



HAL
open science

Geometric and computational hardness of bilevel programming

Jérôme Bolte, Quoc-Tung Le, Edouard Pauwels, Samuel Vaiter

► **To cite this version:**

Jérôme Bolte, Quoc-Tung Le, Edouard Pauwels, Samuel Vaiter. Geometric and computational hardness of bilevel programming. 2024. hal-04649435

HAL Id: hal-04649435

<https://inria.hal.science/hal-04649435v1>

Preprint submitted on 16 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Geometric and computational hardness of bilevel programming

Jérôme Bolte* Tùng Lê* Edouard Pauwels* Samuel Vaiter †

July 16, 2024

Abstract

We first show a simple but striking result in bilevel optimization: unconstrained C^∞ smooth bilevel programming is as hard as general extended-real-valued lower semicontinuous minimization. We then proceed to a worst-case analysis of box-constrained bilevel polynomial optimization. We show in particular that any extended-real-valued semi-algebraic function, possibly non-continuous, can be expressed as the value function of a polynomial bilevel program. Secondly, from a computational complexity perspective, the decision version of polynomial bilevel programming is one level above NP in the polynomial hierarchy (Σ_2^P -hard). Both types of difficulties are uncommon in non-linear programs for which objective functions are typically continuous and belong to the class NP. These results highlight the irremediable hardness attached to general bilevel optimization and the necessity of imposing some form of regularity on the lower level.

1 Introduction

This paper is concerned with the bilevel programming problem (BP), that is formulated as follows:

$$\begin{aligned} & \underset{x \in \mathcal{X}}{\text{Minimize}} && P(x, y) \\ & \text{s.t.} && y \in \Theta(x) := \arg \min_{y \in \mathcal{Y}} Q(x, y) \end{aligned} \tag{BP}$$

where $P, Q : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ are called upper-level and lower-level functions respectively. Similarly, x and y are called upper-level and lower-level variables. We limit ourselves to a lower-level feasible set \mathcal{Y} independent of the upper variable x ¹.

Since $Q(x, \cdot)$ might admit more than one minimizer, the problem (BP) may be ill-posed. There are two major approaches to choose a minimizer in $\Theta(x)$: optimistic and pessimistic². They yield two different *value functions*:

*Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France.

†CNRS & Université Côte d'Azur, Laboratoire J. A. Dieudonné. Nice, France.

¹Due to this independence, the formulation (BP) is sometimes called *simple bilevel optimization* [43, 35]. The general bilevel formulation might replace \mathcal{Y} by $\mathcal{Y}(x)$ where the feasible set depends on the upper variable x .

²Apart from these two models, there exist others such as bilevel optimization under uncertainty [11, 10, 2]. A complete presentation of these models is, however, out of the scope of this paper.

$$\varphi_o: x \mapsto \inf_{y \in \Theta(x)} P(x, y) \quad (\text{BP-O}) \quad \varphi_p: x \mapsto \sup_{y \in \Theta(x)} P(x, y) \quad (\text{BP-P}),$$

where $\Theta(x)$ is defined as in (BP). We call φ_o (resp. φ_p) the optimistic (resp. pessimistic) value function. We emphasize that the term “value function” is different from the “lower-level value function”, which has been used in the literature to construct algorithms, starting with [52]. When $\Theta(x) = \emptyset$, we set $\varphi_o(x) = +\infty$ and $\varphi_p(x) = -\infty$, respectively.

Motivations and objectives

Historically, bilevel optimization has been used to address problems in economics, chemistry, optimal control, and decision-making. For some insights into these traditional applications, we refer readers to [23, Chapter 1], [47, 18]. Recently, there has been a regain of interest in bilevel optimization among the machine learning community due to its applications in hyper-parameters tuning and meta-learning [45, 26].

A large body of work on bilevel programming, especially in the context of machine learning [20, 21, 36, 27, 1, 15, 3, 31], requires strong convexity of the lower-level problem to design scalable and provably convergent algorithms. This is a very favorable situation as the bilevel constraint is essentially equivalent to a *qualified* equality constraint, i.e., a manifold-like constraint. However, for many real-world machine learning problems (e.g., meta-learning and hyper-parameters optimization [26]), the lower-level problem is not strongly convex and even non-smooth and nonconvex. To address the challenges in those situations, existing works [38, 37, 4, 35, 42, 52, 24, 39] have proposed various regularity, variational and ad-hoc assumptions on P, Q, \mathcal{Y} . These assumptions are, however, difficult to verify *a priori* and may fail to hold for typical bilevel problems [28, Section 3.2].

In this work, we adopt a worst-case approach to explore the hardness of bilevel optimization and identify the class of functions that can be represented using a general bilevel problem ((BP-O) and (BP-P)), especially in the setting where the lower-level programs are smooth but possibly nonconvex. Our analysis highlights the difficulties of general bilevel optimization and the necessity of qualification conditions.

Pathological value functions are omnipresent

Bilevel programming falls in the general framework of parametric optimization with constraints parameterized by the upper variable x . In this context, favorable situations, include cases where the constraint $y \in \Theta(x)$ can be equivalently described with a well-behaved equation or defines a smooth or regular mapping. This is typical in the smooth setting if the lower-level is strongly convex, which results in a smooth mapping Θ . A typical difficulty of general bilevel programs is that the resulting value function does not inherit the smoothness properties of its defining data P and Q as the argmin mapping Θ may be poorly structured —precisely because it corresponds to a critical set, here an argmin. The following proposition illustrates this behavior. It is based on the well-known characterization of closed sets as the zero set of smooth (infinitely differentiable) functions due to Whitney (Theorem A.1), the proof is postponed to Section A in the appendix.

Theorem 1.1 (Lower semicontinuous minimization problems are unconstrained C^∞ bilevel problems). *Any proper lower (resp. upper) semicontinuous function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ (resp. $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{-\infty\}$) whose domain $\text{dom}(f) := \{x \in \mathbb{R}^n \mid f(x) \in \mathbb{R}\}$ is closed, is the value function of an optimistic (resp. pessimistic) bilevel problem with C^∞ smooth upper and lower levels P, Q and without lower-level constraint, i.e., $\mathcal{Y} = \mathbb{R}^3$ (3 lower-level variables).*

In particular, any full domain lower semicontinuous function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is the value function of an optimistic bilevel problem with C^∞ smooth upper and lower levels. The proof of Theorem 1.1 is given in Section A. This result highlights the impossibility of dealing with the general C^∞ bilevel problem and raises concerns about the very meaning of a solution algorithm for such problems. Indeed, to have a sense of Proposition 1.1, one can observe that the following monstrous univariate functions admit a representation as a bilevel program with smooth data:

- The *negative* of Thomae’s popcorn function [8, Example 5.1.6h)] is lower semicontinuous and discontinuous on \mathbb{Q} , with value 0 on irrationals and global minimum at $1/2$.
- The characteristic function of the Smith-Volterra-Cantor set [8, Definition 11.1.10] (1 on the set, 0 outside). This is a closed set, its characteristic function is upper semicontinuous, but discontinuities have positive Lebesgue measure.
- The Weierstrass function [8, Remark of Theorem 6.1.2] which is continuous but nowhere differentiable and does not have bounded variations.
- The Cantor staircase [49, Section 6.5.3], which is monotone, nonconstant, and almost everywhere differentiable with null derivative.
- Pathological Lipschitz function for which local minimizers form a dense subset [41], or for which subgradient sequences may fail to have a minimizing behavior [22, 46].

In particular, *any method resembling a gradient algorithm on the value function may encounter insurmountable difficulties **even** if the problem data is arbitrarily smooth.*

Contributions

In light of the preceding examples, it is natural to turn towards more rigid classes $\{P, Q, \mathcal{Y}\}$ to hope for better results without compromising the applicability of the model to concrete problems. Polynomial functions and sets constitute one of the simplest instances of such classes: they have a strict algebraic nature, yet they cover a wealth of concrete real-word applications. We actually pertain to box-constrained polynomial bilevel optimization, i.e., P, Q are polynomials and \mathcal{Y} has a box-like form $[a_1, b_1] \times \dots \times [a_m, b_m]$ where $a_i, b_i \in \mathbb{R} \cup \{\pm\infty\}, i = 1, \dots, m$, independent of the upper level variable x . We investigate the worst case behavior of such bilevel polynomial optimization in two different ways:

1. **Geometric complexity:** Solving a bilevel problem is equivalent to optimizing its value function φ_o or φ_p (cf. (BP-O) and (BP-P)). Loosely speaking, we show that box-constrained polynomial bilevel programming is equivalent to the optimization of an arbitrary real semi-algebraic function (non necessarily continuous). In other words, arbitrary polynomial bilevel programming is not tractable. We actually provide sharp characterizations of the class of value functions of (BP), (BP-O), and (BP-P) in various scenarios, these results are given in Sections 2.
2. **Computational complexity:** We investigate (BP), (BP-O) and (BP-P) along the angle of the classical computational complexity theory [5]. Our second main result asserts that the decision version of bilevel polynomial optimization is Σ_2^P -hard. This implies that bilevel polynomial optimization is more challenging than many **NP**-complete problems.³

³Assuming that the polynomial hierarchy does not collapse on the first level, for which a broad consensus exists.

Connection to existing work

Our study provides an explication for limited theoretical guarantees of many proposed algorithms of bilevel optimization, especially when the lower-level problem is not (strongly) convex. In the literature, many works proposed or analyzed algorithms based on automatic differentiation [38, 37, 4, 39]. In this approach, one replaces the condition $y \in \Theta(x)$, cf. (BP), by an algorithm \mathcal{A} minimizing $Q(x, \cdot)$. Intuitively, if the algorithm \mathcal{A} is differentiable w.r.t. to the upper variable x , then one can also calculate the gradient of the bilevel problem w.r.t. to the upper variable x by the chain rule and use classical first-order methods. The difficulty of this approach is that for general nonconvex functions, most algorithms \mathcal{A} can only find stationary points or local minima of $Q(x, \cdot)$. Therefore, guarantees of these algorithms are established either under strong assumptions (e.g., regularity of $\Theta(x)$, uniform convergence of \mathcal{A}) [38, 39] or for a relaxed model (e.g., y is only required to be a stationary point) [4]. Another approach is based on smoothing techniques [37, 35] for which accumulation points of optimal solutions of the smooth approximation are the stationary points or minimizers of the original/approximate bilevel problems. Since the class of the value functions of general bilevel optimization problems can be large (e.g., Theorem 1.1 and our following results), it is *unsurprising* that standard convergence result of bilevel optimization might require much stronger assumptions to be established.

To avoid pathologies that will be shown in this work, another line of work of bilevel optimization focuses on establishing necessary conditions for locally optimal solutions. These works [52, 24, 42, 28] are mostly based on the reformulation of bilevel problems into single-level ones using the so-called value function of the lower-level problem [19], which is constrained to be non-positive. In [52], the authors proposed to use the so-called *partial calmness* to establish Karush-Kuhn-Tucker (KKT)-like necessary conditions for bilevel optimization. Their proposed qualification constraint plays an important role in this line of research since contrary to classical optimization, other popular constraint qualifications conditions such as linear independence (LICQ), Mangasarian-Fromovitz (MFCQ), and Slater's condition generally fail to hold (see [52, Proposition 4.1] for more discussion). The result in [52] applies only if we view bilevel programming as a classical optimization problem that jointly minimizes P w.r.t. to (x, y) with a special constraint $y \in \Theta(x)$. To achieve necessary conditions of the local solutions of $\varphi_o(x)$, subsequent works required in addition inner semicontinuity of certain set-valued mappings related to $\Theta(x)$ [24, 42, 28]. While these constraint qualification conditions and assumptions are non-trivial and mathematically interesting, it is difficult to identify a class of bilevel problems that satisfy all of them. Our results might partly explain this difficulty, at least in the polynomial setting (see e.g., Theorems 2.12 and 2.15).

Our study on worst-case analysis is different but complementary to works devoted to algorithmic aspects of polynomial bilevel optimization. Using dedicated tools from polynomial optimization, in [30], the authors proposed semidefinite programming (SDP) relaxations for polynomial bilevel optimization with lower-level convex problems. More specifically, thanks to the lower-level convexity (and other reasonable constraint qualifications), one can reformulate polynomial bilevel problems into single-level constrained optimization with polynomial objectives and constraints, using KKT conditions. It allows using SDP relaxation techniques for solving polynomial optimization [33]. When the lower-level problem is not convex, [30, 43] proposed to solve an ϵ -approximate version of the polynomial bilevel problem. Thus, the algorithm consists of two nested loops: the outer one solves the ϵ -approximation problem and the inner one solves a sequential SDP relaxation corresponding to a fixed ϵ . Cluster points of the sequence $\{(x_\epsilon, y_\epsilon)\}_{\epsilon>0}$ (optimal solutions for the ϵ -approximate problem) are the optimal solutions of the original problem (under certain assumptions).

This work also investigates computational complexity, an aspect that is well-studied for the cases of (integer) linear bilevel programming. Indeed, it is known that, while linear bilevel optimization (i.e., P, Q are linear, \mathcal{Y} is a polyhedron) is **NP**-hard [29, 7, 12], integer linear bilevel optimization (i.e., when several variables are constrained to be integers) belongs to the so-called Σ_2^P -hard problems [40, 17, 29]. Essentially, the class Σ_2^P is a generalization of the notions **P** and **NP** to capture the complexity of integer linear bilevel programming (see Section 3.1 for more details). Moreover, if the polynomial hierarchy does not collapse in the first level, i.e. $\mathbf{NP} \subsetneq \Sigma_2^P$, these results imply that the complexity of the Σ_2^P -hard integer linear bilevel optimization problem is much higher than that of any **NP** complete problem. We prove the analog Σ_2^P -hardness of *polynomial bilevel optimization*. This result is natural since single-level polynomial optimization itself is **NP** hard (similarly as linear programming with integer constraints), and bilevel problems typically incur a complexity jump in the polynomial hierarchy.

2 Geometric hardness of polynomial bilevel optimization

Solving an (optimistic/pessimistic) bilevel problem is equivalent to optimizing the corresponding value function (φ_o or φ_p in (BP-O), (BP-P)) which depends solely on the upper-level variables. In this section, we study the complexity of the bilevel polynomial optimization problem by investigating the sets of value functions that can be “expressed” via polynomial bilevel formulations.

First, we distinguish situations for which the feasible set of the lower-level optimization problem, \mathcal{Y} , is a box of the form $[a_1, b_1] \times \dots \times [a_m, b_m]$ which can be bounded or unbounded. More specifically, we say that \mathcal{Y} is an *unbounded box* (or \mathcal{Y} is *unbounded*) if there exists an index $1 \leq i \leq m$ such that $a_i = -\infty$ or $b_i = +\infty$. Otherwise, \mathcal{Y} is a *bounded box* (or simply, *bounded*). It is worth mentioning that, since we are investigating worst case properties, the restriction to box type constraints is actually a strength and our results hold for more general constraint sets.

Second, we restrict our attention to convex lower-level problems. This means that the lower-level objective $Q(x, y)$ is convex with respect to its second argument y and the constraint set \mathcal{Y} is also convex. This represents a natural intermediate situation between strongly convex lower levels and general nonconvex ones. In this setting, we make a distinction between bounded and unbounded box on the one hand, as well as general compact convex sets on the other hand.

Our analysis provides a sharp characterization of the class of functions which can be represented as the value function of polynomial bilevel programs: we show that for the most natural types of bilevel programs, this class is the largest possible. This is described in Section 2.2. Furthermore, our analysis shows that enforcing convexity of the lower-level problems does not reduce much the corresponding class of polynomial bilevel value functions. This is discussed in Section 2.3, which is more technical than the general case for which our results are sharper. In the convex setting, we distinguish between box-constrained convex lower level, which we relate to the class of piecewise polynomial functions, and general convex compact semi-algebraic set for which we obtain a sharper characterization.

The rest of this section is organized as follows. We start in Section 2.1 by providing readers with preliminaries on the mathematical tools required for our analysis: classical results from semi-algebraic geometry and bilevel optimization. The representation results for general polynomial bilevel problems are given in Section 2.2 and the specification to convex lower-level problems is described in Section 2.3. A summary of all the results of this section is provided in Section 2.4.

2.1 Preliminaries

First, we will recall the basic definitions and results of semi-algebraic geometry and their consequences for polynomial bilevel programming. Second, we remind the readers of a simplified version of Berge's Maximum Theorem [14, Section 6.3] and its consequence on semicontinuity properties of bilevel programs. These classical results play a central role in our analysis, and we provide a detailed account for completeness. The informed reader may skip this subsection.

Semi-algebraic geometry and polynomial bilevel value functions

We provide some definitions and facts about semi-algebraic geometry. For an excellent exhibition of this subject, we refer readers to [9].

Definition 2.1 (Semi-algebraic sets and functions). A set $S \in \mathbb{R}^n$ is a *basic* semi-algebraic set if it has the form:

$$S = \{x \in \mathbb{R}^n \mid P(x) = 0 \text{ and } Q_j(x) > 0, j \in J\},$$

where J is a finite index set the functions $P, Q_j, j \in J$ are polynomials. A semi-algebraic set is a finite union of basic semi-algebraic sets.

A function $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a semi-algebraic function if $\text{graph } f := \{(x, f(x)) \in \mathbb{R}^n \times \mathbb{R}^m \mid x \in U\}$ is a semi-algebraic set of \mathbb{R}^{n+m} .

One can even define a semi-algebraic set as a *disjoint union* of finite basic semi-algebraic sets (and not just a *finite union* as in Definition 2.1). In order to be self-contained, we provide a proof for this claim in Section B. In the following, we also consider a definition for extended-real-valued semi-algebraic functions.

Definition 2.2 (Extended-real-valued semi-algebraic functions). A function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is an extended-real-valued semi-algebraic function if three sets $\text{dom}(f) := \{x \in \mathbb{R}^n \mid f(x) \in \mathbb{R}\}$, $\text{dom}(f)^+ := \{x \in \mathbb{R}^n \mid f(x) = +\infty\}$ and $\text{dom}(f)^- := \{x \in \mathbb{R}^n \mid f(x) = -\infty\}$ are semi-algebraic and the function $f|_{\text{dom}(f)}$ (the restriction of f to $\text{dom}(f)$) is semi-algebraic (cf. Definition 2.1). The class of all extended-real-valued semi-algebraic functions is denoted by \mathcal{SA} . For a semi-algebraic function $f : \mathbb{R}^p \rightarrow \mathbb{R}$, we have $\text{dom}(f)^+ = \text{dom}(f)^- = \emptyset$, such functions are called real-valued by opposition to extended-real-valued.

A fundamental result in semi-algebraic geometry is the Tarski-Seidenberg theorem, proving the stability of semi-algebraic sets under projection (and consequently, first-order logic).

Theorem 2.3 (Tarski-Seidenberg theorem). *Let $S \subseteq \mathbb{R}^n$ be a semi-algebraic set, and $\pi : \mathbb{R}^n \rightarrow \mathbb{R}^{n-1}$ be the projection onto the first $n-1$ coordinates. Then $\pi(S) = \{x \in \mathbb{R}^{n-1} \mid \exists y \in \mathbb{R}, (x, y) \in S\}$ is also semi-algebraic.*

Besides general semi-algebraic sets, closed semi-algebraic sets have a more particular form, given in Proposition 2.4.

Proposition 2.4 (Characterization of closed semi-algebraic sets [13, Exercise 2.5.7]). *Every closed semi-algebraic set S in \mathbb{R}^n can be represented in the form:*

$$S = \bigcup_{i \in I} \bigcap_{j \in J} \{x \in \mathbb{R}^n \mid P_{ij}(x) \geq 0\}, \quad (1)$$

where P_{ij} are polynomials and I, J are finite index sets.

Finally, we recall a result on the growth of a semi-algebraic function.

Proposition 2.5 (Growth of semi-algebraic functions [50, Section 4.12]). *For every semi-algebraic function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, there exists a natural number N and a constant C such that $|f(x)| \leq \|x\|^N$ for all $x, \|x\| \geq C$ where $\|\cdot\|$ indicates the Euclidean norm.*

We conclude this section with an application of the Tarski-Seidenberg theorem to justify that optimistic (resp. pessimistic) value functions of polynomial bilevel problems are semi-algebraic (possibly with extended-real values $\pm\infty$). Consider optimistic bilevel optimization: given P, Q, \mathcal{Y} , there is a partition of \mathbb{R}^n into three disjoint sets $\text{dom}(\varphi_o)$, $\text{dom}(\varphi_o)^+$, and $\text{dom}(\varphi_o)^-$ defined as follows:

1. $\text{dom}(\varphi_o) := \{x \in \mathbb{R}^n \mid \exists \delta \in \mathbb{R}, \forall y \in \Theta(x), \delta \leq P(x, y) \text{ and } \forall \epsilon > 0, \exists y' \in \Theta(x), \delta + \epsilon > P(x, y')\}$: if $x \in \text{dom}(\varphi_o)$, then $\varphi_o(x) \in \mathbb{R}$.
2. $\text{dom}(\varphi_o)^+ := \{x \in \mathbb{R}^n \mid \nexists y \in \Theta(x)\}$: if $x \in \text{dom}(\varphi_o)^+$, then $\varphi_o(x) = +\infty$.
3. $\text{dom}(\varphi_o)^- := \{x \in \mathbb{R}^n \mid \forall \delta \in \mathbb{R}, \exists y \in \Theta(x), P(x, y) < \delta\}$: if $x \in \text{dom}(\varphi_o)^-$, then $\varphi_o(x) = -\infty$.

where $\Theta(x) = \{y \in \mathcal{Y} \mid \forall y' \in \mathcal{Y}, Q(x, y) \leq Q(x, y')\}$ is also a first-order logic expression (thus, semi-algebraic). Thus, all three sets are semi-algebraic. Moreover, the graph of the restriction of φ_o to its domain $\text{dom}(\varphi_o)$ is given by:

$$\{(x, P(x, y)) \in \mathbb{R}^n \times \mathbb{R} \mid x \in \text{dom}(\varphi_o), y \in \Theta(x) \text{ and } \forall y' \in \Theta(x), P(x, y) \leq P(x, y')\}.$$

Using Tarski-Seidenberg quantifier elimination, we have the following proposition:

Proposition 2.6 (Semi-algebraicity of value functions). *The value functions of any polynomial bilevel optimization problem with box constraint is semi-algebraic.*

Berge's Maximum Theorem and semicontinuity of bilevel value functions

For the analysis in the bounded setting, we will use Berge's maximum theorem. Its presentation involves the notion of outer semicontinuity of compact set-valued maps.

Definition 2.7 (Outer semicontinuity). A compact set-valued map $\Theta : \mathcal{X} \rightrightarrows \mathcal{Y}$ is called *outer semicontinuous* if, and only if, for all sequences $(x_k)_{k \in \mathbb{N}}$ of \mathcal{X} and $(y_k)_{k \in \mathbb{N}}$ such that $y_k \in \Theta(x_k)$, if $\lim_{k \rightarrow \infty} x_k = x$, $\lim_{k \rightarrow \infty} y_k = y$, then $y \in \Theta(x)$.

In the following, we provide a simplified version of Berge's Maximum Theorem to keep our discussion as simple as possible.

Theorem 2.8 (Berge's Maximum Theorem [14, Section 6.3]). *Consider a continuous function $g : \mathbb{R}^n \times \mathbb{R}^m$ and a compact set \mathcal{Y} . Define $\Theta(x) = \arg \min\{g(x, y) : y \in \mathcal{Y}\}$, we have $\Theta : \mathbb{R}^n \rightrightarrows \mathcal{Y}$ is an outer semicontinuous set-valued mapping with non-empty and compact values.*

As a consequence of Berge's Theorem, one obtains classical semicontinuity properties for bilevel value functions for which we provide a proof for completeness. Recall that a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is lower (resp. upper) semicontinuous if, and only if, its epigraph $\text{epi}(f) := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha\}$ (resp. hypograph $\text{hyp}(f) := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \geq \alpha\}$) is closed.

Corollary 2.9 (Semicontinuity of bilevel value functions with compact lower-level feasible sets). *Let P, Q and \mathcal{Y} as in (BP) be such that P and Q are continuous and \mathcal{Y} is compact. Then, for all $x \in \mathbb{R}^n$, $\Theta(x)$ is non-empty and $P(x, \cdot)$ attains its min/max in $\Theta(x)$. Moreover, the corresponding value function φ_o (resp. φ_p) for the optimistic bilevel problem (BP-O) (resp. pessimistic bilevel problem (BP-P)) is lower (resp. upper) semicontinuous and bounded on every compact.*

Proof. We will prove the proposition for the case of optimistic polynomial bilevel formulation. The pessimistic counterpart is similar. Consider the optimistic formulation (BP-O), we have:

1. Non-emptiness of $\Theta(x)$: if \mathcal{Y} is compact, then the set $\Theta(x)$ is non-empty and compact thanks to Theorem 2.8. Thus, the continuous P attains its minimum/maximum in $\Theta(x)$.
2. Lower-semicontinuity of φ_o : Consider a point $x \in \mathbb{R}^n$ and a sequence $(x_k)_{k \in \mathbb{N}}$ converging to x . Since $\Theta(x)$ is compact, there exists $y_k \in \Theta(x_k)$ such that $\varphi_o(x_k) = P(x_k, y_k)$. We need to prove that: $\varphi_o(x) \leq \liminf_{k \rightarrow \infty} \varphi_o(x_k)$.

In the following, we can assume that $\varphi_o(x_k)$ converges to $\liminf_{k \rightarrow \infty} \varphi_o(x_k)$ and prove this limit is at least $\varphi_o(x)$. Due to the compactness of \mathcal{Y} , the sequence y_k admits at least an accumulation point y . Due to the outer semicontinuity of Θ , $y \in \Theta(x)$. Due to the continuity of P , we also have: $\liminf_{k \rightarrow \infty} \varphi_o(x_k) = P(x, y)$. Therefore,

$$\varphi_o(x) = P(x, y^*) \leq P(x, y) = \liminf_{k \rightarrow \infty} \varphi_o(x_k),$$

where $y^* \in \arg \min\{P(x, y) \mid y \in \Theta(x)\}$.

3. Boundedness on compact sets of φ_o : Since φ_o is lower-semicontinuous, it is lower-bounded in a given compact set C . In addition, $\forall x \in C$, we also have:

$$\max_{x \in C} \varphi_o(x) = \max_{x \in C} \min_{y \in \Theta(x)} P(x, y) \leq \max_{x \in C} \max_{y \in \mathcal{Y}} P(x, y) = \max_{(x, y) \in C \times \mathcal{Y}} P(x, y).$$

Since P is continuous and $C \times \mathcal{Y}$ is compact, there exists a constant C such that $\max_{x \in C} \varphi_o(x) \leq C$. Thus, φ_o is also upper-bounded in C . The proof is concluded. \square

Remark 2.10. As a consequence of Corollary 2.9, semi-algebraic functions that are not lower (resp. upper) semicontinuous cannot be expressed as optimistic (resp. pessimistic) bilevel polynomial optimization with compact lower-level feasible sets.

2.2 Value functions of general polynomial bilevel programs

We introduce the following notation which is a shorthand for value function classes for different types of polynomial bilevel problems: optimistic or pessimistic, bounded or unbounded, etc.

Definition 2.11 (Value functions classes). Given $\text{mode} \in \{\text{optim}, \text{pessim}\}$ (standing for optimistic and pessimistic respectively) and $\text{boxtype} \in \{\text{bounded}, \text{unbound}\}$, we define $\mathcal{P}_{\text{boxtype}}^{\text{mode}}$ the set of value functions that can be represented as a mode (optimistic or pessimistic) bilevel optimization problem with P, Q polynomials and \mathcal{Y} a box satisfying boxtype (bounded or unbounded), i.e.:

$$\mathcal{P}_{\text{boxtype}}^{\text{mode}} := \{h : \mathbb{R}^n \rightarrow \mathbb{R} \mid \exists P, Q \text{ polynomials, } \mathcal{Y} \text{ boxtype satisfy and } \varphi_{\text{mode}} = h\}. \quad (2)$$

with the convention that φ_{optim} and φ_{pessim} are the value functions corresponding to problems (BP-O) and (BP-P) (i.e., φ_o and φ_p), respectively.

For example, if one takes `mode = o`, `boxtype = unbound`, we have:

$$\mathcal{P}_{\text{unbound}}^{\text{optim}} := \{h \mid \exists P, Q \text{ polynomials, } \mathcal{Y} \text{ unbounded and } \varphi_o = h\},$$

where φ_o defined in (BP-O). We consider the cases of unbounded and bounded \mathcal{Y} separately.

With our notation, Proposition 2.6 asserts that $\mathcal{P}_{\text{unbound}}^{\text{optim}} \subset \mathcal{SA}$ and $\mathcal{P}_{\text{unbound}}^{\text{pessim}} \subset \mathcal{SA}$. It is natural to ask whether these inclusions are tight and the following theorem provides a positive answer.

Theorem 2.12 (Value functions of polynomial bilevel programming). *Any extended-real-valued semi-algebraic function is the value function of a polynomial bilevel problem whose lower-level problem is unconstrained. In particular,*

$$\mathcal{P}_{\text{unbound}}^{\text{optim}} = \mathcal{P}_{\text{unbound}}^{\text{pessim}} = \mathcal{SA}.$$

Sketch of proof for Theorem 2.12. We provide a high-level idea of proof here: Given an extended-valued semi-algebraic function $h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$, we partition \mathbb{R}^n into three disjoint semi-algebraic components: $\text{dom}(h)$, $\text{dom}(h)^+$, $\text{dom}(h)^-$ as in Definition 2.2.

For optimistic bilevel optimization (the pessimistic counterpart can be adapted correspondingly), the main idea of our construction is to build upper-level and lower-level functions P, Q such that:

1. If $x \in \text{dom}(h)$, then $\Theta(x) := \arg \min_y Q(x, \cdot)$ is non-empty. Moreover, $\min_{y \in \Theta(x)} P(x, y) = h(x)$ (note that we use min, instead of inf, which implies that the minimum value is attained).
2. If $x \in \text{dom}(h)^+$, then $\Theta(x) = \emptyset$.
3. If $x \in \text{dom}(h)^-$, then $\Theta(x) \neq \emptyset$ but $\inf_{y \in \Theta(x)} P(x, y) = -\infty$.

□

Proof of Theorem 2.12. In this proof, we only consider the optimistic bilevel optimization. The other case can be proved similarly.

Consider an extended-real-valued semi-algebraic function $h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$. By definition $\text{graph } h = \{(x, h(x))\}$ is a semi-algebraic set, and it can be written as:

$$\text{graph } h = \bigcup_{i \in I} \mathcal{S}_i \quad \text{where} \quad \mathcal{S}_i := \{(x, t) \in \mathbb{R}^{n+1} \mid P_i(x, t) = 0 \text{ and } Q_{ij}(x, t) > 0, j \in J\}, \quad (3)$$

where I and J are some finite index sets, the functions $P_i, i \in I$ and $Q_{ij}, (i, j) \in I \times J$ are polynomials w.r.t $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$.

Consider $\text{dom}(h), \text{dom}(h)^+, \text{dom}(h)^-$ defined as in Definition 2.2: WLOG, we also assume that:

$$\begin{aligned} \text{dom}(h) &:= \bigcup_{i \in I} \mathcal{D}_i \quad \text{where} \quad \mathcal{D}_i := \{x \in \mathbb{R}^n \mid R_i(x) = 0 \text{ and } S_{ij}(x) > 0\} \\ \text{dom}(h)^+ &:= \bigcup_{i \in I} \mathcal{D}_i^+ \quad \text{where} \quad \mathcal{D}_i^+ := \{x \in \mathbb{R}^n \mid R_i^+(x) = 0 \text{ and } S_{ij}^+(x) > 0\} \\ \text{dom}(h)^- &:= \bigcup_{i \in I} \mathcal{D}_i^- \quad \text{where} \quad \mathcal{D}_i^- := \{x \in \mathbb{R}^n \mid R_i^-(x) = 0 \text{ and } S_{ij}^-(x) > 0\} \end{aligned} \quad (4)$$

Name	Dimension	Coordinates notation	Type (Upper/Lower variable)
x	n	Not used	Upper
t	1	Not used	Lower
z	$ I \times J $	z_{ij}	Lower
ν	$ I \times J $	ν_{ij}	Lower
ν^+	$ I \times J $	ν_{ij}^+	Lower
ν^-	$ I \times J $	ν_{ij}^-	Lower
u	1	Not used	Lower
v	1	Not used	Lower

Table 1: Specification for the variables of the bilevel formulation.

with the same index sets I, J as in (3) (otherwise, one can add “dummy” polynomial equalities/inequalities, e.g., $0 = 0, 0 < 1$ to match the index sets).

In our construction, we use eight sets of variables that are described in Table 1. In particular, the lower-level variable y is the concatenation of $(t, z, \nu, \nu^+, \nu^-, u, v)$.

Consider the function:

$$F(x, t, z) := \prod_{i \in I} \left(P_i(x, t)^2 + \sum_{j \in J} G_{ij}(x, t, z_{ij}) \right). \quad (5)$$

$$G_{ij}(x, t, z_{ij}) = (1 - Q_{ij}(x, t)z_{ij}^2)^2,$$

where Q_{ij} are the polynomials defined in Equation (3). Let us determine its global minimizers for a fixed value of (x, t) . One gets the closed form for the minimizers z_{ij}^* and the optimal values G_{ij}^* of $G_{ij}(x, t, \cdot)$:

$$z_{ij}^* = \begin{cases} \pm \sqrt{\frac{1}{Q_{ij}(x, t)}} & \text{if } Q_{ij}(x, t) > 0 \\ 0 & \text{otherwise,} \end{cases} \quad G_{ij}^* = \begin{cases} 0 & \text{if } Q_{ij}(x, t) > 0 \\ 1 & \text{otherwise} \end{cases}. \quad (6)$$

Given a fixed value $x \in \text{dom}(h)$, consider two cases:

1. If $(x, t) \in \text{graph } h$ (or equivalently, $t = h(x)$), there exists $i \in I$ such that $P_i(x, t) = 0$ and $Q_{ij}(x, t) > 0, \forall j \in J$. Using (6), we have:

$$P_i(x, t)^2 + \sum_{j \in J} G_{ij}(x, t, z_{ij}^*) = 0.$$

Therefore, $F(x, h(x), z^*) = 0$, in other words z^* is a global minimizer of $F(x, h(x), \cdot)$.

2. If $(x, t) \notin \text{graph } h$ (or equivalently, $t \neq h(x)$), then for all $i \in I$, either $P_i(x, t) \neq 0$ or there is j such that $Q_{ij}(x, t) \leq 0$. In any case, using (6) again, we can conclude that:

$$P_i(x, t)^2 + \sum_{j \in J} G_{ij}(x, t, z_{ij}^*) > 0, \forall i \in I.$$

Therefore, if $t \neq h(x)$, we have $F(x, t, z) > 0, \forall z \in \mathbb{R}^{|I| \times |J|}$.

Analogous to (5), we construct three nearly similar polynomials, using the functions S, R defined in (4).

$$\begin{aligned} H(x, \nu) &:= \prod_{i \in I} \left(R_i(x)^2 + \sum_{j \in J} K_{ij}(x, \nu_{ij}) \right) \quad \text{where} \quad K_{ij}(x, \nu_{ij}) = (1 - S_{ij}(x) \nu_{ij}^2)^2 \\ H^+(x, \nu) &:= \prod_{i \in I} \left(R_i^+(x)^2 + \sum_{j \in J} K_{ij}^+(x, \nu_{ij}^+) \right) \quad \text{where} \quad K_{ij}^+(x, \nu_{ij}^+) = (1 - S_{ij}(x) [\nu_{ij}^+]^2)^2 \\ H^-(x, \nu) &:= \prod_{i \in I} \left(R_i^-(x)^2 + \sum_{j \in J} K_{ij}^-(x, \nu_{ij}^-) \right) \quad \text{where} \quad K_{ij}^-(x, \nu_{ij}^-) = (1 - S_{ij}(x) [\nu_{ij}^-]^2)^2 \end{aligned} \quad (7)$$

Using a similar argument for F , we can conclude that:

1. if $x \in \text{dom}(h)$ (resp. $\text{dom}(h)^+, \text{dom}(h)^-$), there exists ν (resp. ν^+, ν^-) such that $H = 0$ (resp. $H^+ = 0, H^- = 0$);
2. Otherwise, $H(x, \nu) > 0$ (resp. $H^+(x, \nu^+) > 0, H^-(x, \nu^-) > 0$) for all ν (resp. ν^+, ν^-).

Using F, H, H^+, H^- , we construct the upper-level and lower-level polynomials P and Q as follows:

$$\begin{aligned} P(x, y) &= t, \\ Q(x, y) &= \underbrace{H^+(x, \nu^+) H^-(x, \nu^-) F(x, t, z)}_{Q_1(x, t, z, \nu^+, \nu^-)} + \underbrace{H(x, \nu) H^-(x, \nu^-) u^2 + (1 - uv)^2}_{Q_2(x, \nu, \nu^-, u, v)}. \end{aligned}$$

By construction, all functions F, H, H^+, H^- are sums of squares. Therefore, given a fixed value x , if $Q(x, y) = 0$, then y belongs to the set of minimizers of $Q(x, \cdot, \cdot)$. We consider three cases corresponding to the partition of \mathbb{R}^n :

1. If $x \in \text{dom}(h)$: on the one hand, $\min Q(x, y) = 0$ and it is attained since we can choose (t, z, ν, u, v) such that $F(x, t, z) = 0$ (see eq. (6)), $H(x, \nu) = 0$ (see remark after eq. (7)), $uv = 1$. On the other hand, if y minimizes $Q(x, \cdot, \cdot)$, that is $y \in \Theta(x)$, then $t = h(x)$ due to our analysis of $F(x, t, z)$ and the fact that $H^+(x, \nu^+) H^-(x, \nu^-) > 0$, for any ν^+, ν^- . Thus, $P(x, y) = t = h(x)$.
2. If $x \in \text{dom}(h)^+$: the infimum $\inf Q(x, y) = 0$ but it is not attained. Indeed, by choosing ν^+, u, v such that $H^+(x, \nu^+) = 0$ (see remark after eq. (7)), $u \rightarrow 0$ and $v = 1/u \rightarrow +\infty$, $Q(x, y)$ can get arbitrarily close to zero. Nevertheless, the minimum is not attained since for any $\nu, \nu^-, H(x, \nu) H^-(x, \nu^-) > 0$ and one can verify that $au^2 + (1 - uv)^2 > 0$, for any u, v for any $a > 0$, which concludes the proof for this case.
3. If $x \in \text{dom}(h)^-$: $\min Q(x, y) = 0$ and it is attained for any $t \in \mathbb{R}$, by choosing ν^-, u, v such that $H^-(x, \nu^-) = 0$ (see remark after eq. (7)) and $uv = 1$. Thus, the optimistic bilevel optimization results in $\inf P(x, y) = -\infty$.

That concludes the proof. \square

Remark 2.13. We remark that the degree of the constructed polynomial Q is linear in the degrees of polynomials P_i, Q_{ij} defining the graph of the target function h . Since all the information about a bilevel problem is encoded in a single pair of polynomials (P, Q) it is natural that their degree increases depending on the complexity of the underlying representation. A similar comment holds for all the constructions of this section. We leave more quantitative discussions about this representation for future work.

Remark 2.14. While we focus on bilevel programming, the above results actually characterize semi-algebraic functions as a polynomial arg min since the upper level is just the projection on the first coordinate. We also remark that the proof allows to obtain a representation of semi-algebraic sets using a non-negative polynomial argmin or equivalently the zero locus of a polynomial.

Denote respectively by $\mathcal{LSC}, \mathcal{USC}, \mathcal{CB}$ the sets of functions that are lower semicontinuous, upper semicontinuous, and bounded on any compact set (cf. Corollary 2.9). Similar to the unbounded case, combining Proposition 2.6 and Corollary 2.9, we have that asserts that $\mathcal{P}_{\text{bounded}}^{\text{optim}} \subset \mathcal{SA} \cap \mathcal{LSC} \cap \mathcal{CB}$ and $\mathcal{P}_{\text{bounded}}^{\text{pessim}} \subset \mathcal{SA} \cap \mathcal{USC} \cap \mathcal{CB}$. The following shows that these inclusions are tight. Note that, by definition, functions in \mathcal{CB} have full domain (they do not take value $\pm\infty$).

Theorem 2.15 (Value functions of box-constrained polynomial bilevel programming). *Any function which is semi-algebraic, lower (resp. upper) semicontinuous, and bounded on compact sets is the value function of an optimistic (resp. pessimistic) polynomial bilevel problem whose lower-level feasible set \mathcal{Y} is a bounded box. In other words:*

$$\mathcal{P}_{\text{bounded}}^{\text{optim}} = \mathcal{SA} \cap \mathcal{LSC} \cap \mathcal{CB}, \quad \mathcal{P}_{\text{bounded}}^{\text{pessim}} = \mathcal{SA} \cap \mathcal{USC} \cap \mathcal{CB}.$$

Proof. To prove the equalities in Theorem 2.21, we notice that Proposition 2.6 and Corollary 2.9 imply the following:

$$\begin{aligned} \mathcal{P}_{\text{bounded}}^{\text{optim}} &\subseteq \mathcal{SA} \cap \mathcal{LSC} \cap \mathcal{CB}, \\ \mathcal{P}_{\text{bounded}}^{\text{pessim}} &\subseteq \mathcal{SA} \cap \mathcal{USC} \cap \mathcal{CB}. \end{aligned}$$

Therefore, it is sufficient to prove the first claim of Theorem 2.21: for any element $f : \mathbb{R}^n \rightarrow \mathbb{R}$ of $\mathcal{SA} \cap \mathcal{LSC} \cap \mathcal{CB}$ (resp. $\mathcal{SA} \cap \mathcal{USC} \cap \mathcal{CB}$), there exists an optimistic (resp. pessimistic) polynomial problem with a bounded box \mathcal{Y} that equals f . Note that if f is bounded on any compact, $f(x) \in \mathbb{R}, \forall x \in \mathbb{R}^n$.

In this proof, we deal with the case of optimistic formulation. The other case can be dealt with similarly. Before constructing P, Q and the bounded set \mathcal{Y} , we state three key observations concerning f .

1. Since f is semi-algebraic, $\text{graph } f$ is also semi-algebraic. Consequently, the closure $\overline{\text{graph } f}$ is also semi-algebraic (see, for example, [9, Proposition 3.1]). By Proposition 2.4, $\overline{\text{graph } f}$ can be represented as:

$$\overline{\text{graph } f} = \bigcup_{i \in I} \bigcap_{j \in J} \{(x, t) \in \mathbb{R}^{n+1} \mid P_{ij}(x, t) \geq 0\}, \quad (8)$$

where P_{ij} are polynomials with $n + 1$ variables, and I, J are two finite index sets.

2. Since f is lower-semicontinuous one has $f(x) = \min\{t \mid t \in \mathbb{R}, (x, t) \in \overline{\text{graph } f}\}$.

3. Since f is semi-algebraic, by Proposition 2.5, there exist a positive constant C and an integer N such that $|f(x)| \leq \|x\|^N, \forall x, \|x\| \geq C$. We can choose N to be an even integer so that $\|x\|^N$ is a polynomial. On the other hand, since f is bounded on compact sets, there exists another positive constant B such that $|f(x)| \leq B, \forall x, \|x\| \leq C$. By combining these two observations, we have:

$$f(x) \leq B + \|x\|^N, \forall x \in \mathbb{R}^n.$$

Using this remark, we may assume that f is bounded in $[-1, 1]$.

[Indeed, consider the function: $h(x) = f(x)/(B + \|x\|^N)$. Since the polynomial $B + \|x\|^N \geq B, \forall x \in \mathbb{R}$, $h(x)$ is well-defined, and it remains semi-algebraic, lower-semicontinuous. Moreover, h is bounded in $[-1, 1]$. If one can construct P, Q and \mathcal{Y} such that φ_o in (BP-O) is such that $\varphi_o = h$, then the optimistic bilevel problem with $(P(B + \|x\|^N), Q, \mathcal{Y})$ has a value function equal to f .]

In our construction, we use three sets of variables. They are described as in Table 2.

Name	Dimension	Coordinates notation	Type (Upper/Lower variable)	Feasible set
x	n	Not used	Upper	N/A
t	1	Not used	Lower	$[-1, 1]$
z	$ I \times J $	z_{ij}	Lower	$[0, 1/2]^{ I \times J }$

Table 2: Specification for the variables of the bilevel formulation.

We introduce the building block of our polynomials P and Q . Consider:

$$G_{ij}(x, t, z_{ij}) = (P_{ij}(x, t) - (P_{ij}(x, t)^2 + 1)z_{ij})^2,$$

where P_{ij} are the polynomials defined in (8).

Given a fixed value of (x, t) , optimizing $G_{ij}(x, t, z_{ij})$ w.r.t $z_{ij} \in [0, 1/2]$, the optimal value and minimizer of G_{ij} are given by:

$$z_{ij}^* = \begin{cases} \frac{P_{ij}(x, t)}{P_{ij}(x, t)^2 + 1} & \text{if } P_{ij}(x, t) \geq 0 \\ 0 & \text{otherwise} \end{cases}, \quad G_{ij}^*(x, t) = \begin{cases} 0 & \text{if } P_{ij}(x, t) \geq 0 \\ P_{ij}(x, t)^2 & \text{otherwise} \end{cases} \quad (9)$$

Note indeed that $z_{ij}^* \in [0, 1/2]$ because $t/(1 + t^2) \in [0, 1/2], \forall t \geq 0$. As a consequence $G_{ij}^*(x, t) = 0$ if and only if $P_{ij}(x, t) \geq 0$.

Let us now define P, Q and \mathcal{Y} as:

$$\begin{aligned} P(x, t, z) &= t, \\ Q(x, t, z) &= \prod_{i \in I} \left(\sum_{j \in J} G_{ij}(x, t, z_{ij}) \right), \\ \mathcal{Y} &= [-1, 1] \times [0, 1/2]^{|I| \times |J|}. \end{aligned} \quad (10)$$

Note that Q is the product of sums of squares. Hence, $Q(x, t, z) \geq 0$. Consequently, if $Q(x, t, z) = 0$, then $(t, z) \in \arg \min_{\mathcal{Y}} Q(x, \cdot, \cdot)$.

Consider a point $x \in \mathbb{R}^n$, there are two possibilities:

1. If $(x, t) \in \overline{\text{graph } f}$, then there exists $i \in I$ such that $P_{ij}(x, t) \geq 0, \forall j \in J$. By (9), we have:

$$\sum_{j \in J} G_{ij}^*(x, t, z_{ij}) = 0.$$

Thus, the optimal value of Q in this case is zero.

2. If $(x, t) \notin \overline{\text{graph } f}$, then for all $i \in I$, there exists at least an index $j_i \in J$ such that $P_{ij_i}(x, t) < 0$. Hence,

$$\sum_{j \in J} G_{ij}(x, t, z_{ij}^*) \geq P_{ij_i}(x, t)^2 > 0, \forall i \in I.$$

Thus, the optimal value of Q is at least $\prod_{ij} P_{ij_i}(x, t)^2 > 0$.

Therefore, for a minimizer (t, z) of $Q(x, \cdot, \cdot)$, $(x, t) \in \overline{\text{graph } f}$. We emphasize that such t always exists and belongs to the interval $[-1, 1]$ due to the hypothesis of boundedness of f . Finally, among $\{t \in [-1, 1] \mid (x, t) \in \overline{\text{graph } f}\}$, the optimistic formulation will choose the smallest t , which is exactly the value of $f(x)$ due to the lower-semicontinuity of f . \square

2.3 Polynomial bilevel problems with convex lower-level

We also investigate the role of convexity of the lower-level problem in the set of expressible functions. Analogously to Definition 2.11, under the assumption that the lower-level problem is convex, we study the following value function class, where the letter \mathcal{C} highlights convexity.

Definition 2.16 (Value functions with lower-level convexity). Under the same notations as in Definition 2.11, the sets $\mathcal{C}_{\text{boxtype}}^{\text{mode}}$ are analogously defined to $\mathcal{P}_{\text{boxtype}}^{\text{mode}}$, except that the lower-level problem is constrained to be convex in the definition of the former, i.e.:

$$\mathcal{C}_{\text{boxtype}}^{\text{mode}} := \{h : \mathbb{R}^n \rightarrow \mathbb{R} \mid \exists P, Q \text{ polynomials, } Q \text{ convex w.r.t } y, \mathcal{Y} \text{ boxtype such that } \varphi_{\text{mode}} = h\}. \quad (11)$$

Similar to the previous section, we treat the case of bounded and unbounded lower-level constraints separately. This section is concluded with a third result for which we allow the lower-level constraint set to be an arbitrary convex compact semi-algebraic set.

From Definition 2.11 and Definition 2.16, we clearly have that:

$$\mathcal{C}_{\text{boxtype}}^{\text{mode}} \subseteq \mathcal{P}_{\text{boxtype}}^{\text{mode}},$$

for $\text{mode} \in \{\text{optim}, \text{pessim}\}, \text{boxtype} \in \{\text{unbound}, \text{bounded}\}$. However, we do not know if this inclusion is strict. Nevertheless, in this section, we show that the class of value functions in $\mathcal{C}_{\text{unbound}}^{\text{mode}}, \text{mode} \in \{o, p\}$ is very large as it contains all *piecewise polynomial functions*. We denote by $\mathbf{1}_S$ be the characteristic function of a subset $S \subseteq \mathbb{R}^n$ (with value 1 on S and 0 elsewhere), a piecewise polynomial can be defined as follows:

Definition 2.17 (Piecewise polynomial functions). A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is called *piece-wise polynomial* if there exist a semi-algebraic partition $S := \{S_i, i = 1, \dots, N\}$ of \mathbb{R}^n (i.e., S_i are semi-algebraic, pairwise disjoint and their union is equal to \mathbb{R}^n) and N polynomials $P_i : \mathbb{R}^n \rightarrow \mathbb{R}$ such that:

$$f(x) = \sum_{i=1}^N \mathbf{1}_{x \in S_i} P_i(x), \forall x \in \mathbb{R}^n.$$

Equivalently, $f(x) = P_i(x)$ for $x \in S_i$. We call $(S, (P_i)_{i=1}^N)$ the representation of f and write $f = (S, (P_i)_{i=1}^N)$ by an abuse of notations. We use the shorthand \mathcal{PP} to denote the set of piecewise polynomial functions.

In the following, we show that the set of piecewise polynomial functions is contained in the set of value functions with convex lower-level and simple box constraints. This result illustrates that relaxing strong convexity but maintaining convex lower levels in bilevel programs allows to represent the large class of piecewise polynomial functions. While this is a strictly smaller class compared to semi-algebraic functions, this is still a very large class which contains functions which are generally discontinuous with an arbitrary number of discontinuities, and gradient type methods are not adapted to such functions [16].

Theorem 2.18 (Piecewise polynomials are value functions with convex lower-level). *Any piecewise polynomial function is the value function of an optimistic (resp. pessimistic) polynomial bilevel problem with lower objective convex over a box. More specifically:*

$$\begin{aligned} \mathcal{PP} &\subseteq \mathcal{C}_{\text{unbound}}^{\text{optim}} \subseteq \mathcal{SA}, \\ \mathcal{PP} &\subseteq \mathcal{C}_{\text{unbound}}^{\text{pessim}} \subseteq \mathcal{SA}. \end{aligned}$$

The proof of Theorem 2.18 is based on the following lemma.

Lemma 2.19 (Bilevel formulation for semi-algebraic characteristic functions). *Consider a basic semi-algebraic set S (cf. Definition 2.1). There exists an optimistic (resp. pessimistic) polynomial bilevel formulation with a convex lower-level problem whose value function is the characteristic function $\mathbf{1}_S$.*

Proof. Let S be of the form:

$$S := \{P(x) = 0 \text{ and } Q_j(x) > 0, j \in J\} \subseteq \mathbb{R}^n, \quad (12)$$

where P and $Q_j, j \in J$ are polynomials. Consider the following bilevel problem:

$$\begin{aligned} \min_x \quad & F(x, w, z, t) := (1 - tP(x)) \prod_{j=1}^{|J|} (w_j z_j Q_j(x)) \\ \text{s.t.} \quad & (w, z, t) \in \arg \min_{w \in [0,1]^{|J|}, z \in \mathbb{R}^{|J|}, t \in \mathbb{R}} G(x, w, z, t) := (1 - tP(x))^2 + \sum_{j \in J} (1 - Q_j(x) z_j)^2 - Q_j(x) w_j \end{aligned}$$

where w_j, z_j indicate respectively the j th coordinate of the vectors w and z in $\mathbb{R}^{|J|}$.

By construction, F, G are polynomials. In addition, the lower-level problem is also clearly convex because given a fixed x , $G(x, \cdot, \cdot, \cdot)$ is linear w.r.t to w and (semi-definite positive) quadratic w.r.t z and t .

It remains to prove that the constructed bilevel problem has value function equal to $\mathbf{1}_S$. As we will see, given a fixed x , although there are multiple minimizers (w, z, t) , they all yield the same value $F(x, w, z, t)$. Therefore, this construction is valid for both optimistic and pessimistic formulations.

Due to the convexity and – more importantly – the separation of variables w, z, t , we can specify the optimal solution $(w^*(x), z^*(x), t^*(x))$ of $G(x, \cdot, \cdot, \cdot)$ for each fixed $x \in \mathbb{R}^n$ as follows:

$$w_j^*(x) = \begin{cases} 1 & \text{if } Q_j(x) > 0 \\ [0, 1] & \text{if } Q_j(x) = 0, j \in J, \\ 0 & \text{if } Q_j(x) < 0 \end{cases}$$

$$z_j^*(x) = \begin{cases} 1/Q_j(x) & \text{if } Q_j(x) \neq 0, j \in J, \\ \mathbb{R} & \text{otherwise} \end{cases}, \quad t^*(x) = \begin{cases} 1/P(x) & \text{if } P(x) \neq 0 \\ \mathbb{R} & \text{otherwise} \end{cases}.$$

Therefore,

$$w_j^*(x)z_j^*(x)Q_j(x) = \begin{cases} 1 & \text{if } Q_j(x) > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Similarly, we have:

$$1 - t^*(x)P(x) = \begin{cases} 1 & \text{if } P(x) = 0 \\ 0 & \text{otherwise} \end{cases}.$$

The result follows immediately from the two above equalities. \square

Proof of Theorem 2.18. Consider the piecewise polynomial $f = (S, (P_i)_{i=1}^N)$ as in Definition 2.17. Due to the definition of semi-algebraic sets (cf. Definition 2.1), for all $i = 1, \dots, N$, S_i can be written as:

$$S_i = \bigcup_{j \in J} T_{ij},$$

where $T_{ij}, i = 1, \dots, N, j \in J$ are basic pairwise disjoint semi-algebraic – use Proposition B.1 in Section B and the fact that the $S_i, 1 \leq i \leq N$ are disjoint. For each T_{ij} , we take two polynomials F_{ij}, G_{ij} (where G_{ij} is convex w.r.t y) and an unbounded box \mathcal{Y}_{ij} such that both optimistic and pessimistic versions of the following bilevel problem:

$$\begin{aligned} \min_x & F_{ij}(x, y(x)) \\ \text{s.t.} & y(x) \in \arg \min_{y \in \mathcal{Y}_{ij}} G_{ij}(x, y). \end{aligned}$$

are equal to the characteristic function $\mathbf{1}_{T_{ij}}$. Such polynomials and sets exist by the representability result for characteristic functions Lemma 2.19.

Consider the following polynomial bilevel problem:

$$\begin{aligned} \min_x & P(x, y) := \sum_{i=1}^N P_i(x) \left(\sum_{j \in J} F_{ij}(x, y_{ij}(x)) \right) \\ \text{s.t.} & y(x) := (y_{ij}(x))_{1 \leq i \leq N, j \in J} \in \arg \min_{y_{ij} \in \mathcal{Y}_{ij}} Q(x, y) := \sum_{i,j} G_{ij}(x, y_{ij}). \end{aligned}$$

Due to the separation of variables y_{ij} , we have:

$$\arg \min_{y_{ij} \in \mathcal{Y}_{ij}} \sum_{i,j} G_{ij}(x, y_{ij}) = \bigotimes_{i,j} \arg \min_{y_{ij}} G_{ij}(x, y_{ij}).$$

Therefore, by Lemma 2.19, we have for all $i \in I$:

$$\sum_{j \in J} F_{ij}(x, y_{ij}(x)) = \sum_{j \in J} \mathbf{1}_{T_{ij}} = \mathbf{1}_{S_i},$$

where the second equality holds because the S_{ij} are pairwise disjoint. Therefore,

$$P(x, y(x)) = \sum_{i=1}^N \mathbf{1}_{S_i}(x) P_i(x) = f(x)$$

as we desire. This argument is valid for both versions: optimistic and pessimistic. In addition, $Q(x, y)$ is convex since it is equal to the sum of convex functions. This concludes the proof. \square

Analogous to Theorem 2.18, we provide a class of functions that can be expressed using polynomial bilevel problems with convex lower-level problems and bounded box \mathcal{Y} . Similarly, the boundedness assumption on the lower-level imposes a semicontinuity restriction on the underlying value function class. Note that piecewise polynomial functions are bounded on compact sets by construction.

Theorem 2.20 (Semicontinuous piecewise polynomials are value functions with box constraints and convex lower-level). *Any lower semicontinuous (resp. upper semicontinuous) piece-wise polynomial function is the value function of an optimistic (resp. pessimistic) polynomial bilevel problem with lower-level objective convex over a bounded box, i.e.*

$$\begin{aligned} \mathcal{PP} \cap \mathcal{LSC} &\subseteq \mathcal{C}_{\text{bounded}}^{\text{optim}} \subseteq \mathcal{SA} \cap \mathcal{LSC} \cap \mathcal{CB}, \\ \mathcal{PP} \cap \mathcal{USC} &\subseteq \mathcal{C}_{\text{bounded}}^{\text{pessim}} \subseteq \mathcal{SA} \cap \mathcal{USC} \cap \mathcal{CB}. \end{aligned}$$

Proof. We only prove the theorem for the optimistic case. The other case can be dealt with analogously. In the following, we show that if $f \in \mathcal{PP} \cap \mathcal{LSC}$, then $f \in \mathcal{C}_{\text{bounded}}^{\text{optim}}$, which is the first inclusion. The second inclusion was already justified in Theorem 2.15.

Consider $f = (S, (P_i)_{i=1}^N)$ a lower semicontinuous piecewise polynomial function. Our construction is based on two observations:

1. Given a point $x \in \mathbb{R}^n$, define $\mathcal{I}(x) := \{i \in \mathcal{I} \mid x \in \overline{S_i}\}$ the subset of indices in which x belongs to the closure of corresponding sets. Since f is lower-semicontinuous, we have:

$$f(x) = \min_{i \in \mathcal{I}(x)} P_i(x). \quad (13)$$

2. Similar to the proof of Theorem 2.15, there exists a pair integer N and a positive constant $B > 0$ such that $\max_{i=1, \dots, N} P_i(x) \leq B + \|x\|^N$. Thus, in the following construction, WLOG, one can assume that $P_i(x) \leq 0, \forall i = 1, \dots, N, \forall x \in \mathbb{R}^n$. Otherwise, we can consider the function $r(x) = f(x) - B - \|x\|^N$, which is also an element of \mathcal{S} , lower-semicontinuous and all $r_i(x) = P_i(x) - B - \|x\|^N \leq 0$ for all x . If we can find functions P, Q and a bounded box so that the corresponding optimistic bilevel problem has value function equal to $r(x)$, then that of $(P + B + \|x\|^N, Q, \mathcal{Y})$ is equal to h .

For each $i = 1, \dots, N$, since S_i is semi-algebraic, so is its closure. By Proposition 2.4, they can be represented by:

$$\bar{S}_i = \bigcup_{j \in J} \bigcap_{k \in K} \{P_{jk}^i(x) \geq 0\}. \quad (14)$$

for some index sets J, K and polynomials P_{jk}^i . These polynomials will appear in our construction.

Similar to the proofs of other results, we introduce the sets of variables that will be used in our construction.

Name	Dimension	Coordinates notation	Type (Upper/Lower variable)	Feasible set
x	n	Not used	Upper	N/A
z	$ I J K $	z_{jk}^i	Lower	$[0, 1]^{ I \times J \times K }$
s	$ I J $	s_{ij}	Lower	$[0, 1]^{ I \times J }$

Table 3: Specification for the variables of the bilevel formulation.

Denote $z_j^i = (z_{jk}^i)_{k \in K}$, consider the optimization problem:

$$\text{Minimize}_{z_j^i \in [0, 1]^{|K|}} \quad G_j^i(x, z_j^i) = - \sum_{k \in K} P_{jk}^i(x) z_{jk}^i,$$

where P_{jk}^i are polynomials defined in (14). Given a fixed value of x , the optimal value of $(z^i)^*$ is given by:

$$(z_{jk}^i)^* = \begin{cases} 0 & \text{if } P_{jk}^i(x) < 0 \\ [0, 1] & \text{if } P_{jk}^i(x) = 0 \\ 1 & \text{if } P_{jk}^i(x) > 0 \end{cases}.$$

Therefore, we can conclude that:

$$\prod_{k \in K} (z_{jk}^i)^* = \begin{cases} 0 & \text{if } x \notin \bigcap_{k \in K} \{P_{jk}^i(x) \geq 0\} \\ [0, 1] & \text{if } x \in \bigcap_{k \in K} \{P_{jk}^i(x) \geq 0\} \text{ and } \exists k \in K, P_{jk}^i(x) = 0 \\ 1 & \text{if } x \in \bigcap_{k \in K} \{P_{jk}^i(x) \geq 0\} \text{ and } \forall k \in K, P_{jk}^i(x) > 0 \end{cases}. \quad (15)$$

We consider the following bilevel problem:

$$\begin{aligned} P &= \sum_{(i,j) \in I \times J} P_i(x) \left(s_{ij} \prod_{k \in K} z_{jk}^i \right), \\ Q &= \left(1 - \sum_{(i,j) \in I \times J} s_{ij} \right)^2 + \sum_{(i,j) \in I \times J} G_j^i(x, z_j^i), \\ s_{ij} &\in [0, 1], \forall (i, j) \in I \times J, \\ z_{jk}^i &\in [0, 1], \forall (i, j, k) \in I \times J \times K. \end{aligned} \quad (16)$$

The lower-level problem is obviously convex since $(1 - \sum_{(i,j) \in I \times J} s_{ij})^2$ is convex w.r.t s_{ij} , G_j^i is linear w.r.t z_j^i and the feasible set is a hypercube.

In particular, $P_i(x)s_{ij} \leq 0$ since $s_{ij} \in [0, 1]$ and $P_i(x)$ is assumed to be negative. Due to the optimistic nature of the bilevel problem and (15), we get:

$$P_i(x) \prod_{k \in K} z_{jk}^i = \begin{cases} 0 & \text{if } x \notin \bigcap_{k \in K} \{P_{jk}^i(x) \geq 0\} \\ P_i(x) & \text{if } x \in \bigcap_{k \in K} \{P_{jk}^i(x) \geq 0\} \end{cases}. \quad (17)$$

Moreover, any $(s_{ij})_{(i,j) \in |I| \times |J|} \in [0, 1]^{|I| \times |J|}$ such that $\sum_{i,j} s_{ij} = 1$ is optimal for Q . Since we consider the optimistic bilevel formulation, the value of P in (16) becomes:

$$\min_{s_{i,j} \geq 0, \sum s_{i,j} = 1} \sum_{(i,j) \in I \times J} s_{i,j} \left(P_i(x) \prod_{k \in K} z_{jk}^i \right).$$

Thus, the value of P will be equal to the smallest value of $P_i(x) \prod_{k \in K} z_{jk}^i$. This value will be equal to $f(x)$ because:

$$\begin{aligned} \min_{(i,j) \in I \times J} \left\{ P_i(x) \prod_{k \in K} z_{jk}^i \right\} &\stackrel{(17)}{=} \min_{(i,j) \in I \times J} \left\{ P_i(x) \mathbf{1}_{x \in \bigcap_{k \in K} \{P_{jk}^i(x) \geq 0\}} \right\} \\ &\stackrel{(14)}{=} \min_{i \in I} \{ P_i(x) \mathbf{1}_{x \in \bar{s}_i} \} \\ &\stackrel{(13)}{=} f(x). \end{aligned}$$

This concludes the proof. \square

2.3.1 Extension to arbitrary convex, compact, semi-algebraic lower-level constraints

Theorem 2.18 and Theorem 2.20 are limited to box lower-level constraint set \mathcal{Y} . While these constraints are explicit, this leaves open the question of the tightness of the corresponding inclusions. We will not answer this precise question here, but will consider a related question by allowing the lower-level constraint set \mathcal{Y} to be an arbitrary convex compact semi-algebraic set. We remark that for such a lower-level constraint set \mathcal{Y} , the rightmost inclusion in Theorem 2.20 is strict. For example the Euclidean norm can be expressed as a maximum over a ball and is not piecewise polynomial.

We will use the following shorthand to describe the corresponding class of value functions which admit the required bilevel representation:

$$\mathcal{C}_{cc}^{\text{mode}} := \{h \mid \exists P, Q \text{ polynomials, } Q \text{ convex w.r.t. } y, \mathcal{Y} \text{ compact, convex and semi-algebraic s.t. } \varphi_{\text{mode}} = h\} \quad (18)$$

where cc stands for ‘‘compact convex’’. Using Theorem 2.15, we get an immediate relation:

$$\begin{aligned} \mathcal{C}_{\text{bounded}}^{\text{optim}} &\subseteq \mathcal{C}_{cc}^{\text{optim}} \subseteq \mathcal{SA} \cap \mathcal{LSC} \cap \mathcal{CB}, \\ \mathcal{C}_{\text{bounded}}^{\text{pessim}} &\subseteq \mathcal{C}_{cc}^{\text{pessim}} \subseteq \mathcal{SA} \cap \mathcal{USC} \cap \mathcal{CB}. \end{aligned}$$

since bounded boxes are compact, convex, and semi-algebraic. Again, it is natural to investigate whether these inclusions strict. Our following result show that the second inclusion (in the above equations) is actually an equality.

Theorem 2.21 (Value functions of polynomial bilevel programming with convex, compact and semi-algebraic lower-level constraints). *Any semi-algebraic, lower (resp. upper) semicontinuous function which is bounded on compact sets is the value function of an optimistic (resp. pessimistic) polynomial bilevel problem with lower objective convex over a compact, convex and semi-algebraic set. In particular,*

$$\begin{aligned}\mathcal{C}_{cc}^{\text{optim}} &= \mathcal{SA} \cap \mathcal{LSC} \cap \mathcal{CB}, \\ \mathcal{C}_{cc}^{\text{pessim}} &= \mathcal{SA} \cap \mathcal{USC} \cap \mathcal{CB}.\end{aligned}$$

Proof. This proof is based on the following fact: one can equivalently reformulate the lower-level polynomial optimization problem in the proof of Theorem 2.15, by a convex optimization problem with a compact, convex, and semi-algebraic feasible set. This reformulation leads to a modified argmin correspondence in the lower level which does not change the value function overall.

Let us provide details: consider the following optimization problem:

$$\underset{\omega \in \Omega}{\text{Minimize}} \quad F(\omega),$$

where $\Omega \subseteq \mathbb{R}^n$ is a compact, semi-algebraic set and F is a polynomial. Let $d = \deg(F)$ be the highest degrees of a monomial (a product of powers of variables with nonnegative integer exponents $x_1^{d_1} \dots x_n^{d_n}$, $d_i \in \mathbb{N}, \forall 1 \leq i \leq n$) of F . Consider the function $M_d : \Omega \rightarrow \mathbb{R}^{K_d}$ that maps a point $\omega \in \mathbb{R}^n$ to the vector of monomials up to degree d (the constant $K_d = \binom{n+d}{n}$ is the number of such monomials). Since F is a polynomial, it can be written as: $F(\omega) = c^\top M_d(\omega)$ for some vector $c \in \mathbb{R}^{K_d}$, i.e., $F(x)$ is linear w.r.t $M_d(\omega)$. Thus, the original polynomial optimization problem can be written equivalently as:

$$\text{Minimize} \quad c^\top \lambda \quad \text{such that} \quad \lambda \in \text{Conv}(M_d(\Omega)),$$

where $M_d(\Omega) \subseteq \mathbb{R}^{K_d}$ is the image of Ω via the map M_d and $\text{Conv}(\cdot)$ is the convex hull of a set. Since Ω is compact, so are $M_d(\Omega)$ and its convex hull. The semi-algebraicity can be argued similarly (using Carathéodory's theorem for convex hull [51, Theorem 0.0.1] and Theorem 2.3). Thus, this new formulation has a linear objective function with a compact, convex, and semi-algebraic feasible set.

Now, we will plug this reformulation into the construction in the proof of Theorem 2.15. Again, we will only treat the optimistic version (the pessimistic version can be treated similarly). Consider $f \in \mathcal{SA} \cap \mathcal{LSC} \cap \mathcal{CB}$. Using the construction in (10), we remind readers that there exists two functions $P(x, t, z)$ and $Q(x, t, z)$ (x is the upper-level variable, t, z are lower-level variables) and a bounded box \mathcal{Y} such that:

1. The optimistic value function of the associated bilevel optimization problem equals f .
2. The function $P(x, t, z) = t$.
3. For all x , the minimum value of y such that there exists z satisfying $(t, z) \in \arg \min Q(x, \cdot, \cdot)$ is $f(x)$.

Using this information, we can construct a new bilevel optimization problem as follows: instead of using t, z as lower-level variables, we will use λ , a (vector-valued) variable representing all the monomials of the concatenation (t, z) up to the degree of interest. We write $\lambda[t]$, for the coordinate

entry of λ corresponding to the degree-one monomial t . The upper-level, lower-level functions, and the lower-level feasible set of the new bilevel problem are respectively given by:

$$\begin{aligned} P'(x, \lambda) &= \lambda[t], \\ Q'(x, \lambda) &= c_Q^\top \lambda \\ \mathcal{Y}' &= \text{Conv}(M_d(\mathcal{Y})), \end{aligned}$$

where c_Q is the vector containing the coefficients of the monomials of Q . For a given x , consider λ^* is an element of $\text{Conv}(M_d(\mathcal{Y}))$ that satisfies:

1. $\lambda^* \in \arg \min_{\lambda \in \mathcal{Y}'} Q'(x, \lambda)$.
2. $\lambda[t]^*$ attains the minimum value among elements in $\arg \min_{\lambda \in \mathcal{Y}'} Q'(x, \lambda)$ (since $P'(x, \lambda) = \lambda[t]$ and we are considering optimistic bilevel problem).

Again, by Carathéodory's theorem [51, Theorem 0.0.1], an element of \mathcal{Y}' must be written as a convex combination of at most $C = \dim(\lambda) + 1$ ($\dim(\lambda)$ is the dimension of λ) elements of $M_d(\mathcal{Y})$. Therefore, there exists $\lambda_1, \dots, \lambda_C \in M_d(\mathcal{Y})$ and $c_1, \dots, c_C \geq 0, \sum_{i=1}^C c_i = 1$ such that:

$$\lambda^* = \sum_{i=1}^C c_i \lambda_i.$$

By the linearity of Q' and P' with respect to λ , we can conclude that for all $i = 1, \dots, C$, $\lambda_i \in \arg \min_{\lambda \in \mathcal{Y}'} Q'(x, \lambda)$ and $\lambda_i[t]$ also attains minimum value among the elements of $\arg \min Q'(x, \cdot)$. Since $\lambda_i \in M_d(\mathcal{Y})$, $\lambda_i = M_d((t_i, z_i))$ where $(t_i, z_i) \in \arg \min_{(t,z) \in \mathcal{Y}} Q(x, t, z)$ and t_i is the smallest possible such value. It implies that $t_i = f(x), \forall i = 1, \dots, C$. Hence,

$$P'(x, \lambda^*) = \lambda^*[t] = \sum_{i=1}^C c_i \lambda_i[t] = \left(\sum_{i=1}^C c_i \right) f(x) = f(x).$$

This concludes the proof. □

Let us emphasize that the constraint set resulting from the proof of this result does not have an explicit construction and only allows for a looser control of the dimensionality, unlike previous results. Indeed, although \mathcal{Y} is convex, compact and semi-algebraic, we do not know how to *explicitly and efficiently* represent it (using polynomial equalities and inequalities). Such representation is important in polynomial optimization [6] and closely related to (but not quite the same as) the SOS relaxation [33, 44]. We refer readers to [6] for a more dedicated discussion. Note that the same idea (if one allows \mathcal{Y} to be an arbitrary convex, closed semi-algebraic set) does not work for unbounded cases because the convex hull of a closed set is not necessarily closed. We did not find a way around this issue, and leave this question for future work. We also leave open the possibility of obtaining similar results for simple lower-level constraints set \mathcal{Y} such as balls or boxes.

2.4 Summary of the results

Table 4 summarizes below the various representation results mentioned above.

Type of bilevel problem		Expressivity result	
No lower-level convexity			
<i>Optimistic / Pessimistic</i>	<i>Lower-level constraints</i>	<i>Expressible functions</i>	<i>Value function class</i>
Either	unbounded box	(SA)	=
Optimistic	bounded box	(SA) + (LSC) + (CB)	=
Pessimistic	bounded box	(SA) + (USC) + (CB)	=
Convex lower level			
<i>Optimistic / Pessimistic</i>	<i>Lower-level constraints</i>	<i>Expressible functions</i>	<i>Value function class</i>
Either	unbounded box	(PP)	\subset
Optimistic	bounded box	(PP) + (LSC)	\subset
Pessimistic	bounded box	(PP) + (USC)	\subset
Pessimistic	convex, compact, semi-algebraic	(SA) + (USC) + (CB)	=
Optimistic	convex, compact, semi-algebraic	(SA) + (LSC) + (CB)	=

Table 4: Summary of all expressivity results. Abbreviations (**PP**): *piecewise polynomial*, (**SA**): *semi-algebraic*, (**LSC**): *lower-semicontinuous*, (**USC**): *upper-semicontinuous*, (**CB**): *bounded on every compact*. All our results state that the class of expressible functions (third column) is contained in the class of value functions corresponding to the considered type of bilevel problem. The last column indicates whether this inclusion is actually an equality or not.

3 Computational hardness of polynomial bilevel optimization

3.1 Preliminaries on the polynomial hierarchy

This section reminds readers of the polynomial hierarchy, a classification of problems based on their “hardness” in computational complexity. We also discuss the subset sum interval problem – a classical Σ_2^P -hard problem that will play an important role in our analysis.

Polynomial hierarchy

A classical definition of **P** and **NP**, two important concepts of complexity theory, is based on the Turing machine models: **P** and **NP** is the set of *decision* problems that can be solved in polynomial time using deterministic and non-deterministic Turing machines, respectively. Alternatively, one can define a problem belonging to **NP** if for any instance whose answer is *yes*, there exists a proof verifiable in polynomial time (using a deterministic Turing machine). In other words, verifying a *yes* instance of a **NP** problem is a **P** problem. The negated class of **NP** - **coNP** - contains those whose instances with *no* answer can be verifiable in polynomial time. Thus, different from the definitions based on the non-deterministic Turing machine, one can define two classes **NP** and **coNP** based on **P**.

A natural generalization of this approach gives us the polynomial hierarchy. Following [48, Theorem 3.1], one can define the complexity class Σ_k^P as the set of decision problems that can be written in the form:

$$(\exists y_1), (\forall y_2), (\exists y_3), \dots, (Q_k y_k), R(x, y_1, \dots, y_k),$$

where the quantifiers $Q_\ell \in \{\exists, \forall\}$, $1 \leq \ell \leq k$ alternate and R is a boolean formula that can be evaluated in polynomial time (or equivalently, it is a problem in \mathbf{P}). In particular, $\Sigma_1^p = \mathbf{NP}$.

Analogously, one can exchange the role of \exists and \forall in the definition of the class Σ_k^p to define the class Π_k^p as:

$$(\forall y_1), (\exists y_2), (\forall y_3), \dots, (Q_k y_k), R(x, y_1, \dots, y_k).$$

Similarly, we also have $\Pi_1^p = \mathbf{coNP}$. By convention, $\mathbf{P} = \Sigma_0^p = \Pi_0^p$. Thus, similar to the relation $\mathbf{P} \subseteq \mathbf{NP} \cap \mathbf{coNP}$, we have the generalized version:

$$\Sigma_k^p \subseteq \Sigma_{k+1}^p \cap \Pi_{k+1}^p \quad \text{and} \quad \Pi_k^p \subseteq \Sigma_{k+1}^p \cap \Pi_{k+1}^p.$$

For any $k \in \mathbb{N}$, it remains unknown whether $\Sigma_k^p \neq \Sigma_{k+1}^p$ or $\Sigma_k^p = \Sigma_{k+1}^p$, and similarly for Π_k^p and Π_{k+1}^p . Furthermore, if there is equality for a given k_0 , then there is equality for all $k \geq k_0$ [5, Theorem 5.4]. This is called the *collapse* of polynomial hierarchy at the k_0 -th level, a possibility which is considered unlikely and is often used as an assumption in complexity theoretic proofs, see discussions in [5, Chapter 5]). Readers can view an illustration of the polynomial hierarchy and the relations between their components in Figure 1.

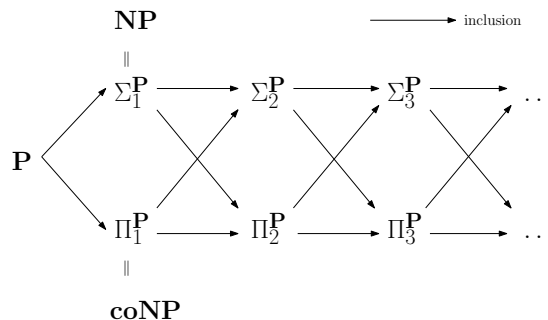


Figure 1: Diagram describes the polynomial hierarchy. Arrows represent the inclusion relation.

A famous tool to study the relation between \mathbf{P} and \mathbf{NP} is the notion of \mathbf{NP} -hardness. A problem \mathcal{A} is called \mathbf{NP} -hard if for any problem \mathcal{B} in class \mathbf{NP} , there exists a transformation (also known as *reduction*) running in polynomial time that turns an instance of \mathcal{B} into that of \mathcal{A} and both share the same answer (*yes* and *no*). As such, if one comes up with a polynomial algorithm for \mathcal{A} , then he can solve all \mathbf{NP} problems in polynomial time. On the other hand, to prove that a problem \mathcal{A} is \mathbf{NP} -hard, it is sufficient to construct a polynomial reduction from any \mathbf{NP} -hard problem to \mathcal{A} .

Analogously, we can define Σ_k^p/Π_k^p -hardness using the same principle for any $k \in \mathbb{N}$. We will use the concept of Σ_2^p hardness. If a problem is Σ_2^p -hard and the polynomial hierarchy does not collapse at the first level, i.e. $\mathbf{NP} \neq \Sigma_2^p$, then the intrinsic computational complexity of this problem is strictly higher than any problem in the class $\mathbf{NP} \cup \mathbf{coNP}$. In the following, we will show that polynomial bilevel optimization is Σ_2^p -hard. This highlights the possibility that bilevel polynomial optimization is actually much harder than problems in \mathbf{NP} which are themselves computationally hard. Note that it is a common assumption in computational complexity theory that the polynomial hierarchy does not collapse at all, and specially not at the first level (see discussions in [5, Section 5.2.1]).

The subset sum interval problem

To show the Σ_2^p -hardness of polynomial bilevel optimization, we need to perform reduction to a known Σ_2^p -hard problem. We choose the *subset sum interval problem* [17], defined as follows:

Problem 3.1 (Subset sum interval problem). Given a finite number of positive integers q_1, \dots, q_k , and two positive integers R and r with $r \leq k$, decide whether there exists an integer $R \leq S \leq R + 2^r - 1$ such that none of the subsets $I \subseteq \{1, \dots, k\}$ satisfies $\sum_{i \in I} q_i = S$.

If $r > k$, the problem is trivial because there are at most 2^k possible values of $\sum_{i \in I} q_i$ but there are already 2^r possible values for S . Using logic notations, Problem 3.1 is equivalent to deciding the correctness of this first-order logic formula:

$$\exists S, \forall I \subseteq \{1, \dots, k\}, R \leq S \leq R + 2^r - 1 \text{ and } S \neq \sum_{i \in I} q_i.$$

The Σ_2^p -hardness of this problem follows from [25], see Section C for more details.

3.2 Polynomial bilevel optimization is Σ_2^p -hard

The main result of this section is to prove the Σ_2^p -hardness of the decision version of the optimistic polynomial bilevel.

Theorem 3.2 (Hardness of polynomial bilevel optimization). *Given two polynomials P and Q of degree at most five, a bounded box lower-level feasible set \mathcal{Y} and a constant c , deciding whether the optimal value of the corresponding optimistic bilevel problem of (P, Q, \mathcal{Y}) is strictly smaller than c is Σ_2^p -hard.*

Theorem 3.2 is an immediate result of the following lemma.

Lemma 3.3. *Given an instance of Problem 3.1, there exist two polynomials of degree five P and Q whose coefficients are integers in the interval $[-2(M^2+1), 2(M^2+1)]$ where $M = \max(\max_{i=1}^k q_i, 2^{r-1}, R)$ and variable (x, y) , where $x \in \mathbb{R}^r, y \in \mathbb{R}^{k+1}, \mathcal{Y} = [0, 1]^{k+1}$, such that the optimal value of (BP-O) is strictly smaller than zero if the answer for the instance of Problem 3.1 is YES, and exactly equal to zero if the answer is NO.*

Before proving the lemma, we argue that the “description” of the polynomials P, Q in Lemma 3.3 is at most polynomial w.r.t. the size of the inputs of Problem 3.1. Indeed, the polynomials P, Q have degree five and all coefficients can be expressed by $\log M$ bits. We remark that the representation of each coefficient might use $O(r)$ bits, from the bound 2^{r-1} , which is exponential w.r.t. the number of bits representing the number r itself. In addition, we have polynomials of r variables, which represent a number of coefficient polynomial in r . This dependency in r is still polynomial overall because any bit representation of Problem 3.1 must use at least $k \geq r$ bits to represent to numbers $q_i, i = 1, \dots, k$. Regarding the other quantities appearing in the definition of M , it is clear that an integer smaller than $\max_{i=1}^k q_i$ has a bit representation of size bounded by that of the collection q_1, \dots, q_k and similarly for R .

Proof of Lemma 3.3. We will use $x \in \mathbb{R}^r$ and $y = (z, t), z \in \mathbb{R}^r, t \in \mathbb{R}^k$ (which implies $y \in \mathbb{R}^{r+k}$) to indicate the upper-level and lower-level variables respectively. We also use $x_i, z_i, t_i \in \mathbb{R}$ to indicate

the i th coordinate of the variables x, z, t respectively. For $x \in \mathbb{R}^r$, define:

$$F(x) := R + \sum_{j=1}^r 2^{j-1} x_j,$$

a linear combination (thus, a polynomial) of x . Intuitively, $F(x)$ is the binary representation of a number in the range $[R, R + 2^r - 1]$ (if all variables x_j are binary).

Moreover, for $t \in \mathbb{R}^k$, we also define:

$$G(t) := \sum_{i=1}^k t_i q_i,$$

where q_i is the integers that appear in the given instance of Problem 3.1. Intuitively, $G(t)$ is equal to the sum of some subsets of $\{q_i, i = 1, \dots, k\}$ (if all variables t_i are binary). Finally, we define:

$$H(x, t) = \left(\sum_{j=1}^k (t_j(1 - t_j))^2 \right) + (F(x) - G(t))^2$$

Note that $H(x, t) = 0$ if and only if t is a binary vector and $F(x) = G(t)$. This mimics the situation where a number $F(x) \in [R, R + 2^r - 1]$ can be written as a subset-sum of the array q .

The polynomials P, Q and the lower-level feasible set can be constructed as follows: We constrain the lower-level variable $y = (z, t) \in \mathcal{Y} = [0, 1]^{k+1}$, or equivalently, $z \in [0, 1], t \in [0, 1]^k$. We define P and Q as follows:

$$\begin{aligned} P &= 1 - z(H(x, t) + 1), \\ Q &= z \underbrace{\left[(z - 1)^2 + \sum_{i=1}^r ((1 - x_i)x_i)^2 \right]}_{L(x, z)} + H(x, t). \end{aligned} \tag{19}$$

By construction, $\deg(P) = \deg(Q) = 5$ as stated. In addition, it can be shown that the coefficients of monomials of P and Q have their absolute values bounded by $2(M^2 + 1)$. Indeed, all the coefficients of $L(x, z)$ in (19) belong to $\{0, 1, -2\}$. A direct calculation also shows that absolute values of the coefficients of $H(x, t)$ are bounded by $2(M^2 + 1)$. More importantly, these two polynomials do not have any common monomial (H does not have variable z but every monomial of L has at least one z). Thus, we can conclude that the coefficients of P and Q lie in the interval $[-2(M^2 + 1), 2(M^2 + 1)]$.

Due to the separation of z and t , given a fixed x , we have:

$$\arg \min Q(x, \cdot, \cdot) = \arg \min L(x, \cdot) \times \arg \min H(x, \cdot). \tag{20}$$

where \times is the Cartesian product between two sets. The key idea in this construction is based on a sequence of observations.

Observation 1 With the construction of P and Q as in (19), we have:

1. If $x \notin \{0, 1\}^r$, $\varphi_o(x) = 1$.
2. If $x \in \{0, 1\}^r$, $\varphi_o(x) \leq 0$.

Indeed, we have: $L(x, z) \geq 0$ because $z \in [0, 1]$ and the other factor is a sum of squares. Thus, $z = 0$ is always a minimizer of $L(x, \cdot)$ on $[0, 1]$. Another possibility to attain the global minimum value of $L(x, z)$ is to have $x \in \{0, 1\}^r$ and $z = 1$. Therefore,

$$z(x) := \arg \min_z L(x, z) = \begin{cases} 0 & \text{if } x \notin \{0, 1\}^r \\ \{0, 1\} & \text{otherwise} \end{cases} \implies \varphi_o(x) = \begin{cases} 1 & \text{if } x \notin \{0, 1\}^r \\ -H(x, t(x)) & \text{otherwise} \end{cases}. \quad (21)$$

where $(z(x), t(x)) \in \arg \min Q(x, \cdot, \cdot)$. Since H is a sum of squares, we proved the first observation.

To finish the proof, we need a second observation.

Observation 2 With the construction of P and Q as in (19), we have:

1. **First case:** If there exists an integer $S \in [R, R + 2^r - 1]$ such that it is not equal to the sum of any subset of $\{q_i, i = 1, \dots, k\}$, then $\inf_x \varphi_o(x) < 0$.
2. **Second case:** Otherwise, $\varphi_o(x) = 0, \forall x \in \{0, 1\}^r$. Hence, $\min_x \varphi_o(x) = 0$.

Due to the previous observation, it is sufficient to consider only binary inputs x . In that case, $\varphi_o(x) \in -H(x, t(x)) = -\min_{t \in [0, 1]^k} H(x, t)$ (cf. Equation (21)). In addition, $F(x)$ is an integer belonging to the interval $[R, R + 2^r - 1]$. We consider two cases one by one:

1. **First case:** We choose $x \in \{0, 1\}^r$ such that $S = F(x)$. Moreover, due to the property of S , $H(x, t) > 0$ because $H(x, t) = 0$ if and only if t is a binary vector and $F(x) = G(t)$. However, that will be equivalent to the statement that S equals the sum of some subsets of $\{q_i, i = 1, \dots, k\}$, a contradiction. Therefore, $\varphi_o(x) = -\min_{t \in [0, 1]^k} H(x, t) < 0$.
2. **Second case:** In this case, for any $R \leq S = F(x) \leq R + 2^r - 1$, there exists a binary vector t such that $F(x) = G(t)$. Thus, in the lower-level problem, when we minimize $H(x, t)$, we will get a binary vector $t(x)$ such that $H(x, t(x)) = 0$. It allows us to conclude that $\varphi_o(x) = -H(x, t(x)) = 0, \forall x \in \{0, 1\}^r$.

Combining two cases yields the proof. □

For interested readers, the whole proof of Lemma 3.3 is to find two polynomials whose corresponding optimistic bilevel problem equals to the lower-semicontinuous function φ_o in (21). Thus, our construction is similar to the proof of Theorem 2.15, with some simplification adapted to the structure of φ_o to minimize the degrees of P, Q . This proof cannot be extended for the pessimistic version since a pessimistic bilevel problem with a bounded box lower-level constraint can only be equivalent to an upper semicontinuous function. Finally, one might wonder if we can construct a “difficult” instance with a convex lower-level problem. Nevertheless, our technique in Theorem 2.21 cannot be applied since the reduction is not guaranteed to be polynomial.

Appendix A The difficulty of smooth bilevel optimization

In this section, we provide the proof of Theorem 1.1. The proof uses the Whitney representation of closed set [32, Section 3.10, 2] [34, Theorem 2.29].

Theorem A.1 (Whitney representation of closed sets). *Any closed set of \mathbb{R}^n is the set zeros of a smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$.*

We remind readers of an important property of lower (resp. upper) semicontinuous functions.

Proposition A.2. *A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is lower (resp. upper) semicontinuous if and only if its epigraph (resp. hypograph) is closed where epigraph and hypograph of f are defined respectively as:*

$$\begin{aligned} \text{epi}(f) &:= \{(x, \alpha) \mid f(x) \leq \alpha\} \subseteq \mathbb{R}^{n+1} \\ \text{hyp}(f) &:= \{(x, \alpha) \mid f(x) \geq \alpha\} \subseteq \mathbb{R}^{n+1}. \end{aligned}$$

Proof of Theorem 1.1. In this proof, we deal with the lower semicontinuous case. The other one can be proved similarly.

If f is lower semicontinuous, the epigraph of f – $\text{epi}(f)$ – is closed. Using Theorem A.1, there exists a smooth function $h : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ such that its zeros set equals $\text{epi}(f)$.

Since the domain $\text{dom}(f) := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$ is closed (due to our assumption), by Theorem A.1, there exists $g : \mathbb{R}^n \rightarrow \mathbb{R}$ such that its zeros equals $\text{dom}(f)$.

Consider the upper-level and lower-level variables $x \in \mathbb{R}^n$ and $y = (y_1, y_2, y_3) \in \mathbb{R}^3$, define:

$$\begin{aligned} P &:= y_1 \\ Q &:= h(x, y_1)^2 + \underbrace{(g(x)y_2)^2 + (1 - y_2y_3)^2}_{p(x, y_2, y_3)}. \end{aligned}$$

We argue that with this choice of P, Q , the function φ_o of the optimistic version equals f . Indeed, due to the separation of y_1 and (y_2, y_3) , we have:

$$\arg \min_y Q = \arg \min_{y_1} h(x, \cdot) \times \arg \min_{y_2, y_3} p(x, \cdot, \cdot).$$

Consider two cases:

1. If $x \in \text{dom}(f)$: $\Theta(x) = \arg \min Q(x, \cdot)$ is the set of y such that $h(x, y_1) = 0$ and $y_2y_3 = 1$. Therefore, $\Theta(x) = \text{epi}(f) \cap \{x\} \times \mathbb{R}$. Since we deal with the optimistic version, $\varphi_o(x) = \min\{y \mid y \in \text{epi}(f) \cap \{x\} \times \mathbb{R}\} = f(x)$ by definition of epigraph. This concludes the proof.
2. If $x \notin \text{dom}(f)$, then $g(x) \neq 0$. We claim that the set $\arg \min_{y_2, y_3} p(x, \cdot, \cdot)$ is empty and that will conclude the proof. Indeed, the infimum of p is zero by taking $y_2 \rightarrow 0$ and $y_3 = 1/y_2$. However, this infimum cannot be attained since both squares cannot equal zero simultaneously. That concludes the proof.

□

Appendix B An alternative definition of semi-algebraic sets

Proposition B.1. *Any semi-algebraic set $S \in \mathbb{R}^n$ can be written as a finite union of disjoint basic semi-algebraic sets.*

Proof. By definition of semi-algebraic sets (cf. Definition 2.1), S has the following form:

$$S = \bigcup_{i \in I} S_i \quad \text{where} \quad S_i := \{x \mid P_i(x) = 0 \text{ and } Q_{ij}(x) > 0, j \in J\}. \quad (22)$$

(We assume the same J for all $i \in I$ for simplicity, but one can increase the number of inequalities in the definition of each S_i to make this assumption valid).

To this end, let $\mathcal{H} := \{P_i, i = 1 \in I\} \cup \{Q_{ij}, (i, j) \in I \times J\}$ the set of all polynomials appearing in (22). It is noteworthy that \mathcal{H} is finite, i.e., $|\mathcal{H}| < +\infty$, thus we can write $\mathcal{H} = \{h_1, \dots, h_{|\mathcal{H}|}\}$ where h_i are polynomials. Consider the following collections of $3^{|\mathcal{H}|}$ basic semi-algebraic sets, indexed by $\mathcal{I} \in \{<, >, =\}^{|\mathcal{H}|}$, and defined as:

$$\mathcal{T}_{\mathcal{I}} := \{x \mid h_i \Delta_i 0, i = 1, \dots, \mathcal{H}\},$$

where Δ_i can receive three possible values $\{>, <, =\}$, encoded by the index \mathcal{I} . These sets are semi-algebraic and they are inherently disjoint. Moreover, for each $\mathcal{T}_{\mathcal{I}}$ is either disjoint and included in S_i for $i \in I$. Finally, the union of these sets equals to \mathbb{R}^n . Therefore, there must exist a subset of indices $I \subseteq \{<, >, =\}^{|\mathcal{H}|}$ such that:

$$S = \bigcup_{\mathcal{I} \in I} \mathcal{T}_{\mathcal{I}}.$$

This concludes the proof. □

Appendix C Σ_2^p -hardness of the subset sum interval problem

The Σ_2^p -hardness of the subset sum interval problem is mentioned in [17], and was credited to [25]. However, in [25], the authors did not prove that the subset sum interval problem is Σ_2^p -hard. Instead, they proved that a problem called *captive queen* is Π_2^p -hard. The goal of this section is to clarify the link the work of [25] and the subset sum interval problem: an intermediate result in [25] imply its Σ_2^p -hardness. But this implication is non-trivial and we describe the arguments here for self-containedness of our presentation. We do *not* claim any scientific contribution as these results are due to [25, 17]. We thank Carvalho M. ⁴ for insightful discussions on this question.

We start with the quantified 3-CNF-SAT problem: given a boolean formula $\phi(X, Y)$ over variables X and Y , in conjunctive normal form with each clause with exactly three literals, decide if $\forall X, \exists Y, \phi(X, Y)$ is true. This is a Π_2^p -complete problem [5, Example 5.9].

Given such a ϕ , [25, Lemma 3.3 and Lemma 3.4] describe the construction of a polynomial time reduction to an instance of the so-called *captive queen problem* [25] of the following forms: there exists k integers q_1, \dots, q_k and two positive integers $R, r \leq k$ such that $\forall X, \exists Y, \phi(X, Y)$ holds, if and only if

$$\forall S, \exists I \subseteq \{1, \dots, k\}, R \leq S \leq R + 2^r - 1 \text{ and } S = \sum_{i \in I} q_i.$$

This is the negation of the truth value of the subset sum interval problem. Since the 3-CNF-SAT problem is Π_2^p complete, we have that negation of the subset sum interval problem is Π_2^p -complete, and the problem itself Σ_2^p -complete [5, Section 5.1].

Acknowledgements

JB, TL, EP thank AI Interdisciplinary Institute ANITI funding, through the French “Investments for the Future – PIA3” program under the grant agreement ANR-19-PI3A0004, Air Force Office of

⁴<https://margaridacarvalho.org/>

Scientific Research, Air Force Material Command, USAF, under grant numbers FA8655-22-1-7012. JB, EP and SV acknowledge support from ANR MAD. JB and EP thank TSE-P and acknowledge support from ANR Chess, grant ANR-17-EURE-0010, ANR Regulia. EP acknowledges support from IUF and ANR Bonsai, grant ANR-23-CE23-0012-01.

References

- [1] Pierre Ablin, Gabriel Peyré, and Thomas Moreau. Super-efficiency of automatic differentiation for functions defined as a minimum. In *Proceedings of the 37th International Conference on Machine Learning, ICML'20*, virtual, 2020. JMLR.org.
- [2] Abdelmalek Aboussoror and Pierre Loridan. Strong-weak stackelberg problems in finite dimensional spaces. *Serdica. Mathematical Journal*, 21:151–170, 1995.
- [3] Michael Arbel and Julien Mairal. Amortized implicit differentiation for stochastic bilevel optimization. In *International Conference on Learning Representations*, Virtual, 2022.
- [4] Michael Arbel and Julien Mairal. Non-Convex Bilevel Games with Critical Point Selection Maps. In *NeurIPS 2022 - 36th Conference on Neural Information Processing Systems*, Advances in Neural Information Processing Systems (NeurIPS) 2022, pages 1–34, New Orleans, United States, November 2022.
- [5] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, Cambridge, 2009.
- [6] Francis Bach and Alessandro Rudi. Exponential convergence of sum-of-squares hierarchies for trigonometric polynomials. *SIAM Journal on Optimization*, 33(3):2137–2159, 2023.
- [7] Jonathan Bard. Some properties of the bilevel programming problem. *Journal of Optimization Theory and Applications*, 68:371–378, 02 1991.
- [8] Robert Gardner Bartle and Donald R Sherbert. *Introduction to real analysis*. Wiley,, New Delhi, 2011.
- [9] Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in Real Algebraic Geometry (Algorithms and Computation in Mathematics)*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [10] Yasmine Beck, Ivana Ljubić, and Martin Schmidt. A survey on bilevel optimization under uncertainty. *European Journal of Operational Research*, 311(2):401–426, 2023.
- [11] Yasmine Beck and Martin Schmidt. A robust approach for modeling limited observability in bilevel optimization. *Operations Research Letters*, 49(5):752–758, 2021.
- [12] Omar Ben-Ayed and E. Blair. Computational difficulties of bilevel linear programming. *Operations Research*, 38:556–560, 06 1990.
- [13] R. Benedetti and J.J. Risler. *Real Algebraic and Semi-algebraic Sets*. Actuelles mathématiques. Hermann, Paris, 1990.

- [14] C. Berge. Topological spaces. *Proceedings of the Edinburgh Mathematical Society*, 13, 1963.
- [15] Mathieu Blondel, Quentin Berthet, Marco Cuturi, Roy Frostig, Stephan Hoyer, Felipe Llinares-López, Fabian Pedregosa, and Jean-Philippe Vert. Efficient and modular implicit differentiation. *ArXiv*, abs/2105.15183, 2021.
- [16] Jérôme Bolte, Tam Le, Edouard Pauwels, and Antonio Silvetti-Falls. Nonsmooth Implicit Differentiation for Machine Learning and Optimization. In *Advances in Neural Information Processing Systems*, Online, France, December 2021.
- [17] Alberto Caprara, Margarida Carvalho, Andrea Lodi, and Gerhard J. Woeginger. A study on the computational complexity of the bilevel knapsack problem. *SIAM Journal on Optimization*, 24(2):823–838, 2014.
- [18] Martina Cerulli. *Bilevel optimization and applications*. Theses, Institut Polytechnique de Paris, December 2021.
- [19] Y. Chen and M. Florian. The nonlinear bilevel programming problem:formulations,regularity and optimality conditions. *Optimization*, 32(3):193–209, 1995.
- [20] Mathieu Dagréou, Pierre Ablin, Samuel Vaiter, and Thomas Moreau. A framework for bilevel optimization that enables stochastic and global variance reduction algorithms. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [21] Mathieu Dagréou, Thomas Moreau, Samuel Vaiter, and Pierre Ablin. A Lower Bound and a Near-Optimal Algorithm for Bilevel Empirical Risk Minimization. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2024.
- [22] Aris Daniilidis and Dmitriy Drusvyatskiy. Pathological subgradient dynamics. *SIAM Journal on Optimization*, 30(2):1327–1338, 2020.
- [23] S. Dempe. *Foundations of Bilevel Programming*. Nonconvex Optimization and Its Applications. Springer US, US, 2002.
- [24] S. Dempe, B. S. Mordukhovich, and A. B. Zemkoho. Sensitivity analysis for two-level value functions with applications to bilevel programming. *SIAM Journal on Optimization*, 22(4):1309–1343, 2012.
- [25] Christian E.J. Eggermont and Gerhard J. Woeginger. Motion planning with pulley, rope, and baskets. In Thomas Wilke Christoph Dürr, editor, *STACS’12 (29th Symposium on Theoretical Aspects of Computer Science)*, volume 14, pages 374–383, Paris, France, 2012. LIPIcs.
- [26] Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazi, and Massimiliano Pontil. Bilevel programming for hyperparameter optimization and meta-learning. In *International Conference on Machine Learning*, 2018.
- [27] Stephen Gould, Basura Fernando, Anoop Cherian, Peter Anderson, Rodrigo Santa Cruz, and Edison Guo. On differentiating parameterized argmin and argmax problems with application to bi-level optimization. *CoRR*, abs/1607.05447, 2016.
- [28] René Henrion and Thomas Surowiec. On calmness conditions in convex bilevel programming. *Applicable Analysis*, 90:951–970, 06 2011.

- [29] R. Jeroslow. The polynomial hierarchy and a simple model for competitive analysis. *Mathematical programming*, 32:146–164, 1985.
- [30] V. Jeyakumar, J. B. Lasserre, G. Li, and T. S. Pham. Convergent semidefinite programming relaxations for global bilevel polynomial optimization problems. *SIAM Journal on Optimization*, 26(1):753–780, 2016.
- [31] Kaiyi Ji, Junjie Yang, and Yingbin Liang. Bilevel optimization: Convergence analysis and enhanced design. In *International Conference on Machine Learning*, Vienna, Austria, 2020.
- [32] Jacques Lafontaine et al. *An introduction to differential manifolds*. Springer, 2015.
- [33] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.
- [34] John M. Lee. *Introduction to Smooth Manifolds*. Springer, Springer, New York, NY, 2000.
- [35] Gui-Hua Lin, Mengwei Xu, and Jane Ye. On solving simple bilevel programs with a nonconvex lower level program. *Mathematical Programming*, 144, 04 2014.
- [36] R. Liu, J. Gao, J. Zhang, D. Meng, and Z. Lin. Investigating bi-level optimization for learning and vision from a unified perspective: A survey and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):10045–10067, dec 2022.
- [37] Risheng Liu, Xuan Liu, Xiaoming Yuan, Shangzhi Zeng, and Jin Zhang. A value-function-based interior-point method for non-convex bi-level optimization. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 6882–6892, virtual, 2021. PMLR.
- [38] Risheng Liu, Yaohua Liu, Shangzhi Zeng, and Jin Zhang. Towards gradient-based bilevel optimization with non-convex followers and beyond. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.
- [39] Risheng Liu, Pan Mu, Xiaoming Yuan, Shangzhi Zeng, and Jin Zhang. A generic first-order algorithmic framework for bi-level programming beyond lower-level singleton. In *International Conference on Machine Learning*, Vienna, Austria, 2020.
- [40] Andrea Lodi, Ted Ralphs, and Gerhard Woeginger. Bilevel programming and the separation problem. *Mathematical Programming*, 146, 08 2013.
- [41] Philip D. Loewen and Xianfu Wang. Typical Properties of Lipschitz Functions. *Real Analysis Exchange*, 26(2):717 – 726, 2000.
- [42] Boris S. Mordukhovich. *Bilevel Optimization and Variational Analysis*. Springer International Publishing, Cham, 2020.
- [43] Jiawang Nie, Li Wang, and Jane J. Ye. Bilevel polynomial programs and semidefinite relaxation methods. *SIAM Journal on Optimization*, 27(3):1728–1757, 2017.
- [44] Pablo A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96:293–320, 2003.

- [45] Aravind Rajeswaran, Chelsea Finn, Sham M Kakade, and Sergey Levine. Meta-learning with implicit gradients. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
- [46] Rodolfo Ríos-Zertuche. Examples of pathological dynamics of the subgradient method for lipschitz path-differentiable functions. *Mathematics of Operations Research*, 47(4):3184–3206, 2022.
- [47] Ankur Sinha, Pekka Malo, and Kalyanmoy Deb. A review on bilevel optimization: From classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, 22(2):276–295, 2018.
- [48] Larry J. Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3(1):1–22, 1976.
- [49] B.S. Thomson, J.B. Bruckner, and A.M. Bruckner. *Elementary Real Analysis*. www.classicalrealanalysis.com, 2008.
- [50] Lou van den Dries and Chris Miller. Geometric categories and o-minimal structures. *Duke Mathematical Journal*, 84(2):497 – 540, 1996.
- [51] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Number 47 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.
- [52] J. J. Ye and D. L. Zhu. Optimality conditions for bilevel programming problems. *Optimization*, 33(1):9–27, 1995.