A true single–level reformulation for pessimistic bilevel optimization*

OLIVER STEIN † AND ALAIN ZEMKOHO ‡

Abstract. We propose a single-level reformulation (SLR) for pessimistic bilevel optimization that does not rely on complementarity conditions or optimal value functions. For this reason, we refer to it as a *true* single-level reformulation (tSLR). A remarkable consequence is that this formulation can satisfy the classical linear independence constraint qualification, despite the fact that even the weaker Mangasarian–Fromovitz constraint qualification is known to systematically fail for standard single-level reformulations of both optimistic and pessimistic bilevel programs. Furthermore, the necessary optimality conditions induced by the tSLR can be expressed as a square system of equations, making them directly computable by any semismooth Newton solver. Leveraging this reformulation, we also derive the first sufficient conditions guaranteeing that a Karush–Kuhn–Tucker point of a pessimistic bilevel program is a local optimum. Preliminary numerical experiments demonstrate that algorithms based on the proposed tSLR can outperform existing approaches for pessimistic bilevel optimization. Overall, the proposed framework suggests that pessimistic bilevel programs may be considerably more tractable than previously believed and need not be inherently more difficult to solve than their optimistic counterparts.

Key words. pessimistic bilevel optimization, single-level reformulation, optimality conditions

AMS subject classifications. 90C26, 90C30, 90C46, 90C47

1 Introduction. In this paper, we consider the bilevel optimization problem

$$\text{(BOP)} \qquad \text{“}\min_x\text{”}\ F(x,y) \ \text{ s.t. } \ x \in X, \ \ y \in S_L(x) := \operatorname*{argmin}_{z\in Y(x)} \ f(x,z),$$

where the vector x (resp. $X \subseteq \mathbb{R}^n$, $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$) represents the upper-level *variable* (resp. *feasible set*, *objective function*). Similarly, y (resp. $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$) corresponds to the lower-level *variable* (resp. *objective function*), while the set-valued mapping $Y : X \rightrightarrows \mathbb{R}^m$ describes the lower-level feasible set that we set as

$$Y(x) := \{y \in \mathbb{R}^m \mid \ g(x,y) \le 0, \ \ h(x,y) = 0\} \tag{1.1}$$

for all $x \in X$. Here, g and h are vector-valued functions from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. Note that the upper-level (resp. lower-level) player is often also called the *leader* (resp. *follower*).

Overall, problem (BOP) corresponds to the upper-level problem, while the set-valued mapping $S_L : X \rightrightarrows \mathbb{R}^m$ collects all the optimal solutions of the lower-level problem

$$(LL(x)) \qquad \min_z \ f(x,z) \ \text{ s.t. } \ z \in Y(x)$$

for a given upper-level variable $x \in X$.

If the lower-level player has a unique optimal solution for all choices of the upper-level, i.e., $|S_L(x)| = 1$ (with $|.|$ denoting the cardinality of the corresponding set) for all $x \in X$, then problem (BOP) can be reduced to the optimization problem

$$\min \ \mathcal{F}(x) \ \text{ s.t. } \ x \in X \tag{1.2}$$

with real-valued objective $\mathcal{F}(x) := F(x, y(x))$ and $S_L(x) = \{y(x)\}$ for all $x \in X$. Given that the analytical expression of the function $y(\cdot) : X \to \mathbb{R}^m$ is generally unknown, (1.2) is usually referred to as the implicit function reformulation of problem (BOP). Problem (1.2) has been one of the main frameworks to develop solution algorithms to solve bilevel programs (see, e.g., [12]) and has become even more widely used recently in the context of machine learning applications [25].

If it is not possible to ensure that the lower-level problem has a unique optimal solution for all upper-level variables, then problem (BOP) is not well-posed; hence, the quotation marks on the min operator in (BOP) are commonly used to reflect the ambiguity in the upper-level minimization.

---

*This is the preprint of the paper, which is under review.

Funding: The work of the second author is partly funded by an Alexander von Humboldt Research Fellowship for Experienced Researchers within the Continuous Optimization Chair at the Institute for Operations Research, Karlsruhe Institute of Technology (KIT).

†Institute for Operations Research (IOR), Karlsruhe Institute of Technology (stein@kit.edu).

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

To make the problem mathematically tractable in this case, two interpretations to capture the nature of the interaction between the upper- and lower-level players have been widely used in the literature. The most popular one is the optimistic model

$$(P_o) \qquad \min\ \varphi_o(x) \ \text{ s.t. } \ x \in X$$

with

$$\varphi_o(x) := \min_{y\in S_L(x)} F(x,y), \tag{1.3}$$

where it is assumed that whenever the lower-level player has more than one option for a given choice of the upper-level player, they pick one that is in favor of the latter; hence, problem (P_o) captures the cooperation between both players.

If cooperation is not possible between the two players, then the leader, as a risk-averse player, will try to protect themself against potential worst choices from the follower by solving the pessimistic bilevel optimization problem

$$(P_p) \qquad \min\ \varphi_p(x) \ \text{ s.t. } \ x \in X$$

with

$$\varphi_p(x) := \max_{y\in S_L(x)} F(x,y). \tag{1.4}$$

The positions considered in problems (P_o) and (P_p) can be seen as extreme, as they reflect either a situation where there is cooperation or not, as represented by problems (P_o) and (P_p), respectively. In consideration of this, many papers have recently considered a partial cooperation model, which could be obtained by minimizing a convex combination of φ_o and φ_p under the upper-level constraint. For more details on such models; see, e.g., [1]. Also see [48] for a set-valued optimization approach to tackle problem (BOP) when $|S_L(x)| > 1$ for some $x \in X$.

Overall, in terms of approaches to tackle (BOP), the most widely used models are (1.2), especially since the recent breakthroughs in machine learning, and problem (P_o) through its standard optimistic reformulation [20]. As for the pessimistic bilevel program (P_p), it has not attracted the same level of attention as the optimistic one. However, recently, there has started to be some growing interest in the development of algorithms for problem (P_p); the next section of this paper provides a detailed overview of the state of the literature on approaches to tackle the problem. As it will become clear there, the existing approaches to tackle problem (P_p) can be organized in two main categories: (i) moving the difficult component of the problem (i.e., in particular, the inclusion $y \in S_L(x)$) to the feasible set of a constrained optimization-based transformation of the problem through a semi-infinite programming model [43] or models closely related to the standard pessimistic problem introduced in [35]; and (ii) viewing the problem (P_p) as that of effectively minimizing the two-level value function (TLVF) φ_p subject $x \in X$; see, e.g., [5, 8, 9].

Fewer works have been dedicated to the development of methods for (P_p) because it is seen as a very difficult problem class. A main challenge in solving a pessimistic bilevel problem (P_p) lies in the fact that φ_p may not be lower semicontinuous, resulting in unsolvability due to a non-attained finite infimum [5, 48]. We discuss this issue and its relation to the proposed single–level reformulation in Subsection 4.4. Hence, it is very easy to find examples of bilevel programs where problem (P_p) has no optimal solutions, while its optimistic counterpart (P_o) does have optimal solutions. In this paper, we introduce a reformulation of problem (P_p), which is likely to change this perception. In fact, leveraging the Wolfe duality associated to the intermediate-level problem

$$(IL(x)) \qquad \max_y\ F(x,y) \ \text{ s.t. } \ y \in S_L(x),$$

we construct a new reformulation of the pessimistic bilevel program with the following key features:

(A) Our reformulation of problem (P_p) is a constrained optimization problem that involves neither complementarity constraints nor an optimal value function. For this reason, we refer to it as a *true single-level reformulation* (tSLR). To the best of our knowledge, all existing single-level reformulations of (P_p) rely either on complementarity conditions or on an optimal value function; see Section 2. The same observation applies to optimistic bilevel optimization, where existing reformulations are based either on complementarity conditions or on the lower-level value function

$$\varphi_L(x) := \min_{z\in Y(x)} f(x,z). \tag{1.5}$$

(B) The tSLR introduced in this paper does not involve implicit variables; that is, variables that are part of the constraint set but not the objective function. Such variables are a major source of numerical difficulties when solving several existing SLRs of the standard optimistic bilevel optimization problem, as reported, for example, in [15].

(C) A remarkable feature of the proposed tSLR is its compatibility with classical constraint qualifications. In particular, we prove that the linear independence constraint qualification (LICQ) can hold for a broad class of pessimistic bilevel programs. This stands in sharp contrast with existing SLRs of bilevel optimization, for which the weaker Mangasarian–Fromovitz constraint qualification (MFCQ) is known to fail systematically, both in the optimistic and pessimistic settings.

(D) The proposed tSLR yields necessary optimality conditions of a standard nonlinear programming type, involving only the usual complementarity relations associated with constrained optimization. Moreover, these conditions can be expressed as a square system of equations, making it possible to directly solve them with any semismooth Newton-type algorithm. To the best of our knowledge, such a feature is not readily available for existing reformulations of either the optimistic or pessimistic bilevel optimization problems.

(E) The tSLR framework also enables the derivation of second-order sufficient optimality conditions for a class of local solutions of problem (P_p). To the best of our knowledge, these constitute the first sufficient conditions established for pessimistic bilevel optimization.

Before continuing further, note that the set-valued mapping $S_p : X \rightrightarrows \mathbb{R}^m$ collects all the optimal solutions of the intermediate-level problem ($IL(x)$) for any given $x \in X$; that is,

$$S_p(x) := \underset{y \in S_L(x)}{\arg\max}\, F(x, y). \tag{1.6}$$

To focus our attention on the main ideas, we use the following blanket assumption throughout the paper. Sufficient conditions for it in terms of the problem data may be found, e.g., in [35].

Assumption 1.1. *For all $x \in X$, it holds $S_p(x) \neq \emptyset$.*

Under Assumption 1.1 not only the intermediate-level maximal value function φ_p from (1.4) is real-valued on X but, in view of $S_p(x) \subset S_L(x)$ for all $x \in X$, this is also the case for the lower-level value function φ_L from (1.5).

1.1 Solution concepts for the pessimistic bilevel program. The first definition below introduces the basic optimal solution concept for problem (P_p) used throughout this paper. Of course, this is not the only notion of solution concept for the problem; for a detailed study of solution notions for the problem, interested readers are referred to [3].

Definition 1.2. *A point $\bar{x} \in X$ will be said to be a local optimal solution of problem (P_p) if there exists a neighborhood U of $\bar{x}$ such that*

$$\varphi_p(\bar{x}) \leq \varphi_p(x) \quad \textit{for all} \quad x \in X \cap U. \tag{1.7}$$

Similarly, $\bar{x} \in X$ will be said to be a global optimal solution for (P_p) if (1.7) holds with $U = \mathbb{R}^n$.

There is an important observation that can be made w.r.t. an optimal solution of problem (P_p). The premise for the pessimistic bilevel program (P_p) to practically make sense is that there exists at least one vector $\hat{x} \in X$ such that $|S_L(\hat{x})| > 1$. Otherwise, one arrives in the situation of problem (1.2) which may likewise be interpreted as an optimistic bilevel optimization problem. However, Definition 1.2 does not necessarily imply that the condition $|S_L(\hat{x})| > 1$ has to hold at a point $\hat{x}$ that is an optimal solution of (P_p). This gives rise to two sub-categories of the notion of optimal solution introduced in Definition 1.2:

Definition 1.3. *A point $\bar{x} \in X$ will be said to be a local (resp. global) optimal solution of problem (P_p) of Category I if $\bar{x}$ is locally (resp. globally) optimal in the sense of Definition 1.2 and it holds that $|S_L(\bar{x})| = 1$. Otherwise (i.e., $|S_L(\bar{x})| > 1$), the point $\bar{x}$ will be said to be a local (resp. global) optimal solution of problem (P_p) of Category II.*

Examples of optimal solutions of Categories I and II are given in Examples 1 and 2, respectively.

Example 1. *Consider the problem (P_p) with the data*

$$F(x, y) := y, \quad f(x, y) := x, \quad X := \mathbb{R}, \quad \textit{and} \quad Y(x) := [-1, 1] \cap [-x, x].$$

Clearly, the global optimal solution $\bar{x} = 0$ of (P_p) is of Category I, while this point is instead the global maximum of the corresponding optimistic bilevel program (P_o); see Fig. 1.

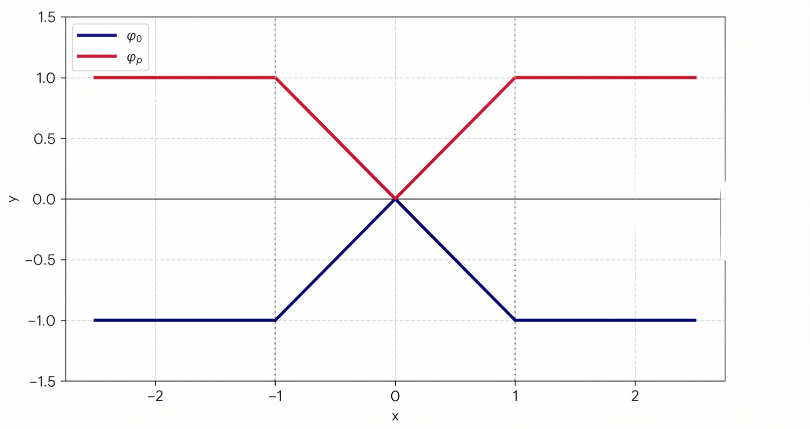


Fig. 1: Graphs of φ_o from (1.3) and φ_p from (1.4) in Example 1.

EXAMPLE 2. *In the case where*

$$\begin{aligned} &F(x,y) := -5x_1 - 8x_2 - 4y_1 + y_2, \quad f(x,y) := -y_1 - y_2, \\ &X := \left\{ x \in \mathbb{R}^2_+ \,\middle|\, x_1 + x_2 \le 6 \right\}, \\ &Y(x) := \left\{ y \in \mathbb{R}^2_+ \,\middle|\, y_1 + y_2 \le 15 - x_1 - 2x_2 \right\}, \end{aligned}$$

the point $\bar{x} = (0, 6)$ *is a global optimal solution for problem* (P_p) *of Category II, given that*

$$S_L(0,6) := \left\{ y \in \mathbb{R}^2_+ \mid y_1 + y_2 = 3 \right\}.$$

It might be worth nothing that here, the point $(6, 0)$ *is the global optimal solution for the original optimistic bilevel optimization problem.*

The importance of distinguishing these two categories of optimal solutions will be clear in Subsection 5.3, where we provide second order sufficient conditions ensuring that a Karush-Kuhn-Tucker (KKT) point of problem (P_p) is specifically a local optimal solution of Category I.

1.2 An illustrative example for the proposed tSLR. To get a flavor of the extraordinary nature of the single-level reformulation introduced in this paper, we consider a class of problem (P_p) with the upper- and lower-level objective functions $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ respectively defined by

$$F(x,y) := \frac{1}{2} \begin{bmatrix} x^\top & y^\top \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^\top & Q_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - c^\top \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad f(x,y) := (c^f)^\top y \tag{1.8}$$

with the data vectors $c := (c_x^\top, c_y^\top)^\top \in \mathbb{R}^{n+m}$ and $c^f \in \mathbb{R}^m$, the matrices $Q_{11} \in \mathbb{R}^{n\times n}$, $Q_{12} \in \mathbb{R}^{n\times m}$, as well as a negative semi-definite symmetric matrix $Q_{22} \in \mathbb{R}^{m\times m}$. Additionally, let the upper- and lower-level feasible sets be given by

$$X := \left\{ x \in \mathbb{R}^n \,\middle|\, A^G x \le b^G \right\} \quad \text{and} \quad Y(x) := \{ y \in \mathbb{R}^m \mid A^g x + B^g y \le b^g \}, \tag{1.9}$$

respectively, with the data vectors $b^G \in \mathbb{R}^r$ and $b^g \in \mathbb{R}^p$, as well as the matrices $A^G \in \mathbb{R}^{r\times n}$, $A^g \in \mathbb{R}^{p\times n}$, and $B^g \in \mathbb{R}^{p\times m}$. Our tSLR reduces in this case to the following classical-type quadratic optimization problem with linear constraints:

$$\begin{aligned} \min_{x,y,z,u,w} \quad & F(x,y) - u^\top (A^g x + B^g z - b^g) - w(y - z)^\top c^f \\ \text{s.t.} \quad & Q_{12}^\top x + Q_{22} y - (B^g)^\top u = c_y \\ & A^G x \le b^G, \;\; A^g x + B^g z \le b^g, \;\; u \ge 0, \;\; w \ge 0. \end{aligned} \tag{1.10}$$

Without any additional assumption, the x-component of any global optimal solution of this problem is a global optimal solution of the corresponding version of problem (P_p), and conversely, for any global optimal solution x of the corresponding (P_p), there are possibly many choices of the quadruple (y, z, u, w) such that (x, y, z, u, w) is globally optimal for (1.10). Obviously, (1.10) can be solved by any quadratic optimization solver. A reformulation like this (without any complementarity constraint or optimal value function) is not possible in optimistic bilevel optimization, and does not seem to have been obtained before for pessimistic bilevel optimization.

In Section 4, we introduce a general version of our tSLR model and conduct a rigorous analysis establishing global and local relationships between it and the original problem (P_p).

1.3 Organization of the paper with summary of main contributions. In the next section, we conduct a detailed survey of the different reformulations and solution methods to tackle problem (P_p). Subsequently, Section 3 covers the basic mathematical tools that will be used for the analysis in Sections 4–6. More specifically, in Section 4, we introduce the motivational background and construction process for our tSLR model, establish the global and local relationship with problem (P_p), and address insights in terms of the behavior of this reformulation and some practical implications. Subsequently, Section 5 is dedicated to the analysis of the behavior of classical MFCQ and LICQ when applied to tSLR; in particular, we construct tractable frameworks ensuring the automatic fulfillment of these CQs. Subsequently, we use the MFCQ to derive completely new necessary optimality conditions of problem (P_p) under conditions not affordable even in the context of the optimistic bilevel program. For the first time in the literature, in Section 5, we also provide sufficient conditions for local optimality of stationary points for the pessimistic bilevel programs (for Category I type). To demonstrate the potential of our tSLR reformulation, Section 6 provides some basic numerical illustrations. Our experiments on a selection of problems show that overall, our model leads to better numerical performance, in comparison to the global approaches from the existing literature, which are presented in the next section.

2 Existing reformulations and methods. Based on the lower-level optimal value function from (1.5), the lower-level solution set-valued mapping can be written as

$$S_L(x) = \{z \in Y(x) \mid f(x, z) \le \varphi_L(x)\}$$

for each $x \in X$. With the function $G_L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{p+1}$ defined by

$$G_L(x, y) := \begin{pmatrix} g(x, y) \\ f(x, y) - \varphi_L(x) \end{pmatrix},$$

we have

$$S_L(x) = Z(x) := \{y \in \mathbb{R}^m \mid G_L(x, y) \le 0,\ h(x, y) = 0\}, \tag{2.1}$$

and may therefore write the intermediate-level maximal value function from (1.4) as

$$\varphi_p(x) = \max_{y \in Z(x)} F(x, y).$$

Considering the pessimistic bilevel program (P_p), we introduce the following assumption; see next section for the definition of the Guignard constraint qualification (GCQ).

Assumption 2.1. *It holds that:*
(1) $\forall x \in X$, $F(x, \cdot)$ *is a concave function.*
(2) $\forall x \in X$, $f(x, \cdot)$ *is a convex function.*
(3) $\forall x \in X$, $g_i(x, \cdot)$ *is a convex function for* $i = 1, \dots, p$.
(4) $\forall x \in X$, $h(x, \cdot)$ *is affine linear.*
(5) $\forall x \in X$, (GCQ) *holds in* $Z(x)$ *at each* $y \in S_p(x)$.

We suppose throughout this section that Assumption 2.1 holds and all the functions F, f, g, and h are at least once continuously differentiable.

The most natural interpretation of problem (P_p) is to view it as a minmax problem, but with a very complex coupled and implicitly defined inner feasible set $S_L(x)$. Standard minmax programming algorithms already struggle with simple linear constraints (see survey in [10]). [49] addresses the implicit nature of the coupled constraint in (P_p) with the new minmax model

$$\min_{x \in X,\, z \in Y(x)}\ \max_{y \in \mathcal{S}_{\mathcal{L}}(x,z)} F(x, y), \tag{MM}$$

where the new set-valued mapping $\mathcal{S}_{\mathcal{L}}$ that replaces S_L is given by

$$\mathcal{S}_{\mathcal{L}}(x,z) := \{y \in Y(x) \mid f(x,y) \leq f(x,z)\}.$$

Let A denote the set of global optimal solutions of problem (P_p) in the sense of Definition 1.2 and similarly, let B collect all the global optimal solutions of problem (MM). Then it can be shown (see [49] or [4, Theorem 2.5]) that $A = \mathrm{proj}_x B$, where *proj* stands for the parallel projection mapping.

Observe that problem (MM) is globally equivalent to the constrained optimization problem

$$\min_{x,y,z} F(x,y) \text{ s.t. } x \in X,\ z \in Y(x),\quad y \in K(x,z) := \operatorname*{arg\,max}_{y \in \mathcal{S}_{\mathcal{L}}(x,z)} F(x,y), \tag{2.2}$$

If Assumption 2.1 holds, then problem (2.2) is global equivalent to the problem

$$\begin{aligned} \min_{x,y,z,u,v,w} \quad & F(x,y) \\ \text{s.t.} \quad & x \in X,\ z \in Y(x), \\ & \nabla_y F(x,y) - \nabla_y g(x,y)u - \nabla_y h(x,y)v - w\nabla_y f(x,y) = 0, \\ & w \geq 0,\ f(x,y) - f(x,z) \leq 0,\ w(f(x,y) - f(x,z)) = 0, \\ & u \geq 0,\ g(x,y) \leq 0,\ u^\top g(x,y) = 0, \\ & h(x,y) = 0, \end{aligned} \tag{MM-CC}$$

which is a special class of the mathematical program with equilibrium constraints (MPEC). It is important to note that due to the presence of the constraint $f(x,y) \leq f(x,z)$, which is a proxy for the lower-level value function constraint $f(x,y) \leq \varphi_L(x)$, stronger constraint qualifications such as the MFCQ will fail for problem (2.2). Hence, the much weaker Assumption 2.1(5) is more appropriate here. For some linear cases of problem (P_p), transformations of the form (MM-CC) are used for numerical methods in the articles [36, 49].

Obviously, the SLR (MM-CC), as well as the standard pessimistic and closely related ones introduced in the next subsection, bring the pessimistic bilevel program in the realm of MPECs, and therefore techniques commonly used for the classical KKT reformulation of the standard optimistic bilevel optimization problem might be possible paths for numerical methods for the problem. However, there are some key differences between reformulation (MM-CC) and the KKT reformulation for optimistic bilevel optimization. First, the feasible set of the former involves the leader's objective function, while the follower's objective function is part of the complementarity constraints. This potentially makes problem (MM-CC) much more difficult to solve. Moreover, no rigorous analysis of the relationship between this problem and the original problem (P_p) has been conducted yet, especially w.r.t. local optimal solutions. Additionally, it is unclear whether the MPEC theory w.r.t. constraint qualifications, as well as necessary and sufficient optimality conditions (in terms of S-, M-, and C-type constraint qualifications and stationarity) can work as it is the case for the KKT reformulation for the optimistic bilevel program; see, e.g., [21, 28, 27].

Another notable issue is that problem (MM-CC) involves two more *implicit variable categories* (i.e., variables present in the feasible set but not in the objective function) than the KKT reformulation of the standard optimistic bilevel program; namely, the second presence of the lower-level variable via z and the Lagrange multiplier w associated to the constraint $f(x,y) \leq f(x,z)$. As it has been extensively studied in the literature (see, e.g., [15] for analysis related to SLRs for optimistic bilevel programs), implicit variables are one of the main causes of challenges involved in solving the KKT reformulation of the standard optimistic bilevel optimization problem. As it can be sensed from the introductory example in (1.10), the tSLR introduced in this paper does not involve any implicit variable or complementarity constraint.

2.1 Standard pessimistic-type reformulations. The following *standard pessimistic* version of problem (P_p) was formally introduced in [35]:

$$\min_{x,y} F(x,y) \text{ s.t. } x \in X,\quad y \in S_p(x). \tag{SP}$$

(P_p) is globally equivalent to (SP) in the sense of [35, Proposition 4.1]. In [35], the "bridge construction" for simple bilevel programs from [34] is applied to reformulate the "vertical" bilevel connection between the intermediate ($IL(x)$) and the lower-level ($LL(x)$) as the "horizontal" connection of a generalized Nash equilibrium problem. This reformulates (SP) into the multi-follower game

$$\min_{x,y,z} F(x,y) \text{ s.t. } x \in X,\quad (y,z) \in E(x), \tag{MFG}$$

where, for $x \in X$, $E(x)$ denotes the set of generalized Nash equilibria of the two player problem that can be written as

$$\text{(GNEP)} \qquad \begin{array}{ll} \min\limits_{y} & -F(x,y) \\ \text{s.t.} & y \in Y(x),\ f(x,y) \leq f(x,z) \end{array} \qquad\qquad \begin{array}{ll} \min\limits_{z} & f(x,z) \\ \text{s.t.} & z \in Y(x). \end{array}$$

The right-hand side problem in (GNEP) is convex under Assumption 2.1, and since z acts as a parameter, also the left-hand side problem is convex. Since in equilibrium points the left-hand side player's constraint $f(x,y) \leq f(x,z)$ becomes $f(x,y) \leq \varphi_L(x)$, the (GCQ) is satisfied at all optimal points of the left-hand side problem under Assumption 2.1(5). Under the mild additional assumption of (GCQ) at each optimal point of the right-hand side problem, in the equilibrium points the respective KKT conditions are necessary and sufficient for optimality in the two player problems (note that, as opposed to LICQ and MFCQ, for nonpolyhedral convex sets neither the ACQ nor the GCQ are necessarily preserved under dropping of constraints).

Replacing optimality in the two player problems by their respective KKT systems results in the true single-level mathematical program with complementarity constraints

$$\text{(SP-CC)} \qquad \begin{array}{rl} \min\limits_{x,y,z,u,v,w,\lambda,\mu} & F(x,y) \\ \text{s.t.} & x \in X, \\ & \nabla_y F(x,y) - \nabla_y g(x,y)u - \nabla_y h(x,y)v - w\nabla_y f(x,y) = 0, \\ & w \geq 0,\ f(x,y) - f(x,z) \leq 0,\ w(f(x,y) - f(x,z)) = 0, \\ & u \geq 0,\ g(x,y) \leq 0,\ u^\top g(x,y) = 0, \\ & h(x,y) = 0, \\ & \nabla_z f(x,z) + \nabla_z g(x,z)\lambda + \nabla_z h(x,z)\mu = 0, \\ & \lambda \geq 0,\ g(x,z) \leq 0,\ \lambda^\top g(x,z) = 0, \\ & h(x,z) = 0. \end{array}$$

The global equivalence of (SP-CC) to (MFG) in the sense of [35, Proposition. 5.4] yields also the global equivalence of (SP-CC) to (P_p). The main difference between this SLR and (MM-CC) is the presence of the lower-level KKT conditions in the feasible set of (SP-CC); therefore leading to an increase in the number implicit variable with the additional presence of lower-level Lagrange multipliers. Clearly, the number of variables and constraints in (SP-CC) is significantly larger, and with an an additional algorithmically challenging (lower-level) complementarity constraints appear. We also remark that, due to the latter issue, in [35] (SP-CC) is also solved in a mixed-integer reformulation using additional binary variables. Although usually from completely different angles, reformulations of problem (P_p) of a flavor similar to (SP-CC) have been derived in other papers [14, 31, 7] and used to solve the problem, usually from the perspective of established techniques in optimistic bilevel optimization such as the the mixed-integer reformulation mentioned above.

Recall that in problem (SP-CC), the second complementarity system in the feasible set, counting from the top, is a kind of proxy expression representing the value function constraint if the lower-level value function reformulation is applied to the intermediate-level problem ($IL(x)$). Instead of the (MFG) reformulation, problem (SP) is also globally equivalent to

$$\min_{x,y} F(x,y) \ \text{ s.t. } \ x \in X,\ \ \nabla_y F(x,y) \in N_{S_L(x)}(y), \tag{2.3}$$

under Assumption 2.1(1)–(4). Here, $N_{S_L(x)}(y)$ represents the normal cone to $S_L(x)$ at y, in the sense of convex analysis. If additionally, Assumption 2.1(5) holds in $S_L(x)$ at y for every $x \in X$, then this problem is globally equivalent, in a suitable sense, to the problem

$$\text{(SP-LF-CC)} \qquad \begin{array}{rl} \min\limits_{x,y,u,v,w} & F(x,y) \\ \text{s.t.} & x \in X,\ h(x,y) = 0, \\ & w \geq 0,\ f(x,y) - \varphi_L(x) \leq 0, \\ & u \geq 0,\ g(x,y) \leq 0,\ u^\top g(x,y) = 0, \\ & \nabla_y F(x,y) - \nabla_y g(x,y)u - \nabla_y h(x,y)v - w\nabla_y f(x,y) = 0. \end{array}$$

This reformulation is used in [26] to develop a proximal point-type algorithm for the special case of problem (P_p), where the lower-level problem is unconstrained. It might also be useful to mention that the feasible set of (SP-LF-CC) is reminiscent of the so-called *combined MPEC and the value function approach* introduced for the standard optimistic bilevel program introduced in [46].

Instead of using the generalized equation reformulation of the intermediate-level problem in (2.3), the two-level value function (1.4) could also be used, and this would lead to a double-value function reformulation for problem (SP), rather than the model in (SP-LF-CC). Such a transformation is the basis of the heuristic-type method developed in [2].

2.2 Semi-infinite programming-based reformulation. It is well-known (see, e.g., [47] and references therein) that problem (P_p) is globally equivalent to the generalized semi-infinite programming problem

$$\min_{x,t} \; t \quad \text{s.t.} \quad x \in X, \quad F(x,y) \le t \quad \forall y \in S_L(x). \tag{2.4}$$

In the paper [43], the standard semi-infinite programming approximation

$$\begin{aligned} &\min_{x,z,t} && t \\ &\text{s.t.} && x \in X,\; z \in Y, \lambda : Y \mapsto [0,1], \\ &&& \lambda(y)\left[f(x,z) - f(x,y) + \epsilon\right] + (1-\lambda(y))\left[F(x,y) - t\right] \le 0 \quad \forall y \in Y \end{aligned} \tag{SIP}$$

of problem (2.4), where, $\lambda : Y \mapsto [0,1]$ defines a decision variable, is used to construct a discretization-type algorithm to compute approximate global optimal solutions for problem (P_p). Practical approaches to generate λ can be found in the latter reference.

Note that semi-infinite programming-based techniques have been used to developed algorithms to globally solve optimistic bilevel optimization problems in many papers; see, e.g., [38, 30, 32].

2.3 Two-level value function-based reformulations. Unlike in the context of the previous approaches, where methods are built based on transformations of problem (P_p) into constrained single-level reformulations of the problem, there is a stream of methods that rely preliminarily on directly minimizing the two-level value function φ_p or constructing a modification of the function that serves as base for numerical algorithms. We start here by referring to the work in [9], where a derivative-free optimization (DFO) method based on off-the-shelf tools is applied to approximate the values of the two-level function φ_p in order to estimate the derivatives of the function to design an iterative process to solve problem (P_p).

In [42], the idea of Molodtsov [39, 40] is used as preliminary step to design a method to approximate solutions for problem (P_p). Note that the idea of Molodtsov can be viewed as approximating the maximization two-level value function φ_p with the minimization one

$$\varphi_o^{\theta,\epsilon}(x) := \min_y \left\{ F(x,y) \,\middle|\; y \in S_L^{\theta,\epsilon}(x) \right\}, \tag{2.5}$$

where $\theta > 0$ is penalization parameter and $\varepsilon > 0$ is a relaxation parameter, such that $S_L^{\theta,\epsilon}$ represents a relaxation of the optimal solution set-valued mapping of the regularized lower-level problem obtained by replacing the function f in ($LL(x)$) by $f - \theta F$; i.e.,

$$S_L^{\theta,\epsilon}(x) := \left\{ y \in Y(x) \,\middle|\; f(x,y) - \theta F(x,y) \le \varphi_L^\theta(x) + \epsilon \right\},$$

where φ_L^θ is the corresponding optimal value function defined by

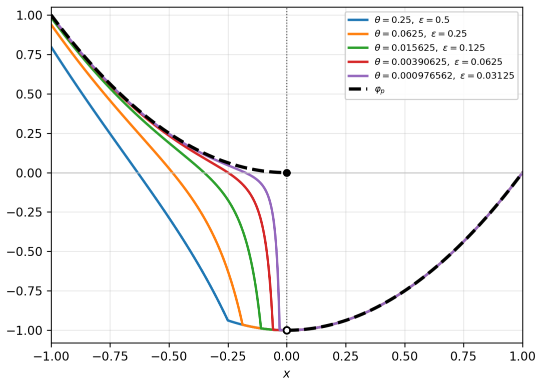
$$\varphi_L^\theta(x) := \min_y \left\{ f(x,y) - \theta F(x,y) \,|\; y \in Y(x) \right\}.$$

Let the optimal values of the pessimistic bilevel program (P_p) and the corresponding regularized problem based on (2.5) be denoted by

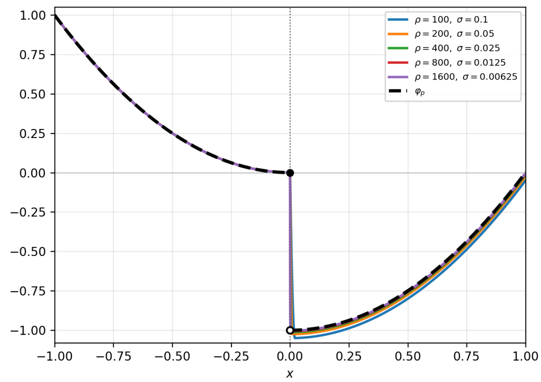
$$\bar{\varphi}_p := \min_{x \in X} \varphi_p(x) \quad \text{and} \quad \bar{\varphi}_o^{\theta,\varepsilon} := \min_{x \in X} \varphi_o^{\theta,\epsilon}(x),$$

respectively. If we assume that the lower-level feasible set is fixed (i.e., $Y(x) := Y$ for all $x \in X$) with X and Y being compact metric spaces, while the functions F and f are *even just* continuous on $X \times Y$, then it is shown in [39] that

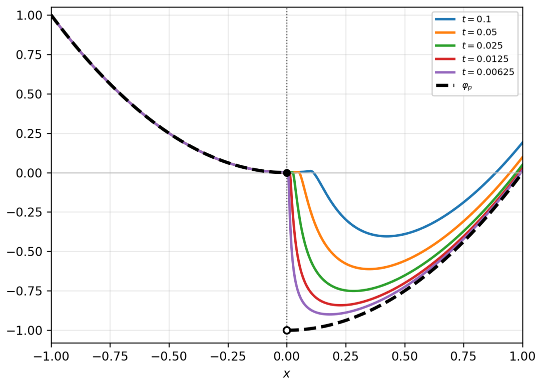
$$\bar{\varphi}_p = \lim_{n \to \infty} \bar{\varphi}_o^{\theta_n,\varepsilon_n} \quad \text{if} \quad \theta_n \to 0^+, \quad \varepsilon_n \to 0^+ \;\; \text{with} \;\; \frac{\theta_n}{\varepsilon_n} \to 0^+.$$



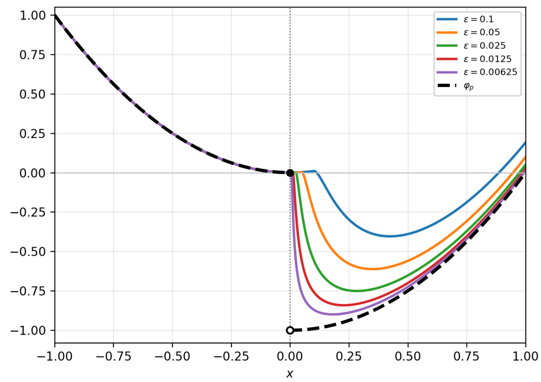
(a) Molodtsov approximation $\varphi_o^{\theta,\epsilon}$ (2.5)



(b) minmax approximation $\phi_{\rho,\sigma}$ (2.6)



(c) Scholtes relaxation $\psi_{\mathcal{S}}^t$ (2.7)



(d) Lower-level value function relaxation φ_p^{ε} (2.10)

Fig. 2: Approximations of the two-level value function φ_p (1.4) for the pessimistic bilevel program (P_p) with $X = [-1, 1]$, $Y = [0, 1]$, $F(x, y) = x^2 - y^2$, and $f(x, y) = -xy$. The reference value function is $\varphi_p(x) = x^2$ for $x \le 0$ and $\varphi_p(x) = x^2 - 1$ for $x > 0$, with a jump discontinuity at $x = 0$.

This result is established under weaker assumptions in [37] and a similar framework for optimistic bilevel optimization is studied in the paper [19].

One that is clear from the Molotsov framework is that the pessimistic bilevel optimization problem is approximated by an optimistic bilevel program; and in fact for $\theta = 0$ and $\varepsilon = 0$, the optimistic bilevel program is recovred. Hence, in [42], the core optimization problem solved is a standard optimistic-type transformation of resulting from (2.5).

Under the assumption that $Y(x) := Y$ for all $x \in X$, the article [8] suggests an approximation of the two-level value function φ_p by the minmax value function

$$\phi_{\rho,\sigma}(x) := \min_{z \in Y} \max_{y \in Y} F(x, y) - \rho\left(f(x, y) - f(x, z)\right) + \frac{\sigma}{2}\|z\|^2 - \sigma y^\top z, \tag{2.6}$$

(with the penalization and regularization parameters ρ and σ being positive) and a framework is constructed to ensure that this function is differentiable and it holds that

$$\lim_{k \to \infty} \left[\inf_{x \in X} \phi_{\rho_k,\sigma_k}(x)\right] = \inf_{x \in X} \varphi_p(x)$$

if X or Y is a bounded set and the sequences $\{\rho_k\}$ and $\{\sigma_k\}$ are such that $\rho_k \to \infty$ and $\sigma_k \to 0$. A gradient descent-type iterative process based on the derivative of the function $\phi_{\rho,\sigma}$ (for $\rho > 0$ and $\sigma > 0$) is proposed and studied in [8].

Finally, in the series of papers [5, 6], considering the KKT reformulation for the lower-level problem, under Assumption 2.1(2)–(4) and the fulfillment of the GCQ in $Y(x)$ at all $y \in S_L(x)$

and $x \in X$, the two-level value function φ_p can be approximated by

$$\psi^t_{\mathcal{R}}(x) := \max_{(y,u)\in\mathcal{D}^t_{\mathcal{R}}(x)} F(x,y), \tag{2.7}$$

where, for $t > 0$ and $x \in X$, $\mathcal{D}^t_{\mathcal{R}}(x)$ can be any classical relaxation (labelled as $\mathcal{R}$) w.r.t. complementarity conditions of set of the KKT points of the lower-level problem ($LL(x)$); i.e.,

$$\mathcal{D}^t_{\mathcal{R}}(x) := \left\{(y,u) \in \mathbb{R}^{m+q} \,\middle|\, \nabla_y \ell(x,y,u) = 0,\ \ \phi^t_{i,\mathcal{R}}(x,y,u) \le 0,\ \ i = 1,\ldots,q\right\}, \tag{2.8}$$

where $\ell(x,y,u) := f(x,y) + u^\top g(x,y)$ denotes the lower-level Lagrangian function with the assumption that $Y(x) := \{y \in \mathbb{R}^m \mid\ g(x,y) \le 0\}$. In particular, for illustration, we assume here that $\mathcal{R}$ corresponds to the *Scholtes* relaxation (denoted by $\mathcal{R} := \mathcal{S}$); then for all $t > 0$ and $i = 1,\ldots,q$, the function $\phi^t_{i,\mathcal{S}}$ is defined for a triplet (x,y,u) by

$$\phi^t_{i,\mathcal{S}}(x,y,u) := \begin{pmatrix} g_i(x,y) \\ -u_i \\ -u_i g_i(x,y) - t \end{pmatrix}.$$

Such a relaxation creates advantages; for instance, (2.7) provides an upper bound for φ_p an if the function $t \mapsto \psi^t_{\mathcal{R}}(x)$ is upper semicontinuous at 0^+ for any $x \in \mathbb{R}^n$ and $(t_k) \downarrow 0$, then for any $x \in \mathbb{R}^n$, we have $\psi^{t_k}_{\mathcal{R}}(x) \to \psi_p(x)$ as $k \to \infty$. Furthermore, let $(t_k) \downarrow 0$ and (x^k) be a sequence such that the point x^k is a global optimal solution of problem

$$\min_{x\in X}\ \psi^t_{\mathcal{R}}(x) \tag{2.9}$$

for $t := t_k$. If $x^k \to \bar{x}$ as $k \to \infty$, then $\bar{x}$ is a global optimal solution of (P_p) provided that the function $t \mapsto \psi^t_{\mathcal{R}}(x)$ is upper semicontinuous at 0^+ for any $x \in \mathbb{R}^n$, and the function $x \mapsto \psi_p(x)$ is lower semicontinuous at $\bar{x}$. Considering this, a method to approximate stationary points of (P_p) via the computation of those of problem (2.9) is introduced and studied in the articles [5, 6].

To conclude this section, it is important to note that a common point between all the approximations of φ_p introduced here is that they lead to functions with much better behavior, as illustrated by the example in Fig. 2. In particular, as in this example, φ_p is typically only upper semicontinuous. This highlights a possible drawback of the two-level value function-based numerical methods, as it is very likely that computed points are not optimal for (P_p) or at best would be optimistic optimal solutions (as it is very likely to be the case for the example in Fig. 2).

Note that in Fig. 2, we also included the value function-based approximation of φ_p defined by

$$\varphi^\varepsilon_p(x) := \max_y\ \{F(x,y) \,|\, y \in Y(x),\ \ f(x,y) - \varphi_L(x) \le \varepsilon\}, \tag{2.10}$$

where the function φ_L is defined in (1.5) and with the relaxation parameter $\varepsilon > 0$, which is commonly used in the literature to tackle the bilevel optimization problem.

3 Preliminary mathematical background.

In this section, we introduce some basic results and concepts that will be used throughout the paper.

3.1 Constraint qualifications, optimality conditions, and duality.

The material presented in this subsection can be mainly found in any standard book on continuous optimization; see, e.g., [23, 45]. The focus in this subsection will be the constrained optimization problem

$$\begin{aligned} \min\quad & \mathfrak{f}(x) \\ \text{s.t.}\quad & x \in \mathcal{C} := \{x \in \mathbb{R}^n \mid\ \mathfrak{g}(x) \le 0,\ \ \mathfrak{h}(x) = 0\} \end{aligned} \tag{3.1}$$

with continuously differentiable functions $\mathfrak{f} : \mathbb{R}^n \to \mathbb{R}$, $\mathfrak{g} : \mathbb{R}^n \to \mathbb{R}^p$, and $\mathfrak{h} : \mathbb{R}^n \to \mathbb{R}^q$. Let $\mathcal{I}(\bar{x}) := \{i := 1,\ldots,p \mid\ \mathfrak{g}_i(\bar{x}) = 0\}$. For ease of notation, in the sequel, we use $\mathcal{I}$ instead of $\mathcal{I}(\bar{x})$. Let $\nabla g_{\mathcal{I}}(\bar{x})$ be the submatrix of the Jacobian matrix $\nabla g(\bar{x})$ made only of the rows with index $i \in \mathcal{I}$. The linear independence constraint qualification (LICQ) will be said to hold at $\bar{x} \in \mathcal{C}$ if

$$\nabla \mathfrak{g}_{\mathcal{I}}(\bar{x})^\top \alpha + \nabla \mathfrak{h}(\bar{x})^\top \beta = 0 \quad \Longrightarrow \quad [\alpha = 0,\ \ \beta = 0]\,. \tag{LICQ}$$

The Mangasarian-Fromovitz constraint qualification (MFCQ) will be said to hold at a point $\bar{x} \in \mathcal{C}$ if the following condition is satisfied:

$$\text{(MFCQ)} \qquad \left.\begin{array}{r} \nabla \mathfrak{g}(\bar{x})^\top \alpha + \nabla \mathfrak{h}(\bar{x})^\top \beta = 0 \\ \alpha \geq 0, \;\; \alpha^\top \mathfrak{g}(\bar{x}) = 0 \end{array}\right\} \Longrightarrow \quad [\alpha = 0, \;\; \beta = 0]\,.$$

Another constraint qualification that will be useful in the sequel is the Abadie constraint qualification (ACQ), which is weaker than the MFCQ. To introduce it, we first recall the tangent cone to $\mathcal{C}$ at one of its points $\bar{x}$:

$$T_{\mathcal{C}}(\bar{x}) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \exists \{x^k\}_{k\in\mathbb{N}} \subset \mathcal{C}, \;\; \exists \{t_k\}_{k\in\mathbb{N}} \subset (0, \;\infty): \\ x^k \to \bar{x}, \;\; t_k \downarrow 0, \;\; \left(x^k - \bar{x}\right)/t_k \to d \end{array} \right\}.$$

The ACQ will be said to hold at a point $\bar{x} \in \mathcal{C}$ if this tangent cone coincides with the linearized tangent cone to $\mathcal{C}$ at the same point $\bar{x}$; i.e.,

$$\text{(ACQ)} \qquad T_{\mathcal{C}}(\bar{x}) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \nabla \mathfrak{g}_i(\bar{x})^\top d \leq 0 \;\; \forall i: \;\; \mathfrak{g}_i(\bar{x}) = 0 \\ \nabla \mathfrak{h}_j(\bar{x})^\top d = 0 \;\; \forall j = 1, \ldots, q \end{array} \right\} =: L_{\mathcal{C}}(\bar{x}).$$

And finally, we introduce a constraint qualification weaker than the ACQ. To proceed, note that for a given cone $\mathcal{K} \subset \mathbb{R}^n$, its dual is the cone

$$\mathcal{K}^* := \left\{ v \in \mathbb{R}^n | \;\; v^\top d \geq 0 \text{ for all } d \in \mathcal{K} \right\}.$$

The Guignard constraint qualification (GCQ) will be said to hold at $\bar{x} \in \mathcal{C}$ if the equality is preserved if the dual is applied on both sides of (ACQ); i.e.,

$$\text{(GCQ)} \qquad \left(T_{\mathcal{C}}(\bar{x})\right)^* = \left(L_{\mathcal{C}}(\bar{x})\right)^*.$$

It is well-known that the GCQ is strictly weaker than the ACQ. Overall, in summary, at a given point $\bar{x} \in \mathcal{C}$, we have the following chain of implications:

$$\text{(LICQ)} \implies \text{(MFCQ)} \implies \text{(ACQ)} \implies \text{(GCQ)}.$$

If $\bar{x}$ is a local optimal solution of problem (3.1) and the GCQ holds at $\bar{x}$, then we can find Lagrange multipliers $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$ such

$$\nabla \mathfrak{f}(\bar{x}) + \nabla \mathfrak{g}(\bar{x})^\top \alpha + \nabla \mathfrak{h}(\bar{x})^\top \beta = 0, \tag{3.2}$$

$$\alpha \geq 0, \;\; \mathfrak{g}(\bar{x}) \leq 0, \;\; \alpha^\top \mathfrak{g}(\bar{x}) = 0, \tag{3.3}$$

$$\mathfrak{h}(x) = 0. \tag{3.4}$$

To state the sufficient optimality condition for problem (3.1), we introduce the Lagrangian function associated to the problem

$$\ell(x, \alpha, \beta) := \mathfrak{f}(x) + \alpha^\top \mathfrak{g}(x) + \beta^\top \mathfrak{h}(x).$$

Now, let $\bar{x}$ be a point such that we can find Lagrange multipliers $\bar{\alpha} \in \mathbb{R}^p$ and $\bar{\beta} \in \mathbb{R}^q$ such that the necessary optimality conditions (3.2)–(3.4) are satisfied. If we assume that the second order sufficient condition (SOSC)

$$d^\top \nabla^2_{xx} \ell\left(\bar{x}, \bar{\alpha}, \bar{\beta}\right) d > 0 \quad \text{for all} \quad d \in \mathfrak{C}(\bar{x}, \bar{\alpha}) \setminus \{0\} \tag{3.5}$$

holds, then $\bar{x}$ is a strict local optimal solution of problem (3.1). Note that here, $\mathfrak{C}(\bar{x}, \bar{\alpha})$ denotes the critical cone associated to problem (3.1) defined by

$$\mathfrak{C}(\bar{x}, \bar{\alpha}) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \nabla \mathfrak{g}_i(\bar{x})^\top d = 0 \;\; \forall i: \;\; \mathfrak{g}_i(\bar{x}) = 0, \;\; \bar{\alpha}_i > 0 \\ \nabla \mathfrak{g}_i(\bar{x})^\top d \leq 0 \;\; \forall i: \;\; \mathfrak{g}_i(\bar{x}) = 0, \;\; \bar{\alpha}_i = 0 \\ \nabla \mathfrak{h}_j(\bar{x})^\top d = 0 \;\; \forall j = 1, \ldots, q \end{array} \right\}.$$

To close this section, consider the following Wolfe dual of problem (3.1):

$$\begin{aligned} \max_{x,\,\alpha,\,\beta} \quad & \ell(x,\alpha,\beta) \\ \text{s.t.} \quad & \nabla_x \ell(x,\alpha,\beta) = 0, \quad \alpha \geq 0. \end{aligned} \tag{3.6}$$

Then, we can state the following strong Wolfe duality result from [44]:

LEMMA 3.1. *For problem (3.1), let the functions $\mathfrak{f}$ and $\mathfrak{g}_i$, for $i = 1,\dots,p$, be convex, while $\mathfrak{h}$ is affine linear. Furthermore, let $\bar{x}$ be an optimal solution of problem (3.1) that satisfies (GCQ). Then, there exist Lagrange multipliers $\bar{\alpha} \in \mathbb{R}^p$ and $\bar{\beta} \in \mathbb{R}^q$ such that $\left(\bar{x},\bar{\alpha},\bar{\beta}\right)$ is an optimal solution of problem (3.6) and $\mathfrak{f}(\bar{x}) = \ell\left(\bar{x},\bar{\alpha},\bar{\beta}\right)$.*

3.2 Parametric and minmin optimization. The focus of this subsection will be on the minmin optimization problem

$$\min_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}(x)} \mathfrak{f}(x,y), \tag{$\mathcal{P}_{2m}$}$$

which involves an *outer* minimization w.r.t. to $x \in \mathbb{R}^n$ (outer variable) and the *inner* minimization w.r.t. $y \in \mathbb{R}^m$ (inner variable). The objective function $\mathfrak{f}$ of the problem is a real-valued function defined on $\mathbb{R}^n \times \mathbb{R}^m$, while $\mathcal{X} \subseteq \mathbb{R}^n$ represents the outer feasible set and the set-valued mapping $\mathcal{Y} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ describes the inner feasible set.

Considering the inner problem in ($\mathcal{P}_{2m}$), which is obviously a parametric optimization problem (in the outer variable), two objects will play an important role in our analysis. That is, we need the optimal solution set-valued mapping $\mathcal{S} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$:

$$\mathcal{S}(x) := \underset{y\in\mathcal{Y}(x)}{\arg\min}\ \mathfrak{f}(x,y) \tag{3.7}$$

and the corresponding optimal value function defined by

$$\phi(x) := \min_{y\in\mathcal{Y}(x)} \mathfrak{f}(x,y).$$

Based on this concept, note that problem ($\mathcal{P}_{2m}$) can be equivalently written as

$$\min_{x\in\mathcal{X}} \phi(x).$$

Hence, throughout this section we will use the following concepts of solution:

DEFINITION 3.2. *A point $\bar{x} \in \mathcal{X}$ will be said to be a local optimal solution of problem ($\mathcal{P}_{2m}$) if there exists a neighborhood U of $\bar{x}$ such that*

$$\phi(\bar{x}) \leq \phi(x) \quad \textit{for all} \;\; x \in X \cap U.$$

As usual, if $U = \mathbb{R}^n$, then the point is a global optimal solution.

Next, we introduce the single-min operator problem associated to problem ($\mathcal{P}_{2m}$):

$$\begin{aligned} \min_{x,y} \quad & \mathfrak{f}(x,y) \\ \text{s.t.} \quad & x \in \mathcal{X}, \quad y \in \mathcal{Y}(x). \end{aligned} \tag{$\mathcal{P}_{1m}$}$$

Denote by $\Omega := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \,|\, x \in \mathcal{X},\ y \in \mathcal{Y}(x)\}$; $(\bar{x},\bar{y})$ will be said to be a local optimal solution of ($\mathcal{P}_{1m}$) if there exists a neighborhood W of $(\bar{x},\bar{y})$ such that

$$\mathfrak{f}(\bar{x},\bar{y}) \leq \mathfrak{f}(x,y) \quad \text{for all} \;\; (x,y) \in \Omega \cap W.$$

Similarly, if $W = \mathbb{R}^n \times \mathbb{R}^m$, then $(\bar{x},\bar{y})$ is a global optimal solution of problem ($\mathcal{P}_{1m}$). Since ϕ is the global minimal value function of the inner problem, we will only be interested in local optimal solutions $(\bar{x},\bar{y})$ of ($\mathcal{P}_{1m}$) with $\bar{y} \in \mathcal{S}(\bar{x})$ when we compare local optimal solutions of ($\mathcal{P}_{1m}$) with those of ($\mathcal{P}_{2m}$) (cf. Lemma 3.4).

Next, we state the global relationship between problems ($\mathcal{P}_{2m}$) and ($\mathcal{P}_{1m}$).

LEMMA 3.3. *The following statements hold true:*

(a) *Let $\bar{x}$ be a global optimal solution of problem* ($\mathcal{P}_{2m}$). *Then, for all $\bar{y} \in \mathcal{S}(\bar{x})$, the point $(\bar{x}, \bar{y})$ is a global optimal solution of problem* ($\mathcal{P}_{1m}$).

(b) *Let $(\bar{x}, \bar{y})$ be globally optimal for* ($\mathcal{P}_{1m}$). *Then $\bar{x}$ is a global optimal solution of* ($\mathcal{P}_{2m}$).

Proof. (a) For any $\bar{y} \in \mathcal{S}(\bar{x})$ and any couple (x, y) such that $x \in \mathcal{X}$ and $y \in \mathcal{Y}(x)$,

$$\mathfrak{f}(\bar{x}, \bar{y}) = \phi(\bar{x}) \leq \phi(x) \leq \mathfrak{f}(x, y).$$

Hence, $(\bar{x}, \bar{y})$ is a global optimal solution of problem ($\mathcal{P}_{1m}$).

As for (b), first note that $(\bar{x}, \bar{y})$ being a global optimal solution of problem ($\mathcal{P}_{1m}$), we automatically have $\bar{y} \in \mathcal{S}(\bar{x})$. Otherwise, we can find $\tilde{y} \in \mathcal{Y}(\bar{x})$ such that

$$\mathfrak{f}(\bar{x}, \bar{y}) > \mathfrak{f}(\bar{x}, \tilde{y}).$$

Note that $(\bar{x}, \tilde{y})$ is a feasible point to problem ($\mathcal{P}_{1m}$). Hence, we have a contradiction. It therefore follows that for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}(x)$, we have

$$\phi(\bar{x}) = \mathfrak{f}(\bar{x}, \bar{y}) \leq \mathfrak{f}(x, y), \tag{3.8}$$

considering the fact that $\bar{y} \in \mathcal{S}(\bar{x})$. Then given that

$$\phi(x) = \begin{cases} +\infty & \text{if} \quad \mathcal{S}(x) = \emptyset, \\ f(x, y^*) \text{ (for some } y^* \in \mathcal{S}(x) \subset \mathcal{Y}(x)) & \text{otherwise,} \end{cases}$$

it follows from (3.8) and the arbitrary choice of $y \in \mathcal{Y}(x)$ that

$$\phi(\bar{x}) = \mathfrak{f}(\bar{x}, \bar{y}) \leq \phi(x).$$

Therefore, $\bar{x}$ is a global optimal solution of problem ($\mathcal{P}_{2m}$). □

To establish the local relationship between the two problems, we need the inner semicontinuity of the inner optimal solution set-valued mapping $\mathcal{S}$. So, $\mathcal{S}$ will be said to be inner semicontinuous at a point $(\bar{x}, \bar{y}) \in \operatorname{gph} \mathcal{S}$ if for every sequence $x^k \to \bar{x}$, there exists a sequence $y^k \in \mathcal{S}(x^k)$ such that $y^k \to \bar{y}$. Note that for any given set-valued mapping Ψ, $(x, y) \in \operatorname{gph} \Psi$ iff $y \in \Psi(x)$. It is worth to mention that the concept of inner semicontinuity holds if the corresponding set-valued mapping is lower semicontinuous in the usual sense; see, e.g., [13] for relevant discussion, some references, and some sufficient conditions that ensure the fulfillment of lower semicontinuity of set-valued mappings that are relevant to $\mathcal{S}$, as described by (3.7).

LEMMA 3.4. *The following statements hold true:*

(a) *Let $\bar{x}$ be a local optimal solution of problem* ($\mathcal{P}_{2m}$). *Then, for all $\bar{y} \in \mathcal{S}(\bar{x})$, $(\bar{x}, \bar{y})$ is a local optimal solution of problem* ($\mathcal{P}_{1m}$).

(b) *Let the point $(\bar{x}, \bar{y}) \in gph \mathcal{S}$, where $\mathcal{S}$ is inner semicontinuous, be a local optimal solution of* ($\mathcal{P}_{1m}$). *Then $\bar{x}$ is a local optimal solution of* ($\mathcal{P}_{2m}$).

Proof. (a) Assume that there is some $\bar{y} \in \mathcal{S}(\bar{x})$ such that $(\bar{x}, \bar{y})$ is not a local optimal solution of ($\mathcal{P}_{1m}$). Then we can find a sequence (x^k, y^k) from Ω with $x^k \to \bar{x}$ and $y^k \to \bar{y}$ such that

$$\mathfrak{f}(x^k, y^k) < \mathfrak{f}(\bar{x}, \bar{y}) = \phi(\bar{x}) \text{ for all } k.$$

Then considering the definition of ϕ, it follows that

$$\phi(x^k) \leq \mathfrak{f}(x^k, y^k) < \mathfrak{f}(\bar{x}, \bar{y}) = \phi(\bar{x}) \text{ for all } k.$$

Therefore, contradicting the fact that the point $\bar{x}$ is a local optimal solution of problem ($\mathcal{P}_{2m}$), given that we have $x^k \in \mathcal{X}$ for all k.

(b) If $\bar{x}$ is not a local optimal solution of problem ($\mathcal{P}_{2m}$), then we can find a feasible sequence $x^k \to \bar{x}$ such that $\phi(\bar{x}) > \phi(x^k)$ for all k. As the set-valued mapping $\mathcal{S}$ is inner semicontinuous at $(\bar{x}, \bar{y})$, we can find a sequence $y^k \in \mathcal{S}(x^k)$ that converges to $\bar{y}$. It follows by the construction that

$$\mathfrak{f}(\bar{x}, \bar{y}) = \phi(\bar{x}) > \phi(x^k) = \mathfrak{f}\left(x^k, y^k\right)$$

with $x^k \in X$, $y^k \in \mathcal{Y}(x^k)$ for all k. This implies that $(\bar{x}, \bar{y})$ is not locally optimal for ($\mathcal{P}_{1m}$). □

The proof of Lemma 3.4 follows along the lines of [16, Theorem 6.9], in the context of the link between the original and standard optimistic bilevel programs. But we include it here for completeness. Also consider the following example based on Example 6.10 from the latter reference:

$$\mathfrak{f}(x,y) := x, \quad \mathcal{X} := [-1,\ 1], \quad \text{and} \quad \mathcal{Y}(x) := \begin{cases} [0,\ 1] & \text{if} \quad x = 0, \\ \{0\} & \text{if} \quad x > 0, \\ \{1\} & \text{if} \quad x < 0. \end{cases}$$

We can easily check that $(0,0)$ is a local optimal solution of the corresponding problem $(\mathcal{P}_{1m})$, while 0 is not a local optimal solution of $(\mathcal{P}_{2m})$. Furthermore, as $\mathcal{S}(x) = \mathcal{Y}(x)$ for all $x \in \mathcal{X}$, $\mathcal{S}$ is not inner semicontinuous at $(0,0)$. This confirms the importance of the inner semicontinuity assumption in part (b) of Lemma 3.4.

4 Duality and single-level reformulations. In this section, we will introduce a true single-level reformulation for problem (P_p) and establish suitable global and local relationships. As mentioned in Section 1, by *true* we mean that the reformulation is a standard nonlinear optimization problem involving neither complementarity constraints nor value functions, and such that classical constraint qualifications hold in solution points. Throughout the remainder of this paper, in addition to Assumption 1.1 we will also consider Assumption 2.1 a blanket assumption.

4.1 A dual reformulation of the intermediate problem. For the subequent analysis it is crucial to note that the pessimistic bilevel problem (P_p) actually possesses a three-level structure. In the lower-level probem $(LL(x))$ $f(x,\cdot)$ is minimized over $Y(x)$, in the intermediate-level problem $(IL(x))$ $F(x,\cdot)$ is maximized over $S_L(x)$, and in the upper-level the function φ_p from (1.4) is minimized over X. The main idea leading to single-level reformulations will, under appropriate convexity assumptions, be based on a dual description of the maximal value $\varphi_p(x)$ of $(IL(x))$ for each fixed $x \in X$ and, hence, of the upper-level's objective function φ_p. To this end, we write the intermediate problem in its equivalent optimal-value function formulation with $Z(x)$ from (2.1),

$$(IL^v(x)) \qquad \max_y F(x,y) \ \text{ s.t. } \ y \in Z(x) = \{y \in \mathbb{R}^m \,|y \in Y(x),\ f(x,y) - \varphi_L(x) \le 0\}.$$

Since for fixed $x \in X$ also $\varphi_L(x)$ is a constant, the above optimal-value function formulation is rather an 'optimal value formulation', which reflects the structure of a so-called *simple bilevel program*; see, e.g., [41], for an overview on the subject. This shall promote a beneficial structure of the single-level reformulation to be introduced below. We emphasize, however, that for each $x \in X$ the MFCQ is violated everywhere in $Z(x)$. Indeed, in view of $Z(x) = S_L(x)$, each $y \in Z(x)$ is a minimal point of the lower-level problem and, thus, satisfies the necessary optimality condition of Fritz-John. The latter prevents the MFCQ from holding at y.

Since Lemma 3.1 does not need the MFCQ but holds under the weaker GCQ, it still makes sense to consider the Wolfe dual (3.6) to $(IL^v(x))$ for fixed $x \in X$. Indeed, this Wolfe dual consists in the minimization of the Lagrangian

$$\mathcal{L}^L_{p_D}(x,y,u,v,w) := F(x,y) - u^\top g(x,y) - v^\top h(x,y) - w(f(x,y) - \varphi_L(x)), \tag{4.1}$$

defined from $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}$ to $\mathbb{R}$, over the set

$$\Lambda_{p_D}(x) := \left\{ (y,u,v,w) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R} \;\middle|\; \begin{array}{l} u \ge 0, \quad w \ge 0 \\ \nabla_y \mathcal{L}^L_{p_D}(x,y,u,v,w) = 0 \end{array} \right\}.$$

By

$$\varphi_{p_D}(x) := \min_{(y,u,v,w)\in\Lambda_{p_D}(x)} \mathcal{L}^L_{p_D}(x,y,u,v,w),$$

we denote the minimal value function of the Wolfe dual, defined on X. Corresponding to the problem (P_p), where φ_p is minimized over X, we introduce the problem

$$(P_{p_D}) \qquad \min_{x\in X} \varphi_{p_D}(x).$$

For $x \in X$, we denote the set of minimal points associated to $\varphi_{p_D}(x)$ by

$$S^L_{p_D}(x) := \operatorname*{arg\,min}_{(y,u,v,w)\in\Lambda_{p_D}(x)} \mathcal{L}^L_{p_D}(x,y,u,v,w). \tag{4.2}$$

Observe that the Lagrangian $\mathcal{L}^L_{p_D}$ from (4.1) relies on the lower-level value function φ_L, which is implicitly defined. Considering potential challenges that could arise in its algorithmic treatment, and to arrive at a true single-level reformulation, let us also introduce the Lagrangian-type real-valued function $\mathcal{L}_{p_D}$ defined by

$$\forall (x,y,z,u,v,w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}:$$
$$\mathcal{L}_{p_D}(x,y,z,u,v,w) := F(x,y) - u^\top g(x,y) - v^\top h(x,y) - w\left(f(x,y) - f(x,z)\right).$$

We can easily observe that for any quintuple $(x,y,u,v,w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}$ with $w \geq 0$, $\mathcal{L}^L_{p_D}$ can be rewritten as

$$\mathcal{L}^L_{p_D}(x,y,u,v,w) = \min_{z\in Y(x)} \mathcal{L}_{p_D}(x,y,z,u,v,w), \tag{4.3}$$

and that

$$\begin{aligned} \nabla_y \mathcal{L}^L_{p_D}(x,y,u,v,w) &= \nabla_y \mathcal{L}_{p_D}(x,y,z,u,v,w) \\ &= \nabla_y F(x,y) - \nabla_y g(x,y)u - \nabla_y h(x,y)v - w\nabla_y f(x,y) \end{aligned} \tag{4.4}$$

holds for any $(x,y,z,u,v,w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}$.

The following lemma will play a crucial role in building the relationships between the problems (P_p), (P_{p_D}), and the single-level reformulations (pSLR), and (tSLR) to be introduced in Section 4.2.

Lemma 4.1. *For all $x \in X$, it holds that*

$$\varphi_p(x) = \varphi_{p_D}(x) = \min_{(y,u,v,w)\in\Lambda_{p_D}(x)} \min_{z\in Y(x)} \mathcal{L}_{p_D}(x,y,z,u,v,w). \tag{4.5}$$

Proof. Under Assumptions 1.1 and 2.1, for each $x \in X$, ($IL^v(x)$) possesses an optimal point $y \in S_p(x)$ which is also a KKT point with corresponding multipliers (u,v,w). By the strong Wolfe duality result from Lemma 3.1, this yields the minimality of (y,u,v,w) for the Wolfe dual and, thus, $\varphi_p(x) = F(x,y) = \mathcal{L}^L_{p_D}(x,y,u,v,w) = \varphi_{p_D}(x)$. The result therefore holds in view of (4.3). □

Remark 4.2. The proof of Lemma 4.1 uses the fact that each KKT point y of problem ($IL^v(x)$) with corresponding multipliers (u,v,w) generates a dually optimal point $(y,u,v,w) \in S^L_{p_D}(x)$. We emphasize that *not all* elements of $S^L_{p_D}(x)$ need to correspond to KKT points of problem ($IL^v(x)$), since the primal feasibility condition $y \in Z(x)$ is not part of the definition of the Wolfe dual. In particular, for an optimal point (y,u,v,w) of the Wolfe dual, the point y need not lie in $S_p(x)$ unless y is primally feasible. Nor do u, v, w need to satisfy complementary slackness conditions with the corresponding constraint functions.

Under the assumptions of Lemma 4.1 the problems (P_p) and (P_{p_D}) are globally and locally equivalent, with the understanding that solution concepts for problem (P_{p_D}) are analogous to those of (P_p) (see Definition 1.2 or Definition 3.2 for a general framework for such an optimal solution notion). While, in particular, for each $x \in X$, the set $\Lambda_{p_D}(x)$ is nonempty, we emphasize that it is also unbounded. Indeed, since the MFCQ is violated everywhere in $Z(x)$, by [24] the set of Lagrange multipliers corresponding to $y \in S_p(x)$ is empty or unbounded, where the first alternative is ruled out by the assumption of the GCQ at y. Since the set of $y \in S_p(x)$ with corresponding Lagrange multipliers forms a subset of $\Lambda_{p_D}(x)$, also the latter is unbounded. This unboundedness is inherited by the feasible sets of the single-level problems (pSLR) and (tSLR) introduced in the subsequent subsection (which, of course, does not entail that also the objective functions of these problems are unbounded on the respective feasible sets).

4.2 Single-level reformulations. Next, we introduce two single-level optimization problems associated to the pessimistic bilevel optimization problem (P_p). We start with the preliminary single-level reformulation defined by

$$\begin{aligned} \min_{x,y,u,v,w} \quad & \mathcal{L}^L_{p_D}(x,y,u,v,w) \\ \text{s.t.} \quad & x \in X,\ \ u \geq 0,\ \ w \geq 0, \\ & \nabla_y F(x,y) - \nabla_y g(x,y)u - \nabla_y h(x,y)v - w\nabla_y f(x,y) = 0 \end{aligned} \tag{pSLR}$$

in whose objective function the implicitly defined and algorithmically potentially challenging function φ_L appears. For the latter reason, we also introduce the true single-level reformulation

$$\text{(tSLR)} \qquad \begin{array}{rl} \min\limits_{x,y,z,u,v,w} & \mathcal{L}_{p_D}(x,y,z,u,v,w) \\ \text{s.t.} & x \in X,\ \ g(x,z) \leq 0,\ \ h(x,z) = 0,\ \ u \geq 0,\ \ w \geq 0, \\ & \nabla_y F(x,y) - \nabla_y g(x,y)u - \nabla_y h(x,y)v - w\nabla_y f(x,y) = 0. \end{array}$$

We remark that, while the last equality constraint in problem (tSLR) originates from the condition $\nabla_y \mathcal{L}^L_{p_D}(x,y,u,v,w) = 0$, in view of (4.4) it may as well be written as $\nabla_y \mathcal{L}_{p_D}(x,y,z,u,v,w) = 0$. Hence, for each fixed (x,z,u,v,w), y is a critical point of $\mathcal{L}_{p_D}(x,\cdot,z,u,v,w)$. Since this function is concave, y is even a global maximal point of $\mathcal{L}_{p_D}(x,\cdot,z,u,v,w)$ over $\mathbb{R}^m$. Therefore, as in the original Wolfe duality argument, the objective function of (tSLR) may be replaced by $\max_{y\in\mathbb{R}^m} \mathcal{L}_{p_D}(x,y,z,u,v,w)$, and the last equality constraint of (tSLR) may instead be dropped. Under appropriate additional linearity assumptions on the defining functions, the variable y can indeed be eliminated completely from (tSLR), see Section 4.3.

Next, we provide links between problems (P_{p_D}) and (pSLR).

Theorem 4.3. *It holds that:*

(a) If $\bar{x}$ is globally optimal for (P_{p_D}), then for all $(\bar{y},\bar{u},\bar{v},\bar{w}) \in S^L_{p_D}(\bar{x})$, the point $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$ is globally optimal for (pSLR). Conversely, if $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$ is a global optimal solution of (pSLR), then $\bar{x}$ is a globally optimal for (P_{p_D}).

(b) If $\bar{x}$ is a local optimal solution of (P_{p_D}), then for all $(\bar{y},\bar{u},\bar{v},\bar{w}) \in S^L_{p_D}(\bar{x})$, the point $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$ is a local optimal solution of (pSLR). Conversely, let the point $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$, where $S^L_{p_D}$ from (4.2) is inner semicontinuous, be a local optimal solution of (pSLR), then $\bar{x}$ is a local optimal solution of problem (P_{p_D}).

Proof. (a) follows from Lemma 3.3, while (b) results from Lemma 3.4. □

Subsequently, we have the following link between problems (pSLR) and (tSLR). To proceed, we introduce the set-valued mapping $S^*_L : \mathbb{R}^n \times \mathbb{R} \rightrightarrows \mathbb{R}^m$ defined by

$$S^*_L(x,w) := \operatorname*{arg\,min}_{z\in Y(x)}\ wf(x,z). \tag{4.6}$$

Obviously, $S^*_L(x,w) = S_L(x)$ if $w > 0$ and $S^*_L(x,w) = Y(x)$ if $w = 0$.

Theorem 4.4. *It holds that:*

(a) If $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$ is globally optimal for (pSLR), then for all $\bar{z} \in S^*_L(\bar{x},\bar{w})$, the point $(\bar{x},\bar{y},\bar{z},\bar{u},\bar{v},\bar{w})$ is globally optimal for (tSLR). Conversely, if $(\bar{x},\bar{y},\bar{z},\bar{u},\bar{v},\bar{w})$ is a global optimal solution of problem (tSLR), then $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$ is a global optimal solution of (pSLR).

(b) If $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$ is locally optimal for (pSLR), then for all $\bar{z} \in S^*_L(\bar{x},\bar{w})$, $(\bar{x},\bar{y},\bar{z},\bar{u},\bar{v},\bar{w})$ is locally optimal for (tSLR). Conversely, let $(\bar{x},\bar{y},\bar{z},\bar{u},\bar{v},\bar{w})$, with S^*_L inner semicontinuous at $(\bar{x},\bar{w},\bar{z})$, be locally optimal for (tSLR), then $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$ is locally optimal for (pSLR).

Proof. Problem (pSLR) is globally and locally equivalent to the problem

$$\text{(pSLR2)} \qquad \min_{(x,y,u,v)\in\Omega^L}\ \min_{z\in Y(x)} \mathcal{L}_{p_D}(x,y,z,u,v,w),$$

where the concept of optimal solution is understood in the same sense as in Definition 3.2, given that we have $\mathcal{L}^L_{p_D}(x,y,u,v,w) = \min\limits_{z\in Y(x)} \mathcal{L}_{p_D}(x,y,z,u,v,w)$. Note that the outer feasible set Ω^L in problem (pSLR2) is given by

$$\Omega^L := \{(x,y,u,v,w) \mid x \in X,\ (y,u,v,w) \in \Lambda_{p_D}(x)\}.$$

Subsequently, considering the fact that

$$S^*_L(x,w) := \operatorname*{arg\,min}_{z\in Y(x)} \mathcal{L}_{p_D}(x,y,z,u,v,w), \tag{4.7}$$

(a) and (b) follow from Lemma 3.3 and Lemma 3.4, respectively. □

Corollary 4.5. *It holds that:*

(a) If the point $\bar{x}$ is globally optimal for problem (P_p), then for all $(\bar{y},\bar{u},\bar{v},\bar{w}) \in S^L_{p_D}(\bar{x})$ and $\bar{z} \in S^*_L(\bar{x},\bar{w})$, the point $(\bar{x},\bar{y},\bar{z},\bar{u},\bar{v},\bar{w})$ is globally optimal for (tSLR). Conversely, if the point $(\bar{x},\bar{y},\bar{z},\bar{u},\bar{v},\bar{w})$ is globally optimal for (tSLR), then $\bar{x}$ is globally optimal (P_p).

(b) If $\bar{x}$ is locally optimal for (P_p), then for all $(\bar{y},\bar{u},\bar{v},\bar{w}) \in S^L_{p_D}(\bar{x})$ and $\bar{z} \in S^*_L(\bar{x},\bar{w})$, the point $(\bar{x},\bar{y},\bar{z},\bar{u},\bar{v},\bar{w})$ is locally optimal for (tSLR). Conversely, let $(\bar{x},\bar{y},\bar{z},\bar{u},\bar{v},\bar{w})$, which is such that $S^L_{p_D}$ (resp. S^*_L) is inner semicontinuous at $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$ (resp. $(\bar{x},\bar{w},\bar{z})$), be locally optimal for (tSLR), then $\bar{x}$ is a local optimal solution of problem (P_p).

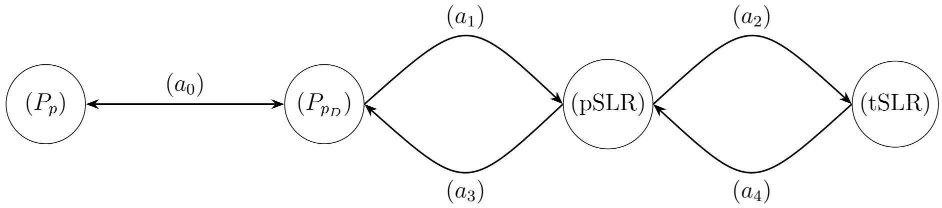


Fig. 3: Links between problems (P_p), (P_{p_D}), (pSLR), and (tSLR) with assumptions (a_1) and (a_2) needed for both the corresponding global and local implications, while (a_3) and (a_4) are the required only for the corresponding local relationships.

The relationships studied here are summarized in Fig. 3, where assumptions (a_0), (a_1), (a_2), (a_3), and (a_4) are respectively defined as follows:

(a_0) Assumption 2.1;
(a_1) $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$ is such that $(\bar{y},\bar{u},\bar{v},\bar{w}) \in S^L_{p_D}(\bar{x})$;
(a_2) $(\bar{x},\bar{y},\bar{z},\bar{u},\bar{v},\bar{w})$ is such that $\bar{z} \in S^*_L(\bar{x},\bar{w})$;
(a_3) $S^L_{p_D}$ is inner semicontinuos at $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$;
(a_4) S^*_L is inner semicontinuos at $(\bar{x},\bar{w},\bar{z})$.

Note that Assumption 1.1 implies that $S^L_{p_D}(\bar{x}) \neq \emptyset$. So, the fulfillment of (a_1) is not a problem, but the requirement rather emphasizes that for the first implications in Theorem 4.3(a) and (b), the point $(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})$ has to be chosen in a specific way; i.e., such that $(\bar{y},\bar{u},\bar{v},\bar{w}) \in S^L_{p_D}(\bar{x})$. An analogous observation can be made for (a_2), while considering the fact that for all $x \in X$, $S^*_L(x,w) = S_L(x)$ if $w > 0$ and $S^*_L(x,w) = Y(x)$ if $w = 0$, while $S_p(x) \subset S_L(x) \subset Y(x)$. As for (a_3) and (a_4), they are only needed for the corresponding local relationships.

Remark 4.6. Due to the observation from Remark 4.2, for a solution $(\bar{x},\bar{y},\bar{z},\bar{u},\bar{v},\bar{w})$ of (tSLR) Corollary 4.5 neither implies optimality of $\bar{y}$ for the intermediate problem ($IL(x)$) nor can optimality of $\bar{z}$ for the lower-level problem ($LL(x)$) be expected. However, once $\bar{x}$ has been computed from a solution of (tSLR), a corresponding optimal point $\bar{z}$ can be generated by solving the convex lower-level problem $(LL(\bar{x}))$, and the solution of the simple bilevel problem $(IL(\bar{x}))$ yields a corresponding optimal point $\bar{y}$ (cf. [41] and the references therein for algorithmic approaches).

4.3 The linear case with respect to the lower-level variable. This section illustrates how the above constructions can be streamlined under appropriate linearity assumptions. Indeed, we shall employ the following assumption, where the occurring functions c, γ, A, b and d map from X to spaces of appropriate dimensions, respectively.

Assumption 4.7. *It holds that:*
(1) $\forall x \in X,\ y \in \mathbb{R}^m:\ F(x,y) = c(x)^\top y + \gamma(x)$.
(2) $\forall x \in X,\ y \in \mathbb{R}^m:\ g(x,y) = A(x)y - b(x)$.
(3) $\forall x \in X,\ y \in \mathbb{R}^m:\ f(x,y) = d(x)^\top y$.

Possibly present equality constraints in the description of $Y(x)$ are also assumed to be affine-linear with respect to y like in Assumption 2.1(4), but here they may be subsumed in the system $A(x)y \leq b$. Observe that, due the the linearity assumptions, for each $x \in X$ (ACQ) holds at each $y \in Z(x)$ so that, altogether, Assumption 4.7 implies Assumption 2.1. Hence, with

$$\mathcal{L}_{p_D}(x,y,z,u,w) = c(x)^\top y + \gamma(x) - u^\top (A(x)y - b(x)) - w\, d(x)^\top (y - z)$$

and

$$\nabla_y \mathcal{L}_{p_D}(x,y,z,u,w) = c(x) - A(x)^\top u - w\, d(x)$$

the true single-level reformulation (tSLR) reads

$$\begin{array}{rl} \min\limits_{x,y,z,u,w} & \mathcal{L}_{p_D}(x,y,z,u,w) \\ \text{s.t.} & x \in X,\ \ g(x,z) \le 0,\ \ u \ge 0,\ \ w \ge 0, \\ & \nabla_y \mathcal{L}_{p_D}(x,y,z,u,w) = 0. \end{array}$$

In analogy to the fact that Wolfe duality collapses to LP duality in the polyhedral case, the equality constraint of this problem can be used to simplify the objective function. This finally results in the problem

$$\text{(tSLR-LP)} \qquad \begin{array}{rl} \min\limits_{x,z,u,w} & \gamma(x) + u^\top b(x) + w\, d(x)^\top z \\ \text{s.t.} & x \in X,\ \ A(x)z \le b(x),\ \ u \ge 0,\ \ w \ge 0, \\ & c(x) - A(x)^\top u - w\, d(x) = 0, \end{array}$$

in which the dependence on y has been eliminated. Replacing the term $w\, d(x)$ in the objective function by another application of the equality constraint yields the alternative expression

$$(c(x)^\top z + \gamma(x)) - u^\top (A(x)z - b(x))$$

for the objective function of problem (tSLR-LP).

Remark 4.8. A special case of Assumption 4.7 is complete linearity, i.e., the affine-linearity of F, g and f in (x, y). This happens for constant functions c, A, and d, and affine-linear γ and b. In this case, and if X is a polyhedral set, the feasible set of (tSLR-LP) is polyhedral as well. Since the objective functions of (tSLR-LP) is a sum of linear and bilinear terms, the problem is nonconvex quadratic. This means that every linear pessimistic bilevel program may be rewritten as a nonconvex quadratic program. Furthermore, observe choosing X as in (1.9), as well as

$$A(x) := A,\ \ b(x) := -Bx + b,\ \ c(x) := x^\top Q_{12} - c_2^\top,\ \ \gamma(x) := \frac{1}{2} x^\top Q_{11} x - c_1^\top x,\ \ d(x) := d,$$

where A, B, b, Q_{12}, c_2, Q_{11}, c_1, and d are constant matrices/vectors, then the problem described in Assumption 4.7 corresponds to the special case of the problem in Subsection 1.2 with $Q_{22} = 0$.

4.4 Problems with discontinuities. One of the main challenges in solving a pessimistic bilevel optimization problem (P_p) is the fact that φ_p may only be upper semicontinuous. This raises the question whether also the single-level reformulations may possess some hidden, but algorithmically unfavorable properties. We prepare our answer by considering the following example.

EXAMPLE 3. *Inspired by [5, Example 2.1], we consider the problem (P_p) with $n = m = 1$, $F(x, y) = x + y$, $f(x, y) = xy$, $X = [\xi, 1]$, $Y = [0, 1]$ and a parameter $\xi < 0$. One can easily check*

$$S_L(x) = \begin{cases} \{1\}, & x \in [\xi, 0), \\ [0,1], & x = 0, \\ \{0\}, & x \in (0,1], \end{cases} \qquad \textit{and} \qquad \varphi_L(x) = \begin{cases} x, & x \in [\xi, 0], \\ 0, & x \in (0,1] \end{cases}$$

as well as

$$S_p(x) = \begin{cases} \{1\}, & x \in [\xi, 0], \\ \{0\}, & x \in (0,1] \end{cases} \qquad \textit{and} \qquad \varphi_p(x) = \begin{cases} x+1, & x \in [\xi, 0], \\ x, & x \in (0,1], \end{cases}$$

so that (P_p) in particular satisfies Assumption 1.1 for all $\xi < 0$. With the functional description $Y = \{y \in \mathbb{R} \mid 0 \le y \le 1\}$, also Assumption 4.7 is satisfied. Recall that (ACQ) holds everywhere in $Z(x)$ since $Z(x) = \{y \in \mathbb{R} \mid 0 \le y \le 1,\ xy \le \varphi_L(x)\}$ is polyhedral for all $x \in X$.

For all $\xi \le -1$, (P_p) possesses a global optimal solution at $\bar{x} = \xi$, while for all $\xi \in (-1, 0)$ it is not solvable since its infinimum zero is not attained. In the latter case the point $\bar{x} = \xi$ is, however, a local optimal solution. The case $\xi \in (-1, 0)$ shows, in particular, that Assumptions 1.1 and 4.7 (let alone the more general Assumption 2.1) are not sufficient for solvability of (P_p).

With $\mathcal{L}_{p_D}(x, y, z, u, w) = x + (1 + u_1 - u_2 - wx)y + u_2 + wxz$, the corresponding (tSLR-LP) is

$$\begin{array}{rl} \min\limits_{x,z,u,w} & x + u_2 + wxz \\ \textit{s.t.} & \xi \le x \le 1,\ \ 0 \le z \le 1,\ \ u_1, u_2, w \ge 0,\ \ 1 + u_1 - u_2 - wx = 0. \end{array}$$

Although at first glance this problem may look well-behaved, by Corollary 4.5(a), it is solvable only for $\xi \leq -1$, while it cannot possess a global optimal solution for $\xi \in (-1, 0)$. In the latter case, like (P_p), the corresponding (tSLR-LP) possesses the non-attained infinimum zero. In fact, the feasible points $(x^k, y^k, z^k, u_1^k, u_2^k, w^k) = (1/k, 0, 0, 0, 0, k)$ yield the objective values $1/k$ for all $k \in \mathbb{N}$, and the following case distinction shows that all feasible points possess a positive objective value: In effect, for all $x \in [\xi, 0]$, we have

$$\begin{aligned} x + u_2 + wxz &= x + (1 + u_1 - wx) + wxz = x + 1 + u_1 + wx(z - 1) \\ &\geq x + 1 \geq \xi + 1 > 0, \end{aligned}$$

and $x \in (0, 1]$ entails $x + u_2 + wxz \geq x > 0$.

We point out that the Weierstrass theorem guarantees solvability of (P_p) if X is nonempty and compact, and if φ_p is lower semicontinuous on X. Given the continuity of F, a standard result from parametric optimization yields the lower semicontinuity of φ_p if the set-valued mapping S_L is inner semicontinuous on X. In Example 3, S_L is not inner semicontinuous at $x = 0$ and, in fact, φ_p is not lower semicontinuous at $x = 0$. For $\xi \in (-1, 0)$ the problem (P_p) is actually not solvable.

Likewise, after the dualization of φ_p to φ_{p_D}, the lower semicontinuity of $\varphi_p = \varphi_{p_D}$ would follow from the lower semicontinuity of φ_L as well the as outer semicontinuity and local boundedness of the set-valued mapping Λ_{p_D} on X. Sufficient conditions for the lower semicontinuity of φ_L are the continuity of f together with the outer semicontinuity and local boundedness of the set-valued mapping Y, which may all be considered mild assumptions. Moreover, it is not hard to see that Λ_{p_D} is outer semicontinuous on X. However, since $\Lambda_{p_D}(x)$ is unbounded (even for all $x \in X$), the local boundedness assumption for Λ_{p_D} fails. This shows that typical unsolvability issues in the pessimistic bilevel problem (P_p) are inherited by the single-level problem (tSLR). They can be ruled out by additional assumptions like the inner semicontinuity of S_L on X.

5 Optimality conditions. In this section, we derive necessary and sufficient optimality conditions for problem (tSLR) and show how to leverage on them to obtain necessary and sufficient optimality conditions for the pessimistic bilevel program (P_p). Throughout this section, we assume here that in problem (tSLR), the upper-level feasible set is described as

$$X := \{x \in \mathbb{R}^n \mid G(x) \leq 0, \quad H(x) = 0\} \tag{5.1}$$

with the functions $G : \mathbb{R}^n \to \mathbb{R}^r$ and $H : \mathbb{R}^n \to \mathbb{R}^s$ being continuously differentiable. Furthermore, the functions F, f, and g are assumed to be twice continuously differentiable w.r.t. y.

We start in the next subsection with the construction of tractable sufficient conditions to ensure the fulfillment of the constraint qualifications (MFCQ) and (LICQ) for problem (tSLR). Subsequently, in Subsection 5.2, we derive necessary optimality conditions for problem (P_p) based on the corresponding conditions for problem (tSLR).

5.1 Constraint qualifications. To proceed here, note that the *upper-level regularity* will be said to hold at x is the MFCQ, as defined in (MFCQ), holds at this point for the constraint system defining the set X. Similarly, the *lower-level regularity* will be said to be satisfied at (x, z) if the MFCQ is satisfied at this point for the constraint system describing the set $Y(x)$. From now on, as necessary, we will use the notation $\zeta := (x, y, z, u, v, w)$. The next result provides sufficient conditions for the fulfillment of the MFCQ for problem (tSLR).

THEOREM 5.1. *The MFCQ is satisfied at a feasible point $\zeta := (x, y, z, u, v, w)$ of problem (tSLR) if the upper-level regularity (resp. lower-level regularity) holds at the point $\bar{x}$ (resp. (x, z)) and the matrix $\nabla^2_{yy}\mathcal{L}^L_{p_D}(x, y, u, v, w)$ is full rank.*

Proof. We start with the notation

$$\alpha := \begin{bmatrix} \alpha_G \\ \alpha_g \\ \alpha_u \\ \alpha_w \end{bmatrix}, \ \beta := \begin{bmatrix} \beta_H \\ \beta_h \\ \beta_{\mathcal{L}_{p_D}} \end{bmatrix}, \ \tilde{G}(\zeta) := \begin{bmatrix} G(x) \\ g(x, z) \\ -u \\ -w \end{bmatrix}, \ \tilde{H}(\zeta) := \begin{bmatrix} H(x) \\ h(x, z) \\ \nabla_y \mathcal{L}_{p_D}(\zeta) \end{bmatrix}. \tag{5.2}$$

Then, we can easily check that the condition $\nabla_\zeta \tilde{G}(\zeta)^\top \alpha + \nabla_\zeta \tilde{H}(\zeta)^\top \beta = 0$ is equivalent to the following system of equations:

$$\nabla G(x)^\top \alpha_G + \nabla H(x)^\top \beta_H + \nabla_x g(x,z)^\top \alpha_g + \nabla_x h(x,z)^\top \beta_h$$

$$+\ \nabla^2_{xy}\mathcal{L}_{p_D}(\zeta)^\top \beta_{\mathcal{L}_{p_D}} = 0, \tag{5.3}$$

$$\nabla_z g(x,z)^\top \alpha_g + \nabla_z h(x,z)^\top \beta_h = 0, \tag{5.4}$$

$$\nabla^2_{yy}\mathcal{L}_{p_D}(\zeta)^\top \beta_{\mathcal{L}_{p_D}} = 0, \qquad \nabla^2_{vy}\mathcal{L}_{p_D}(\zeta)^\top \beta_{\mathcal{L}_{p_D}} = 0, \tag{5.5}$$

$$\nabla^2_{uy}\mathcal{L}_{p_D}(\zeta)^\top \beta_{\mathcal{L}_{p_D}} = \alpha_u, \qquad \nabla^2_{wy}\mathcal{L}_{p_D}(\zeta)^\top \beta_{\mathcal{L}_{p_D}} = \alpha_w. \tag{5.6}$$

With this system, if we formally write the dual form of the MFCQ from Section 3.1 for (tSLR) at $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$, we can easily check that from the lower-level regularity at $(\bar{x}, \bar{z})$, condition (5.4) will imply that $\alpha_g = 0$ and $\beta_h = 0$, while the first equation in (5.5) will lead to $\beta_{\mathcal{L}_{p_D}} = 0$ under the full rank condition imposed in the statement of the lemma. The latter will also imply, considering (5.6), that we have $\alpha_u = 0$ and $\alpha_w = 0$. Finally, with $\alpha_g = 0$, $\beta_h = 0$, and $\beta_{\mathcal{L}_{p_D}} = 0$, it will follow from (5.3) that $\alpha_G = 0$ and $\alpha_H = 0$ under the fulfillment of the upper-level regularity at $\bar{x}$. ∎

Similarly, we next provide sufficient conditions for the fulfillment of the LICQ for (tSLR). The *upper-level (resp. lower-level) LICQ* will be said to hold at x (resp. (x,z)) if LICQ holds at this point for the constraint system defining the upper-level (resp. lower-level) feasible X (resp. $Y(x)$).

THEOREM 5.2. *The LICQ is satisfied at a feasible point $\zeta := (x, y, z, u, v, w)$ of (tSLR) if the upper-level (resp. lower-level) LICQ holds at x (resp. (x,z)) and $\nabla^2_{yy}\mathcal{L}^L_{p_D}(x, y, u, v, w)$ is full rank.*

Proof. Follows along the same line as in the proof of Theorem 5.1. ∎

Next, we provide a framework for the fulfillment of the full rank condition in the result above.

PROPOSITION 5.3. *If $\zeta := (x, y, z, u, v, w)$ is feasible for (tSLR), then $\nabla^2_{yy}\mathcal{L}^L_{p_D}(x, y, u, v, w)$ is full rank, provided that one of the following assumptions is satisfied:*
(a) $\nabla^2_{yy}F(x,y) \prec 0$;
(b) $\nabla^2_{yy}f(x,y) \succ 0$ and $w > 0$;
(c) $\nabla^2_{yy}g_i(x,y) \succ 0$ and $u_i > 0$ for some $i = 1, \ldots, p$.

Proof. Start by observing that based on Assumption 2.1(1)–(4), it holds that

$$\nabla^2_{yy}\mathcal{L}^L_{p_D}(x, y, u, v, w) = -\left(-\nabla^2_{yy}F(x,y) + w\nabla^2_{yy}f(x,y) + \sum_{i=1}^{p} u_i \nabla^2_{yy}g_i(x,y)\right) \preceq 0 \tag{5.7}$$

given that $w \geq 0$ and $u_i \geq 0$ for $i = 1, \ldots, p$ (thanks to the feasibility of $\zeta := (x, y, z, u, v, w)$) and the fact that the functions F, f, and g are twice continuously differentiable w.r.t. y. Therefore, if assumption (a), (b), or (c) of the statement holds, then we have from (5.7) that

$$\nabla^2_{yy}\mathcal{L}^L_{p_D}(x, y, u, v, w) \prec 0.$$

Hence, ensuring that the matrix $\nabla^2_{yy}\mathcal{L}^L_{p_D}(x, y, u, v, w)$ is full rank. ∎

As for the upper- and lower-level regularity and LICQ, we can easily construct examples of functions describing the set X (5.1) and the set-valued mapping Y (1.1), under the framework of Proposition 5.3, such that they are satisfied. This is one of the strengths of reformulation (tSLR), as none of the SLRs of (P_p) introduced in Section 2 can permit the fulfillment of the MFCQ or LICQ. This is neither possible for the optimistic bilevel program, as widely documented in the literature [12, 20].

5.2 Necessary optimality conditions. We start here by providing a framework for the fulfillment of the necessary optimality conditions of problem (tSLR).

THEOREM 5.4. *Let (x, y, z, u, v, w) be a local optimal solution of problem (tSLR), where all the assumptions of Theorem 5.1 are satisfied. Then, there exist Lagrange multipliers α_G, α_g, β_H, β_h,* *and* $\beta_{\mathcal{L}_{p_D}}$ *such that the following conditions are satisfied:*

$$\nabla G(x)^\top \alpha_G + \nabla H(x)^\top \beta_H + \nabla_x g(x,z)^\top \alpha_g + \nabla_x h(x,z)^\top \beta_h$$

$$+\ \nabla_x \mathcal{L}_{p_D}(\zeta)\ +\ \nabla^2_{xy}\mathcal{L}_{p_D}(\zeta)^\top \beta_{\mathcal{L}_{p_D}} = 0, \tag{5.8}$$

$$\alpha_G \geq 0, \quad G(x) \leq 0, \quad \alpha_G^\top G(x) = 0, \tag{5.9}$$

$$H(x) = 0, \tag{5.10}$$

$$\nabla^2_{yy}\mathcal{L}_{p_D}(\zeta)^\top \beta_{\mathcal{L}_{p_D}} = 0,\ \ \nabla_y \mathcal{L}_{p_D}(\zeta) = 0,\ \ h(x,y) + \nabla_y h(x,y)\beta_{\mathcal{L}_{p_D}} = 0, \tag{5.11}$$

$$u \geq 0,\ \ g(x,y) + \nabla_y g(x,y)\beta_{\mathcal{L}_{p_D}} \leq 0,\ \ u^\top \left(g(x,y) + \nabla_y g(x,y)\beta_{\mathcal{L}_{p_D}}\right) = 0, \tag{5.12}$$

$$w \geq 0,\ \ f(x,y) - f(x,z) + \nabla_y f(x,y)\beta_{\mathcal{L}_{p_D}} \leq 0,$$

$$w\left(f(x,y) - f(x,z) + \nabla_y f(x,y)\beta_{\mathcal{L}_{p_D}}\right) = 0, \tag{5.13}$$

$$w\nabla_z f(x,z) + \nabla_z g(x,z)^\top \alpha_g + \nabla_z h(x,z)^\top \beta_h = 0, \tag{5.14}$$

$$\alpha_g \geq 0, \quad g(x,z) \leq 0, \quad \alpha_g^\top g(x,z) = 0, \tag{5.15}$$

$$h(x,z) = 0. \tag{5.16}$$

Proof. It follows straightforwardly from the application of the classical Lagrange multiplier rule to problem (tSLR), while taking into account the fact that under the assumptions of Theorem 5.1, the MFCQ holds at $(\bar{x},\bar{y},\bar{z},\bar{u},\bar{v},\bar{w})$, as a feasible point of (tSLR), and also observing that

$$\alpha_u := -\left(g(x,y) + \nabla_y g(x,y)\beta_{\mathcal{L}_{p_D}}\right) \text{ and } \alpha_w := -\left(f(x,y) - f(x,z) + \nabla_y f(x,y)\beta_{\mathcal{L}_{p_D}}\right),$$

respectively, based on the corresponding definitions in (5.2). ∎

Note that the system (5.8)–(5.16) corresponds to the KKT conditions of problem (tSLR).

Remark 5.5. Based on Assumption 2.1(2)-(4), with $w \geq 0$, the existence of the Lagrange multipliers α_g and β_h such that the block (5.14)–(5.16) of this system holds is equivalent to the inclusion $z \in S^*_L(x,w)$. Therefore, in some sense, this represents the lower-level problem in the KKT conditions (5.8)–(5.16). As for the block (5.11)–(5.13), it corresponds to necessary conditions for $(y,u,v,w) \in S^L_{p_D}(x)$; hence, meaning that this part of the KKT conditions of problem (tSLR) represents the Wolfe dual of the intermediate problem ($IL(x)$). In the same vein, (5.8)–(5.10) can be viewed as representing the upper-level problem described in (P_p) or (P_{p_D}).

Corollary 5.6. *Let x be a local optimal solution of (P_p), and assume that there exist points $(y,u,v,w) \in S^L_{p_D}(x)$ and $z \in S^*_L(x,w)$ such that the upper-level regularity (resp. lower-level regularity) holds at x (resp. (x,z)) and $\nabla^2_{yy}\mathcal{L}^L_{p_D}(x,y,u,v,w)$ is full rank. Then, there exist Lagrange multipliers α_G, α_g, β_H, β_h, and $\beta_{\mathcal{L}_{p_D}}$ such that the KKT conditions (5.8)–(5.16) are satisfied.*

Proof. First note that considering Assumption 1.1, it holds that $S^L_{p_D}(x) \neq \emptyset$, while accounting for the fulfillment of the Wolfe duality result, which in turn holds thanks to Assumption 2.1 (for reference, see Lemma 3.1 and Lemma 4.5). Then observe that also due to Assumption 1.1, it holds that $S^*_L(x,w) \neq \emptyset$. Subsequently, the overall conclusion of the result follows from a combination of the first part of Corollary 4.5(b) and Theorem 5.4. ∎

It follows from Remark 5.5 that the inclusions $(y,u,v,w) \in S^L_{p_D}(x)$ and $z \in S^*_L(x,w)$ are already represented in the KKT conditions (5.8)–(5.16), with $z \in S^*_L(x,w)$ equivalently via (5.14)–(5.16) and $(y,u,v,w) \in S^L_{p_D}(x)$ necessarily with the presence of (5.11)–(5.13). Therefore, they are somehow redundant, and do not necessarily need to be accounted for while referring to the necessary optimality conditions of problem (P_p) obtained here via (tSLR).

The result in Corollary 5.6 represents a fundamental paradigm shift in terms of the construction of necessary optimality conditions for the pessimistic bilevel optimization problem (P_p), and two main observations could made to compare it with existing ones from the literature:

(i) The existing approaches to derive necessary optimality conditions for problem (P_p) are based on calculations of upper estimates for the subdifferential of φ_p [17, 16, 18]. Hence, the required qualification conditions involve assumptions to ensure that this function is Lipschitz continuous near the point of interest. In particular, it is usually required that the set-valued mapping S_p (1.6) or its suitable transformation, depending on the context, satisfies some continuity properties such as the inner semicontinuity, which is not only a strong requirement, but also a very difficult condition to verify in practice. On the other hand, Proposition 5.3 provides a base for a large class of problems for which all the requirements of Corollary 5.6 are automatically satisfied.

(ii) The existing optimality conditions for (P_p) usually involve combinatorial structures such as the S-, M-, and C-type necessary optimality conditions, which are difficult to check or compute in practice. Additionally, most of the existing optimality conditions can give rise to quite large systems, given that they involve convex combinations due to the convex hull structure that intervenes by virtue of the process to compute elements from the Clarke subdifferential of φ_p. On the contrary, our necessary optimality conditions (5.11)–(5.13) are of the usual KKT-type, as they involve only complementarity conditions. Furthermore, we can easily check that (5.11)–(5.13) can be written as a $(n + 3m + 2p + 2q + r + s + 1) \times (n + 3m + 2p + 2q + r + s + 1)$ square system of equations, and is therefore amenable to the semismooth Newton methods; see. e.g., [11, 22].

5.3 Second order sufficient conditions. We close this section by discussing second order sufficient conditions for a point to be a local optimal solution for problem (P_p). Considering the nature of the feasible of problem (tSLR), we assume in this subsection that the functions F, f, g, and h are thrice continuously differentiable.

THEOREM 5.7. *Let the point $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ satisfy the KKT conditions (5.8)–(5.16) with the Lagrange multipliers α and β. Furthermore, assume that the following conditions hold:*
(a) the SOSC (3.5) for (tSLR) is satisfied at $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$;
(b) $S^L_{p_D}$ is inner semicontinuous at $(\bar{x}, \bar{y}, \bar{u}, \bar{v}, \bar{w})$;
(c) S^*_L is inner semicontinuous at $(\bar{x}, \bar{w}, \bar{z})$.
Then, $\bar{x}$ is a strict local optimal solution of problem (P_p) of Category I.

Proof. Start by noting that $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ is a strict local optimal solution of problem (tSLR) by direct application of the SOSC (3.5) to the problem; cf. assumption (a). The local optimality of $\bar{x}$ for problem (P_p) follows directly by virtue of the of the second part of Corollary 4.5(b), thanks to assumptions (b) and (c) of this theorem. However, the strictness and the category of the local optimality of $\bar{x}$ for problem (P_p) still need to be established.

On the strictness, suppose, by contradiction, that $\bar{x}$ is not a strict local optimal solution of problem (P_p). Then, there exists a sequence $\{x^k\} \subset X \setminus \{\bar{x}\}$ satisfying $x^k \to \bar{x}$ such that $\varphi_p(x^k) \leq \varphi_p(\bar{x})$ for all k. Subsequently, by the inner semicontinuity of $S^L_{p_D}$ at $(\bar{x}, \bar{y}, \bar{u}, \bar{v}, \bar{w})$, there exists a sequence $(y^k, u^k, v^k, w^k) \in S^L_{p_D}(x^k)$ such that $(y^k, u^k, v^k, w^k) \to (\bar{y}, \bar{u}, \bar{v}, \bar{w})$. Similarly, by the inner semicontinuity of S^*_L at $(\bar{x}, \bar{w}, \bar{z})$, there exists a sequence $z^k \in S^*_L(x^k, w^k)$ such that we have $z^k \to \bar{z}$. It follows from equations (4.5) and (4.7), together with (4.3), that

$$\mathcal{L}^L_{p_D}(x^k, y^k, u^k, v^k, w^k) = \varphi_p(x^k) \quad \text{and} \quad \mathcal{L}_{p_D}(x^k, y^k, z^k, u^k, v^k, w^k) = \mathcal{L}^L_{p_D}(x^k, y^k, u^k, v^k, w^k),$$

respectively. Therefore, $\mathcal{L}_{p_D}(x^k, y^k, z^k, u^k, v^k, w^k) = \varphi_p(x^k) \leq \varphi_p(\bar{x})$.

On the other hand, we also implicitly have from assumptions (b) and (c) of the theorem that $(\bar{y}, \bar{u}, \bar{v}, \bar{w}) \in S^L_{p_D}(\bar{x})$ and $\bar{z} \in S^*_L(\bar{x}, \bar{w})$, respectively, are satisfied. Hence, it follows once more from (4.5) and (4.7), together with (4.3), that

$$\mathcal{L}_{p_D}(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w}) = \mathcal{L}^L_{p_D}(\bar{x}, \bar{y}, \bar{u}, \bar{v}, \bar{w}) = \varphi_p(\bar{x}).$$

Subsequently, we get that for every k,

$$\mathcal{L}_{p_D}(x^k, y^k, z^k, u^k, v^k, w^k) \leq \mathcal{L}_{p_D}(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w}).$$

Since $(x^k, y^k, z^k, u^k, v^k, w^k) \to (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$, this contradicts the strict local optimality of the point $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ for problem (tSLR). Therefore, $\bar{x}$ is a strict local optimal solution of (P_p).

Recall that to show that x is a local optimal solution of (P_p) of Category I, it suffices to prove that $S_L(x)$ is a singleton. To proceed, we distinguish the cases $w > 0$ and $w = 0$. For $w > 0$ the KKT conditions for ζ with multipliers α and β imply that z is a KKT point of ($LL(x)$) with multipliers $\alpha'_g := \alpha_g/w$ and $\beta'_h := \beta_h/w$. The convexity of ($LL(x)$) implies $z \in S_L(x)$. Moreover, it is not hard to see that the vectors $d_z \in \mathbb{R}^m$ with $(0,0,d_z,0,0,0) \in \mathfrak{C}(\zeta,\alpha)$ form the critical cone of ($LL(x)$) at z with respect to the multiplier vector α'_g. Therefore the SOSC with the Lagrangian

$$\begin{aligned} L(\zeta,\alpha,\beta) \quad &:= \quad \mathcal{L}_{p_D}(\zeta) + \alpha_G^\top G(x) + \alpha_g^\top g(x,z) - \alpha_u^\top u - \alpha_w w \\ &\qquad + \beta_H^\top H(x) + \beta_h^\top h(x,z) + \textstyle\sum_{j=1}^m \beta^j_{\mathcal{L}_{p_D}} \nabla_{y_j}\mathcal{L}_{p_D}(\zeta) \end{aligned}$$

of (tSLR) implies $d_z^\top \nabla^2_{zz} L(\zeta,\alpha,\beta) d_z > 0$ for all $d_z \in \mathfrak{C}(z,\alpha'_g)\setminus\{0\}$. In view of

$$\begin{aligned} \nabla^2_{zz} L(\zeta,\alpha,\beta) &= w\nabla^2_z f(x,z) + \nabla^2_z g(x,z)\alpha_g + \nabla^2_z h(x,z)\beta_h \\ &= w\left(\nabla^2_z f(x,z) + \nabla^2_z g(x,z)\alpha'_g + \nabla^2_z h(x,z)\beta'_h\right) \end{aligned}$$

one obtains the SOSC for z to be a unique local minimal point of ($LL(x)$). In view of $z \in S_L(x)$, the latter set is thus a singleton.

In the case $w = 0$, with analogous arguments as above one sees that z is a KKT point of the lower-level feasibility problem

$$(LL^0(x)) \qquad \min_z 0 \ \text{ s.t. } \ z \in Y(x)$$

with multipliers α_g and β_h, that it is also a global minimal point, that the vectors $d_z \in \mathbb{R}^m$ with $(0,0,d_z,0,0,0) \in \mathfrak{C}(\zeta,\alpha)$ form the critical cone of ($LL^0(x)$) at z and that one obtains the SOSC for z to be a unique local minimal point of ($LL^0(x)$). This is only possible if $Y(x)$ is a singleton, implying that also its subset $S_L(x)$ is a singleton. ∎

To the best of our knowledge this is the first result providing sufficient optimality conditions for the pessimistic bilevel optimization problem (P_p). And clearly, this is possible again thanks to the relationship between this problem and the true single-level reformulation (tSLR).

It is important to note that assumptions (b) and (c) are not there just by default to ensure the local optimality the x-component of a local optimal solution of problem (tSLR) for problem (P_p); these assumptions are also crucial to establish the strictness of such a local optimal solution.

Finally, we provide an example of problem where all the assumptions of Theorem 5.7 hold.

Example 4. *For the problem* (P_p) *with*

$$F(x,y) := y, \quad f(x,y) := x, \quad X := \mathbb{R}, \quad \textit{and} \quad Y(x) := [-x^2, x^2],$$

one obtains $\varphi_p(x) = x^2$, so that $\bar{x} = 0$ is a strict local optimal solution of (P_p). This also follows from Theorem 5.7, since all its assumptions can be satisfied.

Indeed, first note that Assumption 2.1 holds when we put $g_1(x,y) = -x^2 - y$, $g_2(x,y) = -x^2 + y$ and $g_3(x,y) = x - \varphi_L(x)$ with $\varphi_L(x) = x$, i.e. $g_3(x,y) = 0$. In this example we will not remove such redundant constraints in order to account for possible degeneracies in the SOSC. Also note that Assumption 4.7 would be satisfied if we equivalently defined $f(x,y) = 0$, but that here we do not intend to make use of the special polyhedral structure with respect to y and, in particular, not formulate the problem (tSLR-LP). Instead, the corresponding problem (tSLR) is

$$\begin{aligned} \min_{x,y,z,u_1,u_2,u_3,w} \ & y + u_1(x^2+y) + u_2(x^2-y) \\ s.t. \quad & -x^2 - z \le 0, \ -x^2 + z \le 0, \ 0 \le 0, \ u, w \ge 0, \\ & 1 + u_1 - u_2 = 0. \end{aligned}$$

The point $\bar{\zeta} = (\bar{x},\bar{y},\bar{z},\bar{u},\bar{w})$ with $\bar{x} = \bar{y} = \bar{z} = 0$, $\bar{u} = (0,1,1)^\top$, $\bar{w} = 1$ is a KKT point with multipliers $\bar{\alpha}_g = (1/4,1/4,1)^\top$, $\bar{\alpha}_u = (0,0,0)^\top$, $\bar{\alpha}_w = \beta = 0$. The critical cone is computed to be

$$\mathfrak{C}(\bar{\zeta},\bar{\alpha}) = \{(d_x, d_y, 0, d_{u_1}, d_{u_1}, d_{u_3}, d_w)^\top \mid d_x, d_y, d_{u_3}, d_w \in \mathbb{R},\ d_{u_1} \ge 0\},$$

and

$$\nabla^2_\zeta L(\bar{\zeta},\bar{\alpha}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

This yields $d^\top \nabla^2_\zeta L(\bar\zeta,\bar\alpha)d = d_x^2 > 0$ *for all* $d \in \mathfrak{C}(\bar\zeta,\bar\alpha)\setminus\{0\}$*, that is, the SOSC is satisfied at* $\bar\zeta$*.*

With respect to the inner semicontinuity of $S^L_{P_D}$ *at* $(\bar x,\bar y,\bar u,\bar w)$ *observe that*

$$\Lambda_{P_D}(x) = \{(y,u,w) \mid u \geq 0,\ w \geq 0,\ 1+u_1-u_2 = 0\}$$

and $\varphi_{P_D}(x) = \varphi_p(x) = x^2$ *yield*

$$S^L_{P_D}(x) = \{(y,u,w) \mid u \geq 0,\ w \geq 0,\ 1+u_1-u_2=0,\ (u_1+u_2)x^2 = x^2\}.$$

From $(\bar y,\bar u,\bar w) \in S^L_{P_D}(x)$ *for all* $x \in \mathbb{R}$ *we obtain the required inner semicontinuity assumption.*

Finally, S^*_L *is inner semicontinuous at* $(\bar x,\bar w,\bar z) = (0,1,0)$ *in view of* $S^*_L(x) = Y(x)$ *for all* $x \in \mathbb{R}$ *and the inner semicontinuity of* Y *at* $(0,0)$*.*

6 Numerical illustrations. Considering the relationships established in Section 4 between problems (P_p) and (tSLR), we use 6 well-known examples from the literature to illustrate this connection between the two problems. We only consider the global relationship here; cf. Corollary 4.5(a). To proceed, we use the DIRECT algorithm introduced in [29]. However, since this algorithm can only handle box constraints, it is combined with a (quadratic) penalization (to get an unconstrained problem). More precisely, assuming that we are solving the problem (3.1), the constrained problem (3.1) is approximated by the bound-constrained penalized problem

$$\min_{x\in[\ell,u]} \Psi_\rho(x) := \mathfrak{f}(x) + \rho\mathcal{V}(x), \tag{6.1}$$

where, $\rho > 0$ is the penalty parameter and $[\ell,u]$ is a finite search box. For variables that are originally unbounded or one-sided bounded, sufficiently large artificial bounds, $-B \leq x_i \leq B$ and $0 \leq x_i \leq B$, are introduced, as appropriate. The value of B is increased whenever a candidate solution lies close to an artificial boundary. Note that in problem (6.1), $\mathcal{V}$ denotes the quadratic constraint-violation function

$$\mathcal{V}(x) := \sum_{i=1}^{p}\big[\max\{0,\mathfrak{g}_i(x)\}\big]^2 + \sum_{j=1}^{q}\mathfrak{h}_j(x)^2. \tag{6.2}$$

Observe that $\mathcal{V}(x) = 0$ if and only if x is feasible for (3.1). Hence, the global-search procedure consists of two phases. First, DIRECT is applied to

$$\min_{x\in[\ell,u]} \mathcal{V}(x)$$

to identify a point with small constraint violation. Second, DIRECT is applied to the penalized objective Ψ_ρ (6.1). The first phase prevents an improvement in $\mathfrak{f}$ from concealing a large infeasibility during the penalized search. DIRECT partitions the normalized box into hyperrectangles and samples their centers. Hyperrectangles that are potentially optimal, based on both their sampled objective values and their sizes, are subdivided. The method therefore balances global exploration of large unexamined regions with local refinement around promising points. This local refinement is done by applying the well-known sequential least squares programming (SLSQP) method [33] to (3.1). Considering the combination of the *D*IRECT method with the *P*enalization and *S*LSQP algorithm, we label the computational tool used for the experiments presented here as *DPS*.

Let $\bar x$ denote the candidate returned by the two DIRECT phases. It is accepted as feasible when $\sqrt{\mathcal{V}(\bar x)} \leq \varepsilon$, where $\varepsilon > 0$ represents the feasibility tolerance. Overall, DPS is just a deterministic global-search heuristic for the constrained problem (3.1). DIRECT has global convergence properties for the bound-constrained penalized problem under its standard assumptions [29]. However, a quadratic penalty with finite ρ is not generally exact. Consequently, global minimization of (6.1) alone does not provide an unconditional certificate of global optimality for (3.1). But, as it will be shown below, the performance of DPS confirms the finding of Corollary 4.5(a) that problem (P_p) can effectively be solved via the tSLR model introduced in this paper.

The reason for using the DPS scheme rather than established global solvers for constrained optimization is the fact that it can be implemented with fully open-source tools available within the `scipy.optimize` Python environment. We apply this framework to six well-known toy examples from the literature, and for which global optimal solutions have been reported. We label these problems as P1 (see, e.g., [5, Example 2.1]), P2 (see, e.g., [6, Example 3.1]), P3 (see, e.g., [2, Principal Agent problem in Table 5]), P4 (see, e.g., [14, Example 4.1]), P5 (see, e.g., [49, Example 1]), and P6 (see, e.g., [49, Example 2]). All the six examples satisfy Assumption 4.7; hence, we also apply DPS to the corresponding versions of the model (tSLR-LP). Additionally, we apply DPS to the SLRs (MM-CC) and (SP-CC) to enable a comparison of their performance with that of our tSLRs. Out of all the existing methods in the literature presented in Section 2, the SIP-PBDA [43] and TLVF-GSS1 [42] algorithms seem to be the only ones that can directly approximate global optimal solutions. Hence, we also apply them to applicable problems from our six examples (i.e., P1 and P2 for SIP-PBDA and TLVF-GSS1 for the rest). Note that here, we also use the DPS scheme as global solver for the inner problem for SIP-PBDA (Algorithm 1 in [43]).

Method	P1	P2	P3	P4	P5	P6
MM-CC	✓	✓	✓	✓	✗	✗
SP-CC	✗	✓	✗†	✓	✓	✗
tSLR	✓	✓	✓	✓	✓	✓
tSLR-LP	✓	✓	✓	✓	✓	✓
SIP-PBDA	✓	✓	–	–	–	–
TLVF-GSS1	–	–	✓	✓	✓	✗

Table 1: Success (✓, $x_{\text{err}} \leq 10^{-3}$ and feasible) or failure (✗) on each of of the problems P1, . . . , P6; "–" marks a method outside its applicability scope for that problem. Exact x_{err} values are given in Fig. 4, not repeated here. (†: SP-CC on P3 has small x_{err} (9.6×10^{-14}) but is marked as failure because the returned point does not satisfy the 10^{-4} feasibility tolerance.)

DPS initialization. Each DPS solve for the corresponding versions of MM-CC, SP-CC, tSLR, and tSLR-LP uses 12 deterministic starting points for the local SLSQP polish stage: one from a global DIRECT search on the exterior-penalized objective (penalty weight $\rho = 10^5$, evaluation budget $N_{\max} = \min\{3000\hat{n}, 80000\}$ with $\hat{n}$ being the number of variables of the corresponding problems), plus 11 further points obtained by scaling every coordinate to the same fraction $\frac{k}{12}$, where $k = 1, \dots, 11$, of its bound interval–a fixed, reproducible design with no random seed. The best feasible SLSQP result across all 12 starts is retained.

Method	Scope	Success rate	Mean x_{err}	Mean time (s)
tSLR	6	**100.0%**	2.4×10^{-11}	1.06
tSLR-LP	6	**100.0%**	4.5×10^{-12}	0.53
SIP-PBDA	2	**100.0%**	1.8×10^{-15}	0.11
TLVF-GSS1	4	75.0%	2.50	1.00
MM-CC	6	66.7%	0.37	1.09
SP-CC	6	50.0%	0.91	7.78

Table 2: Aggregate performance over each method's applicable scope within problems P1, . . . , P6. Scope refers to the number of problems that the corresponding method can solve.

Performance measures. Considering the focus of Corollary 4.5(a) on computing optimal solutions of problem (P_p) from the tSLR, the primary measure is the leader solution error

$$x_{\text{err}} := \|\tilde{x} - x^*\|_\infty,$$

the distance between the computed leader decision $\tilde{x}$ and the known true optimizer x^*. A run of the DPS scheme will be said to be a *success* if $x_{\text{err}} \leq 10^{-3}$ and $\tilde{x}$ is feasible. Table 1 presents the behavior of the methods w.r.t. their success or not when applied to the different problems (P1, . . . , P6), while Table 2 provides the average error and time performances.

Discussion. tSLR and tSLR-LP are the clear winners: both reach 100% success with error at or below machine precision on every problem, matching the scope-limited bespoke methods SIP-PBDA and TLVF-GSS1 without their applicability restrictions. TLVF-GSS1 is close behind at 75%, missing only P6–consistent with a known local-trap limitation of its global optimality conditions (GOC)–escape mechanism on that problem's degenerate lower-level reported [42]. MM-CC

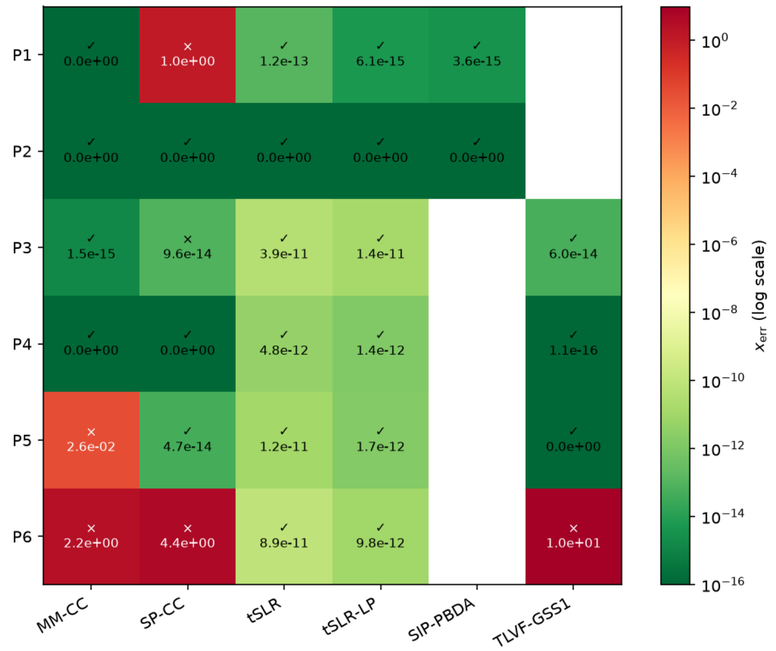


Fig. 4: Leader's solution error $x_{\rm err}$ across all problems and all six methods (log color scale; green = accurate, red = inaccurate). Empty cells represent cases outside a method's applicability scope.

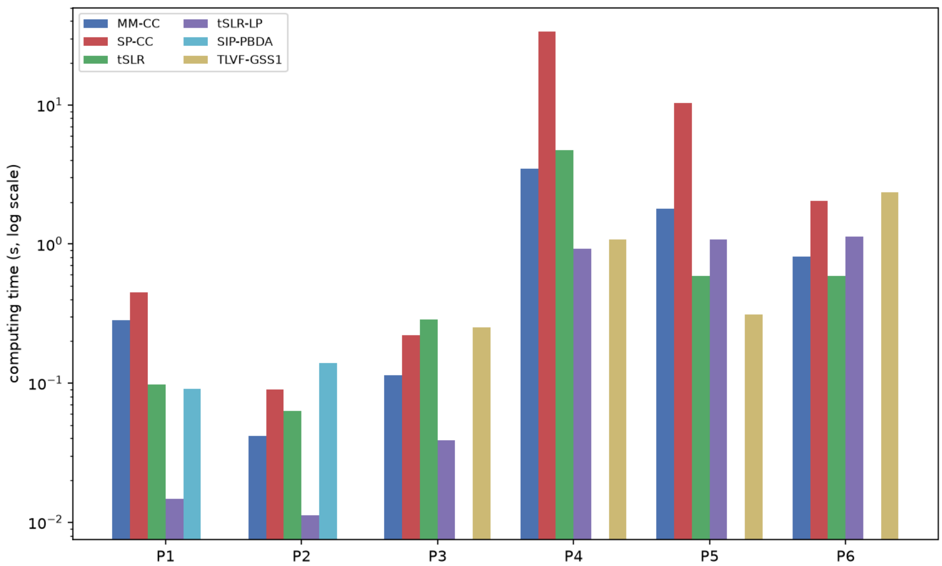


Fig. 5: A missing bar means the method is outside its applicability scope for that problem. SP-CC's cost is the clear outlier, most visibly on P4 (33.5s, an order of magnitude above every other method on that problem), driven by its larger multiplier-variable count inflating each SLSQP polish; tSLR-LP is consistently among the cheapest methods run.

and SP-CC, the two generic complementarity-penalty reformulations, are markedly less reliable (66.7% and 50.0%): both fail on problem P6, and each additionally fails one of P1, P3, or P5 by landing on a spurious stationary point of the MPCC reformulation rather than the true bilevel optimum–a known hazard of penalty-based reformulations that exact gradients and multi-start restarts substantially mitigate but do not eliminate on this problem set. SP-CC is additionally the slowest method by a wide margin (mean 7.78s vs. *less than* 1.1s for every other method), owing to its larger multiplier-variable count inflating the cost of each SLSQP polish – most visibly on P4, where it takes 33.5s against at most a few seconds for every other method. tSLR-LP is both the most reliable method and consistently one of the cheapest, making it the strongest method overall on this problem range by every measure reported here. The individual computing time for each method, where applicable, is provided in Fig. 5; and it is clear that except from P5, tSLR or tSLR-LP has the best computing time on each problem.

7 Concluding comments. In this paper, we introduce a single-level reformulation (SLR) for the pessimistic bilevel optimization problem, which we call a *true single-level reformulation* (tSLR) because, unlike conventional SLRs for bilevel optimization, it involves neither complementarity conditions nor value functions. We establish rigorous global and local relationships between the tSLR and the original problem (P_p). The proposed reformulation offers several important advantages. In particular, it contains no implicit variables and can satisfy the linear independence constraint qualification (LICQ), whereas even the weaker Mangasarian–Fromovitz constraint qualification (MFCQ) systematically fails for all known SLRs of optimistic and pessimistic bilevel programs. Moreover, the tSLR enables us to derive the first sufficient optimality conditions for the pessimistic bilevel optimization problem. On the numerical side, we provide a proof of concept using six illustrative examples, demonstrating the potential of the tSLR for computing globally optimal solutions. Future work will focus on developing numerical methods tailored to problem (tSLR) and systematically comparing them on suitable test problems with existing algorithms for pessimistic bilevel optimization reviewed in Section 2.

REFERENCES

[1] A. Aboussoror, S. Adly, and F. E. Saissi, *Strong-weak nonlinear bilevel problems: existence of solutions in a sequential setting*, Set-Valued and Variational Analysis, 25 (2017), pp. 113–132.
[2] M. Antoniou, A. Sinha, and G. Papa, *δ-perturbation of bilevel optimization problems: An error bound analysis*, Operations Research Perspectives, 13 (2024), p. 100315.
[3] D. Aussel and A. Svensson, *Is pessimistic bilevel programming a special case of a mathematical program with complementarity constraints?*, Journal of Optimization Theory and Applications, 181 (2019), pp. 504–520.
[4] Y. Beck, I. Ljubić, and M. Schmidt, *Linear and Mixed-Integer Bilevel Optimization*, Cambridge University Press, 2027.
[5] I. Benchouk, L. Jolaoso, K. Nachi, and A. Zemkoho, *Scholtes relaxation method for pessimistic bilevel optimization*, Set-Valued and Variational Analysis, 33 (2025), p. 10.
[6] I. Benchouk, L. Jolaoso, K. Nachi, and A. Zemkoho, *Relaxation methods for pessimistic bilevel optimization*, Set-Valued and Variational Analysis, 34 (2026), p. 1.
[7] H. I. Calvete, C. Galé, A. Hernández, and J. A. Iranzo, *A novel approach to pessimistic bilevel problems. an application to the rank pricing problem with ties*, Optimization, 74 (2025), pp. 2823–2856.
[8] Q. Cao, S. Zeng, and J. Zhang, *A single-loop gradient algorithm for pessimistic bilevel optimization via smooth approximation*, Advances in Neural Information Processing Systems, 38 (2026), pp. 23039–23080.
[9] M. Červinka, C. Matonoha, and J. V. Outrata, *On the computation of relaxed pessimistic solutions to mpecs*, Optimization Methods and Software, 28 (2013), pp. 186–206.
[10] S. Cipolla, O. Stein, and A. Zemkoho, *A single loop method for quadratic minmax optimization*, (2026), https://optimization-online.org/?p=33225.
[11] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Mathematical programming, 75 (1996), pp. 407–439.
[12] S. Dempe, *Foundations of bilevel programming*, Kluwer Academic Publishers, 2002.
[13] S. Dempe, J. Dutta, and B. Mordukhovich, *New necessary optimality conditions in optimistic bilevel programming*, Optimization, 56 (2007), pp. 577–604.
[14] S. Dempe, G. Luo, and S. Franke, *Pessimistic bilevel linear optimization*, Journal of Nepal Mathematical Society, 1 (2018), pp. 1–10.
[15] S. Dempe and P. Mehlitz, *Duality-based single-level reformulations of bilevel optimization problems*, Journal of Optimization Theory and Applications, 205 (2025), p. 26.
[16] S. Dempe, B. S. Mordukhovich, and A. B. Zemkoho, *Sensitivity analysis for two-level value functions with applications to bilevel programming*, SIAM Journal on Optimization, 22 (2012), pp. 1309–1343.
[17] S. Dempe, B. S. Mordukhovich, and A. B. Zemkoho, *Necessary optimality conditions in pessimistic bilevel programming*, Optimization, 63 (2014), pp. 505–533.
[18] S. Dempe, B. S. Mordukhovich, and A. B. Zemkoho, *Two-level value function approach to non-smooth optimistic and pessimistic bilevel programs*, Optimization, 68 (2019), pp. 433–455.
[19] S. Dempe and H. Schmidt, *On an algorithm solving two-level programming problems with nonunique lower level solutions*, Computational Optimization and Applications, 6 (1996), pp. 227–249.
[20] S. Dempe and A. Zemkoho, *Bilevel optimization*, in Springer optimization and its applications, vol. 161, Springer, 2020, pp. 1–672.
[21] S. Dempe and A. B. Zemkoho, *On the karush–kuhn–tucker reformulation of the bilevel optimization problem*, Nonlinear Analysis: Theory, Methods & Applications, 75 (2012), pp. 1202–1218.
[22] A. Fischer, A. B. Zemkoho, and S. Zhou, *Semismooth newton-type method for bilevel optimization: global convergence and extensive numerical experiments*, Optimization methods and software, 37 (2022),

pp. 1770–1804.

[23] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, 2000.

[24] J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Mathematical Programming, 12 (1977), pp. 136–138.

[25] R. Grazzi, M. Pontil, S. Salzo, and A. Zemkoho, *Bilevel learning*, arXiv preprint arXiv:2605.01621, (2026).

[26] Z. Guan, D. Sow, S. Lin, and Y. Liang, *Adaprox: A novel method for bilevel optimization under pessimistic framework*, in The Second Conference on Parsimony and Learning (Proceedings Track), 2025.

[27] L. Guo, G.-H. Lin, and J. J. Ye, *Second-order optimality conditions for mathematical programs with equilibrium constraints*, Journal of Optimization Theory and Applications, 158 (2013), pp. 33–64.

[28] J. Y. Jane, *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints*, Journal of Mathematical Analysis and Applications, 307 (2005), pp. 350–369.

[29] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, *Lipschitzian optimization without the lipschitz constant*, Journal of Global Optimization, 3 (1993), pp. 157–181, https://doi.org/10.1007/BF00941892.

[30] D. Jungen, A. Zingler, H. Djelassi, and A. Mitsos, *libdips–discretization-based semi-infinite and bilevel programming solvers: D. jungen et al.*, Mathematical programming computation, (2026), pp. 1–56.

[31] T. Kis, A. Kovács, and C. Mészáros, *On optimistic and pessimistic bilevel optimization models for demand response management*, Energies, 14 (2021), p. 2095.

[32] P.-M. Kleniati and C. S. Adjiman, *Branch-and-sandwich: a deterministic global optimization algorithm for optimistic bilevel programming problems. part i: Theoretical development*, Journal of Global Optimization, 60 (2014), pp. 425–458.

[33] D. Kraft, *A software package for sequential quadratic programming*, Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt, (1988).

[34] L. Lampariello and S. Sagratella, *A bridge between bilevel programs and nash games*, Journal of Optimization Theory and Applications, 174 (2017), pp. 613–635.

[35] L. Lampariello, S. Sagratella, and O. Stein, *The standard pessimistic bilevel problem*, SIAM Journal on Optimization, 29 (2019), pp. 1634–1656.

[36] J. Liu, Y. Hong, and Y. Zheng, *A new variant of penalty method for weak linear bilevel programming problems*, Wuhan University Journal of Natural Sciences, 23 (2018), pp. 328–332.

[37] P. Loridan and J. Morgan, *Weak via strong stackelberg problem: new results*, Journal of global Optimization, 8 (1996), pp. 263–287.

[38] A. Mitsos, P. Lemonidis, and P. I. Barton, *Global solution of bilevel programs with a nonconvex inner program*, Journal of Global Optimization, 42 (2008), pp. 475–513.

[39] D. Molodtsov, *The solution of a class of non-antagonistic games*, USSR Computational Mathematics and Mathematical Physics, 16 (1976), pp. 67–72, https://doi.org/https://doi.org/10.1016/0041-5553(76)90042-2, https://www.sciencedirect.com/science/article/pii/0041555376900422.

[40] D. Molodtsov and V. Fedorov, *Approximation of two-person games with information exchange*, USSR Computational Mathematics and Mathematical Physics, 13 (1973), pp. 123–142, https://doi.org/https://doi.org/10.1016/0041-5553(73)90010-4, https://www.sciencedirect.com/science/article/pii/0041555373900104.

[41] Y. Shehu, P. T. Vuong, and A. Zemkoho, *An inertial extrapolation method for convex simple bilevel optimization*, Optimization Methods and Software, 36 (2021), pp. 1–19.

[42] A. Strekalovsky, *One way to solve pessimistic bilevel optimization problems*, Optimization, 74 (2025), pp. 4795–4825.

[43] W. Wiesemann, A. Tsoukalas, P.-M. Kleniati, and B. Rustem, *Pessimistic bilevel optimization*, SIAM Journal on Optimization, 23 (2013), pp. 353–380.

[44] P. Wolfe, *A duality theorem for non-linear programming*, Quarterly of applied mathematics, 19 (1961), pp. 239–244.

[45] S. Wright, J. Nocedal, et al., *Numerical optimization*, Springer Science, 35 (1999), p. 7.

[46] J. J. Ye and D. Zhu, *New necessary optimality conditions for bilevel programs by combining the mpec and value function approaches*, SIAM Journal on Optimization, 20 (2010), pp. 1885–1905.

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

[48] A. B. Zemkoho, *Solving ill-posed bilevel programs*, Set-Valued and Variational Analysis, 24 (2016), pp. 423–448.

[49] B. Zeng, *A practical scheme to compute the pessimistic bilevel optimization problem*, INFORMS Journal on Computing, 32 (2020), pp. 1128–1142.