Semismooth Newton-type method for bilevel optimization: global convergence and extensive numerical experiments

Andreas Fischer, Alain B. Zemkoho & Shenglong Zhou

Published online: 02 Dec 2021.

Submit your article to this journal ⬈

Article views: 64

View related articles ⬈

View Crossmark data ⬈

Semismooth Newton-type method for bilevel optimization: global convergence and extensive numerical experiments

Andreas Fischer [a], Alain B. Zemkoho [b] and Shenglong Zhou[c]

[a]Faculty of Mathematics, Technische Universität Dresden, Dresden, Germany; [b]School of Mathematics, University of Southampton, Southampton, UK; [c]Department of EEE, Imperial College London, London, UK

ABSTRACT

We consider the standard optimistic bilevel optimization problem, in particular upper- and lower-level constraints can be coupled. By means of the lower-level value function, the problem is transformed into a single-level optimization problem with a penalization of the value function constraint. For treating the latter problem, we develop a framework that does not rely on the direct computation of the lower-level value function or its derivatives. For each penalty parameter, the framework leads to a semismooth system of equations. This allows us to extend the semismooth Newton method to bilevel optimization. Besides global convergence properties of the method, we focus on achieving local superlinear convergence to a solution of the semismooth system. To this end, we formulate an appropriate CD-regularity assumption and derive sufficient conditions so that it is fulfilled. Moreover, we develop conditions to guarantee that a solution of the semismooth system is a local solution of the bilevel optimization problem. Extensive numerical experiments on 124 examples of nonlinear bilevel optimization problems from the literature show that this approach exhibits a remarkable performance, where only a few penalty parameters need to be considered.

1. Introduction

We consider the standard optimistic bilevel optimization problem

$$\min_{x,y} \ F(x,y) \quad \text{s.t.} \quad G(x,y) \leq 0, \quad y \in S(x), \tag{1}$$

also known as the upper-level problem, where the set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ describes the optimal solution set of the lower-level problem

$$\min_z f(x,z) \quad \text{s.t.} \quad g(x,z) \leq 0, \tag{2}$$

i.e.

$$S(x) := \operatorname*{argmin}_z \left\{ f(x,z) \mid g(x,z) \leq 0 \right\}. \tag{3}$$

Throughout the paper, the functions $F, f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$, and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$ are assumed to be twice continuously differentiable. As usual, we call F (resp. f) the upper-level (resp. lower-level) objective function, whereas G and g are called upper-level and lower-level constraint functions, respectively. Finally, x and y represent upper-level and lower-level variables.

In order to derive optimality conditions for the bilevel optimization problem or to treat it numerically, two main approaches for reformulating (1) as a single-level problem exist. One approach is to replace the lower-level problem by its Karush-Kuhn-Tucker (KKT) conditions. This leads to a mathematical program with complementarity constraints (MPCC). However, quite strong assumptions are needed to show that a local (or global) optimal solution of the MPCC yields a local (or global) optimal solution of the corresponding bilevel optimization problem; for details, the reader is referred to [7]. Therefore, we do not pursue this approach here. Nevertheless, we would like to underline that solving an MPCC is challenging, and only a few Newton-type methods with global or local fast convergence exist [28,34]. Even if we disregard the discrepancies between (local or global) solutions of the bilevel problem and the MPCC for a moment, it is an open question which conditions in the context of bilevel problems are implied by assumptions needed for the local convergence analysis of a Newton-type method for the corresponding KKT reformulation.

The second main approach to transform a bilevel program into a single-level optimization problem is the lower-level value function (LLVF) reformulation [40,55], which provides the basis for the developments in this paper. More in detail, the LLVF $\varphi : \mathbb{R}^n \to \mathbb{R}$ is given by

$$\varphi(x) := \min_z \left\{ f(x, z) \mid g(x, z) \leq 0 \right\}, \tag{4}$$

where, for the sake of simplicity, we assume throughout the paper that $S(x) \neq \emptyset$ for all $x \in \mathbb{R}^n$ so that φ is indeed finite-valued on $\mathbb{R}^n$. Then, w.r.t. local and global minimizers, the bilevel program (1) is equivalent to the optimization problem

$$\min_{x,y} F(x, y) \quad \text{s.t.} \quad G(x, y) \leq 0, \quad g(x, y) \leq 0, f(x, y) \leq \varphi(x). \tag{5}$$

In general, this is a nonconvex constrained optimization problem containing the typically nondifferentiable LLVF φ. Even if all the functions involved in (1) are fully convex (i.e. convex w.r.t. (x, y)), the feasible set of problem (5) is generally nonconvex.

Several algorithms for computing a stationary point of the LLVF reformulation were already designed and analyzed. For the case where the feasible set of the lower-level problem does not depend on the upper-level variable x such methods can be found in [35,53,54]. The algorithms in [8,9] were suggested to solve special cases of problem (5), where relaxation schemes are used to deal with the value function (4). Finally, the authors of [32] proposed numerical methods to solve special bilevel programs by exploiting a connection between problem (5) and a generalized Nash equilibrium problem. In addition, we note that the LLVF reformulation has been used recently for methods for certain classes of mixed integer bilevel programming problems, see, e.g. [22,36,51].

In this paper, we develop a framework for solving problem (5), which does not rely on the direct computation of the LLVF (4) or its derivatives, as it is the case in [35,53,54]. Thanks to this framework, we are able to extend the semismooth Newton method to bilevel optimization, for the first time in the literature. The ingredients used to construct the framework and to establish global convergence of the method are well-known in the

literature. At first, we use partial calmness [55] to move the value function constraint, i.e. $f(x, y) \leq \varphi(x)$, to the upper-level objective function, by means of partial exact penalization. For the resulting problem, necessary KKT-type optimality conditions are derived and reformulated as a square nonsmooth semismooth system of equations which depends on the penalty parameter.

For the aforementioned system of equations, global convergence of a semismooth Newton algorithm is established based on [4], see also [41,42,44] for related results. Obtaining local superlinear convergence of a semismooth Newton algorithm usually relies on the nonsingularity of some generalized Jacobian of the system of equations, see [17] and more general results in [14,31,43]. For our setting, we will derive conditions that ensure an appropriate nonsingularity property. To this end, upper- and lower-level linear independence conditions as well as a certain SSOSC (strong second order sufficient condition)-type condition will come into play. In standard nonlinear optimization, a similar setup guaranties that the reference point, where these conditions are satisfied, also corresponds to a locally optimal solution [45]; this is known as Robinson condition. We will derive a Robinson-type condition which guaranties that the reference point corresponds to a locally optimal solution for the penalized bilevel program. For this, the LLVF (4) is required to be second order directionally differentiable in the sense of [1,49].

For the algorithm studied in this paper, we have conducted detailed numerical experiments using the BOLIB (Bilevel Optimization LIBrary) [58] made of 124 examples of nonlinear bilevel optimization problems from the literature. The true optimal solutions are known for 70% of these problems. We were able to recover these solutions using a selection of just 9 values of the involved penalization parameter. However, it is important to emphasize that the primary goal of the method is to compute stationary points based on problem (5). For each of the 9 values of the penalization parameter, the method converges for at least 87% of the problems. The algorithm also exhibits a good experimental order of convergence (at least greater or equal to 1.5 for at least about 70% of the problems) for different values of the parameters. To the best of our knowledge, this is the first time where an algorithm is proposed for nonlinear bilevel optimization, with computational experiments at such a scale and such a level of success.

The paper is organized as follows. In the next section, we recall some basic notions and properties, centred around the generalized first and second order differentiation of the LLVF (4). Section 3 discusses the stationarity concept that will be the basis for the semismooth system suggested later on. There, we recall some basic tools and the general framework for deriving optimality conditions for the LLVF reformulation (5). In Section 4, based on the reformulation in [17], we suggest a semismooth system of equations to rewrite the stationarity conditions and establish a semismooth Newton method for computing stationary points. In Section 5, we derive sufficient conditions for the CD-regularity of this semismooth system at a solution in order to guarantee superlinear or quadratic convergence, and in Section 6, the Robinson-type condition. Finally, Section 7 presents results of a numerical study of the semismooth Newton method on the test problems in the BOLIB library [58].

2. Preliminaries

We first introduce some basic notation that will be used throughout. Depending on the situation, both y and z will be used as lower-level variables. According to this, for some

$(\bar{x}, \bar{y})$ and $(\bar{x}, \bar{z})$, the index sets

$$I^1 := I^G(\bar{x}, \bar{y}) := \left\{ i \mid G_i(\bar{x}, \bar{y}) = 0 \right\} \tag{6}$$

of the active upper-level constraints and

$$I^2 := I^g(\bar{x}, \bar{y}) := \left\{ j \mid g_j(\bar{x}, \bar{y}) = 0 \right\} \quad \text{and} \quad I^3 := I^g(\bar{x}, \bar{z}) \tag{7}$$

for the active lower-level constraints are defined. Moreover, given a multiplier $\bar{u} \in \mathbb{R}^p$ for the upper-level constraint function G and multipliers $\bar{v}, \bar{w} \in \mathbb{R}^q$, each for the lower-level constraint function g, then the index sets

$$
\begin{aligned}
\eta^1 &:= \eta^G(\bar{x}, \bar{y}, \bar{u}) := \{ i \mid \bar{u}_i = 0, G_i(\bar{x}, \bar{y}) < 0 \}, \\
\theta^1 &:= \theta^G(\bar{x}, \bar{y}, \bar{u}) := \{ i \mid \bar{u}_i = 0, G_i(\bar{x}, \bar{y}) = 0 \}, \\
\nu^1 &:= \nu^G(\bar{x}, \bar{y}, \bar{u}) := \{ i \mid \bar{u}_i > 0, G_i(\bar{x}, \bar{y}) = 0 \}
\end{aligned}
\tag{8}
$$

are defined as subsets of the upper-level indices $\{1, \ldots, p\}$ and, similarly,

$$
\begin{aligned}
\eta^2 &:= \eta^g(\bar{x}, \bar{y}, \bar{v}), \quad \theta^2 := \theta^g(\bar{x}, \bar{y}, \bar{v}), \quad \text{and} \quad \nu^2 := \nu^g(\bar{x}, \bar{y}, \bar{v}), \\
\eta^3 &:= \eta^g(\bar{x}, \bar{z}, \bar{w}), \quad \theta^3 := \theta^g(\bar{x}, \bar{z}, \bar{w}), \quad \text{and} \quad \nu^3 := \nu^g(\bar{x}, \bar{z}, \bar{w})
\end{aligned}
\tag{9}
$$

are defined as subsets of the lower-level indices $\{1, \ldots, q\}$. The lower-level Lagrangian function ℓ defined from $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q$ to $\mathbb{R}$ is given by

$$\ell(x, z, w) := f(x, z) + w^\top g(x, z). \tag{10}$$

For a function depending on upper-level and (or) lower-level variables, we will often use the notations ∇_1 and ∇_2 to denote the gradients of this function w.r.t. the upper-level (resp. lower-level) variable, to avoid any potential confusion. Similarly, ∇_2^2, ∇_1^2, and ∇_{12}^2 could be used when referring to second order derivatives.

Since the optimal value function φ (4) is nondifferentiable in general, we need generalized concepts of differentiability. Let us first recall the usual notion of the directional derivative. For a function $\psi : \mathbb{R}^n \to \mathbb{R}^m$, its directional derivative at $\bar{x} \in \mathbb{R}^n$ in the direction $d \in \mathbb{R}^n$ is the limit

$$\psi'(\bar{x}; d) := \lim_{t \downarrow 0} \frac{1}{t} \left[\psi(\bar{x} + td) - \psi(x) \right], \tag{11}$$

provided it exists. Differentiable and componentwise convex (not necessarily differentiable) functions are examples of directionally differentiable functions [46]. The optimal value function φ (4) can be directionally differentiable without being differentiable or convex. To underline this, we are going to recall a result from [23], that will play an important role in this paper. To proceed, we will say that the lower-level linear independence

constraint qualification (LLICQ) holds at a point $(\bar{x}, \bar{z})$ if the family of vectors

$$\left\{\nabla_2 g_i(\bar{x}, \bar{z}) \mid g_i(\bar{x}, \bar{z}) = 0\right\} \tag{12}$$

is linearly independent. For any (x, z), the set of lower-level Lagrange multipliers is given by

$$\Lambda(x, z) := \left\{w \in \mathbb{R}^q \mid \nabla_2 \ell(x, z, w) = 0, \ w \geq 0, \ g(x, z) \leq 0, \ w^\top g(x, z) = 0\right\}. \tag{13}$$

Also, let us define the set-valued map $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ by

$$K(x) := \left\{z \in \mathbb{R}^m \mid g(x, z) \leq 0\right\}, \tag{14}$$

which provides all lower-level feasible points for a given value of x.

Theorem 2.1 ([23, Corollary 4.4]): *Let K (14) be uniformly compact near $\bar{x}$ (i.e. there is a neighbourhood U of $\bar{x}$ such that the closure of the set $\bigcup_{x \in U} K(x)$ is compact) and $K(\bar{x}) \neq \emptyset$. Suppose that LLICQ (12) holds at $(\bar{x}, z)$ for all $z \in S(\bar{x})$. Then, φ is directionally differentiable at $\bar{x}$ in any direction $d \in \mathbb{R}^n$ with*

$$\varphi'(\bar{x}; d) = \min_{z \in S(\bar{x})} \nabla_1 \ell(\bar{x}, z, w_z)^\top d, \tag{15}$$

where, for any $z \in S(\bar{x})$, w_z is the unique element in $\Lambda(\bar{x}, z)$.

Note that the set of lower-level Lagrange multipliers $\Lambda(\bar{x}, z)$ is a singleton for all $z \in S(\bar{x})$ because LLICQ is assumed for all these z. If the lower-level problem is convex (i.e. f and g_i, $i = 1, \ldots, p$ are convex functions), a version of formula (15) is given in [25] without requiring uniqueness of lower-level Lagrange multipliers. In the absence of convexity, corresponding formulas of $\varphi'(\bar{x}; d)$ not requiring single-valuedness of Λ are given in [47]. However, in the latter case, some second order assumptions are imposed on problem (2). Also, the uniform compactness of K can be replaced by other types of assumptions, i.e. the inner semicontinuity or compactness of the lower-level solution map S (3); see [38] for details.

As it is shown in [1], one can define a second order directional derivative for $\psi : \mathbb{R}^n \to \mathbb{R}$ at $\bar{x}$ in directions d and e by

$$\psi''(\bar{x}; d, e) := \lim_{t \downarrow 0} \frac{1}{t^2} \left[\psi\left(\bar{x} + td + \frac{1}{2}t^2 e\right) - \psi(\bar{x}) - t\psi'(\bar{x}; d)\right], \tag{16}$$

provided the limit exists. Next, we recall a formula for the second order directional derivative of the LLVF φ (4) developed in [49]. For this purpose, a few more assumptions are needed. A point $(\bar{x}, \bar{z}, \bar{w})$ is said to satisfy the lower-level strict complementarity condition (LSCC) if

$$\theta^3 = \{j \mid \bar{w}_j = g_j(\bar{x}, \bar{z}) = 0\} = \emptyset. \tag{17}$$

The lower-level submanifold property (LSMP) is said to hold at $(\bar{x}, d)$ if for all $\bar{z}$ such that

$$\bar{z} \in S_1(\bar{x}; d) := \underset{z \in S(\bar{x})}{\operatorname{argmin}} \nabla_1 \ell(\bar{x}, z, w_z)^\top d, \tag{18}$$

the restriction of $S(\bar{x})$ on a neighbourhood of $\bar{z}$ is a smooth submanifold of $\{z \mid g_j(\bar{x}, z) = 0 \text{ for all } j \in I^3\}$. Finally, the lower-level second order sufficient condition (LSOSC) is said

to hold at $(\bar{x}, d)$ if

$$e^\top \nabla_2^2 \ell(\bar{x}, z, w_z) e > 0$$

for all $z \in S_1(\bar{x}; d)$ and all $e \in [T_\circ(z)]^\perp \setminus \{0\}$ such that $\nabla_2 g_j(\bar{x}, z)^\top e = 0$ for all $j \in I^3$. Here $T_\circ(z)$ denotes the tangent space [33] of $S(\bar{x})$ at z. For further details and discussions on these assumptions and related results, including the following one, see [2,49].

Theorem 2.2: *Let the assumptions of Theorem 2.1 be satisfied and suppose that LSCC holds at $(\bar{x}, z, w)$ for all $z \in S_1(\bar{x}; d)$. If, additionally, the LSMP and LSOSC hold at $(\bar{x}, d)$, then φ (4) is second order directionally differentiable at $\bar{x}$ in directions d and e with*

$$\varphi''(\bar{x}; d, e) = \min_{z \in S_1(\bar{x}; d)} \left\{ \nabla_1 \ell(\bar{x}, z, w_z)^\top e + \xi_d(\bar{x}, z) \right\}, \tag{19}$$

with $S_1(\bar{x}; d)$ given in (18) while $\xi_d(\bar{x}, z)$ is defined by

$$\begin{cases} \xi_d(\bar{x}, z) := \min_{e \in \mathcal{Z}_d(\bar{x}, z)} (d^\top, e^\top) \nabla^2 \ell(\bar{x}, z, w_z)(d^\top, e^\top)^\top, \\ \mathcal{Z}_d(\bar{x}, z) := \left\{ e \in \mathbb{R}^m \mid \nabla_1 g_i(\bar{x}, z)^\top d + \nabla_2 g_i(\bar{x}, z)^\top e = 0, \; i \in I^3 \right\}. \end{cases} \tag{20}$$

Note that $\nabla^2 \ell$ stands for the Hessian w.r.t. the vector made of the first and second variables of the function ℓ (10). We can simplify the assumptions needed in this theorem if we impose that the lower-level problem (3) has a unique optimal solution for $x := \bar{x}$; cf. the following theorem from [48].

Theorem 2.3: *Let the assumptions of Theorem 2.1 be satisfied with $S(\bar{x}) = \{\bar{y}\}$ and suppose that*

$$e^\top \nabla_2^2 \ell(\bar{x}, \bar{y}, \bar{w}) e > 0, \quad \forall e \neq 0 \quad s.t. \quad \begin{cases} \nabla_2 g_i(\bar{x}, \bar{y})^\top e = 0 & \text{for } i \in \nu^3, \\ \nabla_2 g_i(\bar{x}, \bar{y})^\top e \leq 0 & \text{for } i \in \theta^3. \end{cases}$$

Then, we have the expression

$$\varphi''(\bar{x}; d, e) = \nabla_x \ell(\bar{x}, \bar{y}, \bar{w})^\top e + \xi_d(\bar{x}, \bar{y}), \tag{21}$$

where $\xi_d(\bar{x}, \bar{y})$ is defined as in (20), with $z = \bar{y}$, and $\mathcal{Z}_d(\bar{x}, \bar{y})$ given by

$$\mathcal{Z}_d(\bar{x}, \bar{y}) := \left\{ e \mid \nabla_1 g_i(\bar{x}, \bar{y})^\top d + \nabla_2 g_i(\bar{x}, \bar{y})^\top e = 0 \text{ for } i \in \nu^3, \right.$$

$$\left. \nabla_1 g_i(\bar{x}, \bar{y})^\top d + \nabla_2 g_i(\bar{x}, \bar{y})^\top e \leq 0 \text{ for } i \in \theta^3 \right\}.$$

Next, recall that a function $\psi : \mathbb{R}^n \to \mathbb{R}^m$, which is Lipschitz continuous around $\bar{x}$, is differentiable almost everywhere around this point. Let D_ψ denote the set of points, where ψ is differentiable. Then, the generalized Jacobian in the sense of Clarke [3] is defined by

$$\partial \psi(\bar{x}) := \text{co} \partial_B \psi(\bar{x}) \quad \text{with } \partial_B \psi(\bar{x}) := \left\{ \lim_{n \to \infty} \nabla \psi(x^n) \mid x^n \to \bar{x}, \; x^n \in D_\psi \right\}, \tag{22}$$

where 'co' stands for the convex hull and $\partial_B \psi(\bar{x})$ is called B-subdifferential of ψ at $\bar{x}$. If ψ is a real-valued function, then $\partial \psi(\bar{x})$ reduces to the Clarke subdifferential.

To complete this section, we briefly review the concept of semismoothness, which is used for the convergence result of the Newton method to be discussed in this paper. Let a function $\psi : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous around $\bar{x}$. Then, ψ is called semismooth at $\bar{x}$ if the limit

$$\lim \left\{ Vd' \mid V \in \partial \psi (\bar{x} + td'),\ d' \to d,\ t \downarrow 0 \right\} \tag{23}$$

exists for all $d \in \mathbb{R}^n$. Moreover, if ψ is semismooth at $\bar{x}$, the limit is equal to the directional derivative $\psi'(\bar{x}; d)$ as defined in (11); see, e.g. [43]. If, in addition to semismoothness of ψ at $\bar{x}$,

$$Vd - \psi'(\bar{x}; d) = O(\|d\|^2)$$

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 concept of a semismooth real-valued function stems from [37]. The extensions above were given in [43]. Finally, ψ is called SC^1 function, if it is continuously differentiable and $\nabla \psi$ is semismooth.

3. Necessary conditions for optimality

The standard approach to derive necessary optimality conditions for the LLVF reformulation (5) of the bilevel optimization problem (1) is to consider the partial penalization

$$\min_{x,y} F(x, y) + \lambda(f(x, y) - \varphi(x)) \quad \text{s.t.} \quad G(x, y) \leq 0, \quad g(x, y) \leq 0, \tag{24}$$

where $\lambda \in (0, \infty)$ denotes the penalization parameter. To deal with the fact that no standard constraint qualification holds for problem (5) [55], the authors of the latter paper introduced the partial calmness concept and showed its benefit for obtaining KKT conditions for (5). More in detail, problem (5) is said to be partially calm at one of its feasible points $(\bar{x}, \bar{y})$ if there exist $\lambda \in (0, \infty)$ and a neighbourhood U of $(\bar{x}, \bar{y}, 0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}$ such that

$$F(x, y) - F(\bar{x}, \bar{y}) + \lambda|\varsigma| \geq 0$$
$$\text{for all } (x, y, \varsigma) \in U \text{ with } G(x, y) \leq 0, g(x, y) \leq 0, f(x, y) - \varphi(x) + \varsigma = 0. \tag{25}$$

Based on this, the following result can be easily derived; cf. [55, Proposition 3.3].

Theorem 3.1: *Let $(\bar{x}, \bar{y})$ be locally optimal for problem* (5). *Then, the latter problem is partially calm at $(\bar{x}, \bar{y})$ if and only if there exists $\lambda \in (0, \infty)$ such that $(\bar{x}, \bar{y})$ is locally optimal for problem* (24).

Let $(\bar{x}, \bar{y})$ be a locally optimal solution of problem (5) at which the latter problem is partially calm. Then, the previous theorem not only ensures that (24) is locally optimal at $(\bar{x}, \bar{y})$ for some $\lambda_0 > 0$. Rather, one can easily check that this holds for any $\lambda \geq \lambda_0$.

Using Theorem 3.1, we are now going to establish the necessary optimality conditions that will be the basis of the Newton method in this paper. To proceed, we first need two

constraint qualifications. The upper-level Mangasarian-Fromovitz constraint qualification (UMFCQ) will be said to hold at $(\bar{x}, \bar{y})$ if there exists $d \in \mathbb{R}^{n+m}$ such that

$$\nabla G_i(\bar{x}, \bar{y})^\top d < 0 \text{ for all } i \in I^1 \quad \text{and} \quad \nabla g_j(\bar{x}, \bar{y})^\top d < 0 \text{ for all } j \in I^2. \tag{26}$$

The lower-level Mangasarian-Fromovitz constraint qualification (LMFCQ) is satisfied at $\bar{x}$ if, for any $z \in S(\bar{x})$, there exists $d \in \mathbb{R}^m$ such that

$$\nabla_2 g_j(\bar{x}, z)^\top d < 0 \quad \text{for all } j \text{ with } g_j(\bar{x}, z) = 0. \tag{27}$$

Further note that a real-valued function $(x, y) \mapsto \psi(x, y)$ will be said to be fully convex if it is convex w.r.t. to all variables. The lower-level problem is usually said to be convex if f and g_j ($j = 1, \ldots, q$) are convex just w.r.t. the lower-level variable. One can easily check that if a function $(x, y) \mapsto \psi(x, y)$ is fully convex, then it is convex w.r.t. x and it is convex w.r.t. y.

Theorem 3.2: *Let $(\bar{x}, \bar{y})$ be a local optimal solution of problem (5), assumed to be partially calm at $(\bar{x}, \bar{y})$. The functions f and $g_1, \ldots, g_q$ are assumed to be fully convex. Furthermore, suppose that LMFCQ holds at $\bar{x}$, and that UMFCQ holds at $(\bar{x}, \bar{y})$. Then, there exist $\lambda \in (0, \infty)$, $u \in \mathbb{R}^p$, $(v, w) \in \mathbb{R}^q \times \mathbb{R}^q$, and $z \in \mathbb{R}^m$ such that the following system holds for $(x, y) = (\bar{x}, \bar{y})$:*

$$\nabla_1 F(x, y) + \nabla_1 G(x, y)u + \nabla_1 g(x, y)v + \lambda \nabla_1 f(x, y) - \lambda \nabla_1 \ell(x, z, w) = 0, \tag{28}$$

$$\nabla_2 F(x, y) + \nabla_2 G(x, y)u + \nabla_2 g(x, y)v + \lambda \nabla_2 f(x, y) = 0, \tag{29}$$

$$\nabla_2 f(x, z) + \nabla_2 g(x, z)w = 0, \tag{30}$$

$$u \geq 0, \quad G(x, y) \leq 0, \quad u^\top G(x, y) = 0, \tag{31}$$

$$v \geq 0, \quad g(x, y) \leq 0, \quad v^\top g(x, y) = 0, \tag{32}$$

$$w \geq 0, \quad g(x, z) \leq 0, \quad w^\top g(x, z) = 0. \tag{33}$$

Proof: Note that similar proof techniques for related results can be found in [11,13,55]. Since the functions $f, g_1, \ldots, g_q$ involved in the lower-level problem are fully convex and sufficiently smooth, the optimal value function φ (4) is convex [52, Lemma 2.1] and, hence, locally Lipschitz continuous around $\bar{x}$ ($|\varphi(x)| < \infty$ for any x was assumed throughout). Therefore, for any $\lambda > 0$, problem (24) is a Lipschitz continuous optimization problem. Moreover, under the partial calmness condition, Theorem 3.1 guarantees that $\lambda > 0$ exists such that $(\bar{x}, \bar{y})$ is a local optimal solution of problem (24). Now, applying the necessary optimality conditions for locally Lipschitz optimization problems to (24) and taking into account that UMFCQ (26) holds at $(\bar{x}, \bar{y})$, we obtain the existence of $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^q$ such that (31), (32), and

$$0 \in \nabla F(\bar{x}, \bar{y}) + \nabla G(\bar{x}, \bar{y})u + \nabla g(\bar{x}, \bar{y})v + \lambda \nabla f(\bar{x}, \bar{y}) + \lambda \begin{pmatrix} \partial(-\varphi)(\bar{x}) \\ \{0\} \end{pmatrix} \tag{34}$$

hold for $(x, y) = (\bar{x}, \bar{y})$. It is clear that (29) for $(x, y) = (\bar{x}, \bar{y})$ follows from the last m components of this inclusion. Moreover, considering the first n components of (34), we

get

$$\nabla_1 F(\bar{x}, \bar{y}) + \nabla_1 G(\bar{x}, \bar{y})u + \nabla_1 g(\bar{x}, \bar{y})v + \lambda \nabla_1 f(\bar{x}, \bar{y}) \in \lambda \partial \varphi(\bar{x})$$

because of $\partial(-\varphi)(\bar{x}) = -\partial \varphi(\bar{x})$. Now, since the lower-level functions are fully convex and LMFCQ holds at $\bar{x}$, it follows from [52, Theorem 2.1] that we can find some $z \in S(\bar{x})$ and $w \in \Lambda(\bar{x}, z)$ such that (28) holds for $(x, y) = (\bar{x}, \bar{y})$. Furthermore, observe that the conditions (30) and (33) for $(x, y) = (\bar{x}, \bar{y})$ result from the definition of $w \in \Lambda(\bar{x}, z)$; cf. (13). ∎

Remark 3.1: In the literature, for example, see [11,13] and references therein, an upper-level regularity condition is often used in combination with LMFCQ to derive (31), (32), and (34). Here, we employ UMFCQ instead as problem (1) involves coupled upper-level constraints (i.e. depending on both the upper and lower-level variables), which is not the case in the aforementioned papers. Moreover, due to the convexity of the lower-level problem, the explicit use of inclusion $z \in S(\bar{x})$ is avoided, given that its fulfilment follows from the conditions (30) and (33).

Remark 3.2: If $y = z$ in the optimality conditions in Theorem 3.2, then we arrive at another well-known type of conditions consisting of (29)–(33) and

$$\nabla_1 F(x, y) + \nabla_1 G(x, y)u + \nabla_1 g(x, y)(v - \lambda w) + \lambda \nabla_1 f(x, y) = 0 \qquad (35)$$

with z replaced by y in (30) and (33). Instead of the full convexity assumption, the inner-semicontinuity concept can also allow one to derive these conditions. To see this, note that if the LMFCQ and inner semicontinuity both hold at $(\bar{x}, z)$, then the Clarke subdifferential of φ can be estimated as

$$\partial \varphi(\bar{x}) \subseteq \{ \nabla_1 \ell(\bar{x}, z, w) \mid w \in \Lambda(\bar{x}, z) \};$$

see [11,13] for related details and references. Note that various other stationarity concepts for the bilevel programs based on the LLVF reformulation are possible; see the latter references for related details. However, it is important to note for any of these conditions, the convexity assumption will still be required for the lower-level problem for inclusion $y \in S(x)$ to be ignored in these conditions.

For conditions ensuring that the assumptions in Theorem 3.2, in particular partial calmness, we refer to [11–13,55,56] and references therein. Keeping the penalty parameter $\lambda > 0$ fixed, the optimality conditions in Theorem 3.2 can be regarded as a mixed complementarity system. The latter will be reformulated as a square system of nonsmooth equations, so that a semismooth Newton method can be applied [4,17,31,41–43]. Dealing with other optimality conditions, like in [11,13,55], in a similar way, one is led to more difficult nonsmooth systems for which more sophisticated Newton-type methods [15,19,20] might be helpful.

4. The algorithm

To apply the semismooth Newton method from [4] to system (28)–(33), for some fixed $\lambda > 0$, the latter system is reformulated by means of the complementarity function $\phi : \mathbb{R}^2 \to \mathbb{R}$

[17] given by

$$\phi(a, b) := \sqrt{a^2 + b^2} - a - b.$$

It can be easily checked that $\phi(a, b) = 0$ if and only if $a \geq 0$, $b \geq 0$, $ab = 0$ is valid. Therefore, to rewrite the complementarity system (31)–(33), we first define functions $\phi^G : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^p$ and $\phi^g : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}^q$ by

$$\phi^G(x, y, u) := \begin{pmatrix} \phi(-G_1(x, y), u_1) \\ \vdots \\ \phi(-G_p(x, y), u_p) \end{pmatrix} \quad \text{and} \quad \phi^g(x, y, v) := \begin{pmatrix} \phi(-g_1(x, y), v_1) \\ \vdots \\ \phi(-g_q(x, y), v_q) \end{pmatrix},$$

respectively. Furthermore, we introduce the Lagrange-type functions $\mathcal{L}^\lambda : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ and $L^\lambda : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$ be respectively defined by

$$\mathcal{L}^\lambda(x, y, u, v) := F(x, y) + u^\top G(x, y) + v^\top g(x, y) + \lambda f(x, y), \tag{36}$$

$$L^\lambda(x, y, z, u, v, w) := \mathcal{L}^\lambda(x, y, u, v) - \lambda \ell(x, z, w), \tag{37}$$

where $\lambda > 0$ is the fixed penalty parameter. Based on these definitions, we now introduce the mapping $\Phi^\lambda : \mathbb{R}^N \to \mathbb{R}^N$ with $N := n + 2m + p + 2q$ by

$$\Phi^\lambda(\zeta) := \begin{bmatrix} \nabla L^\lambda(\zeta) \\ \phi^G(x, y, u) \\ \phi^g(x, y, v) \\ \phi^g(x, z, w) \end{bmatrix} \quad \text{with } \zeta := (x, y, z, u, v, w), \tag{38}$$

where ∇L^λ denotes the gradient of L^λ w.r.t. (x, y, z). Now, keeping (37) and (38) in mind, it can be easily seen that the optimality conditions (28)–(33) in Theorem 3.2 can equivalently be written as

$$\Phi^\lambda(\zeta) = 0. \tag{39}$$

Obviously, this is a square system of N equations and N variables. Moreover, the mapping Φ^λ is strongly semismooth at any solution of (39). In particular, we can apply the semismooth Newton method in [4] with its favorable combination of global and local convergence properties. The latter means superlinear or quadratic convergence based on (strong) semismoothness of Φ^λ and a regularity property of the generalized Jacobians $\partial_B \Phi^\lambda$ or $\partial \Phi^\lambda$ at a solution ζ^* of (39). Sufficient conditions for $\partial \Phi^\lambda(\zeta^*)$ containing only nonsingular matrices will be developed in Section 5. For global convergence, the complementarity function ϕ has the property that ϕ^2 is differentiable with Lipschitz continuous derivative [24]. Due to this, the merit function $\Psi^\lambda : \mathbb{R}^N \to \mathbb{R}$ with

$$\Psi^\lambda(\zeta) := \frac{1}{2} \|\Phi^\lambda(\zeta)\|^2 \tag{40}$$

is continuously differentiable. In particular, this enables to overcome situations where the Newton direction for (39) does not exist or its descent is insufficient. For other complementarity and merit functions as well as their properties we refer to [18,50], for example.

We now present the semismooth Newton algorithm from [4] applied to equation (39) or, in other words, to the optimality conditions from Theorem 3.2.

Algorithm 1 Semismooth Newton algorithm

Step 0: Choose $\lambda > 0$, $\beta > 0$, $\epsilon \geq 0$, $t > 2$, $\rho \in (0,1)$, $\sigma \in \left(0, \frac{1}{2}\right)$, $\zeta^o :=$ $(x^o, y^o, z^o, u^o, v^o, w^o)$.

Set $k := 0$.

Step 1: If $\|\Phi^\lambda(\zeta^k)\| \leq \epsilon$, then stop.

Step 2: Choose $W^k \in \partial_B \Phi^\lambda(\zeta^k)$ and compute a solution d^k of the system

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

If this equation has no solution or if the condition

$$\nabla \Psi^\lambda(\zeta^k)^\top d^k \leq -\beta \|d^k\|^t$$

is violated, then set

$$d^k := -\nabla \Psi^\lambda(\zeta^k).$$

Step 3: Find the smallest integer $s_k \in \{0, 1, 2, 3, \ldots\}$ such that

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

Step 4: Set $\alpha_k := \rho^{s_k}$, $\zeta^{k+1} := \zeta^k + \alpha_k d^k$, $k := k + 1$, and go to Step 1.

Some remarks on Algorithm 1 are in order. Essentially, it is the linesearch semismooth Newton method from [4]. Compared to the latter, we are now dealing with a particular mixed complementarity problem (MCP), see [39] for an extension of [4] to MCPs. The penalization parameter $\lambda > 0$ has to be chosen in Step 0 and is fixed throughout the algorithm. For the definition of the generalized Jacobian $\partial_B \Phi^\lambda(\zeta)$, see Section 2. If all matrices in $\partial_B \Phi^\lambda(\zeta)$ are nonsingular, the function Φ^λ is called BD-regular at ζ. Following [4], global and local convergence properties of Algorithm 1 can be derived as follows.

Theorem 4.1: *Assume that $\bar{\zeta}$ is an accumulation point of a sequence $\{\zeta^k\}$ generated by Algorithm 1 for some $\lambda > 0$. Then, $\bar{\zeta}$ is a stationary point of Ψ^λ, i.e. $\nabla \Psi^\lambda(\bar{\zeta}) = 0$. If $\bar{\zeta}$ solves $\Phi^\lambda(\zeta) = 0$ and Φ^λ is BD-regular at $\bar{\zeta}$, then $\{\zeta^k\}$ converges to $\bar{\zeta}$ superlinearly and, if the functions F, G, f, and g defining problem (1) have locally Lipschitz continuous second order derivatives, the convergence is Q-quadratic.*

The proof of [4, Theorem 11] can be easily extended to the complementarity system $\Phi^\lambda(\zeta) = 0$. Moreover, it is known that the continuity of the second-order derivatives of the problem functions F, G, f, and g (as assumed in Section 1) suffices instead of the semismoothness of these derivatives (as used in [4] for showing that eventually the unit stepsize $\alpha_k = 1$ is accepted).

The assumption that Φ^λ is BD-regular at $\bar{\zeta}$ used in Theorem 4.1 can be replaced by the stronger CD-regularity of Φ^λ at $\bar{\zeta}$. The latter requires the non-singularity of all matrices

in $\partial\Phi^\lambda(\bar\zeta)$. In the next section, we focus on the derivation of sufficient conditions for CD-regularity. The latter also allows a nice connection to the Robinson condition that we are going to discuss in Section 6. Let us finally note that conditions ensuring that a stationary point of a merit function is a solution of the underlying equation were extensively studied for complementarity problems. We do not want to dive into this subject but note that if just one element of $\partial\Phi^\lambda(\bar\zeta)$ is nonsingular then $\nabla\Psi^\lambda(\bar\zeta) = 0$ implies $\Phi^\lambda(\bar\zeta) = 0$, for example see [14, Section 4].

5. CD-regularity

To derive sufficient conditions guaranteeing that CD-regularity holds for Φ^λ (38), we first provide an upper estimate of the generalized Jacobian of Φ^λ in the sense of Clarke (22). Recall that the index sets η^i, ν^i and θ^i with $i = 1, 2, 3$ are defined in (8) and (9).

Theorem 5.1: *Let the functions F, G, f, and g be twice continuously differentiable at $\zeta :=$ (x, y, z, u, v, w). If $\lambda > 0$, then Φ^λ is semismooth at ζ and any matrix $W^\lambda \in \partial\Phi^\lambda(\zeta)$ must take the form*

$$
W^\lambda = \begin{bmatrix}
A_{11}^\lambda & \left(A_{21}^\lambda\right)^\top & -\lambda A_{31}^\top & B_{11}^\top & B_{21}^\top & -\lambda B_{31}^\top \\
A_{21}^\lambda & A_{22}^\lambda & O & B_{12}^\top & B_{22}^\top & O \\
-\lambda A_{31} & O & -\lambda A_{33} & O & O & -\lambda B_{33}^\top \\
\Lambda_1 B_{11} & \Lambda_1 B_{12} & O & \Gamma_1 & O & O \\
\Lambda_2 B_{21} & \Lambda_2 B_{22} & O & O & \Gamma_2 & O \\
\Lambda_3 B_{31} & O & \Lambda_3 B_{33} & O & O & \Gamma_3
\end{bmatrix}
\tag{41}
$$

where the matrices A_{ij} and B_{ij} are respectively defined by

$$
A_{11}^\lambda := \nabla_1^2 \mathcal{L}^\lambda(\zeta) - \lambda \nabla_1^2 \ell(\zeta), \quad A_{21}^\lambda := \nabla_{12}^2 \mathcal{L}^\lambda(\zeta), \quad A_{22}^\lambda := \nabla_2^2 \mathcal{L}^\lambda(\zeta),
$$
$$
A_{31} := \nabla_{12}^2 \ell(\zeta), \quad A_{33} := \nabla_2^2 \ell(\zeta),
$$
$$
B_{11} := \nabla_1 G(x, y), \quad B_{21} := \nabla_1 g(x, y), \quad B_{31} := \nabla_1 g(x, z),
\tag{42}
$$
$$
B_{12} := \nabla_2 G(x, y), \quad B_{22} := \nabla_2 g(x, y), \quad B_{33} := \nabla_2 g(x, z),
$$

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

$$
(a_j^i, b_j^i) \begin{cases}
= (0, -1) & \text{if } j \in \eta^i, \\
= (1, 0) & \text{if } j \in \nu^i, \\
\in \{(\alpha, \beta) : (\alpha - 1)^2 + (\beta + 1)^2 \le 1\} & \text{if } j \in \theta^i.
\end{cases}
\tag{43}
$$

In the next result, we provide conditions ensuring that the function Φ^λ is CD-regular. To proceed, first note that, similarly to the UMFCQ (26) and analogously to the LLICQ (12), the upper-level linear independence constraint qualification (ULICQ) will be said to hold at $(\bar x, \bar y)$ if the following family of vectors is linear independent:

$$
\left\{\nabla G_i(\bar x, \bar y) : i \in I^1\right\} \cup \left\{\nabla g_j(\bar x, \bar y) : j \in I^2\right\}.
\tag{44}
$$

Furthermore, let us introduce the cone of feasible directions for problem (24)

$$
Q(\bar x, \bar y, \bar z) := \Big\{ d \mid \nabla G_i(\bar x, \bar y)^\top d^{1,2} = 0, \ i \in \nu^1,
$$
$$
\nabla g_j(\bar x, \bar y)^\top d^{1,2} = 0, \ j \in \nu^2, \ \nabla g_j(\bar x, \bar z)^\top d^{1,3} = 0, \ j \in \nu^3 \Big\},
\tag{45}
$$

where $d := (d_1^1, \ldots, d_n^1, d_1^2, \ldots, d_m^2, d_1^3, \ldots, d_m^3)^\top$, $d^{1,2} := (d_1^1, \ldots, d_n^1, d_1^2, \ldots, d_m^2)^\top$, and $d^{1,3}$ defined similarly. Recall that for $i = 1, 2, 3$, v^i is defined as in (8)–(9). 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. (x, y) and (x, z), respectively.

Theorem 5.2: *Let $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ satisfy the conditions (28)–(33) for some $\lambda > 0$. Suppose that ULICQ (44) and LLICQ (12) hold at $(\bar{x}, \bar{y})$ and $(\bar{x}, \bar{z})$, respectively. If additionally,*

$$(d^{1,2})^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^{1,2} > \lambda (d^{1,3})^\top \nabla^2 \ell(\bar{\zeta}) d^{1,3} \tag{46}$$

for all $d \in Q(\bar{x}, \bar{y}, \bar{z}) \setminus \{0\}$ and LSCC (17) is also satisfied at $(\bar{x}, \bar{z}, \bar{w})$, 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 5.1, cf. (41)–(43). Hence, it follows that for any $d := (d^1, d^2, d^3, d^4, d^5, d^6)$ with $d^1 \in \mathbb{R}^n, d^2 \in \mathbb{R}^m, d^3 \in \mathbb{R}^m, d^4 \in \mathbb{R}^p, d^5 \in \mathbb{R}^q$ and $d^6 \in \mathbb{R}^q$ such that $W^\lambda d = 0$, we have

$$\nabla_1^2 \mathcal{L}^\lambda(\bar{\zeta}) d^1 - \lambda \nabla_1^2 \ell(\bar{\zeta}) d^1 + \nabla_{21}^2 \mathcal{L}^\lambda(\bar{\zeta}) d^2 - \lambda \nabla_{21}^2 \ell(\bar{\zeta}) d^3$$
$$+ \nabla_1 G(\bar{x}, \bar{y})^\top d^4 + \nabla_1 g(\bar{x}, \bar{y})^\top d^5 - \lambda \nabla_1 g(\bar{x}, \bar{z})^\top d^6 = 0, \tag{47}$$

$$\nabla_{12}^2 \mathcal{L}^\lambda(\bar{\zeta}) d^1 + \nabla_2^2 \mathcal{L}^\lambda(\bar{\zeta}) d^2 + \nabla_2 G(\bar{x}, \bar{y})^\top d^4 + \nabla_2 g(\bar{x}, \bar{y})^\top d^5 = 0, \tag{48}$$

$$-\lambda \nabla_{12}^2 \ell(\bar{\zeta}) d^1 - \lambda \nabla_2^2 \ell(\bar{\zeta}) d^3 - \lambda \nabla_2 g(\bar{x}, \bar{z})^\top d^6 = 0, \tag{49}$$

$$\forall j = 1, \ldots, p, \quad a_j^1 \nabla G_j(\bar{x}, \bar{y})^\top d^{1,2} + b_j^1 d_j^4 = 0, \tag{50}$$

$$\forall j = 1, \ldots, q, \quad a_j^2 \nabla g_j(\bar{x}, \bar{y})^\top d^{1,2} + b_j^2 d_j^5 = 0, \tag{51}$$

$$\forall j = 1, \ldots, q, \quad a_j^3 \nabla g_j(\bar{x}, \bar{z})^\top d^{1,3} + b_j^3 d_j^6 = 0. \tag{52}$$

Recall that p and q represent the number of components of upper- (resp. lower-) constraint functions (1)–(3). For $i = 1, 2, 3$, let $p_1 := p, p_2 := q$ (when g applied to $(\bar{x}, \bar{y})$), and $p_3 := q$ (when g applied to $(\bar{x}, \bar{z})$). Then define P_1^i as the set of indices $j = 1 \ldots, p_i$ such that $a_j^i > 0$ and $b_j^i < 0$; P_2^i as the set of indices $j = 1 \ldots, p_i$ such that $a_j^i = 0$ and $b_j^i = -1$; and P_3^i as the set of indices $j = 1 \ldots, p_i$ such that $a_j^i = 1$ and $b_j^i = 0$. It follows from (50)–(52) that for $j \in P_2^1, j \in P_2^2$, and $j \in P_2^3$,

$$d_j^4 = 0, \quad d_j^5 = 0, \quad \text{and} \quad d_j^6 = 0, \tag{53}$$

respectively. As for $j \in P_3^1, j \in P_3^2$, and $j \in P_3^3$, we respectively get

$$\nabla G_j(\bar{x}, \bar{y})^\top d^{1,2} = 0, \quad \nabla g_j(\bar{x}, \bar{y})^\top d^{1,2} = 0, \quad \text{and} \quad \nabla g_j(\bar{x}, \bar{z})^\top d^{1,3} = 0. \tag{54}$$

Now observe that under the LSCC (17), $\theta^3 = \emptyset$. Hence, from the corresponding counterpart of (43), it follows that $P_1^3 := \emptyset$. We can further check that for $j \in P_1^1$ and $j \in P_1^2$ we

respectively have

$$\nabla G_j(\bar{x}, \bar{y})^\top d^{1,2} = c_j^1 d_j^4 \quad \text{and} \quad \nabla g_j(\bar{x}, \bar{y})^\top d^{1,2} = c_j^2 d_j^5, \tag{55}$$

where $c_j^1 := -\frac{b_j^1}{a_j^1}$ and $c_j^2 := -\frac{b_j^2}{a_j^2}$, respectively. By respectively multiplying (47), (48), and (49) from the left by $(d^1)^\top$, $(d^2)^\top$, and $(d^3)^\top$, and adding the resulting sums together,

$$(d^{1,2})^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^{1,2} - \lambda (d^{1,3})^\top \nabla^2 \ell(\bar{\zeta}) d^{1,3}$$
$$+ (d^4)^\top \nabla G(\bar{x}, \bar{y}) d^{1,2} + (d^5)^\top \nabla g(\bar{x}, \bar{y}) d^{1,2} - \lambda (d^6)^\top \nabla g(\bar{x}, \bar{z}) d^{1,3} = 0. \tag{56}$$

Considering the strict complementarity slackness at $(\bar{x}, \bar{z}, \bar{w})$ again, it follows that

$$(d^6)^\top \nabla g(\bar{x}, \bar{z}) d^{1,3} = \sum_{j \in P_2^3} d_j^6 \nabla g_j(\bar{x}, \bar{z})^\top d^{1,3} + \sum_{j \in P_3^3} d_j^6 \nabla g_j(\bar{x}, \bar{z})^\top d^{1,3} = 0 \tag{57}$$

given that $d_j^6 = 0$ for $j \in P_2^3$ and $\nabla g_j(\bar{x}, \bar{z})^\top d^{1,3} = 0$ for $j \in P_3^3$. Inserting (57) into (56) while taking into account (53)–(55), we get

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

Since by definition, $c_j^1 > 0$ for $j \in P_1^1$ and $c_j^2 > 0$ for $j \in P_1^2$, it follows from condition (46) that $d^1 = 0$, $d^2 = 0$, $d^3 = 0$, $d_j^4 = 0$ for $j \in P_1^1$ and $d_j^5 = 0$ for $j \in P_1^2$, while taking into account (54) and the fact that $\nu^i \subseteq P_3^i$ for $i = 1, 2, 3$. Inserting these values in (47)–(49) and considering (53),

$$\sum_{j \in P_3^1} d_j^4 \nabla_1 G_j(\bar{x}, \bar{y}) + \sum_{j \in P_3^2} d_j^5 \nabla_1 g_j(\bar{x}, \bar{y}) + \sum_{j \in P_3^3} (-\lambda d_j^6) \nabla_1 g_j(\bar{x}, \bar{z}) = 0, \tag{58}$$

$$\sum_{j \in P_3^1} d_j^4 \nabla_2 G_j(\bar{x}, \bar{y}) + \sum_{j \in P_3^2} d_j^5 \nabla_2 g_j(\bar{x}, \bar{y}) = 0, \tag{59}$$

$$\sum_{j \in P_3^3} d_j^6 \nabla_2 g_j(\bar{x}, \bar{z}) = 0. \tag{60}$$

Since the LLICQ (12) is satisfied at $(\bar{x}, \bar{z})$ and $P_3^3 \subseteq I^3$ holds, it follows from (60) that $d_j^6 = 0$ for $j \in P_3^3$. Inserting these values in (58) and combining the resulting equation with (59),

$$\sum_{j \in P_3^1} d_j^4 \nabla G_j(\bar{x}, \bar{y}) + \sum_{j \in P_3^2} d_j^5 \nabla g_j(\bar{x}, \bar{y}) = 0. \tag{61}$$

Considering the fulfilment of the ULICQ (44) at $(\bar{x}, \bar{y})$, and taking into account that $P_3^i \subseteq I^i$ for $i = 1, 2$, we can deduce from (61) that $d_j^4 = 0$ for $j \in P_3^1$ and $d_j^5 = 0$ for $j \in P_3^2$. This concludes the proof as we have shown that all the components of d are zero. ∎

Note that the strict complementarity condition imposed here is restricted to the lower-level problem and does not necessarily imply the local differentiability of the lower-level

optimal solution map as known in earlier results on the Newton method, see, e.g. [29,30] or in the literature on nonlinear parametric optimization, see, e.g. [16]. As we also have the LLICQ, the differentiability of lower-level optimal solution is usually guaranteed when a strong second order sufficient condition (SOSSC)-type condition restricted to the lower-level problem is satisfied as well. The LSCC here only allows us to deal with the minus sign appearing on the lower-value function in problem (5) and is responsible for many of the stationarity concepts for the problem; cf. [11,13,55]. Next we discuss two possible scenarios to avoid imposing the LSCC. In the first case, we assume that the lower-level feasible set is unperturbed.

Theorem 5.3: *Let g in problem* (1) *be independent of the upper-level variable x and suppose that the point* $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ *satisfies conditions* (28)–(33) *for some* $\lambda > 0$*. Furthermore, assume that the family* $\{\nabla_x G_j(\bar{x}, \bar{y}) : j \in I^1\}$ *is linearly independent and the LLICQ holds at* $\bar{y}$ *and* $\bar{z}$*. If additionally,*

$$(d^{1,2})^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^{1,2} > \lambda (d^{1,3})^\top \nabla^2 \ell^*(\bar{\zeta}) d^{1,3} \qquad (62)$$

for all $d \in Q(\bar{\zeta}) \setminus \{0\}$*, where*

$$\nabla^2 \ell^*(\bar{\zeta}) := \begin{bmatrix} \nabla_1^2 \ell(\bar{\zeta}) & \nabla_{21}^2 \ell(\bar{\zeta}) \\ -\nabla_{12}^2 \ell(\bar{\zeta}) & -\nabla_2^2 \ell(\bar{\zeta}) \end{bmatrix}, \qquad (63)$$

then the function Φ^λ *is CD-regular at the point* $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$*.*

Proof: Considering the counterpart of (47)–(52) when g is independent of x and proceeding as in the proof of the previous theorem, we have (53), (54) and

$$\nabla G_j(\bar{x}, \bar{y}) d^{1,2} = c_j^1 d_j^4, \quad \nabla g_j(\bar{y}) d^2 = c_j^2 d_j^5, \quad \text{and} \quad \nabla g_j(\bar{z}) d^3 = c_j^3 d_j^6 \qquad (64)$$

for $j \in P_1^1, j \in P_1^2$, and $j \in P_1^3$, respectively. Here, c_j^1 for $j \in P_1^1$ and c_j^2 for $j \in P_1^2$ are defined as in (54) while $c_j^3 := -\frac{b_j^3}{a_j^3}$ for $j \in P_1^3$, cf. (43). Next, replacing the counterpart of (49) with

$$\nabla_{12}^2 \ell(\bar{\zeta}) d^1 + \nabla_2^2 \ell(\bar{\zeta}) d^3 + \nabla g(\bar{z})^\top d^6 = 0$$

and multiplying this equality, (48), and (47) from the left by $(d^3)^\top$, $(d^2)^\top$, and $(d^1)^\top$, respectively, and adding the resulting sums together, we obtain

$$(d^{1,2})^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^{1,2} - \lambda (d^{1,3})^\top \nabla^2 \ell^*(\bar{\zeta}) d^{1,3}$$

$$+ \sum_{j \in P_1^1} c_j^1 (d_j^4)^2 + \sum_{j \in P_1^2} c_j^2 (d_j^5)^2 + \sum_{j \in P_1^3} c_j^3 (d_j^6)^2 = 0,$$

while taking into account (64) and the counterparts of (53) and (54). Hence, it follows from assumption (62) that $d^1 = 0$, $d^2 = 0$, $d^3 = 0$, $d_j^4 = 0$ for $j \in P_1^1$, $d_j^5 = 0$ for $j \in P_1^2$, and $d_j^6 = 0$ for $j \in P_1^3$. The rest of the proof then follows as that of Theorem 5.2. ∎

For the next result, the LSCC is also not needed, and g does not necessarily have to be independent of the upper-level variable.

Theorem 5.4: *Let $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ satisfy the optimality conditions* (28)–(33) *for some* $\lambda > 0$. *Suppose that the ULICQ and LLICQ hold at* $(\bar{x}, \bar{y})$ *and* $(\bar{x}, \bar{z})$, *respectively. Then* Φ^λ *is CD-regular at* $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ *provided we also have*

$$(d^{1,2})^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^{1,2} > \lambda \left\{ (d^{1,3})^\top \nabla^2 \ell(\bar{\zeta}) d^{1,3} + \sum_{j \in P_1^3} c_j^3 (e_j)^2 \right\} \qquad (65)$$

for all $(d, e) \in [Q(\zeta) \times \mathbb{R}^{|P_1^3|}] \setminus \{0\}$, *with* $c_j^3 := -\frac{b_j^3}{a_j^3}$ *for* $j \in P_1^3$; *cf.* (43).

Proof: Also proceeding as in the proof of Theorem 5.2 while replacing (55) with

$$\nabla G_j(\bar{x}, \bar{y}) d^{1,2} = c_j^1 d_j^4, \quad \nabla g_j(\bar{x}, \bar{y}) d^{1,2} = c_j^2 d_j^5, \quad \text{and} \quad \nabla g_j(\bar{x}, \bar{z}) d^{1,2} = c_j^3 d_j^6 \qquad (66)$$

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

$$(d^{1,2})^\top \nabla^2 \mathcal{L}^\lambda(\zeta) d^{1,2} - \lambda \left\{ (d^{1,3})^\top \nabla^2 \ell(\zeta) d^{1,3} + \sum_{j \in P_1^3} c_j^3 (d_j^6)^2 \right\}$$

$$+ \sum_{j \in P_1^1} c_j^1 (d_j^4)^2 + \sum_{j \in P_1^2} c_j^2 (d_j^5)^2 = 0$$

by inserting (53)–(54) and (66) in the counterpart of (56), as θ^3 is not necessarily empty. Hence, under assumption (65), we get $d^1 = 0, d^2 = 0, d^3 = 0, d_j^4 = 0$ for $j \in P_1^1, d_j^5 = 0$ for $j \in P_1^2$ and $d_j^6 = 0$ for $j \in P_1^3$. Similarly, the rest of the proof then follows as for Theorem 5.2. ∎

Considering the structure of the generalized second order subdifferential of φ (4) (see [57]), condition (65) can be seen as the most natural extension to our problem (24) of the strong second order sufficient condition used for example in [17,42]. To see this, note that condition (65) can be replaced by the following condition, for all $(d^{1,2,3}, e)$ in $[Q(\zeta) \times \mathbb{R}^q] \setminus \{0\}$:

$$(d^{1,2})^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^{1,2} > \lambda \left\{ (d^{1,3})^\top \nabla^2 \ell(\bar{\zeta}) d^{1,3} + e^\top \nabla g(\bar{x}, \bar{z}) d^{1,3} \right\}.$$

Example 5.5: Consider the bilevel optimization problem

$$\min_{x,y} \quad x^2 + y_1^2 + y_2^2$$

$$\text{s.t.} \quad y \in S(x) := \arg\min_y \left\{ \|y - (x, -1)^\top\|^2 : y_1 - y_2 \leq 0, \ -y_1 - y_2 \leq 0 \right\}, \qquad (67)$$

where the lower-level problem is taken from [16, Chapter 1]. The LLVF (4) can be obtained as

$$\varphi(x) = \begin{cases} \frac{1}{2}(1-x)^2 & \text{if } x < -1, \\ 1 + x^2 & \text{if } -1 \leq x \leq 1, \\ \frac{1}{2}(1+x)^2 & \text{if } x > 1. \end{cases}$$

The optimal solution of problem (67) is $(\bar{x}, \bar{y})$ with $\bar{x} = 0$ and $\bar{y} = (0, 0)$. Considering the expression of φ above, one can easily check that $(0, 0, 0, 0)$ satisfies the conditions

$$y_1 - y_2 \le 0, \quad -y_1 - y_2 \le 0, \quad \|y - (x, -1)^\top\|^2 - \varphi(x) + \varsigma = 0.$$

Thus, for problem (67), condition (25) holds with $(\bar{x}, \bar{y}) = (0, 0, 0)$ and $U = \mathbb{R}^4$. Next, note that the LMFCQ (27) holds at any lower-level feasible point. The lower-level optimal solution mapping S is single valued and continuous in this case; hence inner-semicontinuous [13]. Therefore, $(\bar{x}, \bar{y})$ satisfies (28)–(33) and subsequently, the corresponding calculations show that the point $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{v}, \bar{w})$, where $\bar{x} = 0$, $\bar{y} = (0, 0)$, $\bar{z} := (0, 0)$, $\bar{v} := (\lambda, \lambda)$, and $\bar{w} = (1, 1)$ with $\lambda > 0$, solves (38). Furthermore, $Q(\bar{x}, \bar{y}, \bar{z}) = \mathbb{R} \times \{(0, 0, 0, 0)\}$ and for all $d^{1,2,3} \in Q(\bar{x}, \bar{y}, \bar{z}) \setminus \{0\}$,

$$(d^{1,2})^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) d^{1,2} - \lambda (d^{1,3})^\top \nabla^2 \ell(\bar{\zeta}) d^{1,3} = 2(d^1)^2 > 0.$$

Hence, for problem (67), Φ^λ (38) is CD-regular at $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{v}, \bar{w})$, for any value of $\lambda > 0$.

6. Robinson-type condition

For a standard nonlinear optimization problem with twice continuously differentiable functions, the Robinson condition [45] is said to hold at one of its KKT points if LICQ and a strong second order sufficient condition (SSOSC) are satisfied. It was shown in [42] that if a standard nonlinear optimization problem is SC^1 and satisfies the Robinson condition, then the CD-regularity condition holds for the corresponding counterpart of function Φ^λ (38). Considering the structure of the results from the previous section, it can be argued that the combination of assumptions in Theorems 5.2, 5.3 or 5.4 corresponds to an extension of Robinson's condition to the context of bilevel optimization. However, another implication of Robinson's condition, i.e. precisely of the SSOSC, is that it ensures that a given point is a strict local optimal solution for the corresponding nonlinear programming problem. Further analysis on sufficient conditions for strong regularity in the context of mixed complementarity problems can be found in [27].

The aim of this section is to enhance the second order assumption in the previous section so that it can guaranty that points computed by our algorithm are strict local optimal points. To proceed, we introduce the following cone of feasible directions

$$C^\lambda(\bar{x}, \bar{y}) := \{d \mid \nabla G_i(\bar{x}, \bar{y})^\top d = 0 \text{ for } i \in \nu^1, \ \nabla G_i(\bar{x}, \bar{y})^\top d \le 0 \text{ for } i \in \theta^1,$$
$$\nabla g_j(\bar{x}, \bar{y})^\top d = 0 \text{ for } j \in \nu^2, \ \nabla g_j(\bar{x}, \bar{y})^\top d \le 0 \text{ for } j \in \theta^2,$$
$$\nabla F(\bar{x}, \bar{y})^\top d + \lambda \nabla f(\bar{x}, \bar{y})^\top d - \lambda \nabla_1 \ell(\bar{x}, z, w)^\top d^1 \le 0 \text{ for } z \in S(\bar{x})\}, \qquad (68)$$

where $\{w\} := \{w(z)\} = \Lambda(\bar{x}, z)$ for a fixed $z \in S(\bar{x})$, as the LLICQ (12) will be assumed to hold at $(\bar{x}, z)$ for all $z \in S(\bar{x})$. Also note that as in (45), d can be written as $d := ((d^1)^\top, (d^2)^\top)^\top$. Furthermore, we will use the following modified version of the upper-level Lagrangian function (36)

$$\bar{\mathcal{L}}^\lambda_\kappa(x, y, u, v) := \kappa \left(F(x, y) + \lambda f(x, y) \right) + \sum_{i \in I^1(d)} u_i G_i(x, y) + \sum_{j \in I^2(d)} v_j g_j(x, y),$$

where the set $I^1(d)$ (resp. $I^2(d)$) represents the set of indices $i \in I^1$ (resp. $j \in I^2$) such that we have $\nabla G_i(\bar{x}, \bar{y})^\top d = 0$ (resp. $\nabla g_j(\bar{x}, \bar{y})^\top d = 0$). Recall that I^1 and I^2 are given in (6) and (7). In the next result, we first provide slightly general SSOSC-type condition for problem (24).

Theorem 6.1: *Let the point $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ satisfy the conditions (28)–(33) for some $\lambda > 0$. Suppose that the lower-level problem is convex at $\bar{x}$ (i.e. $f(\bar{x}, .)$ and $g_i(\bar{x}, .), i = 1, \ldots, q$, are convex) and the assumptions in Theorem 2.2 hold for all $d \in C^\lambda(\bar{x}, \bar{y})$. Then, $(\bar{x}, \bar{y})$ is a strict local optimal solution of problem (24) provided that, for all $d \in C^\lambda(\bar{x}, \bar{y}) \setminus \{0\}$, the condition*

$$d^\top \nabla^2 \bar{\mathcal{L}}_{\kappa_\circ}^\lambda(\bar{x}, \bar{y}, u, v)d > \lambda \sum_{t=1}^k \kappa_t \xi_{d^1}(\bar{x}, z^t) \tag{69}$$

is satisfied for some vectors u, v, z^t, and κ_t, with $z^t \in S_1(\bar{x}; d^1)$ and $\Lambda(\bar{x}, z^t) = \{w^t\}$, for $t = 1, \ldots, k$, where k is some natural number and $\kappa_\circ := \sum_{i=1}^k \kappa_t$, such that we have

$$\nabla_1 \bar{\mathcal{L}}_{\kappa_\circ}^\lambda(\bar{x}, \bar{y}, u, v) - \lambda \sum_{t=1}^k \kappa_t \nabla_1 \ell(\bar{x}, z^t, w^t) = 0, \quad \nabla_2 \bar{\mathcal{L}}_{\kappa_\circ}^\lambda(\bar{x}, \bar{y}, u, v) = 0 \tag{70}$$

and

$$\kappa_\circ + \sum_{i \in I^1(d)} u_i + \sum_{j \in I^2(d)} v_j = 1, \quad \kappa_\circ \geq 0, \quad u_i \geq 0 \text{ for } i \in I^1(d), \ v_j \geq 0 \text{ for } j \in I^2(d). \tag{71}$$

Proof: First, consider the optimization problem in (24) for the parameter $\lambda >$ for which $\bar{\zeta}$ satisfies the conditions (28)–(33). This problem can obviously be rewritten as

$$\min F(x, y) + \lambda \left(f(x, y) - \varphi(x) \right) \quad \text{s.t.} \quad \psi(x, y) \leq 0,$$

where the function ψ is defined by $\psi(x, y) := \max\{G_1(x, y), \ldots, G_p(x, y), g_1(x, y), \ldots, g_q(x, y)\}$. Next, consider the unconstrained optimization problem

$$\min \phi_\lambda(x, y) := \max \left\{ \psi_\lambda(x, y), \ \psi(x, y) \right\} \tag{72}$$

with $\psi_\lambda(x, y) := F(x, y) - F(\bar{x}, \bar{y}) + \lambda(f(x, y) - f(\bar{x}, \bar{y})) - \lambda(\varphi(x) - \varphi(\bar{x}))$. Based on [2, Chapter 3], it suffices to show that the function ϕ_λ satisfies the following three conditions:

[a] ϕ_λ is directionally differentiable and we have

$$\phi_\lambda'(\bar{x}, \bar{y}; d) \geq 0 \quad \text{for all } d \in \mathbb{R}^{n+m}; \tag{73}$$

[b] ϕ_λ is twice directionally differentiable (16) and fulfils the condition

$$\inf_{e \in \mathbb{R}^{n+m}} \phi_\lambda''(\bar{x}, \bar{y}; d, e) > 0 \quad \text{for all } d \neq 0 \quad \text{s.t.} \quad \phi_\lambda'(\bar{x}, \bar{y}; d) = 0; \tag{74}$$

[c] ϕ_λ satisfies the second order epiregularity condition, i.e. for any $d \in \mathbb{R}^{n+m}, t \geq 0$, and any function (path) e from $\mathbb{R}_+$ to $\mathbb{R}^n \times \mathbb{R}^m$ such that $te(t) \to 0$ as $t \downarrow 0$,

$$\phi_\lambda \left((\bar{x}, \bar{y}) + td + \frac{1}{2}t^2 e(t) \right) \geq \phi_\lambda(\bar{x}, \bar{y}) + t\phi'_\lambda(\bar{x}, \bar{y}; d)$$

$$+ \frac{1}{2}t^2 \phi''_\lambda \left(\bar{x}, \bar{y}; d, e(t) \right) + \circ(t^2). \tag{75}$$

To prove condition [a], first note that ψ is directionally differentiable, as the upper- and lower-level constraint functions are continuously differentiable. In fact,

$$\psi'(\bar{x}, \bar{y}; d) = \max \left\{ \nabla G_i(\bar{x}, \bar{y})^\top d \text{ for } i \in I^1, \ \nabla g_j(\bar{x}, \bar{y})^\top d \text{ for } j \in I^2 \right\} \tag{76}$$

for any $d \in \mathbb{R}^{n+m}$. As for ψ_λ, recalling that $\{w\} := \{w(z)\} = \Lambda(\bar{x}, z)$ for $z \in S(\bar{x})$, thanks to the fulfilment of the LLICQ (12) at $(\bar{x}, z)$ for all $z \in S(\bar{x})$, it follows from Theorem 2.1 that

$$\psi'_\lambda(\bar{x}, \bar{y}; d) = \max \left\{ \nabla(F + \lambda f)(\bar{x}, \bar{y})^\top d - \lambda \nabla_x \ell(\bar{x}, z, w)^\top d^1, \ z \in S(\bar{x}) \right\} \tag{77}$$

for any $d \in \mathbb{R}^{n+m}$. Now, considering the function ϕ_λ (72), it holds that for any $d \in \mathbb{R}^{n+m}$,

$$\phi'_\lambda(\bar{x}, \bar{y}; d) = \max \left\{ \psi'_\lambda(\bar{x}, \bar{y}; d), \ \psi'(\bar{x}, \bar{y}; d) \right\}, \tag{78}$$

as $\phi_\lambda(\bar{x}, \bar{y}) = \psi_\lambda(\bar{x}, \bar{y}) = 0$ and $\psi(\bar{x}, \bar{y}) = 0$ if $I^1 \cup I^2 \neq \emptyset$; $\psi'_\lambda(\bar{x}, \bar{y}; d)$ and $\psi'(\bar{x}, \bar{y}; d)$ are given in (76) and (77), respectively. Next, observe that as $\phi'_\lambda(\bar{x}, \bar{y}; 0) = 0$, condition (73) is equivalent to

$$0 \in \partial_d \phi'_\lambda(\bar{x}, \bar{y}; 0), \tag{79}$$

provided that $\phi'_\lambda(\bar{x}, \bar{y}; .)$ is a convex function. This is indeed the case, as $\psi'(\bar{x}, \bar{y}; .)$ and $\psi'_\lambda(\bar{x}, \bar{y}; .)$ are both convex functions. Recall that in (79), ∂_d represents the subdifferential (in the sense of convex analysis) w.r.t. d. It therefore remains to show that we can find an element from $\partial_d \phi'_\lambda(\bar{x}, \bar{y}; 0)$ which is zero. To proceed, first recall that $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ fulfils (28)–(33) and let

$$\varrho := 1 + \sum_{i \in I^1} \bar{u}_i + \sum_{j \in I^2} \bar{v}_j.$$

We have $\varrho > 0$ and subsequently, it holds that $\kappa_\circ + \sum_{i \in I^1} \bar{u}'_i + \sum_{j \in I^2} \bar{v}'_j = 1$ and

$$\Theta := \kappa_\circ \left(\nabla F(x, y) + \lambda \nabla f(\bar{x}, \bar{y}) - \lambda \begin{bmatrix} \nabla_x \ell(\bar{x}, \bar{z}, \bar{w}) \\ 0 \end{bmatrix} \right)$$

$$+ \sum_{i \in I^1} \bar{u}'_i \nabla G_i(\bar{x}, \bar{y}) + \sum_{j \in I^2} \bar{v}'_j \nabla g_j(\bar{x}, \bar{y})$$

$$= 0,$$

with $\kappa_\circ := \frac{1}{\varrho}, \bar{u}'_i := \frac{1}{\varrho}\bar{u}_i$ for $i \in I^1, \bar{v}'_j := \frac{1}{\varrho}\bar{v}_j$ for $j \in I^2$, and $\bar{w} \in \Lambda(\bar{x}, \bar{z})$. By the convexity of the lower-level problem at $\bar{x}$ and the fulfilment of the LLICQ at $(\bar{x}, \bar{z})$, it follows that inclusion $\bar{w} \in \Lambda(\bar{x}, \bar{z})$ is equivalent to $\bar{z} \in S(\bar{x})$. Furthermore, that we can easily show that (78)

can be rewritten as

$$\phi_\lambda'(\bar{x}, \bar{y}; d) = \max \left\{ \nabla G_i(\bar{x}, \bar{y})^\top d \text{ for } i \in I^1, \right.$$

$$\nabla g_i(\bar{x}, \bar{y})^\top d \text{ for } i \in I^2,$$

$$\left. \nabla (F + \lambda f)(\bar{x}, \bar{y})^\top d - \lambda \nabla_x \ell(\bar{x}, z, w)^\top d^1, \ z \in S(\bar{x}) \right\}. \tag{80}$$

Hence, as $\phi_\lambda'(\bar{x}, \bar{y}; 0) = 0$ and all the items in the max-operator, regarded as functions of d, are convex and zero for $d = 0$, it holds that

$$\partial_d \phi_\lambda'(\bar{x}, \bar{y}; 0) = \operatorname{conv} \left\{ \nabla G_i(\bar{x}, \bar{y}) \text{ for } i \in I^1, \nabla g_j(\bar{x}, \bar{y}) \text{ for } j \in I^2, \right.$$

$$\left. \nabla (F + \lambda f)(\bar{x}, \bar{y}) - \lambda \begin{bmatrix} \nabla_1 \ell(\bar{x}, z, w) \\ 0 \end{bmatrix}, \quad z \in S(\bar{x}) \right\},$$

given that $S(\bar{x})$ is a compact set under the uniform compactness assumption made on the mapping K (14) in Theorem 2.1. It clearly follows that $\Theta \in \partial_d \phi_\lambda'(\bar{x}, \bar{y}; 0)$ and $\Theta = 0$.

To prove condition [b], note that ψ is second order directionally differentiable and

$$\psi''(\bar{x}, \bar{y}; d, e) = \max \left\{ G_i''(\bar{x}, \bar{y}; d, e) \text{ for } i \in I^1(d), \ g_j''(\bar{x}, \bar{y}; d, e) \text{ for } j \in I^2(d) \right\} \tag{81}$$

for all $d, e \in \mathbb{R}^{n+m}$. Furthermore, for all $d, e \in \mathbb{R}^{n+m}$, it holds that

$$\psi_\lambda''(\bar{x}, \bar{y}; d, e) \overset{(1)}{=} \lim_{t \downarrow 0} \frac{2}{t^2} \left\{ \left[(F + \lambda f) \left((\bar{x}, \bar{y}) + td + \frac{1}{2} t^2 e \right) \right. \right.$$

$$\left. - (F + \lambda f)(\bar{x}, \bar{y}) - t(F + \lambda f)'(\bar{x}, \bar{y}; d) \right]$$

$$\left. - \lambda \left[\varphi \left(\bar{x} + td^1 + \frac{1}{2} t^2 e^1 \right) - \varphi(\bar{x}) - t\varphi'(\bar{x}; d^1) \right] \right\}$$

$$\overset{(2)}{=} (F + \lambda f)''(\bar{x}, \bar{y}; d, e) - \lambda \varphi''(\bar{x}; d^1)$$

$$\overset{(3)}{=} F''(\bar{x}, \bar{y}; d, e) + \lambda f''(\bar{x}, \bar{y}; d, e)$$

$$- \lambda \inf_{z \in S_1(\bar{x}; d^1)} \left\{ \nabla_1 \ell(\bar{x}, z, w) e^1 + \xi_{d^1}(\bar{x}, z) \right\}$$

$$\overset{(4)}{=} \sup_{z \in S_1(\bar{x}; d^1)} \left\{ F''(\bar{x}, \bar{y}; d, e) + \lambda f''(\bar{x}, \bar{y}; d, e) \right.$$

$$\left. - \lambda \nabla_1 \ell(\bar{x}, z, w) e^1 - \lambda \xi_{d^1}(\bar{x}, z) \right\}$$

$$\overset{(5)}{=} \max_{z \in S_1(\bar{x}; d^1)} \left\{ F''(\bar{x}, \bar{y}; d, e) + \lambda f''(\bar{x}, \bar{y}; d, e) \right.$$

$$\left. - \lambda \nabla_1 \ell(\bar{x}, z, w) e^1 - \lambda \xi_{d^1}(\bar{x}, z) \right\} \tag{82}$$

with $d := d^{1,2}$, $e := e^{1,2}$ and $\xi_{d^1}(\bar{x}, z)$ is defined as in (19). Equality (1) follows from the definition in (16) and the fact that F and f are continuously differentiable and φ is directional differentiable at $\bar{x}$; cf. Theorem 2.1. Equality (2) is based on the fact that F and f

are twice continuously differentiable and φ is second order directionally differentiable; cf. Theorem 2.2. Equality (3) is based on (19) and (4) is obtained thanks to the independence of $F''(\bar{x}, \bar{y}; d, e)$ and $f''(\bar{x}, \bar{y}; d, e)$ from the variable z. As for the final equality, (5), it results from the compactness of the set $S_1(\bar{x}; d)$, which is satisfied under the framework of Theorem 2.2, see corresponding reference.

As for the second order directional derivative of ϕ''_λ, let $d \in C^\lambda(\bar{x}, \bar{y})$. Then $\phi'_\lambda(\bar{x}, \bar{y}; d) = 0$, considering the fact that (73) holds. Furthermore, as the point $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ satisfies the optimality conditions (28)–(33), it follows that

$$\sum_{i=1}^{p} \bar{u}_i \nabla G_i(\bar{x}, \bar{y})^\top d = \sum_{i \in \eta^1} \bar{u}_i \nabla G_i(\bar{x}, \bar{y})^\top d + \sum_{i \in \theta^1} \bar{u}_i \nabla G_i(\bar{x}, \bar{y})^\top d + \sum_{i \in \nu^1} \bar{u}_i \nabla G_i(\bar{x}, \bar{y})^\top d = 0,$$

considering the fact that $\bar{u}_i = 0$ for $i \in \eta^1 \cup \theta^1$ and $\nabla G_i(\bar{x}, \bar{y})^\top d = 0$ for $i \in \nu^1$ based on the fulfilment of (31), cf. partition in (8), and the definition of $C^\lambda(\bar{x}, \bar{y})$, see (68). Similarly, we have $\sum_{j=1}^{q} \bar{v}_j \nabla g_j(\bar{x}, \bar{y})^\top d = 0$. Hence, from (28)–(29), we have

$$\nabla F(\bar{x}, \bar{y})^\top d + \lambda \nabla f(\bar{x}, \bar{y})^\top d - \lambda \nabla_x \ell(\bar{x}, \bar{z}, \bar{w})^\top d^1 = 0. \tag{83}$$

Coming back to the definition of $C^\lambda(\bar{x}, \bar{y})$, the last line in particular, it follows from (77) that

$$\psi'_\lambda(\bar{x}, \bar{y}; d) = \max \left\{ \nabla F(\bar{x}, \bar{y})^\top d + \lambda \nabla f(\bar{x}, \bar{y})^\top d - \lambda \nabla_x \ell(\bar{x}, z, w)^\top d^1, \ z \in S(\bar{x}) \right\} = 0.$$

Hence, from the expressions in (81) and (82), it holds that for all $d \in C^\lambda(\bar{x}, \bar{y})$ and $e \in \mathbb{R}^{n+m}$,

$$\begin{aligned}
\phi''_\lambda(\bar{x}, \bar{y}; d, e) = \max \Big\{ & \nabla G_i(\bar{x}, \bar{y})^\top e + d^\top \nabla^2 G_i(\bar{x}, \bar{y}) d, \ i \in I^1(d), \\
& \nabla g_j(\bar{x}, \bar{y})^\top e + d^\top \nabla^2 g_j(\bar{x}, \bar{y}) d, \ j \in I^2(d), \\
& (F + \lambda f)''(\bar{x}, \bar{y}; d, e) - \lambda \nabla_1 \ell(\bar{x}, z, w)^\top e^1 - \lambda \xi_{d^1}(\bar{x}, z), \ z \in S_1(\bar{x}; d^1) \Big\}
\end{aligned} \tag{84}$$

with $(F + \lambda f)''(\bar{x}, \bar{y}; d, e) = \nabla(F + \lambda f)(\bar{x}, \bar{y})^\top e + d^\top \nabla^2 (F + \lambda f)(\bar{x}, \bar{y}) d$, as $F + \lambda f$ is twice continuously differentiable. The same can be said for any component of G or g.

It follows from (84) that for $d \in C^\lambda(\bar{x}, \bar{y}) \setminus \{0\}$, the optimization problem in (74), i.e. to minimize the function $\phi''_\lambda(\bar{x}, \bar{y}; d, e)$ with respect to $e \in \mathbb{R}^{n+m}$, can be rewritten as

$$\inf_{\varsigma} \left\{ \varsigma_1 \mid A\varsigma \geq b, \ \underline{a}(z)\varsigma \geq \underline{b}(z), \ z \in B \right\}, \tag{85}$$

where A, b, $\underline{a}(z)$, $\underline{b}(z)$, and B are respectively defined by

$$A := \begin{bmatrix} 1 & -\nabla G_1(\bar{x}, \bar{y})^\top \\ \vdots & \vdots \\ 1 & -\nabla G_{l_1}(\bar{x}, \bar{y})^\top \\ 1 & -\nabla g_1(\bar{x}, \bar{y})^\top \\ \vdots & \vdots \\ 1 & -\nabla g_{l_2}(\bar{x}, \bar{y})^\top \end{bmatrix}, \quad b := \begin{bmatrix} d^\top \nabla^2 G_1(\bar{x}, \bar{y}) d \\ \vdots \\ d^\top \nabla^2 G_{l_1}(\bar{x}, \bar{y}) d \\ d^\top \nabla^2 g_1(\bar{x}, \bar{y}) d \\ \vdots \\ d^\top \nabla^2 g_{l_2}(\bar{x}, \bar{y}) d \end{bmatrix},$$

$$\underline{a}(z) := \left[1, -\nabla F(\bar{x}, \bar{y})^\top - \lambda \nabla f(\bar{x}, \bar{y})^\top + \lambda \left[\nabla_1 \ell(\bar{x}, z, w)^\top, 0 \right] \right],$$

$\underline{b}(z) := d^\top \nabla^2 F(\bar{x}, \bar{y}) d + \lambda d^\top \nabla^2 f(\bar{x}, \bar{y}) d - \lambda \xi_{d^1}(\bar{x}, z)$, and $B := S_1(\bar{x}; d^1)$. Note that ι_1 (resp. ι_2) stands for the cardinality of $I^1(d)$ (resp. $I^2(d)$). Clearly, problem (85) is a semi-infinite optimization problem and for any vector $c \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m$ such that $c_1 > 0$ and $c_i := 0$ for $i = 2, \ldots, n + m + 1$,

$$A_i^\top c > 0, \ i = 1, \ldots, n + m \quad \text{and} \quad \underline{a}(z)^\top c > 0 \text{ for all } z \in B,$$

where A_i represents row i for A. Hence, the extended Mangasarian-Fromovitz constraint qualification (see, e.g. [26] for the definition) holds at any feasible point of problem (85). Combining this with the compactness of $S_1(\bar{x}; d^1)$, it follows from the duality theory of linear semi-infinite optimization (cf. latter reference) that the dual of problem (85) can be obtained as

$$\max_{\kappa, u, v} \quad d^\top \nabla^2 \bar{\mathcal{L}}_{\kappa_\circ}^\lambda(\bar{x}, \bar{y}, u, v) d - \lambda \sum_{t=1}^{k} \kappa_t \xi_{d^1}(\bar{x}, z^t) \tag{86}$$

$$\text{s.t.} \quad (\kappa, u, v) \text{ satisfying } (70)\text{–}(71)$$

for some $z^t \in S_1(\bar{x}; d_x), t = 1, \ldots, k$, with $k \in \mathbb{N}$ and $\kappa_\circ := \sum_{i=1}^{\iota_1} \kappa_t$. Hence, condition (74) holds if the one in (69) is satisfied for some $u, v, \kappa_t, z^t \in S_1(\bar{x}; d^1)$ with $\Lambda(\bar{x}, z^t) = \{w^t\}$, for $t = 1, \ldots, k$, where $k \in \mathbb{N}$ and $\kappa_\circ := \sum_{i=1}^{k} \kappa_t$, such that (70)–(71).

Finally, it remains to show that [c] holds; i.e. condition (75) is satisfied. To proceed, first recall that under the assumptions of Theorem 2.2 that the negative of the optimal value function φ (4) is second order epiregular (see [2, Theorem 4.142]) at $\bar{x}$, i.e.

$$-\varphi \left(\bar{x} + td^1 + \frac{1}{2} t^2 e^1(t) \right) \geq -\varphi(\bar{x}) - t\varphi'(\bar{x}; d^1) - \frac{1}{2} t^2 \varphi'' \left(\bar{x}; d^1, e^1(t) \right) - \circ(t^2) \tag{87}$$

for any $d, t \geq 0$ and any function (path) e from $\mathbb{R}_+$ to $\mathbb{R}^n \times \mathbb{R}^m$ (with $e(t) := (e^1(t), e^2(t))$, where $e^1(t) \in \mathbb{R}^n$) such that $te(t) \to 0$ as $t \downarrow 0$. Furthermore, as the function $(F + \lambda f)$ is twice continuously differentiable, it holds that

$$(F + \lambda f) \left((\bar{x}, \bar{y}) + td + \frac{1}{2} t^2 e(t) \right) \geq (F + \lambda f)(\bar{x}, \bar{y}) + t(F + \lambda f)'(\bar{x}, \bar{y}; d)$$

$$+ \frac{1}{2} t^2 (F + \lambda f)'' \left(\bar{x}, \bar{y}; d, e(t) \right) + \circ(t^2). \tag{88}$$

Multiplying (87) by λ, which is positive, and adding the resulting inequality to (88),

$$\left(F + \lambda(f - \varphi) \right) \left((\bar{x}, \bar{y}) + td + \frac{1}{2} t^2 e(t) \right) \geq \left(F + \lambda(f - \varphi) \right) (\bar{x}, \bar{y})$$

$$+ t \left(F + \lambda(f - \varphi) \right)' (\bar{x}, \bar{y}; d) + \frac{1}{2} t^2 \left(F + \lambda(f - \varphi) \right)'' \left(\bar{x}, \bar{y}; d, e(t) \right) + \circ(t^2). \tag{89}$$

Similarly to (88), the upper- and lower-level constraint functions satisfy the second order epiregularity conditions. Subsequently, condition (75) holds. ∎

Assuming that the lower-level optimal solution mapping S (1) is single-valued at $\bar{x}$ can lead to a much simpler result, closely aligned to Theorem 5.2, and subsequently to the derivation of a Robinson-type condition for problem (24). To proceed, we define $C(\bar{x}, \bar{y}) := Q(\bar{x}, \bar{y}, \bar{y})$; cf. (45).

Theorem 6.2: *Let $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ satisfy (28)–(33) for some $\lambda > 0$, with $S(\bar{x}) = \{\bar{y}\} = \{\bar{z}\}$, and USCC (resp. LSCC) hold at $(\bar{x}, \bar{y}, \bar{u})$ (resp. $(\bar{x}, \bar{y}, \bar{v})$), i.e. $\theta^1 = \emptyset$ (resp. $\theta^2 = \emptyset$). Suppose that the lower-level problem is convex at $\bar{x}$ and the assumptions in Theorem 2.3 are satisfied for all $d \in C(\bar{x}, \bar{y})$. Then, $(\bar{x}, \bar{y})$ is a strict local optimal solution of problem (24) provided that the ULICQ (44) holds at $(\bar{x}, \bar{y})$ and, for all $d \in C(\bar{x}, \bar{y}) \setminus \{0\}$, we have*

$$d^\top \nabla^2 \mathcal{L}^\lambda(\bar{x}, \bar{y}, \bar{u}, \bar{v})d > \lambda \left(d^{1,3}\right)^\top \nabla^2 \ell(\bar{x}, \bar{y}, \bar{w})d^{1,3}. \tag{90}$$

Proof: Proceeding as in the proof of the previous theorem, we have the corresponding expressions of (80) for $d \in \mathbb{R}^{n+m}$, and (81) for $d \in C(\bar{x}, \bar{y})$ and $e \in \mathbb{R}^{n+m}$, given that $S_1(\bar{x}; d^1) \subseteq S(\bar{x}) = \{\bar{y}\}$; cf. (18). Further proceeding as in the previous theorem, the point $(\bar{x}, \bar{y})$ is a strict local optimal solution of problem (24) if for all $d \in C(\bar{x}, \bar{y})$, it holds that

$$d^\top \nabla^2 \bar{\mathcal{L}}^\lambda_{\kappa_\circ}(\bar{x}, \bar{y}, u, v)d > \lambda \kappa_\circ \xi_{d^1}(\bar{x}, \bar{y}) \tag{91}$$

for some $(\kappa_\circ, u, v)$ verifying (71) together with the following conditions, while also taking into account the fact that $S_1(\bar{x}; d^1) \subseteq S(\bar{x}) = \{\bar{y}\}$ and $\{\bar{w}\} = \Lambda(\bar{x}, \bar{y})$ (cf. Theorem 2.3):

$$\nabla_1 \bar{\mathcal{L}}^\lambda_{\kappa_\circ}(\bar{x}, \bar{y}, u, v) - \lambda \kappa_\circ \nabla_1 \ell(\bar{x}, \bar{y}, \bar{w}) = 0, \quad \nabla_2 \bar{\mathcal{L}}^\lambda_{\kappa_\circ}(\bar{x}, \bar{y}, u, v) = 0. \tag{92}$$

One can easily show that the existence of $(\kappa_\circ, u, v)$ verifying (30) and (33) (cf. $\{\bar{w}\} = \Lambda(\bar{x}, \bar{y})$), together with (71) and (92), such that (91) holds, is equivalent to the existence of some $(\kappa'_\circ, u', v') \neq 0$ satisfying (30), (33), (92), and

$$\kappa_\circ \geq 0, \quad u_i \geq 0 \text{ for } i \in I^1(d), v_j \geq 0 \quad \text{for } j \in I^2(d)$$

such that (91) holds, with $(\kappa_\circ, u, v)$ replaced by $(\kappa'_\circ, u', v')$. Subsequently, taking into account the fact that $I^1(d) \subseteq I^1$ and $I^2(d) \subseteq I^2$, it is not difficult to show that with the ULICQ satisfied at $(\bar{x}, \bar{y})$, this point is a strict local optimal solution of problem (24) provided for all $d \in C(\bar{x}, \bar{y})$,

$$d^\top \nabla^2 \bar{\mathcal{L}}^\lambda_1(\bar{x}, \bar{y}, u, v)d > \lambda \xi_{d^1}(\bar{x}, \bar{y}) \tag{93}$$

holds for some (u, v) such that $(\bar{x}, \bar{y}, \bar{z}, u, v, \bar{w})$ satisfies (28)–(33), with $\bar{z} = \bar{y}$. It therefore remains to show that we have $\bar{u} = u$ and $\bar{v} = v$.

To proceed, let us first show that $I^1(d) = I^1$ and $I^2(d) = I^2$ for any $d \in C(\bar{x}, \bar{y})$. Obviously, $I^1(d) \subseteq I^1$ and $I^2(d) \subseteq I^2$, by definition. For the converse, let $i \in I^1$. Then $i \in \nu^1$, given that $\theta^1 = \emptyset$. Furthermore, considering the definition of $C(\bar{x}, \bar{y})$, it follows that $\nabla G_i(\bar{x}, \bar{y})^\top d = 0$. Hence, $i \in I^1(d)$. Subsequently, as $(\bar{x}, \bar{y}, \bar{u}, \bar{v}, \bar{w})$ and $(\bar{x}, \bar{y}, u, v, \bar{w})$ both satisfy (28)–(33) with $\bar{z} = \bar{y}$,

$$\sum_{i \in I^1}(\bar{u}_i - u_i)\nabla G_i(\bar{x}, \bar{y}) + \sum_{j \in I^2}(\bar{v}_j - v_j)\nabla g_j(\bar{x}, \bar{y}) = 0.$$

Based on the ULICQ at $(\bar{x}, \bar{y})$, it follows that $\bar{u} = u$ and $\bar{v} = v$, given that the components of these vectors are all zero when $i \notin I^1$ and $j \notin I^2$, respectively. To conclude the

proof, observe that condition (90) is sufficient for (93) to hold, considering the definition of $\xi_{d^1}(\bar{x}, \bar{y})$ in (20). ∎

It is clear that under the framework of this theorem, the conclusion of Theorem 5.2 is valid, while also guarantying that the resulting point $(\bar{x}, \bar{y})$ is a strict local optimal solution of problem (24). Hence, Theorem 6.2 can be seen as an extension of the Robinson condition to bilevel optimization, in the sense discussed at the beginning of this section.

A common point between Theorems 6.1 and 6.2 is that the SSOSC-type condition, i.e. (69) and (90), respectively, are the main assumptions, as the remaining ones (except from the USCC and LSCC in Theorem 6.2) are mostly technical, helping to ensure that the LLVF φ (4) is second order directionally differentiable. Observe that the USCC and LSCC in Theorem 6.2 help to ensure that points satisfying (70)–(71) coincide with stationarity points in the sense (28)–(33).

Next, we provide a small example illustrating Theorem 6.2, with the corresponding Robinson-type framework.

Example 6.3: Considering the simple bilevel optimization problem

$$\min_{x,y} xy \quad \text{s.t.} \quad x + y \le 2, \; y \in S(x) := \arg\min_y \left\{ y \mid x - y \le 0 \right\},$$

we obviously have $S(x) = \{x\}$ for all $x \in \mathbb{R}$ and one can easily check that the vector $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ with $\bar{x} = \bar{y} = \bar{z} = 0, \bar{u} = 0, \bar{v} = \lambda$, and $\bar{w} = 1$, satisfies the optimality conditions (28)–(33) for any $\lambda > 0$. For this vector, $\theta^1 = v^1 = \emptyset, \theta^2 = \theta^3 = \emptyset$ and $v^2 = \eta^3 = \{1\}$. Hence, the critical cone $C(\bar{x}, \bar{y}, \bar{z}) = \{d \in \mathbb{R}^3 \mid d_1 = d_2 = d_3\}$, and subsequently, for any $d \in C(\bar{x}, \bar{y}, \bar{z}) \setminus \{0\}$,

$$\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}^\top \nabla^2 \mathcal{L}^\lambda(\bar{\zeta}) \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} - \begin{pmatrix} d_1 \\ e_2 \end{pmatrix}^\top \nabla^2 \ell(\bar{\zeta}) \begin{pmatrix} d_1 \\ e_2 \end{pmatrix} = 2d_1^2 > 0.$$

All the other assumptions of Theorem 6.2 are also satisfied. $(0, 0)$ is indeed the unique optimal solution of the problem above and the corresponding penalized version (24) for any $\lambda > 0$.

Note that all the assumptions in Theorem 6.2 also hold for the problem in Example 5.5, given that at $(\bar{x}, \bar{y}, \bar{v}) := (0, 0, 0, \lambda, \lambda)$ and $(\bar{x}, \bar{z}, \bar{w}) := (0, 0, 0, 1, 1), \eta^2 = \emptyset, \theta^2 = \emptyset, v^2 = \{1, 2\}$ and $\eta^3 = \emptyset, \theta^3 = \emptyset, v^3 = \{1, 2\}$, respectively. Hence, having $\nabla_2 g_i(\bar{x}, \bar{z}, \bar{w})^\top (e_2, e_3)^\top = 0$ for $i \in v^3$, is equivalent to $e_2 = e_3 = 0$. Hence, combining this with the calculations in Example 5.5, it is clear that all the assumptions in Theorem 6.2 hold for problem (67).

For examples, where the assumptions in Theorem 6.1 (ensuring the second order directional differentiability of φ (4)) hold without the uniqueness of the lower-level optimal solution, see, e.g. [2,49].

To close this section, we provide an example confirming that the SSOSC-type condition (90) is essential in guaranteeing that a point is locally optimal for problem (24).

Example 6.4: Consider the following example from [6, Chapter 5]:

$$\min_{x,y} (x - 3.5)^2 + (y + 4)^2 \quad \text{s.t.} \quad y \in S(x) := \arg\min_y \left\{ (y - 3)^2 \mid -x + y^2 \le 0 \right\}. \quad (94)$$

The lower-level solution set-valued mapping is single-valued; i.e. precisely,

$$S(x) = \begin{cases} \emptyset & \text{if } x < 0, \\ \{\sqrt{x}\} & \text{if } 0 \leq x \leq 9, \\ \{3\} & \text{if } x > 9. \end{cases}$$

The point $(1, 1)$ is not optimal for (94) but we can easily check that $(1, 1, 1, 2\lambda - 5, 2)$ satisfies the optimality conditions (28)–(33), for $\lambda > \frac{5}{2}$. Obviously, at $(1, 1, 2\lambda - 5)$ and $(1, 1, 2)$, we have

$$\eta^2 = \emptyset, \quad \theta^2 = \emptyset, \quad \nu^2 = \{1\} \quad \text{and} \quad \eta^3 = \emptyset, \quad \theta^3 = \emptyset, \quad \nu^3 = \{1\},$$

respectively. Moreover, $C(1, 1, 1) = \{(d_1, d_2, e_2) \mid d_2 = e_2 = \frac{1}{2}d_1\}$ and for all $(d_1, d_2, e_2) \in C(1, 1, 1)$,

$$\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}^\top \nabla^2 \mathcal{L}^\lambda (1, 1, 1, 2\lambda - 5, 2) \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} - \begin{pmatrix} d_1 \\ e_2 \end{pmatrix}^\top \nabla^2 \ell (1, 1, 2) \begin{pmatrix} d_1 \\ e_2 \end{pmatrix} = 0.$$

We can however check that all the other assumptions of Theorem 6.2 are satisfied for $(1, 1, 1, 2\lambda - 5, 2)$.

Further analysis on the necessity of condition (90) and the corresponding condition in Theorem 6.1, will be studied more closely in a future work. Different types of sufficient optimality conditions for bilevel optimization problems can be found in the papers [5,10].

7. Numerical experiments

Based on our implementation of Algorithm 1 in MATLAB (R2018a), we report and discuss test results obtained for the 124 nonlinear bilevel programs from the current version of the BOLIB library [58].

Recall that the necessary optimality conditions (28)–(33) and their reformulation (39) as nonsmooth system of equations contain the penalization parameter $\lambda > 0$. Based on the construction process of the system (see, e.g. [13,55]), there is no specific rule on how to choose the best value of this parameter. Rather, one may try all λ from a certain finite discrete set in $(0, \infty)$, solve the corresponding optimality conditions, and then choose the best solution in terms of the upper-level objective function value. For our approach, it turned out that a small set of λ-values is sufficient to reach very good results. To be precise, for all our experiments, we just used the nine values of λ in $\Lambda := \lambda \in \{2^{-1}, 2^0, \cdots, 2^6, 2^7\}$.

7.1. Implementation details

Besides the selection of penalization parameters described before, the other parameters needed in Algorithm 1 are set to

$$\beta := 10^{-8}, \quad \epsilon := 10^{-8}, \quad t := 2.1, \quad \rho := 0.5, \quad \sigma := 10^{-4}.$$

For each test example, we only used one starting point. The choices for starting points x^o and y^o are as follows. If an example in literature comes with a starting point, then we

use this point for our experiments. Otherwise, we choose $x^o = \mathbf{1}_n, y^o = \mathbf{1}_m$ except for three examples **No** 20, 119, and 120 because their global optimal solutions are $(\mathbf{1}_n, \mathbf{1}_m)$, where $\mathbf{1}_n := (1, \cdots, 1)^\top \in \mathbb{R}^n$. So, for these three examples we used $x^o = -\mathbf{1}_n, y^o = -\mathbf{1}_m$. Detailed information on starting points can be found in [21]. Moreover, to fully define $\zeta^o = (x^o, y^o, z^o, u^o, v^o, w^o)$, we set

$$z^o := y^o, \quad u^o := \left(|G_1(x^o, y^o)|, \cdots, |G_p(x^o, y^o)|\right)^\top,$$

$$v^o := \left(|g_1(x^o, y^o)|, \cdots, |g_q(x^o, y^o)|\right)^\top$$

and $w^o := v^o$. In addition to the stopping criterion $\|\Phi^\lambda(\zeta^k)\| \leq \epsilon$ used in Algorithm 1, the algorithm is terminated if the iteration index k reaches 2000.

Finally, to pick an element from the generalized B-subdifferential $\partial_B \Phi^\lambda(\zeta^k)$ in Step 2 of Algorithm 1, we adopt the technique in [4].

7.2. Test results

Table 1 lists, for any of the 124 test examples of the BOLIB library [58], values of the leader's objective function F. The column $\boldsymbol{F_{known}}$ shows the best known F-values from the literature. Such a value was not available for 6 of the test problems. This is marked by 'unknown' in the column **Status**. For 83 examples, the best known F-value is even optimal (with status labelled as 'optimal'). For the remaining 35 test problems, the known $\boldsymbol{F}$-value might not be optimal and its status is just set to 'known'. The columns below $\boldsymbol{F_{new}}$ show the F-values obtained by Algorithm 1 for the nine penalization parameters λ.

Note that three examples contain a parameter that should be provided by the user. They are examples **No** 14, 39, and 40. 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 the other three cases, our method still produced the true global optimal solutions. Example **No** 39 has a unique global optimal solution. The result presented in Table 1 is for $c = 0$. We also tested our method when $c = \pm 1$, and obtained the unique optimal solutions as well. Example **No** 40 contains the parameter $M > 1$, and the results presented in Table 1 correspond to $M = 1.5$.

Let us first note that evaluating the performance of an algorithm for the bilevel optimization problem (1) is a difficult task since the decision whether a computed point is (close to) a global solution of (1) basically requires computing the LLVF φ. Therefore, instead of doing this, we suggest the following way of comparing our results for Algorithm 1 with the results from literature known for the test problems. For an approximate solution (x, y) obtained from Algorithm 1, we first compute

$$\delta_F := \frac{F(x, y) - F_{known}}{\max\{1, |F_{known}|\}}, \quad \delta_f := \frac{f(x, y) - f_{known}}{\max\{1, |f_{known}|\}},$$

where F_{known} as above is the best known F-value from literature and f_{known} the lower-level function value which corresponds to F_{known}. Moreover, we set

$$\delta_* := \begin{cases} \max\{|\delta_F|, |\delta_f|\}, & \text{if \textbf{Status} is optimal,} \\ \max\{\delta_F, \delta_f\}, & \text{otherwise.} \end{cases}$$

Table 1. Upper-level objective function values at the solution for different selections of $\lambda \in \Lambda$.

No	Examples	Status	F_{known}	F_{new}									δ_*
				$\lambda = 2^{-1}$	$\lambda = 2^0$	$\lambda = 2^1$	$\lambda = 2^2$	$\lambda = 2^3$	$\lambda = 2^4$	$\lambda = 2^5$	$\lambda = 2^6$	$\lambda = 2^7$	
1	AiyoshiShimizu1984Ex2	optimal	5.00	8.25	1.00	10.31	15.16	14.61	14.80	39.75	14.95	4.97	0.01
2	AllendeStill2013	optimal	1.00	1.44	1.25	1.11	0.82	0.90	0.94	0.97	0.98	0.99	0.01
3	AnEtal2009	optimal	2251.6	2251.6	2251.6	2251.6	2251.6	2251.6	2826.5	2269.5	2251.6	4425.5	0.00
4	Bard1988Ex1	optimal	17.00	49.31	23.13	50.00	53.78	58.55	62.59	65.33	25.00	17.00	0.00
5	Bard1988Ex2	optimal	−6600.0	−6600.0	−6600.0	0.0	0.0	−6600.0	−5641.6	−5641.6	−6600.0	0.0	0.00
6	Bard1988Ex3	optimal	−12.68	−12.68	−14.51	−13.89	−13.40	−13.07	−12.89	−12.78	−12.73	−12.71	0.00
7	Bard1991Ex1	optimal	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00
8	BardBook1998	optimal	0.00	0.00	0.00	0.00	0.00	100.00	11.11	11.11	11.11	11.11	0.00
9	CalamaiVicente1994a	optimal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	CalamaiVicente1994b	optimal	0.31	0.38	0.34	1.06	1.22	1.18	1.46	0.54	1.54	0.31	0.00
11	CalamaiVicente1994c	optimal	0.31	0.37	0.56	1.00	1.22	0.48	0.52	1.30	0.31	1.30	0.00
12	CalveteGale1999P1	optimal	−29.20	−20.00	−20.00	−29.20	−29.20	−13.00	−13.00	−29.20	−23.00	−16.00	0.00
13	ClarkWesterberg1990a	optimal	5.00	9.00	9.00	5.00	9.00	13.00	7.97	9.77	9.86	9.93	0.00
14	Colson2002BIPA1	optimal	250.00	250.00	250.00	250.00	250.00	250.00	250.00	250.00	250.00	250.00	0.00
15	Colson2002BIPA2	known	17.00	49.31	23.13	50.00	53.78	58.55	62.59	65.33	66.95	17.00	0.00
16	Colson2002BIPA3	known	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00
17	Colson2002BIPA4	known	88.79	89.98	88.93	88.79	88.79	88.79	88.79	87.31	88.02	88.39	0.00
18	Colson2002BIPA5	known	2.75	3.42	2.00	2.75	30.64	2.99	3.12	3.19	3.22	3.23	−0.27
19	Dempe1992a	unknown		−2.00	0.00	0.00	−0.25	−0.13	−0.06	−0.03	−0.02	−0.01	
20	Dempe1992b	optimal	31.25	33.98	33.49	32.41	31.25	31.25	31.25	31.25	31.25	31.25	0.00
21	DempeDutta2012Ex24	optimal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	276.17	0.00	0.00
22	DempeDutta2012Ex31	optimal	−1.00	−1.07	−0.66	−0.38	−0.21	−0.11	0.00	−0.03	−0.02	−0.01	0.07
23	DempeFranke2011Ex41	optimal	−1.00	−1.00	−1.00	−1.00	0.00	−1.00	−1.00	0.00	0.00	0.00	0.00
24	DempeFranke2011Ex42	optimal	5.00	−0.88	0.20	1.20	1.80	4.20	5.37	4.83	4.99	10.37	0.00
25	DempeFranke2014Ex38	optimal	2.13	−0.88	−0.50	1.00	2.13	2.15	2.13	2.12	4.00	2.12	0.00
26	DempeEtal2012	optimal	−1.00	−3.00	−1.00	−1.00	−1.00	0.00	−1.00	−1.00	−1.00	−1.00	0.00
27	DempeLohse2011Ex31a	optimal	−6.00	0.13	−5.50	−5.50	−5.50	−0.16	−0.04	0.18	0.33	0.41	0.00
28	DempeLohse2011Ex31b	optimal	−12.00	−12.00	−12.00	−12.00	−12.00	−12.00	−5.38	−5.47	0.35	0.42	0.00
29	DeSilva1978	optimal	−1.00	−0.56	−0.75	−0.89	−1.18	−1.10	−1.06	−1.03	−1.02	−1.01	0.01
30	FalkLiu1995	optimal	−2.20	−1.56	−2.00	−2.16	−2.22	−2.51	−2.38	−2.32	−2.28	−2.27	0.01
31	FloudasEtal2013	optimal	0.00	8.25	0.00	5.81	15.16	39.75	29.75	29.88	14.95	14.98	0.00
32	FloudasZlobec1998	optimal	1.00	100.00	1.00	100.00	100.00	0.92	0.97	100.00	0.99	1.00	0.00
33	GumusFloudas2001Ex1	optimal	2250.0	2250.9	2250.0	2250.0	2500.0	2250.0	2311.2	2308.7	2306.9	2500.0	0.00

(*continued*).

Table 1. Continued

No	Examples	Status	F_{known}	$\lambda = 2^{-1}$	$\lambda = 2^0$	$\lambda = 2^1$	$\lambda = 2^2$	$\lambda = 2^3$	$\lambda = 2^4$	$\lambda = 2^5$	$\lambda = 2^6$	$\lambda = 2^7$	δ_*
								F_{new}						
34	GumusFloudas2001Ex3	optimal	−29.20	−20.00	−20.00	−29.20	−20.00	−13.00	−20.00	−13.00	−48.95	−13.00	0.00
35	GumusFloudas2001Ex4	optimal	9.00	9.00	9.00	9.00	9.00	9.00	7.97	8.46	8.73	8.86	0.00
36	GumusFloudas2001Ex5	optimal	0.19	0.19	0.19	0.19	2.18	0.19	−0.39	0.19	0.19	0.19	0.00
37	HatzEtal2013	optimal	0.00	1.00	0.50	0.25	−0.13	1079.59	648.93	376.50	406.77	217.98	0.13
38	HendersonQuandt1958	known	−3266.7	−3243.6	−3200.0	−3239.7	−3260.5	−3265.2	−3266.3	−3266.6	−3266.6	−3275.2	0.00
39	HenrionSurowiec2011	optimal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40	IshizukaAiyoshi1992a	optimal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
41	KleniatiAdjiman2014Ex3	optimal	−1.00	−1.33	−1.33	−0.59	−0.53	−0.90	−0.75	−0.97	0.99	−0.99	0.01
42	KleniatiAdjiman2014Ex4	known	−10.00	−4.00	−4.48	−5.00	−3.19	−1.16	−2.00	−2.00	−3.31	−6.68	0.52
43	LamparSagrat2017Ex23	optimal	−1.00	−1.00	−1.00	−1.00	−1.00	−1.00	−1.00	−1.00	−1.00	−1.00	0.00
44	LamparSagrat2017Ex31	optimal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
45	LamparSagrat2017Ex32	optimal	0.50	0.63	0.56	0.32	0.40	0.44	0.47	0.48	0.49	0.50	0.00
46	LamparSagrat2017Ex33	optimal	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.00
47	LamparSagrat2017Ex35	optimal	0.80	1.25	0.80	1.00	1.25	1.00	0.80	1.00	0.80	1.25	0.00
48	LucchettiEtal1987	optimal	0.00	0.50	0.75	0.00	0.82	0.90	0.94	0.97	0.98	0.99	0.00
49	LuDebSinha2016a	known	1.14	1.14	1.14	1.52	1.14	1.14	2.24	1.17	2.24	2.24	0.00
50	LuDebSinha2016b	known	0.00	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04
51	LuDebSinha2016c	known	1.12	1.58	1.64	1.60	1.62	1.12	1.64	1.64	1.64	1.12	0.00
52	LuDebSinha2016d	unknown		0.00	−0.29	20.00	−192.00	19.80	−18.83	−192.00	−34.84	−200.75	
53	LuDebSinha2016e	unknown		11.79	7.27	2.09	5.44	2.78	28.65	15.95	13.18	7.24	
54	LuDebSinha2016f	unknown		0.00	0.00	0.00	−18.63	20.00	−165.12	0.00	0.00	0.00	
55	MacalHurter1997	optimal	81.33	81.85	81.46	81.26	81.30	81.31	81.32	81.32	81.33	81.33	0.00
56	Mirrlees1999	optimal	1.00	0.00	0.00	0.00	0.14	4.01	0.01	0.01	0.87	0.01	0.13
57	MitsosBarton2006Ex38	optimal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
58	MitsosBarton2006Ex39	optimal	−1.00	−1.00	−1.00	−1.00	−1.00	−1.00	0.08	0.01	0.00	0.00	0.00
59	MitsosBarton2006Ex310	optimal	0.50	−0.82	−0.71	−0.63	0.42	−0.54	−0.09	−0.07	0.50	0.50	0.00
60	MitsosBarton2006Ex311	optimal	−0.80	−0.80	−0.80	0.51	1.00	−0.09	−0.80	−0.50	0.50	0.18	0.30
61	MitsosBarton2006Ex312	optimal	0.00	10.00	10.00	5.01	−0.26	−0.13	2.58	−0.03	−0.02	7.30	0.02
62	MitsosBarton2006Ex313	optimal	−1.00	−1.58	−1.00	−2.00	−2.00	0.03	0.87	0.19	0.00	0.23	0.00
63	MitsosBarton2006Ex314	optimal	0.25	0.01	0.02	0.04	0.05	0.06	0.06	0.06	0.06	0.06	0.19
64	MitsosBarton2006Ex315	optimal	0.00	−2.00	−2.00	−0.63	0.65	1.59	1.80	1.90	1.95	1.98	0.65
65	MitsosBarton2006Ex316	optimal	−2.00	−3.00	−3.00	−3.00	1.01	0.00	−2.06	0.28	−0.50	−3.00	0.03
66	MitsosBarton2006Ex317	optimal	0.19	0.00	0.00	0.03	0.06	0.13	0.00	0.00	0.00	0.24	0.05
67	MitsosBarton2006Ex318	optimal	−0.25	0.00	−1.00	0.75	0.56	0.36	0.21	0.11	−1.00	0.00	0.25

(*continued*).

Table 1. Continued

No	Examples	Status	F_{known}	F_{new}									δ_*
				$\lambda = 2^{-1}$	$\lambda = 2^0$	$\lambda = 2^1$	$\lambda = 2^2$	$\lambda = 2^3$	$\lambda = 2^4$	$\lambda = 2^5$	$\lambda = 2^6$	$\lambda = 2^7$	
68	MitsosBarton2006Ex319	optimal	−0.26	0.00	0.00	0.00	−0.31	0.47	0.57	0.00	−0.26	−0.02	0.00
69	MitsosBarton2006Ex320	optimal	0.31	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.03	0.03	0.28
70	MitsosBarton2006Ex321	optimal	0.21	0.21	0.21	0.04	0.11	0.36	0.18	0.20	0.20	0.21	0.00
71	MitsosBarton2006Ex322	optimal	0.21	0.21	0.21	0.04	0.11	0.16	0.18	0.20	0.20	0.21	0.00
72	MitsosBarton2006Ex323	optimal	0.18	0.18	0.18	0.18	0.18	0.18	0.15	0.28	0.00	0.25	0.00
73	MitsosBarton2006Ex324	optimal	−1.76	−1.99	−1.75	−0.74	−1.71	−0.94	−0.29	−0.29	−1.76	−0.29	0.00
74	MitsosBarton2006Ex325	known	−1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
75	MitsosBarton2006Ex326	optimal	−2.35	−2.00	−2.00	1.36	0.32	−1.00	−2.00	0.95	1.00	0.32	0.15
76	MitsosBarton2006Ex327	known	2.00	0.04	2.06	0.01	2.06	1.16	0.64	1.00	1.00	1.00	0.05
77	MitsosBarton2006Ex328	known	−10.00	−5.04	−3.70	−3.89	−2.86	−3.00	−1.01	−3.95	−5.00	−2.94	0.97
78	MorganPatrone2006a	optimal	−1.00	−1.00	−1.00	−1.00	0.50	−1.00	0.50	0.94	0.97	0.98	0.00
79	MorganPatrone2006b	optimal	−1.25	0.50	−1.25	−0.50	0.50	0.50	0.63	0.69	−0.50	−0.75	0.00
80	MorganPatrone2006c	optimal	−1.00	−1.00	1.00	1.00	1.00	1.00	1.00	0.75	0.75	1.00	0.00
81	MuuQuy2003Ex1	known	−2.08	−3.26	−2.88	−2.56	−2.34	−2.22	−2.15	−2.11	−2.10	−2.09	−0.55
82	MuuQuy2003Ex2	known	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.01
83	NieWangYe2017Ex34	optimal	2.00	2.00	2.00	2.00	2.00	2.00	3.89	3.80	2.00	4.33	0.00
84	NieWangYe2017Ex52	optimal	−1.71	0.06	−1.39	0.00	0.39	0.00	−0.24	0.00	−0.10	0.00	0.19
85	NieWangYe2017Ex54	optimal	−0.44	0.00	0.00	0.00	0.00	0.00	0.00	−0.15	−0.01	0.00	0.98
86	NieWangYe2017Ex57	known	−2.00	−2.05	−2.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
87	NieWangYe2017Ex58	known	−3.49	−3.53	0.00	0.00	−0.08	0.00	0.00	0.00	0.77	−0.91	0.02
88	NieWangYe2017Ex61	known	−1.02	0.19	−1.11	−1.03	0.39	−1.02	−0.18	0.00	0.00	0.00	0.00
89	Outrata1990Ex1a	known	−8.92	−10.67	−10.02	−9.54	−9.25	−9.09	−9.01	−8.96	0.00	−8.93	0.00
90	Outrata1990Ex1b	known	−7.56	−10.43	−9.46	−8.69	−8.19	−7.90	−7.74	−7.66	−7.62	0.00	0.01
91	Outrata1990Ex1c	known	−12.00	−12.00	−12.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
92	Outrata1990Ex1d	known	−3.60	−4.64	−3.33	−3.67	−3.64	−0.56	−0.47	−3.60	−3.60	−0.37	0.01
93	Outrata1990Ex1e	known	−3.15	−3.93	−3.93	−3.92	−3.92	−0.67	−3.92	−0.39	−0.34	−3.79	−0.10
94	Outrata1990Ex2a	known	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.00
95	Outrata1990Ex2b	known	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.00
96	Outrata1990Ex2c	known	1.86	1.03	1.34	1.57	1.70	2.35	5.82	1.84	1.85	5.49	0.00
97	Outrata1990Ex2d	known	0.92	0.85	0.85	0.85	0.85	0.85	0.85	0.85	9.76	0.85	−0.07
98	Outrata1990Ex2e	known	0.90	0.47	0.63	0.74	0.82	0.86	0.88	0.89	0.89	5.56	0.00
99	Outrata1993Ex31	known	1.56	2.57	2.05	1.45	1.50	1.53	1.55	8.95	1.56	1.56	0.00
100	Outrata1993Ex32	known	3.21	4.02	3.59	3.11	3.16	3.18	3.19	3.20	3.20	3.21	0.00

(*continued*).

Table 1. Continued

No	Examples	Status	F_{known}	F_{new}									δ_*
				$\lambda = 2^{-1}$	$\lambda = 2^0$	$\lambda = 2^1$	$\lambda = 2^2$	$\lambda = 2^3$	$\lambda = 2^4$	$\lambda = 2^5$	$\lambda = 2^6$	$\lambda = 2^7$	
101	Outrata1994Ex31	known	3.21	4.02	3.59	3.11	3.16	3.18	3.19	3.20	3.20	8.01	0.00
102	OutrataCervinka2009	optimal	0.00	0.00	0.00	0.00	−0.13	−0.06	−0.03	−0.02	−0.01	0.00	0.00
103	PaulaviciusEtal2017a	optimal	0.25	0.00	0.78	0.31	0.11	0.00	1.59	1.00	1.00	1.94	0.09
104	PaulaviciusEtal2017b	optimal	−2.00	−2.00	−2.00	−2.00	−2.00	−2.00	−2.00	0.94	−0.22	0.96	0.00
105	SahinCiric1998Ex2	optimal	5.00	9.00	9.00	5.00	9.00	13.00	7.97	9.77	9.86	9.93	0.00
106	ShimizuAiyoshi1981Ex1	optimal	100.00	73.96	83.47	90.50	94.88	97.33	98.64	99.31	99.65	99.83	0.00
107	ShimizuAiyoshi1981Ex2	optimal	225.00	237.00	225.00	225.00	175.00	200.00	212.50	218.75	221.88	1155.9	0.00
108	ShimizuEtal1997a	unknown		25.00	23.13	16.89	25.00	58.55	16.89	16.89	66.95	65.34	
109	ShimizuEtal1997b	optimal	2250.0	2369.0	2269.1	2250.0	2321.7	2250.0	2311.2	2308.6	2426.5	2305.8	0.00
110	SinhaMaloDeb2014TP3	known	−18.68	−9.63	−10.36	−10.36	−5.27	−8.63	−5.27	−9.63	−18.73	−18.71	0.00
111	SinhaMaloDeb2014TP6	known	−1.21	−1.21	−1.21	−1.21	−1.21	2.32	−1.21	1.52	−1.21	−1.21	0.00
112	SinhaMaloDeb2014TP7	known	−1.96	−1.98	−1.96	−1.98	−1.00	−1.98	0.00	0.00	0.00	0.00	0.00
113	SinhaMaloDeb2014TP8	optimal	0.00	50.36	48.58	45.65	0.00	0.00	1225.00	29.31	0.00	388.32	0.00
114	SinhaMaloDeb2014TP9	known	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10.32	10.21	0.00
115	SinhaMaloDeb2014TP10	known	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
116	TuyEtal2007	optimal	22.50	24.73	24.48	24.01	25.00	25.00	25.00	25.00	22.50	22.50	0.00
117	Vogel2012	optimal	1.00	0.44	1.78	2.78	4.00	3.67	3.84	3.92	3.96	3.98	0.78
118	WanWangLv2011	optimal	10.62	12.59	13.78	11.25	11.25	10.94	36.25	7.50	17.00	10.63	0.00
119	YeZhu2010Ex42	optimal	1.00	1.09	1.05	8.78	6.11	5.53	5.39	5.19	5.09	6.11	0.05
120	YeZhu2010Ex43	optimal	1.25	22.99	25.25	12.03	10.50	1.00	9.38	9.19	9.09	9.05	0.20
121	Yezza1996Ex31	optimal	1.50	1.50	1.50	1.50	1.50	1.50	1.50	3.51	3.96	3.49	0.00
122	Yezza1996Ex41	optimal	0.50	0.72	0.62	4.00	4.00	4.00	4.00	4.00	4.00	4.00	0.20
123	Zlobec2001a	optimal	−1.00	0.00	0.50	−0.04	0.00	0.00	−1.00	0.00	0.00	−1.00	0.00
124	Zlobec2001b	unknown		0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	

Table 2. Performance of Algorithm 1 for $\lambda \in \Lambda$.

λ	2^{-1}	2^0	2^1	2^2	2^3	2^4	2^5	2^6	2^7
Number of failures	6	3	2	1	5	9	12	14	16
$\alpha_K = 1$	109	111	113	115	112	108	108	109	103
$y^K \approx z^K$	85	83	75	70	65	64	68	82	96
$v^K \approx w^K$	32	45	39	37	37	41	41	38	42
Average iterations	152.3	84.3	129.1	154.3	194.6	288.9	357.4	375.9	451.3
Average time	0.18	0.08	0.13	0.17	0.21	0.27	0.32	0.34	0.40

In the latter case, δ can become negative. This means that both $F(x, y)$ and $f(x, y)$ are smaller than the values for the point with best F-value known from literature. In the last column of Table 1, marked by δ_* we provide the smallest δ-value among those obtained for all $\lambda \in \Lambda$ for the corresponding test problem.

We then report the performance of Algorithm 1 for $\lambda \in \Lambda$. Let K denote the iteration number, where Algorithm 1 is terminated. The first row in Table 2 shows the number of examples for which Algorithm 1 failed, i.e. it did not reach $\|\Phi^\lambda(\zeta^K)\| \le \epsilon$. Failures were observed only for a very small portion of the 124 examples. The second row in Table 2 reports for how many examples the last step was a full Newton step, i.e. $\alpha_K = 1$. This behaviour could be observed for more than 100 examples regardless of $\lambda \in \Lambda$.

We also want to consider whether $y^K \approx z^K$ occurs. As mentioned in Remark 3.2, $y = z$ generates another well-known type of stationary conditions, i.e. (29)–(33) and (35). Moreover, w is a multiplier associated to the lower-level constraint function g in connection to the lower-level problem, v is associated to the same function g, but in connection to the upper-level problem. The analogy in the complementarity systems (32) and (33) raises the question whether v and w coincide in some examples. In Table 2, the fifth and sixth rows list how many examples, for each λ, were solved with $y^K \approx z^K$ or $v^K \approx w^K$. The symbol '$\approx$' used here means that $\|y^K - z^K\| / \max(1, \|z^K\|) \le 0.01$ and $\|w^K - v^K\| / \max(1, \|v^K\|) \le 0.01$, respectively. More information about the test runs and time of Algorithm 1 is provided by the last two rows of Table 2. The average iterations and CPU time (in seconds) are less than 500 and 0.5 seconds, respectively, over all runs without failure.

Finally, in order to estimate the local behaviour of Algorithm 1 on our test runs, Figure 1 reports on the experimental order of convergence (EOC) defined by

$$\text{EOC} := \max \left\{ \frac{\log \|\Phi^\lambda(\zeta^{K-1})\|}{\log \|\Phi^\lambda(\zeta^{K-2})\|}, \frac{\log \|\Phi^\lambda(\zeta^K)\|}{\log \|\Phi^\lambda(\zeta^{K-1})\|} \right\}.$$

For example, if $\lambda = 2^{-1}$, Algorithm 1 solved 10 examples with EOC < 1.1, further 9 examples with $1 \le \text{EOC} < 1.5$, and 105 examples with EOC ≥ 1.5. Similar results were observed for other $\lambda \in \Lambda$.

More details on the numerical results discussed here can be found in the supplementary material [21].

Let us close this section with a final observation. It is clearly difficult to check whether a (test) problem fulfils the assumptions needed in Theorem 3.2 for stating necessary optimality conditions (or the assumptions used in Section 5 to guarantee the CD-regularity of Φ^λ at some point). Moreover, a great deal of the test problems may not satisfy one or more of those assumptions. Taking into account the remarkable performance of our approach on
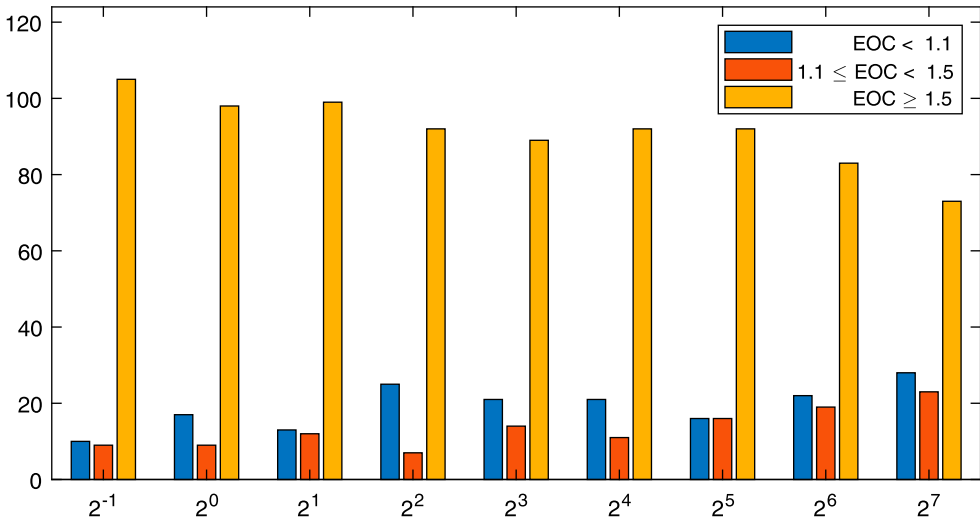
Figure 1. Experimental order of convergence (EOC) of Algorithm 1 for $\lambda \in \Lambda$.

the large test library, we believe that there is enough space for future research, particularly for a relaxation of those assumptions.

Disclosure statement

Funding

Notes on contributors

Andreas Fischer, after completing doctorate (1987) and habilitation (1998) at TU Dresden (Germany), Andreas became associate professor for Optimisation at TU Dortmund (Germany, 1998). Currently, he holds the Chair of Numerical Optimisation at TU Dresden (since 2002) and is director of the Institute of Numerical Mathematics (since 2013). His research interests focus on the design and analysis of efficient algorithms within the field of mathematical programming and nonsmooth equations. Moreover, Andreas is working in applications from engineering sciences with links to discrete optimisation, machine learning, and structural optimisation.

Alain B. Zemkoho is an Associate Professor at the School of Mathematical Sciences within the University of Southampton where he is affiliated to the Operational Research Group and CORMSIS. Prior to joining the University of Southampton, he was a Research Fellow at the University of Birmingham and had previously worked as a Research Associate at the Technical University of Freiberg. He is a Fellow of the Alan Turing Institute for Data Science and Artificial Intelligence, a Fellow of the Institute of Mathematics & Its Applications, and a Fellow of the UK Higher Eduction Academy. His research interests are around the theory and methods for nonconvex, nonsmooth, and bilevel optimization, as well as their applications, including in machine learning, security games, healthcare and medicine.

Shenglong Zhou is a Research Associate in the Department of EEE at Imperial College London. He was a Research Fellow from 2017 to 2020 and a Teaching Fellow from 2020 to 2021 at the University of Southampton. He received his B.Sc. Degree in Information and Computing Science in 2011 and M.Sc. Degree in Operational Research in 2014 both from Beijing Jiaotong University, China, and his Ph.D. degree in Operational Research in 2018 from the University of Southampton, UK. His research interests include the theory and methods of optimization in the fields of the Sparse, Low-rank Matrix, and Bilevel Optimization with applications into Communications, Signal Processing, and Machine Learning.

ORCID

Andreas Fischer 🔴 http://orcid.org/0000-0001-7703-1655
Alain B. Zemkoho 🔴 http://orcid.org/0000-0003-1265-4178

References

[1] A. Ben-Tal and J. Zowe, *Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems*, Math. Program. 24 (1982), pp. 70–91.
[2] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.
[3] F.H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.(originally published in 1983 by John Wiley & Sons).
[4] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Program. 75 (1996), pp. 407–439.
[5] S. Dempe, *A necessary and a sufficient optimality condition for bilevel programming problems*, Optimization 25 (1992), pp. 341–354.
[6] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht, 2002.
[7] S. Dempe and J. Dutta, *Is bilevel programming a special case of mathematical programming with equilibrium constraints?*, Math. Program. 131 (2012), pp. 37–48.
[8] S. Dempe and S. Franke, *Solution algorithm for an optimistic linear Stackelberg problem*, Comput. Oper. Res. 41 (2014), pp. 277–281.
[9] S. Dempe and S. Franke, *On the solution of convex bilevel optimization problems*, Comput. Optim. Appl. 63 (2016), pp. 685–703.
[10] S. Dempe and N. Gadhi, *Second order optimality conditions for bilevel set optimization problems*, J. Global Optim. 47 (2010), pp. 233–245.
[11] S. Dempe and A.B. Zemkoho, *The generalized Mangasarian-Fromowitz constraint qualification and optimality conditions for bilevel programs*, J. Optim. Theory Appl. 148 (2011), pp. 433–441.
[12] S. Dempe and A.B. Zemkoho, *The bilevel programming problem: reformulations, constraint qualifications and optimality conditions*, Math. Program. 138 (2013), pp. 447–473.
[13] S. Dempe, J. Dutta, and B.S. Mordukhovich, *New necessary optimality conditions in optimistic bilevel programming*, Optimization 56 (2007), pp. 577–604.
[14] F. Facchinei, A. Fischer, and C. Kanzow, *Regularity properties of a semismooth reformulation of variational inequalities*, SIAM J. Optim. 8 (1998), pp. 850–869.
[15] F. Facchinei, A. Fischer, and M. Herrich, *A family of Newton methods for nonsmooth constrained systems with nonisolated solutions*, Math. Methods Oper. Res. 77 (2013), pp. 433–443.
[16] A.V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York, 1983.
[17] A. Fischer, *A special Newton-type optimization method*, Optimization 24 (1992), pp. 269–284.
[18] A. Fischer and H. Jiang, *Merit functions for complementarity and related problems: A survey*, Comput. Optim. Appl. 17 (2000), pp. 159–182.
[19] A. Fischer, M. Herrich, A.F. Izmailov, and M.V. Solodov, *A globally convergent LP-Newton method*, SIAM J. Optim. 26 (2016), pp. 2012–2033.

[20] A. Fischer, M. Herrich, A.F. Izmailov, W. Scheck, and M.V. Solodov, *A globally convergent LP-Newton method for piecewise smooth constrained equations: escaping nonstationary accumulation points*, Comput. Optim. Appl. 69 (2018), pp. 325–349.

[21] A. Fischer, A.B. Zemkoho, and S. Zhou, *Detailed numerical experiments for "Semismooth Newton-type method for bilevel optimization: Global convergence and extensive numerical experiments"*, Tech. Rep., School of Mathematics, University of Southampton, Southampton, UK, 2019. Available at http://www.southampton.ac.uk/ abz1e14/papers/fzz-experiments-042021.pdf.

[22] M. Fischetti, I. Ljubic, M. Monaci, and M. Sinnl, *A new general-purpose algorithm for mixed-integer bilevel linear programs*, Oper. Res. 65 (2017), pp. 1615–1637.

[23] J. Gauvin and F. Dubeau, *Differential properties of the marginal function in mathematical programming*, in *Optimality and Stability in Mathematical Programming*, Mathematical Programming Studies 18, M. Guignard, ed., North-Holland, Amsterdam, 1982, pp. 101–119.

[24] C. Geiger and C. Kanzow, *On the resolution of monotone complementarity problems*, Comput. Optim. Appl. 5 (1996), pp. 155–173.

[25] E.G. Golšteĭn *Theory of Convex Programming*, American Mathematical Society, Providence, RI, 2008.

[26] R. Hettich and K.O. Kortanek, *Semi-infinite programming: theory, methods, and applications*, SIAM Rev. 35 (1993), pp. 380–429.

[27] A.F. Izmailov and A.S. Kurennoy, *On regularity conditions for complementarity problems*, Comput. Optim. Appl. 57 (2014), pp. 667–684.

[28] A.F. Izmailov, A.L. Pogosyan, and M.V. Solodov, *Semismooth newton method for the lifted reformulation of mathematical programs with complementarity constraints*, Comput. Optim. Appl. 51 (2012), pp. 199–221.

[29] H. Kleinmichel and K. Schönefeld, *Newton-type methods for nonlinearly constrained programming problems-algorithms and theory*, Optimization 19 (1988), pp. 397–412.

[30] M. Kojima and R. Hirabayashi, *Continuous deformation of nonlinear programs*, in *Sensitivity, Stability and Parametric Analysis*, Mathematical Programming Studies 21, A.V. Fiacco, ed., North-Holland, Amsterdam, 1984, pp. 150–198.

[31] B. Kummer, *Newton's method for non-differentiable functions*, in *Advances in Mathematical Optimization*, J. Guddat et al., eds., Akademie Verlag, Berlin, 1988, pp. 114–125.

[32] L. Lampariello and S. Sagratella, *A bridge between bilevel programs and Nash games*, J. Optim. Theory Appl. 174 (2017), pp. 613–635.

[33] J.M. Lee, *Introduction to Smooth Manifolds*, Springer, New York, 2001.

[34] J. Li, R. Huang, and J. Jian, *A superlinearly convergent qp-free algorithm for mathematical programs with equilibrium constraints*, Appl. Math. Comput. 269 (2015), pp. 885–903.

[35] G.-H. Lin, M. Xu, and J.J. Ye, *On solving simple bilevel programs with a nonconvex lower-level program*, Math. Program. 144 (2014), pp. 277–305.

[36] L. Lozano and J.C. Smith, *A value-function-based exact approach for the bilevel mixed-integer programming problem*, Oper. Res. 65 (2017), pp. 768–786.

[37] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim. 15 (1977), pp. 959–972.

[38] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation. I: Basic Theory. II: Applications*, Springer, Berlin, 2006.

[39] T.S. Munson, F. Facchinei, M.C. Ferris, A. Fischer, and C. Kanzow, *The semismooth algorithm for large scale complementarity problems*, INFORMS J. Comput. 13 (2001), pp. 294–311.

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

[41] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res. 18 (1993), pp. 227–244.

[42] L. Qi and H. Jiang, *Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations*, Math. Oper. Res. 22 (1997), pp. 301–325.

[43] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program. 58 (1993), pp. 353–367.

[44] L. Qi and D. Sun, *A survey of some nonsmooth equations and smoothing newton methods*, in *Progress in Optimization*, A. Eberhard et al., eds., Springer, Boston, 1999, pp. 121–146.

[45] S.M. Robinson, *Generalized equations and their solutions, part II: Applications to nonlinear programming*, in *Optimality and Stability in Mathematical Programming*, Mathematical Programming Studies 19, M. Guignard, ed., North-Holland, Amsterdam, 1982, pp. 200–221.

[46] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[47] R.T. Rockafellar, *Directional differentiability of the optimal value function in a nonlinear programming problem*, in *Sensitivity, Stability and Parametric Analysis*, Mathematical Programming Studies 21, A.V. Fiacco, ed., North-Holland, Amsterdam, 1984, pp. 213–226.

[48] A. Shapiro, *Second order sensitivity analysis and asymptotic theory of parametrized nonlinear programs*, Math. Program. 33 (1985), pp. 280–299.

[49] A. Shapiro, *Perturbation theory of nonlinear programs when the set of optimal solutions is not a singleton*, Appl. Math. Optim. 18 (1988), pp. 215–229.

[50] D. Sun and L. Qi, *On NCP-functions*, Comput. Optim. Appl. 13 (1999), pp. 201–220.

[51] S. Tahernejad, T.K. Ralphs, and S.T. DeNegre, *A branch-and-cut algorithm for mixed integer bilevel linear optimization problems and its implementation*, Math. Program. Comput. 12 (2020), pp. 529–568.

[52] T. Tanino and T. Ogawa, *An algorithm for solving two-level convex optimization problems*, Int. J. Systems Sci. 15 (1984), pp. 163–174.

[53] M. Xu and J.J. Ye, *A smoothing augmented lagrangian method for solving simple bilevel programs*, Comput. Optim. Appl. 59 (2014), pp. 353–377.

[54] M. Xu, J.J. Ye, and L. Zhang, *Smoothing sqp methods for solving degenerate nonsmooth constrained optimization problems with applications to bilevel programs*, SIAM J. Optim. 25 (2015), pp. 1388–1410.

[55] J.J. Ye and D.L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization 33 (1995), pp. 9–27. (with Erratum in Optimization, 39 (1997), pp. 361–366).

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

[57] A.B. Zemkoho, *Estimates of generalized Hessians for optimal value functions in mathematical programming*, 2020. Available at arXiv:1710.05887v2.

[58] S. Zhou, A.B. Zemkoho, and A. Tin, *BOLIB: Bilevel optimization library of test problems*, in *Bilevel Optimization: Advances and Next Challenges*, S. Dempe and A.B. Zemkoho, eds., Springer, Cham, 2020, pp. 563–580.