Convergent Semidefinite Programming Relaxations for Global Bilevel Polynomial Optimization Problems

V. Jeyakumar,[*]   J.B. Lasserre[†]   G. Li[‡]   and   T. S. Phạm[§]

Second Revised Version: January 9, 2016

Abstract

In this paper, we consider a bilevel polynomial optimization problem where the objective and the constraint functions of both the upper and the lower level problems are polynomials. We present methods for finding its global minimizers and global minimum using a sequence of semidefinite programming (SDP) relaxations and provide convergence results for the methods. Our scheme for problems with a convex lower-level problem involves solving a transformed equivalent single-level problem by a sequence of SDP relaxations; whereas our approach for general problems involving a non-convex polynomial lower-level problem solves a sequence of approximation problems via another sequence of SDP relaxations.

Key words: Bilevel programming, global optimization, polynomial optimization, semidefinite programming hierarchies.

1   Introduction

Consider the bilevel polynomial optimization problem

$$(P) \quad \min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \quad f(x,y)$$
$$\text{subject to} \quad g_i(x,y) \leq 0, \ i = 1, \ldots, s,$$
$$y \in Y(x) := \operatorname{argmin}_{w \in \mathbb{R}^m} \{G(x,w) : h_j(w) \leq 0, j = 1, \ldots, r\},$$

[*]Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia.   E-mail: v.jeyakumar@unsw.edu.au, Research was partially supported by the Australian Research Council.

[†]LAAS-CNRS and Institute of Mathematics, LAAS, France, E-mail: lasserre@laas.fr. Research was partially supported by a PGM Grant of the Foundation Mathématique Jacques Hadamard.

[‡]Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia.   E-mail: g.li@unsw.edu.au, Research was partially supported by grant from the Australian Research Council. Part of this work was done while the author was visiting LAAS-CNRS in July 2014.

[§]Institute of Research and Development, Duy Tan University, K7/25, Quang Trung, Danang, Vietnam, and Department of Mathematics, University of Dalat, 1 Phu Dong Thien Vuong, Dalat, Vietnam. E-mail: sonpt@dlu.edu.vn. This author was partially supported by Vietnam National Foundation for Science and Technology Development (NAFOSTED) grant number 101.04-2013.07. Part of the work was done while this author was visiting the Department of Applied Mathematics at University of New South Wales, Australia.

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $g_i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $h_j : \mathbb{R}^m \to \mathbb{R}$ are all polynomials with real coefficients, and we make the blanket assumption that the feasible set of (P) is nonempty, that is, $\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : g_i(x, y) \leq 0, i = 1, \ldots, s, \ y \in Y(x)\} \neq \emptyset$.

Bilevel optimization provides mathematical models for hierarchical decision making processes where the follower's decision depends on the leader's decision. More precisely, if x and y are the decision variables of the leader and the follower respectively then the problem (P) represents the so-called optimistic approach to the leader and follower's game in which the follower is assumed to be co-operative and so, the leader can choose the solution with the lowest cost. We note that, there is another approach, called pessimistic approach, which assumes that the follower may not be co-operative and hence the leader will need to prepare for the worst cost (see for example [11, 44]).

The bilevel optimization problem (P) also requires that the constraints of the lower level problem are independent of the upper level decision variable x (i.e. the functions h_j do not depend on x). This independence assumption guarantees that the optimal value function of the lower level problem is automatically continuous, and so, plays an important role later in establishing convergence of our proposed approximation schemes for finding global optimal solutions of (P). A discussion on this assumption and its possible relaxation is given in Remark 4.9 in Section 4 of the paper.

As noted in [31], the models of the form (P) cover the situations in which the leader can only observe the outcome of the follower's action but not the action itself, and so, has important applications in economics such as the so-called moral hazard model of the principal-agent problem. In particular, in the special case where g_i depends only on x, the sets $\{x \in \mathbb{R}^n : g_i(x) \leq 0\}$ and $\{w \in \mathbb{R}^m : h_j(w) \leq 0\}$ are both convex sets, problem (P) has been studied in [31] and a smoothing projected gradient algorithm has been proposed to find a stationary point of problem (P). On the other hand, the functions f, g_i, G, h_j of (P) in [31] are allowed to be continuously differentiable functions which may not be polynomials in general. For applications and recent developments of solving more general bilevel optimization problems, see [3, 9, 10, 11, 43].

In this paper, in the interest of simplicity, we focus on the *optimistic approach* to the hierarchical decision making process and develop methods for finding a global minimizer and global minimum of (P). We make the following key contributions to bilevel optimization.

- **A novel SDP hierarchy for bilevel polynomial problems.** We propose general purpose schemes for finding global solutions of the bilevel polynomial optimization problem (P) by solving hierarchies of semidefinite programs and establish convergence of the schemes. Our approach makes use of the known techniques of bilevel optimization and the recent developments of (single-level) polynomial optimization, such as the sums-of-squares decomposition and semidefinite programming hierarchy, and does not use any discretization or branch-and-bound techniques as in [17, 37, 44].

- **Convex lower-level problems: Convergence to global solutions.** We first transform the bilevel polynomial optimization problem (P) with a *convex lower-level problem* into an equivalent single level nonconvex polynomial optimization problem. We show that the values of the standard semidefinite programming relaxations of the transformed single level problem converge to the

2

global optimal value of the bilevel problem (P) under a technical assumption that is commonly used in polynomial optimization (see [26] and other references therein).

- **Non-convex lower-level problems: A new convergent approximation scheme.** By examining a sequence of ϵ-approximation (single-level) problems of the bilevel problem (P) with a *not necessarily convex lower level problem*, we present another convergent sequence of SDP relaxations of (P) under suitable conditions. Our approach extends the sequential SDP relaxations, introduced in [27] for parameterized single-level polynomial problems, to bilevel polynomial optimization problems.

It is important to note that local bilevel optimization techniques, studied extensively in the literature [3, 10], apply to broad classes of nonlinear bilevel optimization problems. In the present work, we employ some basic tools and techniques of semi-algebraic geometry to achieve convergence of our semidefinite programming hierarchies of global nonlinear bilevel optimization problems, and so our approaches are limited to studying the class of polynomial bilevel optimization problems.

Moreover, due to the limitation of the SDP programming solvers, our proposed scheme can be used to solve problems with small or moderate size and it may not be able to compete with the ad-hoc (but computationally tractable) techniques, such as branch-and-bound methods and discretization schemes. For instance, underestimation and branch-and-bound techniques were used in [1, 17, 37] and a generalized semi-infinite programming reformulation together with a discretization technique was employed in [44]. See http://bilevel.org/ for other references of computational methods of bilevel optimization.

However, it has recently been shown that, by exploiting sparsity and symmetry, large size problems can be solved efficiently and various numerical packages have been built to solve real-life problems such as the sensor network localization problem [24]. We leave the study of solving large size bilevel problems for future research as it is beyond the scope of this paper.

The outline of the paper is as follows. Section 2 gives preliminary results on polynomials and continuity properties of the solution map of the lower-level problem of (P). Section 3 presents convergence of our sequential SDP relaxation scheme for solving the problem (P) with a convex lower-level problem. Section 4 describes another sequential SDP relaxation scheme and its convergence for solving the general problem (P) with a not necessarily convex lower-level problem. Section 5 reports results of numerical implementations of the proposed methods for solving some bilevel optimization test problems. The appendix provides details of various technical results of semi-algebraic geometry used in the paper and also proofs of certain technical results.

2 Preliminaries

We begin by fixing notation, definitions and preliminaries. Throughout this paper $\mathbb{R}^n$ denotes the Euclidean space with dimension n. The inner product in $\mathbb{R}^n$ is defined by $\langle x, y \rangle := x^T y$ for all $x, y \in \mathbb{R}^n$. The open (resp. closed) ball in $\mathbb{R}^n$ centered at x with radius ρ is denoted by $\mathbb{B}(x, \rho)$ (resp. $\overline{\mathbb{B}}(x, \rho)$).

The non-negative orthant of $\mathbb{R}^n$ is denoted by $\mathbb{R}_+^n$ and is defined by $\mathbb{R}_+^n := \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i \geq 0\}$. Denote by $\mathbb{R}[x]$ the ring of polynomials in $x := (x_1, x_2, \ldots, x_n)$ with real coefficients. For a polynomial f with real coefficients, we use $\deg f$ to denote the degree of f. For a differentiable function f on $\mathbb{R}^n$, ∇f denotes its derivative. For a differentiable function $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, we use $\nabla_x g$ (resp. $\nabla_y g$) to denote the derivative of g with respect to the first variable (resp. second variable). We also use $\mathbb{N}$ (resp. $\mathbb{N}_{>0}$) to denote all the nonnegative (resp. positive) integers. Moreover, for any integer t, let $\mathbb{N}_t^n := \{\alpha \in \mathbb{N}^n : \sum_{i=1}^n \alpha_i \leq t\}$. For a set A in $\mathbb{R}^n$, we use $\mathrm{cl}(A)$ and $\mathrm{int}(A)$ to denote the closure and interior of A. For a given point x, the distance from the point x to a set A is denoted by $d(x, A)$ and is defined by $d(x, A) = \inf\{\|x - a\| : a \in A\}$.

We say that a real polynomial $f \in \mathbb{R}[x]$ is *sum-of-squares* (SOS) if there exist real polynomials $f_j, j = 1, \ldots, r$, such that $f = \sum_{j=1}^r f_j^2$. The set of all sum-of-squares real polynomials in x is denoted by $\Sigma^2[x]$. Moreover, the set of all sum-of-squares real polynomials in x with degree at most d is denoted by $\Sigma_d^2[x]$. We also recall some notions and results of semi-algebraic functions/sets, which can be found in [6, 15].

Definition 2.1 (Semi-algebraic sets and functions) *A subset of $\mathbb{R}^n$ is called* semi-algebraic *if it is a finite union of sets of the form $\{x \in \mathbb{R}^n : f_i(x) = 0, i = 1, \ldots, k; f_i(x) > 0, i = k+1, \ldots, p\}$, where all f_i are real polynomials. If $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^p$ are semi-algebraic sets then the map $f : A \to B$ is said to be* semi-algebraic *if its graph $\{(x, y) \in A \times B : y = f(x)\}$ is a semi-algebraic subset in $\mathbb{R}^n \times \mathbb{R}^p$.*

Semi-algebraic sets and functions are important classes of sets and functions and they have important applications in nonsmooth optimization (for a recent development, see [7]). In particular, they enjoy a number of remarkable properties. Some of these properties, which are used later in the paper, have been summarized in the Appendix A for the convenience of the reader.

We now present a preliminary result on Hölder continuity of the solution mapping of the lower level problem. As a consequence, we provide an existence result of the solution of a bilevel polynomial optimization problem (P).

Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping and let $\bar{y} \in F(\bar{x})$. Recall that F is said to be Hölder continuous (calm) at $(\bar{x}, \bar{y})$ with exponent $\tau \in (0, 1]$ if there exist $\delta, \epsilon, c > 0$ such that

$$d(y, F(\bar{x})) \leq c\|x - \bar{x}\|^\tau \text{ for all } y \in F(x) \cap \mathbb{B}_{\mathbb{R}^m}(\bar{y}, \epsilon) \text{ and } x \in \mathbb{B}_{\mathbb{R}^n}(\bar{x}, \delta).$$

In the case when $\tau = 1$, this property is often refereed as calmness and has been well-studied in nonsmooth analysis (see for example [8]). We first see that even in the case, where G is a continuously differentiable function and the set $\{y \in \mathbb{R}^m : h_j(y) \leq 0\}$ is compact, the solution map $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ of the lower level problem $Y(x) := \mathrm{argmin}_{y \in \mathbb{R}^m}\{G(x, y) : h_j(y) \leq 0, j = 1, \ldots, r\}$ is not necessarily Hölder continuous for any exponent $\tau > 0$.

Example 2.2 (Failure of Hölder continuity for solution map of the lower level problem:

non-polynomial case) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(y) = \begin{cases} e^{-\frac{1}{y^2}} & \text{if} \quad y \neq 0 \\ 0 & \text{if} \quad y = 0. \end{cases}$$

Consider the solution mapping $Y(x) = \operatorname{argmin}_y\{G(x,y) : y^2 \leq 1\}$ for all $x \in [-1,1]$, where $G(x,y) = (x - f(y))^2$. Then, it can be verified that G is a continuously differentiable function (indeed it is a C^∞ function) and

$$Y(x) = \begin{cases} \{\pm\sqrt{\frac{1}{-\ln x}}\} & \text{if} \quad x \in (0,1], \\ \{0\} & \text{if} \quad x \in [-1,0]. \end{cases}$$

We now see that the solution mapping Y is not Hölder continuous at 0 with exponent τ for any $\tau \in (0,1]$. To see this, let $x_k = e^{-k} \to 0$ and $y_k = \sqrt{\frac{1}{k}} \in Y(x_k)$. Then, for any $\tau \in (0,1]$,

$$\frac{|x_k|^\tau}{d(y_k, Y(0))} = \frac{e^{-\tau k}}{\sqrt{\frac{1}{k}}} = \frac{\sqrt{k}}{e^{\tau k}} \to 0.$$

So, the solution mapping is not Hölder continuous at 0 for any $\tau \in (0,1]$.

The Hölder continuity of the solution set of general parametric optimization problems has been established under suitable regularity conditions, for example see [19, 41]. This property plays an important role in establishing the existence of solutions for bilevel programming problems and equilibrium problems (see for example [33] and Corollary 2.5). Next, we show that, the solution map of a lower level problem of a bilevel polynomial optimization problem is always Hölder continuous with an explicit exponent which depends only on the degree of the polynomial involved and the dimension of the underlying space. This result is based on our recent established Łojasiewicz inequality for nonconvex polynomial systems in [30].

For $m, d \in \mathbb{N}$, denote

$$R(m,d) := \begin{cases} 1 & \text{if } d = 1, \\ d(3d-3)^{m-1} & \text{if } d \geq 2. \end{cases} \tag{2.1}$$

Theorem 2.3 (Hölder continuity of solution maps in the lower level problem: polynomial case) *Let h_j, $j = 1, \ldots, r$ and G be polynomials with real coefficients. Denote $d := \max\{\deg h_j, \deg G(x, \cdot)\}$. Suppose that $F := \{y \in \mathbb{R}^m : h_j(y) \leq 0\}$ is compact. Then, the solution map $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ in the lower level problem $Y(x) := \operatorname{argmin}_{y \in \mathbb{R}^m}\{G(x,y) : h_j(y) \leq 0, j = 1, \ldots, r\}$ satisfies the following Hölder continuity property at each point $\bar{x} \in \mathbb{R}^n$: for any $\delta > 0$, there is a constant $c > 0$ such that*

$$Y(x) \subset Y(\bar{x}) + c\,\|x - \bar{x}\|^\tau\,\overline{\mathbb{B}}_{\mathbb{R}^m}(0,1) \quad \text{whenever} \ \|x - \bar{x}\| \leq \delta, \tag{2.2}$$

for some $\tau \in [\tau_0, 1]$ with $\tau_0 = \max\{\frac{1}{R(m+r+1,d+1)}, \frac{2}{R(m+r,2d)}\}$. In particular, Y is Hölder continuous at $\bar{x}$ with exponent τ_0 for any $\bar{x} \in \mathbb{R}^n$.

Proof. For any fixed $x \in \mathbb{R}^n$, define $\Phi(x) = \min_{y \in \mathbb{R}^m} \{G(x,y) : h_j(y) \leq 0, j = 1, \ldots, r\}$ and let

$$\Phi_x(y) := \sum_{i=1}^{r} \big[h_j(y)\big]_+ + |\Phi(x) - G(x,y)|.$$

Then, for all fixed x,

$$
\begin{aligned}
\{y \in \mathbb{R}^n | \, \Phi_x(y) = 0\} &= Y(x) \\
&= \big\{y \in \mathbb{R}^m \big| \, h_j(y) \leq 0 \ \text{ as } \ j = 1, \ldots, s, \ \text{ and } \ \Phi(x) - G(x,y) = 0\big\}.
\end{aligned}
$$

Note that F is compact. Now, the Łojasiewicz inequality for nonconvex polynomial systems [30, Corollary 3.8] gives that there is a constant $c_0 > 0$ such that

$$d\big(y, Y(\overline{x})\big) \leq c_0 \, \Phi_{\overline{x}}(y)^\tau \quad \text{for all } y \in F, \tag{2.3}$$

for some $\tau \in [\tau_0, 1]$ with $\tau_0 = \max\{\frac{1}{R(m+r+1,d+1)}, \frac{2}{R(m+r,2d)}\}$. Further, there is a constant $L > 0$ such that

$$|G(x,y) - G(\overline{x},y)| \leq L\|x - \overline{x}\| \tag{2.4}$$

for all $y \in F$ and for all x with $\|x - \overline{x}\| \leq \delta$. Denote $c := (2\beta^{-1}L)^\tau$ with $\beta := c_0^{-\frac{1}{\tau}} > 0$. For any $y \in Y(x)$ we select now $\overline{y} \in Y(\overline{x})$ satisfying $\|y - \overline{y}\| = d(y, Y(\overline{x}))$. To finish the proof, it suffices to show that

$$\|y - \overline{y}\| \leq c \, \|x - \overline{x}\|^\tau. \tag{2.5}$$

To see this, note that $|\Phi(\overline{x}) - G(\overline{x},y)| = \Phi_{\overline{x}}(y) \geq \beta d\big(y, Y(\overline{x})\big)^{\frac{1}{\tau}} = \beta\|y - \overline{y}\|^{\frac{1}{\tau}}$. Since $\overline{y} \in Y(\overline{x})$, it follows that $G(\overline{x}, \overline{y}) = \Phi(\overline{x}) \leq G(\overline{x}, y)$, and hence

$$\|y - \overline{y}\|^{\frac{1}{\tau}} \leq \beta^{-1}|\Phi(\overline{x}) - G(\overline{x},y)| = \beta^{-1}\big(G(\overline{x},y) - G(\overline{x},\overline{y})\big). \tag{2.6}$$

Furthermore, as $y \in Y(x)$, $G(x,y) \leq G(x,\overline{y})$, and therefore (2.4) gives us that

$$
\begin{aligned}
G(\overline{x},y) - G(\overline{x},\overline{y}) &= \big(G(\overline{x},y) - G(x,y)\big) + \big(G(x,y) - G(x,\overline{y})\big) + \big(G(x,\overline{y}) - G(\overline{x},\overline{y})\big) \\
&\leq \big(G(\overline{x},y) - G(x,y)\big) + \big(G(x,\overline{y}) - G(\overline{x},\overline{y})\big) \\
&\leq 2L\|x - \overline{x}\| \quad \text{as } \ y, \overline{y} \in F.
\end{aligned}
$$

This together with (2.6) yields

$$\|y - \overline{y}\|^{\frac{1}{\tau}} \leq \beta^{-1}\big(G(\overline{x},y) - G(\overline{x},\overline{y})\big) \leq 2\beta^{-1}L\|x - \overline{x}\|.$$

Thus

$$d\big(y, Y(\overline{x})\big) = \|y - \overline{y}\| \leq c \, \|x - \overline{x}\|^\tau,$$

which verifies (2.5) and completes the proof of the theorem.

In general, our lower estimate of the exponent τ will not be tight. We present a simple example to illustrate this.

Example 2.4 Consider the solution mapping $Y(x) = \text{argmin}_{y \in \mathbb{R}}\{(x-y^2)^2 : y^2 \leq 1\}$ for all $x \in [-1, 1]$. Clearly,

$$Y(x) = \begin{cases} \{\pm\sqrt{x}\} & \text{if} \quad x \in [0, 1], \\ \{0\} & \text{if} \quad x \in [-1, 0). \end{cases}$$

So, the solution mapping is Hölder continuous at 0 with exponent $1/2$. On the other hand, our lower estimate gives $\tau_0 = 1/84$. So, the lower estimate is not tight.

Corollary 2.5 (Existence of global minimizer) *For the bilevel polynomial optimization problem* (P), *let* $K = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : g_i(x, y) \leq 0\}$ *and* $F = \{w \in \mathbb{R}^m : h_j(w) \leq 0\}$. *Suppose that* $K_1 = \{x \in \mathbb{R}^n : (x, y) \in K \text{ for some } y \in \mathbb{R}^m\}$ *and* F *are compact sets. Then, a global minimizer for* (P) *exists.*

Proof. Denote the optimal value of problem (P) by $\text{val}(P)$. Let (x_k, y_k) be a minimizing sequence for the bilevel polynomial optimization problem (P) in the sense that $g_i(x_k, y_k) \leq 0$, $i = 1, \ldots, s$, $h_j(y_k) \leq 0$, $j = 1, \ldots, r$, $y_k \in Y(x_k)$ and $f(x_k, y_k) \to \text{val}(P)$. Clearly, $(x_k, y_k) \in K$ (and so, $x_k \in K_1$) and $y_k \in F$. By passing to a subsequence, we may assume that $(x_k, y_k) \to (\bar{x}, \bar{y}) \in K_1 \times F$. By continuity, we have $f(\bar{x}, \bar{y}) = \text{val}(P)$. To see the conclusion, it suffices to show that $\bar{y} \in Y(\bar{x})$. Denote $\epsilon_k = \|x_k - \bar{x}\| \to 0$. Then, by Theorem 2.3, there is $c > 0$ such that

$$Y(x_k) \subseteq Y(\bar{x}) + c\, \epsilon_k^\tau \overline{\mathbb{B}}_{\mathbb{R}^m}(0, 1) \quad \text{for all} \quad k \in \mathbb{N}.$$

As $y_k \in Y(x_k)$, there exists $y_k' \in Y(\bar{x})$ such that

$$\|y_k - y_k'\| \leq 2c\epsilon_k^\tau \to 0. \tag{2.7}$$

Note that $Y(\bar{x}) \subseteq F$, $Y(\bar{x})$ is a closed set and F is compact. It follows that $Y(\bar{x})$ is also a compact set. Passing to the limit in (2.7), we see that $\bar{y} \in Y(\bar{x})$. So, a global minimizer for problem (P) exists.

The following lemma of Putinar ([39]), which provides a characterization for positivity of a polynomial over a system of polynomial inequalities, can also be regarded as a polynomial analog of Farkas' lemma [14]. This lemma has been extensively used in polynomial optimization [26] and plays a key role in the convergence analysis of our proposed method later on.

Lemma 2.6 (Putinar's Positivstellensatz)[39] *Let* f_0 *and* f_i, $i = 1, \ldots, p$ *be real polynomials in* w *on* $\mathbb{R}^v$. *Suppose that there exist* $R > 0$ *and sums-of-squares polynomials* $\hat{\sigma}_1, \ldots, \hat{\sigma}_p \in \Sigma^2[w]$ *such that* $R - \|w\|^2 = \hat{\sigma}_0(w) + \sum_{i=1}^{p} \hat{\sigma}_i f_i(w)$ *for all* $w \in \mathbb{R}^v$. *If* $f_0(w) > 0$ *over the set* $\{w \in \mathbb{R}^v : f_i(w) \geq 0, i = 1, \ldots, p\}$, *then there exist* $\sigma_i \in \Sigma^2[w]$, $i = 0, 1, \ldots, p$ *such that* $f_0 = \sigma_0 + \sum_{i=1}^{p} \sigma_i f_i$.

The following assumption plays a key role throughout the paper.

Assumption 2.1: There exist $R_1, R_2 > 0$ such that the quadratic polynomials $(x, y) \mapsto R_1 - \|(x, y)\|^2$ and $y \mapsto R_2 - \|y\|^2$ can be written as

$$R_1 - \|(x, y)\|^2 = \sigma_0(x, y) - \sum_{i=1}^{s} \sigma_i(x, y)g_i(x, y) \quad \text{and} \quad R_2 - \|y\|^2 = \bar{\sigma}_0(y) - \sum_{j=1}^{r} \bar{\sigma}_j(y)h_j(y),$$

7

for some sums-of-squares polynomials $\sigma_0, \sigma_1, \ldots, \sigma_s \in \Sigma^2[x, y]$ and sums-of-squares polynomials $\bar{\sigma}_0, \bar{\sigma}_1, \ldots, \bar{\sigma}_r \in \Sigma^2[y]$.

We note that Assumption 2.1 implies that both $K = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : g_i(x, y) \leq 0, i = 1, \ldots, s\}$ and $F = \{y \in \mathbb{R}^m : h_j(y) \leq 0, j = 1, \ldots, r\}$ are compact sets [26]. Moreover, Assumption 2.1 can be easily satisfied when K and F are nonempty compact sets, and one knows the bounds N_1 for $\|x\|$ on K and N_2 for $\|y\|$ on F. Indeed, in this case, it suffices to add redundant constraints $g_{s+1}(x, y) = \|(x, y)\|^2 - (N_1^2 + N_2^2)$ and $h_{r+1}(y) = \|y\|^2 - N_2^2$ to the definition of K and F respectively, and Assumption 2.1 is satisfied with $R_1 = N_1^2 + N_2^2$, $R_2 = N_2^2$, $\sigma_i \equiv 0$ for all $1 \leq i \leq s$, $\bar{\sigma}_j \equiv 0$ for all $1 \leq j \leq r$ and $\sigma_{s+1} = \bar{\sigma}_{r+1} \equiv 1$. We also note that, under Assumption 2.1, a solution for problem (P) exists by Corollary 2.5.

3 Convex Lower Level Problems

In this section, we consider the convex polynomial bilevel programming problem (P) where the lower level problem is convex in the sense that, for each $x \in \mathbb{R}^n$, $G(x, \cdot)$ is a convex polynomial, h_j are polynomials, $j = 1, \ldots, r$, and the feasible set of lower level problem $F := \{w \in \mathbb{R}^m : h_j(w) \leq 0, j = 1, \ldots, r\}$ is a convex set. We note that, the representing polynomials h_j which describes the convex feasible set F need not to be convex, in general.

We say that the lower level convex problem of (P) satisfies the *nondegeneracy condition* if for each $j = 1, \ldots, r$,

$$y \in F \text{ and } h_j(y) = 0 \implies \nabla h_j(y) \neq 0.$$

Recall that the lower level convex problem of (P) is said to satisfy the *Slater condition* whenever there exists $y_0 \in \mathbb{R}^m$ such that $h_j(y_0) < 0$, $j = 1, \ldots, r$. Note that, under the Slater condition, the lower level problem automatically satisfies the nondegeneracy condition if each h_j, $j = 1, \ldots, r$ is a convex polynomial.

Let us recall a lemma which provides a link between a KKT point and a minimizer for a convex optimization problem where the representing function of the convex feasible set is not necessarily convex.

Lemma 3.1 *([28, Theorem 2.1]) Let ϕ be a convex function on $\mathbb{R}^m$ and $F := \{w \in \mathbb{R}^m : h_j(w) \leq 0, j = 1, \ldots, r\}$ be a convex set. Suppose that both the nondegeneracy condition and the Slater condition hold. Then, a point y is a global minimizer of $\min\{\phi(w) : w \in F\}$ if and only if y is a KKT point of $\min\{\phi(w) : w \in F\}$, in the sense that, there exist $\lambda_j \geq 0$, $j = 1, \ldots, r$ such that*

$$\nabla\phi(y) + \sum_{j=1}^{r} \lambda_j \nabla h_j(y) = 0, \ \lambda_j h_j(y) = 0, \ h_j(y) \leq 0, \ j = 1, \ldots, r.$$

We see in the following proposition that a polynomial bilevel programming problem with convex lower level problem can be equivalently rewritten as a single level polynomial optimization problem in

8

a higher dimension under the nondegeneracy condition and the Slater condition. In the special case where all the representing polynomials h_j are convex, this lemma has been established in [12].

Proposition 3.2 (Equivalent single-level problem) *Consider problem (P) where the lower level problem is convex. Suppose that the lower level problem satisfies both the nondegeneracy condition and the Slater condition. Then, $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ is a global solution of the bilevel polynomial optimization problem (P) if and only if there exist Lagrange multipliers[1] $\lambda = (\lambda_0, \ldots, \lambda_r) \in \mathbb{R}^{r+1}$ such that $(x, y, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{r+1}$ is a global solution of the following single level polynomial optimization problem:*

$$(\overline{P}) \quad \min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m, \lambda \in \mathbb{R}^{r+1}} \quad f(x, y)$$

$$\text{subject to} \quad g_i(x, y) \leq 0, \ i = 1, \ldots, s, \tag{3.8}$$

$$\lambda_0 \nabla_y G(x, y) + \sum_{j=1}^{r} \lambda_j \nabla h_j(y) = 0,$$

$$\lambda_0 \geq 0, \ \sum_{j=0}^{r} \lambda_j^2 = 1, \ \lambda_j h_j(y) = 0, \ \lambda_j \geq 0, \ h_j(y) \leq 0, \ j = 1, \ldots, r.$$

Proof. Fix any $x \in \mathbb{R}^n$. The conclusion will follow if we show that $y \in Y(x)$ is equivalent to the condition that there exist $\lambda_j \geq 0$, $j = 0, 1, \ldots, r$ such that

$$\lambda_0 \nabla_y G(x, y) + \sum_{j=1}^{r} \lambda_j \nabla h_j(y) = 0,$$

$$\lambda_j h_j(y) = 0, \ \lambda_j \geq 0, \ h_j(y) \leq 0, \ j = 1, \ldots, r, \tag{3.9}$$

$$\lambda_0 \geq 0, \sum_{j=0}^{r} \lambda_j^2 = 1.$$

To see the equivalence, we first assume that $y \in Y(x)$. Under both the nondegeneracy condition and the Slater condition, the preceding Lemma guarantees that there exist $\mu_j \geq 0$, $j = 1, \ldots, r$, such that

$$\nabla_y G(x, y) + \sum_{j=1}^{r} \mu_j \nabla h_j(y) = 0, \ \mu_j h_j(y) = 0, \ \mu_j \geq 0, \ h_j(y) \leq 0, \ j = 1, \ldots, r. \tag{3.10}$$

So, (3.9) holds with $\lambda_0 = \frac{1}{\sqrt{1 + \sum_{j=1}^{r} \mu_j^2}}$ and $\lambda_j = \frac{\mu_j}{\sqrt{1 + \sum_{j=1}^{r} \mu_j^2}}$, $j = 1, \ldots, r$.

Conversely, let (x, y, λ) satisfy (3.9). We now show that $\lambda_0 \neq 0$. Indeed, assume on the contrary that $\lambda_0 = 0$. Then, $\sum_{j=1}^{r} \lambda_j^2 = 1$, $\sum_{j=1}^{r} \lambda_j \nabla h_j(y) = 0$, $\lambda_j h_j(y) = 0$, $\lambda_j \geq 0$ and $h_j(y) \leq 0$ $j = 1, \ldots, r$. Let $J = \{j \in \{1, \ldots, r\} : \lambda_j > 0\} \neq \emptyset$. From the Slater condition, there exists $y_0 \in \mathbb{R}^m$ such that $h_j(y_0) < 0$, $j = 1, \ldots, r$. Then, there exists $\rho > 0$ such that $h_j(w) < 0$ for all $w \in \mathbb{R}^m$ with $\|w - y_0\| \leq \rho$. As $\sum_{j=1}^{r} \lambda_j \nabla h_j(y) = 0$, we obtain

$$\sum_{j \in J} \lambda_j \nabla h_j(y)^T (w - y) = 0 \text{ for all } w \text{ with } \|w - y_0\| \leq \rho. \tag{3.11}$$

---

[1]Indeed, as shown in the proof, $\lambda_0 \neq 0$ always holds under our assumptions. See Remark 3.3 for a detailed discussion.

We now see that $\nabla h_j(y)^T(w - y) \leq 0$ for all w with $\|w - y_0\| \leq \rho$ and for all $j \in J$. (Suppose on the contrary that there exists w_0 with $\|w_0 - y_0\| \leq \rho$ and $j_0 \in J$ such that $\nabla h_{j_0}(y)^T(w_0 - y) > 0$. By continuity, for all small t, $h_{j_0}(y + t(w_0 - y)) > 0$, and hence $y + t(w_0 - y) \notin F$. On the other hand, from our choice of ρ, we see that $h_j(w_0) < 0$ for all $j = 1, \ldots, r$. So, $w_0 \in F$. It then follows from the convexity of F that $y + t(w_0 - y) \in F$ for all small t. This is impossible.) This together with (3.11) and $\lambda_j = 0$ for all $j \notin J$ shows that

$$\nabla h_j(y)^T(w - y) = 0 \text{ for all } w \text{ with } \|w - y_0\| \leq \rho \text{ and } j \in J,$$

and so, $\nabla h_j(y) = 0$ for all $j \in J$. Note that $y \in F$ and $h_j(y) = 0$ for all $j \in J$. This contradicts the non-degeneracy condition, and so, $\lambda_0 \neq 0$. Thus, by dividing λ_0 on both sides of the first relation of (3.9), we see that (3.10) holds. This shows that $y \in Y(x)$ by the preceding lemma again.

Remark 3.3 (Importance of nondegeneracy and Slater's conditions) In Proposition 3.2, we require that the nondegeneracy condition and the Slater condition hold. These assumptions provide us a simple uniform bound for the multipliers $\lambda_0, \ldots, \lambda_r$ in the lower level problem which plays an important role in our convergence analysis later in Theorem 3.5. Indeed, these assumptions ensure that $\lambda_0 \neq 0$, and so, in particular, the equivalence of the following two systems:

$$\begin{bmatrix} \lambda_0 \nabla_y G(x, y) + \sum_{j=1}^{r} \lambda_j \nabla h_j(y) = 0, \\ \lambda_0 \geq 0, \ \lambda_j h_j(y) = 0, \ \lambda_j \geq 0, \ h_j(y) \leq 0, \ j = 1, \ldots, r. \\ \sum_{j=0}^{r} \lambda_j^2 = 1, \end{bmatrix} \Leftrightarrow \begin{bmatrix} \nabla_y G(x, y) + \sum_{j=1}^{r} \mu_j \nabla h_j(y) = 0, \\ \mu_j h_j(y) = 0, \ \mu_j \geq 0, \ h_j(y) \leq 0, \ j = 1, \ldots, r. \end{bmatrix}$$

Note that the non-degeneracy condition is satisfied when the representing functions h_j, $j = 1, \ldots, r$, are convex polynomials and the Slater condition holds. Thus, in this special case, the Slater condition alone is enough for transforming the polynomial bilevel problem with a convex lower level problem to a single-level polynomial optimization problem.

The following simple example illustrates that the preceding Proposition can be applied to the case where h_j's need not be convex polynomials.

Example 3.4 Consider the bilevel problem

$$(EP_1) \quad \min_{x \in \mathbb{R}, y \in \mathbb{R}^2} \quad -x^6 + y_1^2 + y_2^2$$

$$\text{subject to} \quad x^2 + y_1^2 + y_2^2 \leq 2$$

$$y \in Y(x) := \operatorname{argmin}_{w \in \mathbb{R}^2} \{x(w_1 + w_2) : 1 - w_1 w_2 \leq 0, 0 \leq w_1 \leq 1, 0 \leq w_2 \leq 1\}.$$

Clearly, the lower level problem of (EP_1) is convex but the polynomial $(w_1, w_2) \mapsto 1 - w_1 w_2$ is not convex. It can be verified that the non-degeneracy condition and Slater condition hold, and so, (EP_1)

is equivalent to the following single level polynomial optimization problem

$$\min_{x\in\mathbb{R},y\in\mathbb{R}^2,(\lambda_0,\dots,\lambda_5)\in\mathbb{R}^6} \quad -x^6 + y_1^2 + y_2^2$$

$$\text{subject to} \quad x^2 + y_1^2 + y_2^2 \le 2$$

$$1 - y_1 y_2 \le 0, 0 \le y_1 \le 1, 0 \le y_2 \le 1$$

$$\lambda_0 x + \lambda_1(-y_2) - \lambda_2 + \lambda_3 = 0$$

$$\lambda_0 x + \lambda_1(-y_1) - \lambda_4 + \lambda_5 = 0$$

$$\lambda_1(1 - y_1 y_2) = 0, \ \lambda_2 y_1 = 0, \ \lambda_3(1 - y_1) = 0$$

$$\lambda_4 y_2 = 0, \ \lambda_5(1 - y_2) = 0$$

$$\lambda_j \ge 0, j = 0, 1, \dots, 5, \sum_{j=0}^{5} \lambda_j^2 = 1.$$

Proposition 3.2 enables us to construct a sequence of semidefinite programming problems for solving a polynomial bilevel programming problem with a convex lower level problem. To do this, we denote

$$\widehat{G}_p(x, y, \lambda) = \begin{cases} g_p(x, y) & p = 1, \dots, s, \\ h_{p-s}(y) & p = s+1, \dots, s+r, \\ -\lambda_{p-(s+r+1)} & p = s+r+1, \dots, s+2r+1, \end{cases}$$

and

$$\widehat{H}_q(x, y, \lambda) = \begin{cases} \lambda_q h_q(y), & q = 1, \dots, r \\ \left(\lambda_0 \nabla_y G(x, y) + \sum_{j=1}^{r} \lambda_j \nabla h_j(y)\right)_{q-r}, & q = r+1, \dots, r+m \\ \sum_{j=0}^{r} \lambda_j^2 - 1, & q = r+m+1, \end{cases}$$

where $\left(\lambda_0 \nabla_y G(x, y) + \sum_{j=1}^{r} \lambda_j \nabla h_j(y)\right)_i$ is the ith coordinate of $\lambda_0 \nabla_y G(x, y) + \sum_{j=1}^{r} \lambda_j \nabla h_j(y)$, $i = 1, \dots, m$. We also denote the degree of $\widehat{G}_p$ to be u_p and the degree of $\widehat{H}_q$ to be v_q .

We now introduce a sequence of sums-of-squares relaxation problems as follows: for each $k \in \mathbb{N}$,

$$(Q_k) \quad \max_{\mu, \sigma_p} \quad \mu \tag{3.12}$$

$$\text{s.t.} \quad f - \mu = \sigma_0 - \sum_{p=1}^{s+2r+1} \sigma_p \widehat{G}_p - \sum_{q=1}^{r+m+1} \phi_q \widehat{H}_q$$

$$\sigma_p \in \Sigma^2[x, y, \lambda], \ p = 0, 1, \dots, s+2r+1,$$

$$\deg\sigma_0 \le 2k, \ \deg(\sigma_p \widehat{G}_p) \le 2k, p = 1, \dots, s+2r+1,$$

$$\phi_q \in \mathbb{R}[x, y, \lambda], \ q = 1, \dots, r+m+1, \ \deg(\phi_q \widehat{H}_q) \le 2k, q = 1, \dots, r+m+1.$$

It is known that each (Q_k) can be reformulated as a semidefinite programming problem [26].

Theorem 3.5 (Convex lower level problem: Convergence theorem) *Consider the problem (P) where the lower level problem is convex. Suppose that Assumption 2.1 holds and that the lower level problem satisfies both the nondegeneracy condition and the Slater condition. Then,* $\mathrm{val}(Q_k) \leq \mathrm{val}(Q_{k+1})$ *for all* $k \in \mathbb{N}$ *and* $\mathrm{val}(Q_k) \to \mathrm{val}(P)$ *as* $k \to \infty$, *where* $\mathrm{val}(Q_k)$ *and* $\mathrm{val}(P)$ *denote the optimal value of the problems* (Q_k) *and* (P) *respectively.*

Proof. From Corollary 2.5, a global solution of (P) exists. Let (x, y) be a global solution of (P). From Proposition 3.2, there exists $\lambda \in \mathbb{R}^{r+1}$ such that (x, y, λ) is a solution of $(\overline{P})$ and $\mathrm{val}(P) = \mathrm{val}(\overline{P})$.

From the construction of (Q_k), $k \in \mathbb{N}$, it can be easily verified that $\mathrm{val}(Q_k) \leq \mathrm{val}(Q_{k+1}) \leq \mathrm{val}(\overline{P})$ for all $k \in \mathbb{N}$. Let $\epsilon > 0$. Define $\widehat{f}(x, y, \lambda) = f(x, y) - (\mathrm{val}(\overline{P}) - \epsilon)$. Note that the feasible set U of $(\overline{P})$ can be written as

$$U = \{(x, y, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{r+1} \ : \ -\widehat{G}_p(x, y, \lambda) \geq 0, \ p = 1, \ldots, s + 2r + 1,$$
$$-\widehat{H}_q(x, y, \lambda) \geq 0, \widehat{H}_q(x, y, \lambda) \geq 0, \ q = 1, \ldots, r + m + 1\}.$$

Then, we see that $\widehat{f} > 0$ over U. We now verify that the conditions in Putinar's Positivstellensatz (Lemma 2.6) are satisfied. To see this, from Assumption 2.1, there exist $R_1, R_2 > 0$ such that

$$R_1 - \|(x, y)\|^2 = \sigma_0(x, y) - \sum_{i=1}^{s} \sigma_i(x, y) g_i(x, y) \quad \text{and} \quad R_2 - \|y\|^2 = \bar{\sigma}_0(y) - \sum_{j=1}^{r} \bar{\sigma}_j(y) h_j(y),$$

for some sums-of-squares polynomials $\sigma_0, \sigma_1, \ldots, \sigma_s \in \Sigma^2[x, y]$ and sums-of-squares polynomials $\bar{\sigma}_0, \bar{\sigma}_1, \ldots, \bar{\sigma}_r \in \Sigma^2[y]$. Letting $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_r) \in \mathbb{R}^{r+1}$, we obtain that

$$(1 + R_1 + R_2) - \|(x, y, \lambda)\|^2 \ = \ (\sigma_0(x, y) + \bar{\sigma}_0(y)) - \sum_{j=1}^{r} \bar{\sigma}_j(y) h_j(y) - \sum_{i=1}^{s} \sigma_i(x, y) g_i(x, y) + (1 - \sum_{j=0}^{r} \lambda_j^2)$$

$$= \ (\sigma_0(x, y) + \bar{\sigma}_0(y)) - \sum_{j=1}^{r} \bar{\sigma}_j(y) \widehat{G}_{s+j}(x, y, \lambda)$$

$$- \sum_{i=1}^{s} \sigma_i(x, y) \widehat{G}_i(x, y) - \widehat{H}_{r+m+1}(x, y, \lambda).$$

So, applying Putinar's Positivstellensatz (Lemma 2.6) with $w = (x, y, \lambda) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^{r+1}$, there exist sums of squares polynomials $\sigma_p \in \Sigma^2[x, y, \lambda]$, $p = 0, 1, \ldots, s + 2r + 1$ and sums-of-squares polynomials $\phi_{1q}, \phi_{2q} \in \Sigma^2[x, y, \lambda]$, $q = 1, \ldots, r + m + 1$ such that

$$\widehat{f} = \sigma_0 - \sum_{p=1}^{s+2r+1} \sigma_p \widehat{G}_p - \sum_{q=1}^{r+m+1} \phi_{1q} \widehat{H}_q + \sum_{q=1}^{r+m+1} \phi_{2q} \widehat{H}_q.$$

Let $\phi_q \in \mathbb{R}[x, y, \lambda]$ be a real polynomial defined by $\phi_q = \phi_{1q} - \phi_{2q}$, $q = 1, \ldots, r + m + 1$. Then, we have

$$f - (\mathrm{val}(\overline{P}) - \epsilon) = \sigma_0 - \sum_{p=1}^{s+2r+1} \sigma_p \widehat{G}_p - \sum_{q=1}^{r+m+1} \phi_q \widehat{H}_q.$$

Thus, there exists $k \in \mathbb{N}$, $\mathrm{val}(Q_k) \geq \mathrm{val}(\overline{P}) - \epsilon = \mathrm{val}(P) - \epsilon$. Note that, by the construction, $\mathrm{val}(Q_k) \leq \mathrm{val}(\overline{P}) = \mathrm{val}(P)$ for all $k \in \mathbb{N}$. Therefore, $\mathrm{val}(Q_k) \to \mathrm{val}(\overline{P}) = \mathrm{val}(P)$.

Remark 3.6 (Convergence to a global minimizer) It is worth noting that, in addition to the assumptions of Theorem 3.5, if we further assume that the equivalent problem $(\overline{P})$ has a unique solution say $(\bar{x}, \bar{y})$, then we can also find the global minimizer $(\bar{x}, \bar{y})$ with the help of the above sequential SDP relaxation problems. In fact, as each (Q_k) is a semidefinite programming problem, its corresponding dual problem (see [26]) can be formulated as

$$(Q_k^*) \quad \inf_{\mathbf{z} \in \mathbb{N}_{2k}^{n+m+r}} \quad L_{\mathbf{z}}(f)$$
$$\text{subject to} \quad \mathbf{M}_k(\mathbf{z}) \succeq 0, \ \mathbf{z_0} = 1,$$
$$\mathbf{M}_{k-\overline{u_p}}(\widehat{G}_p, \mathbf{z}) \succeq 0, p = 1, \ldots, s + 2r + 1$$
$$\mathbf{M}_{k-\overline{v_q}}(\widehat{H}_q, \mathbf{z}) = 0, q = 1, \ldots, r + m + 1,$$

where $\overline{u_p}$ (resp. $\overline{v_q}$) is the largest integer which is smaller than $\frac{u_p}{2}$ (resp. $\frac{v_q}{2}$), $L_{\mathbf{z}}$ is the Riesz functional defined by $L_{\mathbf{z}}(f) = \sum_\alpha f_\alpha z_\alpha$ with $f(x) = \sum_\alpha f_\alpha x^\alpha$ and, for a polynomial f, $\mathbf{M}_t(f, z)$, $t \in \mathbb{N}$ is the so-called localization matrix defined by $[\mathbf{M}_t(f, \mathbf{z})]_{\alpha,\beta} = \sum_\gamma f_\gamma z_{\alpha+\beta+\gamma}$ for all $\alpha, \beta \in \mathbb{N}_t^{n+m+r}$. From the weak duality, one has $\mathrm{val}(\overline{P}) \geq \mathrm{val}(Q_k^*) \geq \mathrm{val}(Q_k)$. Thus, the preceding theorem together with $\mathrm{val}(P) = \mathrm{val}(\overline{P})$ implies that $\mathrm{val}(Q_k^*) \to \mathrm{val}(P)$. Moreover, it was shown in [25, Theorem 4.2] that if the feasible set of the polynomial optimization problem $(\overline{P})$ has a non-empty interior, then there exists a natural number N_0 such that $\mathrm{val}(Q_k^*) = \mathrm{val}(Q_k)$ for all $k \geq N_0$.

Let $\mathbf{z}_k$ be a solution of (Q_k^*). Then, as $k \to \infty$, we have $(L_{\mathbf{z}_k}(X_1), \ldots, L_{\mathbf{z}_k}(X_n)) \to \bar{x}$, and $(L_{\mathbf{z}_k}(X_{n+1}), \ldots, L_{\mathbf{z}_k}(X_{n+m})) \to \bar{y}$, where X_i denotes the polynomial which maps each vector to its ith coordinate, $i = 1, \ldots, n + m$. The conclusion follows from [40].

The above theorem shows that one can use a sequence of semidefinite programming problems to approximate the global optimal value of a bilevel polynomial optimization problem with convex lower level problem. Moreover, under a sufficient rank condition (see [26, Theorem 5.5]), one can check whether finite convergence has occurred, i.e., by testing whether $\mathrm{val}(Q_{k_0}) = \mathrm{val}(\overline{P})$ for some $k_0 \in \mathbb{N}$. This rank condition has been implemented in the software GloptiPoly 3 [18] along with a linear algebra procedure to extract global minimizers of a polynomial optimization problem.

We now provide a simple example to illustrate how to use sequential SDP relaxations to solve the bilevel polynomial optimization problems with convex lower level problem:

Example 3.7 (Solution by sequential SDP relaxations) Consider the following simple bilevel polynomial optimization problem

$$\min_{(x,y) \in \mathbb{R}^2} \quad xy^5 - y^6$$
$$x^2 + y^2 \leq 2$$
$$y \in Y(x) := \mathrm{argmin}_{w \in \mathbb{R}} \{xw : -1 \leq w \leq 1\}.$$

Direct verification shows that there are two global solutions $(-1, 1)$ and $(1, -1)$ with global optimal value 2. We note that the lower level problem is convex and it is equivalent to the following single

level polynomial optimization problem

$$\min_{(x,y,\lambda_0,\lambda_1,\lambda_2)\in\mathbb{R}^5} \quad xy^5 - y^6$$

$$x^2 + y^2 \leq 2$$
$$\lambda_0 x + \lambda_1 - \lambda_2 = 0$$
$$\lambda_i \geq 0, \lambda_1(y-1) = 0, \lambda_2(-1-y) = 0, -1 \leq y \leq 1$$
$$\lambda_0^2 + \lambda_1^2 + \lambda_2^2 = 1.$$

Solving the converted single level polynomial optimization problem using GloptiPoly 3, the solver extracted two global solutions $(x, y, \lambda_0, \lambda_1, \lambda_2) = (-1.000, 1.000, 0.7071, 0.7071, 0)$ and $(x, y, \lambda_0, \lambda_1, \lambda_2) = (1.000, -1.000, 0.7071, 0, 0.7071)$ with the true global optimal value -2.

Remark 3.8 (Single level polynomial problem) In the case where (P) is a single level problem, Theorem 3.5 yields the known convergence result of the sequential SDP relaxation scheme (often referred to as the Lasserre hierarchy) for solving single level polynomial optimization problems [26]. Indeed, consider a (single level) polynomial optimization problem

$$(P_0) \quad \min_{x\in\mathbb{R}^n}\{f(x) : g_i(x) \leq 0, i = 1, \ldots, s\}.$$

Suppose that there exist $R > 0$ and sums of squares polynomial $\sigma_i \in \Sigma^2[x]$ such that

$$R - \|x\|^2 = \sigma_0(x) - \sum_{i=1}^s \sigma_i(x)g_i(x).$$

Let $\widehat{f}(x,y) = f(x)$, $\widehat{g}_i(x,y) = g_i(x)$, $i = 1, \ldots, s$ and $G(x,y) \equiv 0$ for all $(x,y) \in \mathbb{R}^n \times \mathbb{R}$. We note that $\mathrm{val}(P_0)$ equals the optimal value of the following bilevel polynomial optimization problem

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad \widehat{f}(x,y)$$
$$\text{subject to} \quad \widehat{g}_i(x,y) \leq 0, \ i = 1, \ldots, s,$$
$$y \in Y(x) := \mathrm{argmin}_{w\in\mathbb{R}^m}\{0 : w^2 \leq 1\}$$

Then, Theorem 3.5 yields that $\mathrm{val}(P_0) = \lim_{k\to\infty} \mathrm{val}(Q_k^0)$, where, for each k, the problems (Q_k^0) is given by

$$(Q_k^0) \quad \max_{\mu,\sigma_p} \quad \mu$$

$$\text{s.t.} \quad f - \mu = \sigma_0 - \sum_{p=1}^s \sigma_p g_p,$$

$$\sigma_p \in \Sigma^2[x], \ p = 0, 1, \ldots, s, \ \deg\sigma_0 \leq 2k, \ \deg(\sigma_p g_p) \leq 2k, p = 1, \ldots, s.$$

4 Nonconvex Lower Level Problems

In this section, we examine how to solve a bilevel polynomial optimization problem with a nonconvex lower level problem towards a global minimizer using semi-definite programming hierarchies.

Consider an ϵ-approximation of the general bilevel polynomial problem (P):

$$(P_\epsilon) \quad \min_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m} \quad f(x,y)$$

$$\text{subject to} \quad g_i(x,y) \leq 0, \ i = 1,\ldots,s,$$

$$h_j(y) \leq 0, \ j = 1,\ldots,r,$$

$$G(x,y) - \min_{w\in\mathbb{R}^m}\{G(x,w) : h_j(w) \leq 0, j = 1,\ldots,r\} \leq \epsilon.$$

The above ϵ-approximation problem plays a key role in the so-called value function approach for finding a stationary point of a bilevel programming problems, and has been studied and used widely in the literature (for example see [31, 45]). The main idea of the value function approach is to further approximate the (possibly nonsmooth and nonconvex) function $x \mapsto \min_{w\in\mathbb{R}^m}\{G(x,w) : h_j(w) \leq 0, j = 1,\ldots,r\}$ using smooth functions, and asymptotically solve the problem by using smooth local optimization techniques (such as projected gradient method (PG) and sequential quadratic programming problem (SQP) techniques). For instance, [31] use this approach together with the smoothing projected gradient method to solve the bilevel optimization problem, in the case where g_i depends on x only, $\{x \in \mathbb{R}^n : g_i(x) \leq 0\}$ and $\{w \in \mathbb{R}^m : h_j(w) \leq 0\}$ are convex sets. The algorithm only converges to a stationary point of the original problem (in a suitable sense).

We now introduce a general purpose scheme which enables us to solve (P_ϵ) towards global solutions using SDP hierarchies. The proof techniques for the convergence of this scheme (Theorem 4.6) relies on the joint-marginal method introduced in [27] to approximate a global solution of a parameterized single level polynomial optimization problem. Here, following the approach in [27], we extend the scheme and its convergence analysis to the bilevel polynomial optimization setting.

The following known simple lemma shows that the problem (P_ϵ) indeed approximates the original bilevel polynomial optimization problem as $\epsilon \to 0_+$. To do this, for $\epsilon, \delta \geq 0$, recall that $(\bar{x}, \bar{y})$ is called a δ-global solution of (P_ϵ) if $(\bar{x}, \bar{y})$ is feasible for (P_ϵ) and $f(\bar{x}, \bar{y}) \leq \mathrm{val}(P_\epsilon) + \delta$ where $\mathrm{val}(P_\epsilon)$ is the optimal value of (P_ϵ).

Lemma 4.1 (Approximation lemma cf. [32]) *Suppose that $K := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m : g_i(x,y) \leq 0\}$ and $F = \{w \in \mathbb{R}^m : h_j(w) \leq 0\}$ are compact. Let $\epsilon_k \to 0_+$ and $\delta_k \to 0_+$ as $k \to \infty$. Let $(\bar{x}_k, \bar{y}_k)$ be an δ_k-global solution for (P_{ϵ_k}). Then, $\{(\bar{x}_k, \bar{y}_k)\}_{k\in\mathbb{N}}$ is a bounded sequence and any of its cluster point $(\bar{x}, \bar{y})$ is a solution of the bilevel polynomial optimization problem (P).*

The following lemma explains the analytic property of the function $\epsilon \mapsto \mathrm{val}(P_\epsilon)$, and shows that $\mathrm{val}(P_\epsilon)$ converges to $\mathrm{val}(P)$ in the order of $O(\epsilon^{\frac{1}{q}})$ as $\epsilon \to 0_+$ for some $q \in \mathbb{N}_{>0} := \mathbb{N}\backslash\{0\}$. The proof relies on some important properties and facts on semialgebraic functions/sets and we delay the proof to the Appendix B.

Lemma 4.2 (Analytic property & approximation quality) *Suppose that Assumption 2.1 holds. Let $I \subseteq \mathbb{R}_+ := [0, +\infty)$ be a finite interval. For each $\epsilon \in I$, denote the optimal value of (P_ϵ) by $\mathrm{val}(P_\epsilon)$. Then,*

(i) *the one-dimensional function $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ is a nonincreasing, lower semicontinuous, right-continuous and semialgebraic function on I. In particular, the function $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ is continuous over I except at finitely many points.*

(ii) *There exist $q \in \mathbb{N}_{>0}$, $\epsilon_0 > 0$ and $M > 0$ such that for all $\epsilon \in [0, \epsilon_0)$*

$$\mathrm{val}(P_\epsilon) \leq \mathrm{val}(P) \leq \mathrm{val}(P_\epsilon) + M\epsilon^{\frac{1}{q}}.$$

Now, we present a simple example to illustrate the above lemma. It also implies that, in general, the function $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ can be a discontinuous semialgebraic function.

Example 4.3 Consider the bilevel polynomial optimization problem

$$(EP) \quad \min_{(x,y) \in \mathbb{R}^2} \quad y$$
$$\text{s.t.} \quad x^2 \leq 1,$$
$$y \in \mathrm{argmin}_{w \in \mathbb{R}}\{x^2 + w^2 : w^2(w^2 - 1)^2 \leq 0\}.$$

Note that $J(x) = \min_{w \in \mathbb{R}}\{x^2 + w^2 : w^2(w^2 - 1)^2 \leq 0\} = x^2$. Its ϵ-approximation problem is

$$(EP_\epsilon) \quad \min_{(x,y) \in \mathbb{R}^2} \quad y$$
$$\text{s.t.} \quad x^2 \leq 1,$$
$$y^2(y^2 - 1)^2 \leq 0, \ y^2 \leq \epsilon.$$

It can be verified that

$$\mathrm{val}(EP_\epsilon) = \begin{cases} 0, & \text{if } 0 \leq \epsilon < 1, \\ -1, & \text{if } \epsilon \geq 1. \end{cases}$$

Therefore the function $\epsilon \mapsto \mathrm{val}(EP_\epsilon)$ is nonincreasing, lower semicontinuous, right-continuous and semialgebraic on $[0, +\infty)$. Moreover, it is continuous on $[0, \epsilon_0]$ for any $\epsilon_0 < 1$ and it is discontinuous at 1.

Solving ϵ-approximation problems via sequential SDP relaxations

Here, we describe how to solve an ϵ-approximation problem via a sequence of SDP relaxation problems. One of the key steps is to construct a sequence of polynomials to approximate the optimal value function of the lower level problem $x \mapsto \min_{w \in \mathbb{R}^m}\{G(x, w) : h_j(w) \leq 0, j = 1, \ldots, r\}$. In general, the optimal value function of the lower level problem is merely a continuous function. We now recall a procedure introduced in [27] to approximate this optimal value function by a sequence of polynomials.

Recall that $K = \{x : g_i(x, y) \leq 0, i = 1, \ldots, s\}$. We denote $\mathrm{Pr}_1 K = \{x \in \mathbb{R}^n : (x, y) \in K \text{ for some } y \in \mathbb{R}^m\}$. From Assumption 2.1, K is bounded, and so, $\mathrm{Pr}_1 K$ is also bounded. Let $\mathrm{Pr}_1 K \subseteq \Omega := \{x \in \mathbb{R}^n : \|x\|_\infty \leq M\}$ for some $M > 0$. Let $\theta_l(x) = x_l^2 - M^2$, $l = 1, \ldots, n$. Then $\Omega = \{x : \theta_l(x) \leq 0, l = 1, \ldots, n\}$. Let φ be a probability Borel measure supported on Ω with uniform

distribution on X. We note that all the moments of φ over Ω denoted by $\gamma = (\gamma_\beta)$, $\beta \in \mathbb{N}^n$, defined by

$$\gamma_\beta := \int_\Omega x^\beta d\varphi(x), \ \beta \in \mathbb{N}^n,$$

can be easily computed (see [27]).

For each $k \in \mathbb{N}$ with $k \geq k_0 := \max\{\lceil \frac{\deg f}{2} \rceil, \lceil \frac{\deg h_j}{2} \rceil\}$, set $\mathbb{N}_{2k}^n := \{(\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n : \sum_{l=1}^n \alpha_l \leq 2k\}$ and consider the following optimization problem

$$\max_{\lambda, \sigma_0, \ldots, \sigma_{r+n}} \quad \sum_{\beta \in \mathbb{N}_{2k}^n} \lambda_\beta \gamma_\beta$$

$$\text{s.t.} \quad G(x, y) - \sum_{\beta \in \mathbb{N}_{2k}^n} \lambda_\beta x^\beta = \sigma_0(x, y) - \sum_{j=1}^r \sigma_j(x, y) h_j(y) - \sum_{l=1}^n \sigma_{r+l}(x, y) \theta_l(x)$$

$$\sigma_j \in \Sigma[x, y], \ j = 0, 1, \ldots, r + n \quad (4.1)$$

$$\deg \sigma_0 \leq 2k, \ \deg(\sigma_j h_j) \leq 2k, j = 1, \ldots, r, \ \deg(\sigma_{r+l} \theta_l) \leq 2k, l = 1, \ldots, n,$$

which can be reformulated as a semidefinite programming problem [27]. Then, for any feasible solution $(\lambda, \sigma_0, \sigma_1, \ldots, \sigma_{r+n})$, the polynomial $x \mapsto J_k(x) := \sum_{\beta \in \mathbb{N}_{2k}^n} \lambda_\beta x^\beta$ is of degree $2k$ and it satisfies, for all $x \in \Omega = \{x : \theta_l(x) \leq 0, l = 1, \ldots, n\}$ and $y \in F := \{w : h_j(w) \leq 0, j = 1, \ldots, r\}$,

$$G(x, y) - \sum_{\beta \in \mathbb{N}_{2k}^n} \lambda_\beta x^\beta = \sigma_0(x, y) - \sum_{j=1}^r \sigma_j(x, y) h_j(y) - \sum_{l=1}^n \sigma_{r+l}(x, y) \theta_l(x) \geq 0.$$

So, for every $k \in \mathbb{N}$, $J_k(x) \leq J(x) := \min_{w \in \mathbb{R}^m}\{G(x, w) : h_j(w) \leq 0\}$ for all $x \in \Omega$. Indeed, the next theorem shows that J_k converges to the optimal value function J on Ω, in the L_1-norm sense.

Lemma 4.4 ([27]) *Suppose that Assumption 2.1 holds. For each $k \in \mathbb{N}$, let ρ_k be the optimal value of the semidefinite programming (4.1). Let $\epsilon_k \to 0$ and let $(\lambda, \sigma_0, \sigma_1, \ldots, \sigma_{r+n})$ be an ϵ_k-solution of (4.1) in the sense that $\sum_{\beta \in \mathbb{N}_{2k}^n} \lambda_\beta \gamma_\beta \geq \rho_k - \epsilon_k$. Define $J_k \in \mathbb{R}_{2k}[x]$ by $J_k(x) = \sum_{\beta \in \mathbb{N}_{2k}^n} \lambda_\beta x^\beta$. Then, we have $J_k(x) \leq J(x)$ for all $x \in \Omega$ and*

$$\int_\Omega |J_k(x) - J(x)| d\varphi(x) \to 0 \ as \ k \to \infty.$$

We now introduce a scheme to solve the ϵ-approximation problem for arbitrary $\epsilon > 0$, using sequences of semidefinite programming relaxations.

Algorithm 4.5 (general scheme)

Step 0: Fix $\epsilon > 0$. Set $k = 1$.

Step 1: Solve the semidefinite programming problem (4.1) and obtain the $\frac{1}{k}$-solution (λ^k, σ_j^k) of (4.1). Define $J_k(x) = \sum_{\beta \in \mathbb{N}_{2k}^n} \lambda_\beta^k x^\beta$.

Step 2: Consider the following semialgebraic set

$$S_k := \{(x,y) : g_i(x,y) \leq 0, \ i = 1, \ldots, s, h_j(y) \leq 0, \ j = 1, \ldots, r, G(x,y) - J_k(x) \leq \epsilon\}.$$

If $S_k = \emptyset$, then let $k = k + 1$ and return to Step 1. Otherwise, go to Step 3.

Step 3: Solve the following polynomial optimization problem

$$(P_\epsilon^k) \quad \min_{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m} \quad f(x,y)$$
$$\text{subject to} \quad g_i(x,y) \leq 0, \ i = 1, \ldots, s,$$
$$h_j(y) \leq 0, \ j = 1, \ldots, r,$$
$$G(x,y) - J_k(x) \leq \epsilon.$$

Step 4: Let $v_\epsilon^k = \min_{1 \leq i \leq k} \text{val}(P_\epsilon^i)$. Update $k = k + 1$. Go back to Step 1.

Before we establish the convergence of this procedure, let us comment that the feasibility problem of the semialgebraic set in Step 2 can be tested by a sequence of SDP relaxations via the Positivstellensatz. This was explained in [38] and was implemented in the matlab toolbox SOSTOOLS. As explained before, Step 3 can also be accomplished by solving a sequence of SDP relaxations.

Let us show that there exists a finite number k_0 such that $S_{k_0} \neq \emptyset$, and so, Algorithm 4.5 is well-defined.

Lemma 4.5 Let $\epsilon > 0$. Consider the problem (P_ϵ) and Algorithm 4.5. Let $K = \{(x,y) : g_i(x,y) \leq 0, i = 1, \ldots, s\}$ and $F = \{w : h_j(w) \leq 0, j = 1, \ldots, r\}$. Suppose that Assumption 2.1 holds and $\text{cl}\big(\text{int}(K \cap (\mathbb{R}^n \times F))\big) = K \cap (\mathbb{R}^n \times F)$. Then, there exists a finite number k_0 such that $S_{k_0} \neq \emptyset$ in Step 2 of Algorithm 4.5.

Proof. Note from Corollary 2.5 that a global minimizer $(\bar{x}, \bar{y})$ of (P) exists. In particular, the set $D_0 := \{(x,y) \in K \cap (\mathbb{R}^n \times F) : G(x,y) - J(x) < \epsilon\}$ is an nonempty set as $(\bar{x}, \bar{y}) \in D_0$. Noting from our assumption, we have $\text{cl}\big(\text{int}(K \cap (\mathbb{R}^n \times F))\big) = K \cap (\mathbb{R}^n \times F)$. This together with the fact that $\{(x,y) : G(x,y) - J(x) < \epsilon\}$ is an open set (as the optimal value function of the lower level problem $J(x)$ is continuous) gives us that

$$\widetilde{D} := \{(x,y) \in \text{int}(K \cap (\mathbb{R}^n \times F)) : G(x,y) - J(x) < \epsilon\}$$

is a nonempty open set. Define $D := \text{Pr}_1 \widetilde{D} = \{x \in \mathbb{R}^n : (x,y) \in \widetilde{D} \text{ for some } y \in \mathbb{R}^m\}$. Then, D is also a nonempty open set. Note from Lemma 4.4 that J_k converges to J in $L^1(\Omega, \varphi)$-norm. Hence J_k converges to J almost everywhere on Ω. As $\varphi(\Omega) < +\infty$, the classical Egorov's theorem[2] implies that

---

[2]The Egorov's theorem [13, Theorem 2.2] states that: for a measure space (Ω, φ), let f_k be a sequence of functions on Ω. Suppose that Ω is of finite φ-measure and $\{f_k\}$ converges φ-almost everywhere on Ω to a limit function f. Then, there exists a subsequence l_k such that f_{l_k} converges to f almost uniformly in the sense that, for every $\epsilon > 0$, there exists a measurable subset A of Ω such that $\varphi(A) < \epsilon$, and $\{f_{l_k}\}$ converges to f uniformly on the relative complement $\Omega \backslash A$.

there exists a subsequence l_k such that J_{l_k} converges to J φ-almost uniformly on Ω. So, there exists a Borel set A with $\varphi(A) < \frac{\eta}{2}$ with $\eta := \varphi(D) > 0$ such that

$$J_{l_k} \to J \text{ uniformly over } \Omega \backslash A.$$

We observe that $(\Omega \backslash A) \cap D \neq \emptyset$ (Otherwise, as $D \subseteq \mathrm{Pr}_1 K \subseteq \Omega$, we have $D \subseteq A$. This implies that $\eta = \varphi(D) \leq \varphi(A) = \eta/2$ which is impossible as $\eta > 0$). Let $x_0 \in (\Omega \backslash A) \cap D$. Then, we have $J_{l_k}(x_0) \to J(x_0)$ and there exists $y_0 \in \mathbb{R}^m$ such that $y_0 \in F$, $G(x_0, y_0) - J(x_0) < \epsilon$. In particular, for all k large, $(x_0, y_0) \in S_{l_k}$. Therefore, $S_{l_k} \neq \emptyset$ for all large k, and so, the conclusion follows.

Remark 4.6 The fact that L_1-convergence implies the almost-uniform convergence can also be seen by using Theorem 2.5.1 (L_1-convergence implies convergence in measure) and Theorem 2.5.3 (convergence in measure implies almost-uniform convergence for a subsequence) of [2, Page 92-93] without requiring the measure of Ω to be finite.

We note that the condition "$\mathrm{cl}\big(\mathrm{int}(K \cap (\mathbb{R}^n \times F))\big) = K \cap (\mathbb{R}^n \times F)$" holds when $C := K \cap (\mathbb{R}^n \times F)$ is a finite union of closed convex sets C_i with $\mathrm{int} C_i \neq \emptyset$. Moreover, if the set C is of the form $\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m : G_i(x,y) \leq 0, i = 1, \ldots, l\}$ for some polynomials G_i, $i = 1, \ldots, l$ and $l \in \mathbb{N}$, then the above condition also holds if the commonly used Mangasarian-Fromovitz constraint qualification [34] is satisfied for any $(x,y) \in C$.

We are now ready to state the convergence theorem of the proposed Algorithm 4.5. The proof of it is quite technical and so it is given later in Appendix C.

Theorem 4.7 (General bilevel problem (P): Convergence theorem) *Let $\epsilon > 0$ and consider problem (P_ϵ). Let v_ϵ^k be generated by Algorithm 4.5. Let $K = \{(x,y) : g_i(x,y) \leq 0, i = 1, \ldots, s\}$ and $F = \{w : h_j(w) \leq 0, j = 1, \ldots, r\}$. Suppose that Assumption 2.1 holds and $\mathrm{cl}\big(\mathrm{int}(K \cap (\mathbb{R}^n \times F))\big) = K \cap (\mathbb{R}^n \times F)$. Then,*

(i) *$v_\epsilon^k \to v_\epsilon$ as $k \to \infty$ where $\mathrm{val}(P_\epsilon) \leq v_\epsilon \leq \lim_{\delta \to \epsilon^-} \mathrm{val}(P_\delta)$. In particular, for almost every ϵ, $v_\epsilon^k \to \mathrm{val}(P_\epsilon)$ in the sense that, for all finite intervals $I \subseteq \mathbb{R}_+$, $v_\epsilon = \mathrm{val}(P_\epsilon)$ for all $\epsilon \in I$ except at finitely many points.*

(ii) *There exists $\epsilon_0 > 0$ such that, for all $\epsilon \in (0, \epsilon_0)$, $v_\epsilon^k \to \mathrm{val}(P_\epsilon)$ as $k \to \infty$. Moreover, let $\delta_k \downarrow 0$. Let $v_\epsilon^k = \min_{1 \leq i \leq k} \mathrm{val}(P_\epsilon^i) = \mathrm{val}(P_\epsilon^{i_k})$ and let (x_k, y_k) be a δ_k-solution of $(P_\epsilon^{i_k})$. Then, $\{(x_k, y_k)\}$ is a bounded sequence and any cluster point $(\widehat{x}, \widehat{y})$ of (x_k, y_k) is a global minimizer of (P_ϵ) for all $\epsilon \in (0, \epsilon_0)$.*

We now illustrate how our general scheme can lead to solving a bilevel programming problem with a nonconvex lower level problem towards a global solution. This is done by applying our scheme to a known test problem of the bilevel programming literature.

Example 4.8 (Illustration of our approximation scheme) Consider the following bilevel optimization test problem (for example see [31, 37])

$$\min_{(x,y)\in\mathbb{R}^2} \quad x + y$$

$$\text{subject to} \quad x \in [-1,1], y \in \operatorname{argmin}_{w\in[-1,1]}\{\frac{xw^2}{2} - \frac{w^3}{3}\}.$$

Let $Y(x) := \operatorname{argmin}_{w\in[-1,1]}\{\frac{xw^2}{2} - \frac{w^3}{3}\}$. Clearly, the lower level problem is nonconvex and all the conditions in Theorem 4.7 are satisfied. The optimal value function of the lower level problem is given by

$$J(x) = \min_{w\in[-1,1]}\{\frac{xw^2}{2} - \frac{w^3}{3}\} = \begin{cases} 0, & \text{if} \quad x \in [\frac{2}{3}, 1], \\ \frac{x}{2} - \frac{1}{3}, & \text{if} \quad x \in [-1, \frac{2}{3}), \end{cases}$$

and the solution set of the lower level problem $Y(x)$ can be formulated as

$$Y(x) = \begin{cases} \{0\}, & \text{if} \quad x \in (\frac{2}{3}, 1], \\ \{0, 1\}, & \text{if} \quad x = \frac{2}{3}, \\ \{1\}, & \text{if} \quad x \in [-1, \frac{2}{3}). \end{cases}$$

It is easy to check that the true (unique) global minimizer is $(\bar{x}, \bar{y}) = (-1, 1)^T$ and the true global optimal value is 0.

Now, for $k = 3$, using GloptiPoly 3, we obtain a degree $2k(= 6)$ polynomial approximation of $J(x)$ which is

$$J_3(x) = -0.3338 + 0.5011 * x + 0.0098 * x^2 - 0.0032 * x^3 - 0.0696 * x^4 - 0.1012 * x^5 - 0.0432 * x^6.$$

The following figure depicts the graph of the functions J_3 and J, where the red curve is the graph of the function J and the blue curve is the graph of the degree 6 polynomial J_3. From the graph, we can see that $J_3 \leq J$ over the interval $[-1, 1]$ and provides a reasonably good approximation of the piecewise differentiable (and so, non-polynomial) function $J(x)$.

Figure 1: $J(x)$ and its degree-6 underestimation in Example 4.8
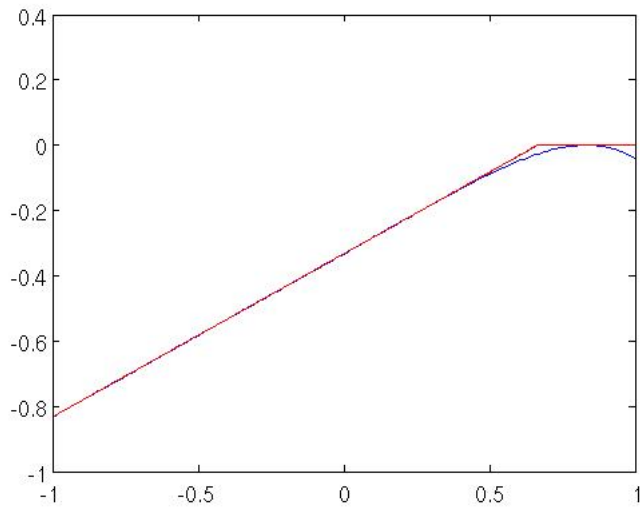
Setting $\epsilon = 0.001$ and solving the following polynomial optimization problem

$$\min_{(x,y)\in\mathbb{R}^2} \quad x + y$$
$$\text{subject to} \quad x \in [-1, 1],$$
$$y \in [-1, 1],$$
$$\frac{xy^2}{2} - \frac{y^3}{3} - J_3(x) \leq 0.001,$$

with GloptiPoly 3, the solver returns the point $(x, y) = (-1.0000, 0.9996)$ with its associated function value $-4.1680e-04$, which is a reasonably good approximation of the true global minimizer and global optimal value of the bilevel programming problem.

Remark 4.9 (Further extensions of the approach) Although we presented our approach for a class of bilevel problems where the constraints of the lower-level problem are independent of the upper-level decision variable x, our approach may be extended to solve the following more general bilevel polynomial optimization problem:

$$(GP) \quad \min_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad f(x, y)$$
$$\text{subject to} \quad g_i(x, y) \leq 0, \ i = 1, \ldots, s,$$
$$y \in Y(x) := \operatorname{argmin}_{w\in\mathbb{R}^m}\{G(x, w) : h_j(x, w) \leq 0, j = 1, \ldots, r\},$$

where the constraints of the lower level problem are allowed to depend on x. In this case, we can construct a sequence of semidefinite programming relaxation for finding a global minimizer and a global minimum of its ϵ-approximation problem and similar convergence results of the scheme can be achieved under an additional technical assumption that the optimal value function of the lower level problem $J(x) := \min_{w\in\mathbb{R}^m}\{G(x, w) : h_j(x, w) \leq 0, j = 1, \ldots, r\}$ is continuous. However, we wish to note that, for the problem (P) discussed in this paper (that is, h_j are independent of x), this condition is automatically satisfied. On the other hand, in general, this condition may fail for the general problem (GP) even when $n = m = 1$. We provide a simple example to illustrate this. Consider the following bilevel programming problem

$$\min_{x\in\mathbb{R}, y\in\mathbb{R}} \quad x^2 + y^2$$
$$\text{subject to} \quad 0 \leq x \leq 1,$$
$$y \in Y(x) := \operatorname{argmin}_{w\in\mathbb{R}}\{(x - w)^2 : x^2 - w^2 \leq 0, w(w - 1) \leq 0, -w(w - 1) \leq 0\}.$$

It can be directly verified that the optimal value function of the lower level problem is given by

$$J(x) := \min_{w\in\mathbb{R}}\{(x - w)^2 : x^2 - w^2 \leq 0, w(w - 1) = 0\} = \begin{cases} 0, & \text{if} \quad x = 0, \\ (x - 1)^2, & \text{if} \quad x \in (0, 1]. \end{cases}$$

and is discontinuous at $x = 0$.

5 Numerical Examples

In this Section, we apply our schemes to solve some bilevel optimization test problems available in the literature and present their results. We conducted the numerical tests on a computer with a 2.8 GHz Intel Core i7 and 8 GB RAM, equipped with Matlab 7.14 (R2012a). We solved bilevel polynomial problems with convex as well as non-convex lower-level problems, where the lower level problems are independent of the upper level decision variables.

We first present results for the following bilevel problems with a convex lower level problem. We note that all the assumptions of Theorem 3.5 are satisfied by these bilevel problems with a convex lower level problem.

Example 5.1 Consider the following bilevel polynomial problem [16]

$$
\begin{cases}
\min_{x,y \in \mathbb{R}} & (x-3)^2 + (y-2)^2 \\
\text{s.t.} & -2x + y - 1 \leq 0 \\
& x - 2y + 2 \leq 0 \\
& x + 2y - 14 \leq 0 \\
& 0 \leq x \leq 8 \\
& 0 \leq y \leq 10 \\
& y \in \mathrm{argmin}_{w \in \mathbb{R}}\{(w-5)^2 : w \in [0,10]\}.
\end{cases}
$$

This problem has a unique global minimizer $(x^*, y^*) = (3, 5)$ and the optimal value $f^* = 9$.

Example 5.2 Consider the following bilevel polynomial problem [16]

$$
\begin{cases}
\min_{x,y \in \mathbb{R}} & -(4x-3)y + (2x+1) \\
\text{s.t.} & 0 \leq x \leq 1 \\
& 0 \leq y \leq 1 \\
& y \in \mathrm{argmin}_{w \in \mathbb{R}}\{-(1-4x)w - (2x+2) : w \in [0,1]\}.
\end{cases}
$$

This problem has a unique global minimizer $(x^*, y^*) = (0.25, 0)$ and the optimal value $f^* = 1.5$.

We first transformed the problems in Example 5.1 and Example 5.2 into equivalent single-level non-convex polynomial optimization problems as proposed in Section 3. Then, we used GloptiPoly 3 [18] and the SDP solver Sedumi [42] to solve the transformed polynomial optimization problems. For these two problems, the second relaxation problem (that is, problem (Q_2)) of the SDP approximation scheme (3.12) returns a solution which agrees with the true solution.

The following table summarizes the results of bilevel problems with a convex lower level problem where (x^*, y^*) and f^* denote the true global minimizer and the true optimal value respectively, (x, y) and f denote the computed minimizer and the computed optimal value respectively and CPU time represents the CPU time (in seconds) used to solve the problems.

Table 1: Convex Lower-Level Problems

Test Problems	Known optimal solutions	Computed solutions
Example 5.1	$(x^*, y^*) = (3, 5)$ $f^* = 9$	$(x, y) = (3.0000, 5.0000)$ $f = 9.0000$ CPU time=0.2511
Example 5.2	$(x^*, y^*) = (0.25, 0)$ $f^* = 1.5$	$(x, y) = (0.2500, 0.0000)$ $f = 1.5000$ CPU time=0.1957

We now solve the following bilevel problems with a non-convex lower level problem. Again, all the assumptions in Theorem 4.7 are satisfied by these bilevel problems with a nonconvex lower level problem.

Example 5.3 Consider the following bilevel polynomial problem [36]

$$\begin{cases} \min_{x,y \in \mathbb{R}} & x \\ \text{s.t.} & -x + y \leq 0 \\ & -10 \leq x \leq 10 \\ & -1 \leq y \leq 1 \\ & y \in \text{argmin}_{w \in \mathbb{R}} \{w^3 : w \in [-1, 1]\}. \end{cases}$$

This problem has a unique global minimizer $(x^*, y^*) = (-1, -1)$ with the optimal value $f^* = -1$.

Example 5.4 Consider the following bilevel polynomial problem [36]

$$\begin{cases} \min_{x,y \in \mathbb{R}} & 2x + y \\ \text{s.t.} & -1 \leq x \leq 1 \\ & -1 \leq y \leq 1 \\ & y \in \text{argmin}_{w \in \mathbb{R}} \{-\frac{1}{2}xw^2 - \frac{1}{4}w^4 : w \in [-1, 1]\}. \end{cases}$$

This problem has two global minimizers $(x_1^*, y_1^*) = (-1, 0)$ and $(x_2^*, y_2^*) = (-1/2, -1)$ with the optimal value $f^* = -2$.

Example 5.5 Consider the following bilevel polynomial problem [36]

$$\begin{cases} \min_{x,y \in \mathbb{R}} & y \\ \text{s.t.} & 0.1 \leq x \leq 1 \\ & -1 \leq y \leq 1 \\ & y \in \text{argmin}_{w \in \mathbb{R}} \{x(16w^4 + 2w^3 + 8w^2 + \frac{3}{2}w + \frac{1}{2}) : w \in [-1, 1]\}. \end{cases}$$

This problem has infinitely many global minimizers $(x^*, y^*) = (a, 0.5)$ for any $a \in [0.1, 1]$ with the optimal value $f^* = 0.5$.

Example 5.6 Consider the following bilevel polynomial problem [36]

$$\begin{cases} \min_{x,y \in \mathbb{R}} & -x + xy + 10y^2 \\ \text{s.t.} & -1 \leq x \leq 1 \\ & -1 \leq y \leq 1 \\ & y \in \text{argmin}_{w \in \mathbb{R}} \{-xw^2 + w^4/2 : w \in [-1, 1]\}. \end{cases}$$

This problem has a unique global minimizer $(x^*, y^*) = (0, 0)$ with the optimal value $f^* = 0$.

We solved these four problems by using the approximation scheme proposed in Section 4 implemented via the software GloptiPoly 3 and the SDP solver Sedumi. For detailed illustration of how the scheme is implemented, see Example 4.8. The numerical results are summarized in the following table. Note that deg denotes the maximum degree of the polynomial underestimation used in a subproblem of our scheme.

Table 2: Non-Convex Lower-Level Problems

Test Problems	Known optimal solutions	Computed solutions
Example 5.3	$(x^*, y^*) = (-1, -1)$ $f^* = -1$	$(x, y) = (-1.0000, -1.0000)$ $f = -1.0000$ CPU time=1.0746 deg=12
Example 5.4	$(x^*, y^*) = (-1, 0)$ or $(-1/2, -1)$ $f^* = -2$	$(x, y) = (-0.9991, -0.0020)$ $f = -2.0002$ CPU time=5.1432 deg=14
Example 5.5	$(x^*, y^*) = (a, 0.5)$ for all $a \in [0.1, 1]$ $f^* = 0.5$	$(x, y) = (0.2299, 0.4990)$ $f = 0.4990$ CPU time=6.8819 deg=12
Example 5.6	$(x^*, y^*) = (0, 0)$ $f^* = 0$	$(x, y) = (0.0034, -0.0002)$ $f = -0.0034$ CPU time= 0.8844 deg=10

6 Conclusion and Further Research

We established that a global minimizer and the global minimum of a bilevel polynomial optimization problem can be found by way of solving a sequence of semidefinite programming relaxations. We first considered a bilevel polynomial optimization problem where the lower level problem is a convex problem. In this case, we proved that the values of the sequence of relaxation problems converge to the global optimal value of the bilevel problem under a mild assumption. This shows that a global solution can simply be found by first transforming the bilevel problem into an equivalent single-level polynomial problem and then solving the resulting single-level problem by the standard sequential SDP relaxations used in the polynomial optimization [26].

We then examined a general bilevel polynomial optimization problem with a not necessarily convex lower-level problem. We established that the global optimal value in this case can be found by way of solving a new sequential semidefinite programming relaxation problems based on the joint-marginal approach proposed in [27]. This was done by using a sequence of semidefinite programming relaxations

of its ϵ-approximation problem under the standard Assumption 2.1 of polynomial optimization, where $\epsilon > 0$ is smaller than a positive threshold.

The convergence of the proposed semidefinite programming approximation scheme relies on Assumption 2.1 which requires that the feasible set of the bilevel problem is bounded. The proposed scheme can also be extended to cover possible unbounded feasible sets by exploiting coercivity of the objective function of the upper/lower level problem as in our recent papers [21, 22, 23] where the convergence of the sequence of semidefinite programming relaxations was established for polynomial optimization problems with unbounded feasible sets.

Our bilevel problem, in the present paper, represents the so-called optimistic approach to the leader and follower's game in which the follower is assumed to be co-operative and so, the leader can choose the solution with the lowest cost. The pessimistic approach assumes that the follower may not be co-operative and hence the leader will need to prepare for the worst cost. Mathematically, the following bilevel problem represents the pessimistic approach:

$$\min_{x \in \mathbb{R}^n} \quad \max_{y \in Y(x)} f(x,y)$$
$$\text{subject to} \quad g_i(x) \leq 0, \ i = 1, \ldots, s,$$

where $Y(x) := \operatorname{argmin}_{w \in \mathbb{R}^m}\{G(x,w) : h_j(w) \leq 0, j = 1, \ldots, r\}$. A possible method to solving this bilevel problem is to construct a polynomial approximation for the optimal value of the problem $x \mapsto \max_{y \in Y(x)} f(x,y)$ using the joint marginal approach of [27] and then design a semidefinite programming approximation method that is similar to the scheme studied in the present paper. This would be an interesting topic for future research.

Appendix A: Semi-algebraic functions and sets

In this appendix, we summarize some of the important properties of semi-algebraic functions which are used in this paper (see [5]).

(i) Finite union (resp. intersection) of semi-algebraic sets is semi-algebraic. The Cartesian product (resp. complement, closure) of semi-algebraic sets is semi-algebraic.

(ii) If f, g are semi-algebraic functions on $\mathbb{R}^n$ and $\lambda \in \mathbb{R}$, then $f + g$, fg and λf are all semi-algebraic functions.

(iii) If f is a semi-algebraic function on $\mathbb{R}^n$ and $\lambda \in \mathbb{R}$, then $\{x : f(x) \leq \lambda\}$ (resp. $\{x : f(x) \leq \lambda\}$, $\{x : f(x) < \lambda\}$ and $\{x : f(x) = \lambda\}$ are all semi-algebraic sets.

(iv) A composition of semi-algebraic maps is a semi-algebraic map.

(v) The image and inverse image of a semi-algebraic set under a semi-algebraic map are semi-algebraic sets. In particular, the projection of a semi-algebraic set is still a semi-algebraic set.

(vi) If S is a compact semi-algebraic set in $\mathbb{R}^m$ and $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a real polynomial, then the function $x \mapsto \min\limits_{y \in \mathbb{R}^m} \{f(x,y) \ : \ y \in S\}$, is also semi-algebraic.

Remark 6.1 If $A \in \mathbb{R}^n, B \in \mathbb{R}^m$ and $S \in \mathbb{R}^n \times \mathbb{R}^m$ are semi-algebraic sets, then we see that $U :=$ $\{x \in A \ : (x,y) \in S, \forall \ y \in B\}$ is also a semi-algebraic set. To see this, from property (v), we see that $\{x \in A \ : \ \exists y \in B, (x,y) \in S\}$ is semialgebraic. As the complement of U is the union of the complement of A and the set $\{x \in A \ : \ \exists y \in B, (x,y) \notin S\}$, it follows that the complement of U is semi-algebraic by property (i). Thus, U is also semi-algebraic by property (i). In general, if we have a finite collection of semi-algebraic sets, then any set obtained from them by a finite chain of quantifiers is also semi-algebraic.

For a one-dimensional semi-algebraic function, we have further the following properties:

Lemma 6.2 (Monotonicity Theorem [15]) *Let f be a semi-algebraic function f on $\mathbb{R}$. Let $a, b \in \mathbb{R}$ with $a < b$. Then, there exists a finite subdivision $a = t_0 < t_1 < \ldots < t_k = b$ such that, on each interval (t_i, t_{i+1}), f is continuous and f either takes a constant value or is strictly monotone.*

Lemma 6.3 (Growth Dichotomy Lemma [35]) *Let $\epsilon_0 > 0$ and let f be a continuous semi-algebraic function f on $[0, \epsilon_0]$ with $f(0) = 0$. Then either f takes a constant value 0 over $[0, \epsilon_0]$ or there exist constants $c \neq 0$ and $p, q \in \mathbb{N}_{>0}$ such that $f(t) = c\, t^{\frac{p}{q}} + o(t^{\frac{p}{q}})$ as $t \to 0_+$.*

Appendix B: Proof of Lemma 4.2

Proof. [Proof of (i)] From the definition of (P_ϵ), it is clear that if $0 \le \epsilon_1 \le \epsilon_2$, then $\mathrm{val}(P_{\epsilon_1}) \ge \mathrm{val}(P_{\epsilon_2})$. Using a similar method of proof as in Lemma 4.1, one can show that $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ is a lower semicontinuous function. Now, let $\epsilon_k \to \epsilon_+$. Then, from the lower semicontinuity,

$$\liminf_{k \to \infty} \mathrm{val}(P_{\epsilon_k}) \ge \mathrm{val}(P_\epsilon).$$

This together with the fact that $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ is nonincreasing shows that $\lim_{k \to \infty} \mathrm{val}(P_{\epsilon_k}) = (P_\epsilon)$. So, this function is right continuous.

Let $J(x) := \min_w \{G(x,w) \ : \ h_j(w) \le 0, j = 1, \ldots, r\}$. By property (vi), J is a semialgebraic function. Let

$$
\begin{aligned}
X \ &:= \ \{(\epsilon, x, y) \in [0, +\infty) \times \mathbb{R}^n \times \mathbb{R}^m \ : \ g_i(x,y) \le 0, i = 1, \ldots, s, \\
&\qquad\qquad h_j(y) \le 0, j = 1, \ldots, r, G(x,y) - J(x) \le \epsilon\}
\end{aligned}
$$

and

$$Y \ := \ \{(\epsilon, x, y) \in X : f(x,y) \le f(a,b), \ \forall (\epsilon, a, b) \in X\}.$$

We can verify that X and Y are semialgebraic sets by properties (ii), (iii) and Remark 6.1. Further, by definition, the graph of the function $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ is given by $\{(\epsilon, f(x,y)) \ : \ (\epsilon, x, y) \in Y\}$. Clearly,

this set is the image of the set Y under the semialgebraic map $(\epsilon, x, y) \mapsto (\epsilon, f(x, y))$, and hence it is a semialgebraic set by property (v). Thus, $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ is a semi-algebraic function on $[0, +\infty)$.

Fix a finite interval $I \subseteq [0, +\infty)$. As $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ is a semialgebraic function, it follows from Lemma 6.2 that the function $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ is continuous over I except at finitely many points.

[Proof of (ii)] Fix a finite interval $I \subseteq [0, +\infty)$. Denote the discontinuity points of $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ on I by $\{\epsilon_1, \ldots, \epsilon_l\}$ for some $l \in \mathbb{N}$. Clearly, $\inf_{1 \leq i \leq l} \epsilon_i > 0$ as $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ is right continuous. Let $\bar{\epsilon} = \min_{1 \leq i \leq l} \{\epsilon_i\}/2 > 0$. Then, $\epsilon \mapsto \mathrm{val}(P_\epsilon)$ is continuous over $[0, \bar{\epsilon}]$. Applying Lemma 6.3 with f replaced by $\epsilon \mapsto \mathrm{val}(P_\epsilon) - \mathrm{val}(P)$ on $[0, \bar{\epsilon}]$, we see that there exist constants $c > 0$, $p, q \in \mathbb{N}_{>0}$ and $\epsilon_0 \in (0, 1)$ with $\epsilon_0 < \bar{\epsilon}$ such that

$$\mathrm{val}(P_\epsilon) \leq \mathrm{val}(P) + c\,\epsilon^{\frac{p}{q}} \leq \mathrm{val}(P) + c\,\epsilon^{\frac{1}{q}}, \quad \text{for all} \quad \epsilon \in [0, \epsilon_0], \tag{6.2}$$

where the last inequality holds as $0 < \epsilon \leq \epsilon_0 < 1$. This, together with the nonincreasing property of $\epsilon \mapsto \mathrm{val}(P_\epsilon)$, yields the last assertion.

Appendix C: Proof of Theorem 4.7 (Convergence of Algorithm 4.5)

Proof. [Proof of (i)] Recall from Lemma 4.4 that $J_k(x) \leq J(x)$ for all $k \in \mathbb{N}$ and for all $x \in \Omega$. So, $\mathrm{val}(P_\epsilon^k) \geq \mathrm{val}(P_\epsilon)$ for all $k \in \mathbb{N}$. This implies that $v_\epsilon^k \geq \mathrm{val}(P_\epsilon)$ for all $k \in \mathbb{N}$. As v_ϵ^k is a non-increasing sequence which is bounded below, $\lim_{k \to \infty} v_\epsilon^k$ exists. Let $v_\epsilon = \lim_{k \to \infty} v_\epsilon^k$. Then,

$$v_\epsilon \geq \mathrm{val}(P_\epsilon). \tag{6.3}$$

Let $\delta \in (0, \epsilon)$ and consider problem $(P_{\epsilon - \delta})$. By Assumption 2.1, K and F are compact sets. From the nonsmooth Danskin Theorem (see [8, Page 86]), we see that the optimal value function of the lower level problem $J(x) := \min_{w \in \mathbb{R}^m} \{G(x, w) : h_j(w) \leq 0, j = 1, \ldots, r\}$ is locally Lipschitz (and so, is continuous). Thus, a global minimizer of $(P_{\epsilon - \delta})$ exists. Let $(\bar{x}, \bar{y})$ be a global minimizer of $(P_{\epsilon - \delta})$. The set $D_0 := \{(x, y) \in K \cap (\mathbb{R}^n \times F) : G(x, y) - J(x) < \epsilon, \ f(x, y) < f(\bar{x}, \bar{y}) + \delta\}$ is a nonempty set as $(\bar{x}, \bar{y}) \in D_0$. Moreover, from our assumption we have $\mathrm{cl}\big(\mathrm{int}(K \cap (\mathbb{R}^n \times F))\big) = K \cap (\mathbb{R}^n \times F)$. This together with the fact that $\{(x, y) : G(x, y) - J(x) < \epsilon, \ f(x, y) < f(\bar{x}, \bar{y}) + \delta\}$ is an open set gives us that

$$\widetilde{D} := \{(x, y) \in \mathrm{int}(K \cap (\mathbb{R}^n \times F)) : G(x, y) - J(x) < \epsilon \text{ and } f(x, y) < f(\bar{x}, \bar{y}) + \delta\}$$

is a nonempty open set. So, $D := \mathrm{Pr}_1 \widetilde{D} = \{x \in \mathbb{R}^n : (x, y) \in \widetilde{D} \text{ for some } y \in \mathbb{R}^m\}$ is also a nonempty open set. Since J_k converges to J in $L^1(\Omega, \varphi)$-norm, J_k converges to J on Ω almost everywhere. Moreover, as $\varphi(\Omega) < +\infty$, the classical Egorov's theorem guarantees that there exists a subsequence l_k such that J_{l_k} converges to J φ-almost uniformly on Ω. So, there exists a Borel set A with $\varphi(A) < \frac{\eta}{2}$ with $\eta := \varphi(D) > 0$ such that $J_{l_k} \to J$ uniformly over $\Omega \backslash A$. As in the proof of Lemma 4.5, we can show that $(\Omega \backslash A) \cap D \neq \emptyset$. Let $x_0 \in (\Omega \backslash A) \cap D$. Then, we have $J_{l_k}(x_0) \to J(x_0)$ and there exists $y_0 \in F$ such that $G(x_0, y_0) - J(x_0) < \epsilon$ and $f(x_0, y_0) < f(\bar{x}, \bar{y}) + \delta$. So, for all large k, $G(x_0, y_0) - J_{l_k}(x_0) < \epsilon$. Thus, for all large k, (x_0, y_0) is feasible for $(P_\epsilon^{l_k})$ and

$$v_\epsilon^{l_k} \leq \mathrm{val}(P_\epsilon^{l_k}) \leq f(x_0, y_0) < f(\bar{x}, \bar{y}) + \delta = \mathrm{val}(P_{\epsilon - \delta}) + \delta.$$

Letting $k \to \infty$, we obtain that $v_\epsilon = \lim_{k\to\infty} v_\epsilon^{l_k} \leq \text{val}(P_{\epsilon-\delta}) + \delta$. Letting $\delta \to 0^+$, we see that

$$v_\epsilon \leq \lim_{\delta\to\epsilon^-} \text{val}(P_\delta). \tag{6.4}$$

Therefore, the inequality $\text{val}(P_\epsilon) \leq v_\epsilon \leq \lim_{\delta\to\epsilon^-} \text{val}(P_\delta)$ follows by combining (6.3) and (6.4). To see the second assertion in (i), we only need to notice from Lemma 4.2(i) that $\epsilon \mapsto \text{val}(P_\epsilon)$ is continuous except finitely many points over a finite interval I.

[Proof of (ii)] From Lemma 4.2(ii), we see that there exists $\epsilon_0 > 0$ such that $\epsilon \mapsto \text{val}(P_\epsilon)$ is continuous over $(0, \epsilon_0)$. Thus, from (i), we have $v_\epsilon^k \to \text{val}(P_\epsilon)$ for all $\epsilon \in (0, \epsilon_0)$. Now, fix any $\epsilon \in (0, \epsilon_0)$, Let $\delta_k \downarrow 0$ as $k \to \infty$. Let $v_\epsilon^k = \min_{1\leq i\leq k} \text{val}(P_\epsilon^i) = \text{val}(P_\epsilon^{i_k})$ and let (x_k, y_k) be a δ_k-solution of $(P_\epsilon^{i_k})$. Then, $\{(x_k, y_k)\} \subseteq K \cap (\mathbb{R}^n \times F)$. As K and F are compact, we see that $\{(x_k, y_k)\}$ is a bounded sequence. Let $(\widehat{x}, \widehat{y})$ be a cluster point of $\{(x_k, y_k)\}$. Clearly, $(\widehat{x}, \widehat{y}) \in K \cap (\mathbb{R}^n \times F)$. As $J_k \leq J$ on Ω for all $k \in \mathbb{N}$, $x_k \in \text{Pr}_1 K \subseteq \Omega$ and (x_k, y_k) is feasible for $(P_\epsilon^{i_k})$. Hence, for each $k \in \mathbb{N}$

$$G(x_k, y_k) - J(x_k) \leq G(x_k, y_k) - J_{i_k}(x_k) \leq \epsilon.$$

Passing to the limit and noting that J is continuous, we get that $G(\widehat{x}, \widehat{y}) - J(\widehat{x}) \leq \epsilon$. So, $(\widehat{x}, \widehat{y})$ is feasible for (P_ϵ). Finally, since $v_\epsilon^k \to \text{val}(P_\epsilon)$, it follows that

$$f(\widehat{x}, \widehat{y}) = \lim_{k\to\infty} f(x_k, y_k) \leq \lim_{k\to\infty} (v_\epsilon^k + \delta_k) = \text{val}(P_\epsilon)$$

and $(\widehat{x}, \widehat{y})$ is a global minimizer of (P_ϵ).

References

[1] C. Audet, P. Hansen, B. Jaumard, and G. Savard, Links between linear bilevel and mixed 0-1 programming problems, 93 (1997), 273-300

[2] R.B. Ash, *Real Analysis & Probability*, Academic Press, (1972).

[3] J.F. Bard, *Practical Bilevel Optimization: Algorithms and Applications*, Kluwer Academic Publications, Dorchester, Netherlands, 1998.

[4] R. Benedetto and J.-J. Risker, *Real algebraic and semialgebraic sets,* Actuality Mathematics, Hermann, 1991.

[5] E. Bierstone and P. Milman, *Semianalytic and subanalytic sets,* Inst. Hautes Études Sci. Publ. Math. 67 (1988), 5-42.

[6] J. Bochnak, M. Coste, and M.-F. Roy, *Real algebraic geometry,* Erg. Math. Grenzgeb., **36**, Springer, 1998.

[7] J. Bolte, H. Attouch and B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel method, *Mathematical Programming*, vol. 137, no. 1-2, (2013), 91-129.

[8] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York (1983).

[9] A. Colson, P. Marcotte, G. Savard, An overview of bilevel programming, *Annals of Operations Research*, 153 (2007) 235-256.

[10] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, 2002.

[11] S. Dempe, Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints, *Optimization*, 52 (2003) 333-359.

[12] S. Dempe and J. Dutta, Is bilevel programming a special case of a mathematical program with complementarity constraints?, *Mathematical Programming* 131 (2012), 37-48.

[13] E. DiBenedetto, *Real Analysis*, Birkhuser Advanced Texts, pp. x+485, (2002),

[14] N. Dinh and V. Jeyakumar, Farkas' lemma: Three decades of generalizations for mathematical optimization, *TOP*, (2014), 22, 1-22.

[15] L. van den Dries and C. Miller, *Geometric categories and o-minimal structures,* Duke Math. J., **84** (1996), 497–540.

[16] C.A. Floudas, M. Pardalos, C.S. Adjiman, W.R. Esposito, Z.H. Gumus, S.T. Harding, J.L. Klepeis, C.A. Meyer and C.A. Schweiger, Handbook of test problems in local and global optimization. Nonconvex Optimization and its Applications, 33. Kluwer Academic Publishers, Dordrecht, 1999.

[17] Z.H. Gümüs and C.A. Floudas, Global Optimization of Nonlinear Bilevel Programming Problems, *Journal of Global Optimization*, 20 (2001) 1-31.

[18] D. Henrion, J. B. Lasserre and J. Loefberg, GloptiPoly 3: moments, optimization and semidefinite programming. *Optimization Methods and Software*, 24 (2009), No. 4-5, 761-779.

[19] H. Gfrerer, Hölder continuity of solutions of perturbed optimization problems under Mangasarian-Fromowitz constraint qualification, *Parametric Optimization and Related Topics* (J. Guddat et al., ed.), Akademie-Verlag, Berlin, 1987, pp. 113-127.

[20] V. Jeyakumar, G.M. Lee and G. Li, Alternative theorems for quadratic inequality systems and global quadratic optimization, *SIAM Journal on Optimization*, 20 (2009), no. 2, 983-1001.

[21] V. Jeyakumar, S. Kim, G. M. Lee and G. Li, Semidefinite programming relaxation methods for global optimization problems with sparse polynomials and unbounded semialgebraic feasible sets, *Journal of Global Optimization*, (2015), DOI 10.1007/s10898-015-0356-6.

[22] V. Jeyakumar, J.B. Lasserre and G. Li, On polynomial optimization over non-compact semi-algebraic sets, *Journal of Optimization Theory and its Applications*, 163 (2014), 707-718.

[23] V. Jeyakumar, T.S. Pham and G. Li, Convergence of the Lasserre hierarchy of SDP relaxations for convex polynomial programs without compactness, *Operations Research Letters*, 42 (2014), 34-40.

[24] S. Kim, M. Kojima, H. Waki, and M. Yamashita, *Algorithm 920: SFSDP: a Sparse Version of Full SemiDefinite Programming relaxation for sensor network localization problems*, ACM Tran. Math. Soft., 38(2012) 1-19.

[25] J.B. Lasserre, Global optimization with polynomials and the problem of moments, *SIAM Journal on Optimization* 11 (2000), 796-817.

[26] J.B. Lasserre, *Moments, Positive Polynomials and Their Applications*, Imperial College Press, 2009.

[27] J.B. Lasserre, A "joint+marginal" approach to parametric polynomial optimization. *SIAM Journal on Optimization*, 20 (2010), 1995-2022.

[28] J. B. Lasserre, On representations of the feasible set in convex optimization. *Optimization Letters* (40) (2010), 1-5.

[29] J.B. Lasserre, Min-max and robust polynomial optimization. *Journal of Global Optimization* 51 (2011), 1–10

[30] G. Li, B.S. Mordukhovich and T.S. Pham, New error bounds for nonconvex polynomial systems with applications to Holderian stability in optimization and spectral theory of tensors, *Mathematical Programming,* 153, no. 2 (2015), 333–362.

[31] G.H. Lin, M.W. Xu and J. J. Ye, On solving simple bilevel programs with a nonconvex lower level program, *Mathematical Programming*, series A, 144 (2014), 277-305.

[32] P. Loridan and J. Morgan, New results on approximate solutions in two-level optimization, *Optimization*, 20 (1989), 819-836.

[33] R. Lucchetti, F. Mignanego, and G. Pieri, Existence theorem of equilibrium points in Stackelberg games with constraints, *Optimization*, 18 (1987), 857-866.

[34] O. L. Mangasarian and S. Fromovitz, The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and its Application* 17, (1967), 37-47.

[35] C. Miller, *Exponentiation is hard to avoid, Proceedings of the American Mathematical Society*, 122 (1994), 257-259.

[36] A. Mitsos and P. I. Barton. A test set for bilevel programs. 2006, avaliable at http://www.researchgate.net/publication/228455291.

[37] A. Mitsos, P. Lemonidis and P. Barton, Global solution of bilevel programs with a nonconvex inner program, *J. Global Optim.*, 42(2008), 475-513.

[38] P.A. Parrilo, S. Lall, Semidefinite programming relaxations and algebraic optimization in Control, *European Journal of Control*, 9 (2003), 307-321.

[39] M. Putinar, Positive polynomials on compact semi-algebraic sets, *Indiana University Mathematics Journal* 41 (1993), 49-95.

[40] M. Schweighofer, Optimization of polynomials on compact semialgebraic sets, *SIAM Journal on Optimization*, 15, No. 3, 805-825 (2005).

[41] S.M. Robinson, Generalized equations and their solutions, part II: Applications to nonlinear programming, *Mathematical Programming Study* 19, 200-221, (1982).

[42] J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11/12, 625-653, (1999).

[43] L.N. Vicente, P.H. Calamai, Bilevel and multilevel programming: a bibliography review, *Journal of Global Optimization*, 5 (1994) 291-306.

[44] W. Wiesemann, A. Tsoukalas, P.M. Kleniati and B. Rustem, Pessimistic bilevel optimization. *SIAM Journal on Optimization,* 23 (2013), no. 1, 353-380.

[45] J.J. Ye and D.L. Zhu, New necessary optimality conditions for bilevel programs by combining MPEC and the value function approach, *SIAM Journal on Optimization*, 20(2010), 1885-1905.