	MorganPatrone2006a	-1.00	-1.00	-1.00	-1.00	-1.50	-1.50
79	MorganPatrone2006b	-1.25	0.53	-0.75	-1.25	-1.50	-1.50
80	MorganPatrone2006c	-1.00	1.00	1.00	1.00	-3.00	-3.00
81	MuuQuy2003Ex1	-2.08	-2.08	1.30	-2.24	-3.01	-3.68
82	MuuQuy2003Ex2	0.64	0.64	0.64	0.64	0.64	0.64
83	NieWangYe2017Ex34	2.00	1.98	2.00	2.00	2.00	2.00
84	NieWangYe2017Ex52	-1.71	-0.44	0.35	-2.14	-0.07	-1.41
85	NieWangYe2017Ex54	-0.44	-0.44	0.00	0.00	0.00	0.00
86	NieWangYe2017Ex57	-2.00	0.00	0.00	-2.04	0.00	-2.06
87	NieWangYe2017Ex58	-3.49	0.00	0.00	0.00	-3.53	-3.54
88	NieWangYe2017Ex61	-1.02	0.00	-0.11	-1.03	0.00	-4.16
89	Outrata1990Ex1a	-8.92	-8.92	-8.92	-8.92	-9.01	-11.15
90	Outrata1990Ex1b	-7.56	-7.58	-7.58	-7.58	-9.25	-11.21
91	Outrata1990Ex1c	-12.00	-6.00	0.00	-12.00	-12.00	-12.00
92	Outrata1990Ex1d	-3.60	-0.38	-0.37	-3.60	-3.88	-6.89
93	Outrata1990Ex1e	-3.15	-3.92	-3.79	-3.92	-4.21	-6.36
94	Outrata1990Ex2a	0.50	7.77	0.50	0.50	0.50	0.50
95	Outrata1990Ex2b	0.50	0.50	0.50	0.50	0.50	0.50
96	Outrata1990Ex2c	1.86	4.34	1.86	1.86	1.86	0.51
97	Outrata1990Ex2d	0.92	0.85	0.85	0.85	0.85	0.35
98	Outrata1990Ex2e	0.90	6.96	6.26	0.90	0.90	0.66
99	Outrata1993Ex31	1.56	1.56	11.96	1.56	1.56	0.88
100	Outrata1993Ex32	3.21	3.20	8.06	3.12	3.01	2.82
101	Outrata1994Ex31	3.21	3.20	3.18	3.12	3.01	2.82
102	OutrataCervinka2009	0.00	0.00	102.79	0.00	0.00	0.00
103	PaulaviciusEtal2017a	0.25	0.00	0.00	0.00	0.00	0.00
104	PaulaviciusEtal2017b	-2.00	-0.20	-2.00	-2.00	-2.00	-2.00
105	SahinCiric1998Ex2	5.00	13.00	5.00	4.58	1.70	0.36
106	ShimizuAiyoshi1981Ex1	100.00	99.09	97.33	92.60	82.00	66.33
107	ShimizuAiyoshi1981Ex2	225.00	225.00	225.00	175.00	118.06	113.12
108	ShimizuEtal1997a	NaN	48.42	30.59	14.04	25.00	25.00
109	ShimizuEtal1997b	2250.00	2329.21	2250.00	2018.86	1863.78	1652.20
110	SinhaMaloDeb2014TP3	-18.68	-5.27	-18.68	-18.79	-24.64	-26.98
111	SinhaMaloDeb2014TP6	-1.21	-1.21	-1.21	-1.21	-1.21	-1.21
112	SinhaMaloDeb2014TP7	-1.96	-1.98	-1.98	-1.98	-1.98	-1.98
113	SinhaMaloDeb2014TP8	0.00	0.00	73.47	25.00	4.63	0.17
114	SinhaMaloDeb2014TP9	0.00	0.00	0.00	0.00	0.00	0.00
115	SinhaMaloDeb2014TP10	0.00	0.00	0.00	0.00	0.00	0.00
116	TuyEtal2007	22.50	24.50	22.50	0.28	0.03	0.00
117	Vogel2012	0.00	0.00	0.00	0.00	0.00	0.00
118	WanWangLv2011	10.62	11.25	11.25	12.28	7.00	12.25
119	YeZhu2010Ex42	1.00	1.00	1.00	1.00	1.00	1.00
120	YeZhu2010Ex43	1.25	1.00	1.00	1.00	1.00	1.00
121	Yezza1996Ex31	1.50	1.50	1.50	1.50	1.00	2.00
122	Yezza1996Ex41	0.50	4.00	0.50	0.50	0.18	0.04
123	Zlobec2001a	-1.00	-1.00	0.00	-1.00	0.00	0.00
124	Zlobec2001b	NaN	0.13	1.00	1.00	0.00	0.08

Table 1: Upper-level objective function values at the solution for different selections of the penalty parameter $\lambda \in \{3^2, 3^1, 3^0, 3^{-1}, 3^{-2}\}$.

In the third column of Table 1, NaN stands for examples with unknown solution. Notice that from the original reference, some examples are associated with a parameter that should be provided by the user; i.e., for example, `CalamaiVicente1994a`, `HenrionSurowiec2011`, and `IshizukaAiyoshi1992a`. The first one is associated with $\rho \geq 1$, which separates the problem into 4 cases: (i) $\rho = 1$, (ii) $1 < \rho < 2$, (iii) $\rho = 2$, and (iv) $\rho > 2$. The results presented in Table 1 correspond to case (i). For other three cases, our method still produced the true global optimal solutions. Problem `HenrionSurowiec2011` has a unique global optimal solution $-0.5c(1, 1)$, where c is the parameter to be provided by the user. The results presented Table 1 is for $c = 0$. We also tested our method when $c = \pm 1$, and obtained the unique optimal solutions as well. Problem `IshizukaAiyoshi1992a` contains a parameter $M > 1$, and the results presented in Table 1 correspond to $M = 4$.

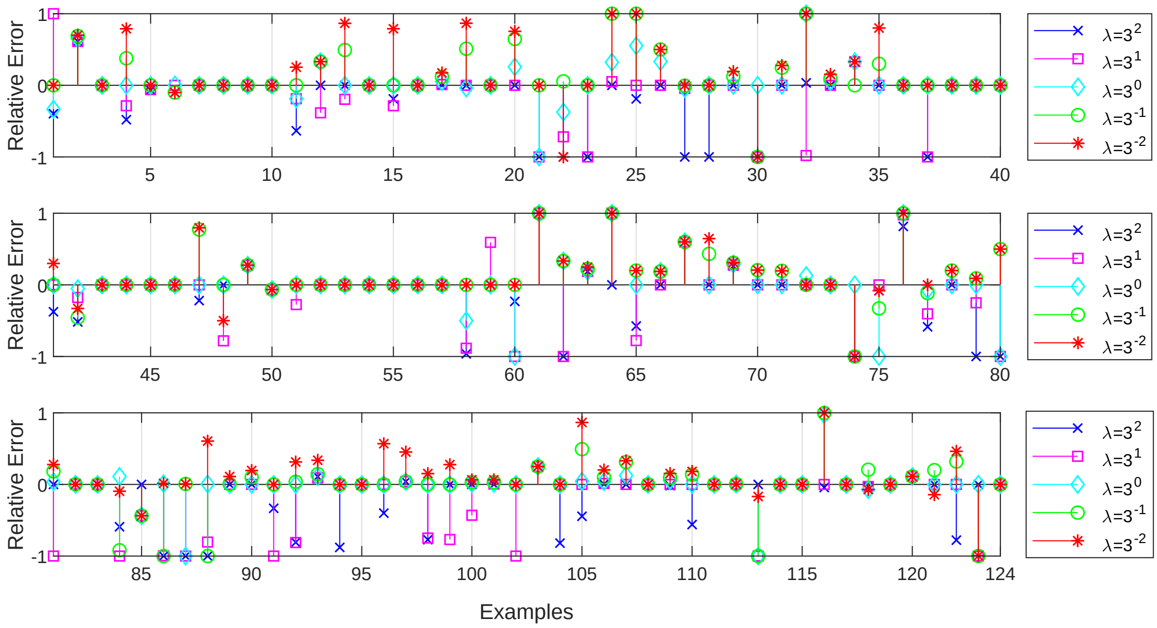

Fig. 4: Relative error of objective function values at the solution.

As the true solution for many of the examples in our test set are known (see supplementary material in [48]), we evaluate the accuracy of Algorithm 4.1 by computing the relative error

$$\text{Relative Error} := \frac{Sol.1 - Sol.2}{\max \ \{1, |Sol.1| + |Sol.2|\}} \tag{4.36}$$

between $Sol.1$ and $Sol.2$. We plot the values in Figure 4, while setting the relative error to zero, if a best solution for the example is unknown. As shown in Figure 4, for each example, if there is one marker among five markers that closely locates on the central line, i.e., Relative Error $= 0$, then our method under at least one λ from $\{3^2, 3^1, 3^0, 3^{-1}, 3^{-2}\}$achieved the true/known best solution $Sol.1$. Therefore one can count that our method closely achieved $Sol.1$ for all the examples except for 16 ones, with |Relative Error|≥ 0.05. They are problems Number 2, 22, 42, 49, 50, 61, 62, 63, 67, 69, 76, 80, 84, 93, 103 and 120 (cf. numbering in the first column of Table 1). However it is worth mentioning that not closely achieving $Sol.1$ does not mean that our method did not find a better solution because some known best solutions may not be the global optimal one. To be more precise, among those 16 examples, the known best solutions of 13 of them are optimal and 3 (i.e., problems number 22, 84, and 93) might not be optimal, which indicates that the solutions produced by our methods such that $Sol.2 < Sol.1$ could be better than the known ones for those examples. For $\lambda \in \{3^2, 3^1, 3^0, 3^{-1}, 3^{-2}\}$, there are respectively 73, 71, 82, 64 and 53 examples (out of 116, excluding the 8 examples with NaN) solved with $|\text{Relative Error}| < 0.05$. This seems to suggest that larger values λ perform better, in terms of approaching the true/known best solutions. However, note that for examples indicated with markers above the central line, we obtained a $Sol.2$s smaller than the corresponding $Sol.1$s, but wich does not necessarily fall within the latter threshold of the relative error. Also observe that for 30 examples out of the 124 (i.e., problems number 3, 7, 9, 14, 16, 36, 38, 39, 40, 43, 44, 45, 46, 49, 50, 55, 56, 57, 67, 73, 82, 95, 103, 111, 112, 114, 115, 117, 119 and 120), we obtain the same value for $Sol.2$ for all the aforementioned values of λ.

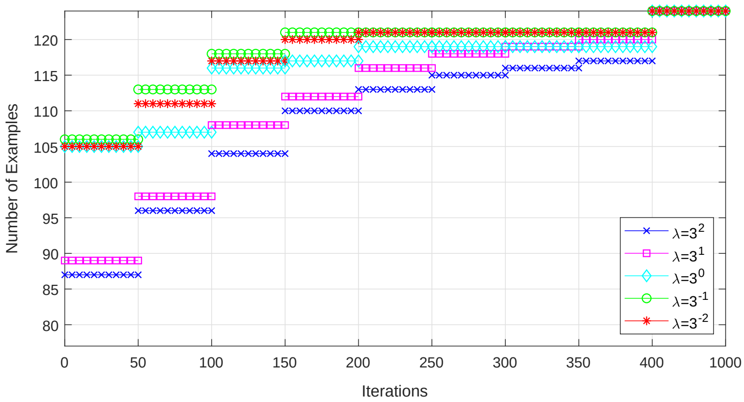

Fig. 5: Relation between the number of examples and iterations

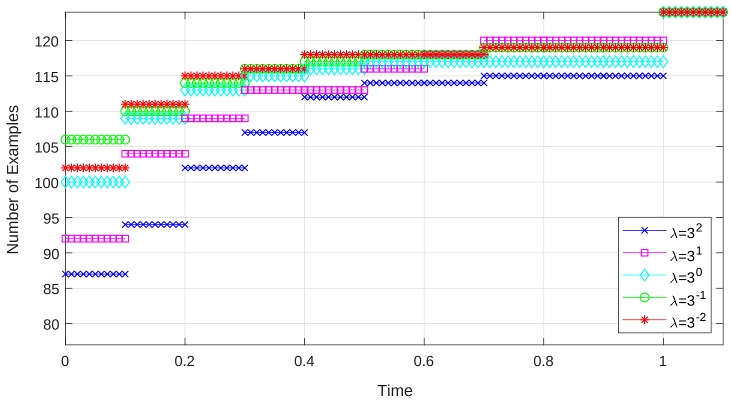

Fig. 6: Relation between the number of examples and CPU time.

For each value of λ considered, we plot the number of iterations in Figure 5. For example, when $\lambda = 3^{-1}$, over 106 examples were solved within 50 iterations, 113 examples within 100 iterations (which means $113 - 106 = 7$ examples were solved by using iterations between 51 and 100). For each λ, over 110 examples were solved within 200 iterations, and 7, 4, 5, 3, 3 examples were addressed by using more than 400 iterations respectively. Basically, for each interval (e.g., [101, 150]), compared with $\lambda \in \{3^2, 3^1\}$, more examples were solved when $\lambda \in \{3^0, 3^{-1}, 3^{-2}\}$.

Similarly, we plot the CPU time (in seconds) in Figure 6. For $\lambda = 3^{-1}$, 106 examples were solved within 0.1 second and 110 examples were solved within 0.2 second(which means $110 - 106 = 4$

examples were solved by using time between 0.1 and 0.2). Basically, for each λ, over 115 examples were solved within 1 second, which means that Algorithm 4.1 is quite fast. Moreover, for each interval (e.g., [0, 0.1]), i.e., under the same CPU time, compared with $\lambda \in \{3^2, 3^1\}$, more examples were solved when $\lambda \in \{3^0, 3^{-1}, 3^{-2}\}$. Considering the local version of Algorithm 4.1 (i.e., the corresponding version without the step size), it might be useful to evaluate how often it terminates with $\alpha^k = 1$. It turns out that for $\lambda = 3^2$ (resp. $\lambda = 3^1$, $\lambda = 3^0$, $\lambda = 3^{-1}$, and $\lambda = 3^{-2}$), 99 (resp. 101, 103, 110, and 109) examples are solved. This confirms that Algorithm 4.1 also have very good local behavior in practice.

Finally, we wanted to assess how often we one could have $y = \xi$ or $v = w$. As mentioned in Section 4.1, $y = \xi$ generates the stationary point in Theorem 4.3, i.e., (4.6)-(4.10). As for v and w, they are not necessarily related, as w is a multiplier associated to the lower-level constraint function but only in connection to the lower-level problem; v is associated is to the same function, g, but in connection to the upper-level problem. But considering the similarity in the complementarity systems (4.9) and (4.10), we wanted to look a bit closely to see how often they could coincide. To proceed, we recorded how many examples, for each λ, were solved with $y^k \approx \xi^k$ or $v^k \approx w^k$ (at the final iteration). The results are summarized in Table 2. For instance, when $\lambda = 3^2$, $42/124 = 33.87\%$ over all examples were solved with $y^k \approx \xi^k$ and $34/124 = 27.42\%$ over all examples was obtained with $v^k \approx w^k$. Note that examples solved with $y^k \approx \xi^k$ may not necessarily be the same solved with $v^k \approx w^k$. Also note that the symbol "$\approx$" used here refers to a very small absolute error (4.36), i.e., |Relative Error|< 0.05.

λ	3^2	3^1	3^0	3^{-1}	3^{-2}
$y^k \approx \xi^k$	42	40	33	32	30
$v^k \approx w^k$	34	34	40	25	22

Table 2: Number of problems solved with $y \approx z$ or $v \approx w$

REFERENCES

[1] L. Adam, R. Henrion, and J. Outrata, *On M-stationarity conditions in MPECs and the associated qualification conditions*, Mathematical Programming, 168 (2018), pp. 229-259.
[2] R. Andreani, G. Haeser, M.L. Schuverdt, and P.J.S. Silva, *A relaxed constant positive linear dependence constraint qualification and applications*, Mathematical Programming, 135(1) (2012), pp. 255-273.
[3] J.-P. Aubin, *Lipschitz behavior of solutions to convex minimization problem*, Mathematics of Operations Research, 9 (1984), pp. 87-111.
[4] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer, *Nonlinear parametric optimization*, Birkhäuser, 1983.
[5] E.M. Bednarczuk, L.I. Minchenko, and K.E. Rutkowski, *On Lipschitz-like continuity of a class of setvalued mappings*, Optimization, 69(12) (2019), pp. 2535-2549.
[6] D. Chan and J.S. Pang, *The generalized quasi-variational inequality problem*, Mathematics of Operations Research, 7(2) (1982), pp. 211-222.
[7] F.H. Clarke, *Optimization and nonsmooth analysis*, SIAM Classics in Applied Mathematics 5, 1994.
[8] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Mathematical Programming, 75(3) (1996), pp. 407-439.
[9] S. Dempe, *Foundations of bilevel programming*, Kluwer Academic Publishers, 2002.
[10] S. Dempe, J. Dutta, and B.S. Mordukhovich, *New necessary optimality conditions in optimistic bileve programming*, Optimization, 56(5-6) (2007), 577-604.
[11] S. Dempe and A.B. Zemkoho, *Bilevel Optimization: Advances and Next Challenges*, Springer, 2020.
[12] S. Dempe and A.B. Zemkoho, *The generalized Mangasarian-Fromowitz constraint qualification and optimality conditions for bilevel programs*, Journal of Optimization Theory and Applications, 148 (2011), pp. 46-68.
[13] A. Dreves, *Uniqueness for quasi-variational inequalities*, Set-Valued and Variational Analysis, 24 (2016), pp. 285-297.
[14] A. Fischer, *A special newton-type optimization method*, Optimization, 24(3-4) (1992), 269-284.
[15] A. Fischer, A.B. Zemkoho, and S. Zhou, *Semismooth Newton-type method for bilevel optimization: Global convergence and extensive numerical experiment*. Optimization Methods and Software, (2021), pp. 1-35, https://doi.org/10.1080/10556788.2021.1977810
[16] S.D. Flam and B. Kummer, *Great fish wars and Nash equilibria*, Technical report WP-0892, Department of Economics, University of Bergen, Norway, 1992.

[17] J. Fliege, A. Tin, and A.B. Zemkoho, *Gauss-Newton-type methods for bilevel optimization*, Computational Optimization and Applications, 78(3) (2021), pp.793-824.
[18] A. Galántai, *Properties and construction of NCP functions*, Computational Optimization and Applications, 52(3) (2012), pp. 805-824.
[19] H. Gfrerer and J.J. Ye, *New constraint qualifications for mathematical programs with equilibrium constraints via variational analysis*. SIAM Journal on Optimization, 27(2) (2017), pp. 842-865.
[20] H. Gfrerer and J.J. Ye, *New sharp necessary optimality conditions for mathematical programs with equilibrium constraints*, Set-Valued and Variational Analysis, 28(2) (2020), pp. 395-426.
[21] P.T. Harker, *Generalized Nash games and quasi-variational inequalities*, European journal of Operational research, 54(1) (1991), pp. 81-94.
[22] R. Henrion, A. Jourani, and J.V. Outrata, *On the calmness of a class of multifunctions*, SIAM Journal on Optimization, 13 (2002), pp. 603-618.
[23] R. Henrion and J.V. Outrata, *A subdifferential condition for calmness of multifunctions*, Journal of Mathematical Analysis and Applications, 258(1) (2001), 110-130.
[24] R. Henrion, J.V. Outrata, and T. Surowiec, *Analysis of M-stationary points to an EPEC modeling oligopolistic competition in an electricity spot market*. ESAIM: Control, Optimisation and Calculus of Variations, 18(2) (2012), pp. 295-317.
[25] R. Henrion and T. Surowiec, *On calmness conditions in convex bilevel programming*, Applicable Analysis, 90(6) (2011), pp. 951-970.
[26] B.F. Hobbs and J.S. Pang, *Nash-Cournot Equilibria in Electric Power Markets with Piecewise Linear Demand Functions and Joint Constraints*, Operations Research, 55(1) (2007), pp. 113-127.
[27] H. Jiang and L. Qi, *Semismooth karush-kuhn-tucker equations and convergence analysis of newton and quasi-newton methods for solving these equations*, Mathematics of Operations Research, 22(2) (1997), 301-325.
[28] P. Mehlitz, and L.I. Minchenko, *R-regularity of set-valued mappings under the relaxed constant positive linear dependence constraint qualification with applications to parametric and bilevel optimization*, Set-Valued and Variational Analysis, 30 (2022), pp. 179-205.
[29] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM Journal on Control and Optimization, 15(6) (1977), 959-972.
[30] B.S. Mordukhovich, *Generalized Differential Calculus for Nonsmooth and Set-Valued Mappings*, Journal of Mathematical Analysis and Applications 183(1) (1994), 250-288.
[31] B.S. Mordukhovich, *Variational analysis and generalized differentiation. I: Basic theory. II: Applications*, Springer, Berlin, 2006.
[32] B.S. Mordukhovich, M.N. Nam and H.M. Phan, *Variational analysis of marginal functions with applications to bilevel programming*, Journal of Optimization Theory and Applications, 152 (2012), pp. 557-586.
[33] B. S. Mordukhovich, and J. V. Outrata, *Coderivative analysis of quasi-variational inequalities with applications to stability and optimization*, SIAM Journal on Optimization, 18 (2007), pp. 389-412.
[34] F.H. Murphy, H.D. Sherali, and A.L. Soyster, *A mathematical programming approach for determining oligopolistic market equilibrium*, Mathematical Programming, 24 (1982), pp. 92-106.
[35] J.V. Outrata and J. Zowe, *A Newton method for a class of quasi-variational inequalities*, Computational Optimization and Applications, 4(1) (1995), pp. 5-21.
[36] J.S. Pang and M. Fukushima, *Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games*, Computational Management Science, 2(1) (2005), pp. 21-56.
[37] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of operations research, 18(1) (1993), pp. 227-244.
[38] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Mathematical Programming, 58 (1993), pp. 353-367.
[39] L. Qi and D. Sun, *A Survey of Some Nonsmooth Equations and Smoothing Newton Methods*, pp. 121-146, Springer US, Boston, MA, 1999.
[40] S.M. Robinson, *Some continuity properties of polyhedral multifunctions*, Mathematical Programming at Oberwolfach, (2009), pp. 206-214, https://doi.org/10.1007/BFb0120929.
[41] R.T. Rockafellar and R.J-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.
[42] G. Still, *Lectures on Parametric Optimization: An Introduction*, Optimization Online, 2018, https://optimization-online.org/2018/04/6587/.
[43] A. Tin and A.B. Zemkoho, *Levenberg-Marquardt method and partial exact penalty parameter selection in bilevel optimization*, Optimization and Engineering, (2022), pp. 1-43, https://doi.org/10.1007/s11081-022-09736-1
[44] J.J. Ye, *New uniform parametric error bounds*, Journal of Optimization Theory and Applications, 98(1) (1998), pp. 197-219.
[45] J.J. Ye, and D.L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization, 33 (1995), pp. 9-27.
[46] A.B. Zemkoho, *Estimates of generalized Hssians for optimal value functions in mathematical programming*, Set-Valued and Variational Analysis, 30 (2022), pp. 847-871.
[47] S. Zhou, A.B. Zemkoho, and A. Tin, *BOLIB: Bilevel Optimization Library of Test Problems*, In Bilevel Optimization (pp. 563-580). Springer, Cham.
[48] S. Zhou and A.B. Zemkoho, *Detailed numerical experiment for "Nonconvex quasi-variational inequalities: stability analysis and application to numerical optimization"*, Technical report, School of Mathematical Sciences, University of Southampton, Southampton, UK, 2022.
[49] A.B. Zemkoho and S. Zhou, *Theoretical and numerical comparison of the Karush-Kuhn-Tucker and value function reformulations in bilevel optimization*, Computational Optimization and Applications, 78(2) (2021), pp. 625-674.