
A true single-level reformulation for pessimistic bilevel optimization

Oliver Stein · Alain Zemkoho

Abstract We propose a single-level reformulation (SLR) for the pessimistic bilevel optimization problem 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 reformulation 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. We leverage on these constraint qualifications to construct new necessary optimality conditions for pessimistic bilevel optimization. The reformulation also has a striking limitation: under the assumptions of our analysis, the classical second-order sufficient condition fails at every Karush–Kuhn–Tucker point of the problem. Nevertheless, 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.

Keywords Pessimistic bilevel optimization · Single-level reformulation · Constraint qualifications · Optimality conditions · Numerical methods

Mathematics Subject Classification (2020) 90C26 · 90C30 · 90C46 · 90C47

Acknowledgements The authors are grateful to Alireza Kabgani and Martin Schmidt for fruitful discussions on an earlier version of this manuscript. The work of the second author was partly funded by the Alexander von Humboldt Foundation, through an Alexander von Humboldt Research Fellowship for Experienced Researchers held within the Continuous Optimization Chair at the Institute for Operations Research, Karlsruhe Institute of Technology (KIT).

1 Introduction

In this paper, we consider the bilevel optimization problem

$$\underset{x}{\text{“min”}} F(x, y) \text{ s.t. } x \in X, y \in S_L(x) := \underset{z \in Y(x)}{\operatorname{argmin}} f(x, z), \quad (\text{BOP})$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ (resp. $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$) corresponds to the upper-level (resp. lower-level) objective function, while the set-valued mapping $Y : X \rightrightarrows \mathbb{R}^m$ describes the lower-level feasible set

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

Oliver Stein
Institute for Operations Research (IOR)
Karlsruhe Institute of Technology
E-mail: stein@kit.edu
ORCID: <https://orcid.org/0000-0001-9514-6317>

Alain Zemkoho
School of Mathematical Sciences
University of Southampton
E-mail: a.b.zemkoho@soton.ac.uk
ORCID: <https://orcid.org/0000-0003-1265-4178>

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*).

In problem (BOP), the vector x (resp. y) represents the upper-level (resp. lower-level) variable. Throughout the paper, the lower-level variable is *continuous*. However, for the upper-level variable, it can be *mixed-integer*; i.e., we could assume that the upper-level feasible set is such that $X \subset \mathbb{Z}^{n_I} \times \mathbb{R}^{n_C}$ with $n_I + n_C = n$ for the construction of the main model proposed in this paper. Hence, throughout the text, unless otherwise stated, it would be assumed that $X \subset \mathbb{Z}^{n_I} \times \mathbb{R}^{n_C}$.

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*

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

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 player, i.e., $|S_L(x)| = 1$ (with $|\cdot|$ denoting the cardinality of the corresponding set) for all $x \in X$, then problem (BOP) can be reduced to the optimization problem

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

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 $y(\cdot) : X \rightarrow \mathbb{R}^m$ 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 in this case. To make the problem mathematically tractable in this situation, 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

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

with

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

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 full 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 themselves against potential worst choices from the follower by solving the pessimistic bilevel optimization problem

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

with

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

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 [49] for a set-valued optimization approach to tackle problem (BOP) when $|S_L(x)| > 1$ for some value(s) of $x \in X$.

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) , however, instead through its standard optimistic reformulation [20]. As for the pessimistic bilevel optimization problem (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 [44] or models closely related to the standard pessimistic problem introduced in [36]; 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., [6, 9, 10].

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 [6, 49]. 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

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

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

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

(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 [15]. Moreover, our proposed tSLR can accommodate mixed-integer upper-level variables.

(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; see, e.g., [20, 12, 47].

Leveraging the possibility that these constraint qualifications can hold for our tSLR, we derive completely new classes of necessary optimality conditions for (P_p) . However, this paper reveals a fundamental limitation for our tSLR: we prove that it systematically fails to satisfy the classical second-order sufficient condition (see Section 3 for the definition of the concept) at its Karush–Kuhn–Tucker points within the framework considered in this paper. Hence, despite its compatibility with classical constraint qualifications, our tSLR violates the classical second-order sufficient condition.

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,

$$S_p(x) := \arg \max_{y \in S_L(x)} F(x, y). \quad (1.6)$$

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 [36].

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

Under this assumption, not only the two-level value function φ_p (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 (1.5).

Throughout the paper, we use the following basic optimal solution concept for problem (P_p) .

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

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

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

Of course, this is not the only optimal solution notion for problem (P_p) ; for a detailed study of solution concepts for the problem, interested readers are referred to the article [3].

The following example gives a flavor of the extraordinary nature of the true single-level reformulation of problem (P_p) introduced in this paper.

Example 1.3 Consider a class of problem (P_p) with the upper- and lower-level objective functions $F : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ and $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ respectively defined by

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

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

$$X := \{x \in \mathbb{Z}^{n_I} \times \mathbb{R}^{n_C} \mid A^G x \leq b^G\} \quad \text{and} \quad Y(x) := \{y \in \mathbb{R}^m \mid A^g x + B^g y \leq b^g\}, \quad (1.9)$$

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:

$$\begin{aligned} \min_{x, y, z, u, w} \quad & F(x, y) - u^\top (A^g x + B^g z - b^g) - w(y - z)^\top c^f \\ \text{s.t.} \quad & Q_{12}^\top x + Q_{22} y - (B^g)^\top u = c_y, \quad x \in \mathbb{Z}^{n_I} \times \mathbb{R}^{n_C}, \\ & A^G x \leq b^G, \quad A^g x + B^g z \leq b^g, \quad u \geq 0, \quad w \geq 0. \end{aligned} \quad (1.10)$$

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 problem (1.10). Obviously, problem (1.10) can be solved by any suitable 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. $\square$

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) .

For the remainder of the paper, note that in the next section, we conduct a detailed survey of the different existing 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 the case of purely continuous upper level variables; in particular, we construct tractable frameworks ensuring the automatic fulfillment of these constraint qualifications. Subsequently, we use the MFCQ and LICQ to derive completely new *first order* and *second order* necessary optimality conditions for problem (P_p) , respectively, under conditions not affordable even in the context of the optimistic bilevel program. In Section 5, however, we also prove that the classical second order sufficient condition systematically fails for the tSLR problem developed in this paper. 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 can lead 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

We start here by recalling that based on the lower-level optimal value function from (1.5), the lower-level solution set-valued mapping can be written as

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

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

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

we have

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

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 [11]). [50] addresses the implicit nature of the coupled constraint in (P_p) with the new minmax model

$$\min_{x \in X, z \in Y(x)} \max_{y \in \mathcal{S}_L(x, z)} F(x, y), \quad (\text{MM})$$

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

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

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 [50] or [5, Theorem 2.5]) that $A = \text{proj}_x B$, where proj stands for the parallel projection mapping.

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

$$\min_{x, y, z} F(x, y) \text{ s.t. } x \in X, z \in Y(x), y \in K(x, z) := \arg \max_{y \in \mathcal{S}_L(x, z)} F(x, y), \quad (2.2)$$

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

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

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 of the classical 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 the development of numerical methods in the articles [37, 50].

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 concepts) can work as it is the case for the KKT reformulation for the optimistic bilevel optimization problem; 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 numerically solving the KKT reformulation of the standard optimistic bilevel optimization problem. As it can be sensed from the introductory example in (1.10), the true single-level reformulation 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 [36]:

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

(P_p) is globally equivalent to (SP) in the sense of [36, Proposition 4.1]. In [36], the “bridge construction” for simple bilevel programs from [35] 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

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

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

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

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

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

The global equivalence of (SP-CC) to (MFG) in the sense of [36, 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 of implicit variables with the additional presence of lower-level Lagrange multipliers. Clearly, the number of variables and constraints in (SP-CC) is significantly larger, and an additional algorithmically challenging (lower-level) complementarity system appears. We also remark that, due to the latter issue, in [36], (SP-CC) is also solved in a mixed-integer reformulation using additional binary variables to address the complementarity constraints. Although usually from completely different angles, reformulations of (P_p) of a flavor similar to (SP-CC) have been derived in other papers [14, 32, 8] and used to solve the problem, usually from the perspective of established techniques in optimistic bilevel optimization such as 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

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

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

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

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 [47].

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., [48] and references therein) that problem (P_p) is globally equivalent to the generalized semi-infinite programming problem

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

In the paper [44], the standard semi-infinite programming approximation

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

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 develop algorithms to globally solve optimistic bilevel optimization problems in many papers; see, e.g., [39, 31, 33].

2.3 Two-level value function-based reformulations

We assume in this subsection that $X \subset \mathbb{R}^n$. 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 [10], 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 (P_p) .

In [43], the idea of Molodtsov [40, 41] 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

$$\varphi_o^{\theta, \epsilon}(x) := \min_y \left\{ F(x, y) \mid y \in S_L^{\theta, \epsilon}(x) \right\}, \quad (2.5)$$

where $\theta > 0$ is penalization parameter and $\epsilon > 0$ 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.,

$$S_L^{\theta, \epsilon}(x) := \{y \in Y(x) \mid f(x, y) - \theta F(x, y) \leq \varphi_L^\theta(x) + \epsilon\},$$

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

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

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

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

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 [40] that

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

This result is established under weaker assumptions in [38] and a similar framework for optimistic bilevel optimization is studied in the paper [19].

One fact that is clear from the Molodtsov framework is that the pessimistic bilevel optimization problem is approximated by an optimistic bilevel program; and in fact for $\theta = 0$ and $\epsilon = 0$, the optimistic bilevel program is recovered. Hence, in [43], the core optimization problem solved is a standard optimistic-type transformation resulting from (2.5).

Under the assumption that $Y(x) := Y$ for all $x \in X$, the article [9] suggests an approximation of the two-level value function φ_p by the minmax value function

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

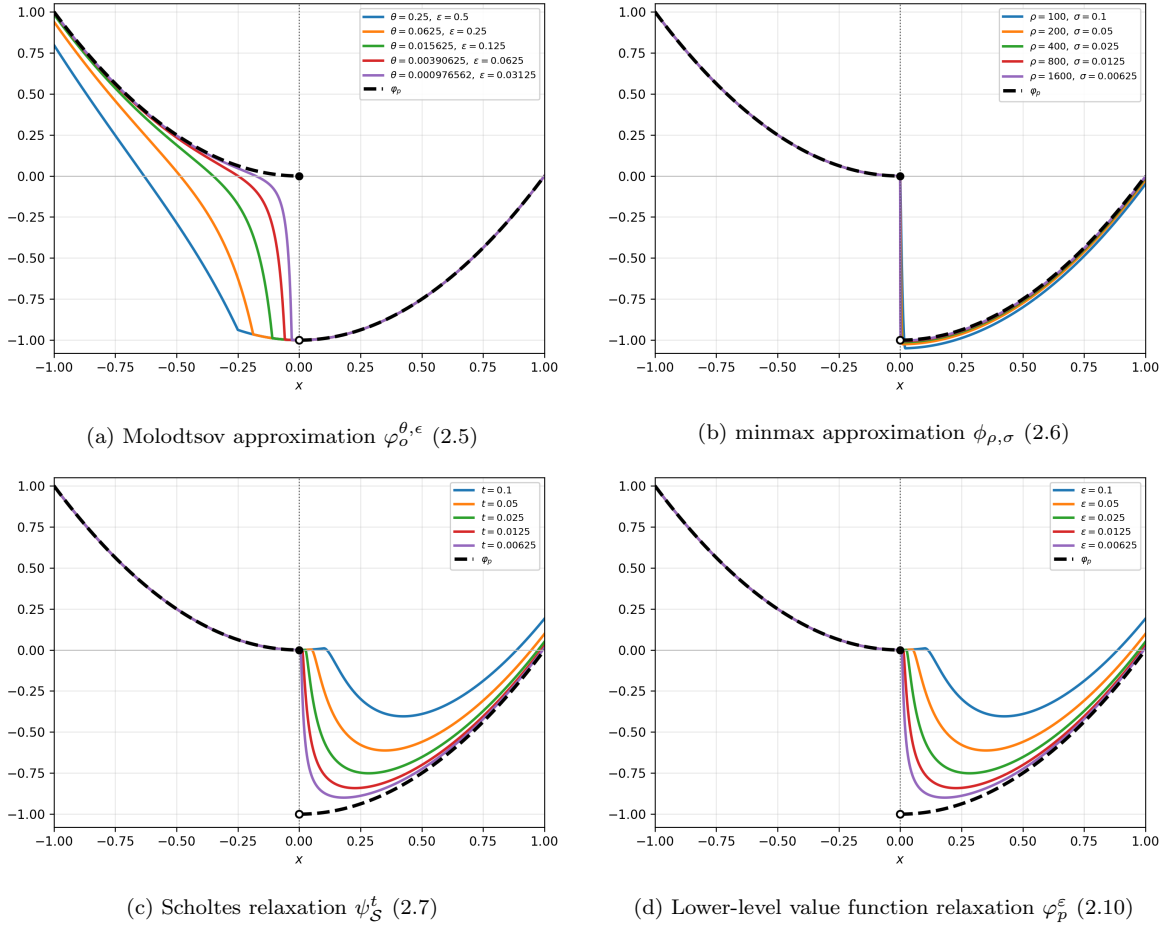


Fig. 1: Approximations of the two-level value function φ_p (1.4) for the pessimistic bilevel program (P_p) with $X = [-1, 1]$, $Y = [0, 1]$, $F(x, y) = x^2 - y^2$, and $f(x, y) = -xy$. The reference value function is $\varphi_p(x) = x^2$ for $x \leq 0$ and $\varphi_p(x) = x^2 - 1$ for $x > 0$, with a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

(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

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

if X or Y is a bounded set and the sequences $\{\rho_k\}$ and $\{\sigma_k\}$ are such that $\rho_k \rightarrow \infty$ and $\sigma_k \rightarrow 0$. 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 [9].

Finally, in the series of papers [6, 7], 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

$$\psi_{\mathcal{R}}^t(x) := \max_{(y, u) \in \mathcal{D}_{\mathcal{R}}^t(x)} F(x, y), \quad (2.7)$$

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

$$\mathcal{D}_{\mathcal{R}}^t(x) := \{(y, u) \in \mathbb{R}^{m+q} \mid \nabla_y \ell(x, y, u) = 0, \phi_{i, \mathcal{R}}^t(x, y, u) \leq 0, i = 1, \dots, q\}, \quad (2.8)$$

where $\ell(x, y, u) := f(x, y) + u^\top g(x, y)$ denotes the lower-level Lagrangian function with the assumption that $Y(x) := \{y \in \mathbb{R}^m \mid g(x, y) \leq 0\}$. 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 $i = 1, \dots, q$, the function $\phi_{i, \mathcal{S}}^t$

is defined for a triplet (x, y, u) by

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

Such a relaxation creates advantages; for instance, (2.7) provides an upper bound for φ_p as if the function $t \mapsto \psi_{\mathcal{R}}^t(x)$ 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 $\psi_{\mathcal{R}}^{t_k}(x) \rightarrow \psi_p(x)$ as $k \rightarrow \infty$. 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

$$\min_{x \in X} \psi_{\mathcal{R}}^t(x) \quad (2.9)$$

for $t := t_k$. If $x^k \rightarrow \bar{x}$ as $k \rightarrow \infty$, then $\bar{x}$ is a global optimal solution of (P_p) provided that the function $t \mapsto \psi_{\mathcal{R}}^t(x)$ 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 [6, 7].

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. 1. 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. 1).

Note that in Fig. 1, we also included the value function-based approximation of φ_p defined by

$$\varphi_p^\varepsilon(x) := \max_y \{F(x, y) \mid y \in Y(x), f(x, y) - \varphi_L(x) \leq \varepsilon\}, \quad (2.10)$$

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 found in any standard book on continuous optimization; see, e.g., [22, 46]. The focus in this subsection will be the constrained optimization problem

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

with continuously differentiable functions $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^p$, and $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^q$. Let $\mathcal{I}(\bar{x}) := \{i := 1, \dots, p \mid \mathbf{g}_i(\bar{x}) = 0\}$. For ease of notation, in the sequel, we use $\mathcal{I}$ instead of $\mathcal{I}(\bar{x})$. Let $\nabla \mathbf{g}_{\mathcal{I}}(\bar{x})$ be the submatrix of the Jacobian matrix $\nabla \mathbf{g}(\bar{x})$ 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

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

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:

$$\left. \begin{aligned} \nabla \mathbf{g}(\bar{x})^\top \alpha + \nabla \mathbf{h}(\bar{x})^\top \beta &= 0 \\ \alpha &\geq 0, \quad \alpha^\top \mathbf{g}(\bar{x}) = 0 \end{aligned} \right\} \implies [\alpha = 0, \beta = 0]. \quad (\text{MFCQ})$$

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 $\bar{x}$:

$$T_{\mathcal{C}}(\bar{x}) := \left\{ d \in \mathbb{R}^n \mid \begin{aligned} &\exists \{x^k\}_{k \in \mathbb{N}} \subset \mathcal{C}, \quad \exists \{t_k\}_{k \in \mathbb{N}} \subset (0, \infty) : \\ &x^k \rightarrow \bar{x}, \quad t_k \downarrow 0, \quad (x^k - \bar{x})/t_k \rightarrow d \end{aligned} \right\}.$$

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.,

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

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

$$\mathcal{K}^* := \{ v \in \mathbb{R}^n \mid v^\top d \geq 0 \text{ for all } d \in \mathcal{K} \}. \quad (3.2)$$

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.,

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

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

$$\nabla \mathbf{f}(\bar{x}) + \nabla \mathbf{g}(\bar{x})^\top \alpha + \nabla \mathbf{h}(\bar{x})^\top \beta = 0, \quad (3.3)$$

$$\alpha \geq 0, \quad \mathbf{g}(\bar{x}) \leq 0, \quad \alpha^\top \mathbf{g}(\bar{x}) = 0, \quad (3.4)$$

$$\mathbf{h}(x) = 0. \quad (3.5)$$

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

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

Now, let x be such that we can find Lagrange multipliers $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$ such that the optimality conditions (3.3)–(3.5) are satisfied. If we assume that the second order sufficient condition (SOSC)

$$d^\top \nabla_{xx}^2 \ell(x, \alpha, \beta) d > 0 \quad \text{for all } d \in \mathfrak{C}(x, \alpha) \setminus \{0\} \quad (\text{SOSC})$$

holds, then x is a strict local optimal solution of problem (3.1); see, e.g., [4] for more details on (SOSC). Note that here, $\mathfrak{C}(x, \alpha)$ denotes the critical cone associated to (3.1), and which is defined by

$$\mathfrak{C}(x, \alpha) := \left\{ d \in \mathbb{R}^n \mid \begin{array}{l} \nabla \mathbf{g}_i(x)^\top d = 0 \quad \forall i : \mathbf{g}_i(x) = 0, \quad \alpha_i > 0 \\ \nabla \mathbf{g}_i(x)^\top d \leq 0 \quad \forall i : \mathbf{g}_i(x) = 0, \quad \alpha_i = 0 \\ \nabla \mathbf{h}_j(x)^\top d = 0 \quad \forall j = 1, \dots, q \end{array} \right\}. \quad (3.6)$$

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

$$\begin{aligned} \max_{x, \alpha, \beta} \quad & \ell(x, \alpha, \beta) \\ \text{s.t.} \quad & \nabla_x \ell(x, \alpha, \beta) = 0, \quad \alpha \geq 0. \end{aligned} \quad (3.7)$$

Then, we can state the following strong Wolfe duality result from [45, 22]:

Lemma 3.1 *For problem (3.1), let the functions $\mathbf{f}$ and $\mathbf{g}_i$, for $i = 1, \dots, p$, be convex, while $\mathbf{h}$ is an affine linear function. Furthermore, let the point x be an optimal solution of problem (3.1) that satisfies (GCQ). Then, there exist Lagrange multipliers $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$ such that (x, α, β) is an optimal solution of problem (3.7) and it holds that $\mathbf{f}(x) = \ell(x, \alpha, \beta)$.*

3.2 Parametric and minmin optimization

The focus of this subsection will be on the *minmin* optimization problem

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

which involves an *outer* (resp. *inner*) minimization w.r.t. to the outer (resp. inner) variable $x \in \mathbb{R}^n$ (resp. $y \in \mathbb{R}^m$). $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ represents the objective function of $(\mathcal{P}_{2m})$, while $\mathcal{X} \subseteq \mathbb{R}^n$ corresponds to 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 $\mathcal{S}: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$:

$$\mathcal{S}(x) := \arg \min_{y \in \mathcal{Y}(x)} f(x, y) \quad (3.8)$$

and the corresponding optimal value function defined by

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

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

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

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 $(\mathcal{P}_{2m})$:

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

Denote by $\Omega := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid x \in \mathcal{X}, y \in \mathcal{Y}(x)\}$; $(\bar{x}, \bar{y})$ 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

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

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 problem $(\mathcal{P}_{1m})$ with $\bar{y} \in \mathcal{S}(\bar{x})$ when we compare local optimal solutions of problem $(\mathcal{P}_{1m})$ with those of problem $(\mathcal{P}_{2m})$ (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 problem $(\mathcal{P}_{1m})$. 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 $y \in \mathcal{Y}(x)$,

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

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

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

Note that $(\bar{x}, \bar{y})$ 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

$$\phi(\bar{x}) = f(\bar{x}, \bar{y}) \leq f(x, y), \quad (3.9)$$

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

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

it follows from (3.9) and the arbitrary choice of $y \in \mathcal{Y}(x)$ that

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

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

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 $(\bar{x}, \bar{y}) \in \text{gph } \mathcal{S}$ if for every sequence $x^k \rightarrow \bar{x}$, there exists a sequence $y^k \in \mathcal{S}(x^k)$ such that $y^k \rightarrow \bar{y}$. Note that for any given set-valued mapping Ψ , $(x, y) \in \text{gph } \Psi$ 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 (3.8).

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

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 $x^k \rightarrow \bar{x}$ and $y^k \rightarrow \bar{y}$ such that

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

Then considering the definition of ϕ , it follows that

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

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 $x^k \rightarrow \bar{x}$ 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

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

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

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:

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

We can easily check that the point $(0, 0)$ is a local optimal solution of the corresponding problem $(\mathcal{P}_{1m})$, while 0 is not a local optimal solution of problem $(\mathcal{P}_{2m})$. 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 new single-level reformulations

In this section, we will introduce a true single-level reformulation for (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 subsequent analysis it is crucial to note that the pessimistic bilevel problem (P_p) actually possesses a three-level structure. In the lower-level problem $(LL(x))$ the function $f(x, \cdot)$ is minimized over $Y(x)$, in the intermediate-level problem $(IL(x))$ the function $F(x, \cdot)$ is maximized over $S_L(x)$, while 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 problem's objective function φ_p . To this end, we write the intermediate problem in its equivalent optimal-value function formulation with $Z(x)$ given in (2.1),

$$\max_y F(x, y) \quad \text{s.t.} \quad y \in Z(x) = \{y \in Y(x) \mid f(x, y) - \varphi_L(x) \leq 0\}. \quad (IL^v(x))$$

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., [42], 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 (3.7) to $(IL^v(x))$ for fixed $x \in X$. Indeed, this Wolfe dual consists in the minimization of the Lagrangian

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

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

$$A_{p_D}(x) := \left\{ (y, u, v, w) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R} \mid \begin{array}{l} u \geq 0, \quad w \geq 0 \\ \nabla_y \mathcal{L}_{p_D}^L(x, y, u, v, w) = 0 \end{array} \right\}.$$

By

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

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

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

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

$$S_{p_D}^L(x) := \arg \min_{(y, u, v, w) \in A_{p_D}(x)} \mathcal{L}_{p_D}^L(x, y, u, v, w). \quad (4.2)$$

Observe that the Lagrangian $\mathcal{L}_{p_D}^L$ 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

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

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 $w \geq 0$, the Lagrangian-type function $\mathcal{L}_{p_D}^L$ can be rewritten as

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

and that

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

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*

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

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, we have $\varphi_p(x) = F(x, y) = \mathcal{L}_{p_D}^L(x, y, u, v, w) = \varphi_{p_D}(x)$. The result therefore holds in view of (4.3). $\square$

Remark 4.2 The proof of Lemma 4.1 uses the fact that each KKT point y of $(IL^v(x))$ with corresponding multipliers (u, v, w) generates a dually optimal point $(y, u, v, w) \in S_{p_D}^L(x)$. We emphasize that *not all* elements of $S_{p_D}^L(x)$ need to correspond to KKT points of $(IL^v(x))$, 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 (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 [23] 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

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

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

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

We remark that, while the last equality constraint in problem (tSLR) originates from the condition $\nabla_y \mathcal{L}_{p_D}^L(x, y, u, v, w) = 0$, 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 problem (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 problem (tSLR) may instead be dropped. Under appropriate additional linearity assumptions on the defining functions, the variable y can indeed be eliminated completely from problem (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 problem (P_{p_D}) , then for all $(\bar{y}, \bar{u}, \bar{v}, \bar{w}) \in S_{p_D}^L(\bar{x})$, the point $(\bar{x}, \bar{y}, \bar{u}, \bar{v}, \bar{w})$ is globally optimal for problem (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 problem (P_{p_D}) .*
- (b) *If $\bar{x}$ is a local optimal solution of (P_{p_D}) , then for all $(\bar{y}, \bar{u}, \bar{v}, \bar{w}) \in S_{p_D}^L(\bar{x})$, the point $(\bar{x}, \bar{y}, \bar{u}, \bar{v}, \bar{w})$ is a local optimal solution of (pSLR). Conversely, let $(\bar{x}, \bar{y}, \bar{u}, \bar{v}, \bar{w})$, where $S_{p_D}^L$ from (4.2) is inner semicontinuous, be a local optimal solution of (pSLR), then $\bar{x}$ is a local optimal solution of (P_{p_D}) .*

Proof (a) follows from Lemma 3.3, while (b) results from Lemma 3.4. $\square$

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

$$S_L^*(x, w) := \arg \min_{z \in Y(x)} wf(x, z). \quad (4.6)$$

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

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

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

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

Subsequently, considering the fact that

$$S_L^*(x, w) := \arg \min_{z \in Y(x)} \mathcal{L}_{p_D}(x, y, z, u, v, w), \quad (4.7)$$

(a) and (b) follow from Lemma 3.3 and Lemma 3.4, respectively. $\square$

Corollary 4.5 *It holds that:*

- (a) *If the point $\bar{x}$ is globally optimal for problem (P_p) , then for all $(\bar{y}, \bar{u}, \bar{v}, \bar{w}) \in S_{p_D}^L(\bar{x})$ 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 $(\bar{y}, \bar{u}, \bar{v}, \bar{w}) \in S_{p_D}^L(\bar{x})$ 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 $S_{p_D}^L$ (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) .*

The relationships studied here are summarized in Fig. 2, where assumptions (a_0) , (a_1) , (a_2) , (a_3) , and (a_4) are respectively defined as follows:

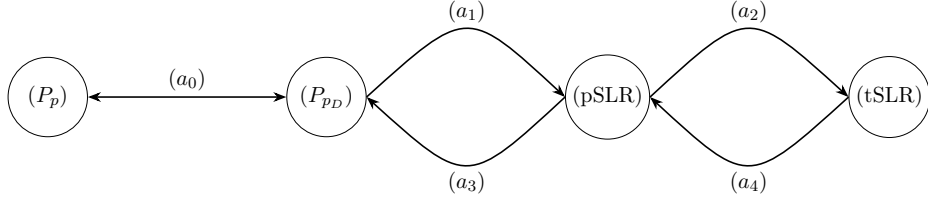


Fig. 2: Links between problems (P_p) , (P_{p_D}) , $(pSLR)$, and $(tSLR)$ with assumptions (a_1) and (a_2) needed for both the corresponding global and local implications, while (a_3) and (a_4) are the required only for the corresponding local relationships.

(a_0) Assumption 2.1;

(a_1) $(\bar{x}, \bar{y}, \bar{u}, \bar{v}, \bar{w})$ is such that $(\bar{y}, \bar{u}, \bar{v}, \bar{w}) \in S_{p_D}^L(\bar{x})$;

(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) $S_{p_D}^L$ is inner semicontinuous at $(\bar{x}, \bar{y}, \bar{u}, \bar{v}, \bar{w})$;

(a_4) S_L^* is inner semicontinuous at $(\bar{x}, \bar{w}, \bar{z})$.

Note that Assumption 1.1 implies that $S_{p_D}^L(\bar{x}) \neq \emptyset$. 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 $(\bar{y}, \bar{u}, \bar{v}, \bar{w}) \in S_{p_D}^L(\bar{x})$. 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. [42] 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 to 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

$$\begin{aligned} \min_{x, y, z, u, w} \quad & \mathcal{L}_{p_D}(x, y, z, u, w) \\ \text{s.t.} \quad & x \in X, \quad g(x, z) \leq 0, \quad u \geq 0, \quad w \geq 0, \\ & \nabla_y \mathcal{L}_{p_D}(x, y, z, u, w) = 0. \end{aligned}$$

In analogy to the fact that Wolfe duality collapses to linear programming 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

$$\begin{aligned} \min_{x,z,u,w} \quad & \gamma(x) + u^\top b(x) + w d(x)^\top z \\ \text{s.t.} \quad & x \in X, \quad A(x)z \leq b(x), \quad u \geq 0, \quad w \geq 0, \\ & c(x) - A(x)^\top u - w d(x) = 0, \end{aligned} \quad (\text{tSLR-LP})$$

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 the continuous relaxation of X is a polyhedral set, the relaxed 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

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

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 Example 1.3 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 (cf. Fig. 1). 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 4.9 Inspired by [6, Example 2.1], we consider the problem (P_p) with $n = m = 1$,

$$F(x, y) = x + y, \quad f(x, y) = xy, \quad X = [\xi, 1], \quad Y = [0, 1],$$

and a parameter $\xi < 0$. One can easily check that

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

as well as

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

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

For all $\xi \leq -1$, (P_p) possesses a global optimal solution at $\bar{x} = \xi$, while for all $\xi \in (-1, 0)$ it is not solvable since its infimum 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 $\mathcal{L}_{pD}(x, y, z, u, w) = x + (1 + u_1 - u_2 - wx)y + u_2 + wxz$, the corresponding (tSLR-LP) is

$$\begin{aligned} \min_{x,z,u,w} \quad & x + u_2 + wxz \\ \text{s.t.} \quad & \xi \leq x \leq 1, \quad 0 \leq z \leq 1, \quad u_1, u_2, w \geq 0, \quad 1 + u_1 - u_2 - wx = 0. \end{aligned}$$

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 problem (tSLR-LP) possesses the non-attained infimum 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 $x + u_2 + wxz \geq x > 0$. $\square$

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 lower semicontinuous on X . In Example 4.9, S_L is not lower 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 as the 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 optimization problem (P_p) are inherited by the single-level problem (tSLR). They can be ruled out by additional assumptions like the lower semicontinuity of S_L on X .

5 Optimality conditions

We assume throughout in this section that $X \subset \mathbb{R}^n$. 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

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

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

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, these sufficient conditions are used to derive necessary optimality conditions for (P_p) , before we study the standard second order sufficient condition in Subsection 5.3.

5.1 Constraint qualifications

To proceed here, note that the *upper-level regularity* will be said to hold at x if 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 (tSLR) if the upper-level (resp. lower-level) regularity holds at x (resp. (x, z)) and the matrix $\nabla_{yy}^2 \mathcal{L}_{p_D}(\zeta)$ is full rank.*

Proof We start with the notation

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

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:

$$\begin{aligned} \nabla G(x)^\top \alpha_G + \nabla H(x)^\top \beta_H + \nabla_x g(x, z)^\top \alpha_g + \nabla_x h(x, z)^\top \beta_h \\ + \nabla_{xy}^2 \mathcal{L}_{p_D}(\zeta)^\top \beta_{\mathcal{L}_{p_D}} = 0, \end{aligned} \quad (5.3)$$

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

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

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

With this system, if we formally write the dual form of the MFCQ from Section 3.1 for (tSLR) at $\zeta := (x, y, z, u, v, w)$, we can easily check that from the lower-level regularity at (x, 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 x . $\square$

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 $\nabla_{yy}^2 \mathcal{L}_{p_D}(\zeta)$ is full rank.*

Proof Follows along the same line as in the proof of Theorem 5.1. $\square$

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

Proposition 5.3 *If $\zeta := (x, y, u, v, w)$ is feasible for (tSLR), then $\nabla_{yy}^2 \mathcal{L}_{p_D}(\zeta)$ is full rank, provided that one of the following assumptions is satisfied:*

- (a) $\nabla_{yy}^2 F(x, y) \prec 0$;
- (b) $\nabla_{yy}^2 f(x, y) \succ 0$ and $w > 0$;
- (c) $\nabla_{yy}^2 g_i(x, y) \succ 0$ and $u_i > 0$ for some $i = 1, \dots, p$.

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

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

given that $w \geq 0$ and $u_i \geq 0$ for $i = 1, \dots, p$ (thanks to the feasibility of $\zeta := (x, y, u, v, w)$) 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 $\nabla_{yy}^2 \mathcal{L}_{p_D}(\zeta) \prec 0$, ensuring that the latter matrix is full rank. $\square$

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 establishing the first order 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:*

$$\begin{aligned} \nabla G(x)^\top \alpha_G + \nabla H(x)^\top \beta_H + \nabla_x g(x, z)^\top \alpha_g + \nabla_x h(x, z)^\top \beta_h \\ + \nabla_x \mathcal{L}_{p_D}(\zeta) + \nabla_{xy}^2 \mathcal{L}_{p_D}(\zeta)^\top \beta_{\mathcal{L}_{p_D}} = 0, \end{aligned} \quad (5.8)$$

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

$$H(x) = 0, \quad (5.10)$$

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

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

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

$$w (f(x, y) - f(x, z) + \nabla_y f(x, y)^\top \beta_{\mathcal{L}_{p_D}}) = 0, \quad (5.13)$$

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

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

$$h(x, z) = 0. \quad (5.16)$$

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

$$\alpha_u := - (g(x, y) + \nabla_y g(x, y) \beta_{\mathcal{L}_{p_D}}) \quad \text{and} \quad \alpha_w := - (f(x, y) - f(x, z) + \nabla_y f(x, y)^\top \beta_{\mathcal{L}_{p_D}}), \quad (5.17)$$

respectively, based on the corresponding definitions in (5.2). $\square$

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 $(y, u, v, w) \in S_{p_D}^L(x)$; 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, the block (5.8)–(5.10) can be viewed as the part of these optimality conditions representing the upper-level problem described in (P_p) or (P_{p_D}) .

Corollary 5.6 *Let x be a local optimal solution of problem (P_p) , and assume that there exist points $(y, u, v, w) \in S_{p_D}^L(x)$ and $z \in S_L^*(x, w)$ such that the upper-level regularity (resp. lower-level regularity) holds at x (resp. (x, z)) and $\nabla_{yy}^2 \mathcal{L}_{p_D}(x, y, z, u, v, w)$ 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 the fulfillment of Assumption 1.1, it holds that $S_{p_D}^L(x) \neq \emptyset$, 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 4.1). 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. $\square$

It follows from Remark 5.5 that the inclusions $(y, u, v, w) \in S_{p_D}^L(x)$ 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 $(y, u, v, w) \in S_{p_D}^L(x)$ necessarily with the presence of (5.11)–(5.13). Therefore, they are somewhat 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 be 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 (5.8)–(5.16) are of the usual KKT-type, as they involve only complementarity conditions. Furthermore, we can easily check that (5.8)–(5.16) can be written as a $(n + 3m + 2p + 2q + r + s + 1) \times (n + 3m + 2p + 2q + r + s + 1)$ square system of equations.

In preparation of the following subsection, we complement the above first order necessary optimality conditions by second order conditions. Considering the nature of the feasible set of problem (tSLR), we need the following assumption.

Assumption 5.7 *The functions $G : \mathbb{R}^n \rightarrow \mathbb{R}^r$ and $H : \mathbb{R}^n \rightarrow \mathbb{R}^s$ are twice continuously differentiable, while F , f , g , and h are thrice continuously differentiable.*

In the subsequent results, we shall use the combined multiplier vectors α, β and the combined constraint functions $\tilde{G}, \tilde{H}$ from (5.2), where α_u and α_w are defined as in (5.17). The Lagrangian function of problem (tSLR) thus is

$$\begin{aligned} \mathcal{L}(\zeta, \alpha, \beta) := & \mathcal{L}_{p_D}(\zeta) + \alpha_G^\top G(x) + \alpha_g^\top g(x, z) - \alpha_u^\top u - \alpha_w^\top w \\ & + \beta_H^\top H(x) + \beta_h^\top h(x, z) + \sum_{j=1}^m \beta_{\mathcal{L}_{p_D}, j} \nabla_{y_j} \mathcal{L}_{p_D}(\zeta), \end{aligned} \quad (5.18)$$

and the critical cone from (3.6) has the form

$$\mathfrak{C}(\zeta, \alpha) := \left\{ d \in \mathbb{R}^{n+2m+p+q+1} \left| \begin{array}{l} \nabla \tilde{G}_i(\zeta)^\top d = 0 \quad \forall i : \tilde{G}_i(\zeta) = 0, \alpha_i > 0 \\ \nabla \tilde{G}_i(\zeta)^\top d \leq 0 \quad \forall i : \tilde{G}_i(\zeta) = 0, \alpha_i = 0 \\ \nabla \tilde{H}_j(\zeta)^\top d = 0 \quad \forall j = 1, \dots, q \end{array} \right. \right\}. \quad (5.19)$$

The following theorem results from an application of the standard second order necessary optimality condition under LICQ to the problem (tSLR).

Theorem 5.8 *Let Assumption 5.7 hold and let $\zeta = (x, y, z, u, v, w)$ be a local optimal solution of problem (tSLR), where all the assumptions of Theorem 5.2 are satisfied. Then, there exist unique multipliers α and β such that the KKT conditions (5.8)–(5.16) are satisfied and such that we have*

$$d^\top \nabla_\zeta^2 \mathcal{L}(\zeta, \alpha, \beta) d \geq 0 \quad \text{for all } d \in \mathfrak{C}(\zeta, \alpha). \quad (5.20)$$

Linking a local optimization solution of problem (P_p) to one of (tSLR) like in the proof of Corollary 5.6, we obtain the following result.

Corollary 5.9 *Let x be a local optimal solution of problem (P_p) , and assume that there exist points $(y, u, v, w) \in S_{p_D}^L(x)$ and $z \in S_L^*(x, w)$ such that the upper-level LICQ (resp. lower-level LICQ) holds at x (resp. (x, z)) and $\nabla_{yy}^2 \mathcal{L}_{p_D}(x, y, z, u, v, w)$ is full rank. Then, there exist unique Lagrange multipliers α and β such that the KKT conditions (5.8)–(5.16) and the second order condition (5.20) are satisfied.*

5.3 Sufficient optimality conditions

In continuous optimization, it is often important to characterize strict local optimal solutions, as done with the second order sufficient condition (SOSC) for problem (3.1). A key question in this subsection is to know whether such a framework is applicable to our true single-level reformulation model (tSLR), and how this could potentially help to build sufficient conditions for strict locality for problem (P_p) . We require Assumption 5.7 to hold throughout this subsection.

We start by establishing a relationship between the strict local optimality of (tSLR) and (P_p) .

Theorem 5.10 *If $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ is a strict local optimal solution of (tSLR), then $\bar{x}$ is a strict local optimal solution of problem (P_p) , provided that the following assumptions are satisfied:*

- (a) $S_{p_D}^L$ is inner semicontinuous at $(\bar{x}, \bar{y}, \bar{u}, \bar{v}, \bar{w})$;
- (b) S_L^* is inner semicontinuous at $(\bar{x}, \bar{w}, \bar{z})$.

Proof Note that based on the hypothesis, the local optimality of $\bar{x}$ for problem (P_p) follows directly by virtue of the second part of Corollary 4.5(b), thanks to assumptions (a) and (b) of this theorem. 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 $x^k \rightarrow \bar{x}$ such that $\varphi_p(x^k) \leq \varphi_p(\bar{x})$ for all k . Subsequently, by the inner semicontinuity of $S_{p_D}^L$ at $(\bar{x}, \bar{y}, \bar{u}, \bar{v}, \bar{w})$, there exists a sequence $(y^k, u^k, v^k, w^k) \in S_{p_D}^L(x^k)$ such that $(y^k, u^k, v^k, w^k) \rightarrow (\bar{y}, \bar{u}, \bar{v}, \bar{w})$. 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 $z^k \rightarrow \bar{z}$. It follows from equations (4.5) and (4.7), together with (4.3), that

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

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 (a) and (b) of the theorem that $(\bar{y}, \bar{u}, \bar{v}, \bar{w}) \in S_{p_D}^L(\bar{x})$ and $\bar{z} \in S_L^*(\bar{x}, \bar{w})$, respectively, are satisfied. Hence, it follows once more from equations (4.5) and (4.7), together with (4.3), that

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

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 $(x^k, y^k, z^k, u^k, v^k, w^k) \rightarrow (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$, 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) . $\square$

It is important to note that assumptions (a) and (b) 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.

We emphasize that the assumptions of Theorem 5.10 cover the ones which guarantee local optimality of $\bar{x}$ for (P_p) in the second part of Corollary 4.5(b). The only additional assumption is the strictness of local optimality in the corresponding problem (tSLR). In the following example, all the assumptions of Theorem 5.10 hold, except the strict local optimality of problem (tSLR). Nevertheless, the local optimality of problem (P_p) is still strict.

Example 5.11 For the problem (P_p) with

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

one obtains $\varphi_p(x) = x^2$, so that $\bar{x} = 0$ is a strict local optimal solution of problem (P_p) .

Assumption 1.1 holds as $S_p(x) = \{x^2\}$ for all $x \in X$. Additionally, Assumption 2.1 also holds, as we can put $g_1(x, y) = -x^2 - y$, $g_2(x, y) = -x^2 + y$, while accounting for the fact that $\varphi_L(x) = x$. Hence, the corresponding problem (tSLR) here is

$$\begin{aligned} \min_{x, y, z, u_1, u_2, w} \quad & \mathcal{L}_{p_D}(x, y, z, u, w) := y + u_1(x^2 + y) + u_2(x^2 - y) \\ \text{s.t.} \quad & -x^2 - z \leq 0, \quad -x^2 + z \leq 0, \quad u, w \geq 0, \\ & 1 + u_1 - u_2 = 0. \end{aligned} \tag{5.21}$$

The point $\bar{\zeta} = (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{w})$, with $\bar{x} = \bar{y} = \bar{z} = 0$, $\bar{u} = (0, 1)^\top$, and $\bar{w} = 1$, is a global, and therefore a local, optimal solution of this problem. In fact, for any feasible point $\zeta = (x, y, z, u, w)$ of (5.21),

$$\mathcal{L}_{p_D}(\zeta) = (1 + u_1 - u_2)y + (u_1 + u_2)x^2 = (u_1 + u_2)x^2 \geq 0 = \mathcal{L}_{p_D}(\bar{\zeta}).$$

However, the family of points $\zeta(t) := (0, t, 0, 0, 1, 1)$ for $t \in \mathbb{R}$, which is feasible for problem (5.21) and satisfies $\zeta(t) \rightarrow \bar{\zeta}$ as $t \rightarrow 0$, is such that for any $t \neq 0$, $\zeta(t) \neq \bar{\zeta}$ and $\mathcal{L}_{p_D}(\zeta(t)) = \mathcal{L}_{p_D}(\bar{\zeta})$. Thus, $\bar{\zeta}$ is not a strict local optimal solution of problem (5.21).

With respect to the inner semicontinuity of $S_{p_D}^L$ at $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$, observe that

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

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

$$S_{p_D}^L(x) = \begin{cases} \{(y, 0, 1, w) \mid y \in \mathbb{R}, w \geq 0\}, & \text{if } x \neq 0, \\ \{(y, u_1, 1 + u_1, w) \mid y \in \mathbb{R}, u_1 \geq 0, w \geq 0\}, & \text{if } x = 0. \end{cases}$$

From $(\bar{y}, \bar{u}, \bar{w}) \in S_{p_D}^L(x)$ 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 $(0, 0)$. $\square$

An interesting feature of this example is the fact that $S_L(\bar{x})$ is a singleton, while $S_L(x)$ is a proper interval for any $x \in X \setminus \{\bar{x}\}$. Moreover, as it can be seen in the graphs of φ_o and φ_p in Fig. 3, it seems quite interesting that $\bar{x}$ is the global optimal solution for the corresponding problem (P_p) , while being the worst point (from the minimization perspective) for the optimistic problem (P_o) .

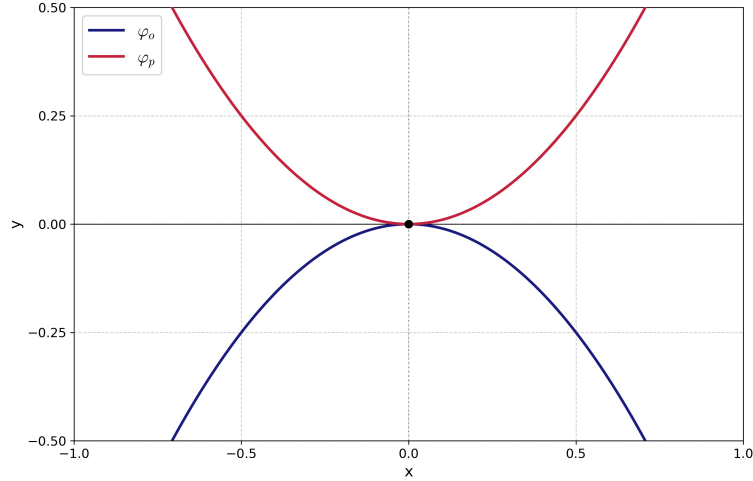


Fig. 3: Graphs of φ_o from (1.3) and φ_p from (1.4) for the problem in Example 5.11.

Example 5.11 also illustrates that the converse of the implication in Theorem 5.10 is not necessarily true. In practice, one would expect to use the SOS of problem (tSLR) as key tool to guarantee that a point is a strict local optimal solution, in order to enable one to leverage on Theorem 5.10 to derive a sufficient condition for strict local optimality for problem (P_p) . Unfortunately, based on the discussion in Subsection 3.1, the point $\bar{\zeta} = (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{w})$ with $\bar{x} = \bar{y} = \bar{z} = 0$, $\bar{u} = (0, 1)^\top$, $\bar{w} = 1$, which can easily be shown to be a KKT point of problem (5.21) with multipliers $\bar{\alpha}_g = (1/4, 1/4)^\top$, $\bar{\alpha}_u = (0, 0)^\top$, $\bar{\alpha}_w = 0$, and $\bar{\beta} := \bar{\beta}_{\mathcal{L}_{p_D}} = 0$ does not satisfy the SOS (as $\bar{\zeta}$ is not a strict local optimal solution of problem (5.21), according to Example 5.11). In fact, the critical cone can be obtained as

$$\mathfrak{C}(\bar{\zeta}, \bar{\alpha}) = \{(d_x, d_y, 0, d_{u_1}, d_{u_2}, d_w)^\top \mid d_x, d_y, d_w \in \mathbb{R}, d_{u_1} \geq 0\},$$

and the Hessian of the Lagrangian function $\mathcal{L}$ of problem (5.21) is

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

This yields $d^\top \nabla_{\zeta\zeta}^2 \mathcal{L}(\bar{\zeta}, \bar{\alpha}, \bar{\beta}) d = d_x^2$ for all $d \in \mathfrak{C}(\bar{\zeta}, \bar{\alpha}) \setminus \{0\}$. Now take, for example, $d_*^\top = (0, 1, 0, 0, 0, 0)$. We have $d_* \in \mathfrak{C}(\bar{\zeta}, \bar{\alpha}) \setminus \{0\}$, and obviously, $d_*^\top \nabla_{\zeta\zeta}^2 \mathcal{L}(\bar{\zeta}, \bar{\alpha}, \bar{\beta}) d_* = 0$. Thus, the SOSC fails at $(\bar{\zeta}, \bar{\alpha}, \bar{\beta})$.

A key observation that could be made for this example is that $\bar{y}, \bar{z} \in S_L(\bar{x})$ and $\Lambda_L(\bar{x}, \bar{y}) \neq \emptyset$, where $\Lambda_L(\bar{x}, \bar{y})$ denotes the set of lower-level Lagrange multipliers, which is defined as

$$\Lambda_L(x, y) := \left\{ (u, v) \in \mathbb{R}^p \times \mathbb{R}^q \left| \begin{array}{l} \nabla_y f(x, y) + \nabla_y g(x, y)^\top u + \nabla_y h(x, y)^\top v = 0 \\ u \geq 0, \quad g(x, y) \leq 0, \quad u^\top g(x, y) = 0 \\ h(x, y) = 0 \end{array} \right. \right\}$$

for any pair (x, y) such that y is feasible for $(LL(x))$. The next result will enable us to show in Remark 5.13 that (SOSC) fails for the problem from Example 5.11 because $\bar{y}, \bar{z} \in S_L(\bar{x})$ and $\Lambda_L(\bar{x}, \bar{y}) \neq \emptyset$.

Proposition 5.12 *Let $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ be a feasible point for problem (tSLR) with $\bar{y}, \bar{z} \in S_L(\bar{x})$ and $\Lambda_L(\bar{x}, \bar{y}) \neq \emptyset$. Then, for every $(u, v) \in \Lambda_L(\bar{x}, \bar{y})$, the path*

$$\zeta(t) := (\bar{x}, \bar{y}, \bar{z}, \bar{u} + tu, \bar{v} + tv, \bar{w} + t) \quad \text{for every } t \geq 0$$

is feasible for problem (tSLR) and satisfies $\mathcal{L}_{p_D}(\zeta(t)) = \mathcal{L}_{p_D}(\bar{\zeta})$ for every $t \geq 0$.

Proof For any $(u, v) \in \Lambda_L(\bar{x}, \bar{y})$, considering the fact that $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ is feasible for (tSLR),

$$\begin{aligned} & \nabla_y F(\bar{x}, \bar{y}) - \nabla_y g(\bar{x}, \bar{y})^\top u(t) - \nabla_y h(\bar{x}, \bar{y})^\top v(t) - w(t) \nabla_y f(\bar{x}, \bar{y}) \\ &= \nabla_y F(\bar{x}, \bar{y}) - \nabla_y g(\bar{x}, \bar{y})^\top \bar{u} - \nabla_y h(\bar{x}, \bar{y})^\top \bar{v} - \bar{w} \nabla_y f(\bar{x}, \bar{y}) \\ & \quad - t [\nabla_y f(\bar{x}, \bar{y}) + \nabla_y g(\bar{x}, \bar{y})^\top u + \nabla_y h(\bar{x}, \bar{y})^\top v] = 0 \end{aligned}$$

for any $t \geq 0$. Hence, $\zeta(t)$ is a feasible point for problem (tSLR) for any $t \geq 0$. Moreover,

$$\mathcal{L}_{p_D}(\zeta(t)) - \mathcal{L}_{p_D}(\bar{\zeta}) = -tu^\top g(\bar{x}, \bar{y}) - tv^\top h(\bar{x}, \bar{y}) - t(f(\bar{x}, \bar{y}) - f(\bar{x}, \bar{z})).$$

Since $(u, v) \in \Lambda_L(\bar{x}, \bar{y})$, it holds that $u^\top g(\bar{x}, \bar{y}) = 0$ and $h(\bar{x}, \bar{y}) = 0$. Additionally, given that we have $\bar{y}, \bar{z} \in S_L(\bar{x})$, it holds that $f(\bar{x}, \bar{y}) = \varphi_L(\bar{x}) = f(\bar{x}, \bar{z})$. Hence, $\mathcal{L}_{p_D}(\zeta(t)) - \mathcal{L}_{p_D}(\bar{\zeta}) = 0$ for all $t \geq 0$. $\square$

Remark 5.13 Another interpretation of this result is that for any feasible point $\zeta := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ of problem (tSLR) with $\bar{y}, \bar{z} \in S_L(\bar{x})$ and $\Lambda_L(\bar{x}, \bar{y}) \neq \emptyset$, there exists a nonconstant feasible path $\zeta(\cdot) : [0, \infty) \rightarrow \mathcal{F}_{\text{tSLR}}$ with origin at 0 (i.e., with $\zeta(0) = \bar{\zeta}$), where the objective function of problem (tSLR) is constant; i.e., $\mathcal{L}_{p_D}(\zeta(t)) = \mathcal{L}_{p_D}(\bar{\zeta})$ for every $t \geq 0$. Note that $\mathcal{F}_{\text{tSLR}}$ denotes the feasible set of problem (tSLR). As a consequence, no feasible point $\zeta := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ of problem (tSLR) can be a strict local optimal solution if $\bar{y}, \bar{z} \in S_L(\bar{x})$ and $\Lambda_L(\bar{x}, \bar{y}) \neq \emptyset$.

For the final discussion of this section, recall that the classical SOSC for (tSLR) is a sufficient condition for the fulfillment of strict local optimality for the problem, as described in (SOSC). Therefore, to potentially leverage on the implication in Theorem 5.10 to construct a second order sufficient optimality condition for problem (P_p) , it is crucial to know whether the second order sufficient condition (SOSC) can hold for (tSLR). In fact, we show that (SOSC) systematically fails for any KKT point of problem (tSLR). The following lemmas will be crucial for this proof.

Lemma 5.14 *Let $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ satisfy the KKT conditions (5.8)–(5.16) with the Lagrange multipliers $\bar{\alpha}$ and $\bar{\beta}$. If the SOSC for (tSLR) holds at $(\bar{\zeta}, \bar{\alpha}, \bar{\beta})$, then $\nabla_{yy}^2 \mathcal{L}_{p_D}(\bar{\zeta})$ is negative definite.*

Proof Based on Assumption 2.1(1)–(4), it follows along the lines of the proof of Proposition 5.3, cf. (5.7), that for any $\zeta := (x, y, z, u, v, w)$ with $u \geq$ and $w \geq 0$, we have $\nabla_{yy}^2 \mathcal{L}_{p_D}(\zeta) \preceq 0$.

Suppose that $\nabla_{yy}^2 \mathcal{L}_{p_D}(\bar{\zeta})$ is singular. Then, there exists $d_y \neq 0$ such that $\nabla_{yy}^2 \mathcal{L}_{p_D}(\bar{\zeta})d_y = 0$. Hence, we can easily check that

$$d := (0, d_y, 0, 0, 0, 0) \in \mathfrak{C}(\bar{\zeta}, \bar{\alpha}),$$

where $\mathfrak{C}(\bar{\zeta}, \bar{\alpha})$ is the critical cone to the feasible set of problem (tSLR) from (5.19).

With the Lagrangian $\mathcal{L}(\zeta, \alpha, \beta)$ of (tSLR) from (5.18), along the direction d above, it holds that

$$d^\top \nabla_{\zeta}^2 \mathcal{L}(\bar{\zeta}, \bar{\alpha}, \bar{\beta})d = d_y^\top \underbrace{\nabla_{yy}^2 \mathcal{L}_{p_D}(\bar{\zeta})d_y}_{=0} + \sum_{j=1}^m \bar{\beta}_{\mathcal{L}_{p_D}, j} d_y^\top \nabla_{yy}^2 (\nabla_{y_j} \mathcal{L}_{p_D})(\bar{\zeta})d_y. \quad (5.22)$$

We introduce the function

$$\psi(y) := d_y^\top \nabla_{yy}^2 \mathcal{L}_{p_D}(\bar{x}, y, \bar{z}, \bar{u}, \bar{v}, \bar{w})d_y.$$

Observe that based on Assumption 2.1(1)–(4), we have

$$\psi(y) \leq 0 = \psi(\bar{y}) \text{ for all } y \text{ near } \bar{y}$$

given that $\nabla_{yy}^2 \mathcal{L}_{p_D}(\bar{\zeta})d_y = 0$. Thus, $\bar{y}$ is a local maximizer of the differentiable function ψ , and consequently, we have $\nabla \psi(\bar{y}) = 0$. Thus,

$$\sum_{j=1}^m \bar{\beta}_{\mathcal{L}_{p_D}, j} d_y^\top \nabla_{yy}^2 (\nabla_{y_j} \mathcal{L}_{p_D})(\bar{\zeta})d_y = \nabla \psi(\bar{y})^\top \bar{\beta}_{\mathcal{L}_{p_D}} = 0.$$

Therefore, (5.22) implies $d^\top \nabla_{\zeta}^2 \mathcal{L}(\bar{\zeta}, \bar{\alpha}, \bar{\beta})d = 0$. This contradicts the fulfillment of the SOSC given that $d \in \mathfrak{C}(\bar{\zeta}, \bar{\alpha}) \setminus \{0\}$. Hence, $\nabla_{yy}^2 \mathcal{L}_{p_D}(\bar{\zeta})$ must be nonsingular. Since this matrix is negative semidefinite, it follows that $\nabla_{yy}^2 \mathcal{L}_{p_D}(\bar{\zeta}) \prec 0$. $\square$

Lemma 5.15 *If $S_L(\bar{x}) = \{\bar{y}\}$ and (GCQ) holds in $Z(\bar{x})$ at $\bar{y}$, then $\Lambda_L(\bar{x}, \bar{y}) \neq \emptyset$.*

Proof Since $Z(\bar{x}) = S_L(\bar{x}) = \{\bar{y}\}$, it holds that $T_{Z(\bar{x})}(\bar{y}) = \{0\}$. Hence, thanks to the GCQ,

$$(L_{Z(\bar{x})}(\bar{y}))^* = (T_{Z(\bar{x})}(\bar{y}))^* = \mathbb{R}^m.$$

Since $L_{Z(\bar{x})}(\bar{y})$ is a closed convex cone, $L_{Z(\bar{x})}(\bar{y}) = \left((L_{Z(\bar{x})}(\bar{y}))^* \right)^* = \{0\}$. Moreover,

$$L_{Z(\bar{x})}(\bar{y}) = \{d_y \in L_{Y(\bar{x})}(\bar{y}) \mid \nabla_y f(\bar{x}, \bar{y})^\top d_y \leq 0\} \quad (5.23)$$

given that the constraint $f(\bar{x}, y) - \varphi_L(\bar{x}) \leq 0$ is active at $\bar{y}$. Combining (5.23) with the fact that $L_{Z(\bar{x})}(\bar{y}) = \{0\}$, it holds that $\nabla_y f(\bar{x}, \bar{y})^\top d_y > 0$ if $d_y \in L_{Y(\bar{x})}(\bar{y}) \setminus \{0\}$. Hence,

$$\nabla_y f(\bar{x}, \bar{y})^\top d_y \geq 0 \text{ for all } d_y \in L_{Y(\bar{x})}(\bar{y}).$$

Thus, based on the dual cone definition (3.2), we clearly have that

$$\nabla_y f(\bar{x}, \bar{y}) \in (L_{Y(\bar{x})}(\bar{y}))^*. \quad (5.24)$$

Furthermore, note that the linearized cone to $Y(\bar{x})$ at $\bar{y}$ is

$$L_{Y(\bar{x})}(\bar{y}) = \left\{ d_y \in \mathbb{R}^m \mid \begin{array}{l} \nabla_y g_i(\bar{x}, \bar{y})^\top d_y \leq 0, \quad i \in I_g(\bar{x}, \bar{y}) \\ \nabla_y h_j(\bar{x}, \bar{y})^\top d_y = 0, \quad j = 1, \dots, q \end{array} \right\},$$

where $I_g(\bar{x}, \bar{y}) := \{i \in \{1, \dots, p\} \mid g_i(\bar{x}, \bar{y}) = 0\}$, and its dual cone can be written as

$$\begin{aligned} (L_{Y(\bar{x})}(\bar{y}))^* &= \left\{ a \in \mathbb{R}^m \mid \exists \lambda_i \geq 0 \text{ for } i \in I_g(\bar{x}, \bar{y}) \text{ and } \exists \mu \in \mathbb{R}^q : \right. \\ &\quad \left. a = - \sum_{i \in I_g(\bar{x}, \bar{y})} \lambda_i \nabla_y g_i(\bar{x}, \bar{y}) - \sum_{j=1}^q \mu_j \nabla_y h_j(\bar{x}, \bar{y}) \right\}. \end{aligned}$$

Combining this with (5.24), we clearly have that $\Lambda_L(\bar{x}, \bar{y}) \neq \emptyset$. $\square$

We are now ready to show that the SOSC for problem (tSLR) fails at any of its KKT points.

Proposition 5.16 *Let $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ satisfy the KKT conditions (5.8)–(5.16) with the Lagrange multipliers $\bar{\alpha}$ and $\bar{\beta}$. Then, the SOSC for problem (tSLR) fails at $(\bar{\zeta}, \bar{\alpha}, \bar{\beta})$.*

Proof By contradiction, we assume that the SOSC for (tSLR) holds at $(\bar{\zeta}, \bar{\alpha}, \bar{\beta})$. Then, it follows from Lemma 5.14 that $\nabla_{yy}^2 \mathcal{L}_{p_D}(\bar{\zeta}) \prec 0$. Hence, the first block of (5.11) implies that $\bar{\beta}_{\mathcal{L}_{p_D}} = 0$.

Case 1: Suppose that $\bar{w} > 0$. Then, combining this with $\bar{\beta}_{\mathcal{L}_{p_D}} = 0$, it follows from (5.11)–(5.13) that we have

$$g(\bar{x}, \bar{y}) \leq 0, \quad \bar{u} \geq 0, \quad \bar{u}^\top g(\bar{x}, \bar{y}) = 0, \quad h(\bar{x}, \bar{y}) = 0, \quad f(\bar{x}, \bar{y}) - f(\bar{x}, \bar{z}) = 0. \quad (5.25)$$

Additionally, considering $\bar{w} > 0$ again, we have from (5.14)–(5.16) that $(\alpha'_g, \beta'_h) := (\bar{\alpha}_g, \bar{\beta}_h)/\bar{w} \in \Lambda_L(\bar{x}, \bar{z})$. Hence, by the convexity of the lower-level problem in Assumption 2.1(2)–(4), it follows that $\bar{z} \in S_L(\bar{x})$. Thus, combining this with the last equation in (5.25),

$$f(\bar{x}, \bar{y}) = f(\bar{x}, \bar{z}) \leq \varphi_L(\bar{x}) \leq f(\bar{x}, \bar{y}). \quad (5.26)$$

Furthermore, with $\bar{y} \in Y(\bar{x})$ from (5.25), we have from (5.26) that $\bar{y} \in S_L(\bar{x})$. Moreover, it is not hard to see that the vectors $d_z \in \mathbb{R}^m$ with $(0, 0, d_z, 0, 0, 0) \in \mathfrak{C}(\bar{\zeta}, \bar{\alpha})$ form the critical cone of $(LL(x))$ for $x := \bar{x}$ at $\bar{z}$ with respect to the multiplier vector α'_g . Therefore, the SOSC with the Lagrangian $\mathcal{L}(\zeta, \alpha, \beta)$ of (tSLR) from (5.18) implies $d_z^\top \nabla_{zz}^2 \mathcal{L}(\bar{\zeta}, \bar{\alpha}, \bar{\beta}) d_z > 0$ for all $d_z \in \mathfrak{C}(\bar{z}, \alpha'_g) \setminus \{0\}$. In view of

$$\begin{aligned} \nabla_{zz}^2 \mathcal{L}(\bar{\zeta}, \bar{\alpha}, \bar{\beta}) &= \bar{w} \nabla_{zz}^2 f(\bar{x}, \bar{z}) + \nabla_{zz}^2 g(\bar{x}, \bar{z}) \bar{\alpha}_g + \nabla_{zz}^2 h(\bar{x}, \bar{z}) \bar{\beta}_h \\ &= \bar{w} (\nabla_{zz}^2 f(\bar{x}, \bar{z}) + \nabla_{zz}^2 g(\bar{x}, \bar{z}) \alpha'_g + \nabla_{zz}^2 h(\bar{x}, \bar{z}) \beta'_h), \end{aligned}$$

one obtains the SOSC for $\bar{z}$ to be a unique local minimal point of problem $(LL(x))$ for $x := \bar{x}$. In view of $\bar{z} \in S_L(\bar{x})$, the latter set is thus a singleton. Therefore, it holds that $\bar{y} = \bar{z}$. Subsequently, considering the system (5.14)–(5.16) once again, it holds that $\Lambda_L(\bar{x}, \bar{y}) \neq \emptyset$. Hence, applying Proposition 5.12, it follows that $\bar{\zeta}$ is not a strict local optimality solution of (tSLR). This contradicts the fulfillment of the SOSC for problem (tSLR) at $(\bar{\zeta}, \bar{\alpha}, \bar{\beta})$.

Case 2: Suppose that $\bar{w} = 0$. Then, proceeding in a way similar to the previous case, it follows that with $\bar{\beta}_{\mathcal{L}_{p_D}} = 0$, we get from equations (5.11)–(5.13) that

$$g(\bar{x}, \bar{y}) \leq 0, \quad \bar{u} \geq 0, \quad \bar{u}^\top g(\bar{x}, \bar{y}) = 0, \quad h(\bar{x}, \bar{y}) = 0, \quad f(\bar{x}, \bar{y}) - f(\bar{x}, \bar{z}) \leq 0. \quad (5.27)$$

Moreover, with $\bar{w} = 0$, it follows from the system (5.14)–(5.16) that $\bar{z}$ is a KKT point of the lower-level feasibility problem

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

with multipliers $\bar{\alpha}_g$ and $\bar{\beta}_h$. Since $Y(\bar{x})$ is convex, $\bar{z}$ is also a global minimal point. Moreover, given that the vectors $d_z \in \mathbb{R}^m$ with $(0, 0, d_z, 0, 0, 0) \in \mathfrak{C}(\bar{\zeta}, \bar{\alpha})$ form the critical cone of $(LL^0(x))$ at $\bar{z}$, we obtain from the fulfillment of the SOSC at $\bar{z}$ that this point is a unique local minimal point of $(LL^0(x))$. This is only possible if $Y(\bar{x})$ is a singleton, implying that also its subset $S_L(\bar{x})$ is a singleton. Hence, $S_p(\bar{x}) = S_L(\bar{x}) = Y(\bar{x}) = \{\bar{z}\}$. Subsequently, $f(\bar{x}, \bar{z}) \leq \varphi_L(\bar{x})$. Combining this with (5.27),

$$g(\bar{x}, \bar{y}) \leq 0, \quad h(\bar{x}, \bar{y}) = 0, \quad \varphi_L(\bar{x}) \leq f(\bar{x}, \bar{y}) \leq f(\bar{x}, \bar{z}) \leq \varphi_L(\bar{x}).$$

This implies that $\bar{y} \in Y(\bar{x})$ and $f(\bar{x}, \bar{y}) = \varphi_L(\bar{x})$. Thus, $\bar{y} \in S_L(\bar{x})$. This implies that $\bar{y} = \bar{z}$. Hence, $S_L(\bar{x}) = \{\bar{y}\}$, and we have from a combination of Assumption 2.1(5) (while accounting for the fact that $S_p(\bar{x}) \subset S_L(\bar{x}) = \{\bar{y}\}$) and Lemma 5.15 that $\Lambda_L(\bar{x}, \bar{y}) \neq \emptyset$. Similarly to the previous case (Case 1), we get a contradiction by applying Proposition 5.12. $\square$

Considering Remark 4.2 and Remark 4.6, it is not necessarily expected that for an optimal solution $\bar{\zeta} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ of problem (tSLR), one would have $\bar{y} \in S_L(\bar{x})$, $\bar{z} \in S_L(\bar{x})$, or $(\bar{u}, \bar{v}) \in \Lambda_L(\bar{x}, \bar{y})$. The proof of Proposition 5.16 shows, however, that all three conditions would be enforced by the SOSC in the optimal solution, leading to a contradiction in view of Proposition 5.12. This illustrates that having these three conditions satisfied simultaneously is actually not a positive situation for problem (tSLR). In particular, the systematic failure of the SOSC must be expected to interfere with the convergence of certain types of numerical methods for (tSLR) as a consequence of the next result, which is a refinement of Theorem 5.8 (implied by Proposition 5.16).

Corollary 5.17 *Under the assumptions of Theorem 5.8, its assertions as well as the following two statements are satisfied:*

- (a) *There exists some $d \in \mathfrak{C}(\zeta, \alpha) \setminus \{0\}$ with $d^\top \nabla_{\zeta\zeta}^2 \mathcal{L}(\zeta, \alpha, \beta) d = 0$.*
- (b) *Let $\alpha_i > 0$ for all i with $\tilde{G}_i(\zeta) = 0$ (i.e., strict complementarity holds) and let the rows of the matrix B collect the derivatives $\nabla G_i(\zeta)^\top$ with $\tilde{G}_i(\zeta) = 0$ and $\nabla \tilde{H}_j(\zeta)^\top$, $j = 1 \dots, q$. Then the Jacobian $\begin{pmatrix} \nabla_{\zeta\zeta}^2 \mathcal{L}(\zeta, \alpha, \beta) & B^\top \\ B & 0 \end{pmatrix}$ of the equations in the KKT system (5.8)–(5.16) is singular.*

Proof The assertion of part (a) immediately follows from Theorem 5.8 and Proposition 5.16. To see part (b), note that the strict complementarity assumption implies that the cone $\mathfrak{C}(\zeta, \alpha)$ is a linear space (the tangent space to the feasible set). Therefore, the assertion of part (a) can be restated as the singularity of the restriction of $\nabla_{\zeta\zeta}^2 \mathcal{L}(\zeta, \alpha, \beta)$ to this tangent space. Since under the assumptions of Theorem 5.8, the LICQ holds at ζ , the matrix B possesses full row rank. The assertion now follows from [24, Lemma 3.4] (see also [30]). $\square$

A take away of this singularity result is that to leverage on Theorem 5.10 for practical sufficient conditions for problem (P_p) , different types of characterizations for sufficient optimality for problem (tSLR) would need to be explored, including possibly first order-type conditions. In addition, one may not expect second order methods to work for the algorithmic solution of (tSLR), but without further modifications of the reformulation one should rather resort to first order methods.

6 Numerical illustrations

Considering the relationships established in Section 4 between (P_p) and (tSLR), we use six 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 constrained problem (3.1), it is approximated by the bound-constrained penalized problem

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

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 (6.1), $\mathcal{V}$ denotes the quadratic constraint-violation function

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

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 [34] to (3.1). Considering the combination of the DIRECT method with the Penalization and SLSQP 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., [6, Example 2.1]), P2 (see, e.g., [7, 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., [50, Example 1]), and P6 (see, e.g., [50, 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 [44] and TLVF-GSS1 [43] 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 [44]).

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 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 $N_{\text{max}} = \min\{3000\hat{n}, 80000\}$ 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

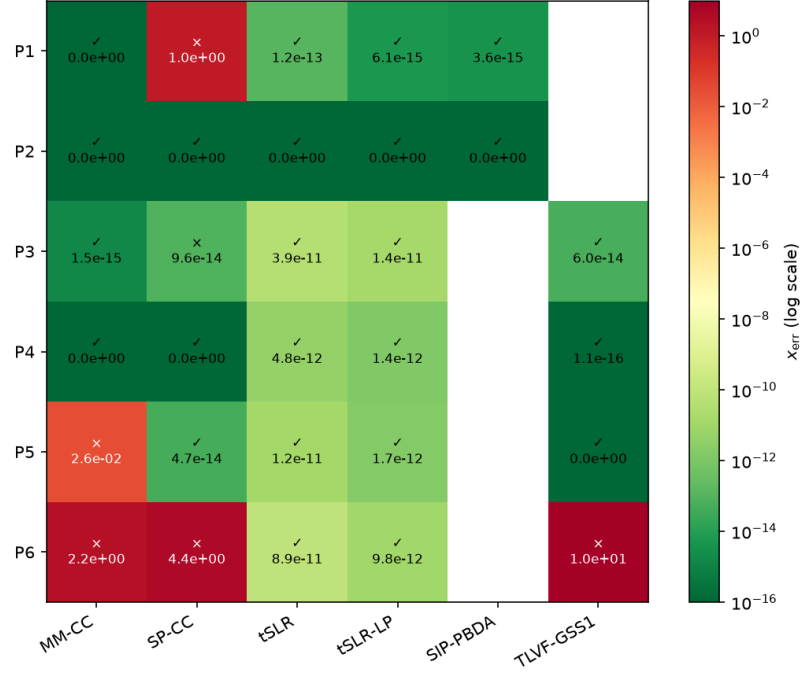


Fig. 4: Leader's solution error x_{err} across all problems and all six methods (log color scale; green = accurate, red = inaccurate). Empty cells represent cases outside a method's applicability scope.

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.

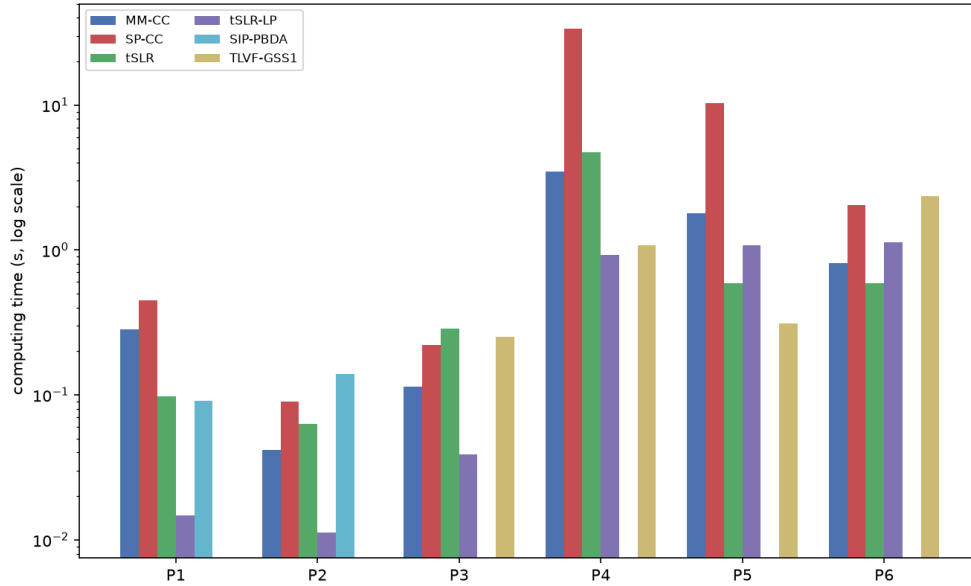


Fig. 5: A missing bar means the method is outside its applicability scope for that problem. SP-CC's cost is the clear outlier, most visibly on P4 (33.5s, an order of magnitude above every other method on that problem), driven by its larger multiplier-variable count inflating each SLSQP polish; tSLR-LP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

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 [43]. MM-CC 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, whereas even the weaker Mangasarian–Fromovitz constraint qualification systematically fails for all known SLRs of optimistic and pessimistic bilevel programs. Moreover, (tSLR) enables the derivation of new necessary optimality conditions for pessimistic bilevel programs. On the numerical side, six illustrative examples provide a proof of concept for computing globally optimal solutions with the tSLR. Future work will focus on developing numerical methods tailored to (tSLR) and systematically comparing them, on suitable test problems, with the existing algorithms for pessimistic bilevel optimization reviewed in Section 2. A major limitation, however, is the systematic failure of the classical second-order sufficient condition at Karush–Kuhn–Tucker points of (tSLR) within the framework of this paper, as shown in Subsection 5.3. Consequently, second-order methods may struggle to solve (tSLR) efficiently, as observed in Corollary 5.17.

As observed throughout Sections 4 and 5, some features of problem (tSLR) require careful interpretation. In particular, variables u , v , and w arise from the Wolfe-dual representation of the intermediate problem, but they are not required to satisfy the complementary-slackness conditions that would ordinarily be associated with the corresponding constraints. Likewise, at an optimal solution (x, y, z, u, v, w) of problem (tSLR), the auxiliary point y need not solve the intermediate problem and z need not solve the original lower-level problem. Their role is to construct an exact single-level representation from which the optimal leader decision x can be recovered. Once an optimal x has been obtained for problem (P_p), corresponding optimal solutions for the intermediate- and lower-level problems ($IL(x)$) and ($LL(x)$), respectively, can be generated (see Remark 4.6).

It is also important to note that the proposed true single-level reformulation does not remove the fundamental existence difficulties of pessimistic bilevel optimization. The assumptions used to establish the tSLR do not, by themselves, guarantee that either the original problem or its reformulation possesses an optimal solution. In particular, the feasible set of the tSLR is generally unbounded, and the discontinuity and nonattainment phenomena caused, for example, by a failure of lower semicontinuity of the lower-level optimal solution set-valued mapping are inherited by the reformulation. This is unavoidable for an exact reformulation: if the original pessimistic problem has a finite but unattained infimum, the tSLR cannot be expected to manufacture an optimal solution that does not exist.

Despite the failure of (SOSC) for problem (tSLR), the broader message of this work is a very *optimistic* one. Pessimistic bilevel programs possess genuine structural difficulties and may fail to admit optimal solutions without additional regularity assumptions. However, whenever an optimal solution does exist, computing it need not be intrinsically more difficult than solving the corresponding optimistic problem. Indeed, the tSLR suggests that, for some problem classes, finding a pessimistic optimal solution may even be easier, because it can be approached through a reformulation without implicit variables, complementarity constraints, or optimal value functions, and which can satisfy the standard constraint qualifications that are unavailable to conventional bilevel optimization reformulations.

These observations open several directions for future research. A first priority could be the development of global and local algorithms specifically tailored to the structure of problem (tSLR), particularly, methods that exploit the close relationship between its objective function and its constraint

$$\nabla_y F(x, y) - \nabla_y g(x, y)^\top u - \nabla_y h(x, y)^\top v - w \nabla_y f(x, y) = 0$$

could be part of such future explorations. Specialized algorithms for the reduced formulation (tSLR-LP), as well as extensions capable of accommodating discrete or mixed-integer upper-level variables, also deserve investigation. Finally, a broader computational study on systematic benchmark collections is needed to compare these methods with the existing algorithms reviewed in Section 2 and to identify the classes of pessimistic bilevel programs for which the tSLR provides the greatest advantage.

Statements and declarations

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The data and code used to produce the numerical results in Section 6 are available from the corresponding author upon reasonable request.

Code availability

The code used to produce the numerical results in Section 6 is available from the corresponding author upon reasonable request.

Author contributions

Both authors contributed to the conception, analysis, and writing of this work, and read and approved the final manuscript.

References

1. Aboussoror, A., Adly, S., Saissi, F.E.: Strong-weak nonlinear bilevel problems: existence of solutions in a sequential setting. *Set-Valued and Variational Analysis* **25**(1), 113–132 (2017)
2. Antoniou, M., Sinha, A., Papa, G.: δ -perturbation of bilevel optimization problems: An error bound analysis. *Operations Research Perspectives* **13**, 100315 (2024). <https://doi.org/10.1016/j.orp.2024.100315>
3. Aussel, D., Svensson, A.: Is pessimistic bilevel programming a special case of a mathematical program with complementarity constraints? *Journal of Optimization Theory and Applications* **181**(2), 504–520 (2019)
4. Bazaraa, M., Shetty, H., C.M., S.: *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, Inc. (2006)
5. Beck, Y., Ljubić, I., Schmidt, M.: *Linear and Mixed-Integer Bilevel Optimization: Theory and Algorithms*. Cambridge University Press (2027). Forthcoming
6. Benchouk, I., Jolaoso, L., Nachi, K., Zemkoho, A.: Scholtes relaxation method for pessimistic bilevel optimization. *Set-Valued and Variational Analysis* **33**(2), 10 (2025). <https://doi.org/10.1007/s11228-025-00747-5>
7. Benchouk, I., Jolaoso, L., Nachi, K., Zemkoho, A.: Relaxation methods for pessimistic bilevel optimization. *Set-Valued and Variational Analysis* **34**(1), 1 (2026). <https://doi.org/10.1007/s11228-025-00787-x>
8. Calvete, H.I., Galé, C., Hernández, A., Iranzo, J.A.: A novel approach to pessimistic bilevel problems. an application to the rank pricing problem with ties. *Optimization* **74**(12), 2823–2856 (2025). <https://doi.org/10.1080/02331934.2024.2388204>
9. Cao, Q., Zeng, S., Zhang, J.: A single-loop gradient algorithm for pessimistic bilevel optimization via smooth approximation. *Advances in Neural Information Processing Systems* **38**, 23039–23080 (2025)
10. Cervinka, M., Matonoha, C., Outrata, J.V.: On the computation of relaxed pessimistic solutions to mpecs. *Optimization Methods and Software* **28**(1), 186–206 (2013)

11. Cipolla, S., Stein, O., Zemkoho, A.: A single loop method for quadratic minmax optimization (2026). URL <https://arxiv.org/abs/2608.17830>
12. Dempe, S.: Foundations of bilevel programming. Kluwer Academic Publishers (2002)
13. Dempe, S., Dutta, J., Mordukhovich, B.: New necessary optimality conditions in optimistic bilevel programming. *Optimization* **56**(5-6), 577–604 (2007)
14. Dempe, S., Luo, G., Franke, S.: Pessimistic bilevel linear optimization. *Journal of Nepal Mathematical Society* **1**(1), 1–10 (2018)
15. Dempe, S., Mehlitz, P.: Duality-based single-level reformulations of bilevel optimization problems. *Journal of Optimization Theory and Applications* **205**(2), 26 (2025). <https://doi.org/10.1007/s10957-025-02627-2>
16. Dempe, S., Mordukhovich, B.S., Zemkoho, A.B.: Sensitivity analysis for two-level value functions with applications to bilevel programming. *SIAM Journal on Optimization* **22**(4), 1309–1343 (2012)
17. Dempe, S., Mordukhovich, B.S., Zemkoho, A.B.: Necessary optimality conditions in pessimistic bilevel programming. *Optimization* **63**(4), 505–533 (2014)
18. Dempe, S., Mordukhovich, B.S., Zemkoho, A.B.: Two-level value function approach to non-smooth optimistic and pessimistic bilevel programs. *Optimization* **68**(2-3), 433–455 (2019)
19. Dempe, S., Schmidt, H.: On an algorithm solving two-level programming problems with nonunique lower level solutions. *Computational Optimization and Applications* **6**(3), 227–249 (1996)
20. Dempe, S., Zemkoho, A.: Bilevel optimization: Advances and next challenges. In: *Springer Optimization and its Applications*, vol. 161, pp. 1–672. Springer (2020)
21. Dempe, S., Zemkoho, A.B.: On the Karush–Kuhn–Tucker reformulation of the bilevel optimization problem. *Non-linear Analysis: Theory, Methods & Applications* **75**(3), 1202–1218 (2012)
22. Fletcher, R.: *Practical Methods of Optimization*. John Wiley & Sons (2000)
23. Gauvin, J.: A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming* **12**(1), 136–138 (1977)
24. Gould, N.I.M.: On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem. *Mathematical Programming* **32**(1), 90–99 (1985). <https://doi.org/10.1007/BF01585660>
25. Grazi, R., Pontil, M., Salzo, S., Zemkoho, A.: Bilevel learning. arXiv preprint arXiv:2605.01621 (2026)
26. Guan, Z., Sow, D., Lin, S., Liang, Y.: Adaprox: A novel method for bilevel optimization under pessimistic framework. In: *The Second Conference on Parsimony and Learning (Proceedings Track)*, vol. 280, pp. 134–164. PMLR (2025)
27. Guo, L., Lin, G.H., Ye, J.J.: Second-order optimality conditions for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications* **158**(1), 33–64 (2013)
28. Jane, J.Y.: Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications* **307**(1), 350–369 (2005)
29. Jones, D.R., Perttunen, C.D., Stuckman, B.E.: Lipschitzian optimization without the lipschitz constant. *Journal of Global Optimization* **3**(2), 157–181 (1993). <https://doi.org/10.1007/BF00941892>
30. Jongen, H.T., Möbert, T., Rückmann, J., Tammer, K.: On inertia and Schur complement in optimization. *Linear Algebra and its Applications* **95**, 97–109 (1987). [https://doi.org/10.1016/0024-3795\(87\)90028-0](https://doi.org/10.1016/0024-3795(87)90028-0)
31. Jungen, D., Zingler, A., Djelassi, H., Mitsos, A.: libDIPS–discretization-based semi-infinite and bilevel programming solvers. *Mathematical Programming Computation* pp. 1–56 (2026). <https://doi.org/10.1007/s12532-026-00308-3>
32. Kis, T., Kovács, A., Mészáros, C.: On optimistic and pessimistic bilevel optimization models for demand response management. *Energies* **14**(8), 2095 (2021)
33. Kleniati, P.M., Adjiman, C.S.: Branch-and-sandwich: a deterministic global optimization algorithm for optimistic bilevel programming problems. part i: Theoretical development. *Journal of Global Optimization* **60**(3), 425–458 (2014)
34. Kraft, D.: A software package for sequential quadratic programming. *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt* (1988)
35. Lampariello, L., Sagratella, S.: A bridge between bilevel programs and nash games. *Journal of Optimization Theory and Applications* **174**(2), 613–635 (2017)
36. Lampariello, L., Sagratella, S., Stein, O.: The standard pessimistic bilevel problem. *SIAM Journal on Optimization* **29**(2), 1634–1656 (2019)
37. Liu, J., Hong, Y., Zheng, Y.: A new variant of penalty method for weak linear bilevel programming problems. *Wuhan University Journal of Natural Sciences* **23**(4), 328–332 (2018)
38. Loridan, P., Morgan, J.: Weak via strong stackelberg problem: new results. *Journal of global Optimization* **8**(3), 263–287 (1996)
39. Mitsos, A., Lemonidis, P., Barton, P.I.: Global solution of bilevel programs with a nonconvex inner program. *Journal of Global Optimization* **42**(4), 475–513 (2008)
40. Molodtsov, D.: The solution of a class of non-antagonistic games. *USSR Computational Mathematics and Mathematical Physics* **16**(6), 67–72 (1976). [https://doi.org/10.1016/0041-5553\(76\)90042-2](https://doi.org/10.1016/0041-5553(76)90042-2). URL <https://www.sciencedirect.com/science/article/pii/0041555376900422>
41. Molodtsov, D., Fedorov, V.: Approximation of two-person games with information exchange. *USSR Computational Mathematics and Mathematical Physics* **13**(6), 123–142 (1973). [https://doi.org/10.1016/0041-5553\(73\)90010-4](https://doi.org/10.1016/0041-5553(73)90010-4). URL <https://www.sciencedirect.com/science/article/pii/0041555373900104>
42. Shehu, Y., Vuong, P.T., Zemkoho, A.: An inertial extrapolation method for convex simple bilevel optimization. *Optimization Methods and Software* **36**(1), 1–19 (2021)
43. Strekalovsky, A.: One way to solve pessimistic bilevel optimization problems. *Optimization* **74**(17), 4795–4825 (2025). <https://doi.org/10.1080/02331934.2025.2487713>
44. Wiesemann, W., Tsoukalas, A., Kleniati, P.M., Rustem, B.: Pessimistic bilevel optimization. *SIAM Journal on Optimization* **23**(1), 353–380 (2013)
45. Wolfe, P.: A duality theorem for non-linear programming. *Quarterly of Applied Mathematics* **19**(3), 239–244 (1961)
46. Wright, S., Nocedal, J., et al.: *Numerical optimization*. Springer Science **35**(67-68), 7 (1999)
47. Ye, J.J., Zhu, D.: New necessary optimality conditions for bilevel programs by combining the MPEC and value function approaches. *SIAM Journal on Optimization* **20**(4), 1885–1905 (2010)
48. Zemkoho, A.B.: A simple approach to optimality conditions in minmax programming. *Optimization* **63**(3), 385–401 (2014)
49. Zemkoho, A.B.: Solving ill-posed bilevel programs. *Set-Valued and Variational Analysis* **24**(3), 423–448 (2016)
50. Zeng, B.: A practical scheme to compute the pessimistic bilevel optimization problem. *INFORMS Journal on Computing* **32**(4), 1128–1142 (2020)