Stability analysis for two-level value functions and application to numerically solve a pessimistic bilevel program*

Alain B. Zemkoho

School of Mathematical Sciences,
University of Southampton, UK

Abstract

Some stability results are presented for a two-level value function, which is the optimal value function of a parametric optimization problem constrained by the optimal solution set of another parameteric optimization problem. It is then shown how to use these stability results to write down (and subsequently compute) stationary points for a pessimistic bilevel optimization problem. It is also demonstrated how the corresponding Scholtes relaxation–based numerical process can be used to calculate local and global–type optimal solutions for the pessimistic bilevel program if one is equipped with a solver for minmax programs involving coupled inner constraints.

Keywords: Two-level value function; Pessimistic bilevel optimization; Karush-Kuhn-Tucker reformulation; Necessary optimality conditions; Scholtes relaxation

Mathematics Subject Classification (2020): 90C31; 90C33; 90C46; 90C47

1. Introduction

Consider the functions $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$ (with the dimensions n, m, $q \in \mathbb{N}$) that are used to introduce the parametric optimization problem

$$\min_y \ \{f(x, y) : \ y \in Y(x)\}, \tag{1.1}$$

where the set-valued mapping $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is defined by $Y(x) := \{y \in \mathbb{R}^m : \ g(x, y) \leq 0\}$. Let $S(x)$ denote the optimal solution of problem (1.1) for a fixed parameter $x \in \mathbb{R}^n$. Subsequently, consider another function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ that associated to the parameteric optimization problem (1.1) enables the definition of the optimal value function

$$\varphi_p(x) := \max_y \ \{F(x, y) : \ y \in S(x)\}. \tag{1.2}$$

This function describes the optimal values of a special class of parametric optimization problem where, unlike in (1.1), the feasible set is defined by the optimal solution set of another parametric optimization problem. Hence, φ_p was labelled in [4] as a two-level (optimal) value function (TLVF), to reflect this hierarchical structure in the underlying optimization problem, where F represents the objective function of the upper-level problem, while f and Y correspond to the lower-level objective function and constraint set-valued mapping, respectively.

There are two primary applications for TLVFs. First, as it will be clear in Section 3, its stability properties can serve as base to solve two very complicated class of optimization problems. That is, the (original) optimistic and pessimistic bilevel optimization problems

$$\min_{x \in X} \varphi_o(x) \quad \text{and} \quad \min_{x \in X} \varphi_p(x), \tag{1.3}$$

*Invited paper for latest edition of the Encyclopedia of Optimization: https://doi.org/10.1007/978-3-030-54621-2

respectively, where $\varphi_o(x) := \min_y \{F(x,y) : \; y \in S(x)\}$ denotes the minimization version of the two-level value function and X corresponds to the upper-level feasible set. Furthermore, note that the labels "o" and "p" attached to the TLVFs in (1.3) refer to the fact that φ_o and φ_p are associated to the optimistic and pessimistic bilevel optimization problems, respectively.

Secondly, stability results for φ_o or φ_p can be directly applied to derive stability or sensitivity results for the simple (see, e.g., [10]) and standard optimistic (according to [12]) bilevel optimization problems. For more details on bilevel optimization and the wide range of its applications, interested readers are referred to the edited volume [7] with the most recent surveys on the topic.

In the next section, the (generalized) differentiation and Lipschitz continuity properties, including estimates of subdifferentials of φ_p, are presented. These results are then used in Section 3 to derive optimality conditions and a basic algorithmic framework for solving the pessimistic bilevel optimization problem. For more details on the results presented here, interested readers are referred to [4, 6, 5, 3, 2].

2. Subdifferential and Lipschitz continuity

Start here by assuming that the following is valid throughout the chapter:

$$\forall x \in \mathbb{R}^n : \; \begin{cases} f(x,.) \text{ and } \; g_j(x,.), \; j = 1,\ldots,p, \; \text{ are convex functions,} \\ \exists y_x \in \mathbb{R}^m : \; g_j(x,y_x) < 0, \; j = 1,\ldots,p. \end{cases} \tag{2.1}$$

Note that the second line corresponds to the well-known Slater constraint qualification. Additionally, it is also assumed that throughout the chapter, F (resp. f and g) is a (resp. are twice) continuously differentiable function (resp. functions).

It results that if the underlying upper-level optimization problem in (1.2) is considered in terms of global optimal solution, then under assumption (2.1), $\varphi_p(x)$ is equal to

$$\psi_p(x) := \max \{F(x,y) : \; (y,u) \in \mathcal{D}(x)\} \tag{KKT}$$

for all $x \in \mathbb{R}^n$. Here, the set-valued mapping $\mathcal{D} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m \times \mathbb{R}^q$ is given by

$$\mathcal{D}(x) := \left\{ (y,u) \in \mathbb{R}^{m+q} \middle| \; \mathcal{L}(x,y,u) = 0, \; u \geq 0, \; g(x,y) \leq 0, \; u^\top g(x,y) = 0 \right\} \tag{2.2}$$

with $\mathcal{L}(x,y,u) := \nabla_y f(x,y) + \nabla_y g(x,y)^\top u$ denoting the gradient w.r.t. y for the Lagrangian function associated to the lower-level problem (1.1). Clearly, the expression (KKT) corresponds to the Karush-Kuhn-Tucker (KKT) reformulation of the TLVF φ_p.

Observe that the underlying parametric optimization problem in (KKT) is a special class of mathematical program with complementarity constraints (MPCC). This therefore gives rise to a wide range of stability results for the function φ_p, depending on the specific transformations that could be done for the involved complementarity constraints. Focus here will mainly be on a result associated to the application of the Clarke subdifferential to *min* operator appearing in the following reformulation of $\mathcal{D}$ (2.2) introduced by [9]:

$$\mathcal{D}(x) := \left\{ (y,u) \in \mathbb{R}^{m+q} \middle| \; \mathcal{L}(x,y,u) = 0, \; \min\{u_j, \; -g_j(x,y)\} = 0, \; j = 1,\ldots,q \right\}. \tag{2.3}$$

This will lead to Clarke (or C–for short)-type upper estimates for the subdifferential for φ_p, and their construction is based on the following C-type qualification conditions. For a point $(\bar{x},\bar{y},\bar{u})$ such that $(\bar{y},\bar{u}) \in \mathcal{D}(\bar{x})$, the C-qualification conditions that are defined by

$$(\beta,\gamma) \in \Lambda^{ec}(\bar{x},\bar{y},\bar{u},0) \implies \beta = 0, \; \gamma = 0, \tag{A_1^c}$$

$$(\beta,\gamma) \in \Lambda_y^{ec}(\bar{x},\bar{y},\bar{u},0) \implies \nabla_x \mathcal{L}(\bar{x},\bar{y},\bar{u})^\top \beta + \nabla_x g(\bar{x},\bar{y})^\top \gamma = 0, \tag{A_2^c}$$

$$(\beta,\gamma) \in \Lambda_y^{ec}(\bar{x},\bar{y},\bar{u},0) \implies \beta = 0, \; \gamma = 0 \tag{A_3^c}$$

with the C-multiplier set $\Lambda^{ec}(\bar{x},\bar{y},\bar{u},0)$ resulting from setting $v = 0$ in the formula

$$\Lambda^{ec}(\bar{x},\bar{y},\bar{u},v) := \left\{ (\beta,\gamma) \in \mathbb{R}^{m+q} \middle| \begin{array}{l} \nabla_y g_\nu(\bar{x},\bar{y})\beta = 0, \;\; \gamma_\eta = 0 \\ \forall i \in \theta : \;\; \gamma_i \nabla_y g_i(\bar{x},\bar{y})\beta \geq 0 \\ v + \nabla_{x,y}\mathcal{L}(\bar{x},\bar{y},\bar{u})^\top \beta + \nabla g(\bar{x},\bar{y})^\top \gamma = 0 \end{array} \right\}, \tag{2.4}$$

while $\Lambda_y^{ec}(\bar{x}, \bar{y}, \bar{u}, 0)$ can be obtained by replacing the derivatives w.r.t. (x, y) in the last equation in the set (2.4) by the derivatives of the same functions w.r.t. y only. Obviously, condition (A_3^c) implies that (A_1^c) holds, and also implies the fulfillment of (A_2^c). Note that in the definitions of (A_1^c), (A_2^c), (A_3^c), and (2.4), the index sets η, θ, and ν, respectively, can be written as follows:

$$\left.\begin{array}{lllll} \eta & := & \eta(\bar{x}, \bar{y}, \bar{u}) & := & \{i \in \{1, ..., q\} \mid\ \bar{u}_i = 0,\ g_i(\bar{x}, \bar{y}) < 0\} \\ \theta & := & \theta(\bar{x}, \bar{y}, \bar{u}) & := & \{i \in \{1, ..., q\} \mid\ \bar{u}_i = 0,\ g_i(\bar{x}, \bar{y}) = 0\} \\ \nu & := & \nu(\bar{x}, \bar{y}, \bar{u}) & := & \{i \in \{1, ..., q\} \mid\ \bar{u}_i > 0,\ g_i(\bar{x}, \bar{y}) = 0\} \end{array}\right\}. \tag{2.5}$$

It might also be useful to note that in (2.4), $\nabla_y g_\nu(\bar{x}, \bar{y})\beta := (\nabla_y g_i(\bar{x}, \bar{y})\beta)_{i\in\nu}$ and $\gamma_\eta := (\gamma_i)_{i\in\eta}$.

To state the main result of this section, the optimal solution set-valued mapping associated to the underlying upper-level parametric optimization problem in (KKT),

$$\mathcal{S}_p(x) := \min_y \ \{F(x, y) :\ (y, u) \in \mathcal{D}(x)\}, \tag{2.6}$$

needs to satisfy the inner semicontinuity property, which will be said to hold at $(\bar{x}, \bar{y}) \in \text{gph}\mathcal{S}_p$ (i.e., $\bar{y} \in \mathcal{S}_p(\bar{x})$) if for every sequence $x^k \to \bar{x}$, there is a sequence $y^k \in \mathcal{S}_p(x^k)$ that converges to $\bar{y}$.

Theorem 1 ([4]). *Let the set-valued mapping $\mathcal{S}_p$ (2.6) be inner semicontinuous at $(\bar{x}, \bar{y}) \in gph\mathcal{S}_p$ and let (A_1^c) hold at $(\bar{x}, \bar{y})$. Then the subdifferential of φ_p can be estimated by*

$$\partial\varphi_p(\bar{x}) \subset \bigcup_{(\beta,\gamma)\in\Lambda_y^{ec}(\bar{x},\bar{y},\bar{u})} \left\{\nabla_x F(\bar{x}, \bar{y}) + \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \beta + \nabla_x g(\bar{x}, \bar{y})^\top \gamma\right\}. \tag{2.7}$$

Furthermore, φ_p is Lipschitz continuous around $\bar{x}$ if (A_2^c) is also satisfied at $(\bar{x}, \bar{y}, \bar{u})$.

Here, $\Lambda_y^{ec}(\bar{x}, \bar{y}, \bar{u}) := \Lambda_y^{ec}(\bar{x}, \bar{y}, \bar{u},\ \nabla_y F(\bar{x}, \bar{y}))$ and $\partial\varphi_p$ represents the concept of basic or Mordukhovich subdifferential, which can be obtained as the Painlevé–Kuratowski outer/upper limit

$$\partial\varphi_p(\bar{x}) := \left\{\vartheta \in \mathbb{R}^n :\ \exists x^k \to \bar{x},\ \vartheta^k \to \vartheta \ \text{ with } \ \vartheta^k \in \hat{\partial}\varphi_p(x^k) \text{ as } k \to \infty\right\}$$

of the the Fréchet/regular subdifferential of φ_p defined by

$$\hat{\partial}\varphi_p(\bar{x}) := \left\{\vartheta \in \mathbb{R}^n :\ \liminf_{x\to\bar{x}} \frac{\varphi_p(x) - \varphi_p(\bar{x}) - \langle\vartheta, x - \bar{x}\rangle}{\|x - \bar{x}\|}\right\}.$$

Recall that if the function φ_p is convex near $\bar{x}$, then $\partial\varphi_p(\bar{x})$ coincides with its subdifferential in the sense of convex analysis. Also, if φ_p is continuously differential at $\bar{x}$, then $\partial\varphi_p(\bar{x}) = \{\nabla\varphi_p(\bar{x})\}$. Conversely, if $\partial\varphi_p(\bar{x})$ is single-valued, the function φ_p is strictly differentiable. Obviously, based on (2.7), φ_p will be strictly differentiable if $\Lambda_y^{ec}(\bar{x}, \bar{y}, \bar{u}) = \{(\beta, \gamma)\}$. For more details on these concepts, interested readers are referred to [8].

Note that the inner semincontinuity requirement on $\mathcal{S}_p$ is strong and closely related to the set-valued mapping concept of lower semicontinuity [1]. This can be weakened to require, for example, that $\mathcal{S}_p$ be uniformly bounded around $\bar{x}$ (i.e., $\mathcal{S}_p(x) \subset B$ for all x in some neighborhood of $\bar{x}$, and with B being a bounded subset of $\mathbb{R}^m \times \mathbb{R}^q$). However, under such a weak assumption, one gets a much larger upper estimate for the subdifferential of φ_p in (2.7). It might be worth mentioning that much sharper upper estimates for $\partial\varphi_p(\bar{x})$ in (2.7) in terms of corresponding M- and S-type multiplier sets can be obtained, provided that M-qualification conditions and MPEC–linear independence constraint qualification, respectively, are satisfied. More details on such results can be found in [4, 5].

3. Pessimistic bilevel optimization

In this section, the pessimistic bilevel optimization problem

$$\min_{x\in\mathbb{R}^n} \varphi_p(x) \tag{P$_p$}$$

with φ_p given in (1.2), but without any upper-level constraint, is considered. Note however that having the upper-level feasible set as $X := \mathbb{R}^n$ is just for simplification of the presentation, as accommodating such constraints is relatively straightforward. As starting point, Theorem 1 will be used to derive necessary optimality conditions for problem (P_p). Note that a point $\bar{x} \in \mathbb{R}^n$ will be said to be a local optimal solution for (P_p) if there is a neighborhood U of $\bar{x}$ such that

$$\forall x \in U: \ \varphi_p(\bar{x}) \leq \varphi_p(x). \tag{3.1}$$

Similarly, $\bar{x} \in \mathbb{R}^n$ will be said to be globally optimal for (P_p) if condition (3.1) holds with $U = \mathbb{R}^n$.

Considering the reformulation of the TLVF φ_p in (KKT), the C-stationarity concept of problem (P_p) will be the main focus of the analysis here.

Definition 3.1. A point $\bar{x}$ will be said to be C-stationary point for problem (P_p) if there exists a vector $(\bar{y}, \bar{u}, \beta, \gamma)$ such that the following relationships are satisfied:

$$(\bar{x}, \bar{y}, \bar{u}) \in \operatorname{gph} \mathcal{S}_p, \tag{3.2}$$

$$\nabla F(\bar{x}, \bar{y}) + \sum_{l=1}^{m} \beta_l \nabla \mathcal{L}_l(\bar{x}, \bar{y}, \bar{u}) + \sum_{i=1}^{q} \gamma_i \nabla g_i(\bar{x}, \bar{y}) = 0, \tag{3.3}$$

$$\forall i \in \nu: \ \sum_{l=1}^{m} \beta_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) = 0, \ \ \forall i \in \eta: \ \gamma_i = 0, \tag{3.4}$$

$$\forall i \in \theta: \ \ \gamma_i \sum_{l=1}^{m} \beta_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) \geq 0. \tag{3.5}$$

Here, the index sets η, θ, and ν are respectively defined in (2.5).

Theorem 2 ([5])**.** *Let the point $\bar{x}$ be a local optimal solution for problem (P_p) and assume that the set-valued mapping $\mathcal{S}_p$ (2.6) is inner semicontinuous at $(\bar{x}, \bar{y}, \bar{u}) \in gph\mathcal{S}_p$, while the conditions (A_1^c) and (A_2^c) also hold at $(\bar{x}, \bar{y}, \bar{u})$. Then $\bar{x}$ is C-stationary.*

Note that assumption (2.1) is needed for the fulfillment of this result. However, as it is assumed to hold throughout the chapter, we do not not explicitly state this in Theorem 2.

An interesting question that could be asked is how can a C-stationary point for problem (P_p) be computed in practice. Considering the combinatorial nature of the system (3.2)–(3.5), especially with the involvement of the index sets in (2.5), it would be quite a tricky task to directly compute such points. However, this can be done indirectly, by instead computing the stationarity points of the following associated Scholtes relaxation problem:

$$\min_{x \in \mathbb{R}^n} \ \psi_p^t(x) := \max_{(y,u) \in \mathcal{D}^t(x)} F(x, y), \tag{KKT_t}$$

Here, $\mathcal{D}^t$ for $t > 0$, represents the Scholtes perturbation of the set-valued mapping $\mathcal{D}$ in (2.2),

$$\mathcal{D}^t(x) := \left\{(y, u) \in \mathbb{R}^{m+q} \,|\, \mathcal{L}(x, y, u) = 0, \ \ u \geq 0, \ g(x, y) \leq 0, \ -u_i g_i(x, y) \leq t, \ i = 1, ..., q \right\}, \tag{3.6}$$

which is assumed to be nonempty throughout the chapter. For a fixed $t > 0$, x^t will be said to be a stationary point for (KKT_t) if there exists a vector $(y^t, u^t, \beta^t, \gamma^t, \mu^t, \delta^t)$ such that

$$(x^t, y^t, u^t) \in \operatorname{gph} \mathcal{S}_p^t, \tag{3.7}$$

$$\nabla F(x^t, y^t) + \nabla \mathcal{L}(x^t, y^t, u^t)^\top \beta^t - \sum_{i=1}^{q} \left(\gamma_i^t - \delta_i^t u_i^t\right) \nabla g_i(x^t, y^t) = 0, \tag{3.8}$$

$$\forall i = 1, \ldots, q: \ \ \nabla_y g_i(x^t, y^t) \beta^t + \mu_i^t + \delta_i^t g_i(x^t, y^t) = 0, \tag{3.9}$$

$$\forall i = 1, \ldots, q: \ \ \gamma_i^t \geq 0, \ \ g_i(x^t, y^t) \leq 0, \ \ \gamma_i^t g_i(x^t, y^t) = 0, \tag{3.10}$$

$$\forall i = 1, \ldots, q: \ \ \mu_i^t \geq 0, \ \ u_i^t \geq 0, \ \ \mu_i^t u_i^t = 0, \tag{3.11}$$

$$\forall i = 1, \ldots, q: \ \ \delta_i^t \geq 0, \ \ -u_i^t g_i(x^t, y^t) \leq t, \ \ \delta_i^t \left(u_i^t g_i(x^t, y^t) + t\right) = 0. \tag{3.12}$$

These necessary optimality conditions can be formally derived from problem (KKT_t) as follows.

Theorem 3 ([11]). *For a given $t > 0$, let x^t be a local optimal solution for problem* (KKT$_t$). *Assume that the set-valued mapping $\mathcal{S}_p^t$ is inner semicontinous at $(x^t, y^t, u^t) \in \operatorname{gph} \mathcal{S}_p^t$ and the following qualification condition holds at the point (x^t, y^t, u^t):*

$$\left.\begin{array}{r}\nabla_y \mathcal{L}(x^t, y^t, u^t)^\top \beta^t - \sum_{i=1}^{q} \nabla_y g_i(x^t, y^t)\,(\gamma_i^t - \delta_i^t u_i^t) = 0\\ \nabla_y g_i(x^t, y^t)\beta^t + \mu_i^t + \delta_i^t g_i(x^t, y^t) = 0\\ \gamma^t \geq 0,\ \ \delta^t \geq 0,\ \ \mu^t \geq 0\\ \mu_i^t u_i^t = 0,\quad \gamma_i^t g_i(x^t, y^t) = 0,\ \ \delta_i^t\,(u_i^t g_i(x^t, y^t) + t) = 0\end{array}\right\} \Longrightarrow \left\{\begin{array}{l}\beta^t = 0_m,\\ \\ \delta^t = \gamma^t = \mu^t = 0_q.\end{array}\right. \tag{3.13}$$

Then x^t is a stationary point for problem (KKT$_t$).

Algorithm 1 Scholtes relaxation method for pessimistic bilevel optimization

---

Step 0: Choose $t^0 > 0$ and set $k := 0$.
Step 1: Compute a stationary point of problem (KKT$_t$) for $t := t^k$.
Step 2: Select $0 < t^{k+1} < t^k$, set $k := k + 1$, and go to Step 1.

To compute C-stationarity points based on the stationarity conditions of (KKT$_t$), the corresponding version of the Scholtes relaxation method provided in Algorithm 1 can be used. For the convergence of this result, the M–counterparts of the qualifications (A_1^c) and (A_2^c), denoted by (A_1^m) and (A_2^m), respectively, could be used. These conditions can be obtained by simply replacing condition (3.5) in the C-multipliers set (2.4) by the following M-one:

$$\forall i \in \theta: \quad (\gamma_i < 0 \wedge \sum_{l=1}^{m} \beta_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) < 0) \vee \gamma_i \sum_{l=1}^{m} \beta_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) = 0.$$

One can easily check that (A_1^m) and (A_2^m) are weaker than (A_1^c) and (A_2^c), respectively.

Theorem 4 ([3]). *Let $t_k \downarrow 0$ and $(x^k)_k$ be a sequence such that x^k is a stationary point of* (KKT$_t$) *for $t := t^k$ with $x^k \to \bar{x}$ as $k \to \infty$ such that $\mathcal{S}_p(\bar{x})$ is nonempty and compact. Let (A_1^m) and (A_2^m) hold at all $(\bar{x}, y, u)$ with $(y, u) \in \mathcal{S}_p(\bar{x})$. Furthermore, let*

$$\lim_{k\to\infty} e\left(\mathcal{S}_p^{t_k}(x^k),\ \mathcal{S}_p(\bar{x})\right) = 0. \tag{3.14}$$

Then $\bar{x}$ is a C-stationary point.

Under semicontinuity properties, in the sense of Hausdorff, associated to the set-valued mappings $\mathcal{D}$ and $\mathcal{D}^t$ (for $t > 0$), the fulfillment of condition (3.14) can be ensured.

Proposition 1 ([3]). *Let $t_k \downarrow 0$ and let $x^k \to \bar{x}$ as $k \to \infty$ such that $\mathcal{D}(\bar{x})$ is nonempty and compact. Assume that the set-valued mapping $x \rightrightarrows \mathcal{D}(x)$ (resp. $(t, x) \rightrightarrows \mathcal{D}^t(x)$) is lower semicontinuous at any $(\bar{x}, y, u) \in gph\,\mathcal{D}$ (resp. upper semicontinuous at $(0^+, \bar{x})$) in the sense of Hausdorff. Then the property* (3.14) *is satisfied.*

For a numerical implementation of Algorithm 1 to compute C-stationary points for (P$_p$) and its application on multiple examples, see [3]. If there is a handy tool to solve the minmax problem (KKT$_t$) globally or locally, in the sense of (3.1), Algorithm 1 could be adapted to solve problem (P$_p$) as well. In this case, problem (KKT$_t$) for $t := t^k$, would instead be solved globally or locally in Step 1 of Algorithm 1. Before stating the convergence of the algorithm in this case, it might be interesting to observe that for a fixed point x, a sequence of objective function values of problem (KKT$_t$) for $t := t_k$ can converge to that of problem (P$_p$) as $t_k \to 0$, while $k \to \infty$.

Proposition 2 ([3]). *Let $t \mapsto \psi_p^t(x)$ be upper semicontinuous at 0^+ for any $x \in \mathbb{R}^n$, and let $t_k \downarrow 0$. Then for any $x \in \mathbb{R}^n$, it holds that $\psi_p^{t_k}(x) \to \psi_p(x)$ as $k \to \infty$.*

Next is the convergence of Algorithm 1 when (KKT$_t$) is solved globally as $t \downarrow 0$.

Theorem 5 ([3]). *Let $x \mapsto \psi_p(x)$ be lower semicontinuous at $\bar{x}$ and $t \mapsto \psi_p^t(x)$ be upper semicontinuous at 0^+ for all $x \in \mathbb{R}^n$. Furthermore, let $t_k \downarrow 0$ and $(x^k)_k$ be a sequence such that x^k is a global optimal solution of problem (KKT_t) for $t := t_k$. If the sequence $(x^k)_k$ admits a subsequence converging to the point $\bar{x}$ as $k \to \infty$, then $\bar{x}$ is a global optimal solution of (P_p).*

Considering the nonconvexity of problem (P_p), it is more likely that in practice, a scheme to solve the problem would only compute local optimal solutions. Hence, below is a result where iteratively solving (KKT_t) can ensure that a local optimal solution for (P_p) is obtained.

Theorem 6 ([3]). *Let $x \mapsto \psi_p(x)$ be lower semicontinuous at $\bar{x}$, $\bar{r} > 0$ be such that $(t, u) \mapsto \psi_p^t(u)$ is upper semicontinuous at $(0, x)$ for any $x \in B(\bar{x}, \bar{r})$, let $t_k \downarrow 0$, $r^k > 0$, and $(x^k)_k$ be a sequence of optimal solutions of (KKT_t) for $t := t^k$ in $B(x^k, r^k)$; i.e.,*

$$\psi_p^{t_k}(x^k) \leq \psi_p^{t_k}(x) \quad \forall x \in B(x^k, r^k).$$

If $\bar{x}$ is a cluster point of $(x^k)_k$ and $\liminf_{k\to\infty} r^k > 0$, then $\bar{x}$ is a local optimal solution of (P_p).

4. Conclusions

In this chapter, a stability analysis for the TLVF φ_p (1.2) is conducted and is is shown how this can be used to solve the pessimistic bilevel optimization problem (P_p). For numerical experiments demonstrating the potential of the Scholtes relaxation method in Algorithm 1, see [3]. It might be useful to mention that the results here can be extended to other types of relaxations for complementarity constraints involved in (2.2); see [2]. The focus here has only been on the use of the KKT reformulation to address φ_p. Other reformulations are possible, with a prominent one being the lower-level (optimal) value reformulation, which consists to observe that $y \in S(x)$ is equivalent to $g(x, y) \leq 0$ and $f(x, y) \leq \varphi(x)$, where φ is the optimal value function associated to the parametric optimization problem in (1.1); for more details on this approach, see, e.g., [4, 5]. For possible extensions of the discussion in this chapter to nonsmooth functions, see [6].

References

[1] Bernd Bank, Jürgen Guddat, Diethard Klatte, Bernd Kummer, and Klaus Tammer. *Non-linear parametric optimization*, volume 58. Walter de Gruyter GmbH & Co KG, 1982.

[2] Imane Benchouk, Lateef Jolaoso, Khadra Nachi, and Alain Zemkoho. Relaxation methods for pessimistic bilevel optimization. arXiv preprint arXiv:2412.11416, 2024.

[3] Imane Benchouk, Lateef Jolaoso, Khadra Nachi, and Alain Zemkoho. Scholtes relaxation method for pessimistic bilevel optimization. *Set-Valued and Variational Analysis*, 33(2):10, 2025.

[4] Stephan Dempe, Boris S Mordukhovich, and Alain B Zemkoho. Sensitivity analysis for two-level value functions with applications to bilevel programming. *SIAM Journal on Optimization*, 22(4):1309–1343, 2012.

[5] Stephan Dempe, Boris S Mordukhovich, and Alain B Zemkoho. Necessary optimality conditions in pessimistic bilevel programming. *Optimization*, 63(4):505–533, 2014.

[6] Stephan Dempe, Boris S Mordukhovich, and Alain B Zemkoho. Two-level value function approach to non-smooth optimistic and pessimistic bilevel programs. *Optimization*, 68(2-3):433–455, 2019.

[7] Stephan Dempe and Alain Zemkoho. Bilevel optimization. In *Springer Optimization and its Applications*, volume 161. Springer, 2020.

[8] Boris S Mordukhovich. *Variational Analysis and Applications*. Springer, 2018.

[9] Holger Scheel and Stefan Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.

[10] Yekini Shehu, Phan Tu Vuong, and Alain Zemkoho. An inertial extrapolation method for convex simple bilevel optimization. *Optimization Methods and Software*, 36(1):1–19, 2021.

[11] Alain B Zemkoho. A simple approach to optimality conditions in minmax programming. *Optimization*, 63(3):385–401, 2014.

[12] Alain B Zemkoho. Solving ill-posed bilevel programs. *Set-Valued and Variational Analysis*, 24(3):423–448, 2016.