Pure and Applied Functional Analysis
Volume 8, Number 5, 2023, 1437–1463

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

A SEMISMOOTH NEWTON-TYPE METHOD FOR BILEVEL PROGRAMS WITH LINEAR LOWER LEVEL PROBLEM

FLORIANE MEFO KUE, THORSTEN RAASCH, AND ALAIN B. ZEMKOHO

Abstract. We consider a bilevel program involving a linear lower level problem with left-hand-side perturbation. We then consider the Karush-Kuhn-Tucker reformulation of the problem and subsequently build a tractable optimization problem with linear constraints by means of a partial exact penalization. A regularized Newton system of equations is then generated from the latter problem and a Newton-type method is developed to solve it. Finally, we illustrate the practical implementation of the algorithm on the optimal toll-setting problem in transportation networks.

1. Introduction

Bilevel programming problems are mathematical problems with a special constraint, which is implicitly determined by another optimization problem. This latter problem, called the follower's or lower level problem, is defined by

$$\min_{\mathbf{y}}\{f(\mathbf{x},\mathbf{y}) \mid g(\mathbf{x},\mathbf{y}) \leq 0\},$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^l$ represent the lower level objective and constraint functions, respectively. The bilevel programming problem can be formally described as

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{y}} && F(\mathbf{x},\mathbf{y}) \\ &\text{s.t.} && G(\mathbf{x}) \leq 0, \quad \mathbf{y} \in \Psi(\mathbf{x}) := \operatorname*{Argmin}_{\mathbf{y}} \{f(\mathbf{x},\mathbf{y}) \mid g(\mathbf{x},\mathbf{y}) \leq 0\}, \end{aligned} \tag{1.1}$$

where, similarly, the functions $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{R}^k$ correspond to the upper level objective and constraint functions, respectively. Problem (1.1) is called the upper level or leader's problem. It is often referred to as the optimistic formulation of the bilevel programming problem [12]. The variables $\mathbf{x}$ and $\mathbf{y}$ are the upper and lower level variables, respectively. In this order of ideas, our focus in this paper will be on the simple bilevel program

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{y}} && F(\mathbf{x},\mathbf{y}) \\ &\text{s.t.} && \mathbf{D}\mathbf{x} \leq \mathbf{d}, \quad \mathbf{y} \in \Psi(\mathbf{x}), \end{aligned} \tag{1.2}$$

2020 *Mathematics Subject Classification.* 90C29, 90C46, 46N10.
Key words and phrases. Bilevel optimization, Newton method, optimal toll-setting.

with the lower level optimal solution set-valued mapping Ψ defined by

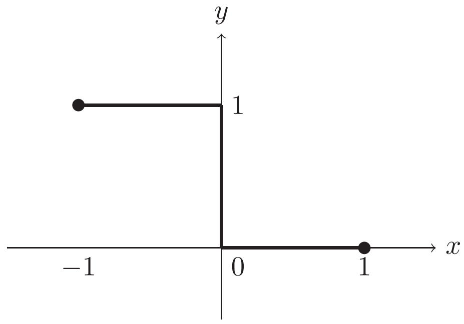
$$\Psi(\mathbf{x}) := \underset{\mathbf{y}}{\mathrm{Argmin}}\, \{\mathbf{x}^\top \mathbf{y} \mid \mathbf{A}\mathbf{y} \leq \mathbf{b}\}, \tag{1.3}$$

where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{d} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^l$ and all the remaining vectors and matrices are of appropriate dimensions. Note that the matrix $\mathbf{A}$ involved in (1.3) does not depend on the parameter $\mathbf{x}$ and the function F is assumed to be twice continuously differentiable. The term *simple bilevel program* for this class of problem was coined in the paper [30], referring to the fact that the upper (resp. lower) level feasible set is independent from $\mathbf{y}$ (resp. $\mathbf{x}$). However, it is important to mention that the same expression was initially used in [11] to name a completely different class of problem, that can be traced back to [31] and has been investigated in various publications; see, e.g., [4, 21, 42, 43].

It is clear that the lower level problem (1.3) is linear (w.r.t. $\mathbf{y}$), but with a left-hand-side perturbation. However, problem (1.2) is still a significantly difficult class of problem; to have a taste of this, consider the following example, where F is any real-valued function and $\mathbf{x}$ and $\mathbf{y}$ are one-dimensional:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & F(\mathbf{x}, \mathbf{y}) \\ \text{s.t.} \quad & \mathbf{x} \in [-1, 1], \quad \mathbf{y} \in \mathrm{Argmin}\, \{\mathbf{x}\mathbf{y} \mid \mathbf{y} \in [0, 1]\}. \end{aligned} \tag{1.4}$$

The feasible set of the problem is depicted in the following picture:



Clearly, the feasible set of problem (1.4), represented by the thick line segments, is the union of convex polyhedral sets; but it is not a convex set itself. Hence, finding a global optimal solution for problem (1.2)–(1.3) is generally not an easy task. This is why our focus in this paper will be on computing stationary points of the problem, which can potentially be locally optimal.

Note that if we replace the lower level problem (1.3) with the following one, where the perturbation is instead in the right-hand-side, we get a much easier optimization problem:

$$\Psi(\mathbf{x}) := \underset{\mathbf{y}}{\mathrm{Argmin}}\, \{\mathbf{c}^\top \mathbf{y} \mid \mathbf{A}\mathbf{y} \leq \mathbf{x}\}. \tag{1.5}$$

Problem (1.5) is *fully convex*, in the sense that the objective and constraint functions are convex in $(\mathbf{x}, \mathbf{y})$. Thanks to this, some specific algorithms work well for (1.2), (1.5), but cannot be implemented on (1.2), (1.3); cf. [12, Chapter 3]. Also, the optimal value function of problem (1.5) is convex and (1.2), (1.5) is partially calm [34]. The latter properties are critical in addressing problem (1.2), (1.5) theoretically and numerically. But unfortunately, they cannot hold for (1.2), (1.3). This probably justifies why problem (1.2), (1.5) has been at the center of attention in the development of solution algorithms for linear bilevel programs; see, e.g., [5] for the most recent literature review on the subject. Note that the approach to be developed in this paper can be applied to problem (1.2), (1.5) as well.

There are various practical bilevel programs with linear lower-level problems that are not covered by the model (1.2), (1.5). Among them, we have some network optimization problems that are commonly studied and used in operations research, such as the shortest path, transportation, minimal-cost network-flow, and toll optimization (also known as optimal toll-setting or network pricing) problems [3,19,29]. This last class of problem will be used in Section 5 to illustrate the practical implementation of the algorithm developed in this paper.

To tackle problem (1.2), (1.3), we consider its Karush-Kuhn-Tucker (KKT) reformulation, as the required lower level convexity and regularity are both automatically satisfied [13]. Then, using a certain tractable transformation process, we construct a partial exact penalization of problem (1.2), (1.3) and show its close link with the KKT reformulation; cf. Section 3. In the context of this penalized problem, we investigate a local convergence method (see, e.g., [20, 36, 39]) in Section 4, where sufficient conditions ensuring its convergence are established and developped a smoothed-regularized Newton method. The method is then implemented on the toll optimization problem in Section 5.

It is important to recall that Newton-type methods based on the KKT or value function reformulations of general versions of the bilevel optimization problem (1.2), (1.3) have been studied recently in [24, 25, 47]. These works typically require the partial calmness property, which does not necessarily hold for problem (1.2), (1.3); cf. [34]. Moreover, for this problem specifically, very little work tailored to it has been done in the literature. We are aware of the paper [15], where an algorithm based on the optimal value reformulation to compute local and global optimal solutions is derived. In [16] the authors propose an approach to solve (1.2), (1.3) in the discrete case through the optimal value reformulation approach as well. For a more general version of the problem, the authors in [30] design smoothing projected gradient algorithm for simple bilevel programs with a nonconvex lower level program using the optimal value reformulation approach. The main difference between those approaches and the one that we are proposing is on the reformulation used to replace the original problem into a single-level optimization problem. In fact, while we use the KKT reformulation, the other works apply the concept of partial calmness through the optimal value reformulation to derive necessary optimality conditions of (1.2), (1.3). A comparison between both reformulations can be found in [47].

Before, we go deep in the analysis as described above, we start with some preliminary elements in the next section.

2. Preliminaries

Here, we introduce some preliminary concepts that will be used later in this paper. First, note that a function $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ is said to be locally Lipschitz continuous around $\mathbf{x}^0 \in \mathbb{R}^n$ if there exist $\alpha > 0$ and $L > 0$ such that the following condition holds:

$$\forall \mathbf{x}, \mathbf{x}' \in \mathbf{x}^0 + \alpha \mathbb{B}_{\mathbb{R}^n}: \quad \|\Phi(\mathbf{x}) - \Phi(\mathbf{x}')\|_{\mathbb{R}^m} \leq L \|\mathbf{x} - \mathbf{x}'\|_{\mathbb{R}^n}.$$

We denote by $\mathbb{B}_{\mathbb{R}^n}$ the closed unit ball in $\mathbb{R}^n$ and the number L is called the Lipschitz constant. Φ will be said to be locally Lipschitz continuous if it is locally Lipschitz continuous around every point of $\mathbb{R}^n$. It is said to be Lipschitz continuous if the above inequality holds with $\alpha = \infty$. Recall that any convex function is locally Lipschitz continuous on the relative interior of its domain.

Consider a locally Lipschitz continuous function $\Phi : \mathbb{R}^n \to \mathbb{R}^m$. According to Rademacher's theorem [7], Φ is differentiable almost everywhere. Let D_Φ denote the set of all the points where Φ is differentiable. Then, we can define the Bouligand subdifferential of Φ at $\mathbf{x} \in \mathbb{R}^n$ by

$$\partial_B \Phi(\mathbf{x}) := \{\mathbf{C} \in \mathbb{R}^{m \times n} \mid \mathbf{C} = \lim_{k \to \infty} \nabla \Phi(\mathbf{x}_k),\ \mathbf{x}_k \to \mathbf{x},\ \mathbf{x}_k \in D_\Phi\}.$$

The generalised Jacobian of Φ in the sense of Clarke [7] is given by

$$\partial \Phi(\mathbf{x}) := \operatorname{conv} \partial_B \Phi(\mathbf{x}),$$

where *conv* stands for the convex hull. The postulated local Lipschitz continuity property of Φ around $\mathbf{x}$ guarantees that the set $\partial \Phi(\mathbf{x})$ is nonempty and compact. If Φ is a convex function, then $\partial \Phi(\mathbf{x})$ coincides with the subdifferential in the sense of convex analysis.

Next we introduce the notion of semismoothness introduced in [35] and extended to vector-valued functions in [37, 39]. A function $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ is said to be semismooth at a given point $\mathbf{x}$ if Φ is directional differentiable at $\mathbf{x}$ and for any $\mathbf{C} \in \partial \Phi(\mathbf{x} + \mathbf{h}), \mathbf{h} \to 0$

$$\Phi(\mathbf{x} + \mathbf{h}) - \Phi(\mathbf{x}) - \mathbf{C}\mathbf{h} = o(\|\mathbf{h}\|).$$

The function Φ is said to be strongly semismooth at $\mathbf{x}$ if Φ is semismooth at $\mathbf{x}$ and it holds that

$$\Phi(\mathbf{x} + \mathbf{h}) - \Phi(\mathbf{x}) - \mathbf{C}\mathbf{h} = O(\|\mathbf{h}\|^2),\ \forall \mathbf{C} \in \partial \Phi(\mathbf{x} + \mathbf{h}), \mathbf{h} \to 0.$$

Any continuously differentiable function is obviously semismooth. But this is not necessarily the case for a locally Lipschitz continuous function [35]. However, piecewise linear functions are strongly semismooth and a special case from such a class, that will be of a particular interest in this paper is $t \mapsto \max(0, t)$. Its Clarke subdifferential can be obtained as

$$v \in \partial \max(0, t) \iff v \in \begin{cases} \{1\} & \text{if } t > 0, \\ \{0\} & \text{if } t < 0, \\ [0, 1] & \text{if } t = 0. \end{cases}$$

Throughout the paper, we use $\mathbb{R}^n_+$ and $\mathbb{R}^{n \times m}$ to denote the cone of n-dimensional real-valued vectors with nonnegative components and the set of real-valued matrices with n rows and m columns, respectively. Furthermore we denote by $\mathbf{0}$, $\mathbf{I}$, and $\mathbf{e}$ the null matrix, the identity matrix and the vector of ones of appropriate dimension, respectively. Boldface lower-case letters represent vectors and boldface upper-case letters represent a matrix.

3. Single-level reformulation

Considering problem (1.2), (1.3), it is obvious that its lower level problem is both convex and linear w.r.t. the lower level variable $\mathbf{y}$. Hence, the KKT reformulation of the problem can be written as follows without any additional requirement:

$$\begin{array}{lll} & \min\limits_{\mathbf{x},\mathbf{y},\mathbf{z}} & F(\mathbf{x},\mathbf{y}) \\ (3.1) & \text{s.t.} & \mathbf{D}\mathbf{x} \le \mathbf{d},\ \ \mathbf{A}^\top\mathbf{z} + \mathbf{x} = 0, \\ & & \mathbf{A}\mathbf{y} \le \mathbf{b},\ \ \mathbf{z} \ge 0,\ \ \mathbf{z}^\top(\mathbf{A}\mathbf{y} - \mathbf{b}) = 0. \end{array}$$

Reference [13] provides a detailed analysis of the relationship between this reformulation and the original problem (1.2), (1.3). One of the main issues in solving problem (3.1) with standard continuous optimization techniques is the presence of the complementarity conditions

$$(3.2) \qquad (\mathbf{b} - \mathbf{A}\mathbf{y})_i \ge 0,\ \ \mathbf{z}_i \ge 0,\ \ \mathbf{z}_i\,(\mathbf{b} - \mathbf{A}\mathbf{y})_i = 0 \quad \text{for} \quad i = 1, \dots, l$$

in the feasible set, which causes the failure of well-known constraint qualifications (see, e.g., [18]). To deal with this issue here, we start by considering the reformulation

$$(3.3) \qquad \min_{(\mathbf{r}_i,\mathbf{s}_i)\in T_i} \mathbf{r}_i\mathbf{z}_i + \mathbf{s}_i(\mathbf{b} - \mathbf{A}\mathbf{y})_i = 0 \quad \text{for} \quad i = 1, \dots, l,$$

of the complementarity conditions (3.2), where, for $i = 1, \dots, l$, the set T_i is defined by

$$(3.4) \qquad T_i := \{(\mathbf{r}_i, \mathbf{s}_i) \in \mathbb{R}^2 \mid \mathbf{r}_i \ge 0,\ \ \mathbf{s}_i \ge 0,\ \ \mathbf{r}_i + \mathbf{s}_i = 1\}.$$

Note that the transformation (3.3) obviously follows in an equivalent way from (3.2), since (3.2) is identical to $\min\,(\mathbf{z}_i,\,(\mathbf{b} - \mathbf{A}\mathbf{y})_i) = 0$ for $i = 1, \dots, l$.

Based on (3.3), we consider the following penalization of (3.1):

$$(\mathrm{P}_\alpha) \qquad \min_{\mathbf{x},\mathbf{y},\mathbf{z}} F(\mathbf{x},\mathbf{y}) + \alpha\pi(\mathbf{y},\mathbf{z}) \quad \text{s.t.} \quad (\mathbf{x},\mathbf{y},\mathbf{z}) \in Z_1,$$

where $\alpha > 0$ represents the penalization parameter and the function π and set Z_1 are respectively defined as follows:

$$\begin{aligned} \pi(\mathbf{y},\mathbf{z}) &:= \textstyle\sum_{i=1}^{l} \min\limits_{(\mathbf{r}_i,\mathbf{s}_i)\in T_i} \mathbf{r}_i\mathbf{z}_i + \mathbf{s}_i(\mathbf{b} - \mathbf{A}\mathbf{y})_i \ge 0, \\ Z_1 &:= \{(\mathbf{x},\mathbf{y},\mathbf{z}) \mid \mathbf{D}\mathbf{x} \le \mathbf{d},\ \mathbf{A}\mathbf{y} \le \mathbf{b},\ \mathbf{A}^\top\mathbf{z} + \mathbf{x} = 0, \mathbf{z} \ge 0\}. \end{aligned}$$

To establish the relationship between (P_α) and (3.1), let $T := \prod_{i=1}^{l} T_i$ and let Z_1^* be the set of vertices of the polyhedra Z_1.

We suppose in the rest of the paper, unless stated otherwise, that there exists at least one optimal solution of the bilevel optimization problem which belongs to the set Z_1^*. We then have the following result inspired by [32, Proposition 3.1].

Proposition 3.1. *Assume that for all $\alpha > 0$, problem (P_α) possesses an optimal solution which belongs to Z_1^*. Then there exists a scalar $\bar{\alpha} > 0$ such that for all $\alpha \ge \bar{\alpha}$ the following two statements are valid:*

(i) *Any global optimal solution $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ of problem (P_α) which belongs to Z_1^* solves problem (3.1) globally.*

(ii) *Any global optimal solution* $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ *of problem* (3.1) *optimally solves problem* (P_α) *globally.*

Proof. For (i), we start by noting that any global solution $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ of (P_α) which belongs to Z_1^* and satisfy $\pi(\bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha) = 0$ solves (3.1) globally. In fact let $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ be a feasible point of (3.1). We have

$$F(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}, \mathbf{y}) + \alpha\pi(\mathbf{y}, \mathbf{z}) \geq F(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha) + \alpha\pi(\bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha) = F(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha).$$

Hence it suffices to show that for all sufficiently large α, every global solution $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ of (P_α) which is an extreme point of Z_1 satisfies $\pi(\bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha) = 0$. We suppose that this is not true. Then there exist sequences $\{\alpha_k\} \to \infty$ and $\{X_k := (\mathbf{x}_k, \mathbf{y}_k, \mathbf{z}_k)\}_k \subseteq Z_1^*$ where X_k is optimal solution of (P_α) and $\pi(\mathbf{y}_k, \mathbf{z}_k) > 0$. Due to the fact that the set of extreme points Z_1^* is finite we have

$$\beta := \inf_k \pi(\mathbf{y}_k, \mathbf{z}_k) > 0 \quad \text{and} \quad \rho := \inf_k F(\mathbf{x}_k, \mathbf{y}_k) > -\infty.$$

Then for any feasible point $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ of (3.1) it holds that

$$F(\mathbf{x}, \mathbf{y}) \geq F(\mathbf{x}_k, \mathbf{y}_k) + \alpha_k\pi(\mathbf{y}_k, \mathbf{z}_k) \geq \rho + \alpha_k\beta.$$

Therefore, for $k \to \infty$ we get $F(\mathbf{x}, \mathbf{y}) = \infty$ and this is a contradiction. Consequently there exists a scalar $\bar{\alpha} > 0$ such that for all $\alpha \geq \bar{\alpha}$ any optimal solution of (P_α) which belongs to Z_1^* solves (3.1).

As for (ii), let $\bar{\alpha} > 0$ be the penalization parameter whose existence has been proved in the first assertion of this proof. Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ be a global optimal solution of (3.1). Let $\alpha \geq \bar{\alpha}$ and $\bar{X}_\alpha := (\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ be a global optimal solution of (P_α) which belongs to Z_1^* (such point exists from our assumption). Then $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ solves (3.1) from the first-proved assertion. We want to show that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ solves (P_α). For any feasible solution $X_\alpha = (\mathbf{x}_\alpha, \mathbf{y}_\alpha, \mathbf{z}_\alpha)$ of (P_α), for $\alpha \geq \bar{\alpha}$,

$$\begin{aligned}
F(\mathbf{x}_\alpha, \mathbf{y}_\alpha) + \alpha\pi(\mathbf{y}_\alpha, \mathbf{z}_\alpha) &\geq F(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha) + \alpha\pi(\bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha) \\
&= F(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha) \\
&= F(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \\
&= F(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + \alpha\pi(\bar{\mathbf{y}}, \bar{\mathbf{z}}),
\end{aligned}$$

where the first and second equality are due to the fact that on the first hand $\bar{X}_\alpha$ belongs to Z_1^* (see also the first assertion) and on the second hand $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ and $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ solve (3.1). Therefore the point $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ solves (P_α). □

Remark 3.2. *One way to ensure the existence of solution of* (P_α) *which belongs to* Z_1^* *is to assume that the function* F *is concave, the set* Z_1 *is nonempty and* F *is bounded below on* Z_1*. In fact since the set* Z_1 *is a polyhedral containing no lines (it is not difficult to verify it since* Z_1 *is supposed to have at least one extreme point) and that for all* α*, the mapping* $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \mapsto F(\mathbf{x}, \mathbf{y}) + \alpha\pi(\mathbf{y}, \mathbf{z})$ *is concave, we conclude using Corollary 32.3.4 in [41] or [1, Theorem 3.4.7] that* (P_α) *is well defined and there exists an optimal solution which belongs to* Z_1^**.*

Proposition 3.3. *Assume that for all* $\alpha > 0$*, problem* (P_α) *possesses an optimal solution which belongs to* Z_1^**. Then there exists a scalar* $\bar{\alpha} > 0$ *such that for all* $\alpha \geq \bar{\alpha}$*, the following two assertions hold:*

(i) *Any local optimal solution* $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ *of* (P_α) *which belongs to* Z_1^* *solves* (3.1) *locally.*

(ii) *Any local optimal solution* $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ *of problem* (3.1) *solves* (P_α) *locally if from any sequence,* $(\bar{\mathbf{x}}_\alpha^n, \bar{\mathbf{y}}_\alpha^n, \bar{\mathbf{z}}_\alpha^n)$*, of feasible points of problem* (P_α) *that converges to* $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$*, there exists a sequence* $(\tilde{\mathbf{y}}_\alpha^n, \tilde{\mathbf{z}}_\alpha^n)$ *converging to* $(\bar{\mathbf{y}}, \bar{\mathbf{z}})$ *such that* $(\bar{\mathbf{x}}_\alpha^n, \tilde{\mathbf{y}}_\alpha^n, \tilde{\mathbf{z}}_\alpha^n)$ *is feasible for* (3.1) *and* $F(\bar{\mathbf{x}}_\alpha^n, \bar{\mathbf{y}}_\alpha^n) = F(\bar{\mathbf{x}}_\alpha^n, \tilde{\mathbf{y}}_\alpha^n)$.

Proof. Let $\bar{\alpha} > 0$ as given in the first part of the proof of Proposition 3.1.

For (i), let $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ be a given local solution of (P_α). Then $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ is also a local solution of (3.1). In fact we suppose that it is not true. Then there exists a sequence of feasible points $(\bar{\mathbf{x}}^n, \bar{\mathbf{y}}^n, \bar{\mathbf{z}}^n)_n$ for (3.1) converging to $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ such that $F(\bar{\mathbf{x}}^n, \bar{\mathbf{y}}^n) < F(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha)$ for all $n \in \mathbb{N}$. Since $(\bar{\mathbf{x}}^n, \bar{\mathbf{y}}^n, \bar{\mathbf{z}}^n)$ is feasible for (3.1), we get $\pi(\bar{\mathbf{y}}^n, \bar{\mathbf{z}}^n) = 0$. Therefore, we have the following inequalities, which lead to a contradiction:

$$F(\bar{\mathbf{x}}^n, \bar{\mathbf{y}}^n) + \alpha\pi(\bar{\mathbf{y}}^n, \bar{\mathbf{z}}^n) = F(\bar{\mathbf{x}}^n, \bar{\mathbf{y}}^n) < F(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha) \leq F(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha) + \alpha\pi(\bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha).$$

About the feasibility of $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ for (3.1), we can show (similarly as in the proof of Proposition 3.1) that for all sufficiently large α, every local solution $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ of (P_α) which is an extreme point of Z_1 satisfies $\pi(\bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha) = 0$.

In the case of (ii), let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ be a local optimal solution of (3.1). We suppose by contradiction that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is not a local optimal solution of (P_α). Then there exists a sequence of feasible points $(\bar{\mathbf{x}}_\alpha^n, \bar{\mathbf{y}}_\alpha^n, \bar{\mathbf{z}}_\alpha^n)_n$ of (P_α) converging to $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ such that

$$F(\bar{\mathbf{x}}_\alpha^n, \bar{\mathbf{y}}_\alpha^n) + \alpha\pi(\bar{\mathbf{y}}_\alpha^n, \bar{\mathbf{z}}_\alpha^n) < F(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + \alpha\pi(\bar{\mathbf{y}}, \bar{\mathbf{z}}), \ \forall n \in \mathbb{N}.$$

Since $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is feasible for (3.1), we get $\pi(\bar{\mathbf{y}}, \bar{\mathbf{z}}) = 0$. We can then deduce that $F(\bar{\mathbf{x}}_\alpha^n, \bar{\mathbf{y}}_\alpha^n) < F(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ because $\pi(\bar{\mathbf{y}}_\alpha^n, \bar{\mathbf{z}}_\alpha^n) \geq 0$. Moreover, we can find a sequence $(\tilde{\mathbf{y}}_\alpha^n, \tilde{\mathbf{z}}_\alpha^n)$ converging to $(\bar{\mathbf{y}}, \bar{\mathbf{z}})$ such that $(\bar{\mathbf{x}}_\alpha^n, \tilde{\mathbf{y}}_\alpha^n, \tilde{\mathbf{z}}_\alpha^n)$ is feasible for (3.1) and

$$F(\bar{\mathbf{x}}_\alpha^n, \bar{\mathbf{y}}_\alpha^n) = F(\bar{\mathbf{x}}_\alpha^n, \tilde{\mathbf{y}}_\alpha^n).$$

This contradicts the fact that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is a local optimal solution of (3.1).

□

Below, we present an example where the assumptions in the second part of Theorem 3.3 are satisfied.

Example 3.4. *Consider the bilevel program*

$$(3.5) \qquad \begin{aligned} &\min_{x,y} && -x^2 + 1 \\ &s.t. && y \in \Psi(x) = \underset{y}{\mathit{Argmin}}\, \{xy \mid 0 \leq y \leq 1\}. \end{aligned}$$

An optimal solution of the related KKT reformulation is given by $(0, 1, 0, 0)$ *and we can observe that for any sequence* $(\bar{\mathbf{x}}_\alpha^n, 1, (\bar{\mathbf{z}}_\alpha)_1^n, (\bar{\mathbf{z}}_\alpha)_2^n)$ *of feasible points of the related* (P_α) *converging to* $(0, 1, 0, 0)$*, the sequence* $(\bar{\mathbf{x}}_\alpha^n, 1, 0, (\bar{\mathbf{z}}_\alpha)_1^n - \bar{\mathbf{x}}_\alpha^n)$ *is feasible for the the KKT reformulation and converges to* $(0, 1, 0, 0)$ *as well. We have obviously* $F(\bar{\mathbf{x}}_\alpha^n, 1) = F(\bar{\mathbf{x}}_\alpha^n, \bar{\mathbf{y}}_\alpha^n)$.

In order to render (P_α) more tractable, we use the following result to get rid of the *min* operator in π, appearing in its objective function.

Proposition 3.5. *For any $\alpha > 0$, problem* (P_α) *is globally equivalent to*

$$(\mathrm{Q}_\alpha) \qquad \min_{\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{r},\mathbf{s}} F(\mathbf{x},\mathbf{y}) + \alpha(\mathbf{r}^\top \mathbf{z} + \mathbf{s}^\top(\mathbf{b} - \mathbf{A}\mathbf{y})) \ \ s.t. \ \ (\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{r},\mathbf{s}) \in Z_1 \times T$$

and the following assertions hold:

(i) *From any local optimal solution* $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ *of problem* (P_α)*, the point* $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha, \bar{\mathbf{r}}_\alpha, \bar{\mathbf{s}}_\alpha)$*, with the component* $(\bar{\mathbf{r}}_\alpha, \bar{\mathbf{s}}_\alpha) \in$ *Argmin* $\pi(\bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$*, is local optimal solution of* (Q_α) *as well.*

(ii) *From any local optimal solution* $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha, \bar{\mathbf{r}}_\alpha, \bar{\mathbf{s}}_\alpha)$ *of problem* (Q_α) *with* $(\bar{\mathbf{r}}_\alpha, \bar{\mathbf{s}}_\alpha) \in$ *Argmin* $\pi(\mathbf{y}, \mathbf{z})$ *for all* $(\mathbf{x},\mathbf{y},\mathbf{z}) \in Z_1$*, the point* $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ *is a local solution of problem* (P_α)*.*

Proof. The initial statement on the global equivalence between problems (P_α) and (Q_α) is obvious. As for (i), let $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ be a local optimal solution of problem (P_α). We want to show that $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha, \bar{\mathbf{r}}_\alpha, \bar{\mathbf{s}}_\alpha)$ with $(\bar{\mathbf{r}}_\alpha, \bar{\mathbf{s}}_\alpha) \in \mathrm{Argmin}\ \pi(\bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ is also a local optimal solution of problem (Q_α). We suppose by contradiction that it is not true. Therefore, there exists a sequence of feasible points $(\mathbf{x}^n_\alpha, \mathbf{y}^n_\alpha, \mathbf{z}^n_\alpha, \mathbf{r}^n_\alpha, \mathbf{s}^n_\alpha)_n$ for (Q_α) converging to the point $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha, \bar{\mathbf{r}}_\alpha, \bar{\mathbf{s}}_\alpha)$ such that for all $n \in \mathbb{N}$,

$$F(\mathbf{x}^n_\alpha, \mathbf{y}^n_\alpha) + \alpha((\mathbf{r}^n_\alpha)^\top \mathbf{z}^n_\alpha + (\mathbf{s}^n_\alpha)^\top(\mathbf{b} - \mathbf{A}\mathbf{y}^n_\alpha)) < F(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha) + \alpha(\bar{\mathbf{r}}_\alpha^\top \bar{\mathbf{z}}_\alpha + \bar{\mathbf{s}}_\alpha^\top(\mathbf{b} - \mathbf{A}\bar{\mathbf{y}}_\alpha)).$$

It is clear that for all $n \in \mathbb{N}$, $(\mathbf{x}^n_\alpha, \mathbf{y}^n_\alpha, \mathbf{z}^n_\alpha)$ is feasible for (P_α) and we get

$$\begin{aligned} F(\mathbf{x}^n_\alpha, \mathbf{y}^n_\alpha) + \alpha\pi(\mathbf{y}^n_\alpha, \mathbf{z}^n_\alpha) &\le F(\mathbf{x}^n_\alpha, \mathbf{y}^n_\alpha) + \alpha((\mathbf{r}^n_\alpha)^\top \mathbf{z}^n_\alpha + (\mathbf{s}^n_\alpha)^\top(\mathbf{b} - \mathbf{A}\mathbf{y}^n_\alpha)) \\ &< F(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha) + \alpha(\bar{\mathbf{r}}_\alpha^\top \bar{\mathbf{z}}_\alpha + \bar{\mathbf{s}}_\alpha^\top(\mathbf{b} - \mathbf{A}\bar{\mathbf{y}}_\alpha)) \\ &= F(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha) + \alpha\pi(\bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha). \end{aligned}$$

Then the point $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ does not solves (P_α) locally and this is absurd.

As for (ii), let $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha, \bar{\mathbf{r}}_\alpha, \bar{\mathbf{s}}_\alpha)$ be a local optimal solution of problem (Q_α). We show that $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ is a local optimal solution of problem (P_α) by making use of the assumption and by observing that from any feasible point $(\mathbf{x}_\alpha, \mathbf{y}_\alpha, \mathbf{z}_\alpha)$ of (P_α) in a neighbourhood $\mathcal{V}$ of $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$, the point $(\mathbf{x}_\alpha, \mathbf{y}_\alpha, \mathbf{z}_\alpha, \bar{\mathbf{r}}_\alpha, \bar{\mathbf{s}}_\alpha)$ is feasible for problem (Q_α) and belongs to $\mathcal{V}$.

□

In the sequel we give an example, where the assumption in (ii) holds.

Example 3.6. *We examine the following bilevel program*

$$\begin{aligned} \min_{x,y} \quad & -x^2 - y \\ s.t. \quad & x \ge 0,\ y \in \Psi(x) = \mathop{Argmin}_y \{xy \mid y \le 1\}. \end{aligned}$$

The penalized reformulation is given by

$$\begin{aligned} \min_{x,y,z} \quad & -x^2 - y + \alpha(\min_{r,s}\{rz + s(1-y) \mid r + s = 1, r, s \in [0,1]\}) \\ s.t. \quad & x \ge 0, x + z = 0, y \le 1, z \ge 0. \end{aligned}$$

The corresponding problem (Q_α) *is then*

$$\begin{aligned} \min_{x,y,z,r,s} \quad & -x^2 - y + \alpha(rz + s(1-y)) \\ s.t. \quad & x \ge 0, x + z = 0, y \le 1, z \ge 0 \\ & r + s = 1, r, s \in [0,1] \end{aligned}$$

An optimal solution of the latter problem is $(0, 1, 0, 1, 0)$ *and we can easily check that* $(1, 0) \in \mathit{Argmin}\, \pi(y, z)$ *for any* $(x, y, z) \in Z_1$.

Combining Propositions 3.1, 3.3 as well as Proposition 3.5 it follows clearly that (3.1) is globally equivalent to (Q_α) if a global optimal solution $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ of problem (P_α) belongs to Z_1^*. With respect to the local solutions, we obtain the following proposition.

Proposition 3.7. *Assume that for all* $\alpha > 0$, *problem* (P_α) *possesses an optimal solution which belongs to* Z_1^*. *There exists a scalar* $\bar{\alpha} > 0$ *such that for all* $\alpha \geq \bar{\alpha}$ *the following two assertions hold:*

(i) *From any local optimal solution* $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha, \bar{\mathbf{r}}_\alpha, \bar{\mathbf{s}}_\alpha)$ *of* (Q_α), $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha)$ *solves* (3.1) *locally if* $(\bar{\mathbf{x}}_\alpha, \bar{\mathbf{y}}_\alpha, \bar{\mathbf{z}}_\alpha) \in Z_1^*$ *and* $(\bar{\mathbf{r}}_\alpha, \bar{\mathbf{s}}_\alpha) \in \mathit{Argmin}\, \pi(\mathbf{y}, \mathbf{z})$ *for all* $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in Z_1$.

(ii) *Suppose that* $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ *is local optimal solution of* (3.1), *then* $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}, \bar{\mathbf{r}}, \bar{\mathbf{s}})$ *with* $(\bar{\mathbf{r}}, \bar{\mathbf{s}}) \in \mathit{Argmin}\, \pi(\bar{\mathbf{y}}, \bar{\mathbf{z}})$ *solves* (Q_α) *locally if from any sequence of feasible points of* (P_α), $(\bar{\mathbf{x}}_\alpha^n, \bar{\mathbf{y}}_\alpha^n, \bar{\mathbf{z}}_\alpha^n)$ *converging to* $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$, *there exists a sequence* $(\tilde{\mathbf{y}}_\alpha^n, \tilde{\mathbf{z}}_\alpha^n)$ *converging to* $(\bar{\mathbf{y}}, \bar{\mathbf{z}})$ *such that* $(\bar{\mathbf{x}}_\alpha^n, \tilde{\mathbf{y}}_\alpha^n, \tilde{\mathbf{z}}_\alpha^n)$ *is feasible for* (3.1) *and* $F(\bar{\mathbf{x}}_\alpha^n, \bar{\mathbf{y}}_\alpha^n) = F(\bar{\mathbf{x}}_\alpha^n, \tilde{\mathbf{y}}_\alpha^n)$.

To close this section, note that (3.3) is not the only way to reformulate the complementarity conditions (3.2). Various other transformations are possible; see [26] for a large number of such functions and related properties. However, unlike most reformulations that can be found in the latter reference, (3.3) enables problem (P_α) to be transformed into the smooth optimization problem (Q_α).

It is possible to use the bilinear function $(\mathbf{y}, \mathbf{z}) \longmapsto \mathbf{z}^\top(\mathbf{A}\mathbf{y} - \mathbf{b})$ as penalization term in (P_α), see, e.g., [44, 47]. However, its utilization requires the fulfilment of the partial calmness condition, introduced in [45], which unfortunately does not necessarily hold for (1.2), (1.3).

It is also worth mentioning that exact penalization has been widely used in the context of bilevel programs with lower level problem of the form (1.5). But it is unclear whether the corresponding results are applicable to (1.2), (1.3). The most recent overview of the topic can be found in **[14]**.

4. The semismooth Newton-type method

Based on the relationship between (3.1) and (Q_α), we construct a framework in this section to compute the stationary points of the latter problem using a regularized Newton method. To begin the process, we write the necessary optimality conditions for (Q_α) by means of the standard Lagrange multipliers rule for smooth optimization problems. As the feasible set of this problem is only described by linear constraints, no constraint qualification is needed.

Theorem 4.1. *Let* $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}, \mathbf{s}) \in \mathbb{R}^{2n+3l}$ *be a local optimal solution of problem* (Q_α) *for a fixed value of* $\alpha > 0$. *Then, there exist* $\boldsymbol{\lambda}_1 \in \mathbb{R}_+^m$, $\boldsymbol{\lambda}_6 \in \mathbb{R}^n$, $\boldsymbol{\lambda}_i \in \mathbb{R}_+^l$,

$i = 2, 3, 4, 5$, $\boldsymbol{\lambda}_7 \in \mathbb{R}^l$ such that

$$0 = \nabla_{\mathbf{x}} F(\mathbf{x}, \mathbf{y}) + \mathbf{D}^\top \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_6, \tag{4.1a}$$
$$0 = \nabla_{\mathbf{y}} F(\mathbf{x}, \mathbf{y}) - \alpha \mathbf{A}^\top \mathbf{s} + \mathbf{A}^\top \boldsymbol{\lambda}_2, \tag{4.1b}$$
$$0 = \alpha \mathbf{r} + \mathbf{A} \boldsymbol{\lambda}_6 - \boldsymbol{\lambda}_3, \tag{4.1c}$$
$$0 = \alpha \mathbf{z} + \boldsymbol{\lambda}_7 - \boldsymbol{\lambda}_4, \tag{4.1d}$$
$$0 = \alpha(\mathbf{b} - \mathbf{A}\mathbf{y}) + \boldsymbol{\lambda}_7 - \boldsymbol{\lambda}_5, \tag{4.1e}$$
$$0 = \mathbf{A}^\top \mathbf{z} + \mathbf{x}, \tag{4.1f}$$
$$0 = \mathbf{r} + \mathbf{s} - \mathbf{e}, \tag{4.1g}$$
$$0 = \boldsymbol{\lambda}_1^\top (\mathbf{D}\mathbf{x} - \mathbf{d}),\ \mathbf{D}\mathbf{x} \leq \mathbf{d},\ \boldsymbol{\lambda}_1 \geq 0, \tag{4.1h}$$
$$0 = \boldsymbol{\lambda}_2^\top (\mathbf{b} - \mathbf{A}\mathbf{y}), \mathbf{b} - \mathbf{A}\mathbf{y} \geq 0,\ \boldsymbol{\lambda}_2 \geq 0, \tag{4.1i}$$
$$0 = \boldsymbol{\lambda}_3^\top \mathbf{z},\ \boldsymbol{\lambda}_3 \geq 0,\ \mathbf{z} \geq 0, \tag{4.1j}$$
$$0 = \boldsymbol{\lambda}_4^\top \mathbf{r},\ \boldsymbol{\lambda}_4 \geq 0,\ \mathbf{r} \geq 0, \tag{4.1k}$$
$$0 = \boldsymbol{\lambda}_5^\top \mathbf{s},\ \boldsymbol{\lambda}_5 \geq 0,\ \mathbf{s} \geq 0. \tag{4.1l}$$

Next, we transform these conditions into a complete system of equations using the following lemma.

Lemma 4.2 (see, e.g., [28]). *For the vectors $\mathbf{y}, \boldsymbol{\lambda} \in \mathbb{R}^n$, the system of complementarity conditions $\mathbf{y} \leq 0$, $\boldsymbol{\lambda} \geq 0$, $\boldsymbol{\lambda}^\top \mathbf{y} = 0$ is equivalent to* $\max\{0, \boldsymbol{\lambda} + t\mathbf{y}\} - \boldsymbol{\lambda} = 0$ *for any $t > 0$.*

In the process of computing points satisfying (4.1a)-(4.1l), the reformulation of the complementary conditions in Lemma 4.2 has a number of advantages. At first, as it will be clear in the sequel (see, e.g., Theorem 4.10), the possibility to choose any value of $t > 0$ provides a level of freedom and flexibility, which can be crucial in solving the resulting system. Also, according to [23], reformulating the complementarity conditions with linear functions by means of the Lemma 4.2 can enable the Newton-type method to exhibit the finite termination property while equally demanding slightly weaker assumptions for superlinear convergence. Furthermore, Lemma 4.2 can enable the application of the Newton method, to be studied here, to problem in infinite dimensions (e.g., bilevel optimal control problems attracting more and more attention [2]), through slant differentiability that generalizes the notion of semismoothness [6, 28].

To solve (4.1a)-(4.1l) in the case where the upper level objective function F is affine linear one, that is, if $F(\mathbf{x}, \mathbf{y}) := \mathbf{k}_1^\top \mathbf{x} + \mathbf{k}_2^\top \mathbf{y} + \mathbf{k}_3$, $\mathbf{k}_1, \mathbf{k}_2 \in \mathbb{R}^n$ and $\mathbf{k}_3 \in \mathbb{R}$, Lemma 4.2 allows us to rewrite it as

$$\begin{aligned} &\mathbf{B}^1 \mathbf{Y} + \mathbf{B}^2 \boldsymbol{\Gamma} = \mathbf{v}, \\ &\boldsymbol{\Gamma} - \max(0, \boldsymbol{\Gamma} + t \cdot \boldsymbol{\Upsilon}) = 0, \end{aligned} \tag{4.2}$$

where $t > 0$ is a vector and the max operator is understood componentwise. The given matrices and variables involved in system (4.2) are respectively defined by

$$\mathbf{B}^1 := \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\alpha\mathbf{A}^\top \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \alpha\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \alpha\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\alpha\mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{A}^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{I} \end{pmatrix}, \quad \mathbf{B}^2 := \begin{pmatrix} \mathbf{D}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

$$\mathbf{v} := \begin{pmatrix} -\mathbf{k}_1 \\ -\mathbf{k}_2 \\ \mathbf{0} \\ \mathbf{0} \\ -\alpha\mathbf{b} \\ \mathbf{0} \\ \mathbf{e} \end{pmatrix}, \quad \boldsymbol{\Upsilon} := \begin{pmatrix} \mathbf{Dx} - \mathbf{d} \\ \mathbf{Ay} - \mathbf{b} \\ -\mathbf{z} \\ -\mathbf{r} \\ -\mathbf{s} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{Y} := \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \\ \mathbf{r} \\ \mathbf{s} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Gamma} := \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \\ \boldsymbol{\lambda}_3 \\ \boldsymbol{\lambda}_4 \\ \boldsymbol{\lambda}_5 \\ \boldsymbol{\lambda}_6 \\ \boldsymbol{\lambda}_7 \end{pmatrix}.$$

Here, we suppose with no loss of generality that $\boldsymbol{\lambda}_6, \boldsymbol{\lambda}_7 \geq 0$, otherwise for e.g. if $\boldsymbol{\lambda}_7 < 0$, we consider the variable $\boldsymbol{\lambda}_7' = -\boldsymbol{\lambda}_7$ and modify the matrix $\mathbf{B}^2$. On the other hand, if the matrix $\mathbf{B}^2$ has a full column rank and $(\mathbf{B}^2)^{-1}\mathbf{B}^1$ is a P-matrix, then it can be possible to develop a locally and globally convergent semismooth Newton scheme to solve (4.2); see, e.g., [27]. However, to expand the number of applications of problem (1.2) (see, e.g., next section), we would like the upper level objective function to be a more general twice continuously differentiable function. In this case, solving (4.1a)-(4.1l) is equivalent to finding the zeros of the equation

$$\Phi^{\alpha,t}(\mathbf{u}) := \begin{pmatrix} \nabla_{\mathbf{x}} F(\mathbf{x}, \mathbf{y}) + \mathbf{D}^\top \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_6 \\ \nabla_{\mathbf{y}} F(\mathbf{x}, \mathbf{y}) - \alpha\mathbf{A}^\top \mathbf{s} + \mathbf{A}^\top \boldsymbol{\lambda}_2 \\ \alpha\mathbf{r} + \mathbf{A}\boldsymbol{\lambda}_6 - \boldsymbol{\lambda}_3 \\ \alpha\mathbf{z} + \boldsymbol{\lambda}_7 - \boldsymbol{\lambda}_4 \\ \alpha(\mathbf{b} - \mathbf{Ay}) + \boldsymbol{\lambda}_7 - \boldsymbol{\lambda}_5 \\ \mathbf{A}^\top \mathbf{z} + \mathbf{x} \\ \mathbf{r} + \mathbf{s} - e \\ \boldsymbol{\lambda}_1 - \max(0, \boldsymbol{\lambda}_1 + t_1(\mathbf{Dx} - \mathbf{d})) \\ \boldsymbol{\lambda}_2 - \max(0, \boldsymbol{\lambda}_2 + t_2(\mathbf{Ay} - \mathbf{b})) \\ \boldsymbol{\lambda}_3 - \max(0, \boldsymbol{\lambda}_3 + t_3(-\mathbf{z})) \\ \boldsymbol{\lambda}_4 - \max(0, \boldsymbol{\lambda}_4 + t_4(-\mathbf{r})) \\ \boldsymbol{\lambda}_5 - \max(0, \boldsymbol{\lambda}_5 + t_5(-\mathbf{s})) \end{pmatrix} = 0 \tag{4.3}$$

with variable $\mathbf{u} := (\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}, \mathbf{s}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4, \boldsymbol{\lambda}_5, \boldsymbol{\lambda}_6, \boldsymbol{\lambda}_7) \in \mathbb{R}^{3n+8l+m}$.

Clearly, $t > 0$ is a parameter in the system of equations in (4.3) that we solve, and theoretically, any positive value should be fine when solving the system. However, as our end-goal is to compute a point that can potentially solve the original bilevel optimization problem, a suitable choice might be necessary.

It can easily be checked that (4.3) is a square $(3n + 8l + m) \times (3n + 8l + m)$ system of equations. Additionally, it is well-known that the involved max function is semismooth. Hence, we can solve the system by means of a generalized Newton method; see, e.g., [38, 39, 46].

Theorem 4.3. *[39] Fix $\alpha > 0$, $t > 0$ and consider a point $\bar{\mathbf{u}}$ such that $\Phi^{\alpha,t}(\bar{\mathbf{u}}) = \mathbf{0}$. Assume that all $V \in \partial\Phi^{\alpha,t}(\bar{\mathbf{u}})$ are nonsingular. Then every sequence generated by*

$$\bar{\mathbf{u}}^{k+1} = \bar{\mathbf{u}}^k - (\mathbf{C}^k)^{-1}\Phi^{\alpha,t}(\bar{\mathbf{u}}^k), \quad \mathbf{C}^k \in \partial\Phi^{\alpha,t}(\bar{\mathbf{u}}^k), k = 0, 1, 2, \ldots \tag{4.4}$$

is superlinearly convergent to $\bar{\mathbf{u}}$, provided that the starting point $\bar{\mathbf{u}}^0$ is sufficiently close to $\bar{\mathbf{u}}$. Moreover the strongly semismoothness of ∇F implies that the convergence rate is quadratic.

Proof. Follows from [39] and by noting that the structure of $\Phi^{\alpha,t}$ ensures that it is semismooth since the upper level objective function F is assumed to be twice continuously differentiable and hence, ∇F is semismooth. Furthermore, $\Phi^{\alpha,t}$ is strongly semismooth considering the twice continuous differentiability of ∇F and the nature of the max operator involved in the function. $\square$

Next, we derive sufficient conditions enabling the nonsingularity of all elements of the generalized Jacobian of $\Phi^{\alpha,t}$. For that, we first provide an upper estimate of the generalized Jacobian of $\Phi^{\alpha,t}$ in the sense of Clarke.

Theorem 4.4. *Fixing $\alpha > 0$ and $t > 0$, the function $\mathbf{u} \mapsto \Phi^{\alpha,t}(\mathbf{u})$ (4.3) is strongly semismooth at any point $\mathbf{u}$ and any matrix from the generalized Jacobian $\partial\Phi^{\alpha,t}(\mathbf{u})$ can be written as*

$$\begin{pmatrix}
\nabla^2_{\mathbf{xx}}F(\mathbf{x},\mathbf{y}) & \nabla^2_{\mathbf{yx}}F(\mathbf{x},\mathbf{y}) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\
\nabla^2_{\mathbf{yx}}F(\mathbf{x},\mathbf{y}) & \nabla^2_{\mathbf{yy}}F(\mathbf{x},\mathbf{y}) & \mathbf{0} & \mathbf{0} & -\alpha\mathbf{A}^\top & \mathbf{0} & \mathbf{A}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \alpha\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{A} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \alpha\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{I} \\
\mathbf{0} & -\alpha\mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{I} \\
\mathbf{I} & \mathbf{0} & \mathbf{A}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
-t_1\mathbf{p}_1 \cdot \mathbf{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{q}_1 \cdot \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & -t_2\mathbf{p}_2 \cdot \mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{q}_2 \cdot \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & t_3\mathbf{p}_3 \cdot \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{q}_3 \cdot \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & t_4\mathbf{p}_4 \cdot \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{q}_4 \cdot \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & t_5\mathbf{p}_5 \cdot \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{q}_5 \cdot \mathbf{I} & \mathbf{0} & \mathbf{0}
\end{pmatrix}$$

with the operation $\cdot$ understood as componentwise multiplication and the vectors $\mathbf{p}_i, \mathbf{q}_i, i \in \{1, \ldots, 5\}$ are defined such that

$$\begin{aligned}
&\mathbf{p}_1^j \in \partial\max(0,\cdot)(\boldsymbol{\lambda}_1 + t_1(\mathbf{D}\mathbf{x} - \mathbf{d}))_j,\ \mathbf{p}_2^k \in \partial\max(0,\cdot)(\boldsymbol{\lambda}_2 + t_2(\mathbf{A}\mathbf{y} - \mathbf{b}))_k \\
&\mathbf{p}_3^k \in \partial\max(0,\cdot)(\boldsymbol{\lambda}_3 - t_3\mathbf{z})_k, \mathbf{p}_4^k \in \partial\max(0,\cdot)(\boldsymbol{\lambda}_4 - t_4\mathbf{r})_k, \\
&\mathbf{p}_5^k \in \partial\max(0,\cdot)(\boldsymbol{\lambda}_5 - t_5\mathbf{s})_k, (j,k) \in \{1, \ldots, m\} \times \{1, \ldots, l\}, \\
&\mathbf{q}_i := \mathbf{e} - \mathbf{p}_i, i \in \{1, \ldots, 5\},
\end{aligned}$$

Proof. It suffices to calculate the generalized derivative of the last five components of $\Phi^{\alpha,t}(\mathbf{u})$ (4.3). On the other hand, we mention that the reasoning below will be done only for the equation

$$Q(\mathbf{u}) := \boldsymbol{\lambda}_1 - \max(0, \boldsymbol{\lambda}_1 + t_1(\mathbf{D}\mathbf{x} - \mathbf{d}))$$

and this can be repeated for the fourth other equations. For that we observe that for $i = 1, \dots, m$

$$Q^i(\mathbf{u}) = \begin{cases} \boldsymbol{\lambda}_1^i & \text{if} \quad \boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i \leq 0, \\ -t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i & \text{if} \quad \boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i > 0. \end{cases}$$

Therein, $\boldsymbol{\lambda}_1^i$ (resp. $(\mathbf{D}\mathbf{x} - \mathbf{d})^i$, $Q^i(\mathbf{u})$) is the i-th component of $\boldsymbol{\lambda}_1$ (resp. $\mathbf{D}\mathbf{x} - \mathbf{d}$, $Q(\mathbf{u})$). We see that the function Q^i is differentiable when $\boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i > 0$ or when $\boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i < 0$. However it is not differentiable when $\boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i = 0$. Hence, by setting $\mathbf{D}^i$ to be the i-th row of $\mathbf{D}$ we get

$$\begin{aligned} \partial Q^i(\mathbf{u}) &= \begin{cases} \mathbf{M}^i := (0, \dots, 0, 1, 0, \dots, 0) & \text{if} \quad \boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i < 0, \\ \mathbf{N}^i = (-t_1\mathbf{D}^i, \dots, 0, \dots, 0) & \text{if} \quad \boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i > 0, \\ \operatorname{conv}(\mathbf{N}^i, \mathbf{M}^i) & \text{if} \quad \boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i = 0 \end{cases} \\ &= \left\{ \xi_i \mathbf{N}^i + (1 - \xi_i)\mathbf{M}^i \;\middle|\; \begin{array}{r} \xi_i = 1 \text{ if } \boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i > 0, \\ \xi_i = 0 \text{ if } \boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i < 0, \\ \xi_i \in [0,1] \text{ if } \boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i = 0 \end{array} \right\}. \end{aligned}$$

Therefore,

$$\partial Q(\mathbf{u}) = \left\{ \mathbf{p}_1 \cdot \mathbf{N} + (\mathbf{e} - \mathbf{p}_1) \cdot \mathbf{M} \;\middle|\; \begin{array}{r} \xi_i = 1 \text{ if } \boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i > 0, \\ \xi_i = 0 \text{ if } \boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i < 0, \\ \xi_i \in [0,1] \text{ if } \boldsymbol{\lambda}_1^i + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})^i = 0, \\ \mathbf{p}_1 = (\xi_1, \dots, \xi_n)^\top \end{array} \right\}$$

with $\mathbf{M}$ and $\mathbf{N}$ respectively defined by

$$\begin{aligned} \mathbf{M} &:= \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \\ \mathbf{N} &:= \begin{pmatrix} -t_1\mathbf{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \end{aligned} \tag{4.5}$$

and $\mathbf{p}_1 \cdot \mathbf{N}$ corresponds to the matrix whose rows are obtained from the multiplication of the $\mathrm{i_{th}}$ row of $\mathbf{N}$ and the vector $\mathbf{p}_1$. The same notation holds similarly for $(\mathbf{e} - \mathbf{p}_1) \cdot \mathbf{M}$. □

Observe that it is also possible to compute the generalized derivative of $\Phi^{\alpha,t}$ by using the following chain rule

$$\partial(g \circ h)(x) = \partial g(h(x))\partial h(x),$$

where $g \colon \mathbb{R}^{m_1} \to \mathbb{R}^1$ is convex and $h \colon \mathbb{R}^{m_2} \to \mathbb{R}^{m_1}$ is continuously differentiable, see [7, Theorem 2.6.6]. Both procedures lead to the same result.

Next, we identify some scenarios, where the nonsingularity assumption on $\partial\Phi^{\alpha,t}$ required in the convergence Theorem 4.3 is satisfied. To proceed, we consider the functions

$$h_1(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}, \mathbf{s}) := \mathbf{D}\mathbf{x} - \mathbf{d}, \quad h_2(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}, \mathbf{s}) := \mathbf{A}\mathbf{y} - \mathbf{b},$$

$$h_3(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}, \mathbf{s}) := -\mathbf{z}, \quad h_4(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}, \mathbf{s}) := -\mathbf{r}, \quad h_5(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}, \mathbf{s}) := -\mathbf{s},$$

and the following sets, for $i = 1, \dots, 5$:

$$P_i := \{j \in \{1, \dots, k_i\} \mid 0 < (\boldsymbol{\lambda}_i + t_i h_i(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}, \mathbf{s}))_j\}$$
$$Q_i := \{j \in \{1, \dots, k_i\} \mid 0 > (\boldsymbol{\lambda}_i + t_i h_i(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}, \mathbf{s}))_j\}$$

with

$$k_1 = m, k_2 = k_3 = k_4 = k_5 = l.$$

Theorem 4.5. *Let $\bar{\mathbf{u}}$ be a point that satisfies the optimality conditions in Theorem 4.1. Suppose that the matrix $\mathbf{A}$ (with $l = n$) is invertible and $\nabla^2_{xx} F(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ has full column rank. Furthermore, assume that the sets P_1, P_3, P_5, Q_2, and Q_4 are empty. Then, all the elements of $\partial\Phi^{\alpha,t}(\bar{\mathbf{u}})$ are nonsingular.*

Proof. Choosing an arbitrary matrix $\mathbf{C} \in \partial\Phi^{\alpha,t}(\bar{\mathbf{u}})$ and considering the homogeneous linear system $\mathbf{Cd} = 0$, with a suitably partitioned vector $\mathbf{d} := (\mathbf{d}_i)_{i=1}^{12}$,

$$\nabla^2_{\mathbf{xx}} F(\bar{\mathbf{x}}, \bar{\mathbf{y}})\mathbf{d}_1 + \nabla^2_{\mathbf{yx}} F(\bar{\mathbf{x}}, \bar{\mathbf{y}})\mathbf{d}_2 + \mathbf{D}^\top \mathbf{d}_6 + \mathbf{d}_{11} = \mathbf{0}, \tag{4.6}$$
$$\nabla^2_{\mathbf{xy}} F(\bar{\mathbf{x}}, \bar{\mathbf{y}})\mathbf{d}_1 + \nabla^2_{\mathbf{yy}} F(\bar{\mathbf{x}}, \bar{\mathbf{y}})\mathbf{d}_2 - \alpha\mathbf{A}^\top \mathbf{d}_5 + \mathbf{A}^\top \mathbf{d}_7 = \mathbf{0}, \tag{4.7}$$
$$\alpha\mathbf{d}_4 + \mathbf{A}\mathbf{d}_{11} - \mathbf{d}_8 = \mathbf{0}, \tag{4.8}$$
$$\alpha\mathbf{d}_3 + \mathbf{d}_{12} - \mathbf{d}_9 = \mathbf{0}, \tag{4.9}$$
$$-\alpha\mathbf{A}\mathbf{d}_2 + \mathbf{d}_{12} - \mathbf{d}_{10} = \mathbf{0}, \tag{4.10}$$
$$\mathbf{d}_1 + \mathbf{A}^\top \mathbf{d}_3 = 0, \tag{4.11}$$
$$\mathbf{d}_4 + \mathbf{d}_5 = \mathbf{0}, \tag{4.12}$$
$$-t_1\mathbf{p}_1 \cdot \mathbf{D}\mathbf{d}_1 + (\mathbf{e} - \mathbf{p}_1)\mathbf{d}_6 = \mathbf{0}, \tag{4.13}$$
$$-t_2\mathbf{p}_2 \cdot \mathbf{A}\mathbf{d}_2 + (\mathbf{e} - \mathbf{p}_2)\mathbf{d}_7 = \mathbf{0}, \tag{4.14}$$
$$t_3\mathbf{p}_3\mathbf{d}_3 + (\mathbf{e} - \mathbf{p}_3)\mathbf{d}_8 = \mathbf{0}, \tag{4.15}$$
$$t_4\mathbf{p}_4\mathbf{d}_4 + (\mathbf{e} - \mathbf{p}_4)\mathbf{d}_9 = \mathbf{0}, \tag{4.16}$$
$$t_5\mathbf{p}_5\mathbf{d}_5 + (\mathbf{e} - \mathbf{p}_5)\mathbf{d}_{10} = \mathbf{0}. \tag{4.17}$$

Since the sets P_1, P_3, P_5, Q_2, Q_4 are empty, we get from the definitions of $\mathbf{p}_1$, $\mathbf{p}_3$, $\mathbf{p}_5$, $\mathbf{p}_2$, and $\mathbf{p}_4$ that $\mathbf{p}_1 = \mathbf{p}_3 = \mathbf{p}_5 = \mathbf{0}$ and $\mathbf{p}_2 = \mathbf{p}_4 = \mathbf{e}$. It follows then from (4.12)–(4.17) and the invertibility of $\mathbf{A}$ that $\mathbf{d}_6 = \mathbf{0} = \mathbf{d}_2 = \mathbf{d}_8 = \mathbf{d}_4 = \mathbf{d}_5 = \mathbf{d}_{10}$. Inserting these values in (4.8) and in (4.6) leads to $\mathbf{d}_{11} = \mathbf{0} = \mathbf{d}_1$ since the matrix $\nabla^2_{\mathbf{xx}} F(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ has full column rank. The reuse of the invertibility of $\mathbf{A}$ in (4.7) and (4.11) gives $\mathbf{d}_7 = \mathbf{0} = \mathbf{d}_3$. The equations (4.9)–(4.10) lead to $\mathbf{d}_{12} = \mathbf{0} = \mathbf{d}_9$. □

Example 4.6. *Consider the bilevel optimization problem*

$$\begin{aligned} \min_{x,y} \quad & -16x + 3x^2 + 3y^2 \\ s.t. \quad & 1 \le x \le \tfrac{8}{3}, \ \ y \in \Psi(x) = \underset{y}{Argmin}\, \{xy \mid y \ge 0\}. \end{aligned} \tag{4.18}$$

The optimal solution of problem (4.18) is $(\bar{x}, \bar{y}) = (\frac{8}{3}, 0)$. One can check that the assumption of Proposition 3.1 is satisfied. Therefore there exists a scalar $\bar{\alpha} > 0$

such that for all $\alpha \geq \bar{\alpha}$ we can find $(\bar{r}, \bar{s}) \in T$ such that the point $(\frac{8}{3}, 0, \frac{8}{3}, \bar{r}, \bar{s})$ solves the problem

$$\begin{array}{ll} \min\limits_{x,y,z,r,s} & -16x + 3x^2 + 3y^2 + \alpha(rz + sy) \\ s.t. & 1 \leq x \leq \frac{8}{3}, \;\; y \geq 0, \;\; x = z \geq 0, \;\; r + s = 1, \;\; r \geq 0, \;\; s \geq 0. \end{array} \tag{4.19}$$

The optimal solution of (4.19) *is* $(\frac{8}{3}, 0, \frac{8}{3}, 0, 1)$. *The corresponding calculations show that the point*

$$\bar{\mathbf{u}} = (\bar{x}, \bar{y}, \bar{z}, \bar{r}, \bar{s}, \bar{\boldsymbol{\lambda}}_1, \bar{\lambda}_2, \bar{\lambda}_3, \bar{\lambda}_4, \bar{\lambda}_5, \bar{\lambda}_6, \bar{\lambda}_7) = \left(\frac{8}{3}, 0, \frac{8}{3}, 0, 1, 0, 0, \alpha, 0, \frac{8}{3}\alpha, 0, 0, 0\right)$$

solves the optimality conditions (4.1a)-(4.1l). *Furthermore, it is clear that the matrices $\mathbf{A} = -1$ and $\nabla^2_{xx} F(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ are invertible, and the following implications hold*

$$\begin{array}{rlrl} \bar{\boldsymbol{\lambda}}_1 + t_1(\mathbf{D}\bar{x} - \mathbf{d}) \leq 0 & \Longrightarrow \;\; \mathbf{p}_1 = \mathbf{0}, & \bar{\lambda}_3 - t_3\bar{z} < 0 & \Longrightarrow \;\; p_3 = 0, \\ \bar{\lambda}_2 + t_2(\mathbf{A}\bar{y} - \mathbf{b}) > 0 & \Longrightarrow \;\; p_2 = 1, & \bar{\lambda}_5 - t_5\bar{s} < 0 & \Longrightarrow \;\; p_5 = 0 \\ \bar{\lambda}_4 - t_4\bar{r} > 0 & \Longrightarrow \;\; p_4 = 1 & & \end{array}$$

Therefore, based on Theorem 4.5, all elements of $\partial\Phi^{\alpha,t}(\bar{\mathbf{u}})$ are nonsingular for any value of $\alpha \geq \bar{\alpha}$.

Obviously, requiring the invertibility of $\mathbf{A}$ in Theorem 4.5 is a very strong assumption, as it can imply that the lower level has a unique feasible point. In the sequel, we present a scenario where we avoid to impose this assumption.

Theorem 4.7. *Let $\bar{\mathbf{u}}$ be a point satisfying the optimality conditions given in Theorem 4.1. Suppose that the matrix $\nabla^2_{yy} F(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ has full column rank. Furthermore, we assume that the followings sets P_1, P_2, P_5, Q_3, and Q_4 are empty. Then, all the elements of $\partial\Phi^{\alpha,t}(\bar{\mathbf{u}})$ are nonsingular.*

Proof. Proceeding as in the proof of Theorem 4.5, it follows from the emptiness of the sets P_1, P_2, P_5, Q_3, Q_4 and (4.12)–(4.17) that $\mathbf{d}_6 = \mathbf{0} = \mathbf{d}_7 = \mathbf{d}_3 = \mathbf{d}_4 = \mathbf{d}_5 = \mathbf{d}_{10}$ because $\mathbf{p}_1 = \mathbf{p}_2 = \mathbf{p}_5 = \mathbf{0}$ and $\mathbf{p}_3 = \mathbf{p}_4 = \mathbf{e}$. Inserting these values in (4.7) and (4.11) leads to $\mathbf{d}_1 = \mathbf{0} = \mathbf{d}_2$ given that the matrix $\nabla^2_{yy} F(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is supposed to have full column rank. Inserting again these values in (4.6), (4.8)–(4.10) imply that $\mathbf{d}_{11} = \mathbf{0} = \mathbf{d}_8 = \mathbf{d}_9 = \mathbf{d}_{12}$. ☐

Example 4.8. *Consider the bilevel optimization problem*

$$\begin{array}{ll} \min\limits_{x,y} & 2(x_1 + x_2) + x_1^2 + y_1^2 + y_2^2 \\ s.t. & x_1 + x_2 \leq 2, \; x_1 \geq 0, \; x_2 \geq 0, \\ & \mathbf{y} \in \Psi(\mathbf{x}) = \underset{\mathbf{y}}{Argmin} \, \{\mathbf{x}^\top \mathbf{y} \mid y_1 = y_2\}. \end{array} \tag{4.20}$$

One can observe that the matrix

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

is not invertible and that an optimistic solution of problem (4.20) *is* $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = (0, 0, 0, 0)$. *One can check that the assumption of Proposition 3.1 is satisfied. Therefore, there exists a scalar $\bar{\alpha} > 0$ such that for all $\alpha \geq \bar{\alpha}$, the point*

$$\bar{\mathbf{u}} = (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}, \bar{\mathbf{r}}, \bar{\mathbf{s}}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4, \boldsymbol{\lambda}_5, \boldsymbol{\lambda}_6, \boldsymbol{\lambda}_7)$$

with $(\bar{\mathbf{r}}, \bar{\mathbf{s}}) = (0,0,1,1)$, $\boldsymbol{\lambda}_1 = (0,0)$, $\boldsymbol{\lambda}_2 = (0,0)$, $\boldsymbol{\lambda}_3 = (0,0)$, $\boldsymbol{\lambda}_4 = (0,0)$, $\boldsymbol{\lambda}_5 = (0,0)$, $\boldsymbol{\lambda}_6 = (-2,-2)$, $\boldsymbol{\lambda}_7 = (0,0)$, *and* $\bar{\mathbf{z}} = (0,0)$ *solves the optimality conditions* (4.1a)-(4.1l). *Furthermore, it is clear that the matrix*

$$\nabla^2_{yy} F(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

is full column rank, and the following implications hold

$$\begin{aligned} \boldsymbol{\lambda}_1 + t_1(\mathbf{D}\bar{x} - \mathbf{d}) \leq 0 &\Longrightarrow \quad \mathbf{p}_1 = \mathbf{0}, \quad \boldsymbol{\lambda}_3 - t_3\bar{\mathbf{z}} \geq 0 \quad \Longrightarrow \quad \mathbf{p}_3 = \mathbf{e}, \\ \boldsymbol{\lambda}_2 + t_2(\mathbf{A}\bar{y} - \mathbf{b}) \leq 0 &\Longrightarrow \quad \mathbf{p}_2 = \mathbf{0}, \quad \boldsymbol{\lambda}_4 - t_4\bar{\mathbf{r}} \geq 0 \quad \Longrightarrow \quad \mathbf{p}_4 = \mathbf{e}, \\ \boldsymbol{\lambda}_5 - t_5\bar{\mathbf{s}} \leq 0 &\Longrightarrow \quad \mathbf{p}_5 = \mathbf{0}. \end{aligned}$$

Therefore, all the assumptions of Theorem 4.7 are fulfilled for $\bar{\mathbf{u}}$ *and any* $\alpha \geq \bar{\alpha}$.

Remark 4.9. *The following observations can be made:*

(1) *The assumptions on the emptiness of* P_i *and* Q_i*, for* $i \in \{1, \ldots, 5\}$ *can be reflected in some situations, where the solution w.r.t. the corresponding constraint is nondegenerate or not. For example,* $P_1 = \emptyset$ *means that* $\boldsymbol{\lambda}_1 + t_1(\mathbf{Dx} - \mathbf{d}) \leq 0$. *Clearly, if* $\mathbf{Dx} - \mathbf{d} < 0$, *we get the emptiness of* P_1 *from the complementarity conditions. Thus, the strict complementarity condition of some constraints is sufficient to ensure the emptiness of some sets* P_i *or* Q_i *for* $i \in \{1, \ldots, 5\}$.
(2) *There could be a chance that the combination of the MPCC-LICQ (linear independence constraint qualification for mathematical program with complementarity constraints) and MPCC-SOSC (strong second-order sufficient condition for mathematical program with complementarity constraints) could work; we have thought about this, but no obvious connection seems apparent. The topic will be further investigated in the future. However, as we have seen, the assumptions of Theorems 4.4 and 4.5 are satisfied in Examples 4.1 and 4.2 but one can easily check that MPCC-LICQ is satisfied w.r.t. Example 4.1 (even MPCC-SOSC) but not w.r.t. Example 4.2.*

We cannot rely on the square of $\Phi^{\alpha,t}$ in order to provide a suitable globalization due to the presence of the function max. This is one of the main challenges of this local convergence and we are aware that the Fischer-Burmeister function [22] could come into play when replacing the complementarity conditions in Theorem 4.1 instead using lemma 4.1. But as mentioned in comments after Lemma 4.1, our primary goal in this paper is to explore the approach considered in the paper, which, to the best of our knowledge, has not been considered before in the context of bilevel optimization. Later on, we will see a way to construct a globalized algorithm for solving (4.3).

To implement the equation (4.4), we should be able to calculate an element of $\partial\Phi^{\alpha,t}(\mathbf{u}^k)$. Our task here is therefore to show how this can be accomplished. To proceed, we define for $t_i > 0$, $i = 1, \ldots, 5$, the function $X : \mathbb{R}^{3n+8l+m} \to \mathbb{R}^{4l+m}$ such that each $\mathbf{u} := (\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}, \mathbf{s}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4, \boldsymbol{\lambda}_5, \boldsymbol{\lambda}_6, \boldsymbol{\lambda}_7)$ assigns

$$X(\mathbf{u}) = (X^1(\mathbf{u}) \ldots X^5(\mathbf{u}))^\top.$$

Therein,

$$\begin{aligned} X^1(\mathbf{u}) &:= \boldsymbol{\lambda}_1 + t_1(\mathbf{D}\mathbf{x} - \mathbf{d}), \qquad X^3(\mathbf{u}) := \boldsymbol{\lambda}_3 - t_3\mathbf{z}, \qquad X^5(\mathbf{u}) := \boldsymbol{\lambda}_5 - t_5\mathbf{s}. \\ X^2(\mathbf{u}) &:= \boldsymbol{\lambda}_2 + t_2(\mathbf{A}\mathbf{y} - \mathbf{b}), \qquad X^4(\mathbf{u}) := \boldsymbol{\lambda}_4 - t_4\mathbf{r}. \end{aligned}$$

Theorem 4.10. *Let* $\mathbf{u} := (\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}, \mathbf{s}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4, \boldsymbol{\lambda}_5, \boldsymbol{\lambda}_6, \boldsymbol{\lambda}_7) \in \mathbb{R}^{3n+8l+m}$. *The following matrix is an element of* $\partial\Phi^{\alpha,t}(\mathbf{u})$.

$$H := \begin{pmatrix} \nabla^2_{\mathbf{xx}}F(\mathbf{x},\mathbf{y}) & \nabla^2_{\mathbf{yx}}F(\mathbf{x},\mathbf{y}) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \nabla^2_{\mathbf{yx}}F(\mathbf{x},\mathbf{y}) & \nabla^2_{\mathbf{yy}}F(\mathbf{x},\mathbf{y}) & \mathbf{0} & \mathbf{0} & -\alpha\mathbf{A}^\top & \mathbf{0} & \mathbf{A}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \alpha\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \alpha\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & -\alpha\mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} & \mathbf{A}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ K^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{K}^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & K^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{K}^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & K^3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{K}^3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & K^4 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{K}^4 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & K^5 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{K}^5 & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

$$K^1 := -\tfrac{t_1}{2}\left(\mathbf{D} + sgn(X^1(\mathbf{u}))\cdot\mathbf{D}\right), \quad K^2 := -\tfrac{t_2}{2}\left(\mathbf{A} + sgn(X^2(\mathbf{u}))\cdot\mathbf{A}\right)$$

$$K^i := \tfrac{t_i}{2}\left(\mathbf{I} + sgn(X^i(\mathbf{u}))\cdot\mathbf{I}\right), i = 3, 4, 5,$$

$$\tilde{K}^i := \tfrac{1}{2}\left(\mathbf{I} - sgn(X^i(\mathbf{u}))\cdot\mathbf{I}\right), \ i = 1, \ldots, 5.$$

Proof. Observe that the first $3n + 4l$ components of $\Phi^{\alpha,t}$ are differentiable. Hence, we focus our attention on the components which contain the max operator. In order to prove that $H \in \partial\Phi^{\alpha,t}(\mathbf{u})$, we will build a sequence of points $(\tilde{\mathbf{u}}^k)_k$ and $(\tilde{\tilde{\mathbf{u}}}^k)_k$, where $\Phi^{\alpha,t}$ is differentiable at the points $\mathbf{u} + \varepsilon^k\tilde{\mathbf{u}}$ and $\mathbf{u} + \varepsilon^k\tilde{\tilde{\mathbf{u}}}^k$ and it holds $\nabla\Phi^{\alpha,t}(\mathbf{u} + \varepsilon^k\tilde{\mathbf{u}}) \to H^1$ and $\nabla\Phi^{\alpha,t}(\mathbf{u} + \varepsilon^k\tilde{\tilde{\mathbf{u}}}^k) \to H^2$ for $k \to \infty$. After that we will show that H is an element of the convex hull of H^1 and H^2. Let

$$\Delta := \left\{ i \in \{1, \ldots, 4l + m\} \mid X^j_i(\mathbf{u}) = 0, \text{ for at least one } j \in \{1, \ldots, 5\} \right\}.$$

We consider some vectors

$$\tilde{\mathbf{u}} := (\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}}, \tilde{\mathbf{r}}, \tilde{\mathbf{s}}, \tilde{\boldsymbol{\lambda}}_1, \tilde{\boldsymbol{\lambda}}_2, \tilde{\boldsymbol{\lambda}}_3, \tilde{\boldsymbol{\lambda}}_4, \tilde{\boldsymbol{\lambda}}_5, \tilde{\boldsymbol{\lambda}}_6, \tilde{\boldsymbol{\lambda}}_7)$$

and

$$\tilde{\tilde{\mathbf{u}}} := (\tilde{\tilde{\mathbf{x}}}, \tilde{\tilde{\mathbf{y}}}, \tilde{\tilde{\mathbf{z}}}, \tilde{\tilde{\mathbf{r}}}, \tilde{\tilde{\mathbf{s}}}, \tilde{\tilde{\boldsymbol{\lambda}}}_1, \tilde{\tilde{\boldsymbol{\lambda}}}_2, \tilde{\tilde{\boldsymbol{\lambda}}}_3, \tilde{\tilde{\boldsymbol{\lambda}}}_4, \tilde{\tilde{\boldsymbol{\lambda}}}_5, \tilde{\tilde{\boldsymbol{\lambda}}}_6, \tilde{\tilde{\boldsymbol{\lambda}}}_7)$$

from $\mathbb{R}^{3n+8l+m}$, such that $\tilde{\mathbf{u}}_i = 0 = \tilde{\tilde{\mathbf{u}}}_i$ for $i \notin \Delta$ and $\tilde{\mathbf{u}}_i = 1$, $\tilde{\tilde{\mathbf{u}}}_i = -1$ for $i \in \Delta$.

For ease of reference we will treat only one equation since the same reasoning can be analogously applied on the other equations.

If there exists $i_0 \in \{1, \dots, 4l+m\}$ such that $X^j_{i_0}(\mathbf{u}) \neq 0$, for all $j = 1, \dots, 5$. We get for example $X^1_{i_0}(\mathbf{u}) \neq 0$. Then $\Phi^{\alpha,t}_{3n+4l+i_0}$ is differentiable at $\mathbf{u}$ and we get from the proof of Theorem 4.4, cf. (4.5), that

$$\partial \Phi^{\alpha,t}_{3n+4l+i_0}(\mathbf{u}) = \begin{cases} \mathbf{N}^{i_0} & \text{if} \quad X^1_{i_0}(\mathbf{u}) > 0, \\ \mathbf{M}^{i_0} & \text{if} \quad X^1_{i_0}(\mathbf{u}) < 0. \end{cases}$$

This gives the row $H_{3n+4l+i_0}$ since $\operatorname{sgn}(X^1_{i_0}(\mathbf{u})) = \pm 1$ and $K^1_{i_0} = -t_1 \mathbf{D}_{i_0}$, $\tilde{K}^1_{i_0} = \mathbf{0}$ if $X^1_{i_0}(\mathbf{u}) > 0$ and $K^1_{i_0} = \mathbf{0}$, $\tilde{K}^1_{i_0} = \mathbf{I}$ if $X^1_{i_0}(\mathbf{u}) < 0$.

Otherwise, suppose that for all $i \in \{1, \dots, 4l+m\}$, $X^j_i(\mathbf{u}) = 0$, for some $j = 1, \dots, 5$. Assume, without loss of generality that $X^1(\mathbf{u}) = \mathbf{0}$ (because we can apply the same reasoning on all $j = 1, \dots, 5$) then $\operatorname{sgn}(X^1(\mathbf{u})) = \mathbf{0}$ (it is worth mentioning that this definition of sgn should be understood componentwise). We consider the sequences $\tilde{\mathbf{u}}^k_i := \tilde{\mathbf{u}}_i + \varepsilon^k$, $\tilde{\tilde{\mathbf{u}}}^k_i := \mathbf{u}_i + \varepsilon^k$, $k \in \mathbb{N}$, and $\varepsilon^k \downarrow 0$. It holds

$$\begin{aligned} X^1(\mathbf{u} + \varepsilon^k \tilde{\mathbf{u}}) &= (\boldsymbol{\lambda}_1 + \varepsilon^k \tilde{\boldsymbol{\lambda}}_1) + t_1(\mathbf{D}(\mathbf{x} + \varepsilon^k \tilde{\mathbf{x}}) - \mathbf{d}) \\ &= \underbrace{\boldsymbol{\lambda}_1 + t_1(\mathbf{D}\mathbf{x} - \mathbf{d})}_{=0} + \varepsilon^k \tilde{\boldsymbol{\lambda}}_1 + t_1 \varepsilon^k \mathbf{D}\tilde{\mathbf{x}} \\ &= \varepsilon^k(\tilde{\boldsymbol{\lambda}}_1 + t_1 \mathbf{D}\tilde{\mathbf{x}}) > 0 \quad \text{for } t_1 \text{ small enough since } \tilde{\boldsymbol{\lambda}}_1 = 1. \end{aligned}$$

In the same vein, we have $X^1(\mathbf{u} + \varepsilon^k \tilde{\tilde{\mathbf{u}}}) = \varepsilon^k(\tilde{\tilde{\boldsymbol{\lambda}}}_1 + t_1 \mathbf{D}\tilde{\tilde{\mathbf{x}}}) < 0$ for t_1 small enough since $\tilde{\tilde{\boldsymbol{\lambda}}}_1 = -1$. Consequently, the mapping $\Phi^{\alpha,t}_1 : \mathbf{u} \mapsto \boldsymbol{\lambda}_1 - \max(0, \boldsymbol{\lambda}_1 + t_1(\mathbf{D}\mathbf{x} - \mathbf{d}))$ is at the points $\mathbf{u} + \varepsilon^k \tilde{\mathbf{u}}$ and $\mathbf{u} + \varepsilon^k \tilde{\tilde{\mathbf{u}}}$ differentiable and we get

$$\begin{aligned} \nabla \Phi^{\alpha,t}_1(\mathbf{u} + \varepsilon^k \tilde{\tilde{\mathbf{u}}}) &\underset{k\to\infty}{\to} \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} := H^1 \\ \nabla \Phi^{\alpha,t}_1(\mathbf{u} + \varepsilon^k \tilde{\mathbf{u}}) &\underset{k\to\infty}{\to} \begin{pmatrix} -t_1\mathbf{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} := H^2 \end{aligned}$$

and clearly the matrix

$$\begin{pmatrix} \frac{-t_1}{2}\mathbf{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{2}\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

belongs to the convex hull of H^1 and H^2. We end the proof by handling analogously the other equations as suggested some lines before.

$\square$

On the other hand, the particularity of the matrix H constructed in the previous theorem is the fact that it is the limit of the gradient of the following differentiable approximation

$$\Phi_\varepsilon^{\alpha,t}(\mathbf{u}) := \begin{pmatrix} \nabla_{\mathbf{x}} F(\mathbf{x},\mathbf{y}) + \mathbf{D}^\top \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_6 \\ \nabla_{\mathbf{y}} F(\mathbf{x},\mathbf{y}) - \alpha \mathbf{A}^\top \mathbf{s} + \mathbf{A}^\top \boldsymbol{\lambda}_2 \\ \alpha \mathbf{r} + \mathbf{A}\boldsymbol{\lambda}_6 - \boldsymbol{\lambda}_3 \\ \alpha \mathbf{z} + \boldsymbol{\lambda}_7 - \boldsymbol{\lambda}_4 \\ \alpha(\mathbf{b} - \mathbf{A}\mathbf{y}) + \boldsymbol{\lambda}_7 - \boldsymbol{\lambda}_5 \\ \mathbf{A}^\top \mathbf{z} + \mathbf{x} \\ \mathbf{r} + \mathbf{s} - e \\ \boldsymbol{\lambda}_1 - \frac{1}{2}(X^1(\mathbf{u}) + \sqrt{(X^1(\mathbf{u}))^2 + \varepsilon}) \\ \boldsymbol{\lambda}_2 - \frac{1}{2}(X^2(\mathbf{u}) + \sqrt{(X^2(\mathbf{u}))^2 + \varepsilon}) \\ \boldsymbol{\lambda}_3 - \frac{1}{2}(X^3(\mathbf{u}) + \sqrt{(X^3(\mathbf{u}))^2 + \varepsilon}) \\ \boldsymbol{\lambda}_4 - \frac{1}{2}(X^4(\mathbf{u}) + \sqrt{(X^4(\mathbf{u}))^2 + \varepsilon}) \\ \boldsymbol{\lambda}_5 - \frac{1}{2}(X^5(\mathbf{u}) + \sqrt{(X^5(\mathbf{u}))^2 + \varepsilon}) \end{pmatrix} \tag{4.21}$$

of the function $\Phi^{\alpha,t}$. Since for $\varepsilon > 0$ and $x \in \mathbb{R}$,

$$\frac{1}{2}(x + \sqrt{x^2 + \varepsilon}) \;\longrightarrow\; \max(0, x) \quad \text{as} \quad \varepsilon \;\longrightarrow\; 0.$$

We recall that the square and the square-root applied on vectors in (4.21) are understood componentwise. It can easily be checked that the gradient of $\Phi_\varepsilon^{\alpha,t}$ can be written as

$$\nabla \Phi_\varepsilon^{\alpha,t}(\mathbf{u}) = \begin{pmatrix} \nabla^2_{\mathbf{xx}} F(\mathbf{x},\mathbf{y}) & \nabla^2_{\mathbf{yx}} F(\mathbf{x},\mathbf{y}) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \nabla^2_{\mathbf{yx}} F(\mathbf{x},\mathbf{y}) & \nabla^2_{\mathbf{yy}} F(\mathbf{x},\mathbf{y}) & \mathbf{0} & \mathbf{0} & -\alpha\mathbf{A}^\top & \mathbf{0} & \mathbf{A}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \alpha\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \alpha\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & -\alpha\mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} & \mathbf{A}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ G^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{G}^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & G^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{G}^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & G^3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{G}^3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & G^4 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{G}^4 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & G^5 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{G}^5 & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where

$$G^1 := \frac{-t_1}{2}(\mathbf{D} + (X^1 \div L^1)\cdot \mathbf{D}), \quad G^i := \frac{t_i}{2}(\mathbf{I} + (X^i \div L^i)\cdot \mathbf{I}),\ i = 3,4,5,$$
$$G^2 := \frac{-t_2}{2}(\mathbf{A} + (X^2 \div L^2)\cdot \mathbf{A}), \quad \tilde{G}^i := \frac{1}{2}(\mathbf{I} - (X^i \div L^i)\cdot \mathbf{I}),\ i = 1,\dots,5.$$

Here, for $i = 1, \ldots, 5$, we define $L^i := (L^i_j)_{j=1}^{\theta_i}$ with $L^i_j := \sqrt{X^i_j + \varepsilon}$ and $\theta_1 = m, \theta_i = l$, for $i = 2, \ldots, 5$. The signs $\div$ and $\cdot$ are used to denote the componentwise division and multiplication, respectively.

The above observation implies that we have two options to evaluate an element from the generalized Jacobian of the function $\Phi^{\alpha,t}$ in (4.4): either we compute a subgradient of $\Phi^{\alpha,t}$ directly if possible or we make use of the derivative of the approximation function $\Phi^{\alpha,t}_\varepsilon$ (which is differentiable, hence more tractable) for small values of ε and by letting ε go to 0. For the numerical application of the method, we use the following algorithm.

Algorithm 1 : Newton algorithm

Step 0: Choose $k = 0$, $\mathbf{u}^0 = (\mathbf{x}^0, \mathbf{y}^0, \mathbf{z}^0, \boldsymbol{\lambda}^0_1, \boldsymbol{\lambda}^0_2, \boldsymbol{\lambda}^0_3, \boldsymbol{\lambda}^0_4, \boldsymbol{\lambda}^0_5, \boldsymbol{\lambda}^0_6, \boldsymbol{\lambda}^0_7)$, $\delta > 0$, $t > 0$, $\varepsilon > 0$, $\beta \in (0, 1/2)$, $\tau^0 = 1$, $\alpha_1 \in (0, 0.5)$ and $\alpha > 0$.
Step 1: Compute $\Phi^{\alpha,t}_\varepsilon(\mathbf{u}^k)$. If $\|\Phi^{\alpha,t}_\varepsilon(\mathbf{u}^k)\| \leq \delta$, stop.
Step 2: Select an element $\mathbf{C}^k \in \partial\Phi^{\alpha,t}_\varepsilon(\mathbf{u}^k)$ and compute $\mathbf{d}^k$ such that

$$\mathbf{C}^k\mathbf{d}^k + \Phi^{\alpha,t}_\varepsilon(\mathbf{u}^k) = 0. \tag{4.22}$$

Step 3: While

$$\Psi^{\alpha,t}_\varepsilon(\mathbf{u}^k - \tau_k\nabla\Psi^{\alpha,t}_\varepsilon(\mathbf{u}^k)) > \Psi^{\alpha,t}_\varepsilon(\mathbf{u}^k) + \alpha_1\tau_k\nabla\Psi^{\alpha,t}_\varepsilon(\mathbf{u}^k)^\top\mathbf{d}^k,$$

set $\tau_k = \beta\tau_k$, with $\Psi^{\alpha,t}_\varepsilon(\mathbf{u}) := \frac{1}{2}\|\Phi^{\alpha,t}_\varepsilon(\mathbf{u})\|^2$.
Step 4: Update $\mathbf{u}^k = \mathbf{u}^k + \tau_k\mathbf{d}^k$.
Step 5: Set $k = k + 1$ and go to Step 1.

The convergence and interesting features of this algorithm have been well investigated in [10].

Remark 4.11. (1) *If the equation* (4.22) *is not solvable or if the nonsingularity assumption for* $\Phi_{\alpha,t}$ *e.g. in Theorem 4.7 is violated, there is a standard trick in numerical optimization which consists to add a factor of the identity matrix to the corresponding matrix from the generalized Jacobian* $\partial\Phi_{\alpha,t}$. *This often leads to a matrix which is nonsingular and is a good approximation of the corresponding element of the Clarke subdifferential.*

(2) *As mentioned in the introductory part, the method developed in this section to solve* (1.2), (1.3) *can be applied to* (1.2), (1.5). *In fact, for the latter problem, the counterpart of the KKT reformulation* (3.1) *can be obtained similarly, without any additional assumption. Hence, we can deduce the optimality conditions and derive a similar generalized derivative as in Theorem 4.4.*

5. Application to the toll-setting problem

We consider the bilevel optimization formulation of the toll-setting problem in transportation. Our aim is to show how the theory discussed in the previous section can be used to solve the problem. First, let us provide a brief description of the problem; for more details, see, e.g., [3, 19, 29]. In this problem, the upper level decision-maker corresponds to a road authority or the owner of a highway system which is allowed to set tolls on a subset of the links of the network. As for the lower level decision-maker, it is represented by the collection of network users minimizing their travel cost. It is assumed that for a toll selection from the road authority, the network users behave selfishly by trying to minimize their own travel cost. Therefore, the road authority desiring to maximize his/her revenues from tolls will solve the bilevel program

$$
\begin{aligned}
&\max_{T_a} \quad \textstyle\sum_{a\in\mathcal{A}_1} T_a y_a \\
&\text{s.t.} \quad \forall a \in \mathcal{A}_1 : \;\; T_a \geq l_a, \\
&\qquad \min_{y_a} \;\; \textstyle\sum_{a\in\mathcal{A}_1}(c_a + T_a)y_a + \sum_{a\in\mathcal{A}_2} c_a y_a \\
&\qquad \forall i \in \mathcal{N},\, \forall (j,k) \in \mathcal{O_D} : \textstyle\sum_{a\in i^+} y_a^{jk} - \sum_{a\in i^-} y_a^{jk} = \begin{cases} 1 & \text{if } i = k_1, \\ -1 & \text{if } i = k_2, \\ 0 & \text{otherwise,} \end{cases} \\
&\qquad \forall a \in \mathcal{A} : \;\; y_a = \textstyle\sum_{(j,k)\in\mathcal{O_D}} d^{jk} y_a^{jk}, \\
&\qquad \forall a \in \mathcal{A},\;\; \forall (j,k) \in \mathcal{O_D} : \quad y_a^{jk} \geq 0,
\end{aligned}
\tag{5.1}
$$

where the involved data and variables are defined as follows:

- $\mathcal{A}_1$: subset of the tolled links,
- $\mathcal{A}$: set of the links,
- $\mathcal{A}_2$: $\mathcal{A} \setminus \mathcal{A}_1$,
- T_a: toll for link $a \in \mathcal{A}_1$,
- c_a: (fixed) travel cost for link $a\,(a \in \mathcal{A})$, exclusive of toll,
- $\mathcal{N}$: set of nodes,
- i^+: set of links exiting from node $i \in \mathcal{N}$,
- i^-: set of links ending at node $i \in \mathcal{N}$,
- $\mathcal{O_D}$: set of origin-destination node pairs,
- y_a^{jk}: traffic flow from origin j to destination k on link a, $(j,k) \in \mathcal{O_D}$, $a \in \mathcal{A}$,
- d^{jk}: proportion of traffic flow demand between origin j and destination k,
- l_a: lower bound on toll for link $a\,(a \in A_1)$,
- y_a: traffic flow for link $a \in \mathcal{A}$.

Problem (5.1) can be written as

$$
\begin{aligned}
&\max_{x_a} \quad \textstyle\sum_{a\in\mathcal{A}_1}(x_a - c_a)y_a \\
&\text{s.t.} \quad \forall a\in\mathcal{A}_1: \quad x_a \geq l_a + c_a, \quad \forall a\in\mathcal{A}_2: \quad x_a = c_a, \\
&\qquad \min_{y_a} \quad \textstyle\sum_{a\in\mathcal{A}} x_a y_a \\
&\qquad \forall i\in\mathcal{N},\ \forall (j,k)\in\mathcal{O}_{\mathcal{D}}: \textstyle\sum_{a\in i^+} y_a^{jk} - \sum_{a\in i^-} y_a^{jk} = \begin{cases} 1 & \text{if}\, i=j, \\ -1 & \text{if}\, i=k, \\ 0 & \text{otherwise}, \end{cases} \\
&\qquad \forall a\in\mathcal{A}: \quad y_a = \textstyle\sum_{(j,k)\in\mathcal{O}_{\mathcal{D}}} d^{jk} y_a^{jk}, \\
&\qquad \forall a\in\mathcal{A}, \quad \forall (j,k)\in\mathcal{O}_{\mathcal{D}}: \quad y_a^{jk} \geq 0,
\end{aligned}
\tag{5.2}
$$

if we make the variable change

$$
x_a := \begin{cases} c_a + T_a & \text{if } a\in\mathcal{A}_1, \\ c_a & \text{if } a\in\mathcal{A}_2, \end{cases}
$$

in order to keep the variable pattern used so far for problem (1.2), (1.3).

Following the discussion in the previous section, the optimality conditions resulting from the application of Theorem 4.1 to problem (5.2) can be written in the form (4.3) with

$$
\frac{\partial F(\mathbf{x},\mathbf{y})}{\partial x_a} = \begin{cases} y_a & \text{if } a\in\mathcal{A}_1, \\ 0 & \text{otherwise}, \end{cases} \qquad \frac{\partial F(\mathbf{x},\mathbf{y})}{\partial y_a} = \begin{cases} x_a - c_a & \text{if } a\in\mathcal{A}_1, \\ 0 & \text{otherwise}. \end{cases}
$$

As for the feasible set of the lower level problem in (5.2), it can be rewritten as $\mathbf{A}_1\mathbf{y} = \mathbf{b}_1,\ \mathbf{y}\geq 0$ with

$$
\mathbf{A}_1 := \begin{pmatrix}
A_{j_1k_1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & A_{j_2k_2} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\
\vdots & \ddots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \dots & \mathbf{0} & A_{j_{M-1}k_{M-1}} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & A_{j_Mk_M} & \mathbf{0} \\
d^{j_1k_1}\mathbf{I} & d^{j_2k_2}\mathbf{I} & \dots & d^{j_{M-1}k_{M-1}}\mathbf{I} & d^{j_Mk_M}\mathbf{I} & -\mathbf{I}
\end{pmatrix}, \ \mathbf{y} := \begin{pmatrix} y_a^{jk} \\ y_a \end{pmatrix}_{a,j,k},
$$

$$
\mathbf{b}_1 := \begin{pmatrix} b^i_{j_sk_s} \\ \mathbf{0} \end{pmatrix}_s
$$

$$
\forall s = 1,\dots,M, \quad \forall (j_s,k_s)\in\mathcal{O}_{\mathcal{D}}, \quad \forall i\in\mathcal{N}: \quad b^i_{j_sk_s} = \begin{cases} 1 & \text{if } i=j_s, \\ -1 & \text{if } i=k_s, \\ 0 & \text{otherwise}. \end{cases}
$$

The matrix $A_{j_s k_s}$, $s = 1, \ldots, M$ represents the incidence matrix of the graph with origin j_s and destination k_s. Here we have set $M := |\mathcal{O}_\mathcal{D}|$ to be the number of start-destination node pairs. We want to draw the attention of the reader on the fact that, for each individual graph an equation of the traffic flow condition can be omitted, i.e. a row in $\mathbf{A}_1$ belonging to the matrix A_{jk} and the corresponding component in the vector $\mathbf{b}_1$ can be deleted, since the systems of equations are over-determined for a single flow. Furthermore the number of variables can reduce if the vehicles from node j to node k are allowed to use during their journey only the road which are useful.

To get the lower level feasible set in the format in (1.2), (1.3), the above described system $\mathbf{A}_1\mathbf{y} = \mathbf{b}_1,\ \mathbf{y} \geq 0$ is completely transformed into inequality conditions. In the same vein, the matrix $\mathbf{D}$ from the upper level constraint is constructed by replacing the equality constraints by inequalities. It is obtained from the constraints $-x_a \leq -l_a - c_a$ for $a \in \mathcal{A}_1$ and $x_a \leq c_a$, $-x_a \leq -c_a$ for $a \in \mathcal{A}_2$. This means that for all $j \in \mathcal{A}$,

$$\mathbf{D}_{ij} := \begin{cases} -\delta_{ij} & \text{if } i \in \mathcal{A}_1, \\ \begin{pmatrix} \delta_{ij} \\ -\delta_{ij} \end{pmatrix} & \text{if } i \in \mathcal{A}_2, \end{cases} \qquad \mathbf{d}_i := \begin{cases} -l_a - c_a & \text{if } i \in \mathcal{A}_1, \\ \begin{pmatrix} c_a \\ -c_a \end{pmatrix} & \text{if } i \in \mathcal{A}_2. \end{cases}$$

It is worth recognizing that transforming an equality into two inequalities is not computationally ideal. Nevertheless, we do so here just to be faithful to the model (1.2), (1.3) and the corresponding analysis conducted in the previous sections.

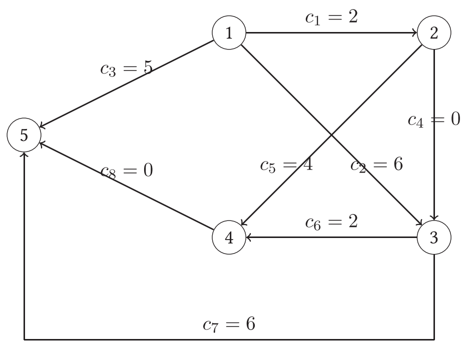


FIGURE 1. Network 1 (network with 5 nodes)

Here in the context of problem (5.2), we can easily observe that Theorems 4.5 and 4.7 are not applicable since $\nabla^2_{xx}F(\mathbf{x}, \mathbf{y})$ and $\nabla^2_{yy}F(\mathbf{x}, \mathbf{y})$ are not full column rank matrices. However, it is possible to prove that if the matrix $\mathbf{A}$ is invertible and the following inequalities hold $X^1(\bar{\mathbf{u}}) < 0$, $X^2(\bar{\mathbf{u}}) > 0$, $X^3(\bar{\mathbf{u}}) > 0$, $X^4(\bar{\mathbf{u}}) > 0$, and $X^5(\bar{\mathbf{u}}) < 0$, then the matrix H is nonsingular.

To illustrate the application of the proposed method to (5.2), we consider the examples of networks in Figure 1 and 2, which are taken from [9].

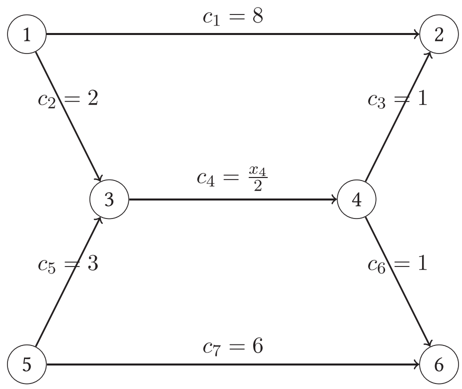


FIGURE 2. Network 2 (network with 6 nodes)

The origin-destination nodes for Network 1 (resp. Network 2) is $(1, 5)$ (resp. $\{(1, 2), (5, 6)\}$). The formulation of the toll problem related to these networks respectively leads to bilevel programming problems explicitly written as

$$\begin{array}{ll}\max\limits_{\mathbf{x},\mathbf{y}} & (x_3-5)y_3 + x_4y_4 + x_8y_8\\ \text{s.t.} & \forall a \in \mathcal{A}_2, x_a = c_a,\\ & \forall a \in \mathcal{A}_1, x_3 \geq 5, x_4 \geq 0, x_8 \geq 0,\\ & \min\limits_{y} x^\top y\\ & y_1 + y_2 + y_3 = 1,\\ & y_8 + y_7 + y_3 = 1,\\ & y_4 + y_5 - y_1 = 0,\\ & y_6 + y_7 - y_2 - y_4 = 0,\\ & y_8 - y_5 - y_6 = 0,\\ & y_i \geq 0,\ \forall i = 1, \ldots, 8,\end{array}$$

and

$$\begin{array}{ll}\max\limits_{\mathbf{x},\mathbf{y}} & \frac{1}{2}x_4y_4\\ \text{s.t.} & \forall a \in \mathcal{A}_2, x_a = c_a,\\ & x_4 = x_8,\\ & \min\limits_{y} x^\top y\\ & y_1 + y_2 = 1,\\ & y_5 + y_7 = 1,\\ & y_2 = y_3 = y_4,\\ & y_5 = y_6 = y_8,\\ & y_i \geq 0,\ \forall i = 1, \ldots, 8.\end{array}$$

We have $\mathcal{A}_1 = \{3, 4, 8\}$ and $\mathcal{A}_2 = \{1, 2, 5, 6, 7\}$ for Network 1 and $\mathcal{A}_1 = \{4\}$ and $\mathcal{A}_2 = \{1, 2, 3, 5, 6, 7\}$ for Network 2. The variables y_8 and y_4 in the mathematical formulation of the toll problem related to the Network 2 denote y_4^{12} and y_4^{56}, respectively, and the travel cost c_a for each link $(a \in \mathcal{A})$ can be obtained on each corresponding network.

To proceed, we consider the parameters: $t(1) = 0.045$, $t(2) = 0.049$, $t(3) = 0.025$, $t(4) = 0.005$, $t(5) = 0.0025$, $\varepsilon = 0.01$, $\delta = 10^{-6}$.

For these examples, the algorithm stops when the iteration index k reaches 3 in Network 1 and 10 in Network 2. As starting point for Network 1, we choose

$$\begin{array}{l}\mathbf{x}_0 = (2, 6, 5, 5, 4, 2, 6, 5),\ \ \mathbf{y}_0 = (0, 1, 0, 0, 0, 1, 0, 0),\ \ \mathbf{z}_0 = \mathbf{A}y_0 - b = \boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_3,\\ \mathbf{r}_0 = \boldsymbol{\lambda}_4 = \boldsymbol{\lambda}_7 = \mathbf{0}_{l\times 1},\ \ \mathbf{s}_0 = \mathbf{e} = \boldsymbol{\lambda}_5,\ \ \boldsymbol{\lambda}_1 = |\mathbf{D}\mathbf{x}_0 - \mathbf{d}|,\ \ \boldsymbol{\lambda}_6 = \mathbf{0}_{n\times 1},\end{array}$$

and for $\alpha = 0.45$, we obtain $F(\mathbf{x}^*, \mathbf{y}^*) = -7.0346$ and $f(\mathbf{x}^*, \mathbf{y}^*) = 12.9927$, while the known values in the literature are $F(\mathbf{x}^*, \mathbf{y}^*) = -7$ and $f(\mathbf{x}^*, \mathbf{y}^*) = 12$, see [8].

For Network 2, we choose

$$\mathbf{x}_0 = (8,2,1,0,3,1,6,0), \ \ \mathbf{y}_0 = \mathbf{0}_{n\times 1}, \ \ \mathbf{z}_0 = \mathbf{A}y_0 - b = \boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_3,$$
$$\mathbf{r}_0 = \boldsymbol{\lambda}_4 = \boldsymbol{\lambda}_7 = \mathbf{0}_{l\times 1}, \ \ \mathbf{s}_0 = \mathbf{e} = \boldsymbol{\lambda}_5, \ \ \boldsymbol{\lambda}_6 = \mathbf{0}_{n\times 1},$$

as a starting point and for $\alpha = 4.791$, we get the values $F(\mathbf{x}^*, \mathbf{y}^*) = -11.7003$ and $f(\mathbf{x}^*, \mathbf{y}^*) = 34.0099$. The values known in the literature are as follows: $F(\mathbf{x}^*, \mathbf{y}^*) = -11$ and $f(\mathbf{x}^*, \mathbf{y}^*) = 34$; see [8].

As you can observe the choice of the parameters t and α have been done randomly and the one of starting points has been deduced from the one of [8]. Selecting any parameters t and α would work in terms of solving the system of equations (4.3). However, as you would be aware, solving this system does not necessary ensure that we get an optimal (local/global) of our original bilevel program. Hence, to select the initialization framework of the algorithm that can ensure that we get an optimal solution of problem (1.2)–(1.3), some care would be needed. The sensitivity/robustness of our method in terms of the initialization framework will be part of future work that we are considering for this method.

6. Conclusion

We considered a bilevel problem involving a linear lower level problem whose objective function is influenced by leader decision variables and the lower-level constraints are independent of the upper-level decisions. Using the KKT conditions, we proposed an equivalent penalty single-level problem to which we applied a smoothed-regularized Newton method that yields stationary points of the problem which are potentially locally optimal. We provided sufficient conditions ensuring local convergence of the method and applied the globalized algorithm to the toll-setting problem.

References

[1] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons 2013.

[2] F. Benita and P. Mehlitz, *Bilevel optimal control with final-state-dependent finite-dimensional lower level*, SIAM J. Optim. **26** (2016), 718–752.

[3] L. Brotcorne, M. Labbé, P. Marcotte and G. Savard *A bilevel model and solution algorithm for a freight tariff-setting problem*, Transportation Science **34** (2000), 289–302.

[4] A. Cabot, *Proximal point algorithm controlled by a slowly vanishing term: applications to hierarchical minimization*, SIAM J. Optim. **15** (2005), 555–572.

[5] H. I. Calvete and Galé, *Algorithms for linear bilevel optimization*, in: Dempe, S., Zemkoho, A. (eds.), Bilevel optimization: advances and next challenges Springer, 2020, pp. 293–312.

[6] X. Chen, Z. Nashed and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal. **38** (2000), 1200–1216.

[7] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM 1990.

[8] B. Colson, P. Marcotte and G. Savard, *A trust-region method for nonlinear bilevel programming: algorithm and computational experience*, Comput. Optim. Appl. **30** (2005), 211–227.

[9] B. Colson, *Bilevel programming with approximation methods: software guide and test problems*, Department of Mathematics, Facultés Universitaires Notre Dame de la Paix 2002.

[10] T. De Luca, F. Facchinei and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Mathematical programming **75** (1996), 407–439.

[11] S. Dempe, N. Dinh and J. Dutta, *Optimality conditions for a simple convex bilevel programming problem*, in: Variational analysis and generalized differentiation in optimization and control, R. S. Burachikm, J. Yao (eds.), Springer, 2010, pp. 149–161.
[12] S. Dempe *Foundations of Bilevel Programming*, Kluver Academic Publishers 2002.
[13] S. Dempe and J. Dutta, *Is bilevel programming a special case of a mathematical program with complementarity constraints?*, Math. Program. **131** (2012), 37–48.
[14] S. Dempe, *Bilevel optimization: theory, algorithms, applications and a bibliography*, Bilevel optimization. Springer, Cham, 2020, pp. 581–672.
[15] S. Dempe and S. Franke, *Solution algorithm for an optimistic linear Stackelberg problem*, Comput. Oper. Res. **41** (2014), 277–281.
[16] S. Dempe and F. M. Kue, *Solving discrete linear bilevel optimization problems using the optimal value reformulation*, J. Global Optim. **68** (2017), 255–277.
[17] S. Dempe and A. Zemkoho, *Bilevel Optimization: Advances and Next Challenges*, Springer, 2020.
[18] S. Dempe and A. B. Zemkoho, *On the Karush–Kuhn–Tucker reformulation of the bilevel optimization problem*, Nonlinear Anal. **75** (2012), 1202–1218.
[19] S. Dewez and Labbé, *New formulations and valid inequalities for a bilevel pricing problem*, Oper. Res. Lett. **36** (2008), 141–149.
[20] F. Facchinei and J. Soares, *A new merit function for nonlinear complementarity problems and a related algorithm*, SIAM J. Optim. **7** (1997), 225–247.
[21] M. C. Ferris and O. L. Mangasarian, *Finite perturbation of convex programs*, Appl. Math. Optim. **23** (1991), 263–273.
[22] A. Fischer, *A special Newton-type optimization method*, Optimization **24** (1992), 269–284.
[23] A. Fischer and H. Jiang, *Merit functions for complementarity and related problems: a survey*, Comput. Optim. Appl. **17** (2000), 159–182.
[24] A. Fischer, A. B. Zemkoho and S. Zhou, *Semismooth Newton-type method for bilevel optimization: global convergence and extensive numerical experiments*, arXiv preprint arXiv:1912.07079 2019.
[25] J. Fliege, A. Tin and A. Zemkoho, *Gauss-Newton-type methods for bilevel optimization*, Comput. Optim. Appl. **78** (2021), 793–824.
[26] A. Galántai, *Properties and construction of NCP functions*, Comput. Optim. Appl. **52** (2012), 805–824.
[27] M. Hintermüller, *Semismooth Newton methods and applications*, Oberwolfach-Seminar Notes on "Mathematics of PDE-Constrained Optimization", Mathematisches Forschungsinstitut in Oberwolfach 2010.
[28] M. Hintermüller, K. Ito and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim. **13** (2002), 865–888.
[29] M. Labbé, P. Marcotte and G. Savard, *A bilevel model of taxation and its application to optimal highway pricing*, Management Science, **44** (1998), 1608–1622.
[30] G. Lin, M. Xu and J. Y. Jane, *On solving simple bilevel programs with a nonconvex lower level program*, Math. Programm. **144** (2014), 277–305.
[31] O. L. Mangasarian and R. Meyer, *Nonlinear perturbation of linear programs*, SIAM J. Control Optim. **17** (1979), 745–752.
[32] O. L. Mangasarian and J. S. Pang, *Exact penalty for mathematical programs with linear complementarity constraints*, Optimization **42** (1997), 1–8.
[33] P. Mehlitz, *Contributions to complementarity and bilevel programming in Banach spaces*, 2017.
[34] P. Mehlitz, L. I. Minchenko and A. B. Zemkoho, *A note on partial calmness for bilevel optimization problems with linear structures at the lower level*, Optim. Lett. in press, doi.org/10.1007/s11590-020-01636, 2020.
[35] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim. **15** (1977), 959–972.
[36] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Springer, 2004.
[37] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res. **18** (1993), 227–244.

[38] H. Qi and D. Sun, *A quadratically convergent Newton method for computing the nearest correlation matrix*, SIAM J. Matrix Anal. Appl. **28** (2006), 360–385.
[39] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programm. **58** (1993), 353–367.
[40] S. M. Robinson, *Generalized equations and their solutions, part II: applications to nonlinear programming*, in: Optimality and stability in mathematical programming, M. Guignard (ed.), Springer, 1982, pp. 200–221.
[41] R. T. Rockafellar, *Convex Analysis* Princeton University Press. 1970.
[42] Y. Shehu, P. T. Vuong and A. Zemkoho, *An inertial extrapolation method for convex simple bilevel optimization*, Optim. Methods Softw. **36** (2021), 1–19.
[43] M. Solodov, *An explicit descent method for bilevel convex optimization*, J. Convex Anal. **14** (2007): 227.
[44] 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), 481–507.
[45] J. J. Ye and D. L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization **33** (1995), 9–27.
[46] J. Yin and Q. Li, *A semismooth Newton method for support vector classification and regression*, Comput. Optim. Appl. **73** (2019), 477–508.
[47] A. Zemkoho and S. Zhou, *Theoretical and numerical comparison of the Karush–Kuhn–Tucker and value function reformulations in bilevel optimization*, Comput. Optim. Appl. **78** (2021), 625–674.

Manuscript received June 18 2022
revised August 27 2022

F. Mefo Kue
Department of Mathematics, University of Siegen, Walter-Flex-Str. 3, D-57068 Siegen, Germany
Department of Mathematics, Higher Training College, University of Yaoundé I, P.O. Box 812, Yaoundé, Cameroon
AIMS-Cameroon Research Center, P.O. Box 608, Crystal Gardens, Limbe, Cameroon
E-mail address: `floriane.kue@aims-cameroon.org`

T. Raasch
Department of Mathematics, University of Siegen, Walter-Flex-Str. 3, D-57068 Siegen, Germany
E-mail address: `raasch@mathematik.uni-siegen.de`

A. B. Zemkoho
School of Mathematical Sciences, University of Southampton, SO17 1BJ Southampton, United Kingdom
E-mail address: `a.b.zemkoho@soton.ac.uk`