The Standard Pessimistic Bilevel Problem

Lorenzo Lampariello[1], Simone Sagratella[2], and Oliver Stein[3]

[1]Department of Business Studies, Roma Tre University, Italy, lorenzo.lampariello@uniroma3.it
[2]Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza University of Rome, Italy, sagratella@diag.uniroma1.it
[3]Institute for Operations Research, Karlsruhe Institute of Technology (KIT), Germany, stein@kit.edu

June 30, 2018

Abstract

Pessimistic bilevel optimization problems, as optimistic ones, possess a structure involving three interrelated optimization problems. Moreover, their finite infima are only attained under strong conditions. We address these difficulties within a framework of moderate assumptions and a perturbation approach which allow us to approximate such finite infima arbitrarily well by minimal values of a sequence of solvable single-level problems.

To this end, as already done for optimistic problems, for the first time in the literature we introduce the standard version of the pessimistic bilevel problem. For its algorithmic treatment, we reformulate it as a standard optimistic bilevel program with a two follower Nash game in the lower level. The latter lower level game, in turn, is replaced by its Karush-Kuhn-Tucker conditions, resulting in a single-level mathematical program with complementarity constraints.

We prove that the perturbed pessimistic bilevel problem, its standard version, the two follower game as well as the mathematical program with complementarity constraints are equivalent with respect to their global minimal points. We study the more intricate connections between their local minimal points in detail. As an illustration, we numerically solve a regulator problem from economics for different values of the perturbation parameters.

Keywords: Pessimistic bilevel programming, standard optimistic bilevel problems, generalized Nash equilibrium problem, mathematical program with complementarity constraints

Preliminary citation: Optimization Online, Preprint ID 2018-06-6688, 2018

The work of the authors was supported by the MIUR-DAAD Joint Mobility Program under Grant 34793/57396680.

1 Introduction

In bilevel optimization an upper level objective function $F(x, y)$ is minimized with respect to $x \in X \subseteq \mathbb{R}^n$, while $y \in \mathbb{R}^p$ is required to be in the optimal point set of some lower level optimization problem which is parametrized by x. We write this lower level parametric optimization problem as

$$\begin{array}{ll} \underset{w}{\text{minimize}} & f(x, w) \\ \text{s.t.} & w \in M(x), \end{array} \tag{1}$$

where $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ is continuously differentiable on its domain and $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ is a set-valued mapping. The x-dependent optimal point set of this lower level problem is denoted by $S(x) \subseteq \mathbb{R}^p$.

In applications, the bilevel problem often is to be solved by a leading agent. The leader passes the decision x to the follower, who then solves the lower level problem (1) and returns some $y \in S(x)$ to the leader. Clearly, if the lower problem was uniquely solvable for each $x \in X$, we would obtain the unique response $S(x) = \{y(x)\}$, and the leader's problem could be reduced to the minimization of $F(x, y(x))$ over X.

The situation becomes more intricate if the lower level problem is not uniquely solvable for each $x \in X$. Depending on the application, the leader then may or may not have the information whether the follower returns the best response y from $S(x)$ in terms of the leader's objective $F(x, y)$. If such a best response is returned, the leader minimizes the function $\inf_{y \in S(x)} F(x, y)$ over $x \in X$ which yields the so-called original optimistic version of the bilevel program. If, on the other hand, the follower is an adversary of the leader, or if the follower is indifferent and the leader wishes to hedge against the worst possible response of the follower, the leader's appropriate objective function becomes $\sup_{y \in S(x)} F(x, y)$. Essentially, both optimistic and pessimistic bilevel problems possess a structure involving three interrelated optimization problems.

The pessimistic version of the bilevel program is the subject of the present paper. To distinguish it from a later reformulation let us call it the original Pessimistic Bilevel Problem and formulate it as

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \sup_{y \in S(x)} F(x, y) \\ \text{s.t.} & x \in X, \end{array} \tag{PBP}$$

where $F : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ is continuously differentiable on its domain, and $X \subseteq \mathbb{R}^n$ is nonempty and compact. More blanket assumptions on the upper and lower level problems will be introduced in section 2.

A comprehensive introduction to bilevel programming as well as its optimistic and pessimistic aspects is given in [4]. While the pessimistic view is more realistic in many practical applications, the optimistic view has received considerably more attention in the literature, at least from an algorithmic standpoint (see the references in [4] and, for more recent developments, [9]). Essentially, the reason is that, resorting to the so-called *standard* version of the optimistic problem [29], the successive minimization of F over x and y may be combined to a joint minimization of F over the points (x, y) with $x \in X$ and $y \in S(x)$. Hence, in the optimistic version, the structure involving three interrelated optimization problems gives rise to a two-level problem.

The theoretical as well as algorithmic treatment of the pessimistic version is more intricate, arguably because of the impossibility to reduce the structure involving three interrelated optimization problems to a two-level problem. For systematic comparisons of the optimistic and pessimistic views we refer to [18]. Necessary optimality conditions for pessimistic bilevel problems were recently obtained in [7, 8] under additional Lipschitz assumptions (which do not necessarily hold in our framework), but without the application to an algorithmic solution procedure.

To the best of our knowledge, so far the only algorithmic approach to a version of pessimistic bilevel problems is presented in [28]. There the assumptions on problem data are independent of ours, since convexity of the defining functions (see section 2) is not necessarily assumed, but the lower level problem's feasible set M needs to be constant instead of x-dependent. Furthermore, the discretization approach proposed in [28] fundamentally differs from the reformulation approach in the present paper.

As mentioned before, in section 2 we introduce a framework of moderate blanket assumptions which particularly guarantee the solvability of the lower level problem (1) for any $x \in X$ as well as the attainment of $\sup_{y \in S(x)} F(x, y)$ as a maximal value. However, an example in section 3 illustrates that these moderate assumptions do *not* entail the attainment of a finite infimum of the upper level objective function $\max_{y \in S(x)} F(x, y)$ over the nonempty and compact set X, that is, solvability of (PBP) is not guaranteed.

While a straightforward remedy would be the additional condition of inner semicontinuity for the set-valued mapping S, such an assumption may not be considered moderate and would restrain the general applicability of a solution approach. Instead, we replace the optimal point set $S(x)$ by the set of ε-optimal points $S_\varepsilon(x)$ and consider the corresponding perturbed problem (PBP_ε). This approach is also used in [28] and goes back at least to [17].

We prove that (PBP_ε) is solvable for any $\varepsilon > 0$ and that, for any zero sequence of perturbation parameters, the optimal points of (PBP_ε) approximate an optimal point of the closure of the original (PBP). In particular, the optimal values of the perturbed problems approximate the finite infimum of the original (PBP), even if it is not attained. Related results are shown in [18, 28], but with a different technique of proof and merely for a constant set-valued mapping M (which would not cover the application we present in section 7).

Starting from section 4 we develop a strategy to solve the perturbed problem (PBP_ε) algorithmically. As commonly done for the optimistic bilevel problem, we introduce the *standard* version, denoted by ($\mathrm{SPBP}_\varepsilon$), also for the pessimistic bilevel problem (PBP_ε), thus filling a gap in the literature. The ($\mathrm{SPBP}_\varepsilon$) turns out to be a three-level problem. We prove an equivalence between the global minimal points of (PBP_ε) and ($\mathrm{SPBP}_\varepsilon$) and we discuss the relations between their local optimal points.

In section 5 we show how the three-level problem ($\mathrm{SPBP}_\varepsilon$) may be replaced by a two-level problem (MFG_ε), where the lower level consists of a generalized Nash game, and how to substitute the latter by a single-level mathematical program with complementarity constraints ($\mathrm{MPCC}_\varepsilon$). For all these problems we prove equivalences between their global minimal points and discuss the relations between their local optimal points. Section 6 briefly surveys the shown interrelations between the four considered problems ((PBP_ε), ($\mathrm{SPBP}_\varepsilon$), ($\mathrm{MFG}_\varepsilon$), and ($\mathrm{MPCC}_\varepsilon$)).

In section 7 we present an example from economics, where a regulator takes care of consumers wellness by bounding production quantities of firms which share common resources. Our numerical results serve as proof of concept for our general framework and illustrate the theoretical approximation results.

Some final remarks are given in section 8. An appendix contains the proof of the main approximation theorem, some complementary material, as well as an alternative reformulation of the problem ($\mathrm{SPBP}_\varepsilon$) as a (single-level) three player generalized Nash equilibrium problem.

This latter approach is, however, not in the focus of the paper since we consider it interesting only for a subclass of pessimistic bilevel problems.

In the paper we employ standard notation. However, for the reader's convenience, we remark that, considering $h : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$, we denote by $\nabla_1 h(x, y)$ the gradient of $h(\bullet, y)$ evaluated at x, while by $\nabla_2 h(x, y)$ the gradient of $h(x, \bullet)$ evaluated at y. As for the definitions of semicontinuity and other properties of single and set-valued mappings such as lower semicontinuity, inner and outer semicontinuity, as well as continuity, and local boundedness, we refer the reader to [21].

2 Blanket assumptions on the original problem

In addition to the assumption of a nonempty and compact set X and the smoothness requirements for the upper and lower level objective functions F and f from section 1, we impose the following additional blanket assumptions on (PBP) throughout this article:

(i) $X \subseteq \operatorname{dom} M$;

(ii) M is continuous relative to X at any point in X, locally bounded and convex-valued at any $x \in X$;

(iii) $F(\bullet, y)$ is convex on $\mathbb{R}^n$ for every $y \in M(X)$;

(iv) $F(x, \bullet)$ is concave on $\mathbb{R}^p$ for every $x \in X$;

(v) $f(x, \bullet)$ is convex on $\mathbb{R}^p$ for every $x \in X$.

These blanket assumptions imply, in particular, that the set $M(x)$ is nonempty and compact for every $x \in X$, so that we obtain $X \subseteq \operatorname{dom} S$ by the Weierstrass Theorem. The optimal value function

$$\varphi(x) \triangleq \min_y \{f(x, y) \mid y \in M(x)\}$$

of the lower level problem (1) thus is real-valued. In view of assumptions (i), (ii) and (v), the function φ [2, Theorem 4.3.3] is continuous relative to X at any point in X. We shall extensively resort to the following equivalent representation of S:

$$S(x) = \{y \in M(x) \mid f(x, y) \leq \varphi(x)\}.$$

The following result implies that the set $S(x)$ is closed for each $x \in X$. It follows immediately from, e.g., [2, Theorem 3.1.1].

Proposition 2.1 *Under our blanket assumptions, the set-valued mapping S is outer semicontinuous relative to X at any point in X.*

Due to assumption (ii) and $S(x) \subseteq M(x)$ for all $x \in X$, the set-valued mapping S is locally bounded. In particular, for each $x \in X$ the set $S(x)$ is nonempty and compact, so that the supremum in the definition of the objective function in (PBP) is attained. Consequently, problem (PBP) may be written as the minimization of the function

$$\psi(x) \triangleq \max_{y \in S(x)} F(x, y)$$

over the set X.

Remark 2.2 Let M possess the functional description

$$M(x) \triangleq \{y \in \mathbb{R}^p \,|\, h(x,y) \le 0\},$$

where $h : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^q$ is continuously differentiable on its domain and such that, for every $x \in X$, $h(x, \bullet)$ is convex and there exists some $\widehat{y}(x) \in \mathbb{R}^p$ for which $h(x, \widehat{y}(x)) < 0$, that is, Slater's constraint qualification holds for $M(x)$. In this case, regarding assumption (ii), the convex-valuedness of M is guaranteed by the convexity of $h(x, \bullet)$ at any $x \in X$; the outer semicontinuity of M relative to X at any point in X follows [2, Theorem 3.1.1] from the continuity of h; finally, the inner semicontinuity of M relative to X at any point in X is due [2, Theorem 3.1.6] to the continuity of h, the convexity of $h(x, \bullet)$ and the Slater's constraint qualification for $M(x)$ at every $x \in X$. We shall resort to such a functional description of M only where necessary. □

3 Solvability and approximation of (PBP)

Solvability of problem (PBP) is a sensitive issue (see, e.g., [4, 19, 28, 29], examples and references therein). The following simple example clearly shows that our blanket assumptions do not imply solvability of (PBP).

Example 3.1 Consider $F(x,y) = x^2 + y$, $f(x,y) = xy$, $X = [-1,1]$, $M(x) = [-1,1]$ for every $x \in X$. All the blanket assumptions hold in this case. On the interval $[-1,0]$ one computes $\psi(x) = x^2 + 1$, while on $(0,1]$ one obtains $\psi(x) = x^2 - 1$. The problem (PBP) hence does not possess an optimal point. Notice that this happens because ψ is not lower semicontinuous at 0. □

Hence, under our assumptions, the function ψ is not even lower semicontinuous. For this reason, we do not consider stationary points for (PBP): assumptions that guarantee stationarity conditions to be necessary for local optimality of (PBP) (see [7, 8]) do not hold in our framework.

At first glance, a natural remedy to obtain solvability would be the requirement of further assumptions. It is well known (or, alternatively, can also be derived by [2, Theorem 4.2.2]) that, in our framework, a sufficient condition for an optimal point of (PBP) to exist is the inner semicontinuity of the set-valued mapping S, as this implies the lower semicontinuity of the objective function ψ of (PBP). However, while S is outer semicontinuous (see Proposition 2.1) under standard mild conditions, requiring S to be inner semicontinuous is quite a strong assumption. Few classes of problems exist, e.g. those for which S is fixed, such that the inner semicontinuity of S may be considered a weak assumption.

Remark 3.2 Whenever S is a fixed set-valued mapping, things become simpler since minimizing ψ over X turns out to be a standard convex-concave saddle point problem. In order to address the latter mathematical program, one can employ a projected subgradient-type algorithm. This procedure aims at minimizing on X the nonsmooth function $\psi(x) = \max_{y \in S} F(x,y)$, which turns out to be convex under our initial assumptions. More specifically, the scheme simply consists in the alternate solution of three (smooth) convex optimization problems: given an initial point $x^0 \in X$ and a suitable positive stepsize γ^k, one has to successively compute

$$w^k \in \arg \min_{w \in M(x^k)} f(x^k, w), \quad y^k \in \arg \max_{y \in M(x^k)} \left\{ F(x^k, y) \,|\, f(x^k, y) \le f(x^k, w^k) \right\}$$

and, in turn, the subgradient-like iteration

$$x^{k+1} = P_X\left(x^k - \gamma^k \nabla_1 F(x^k, y^k)\right).$$

Note that the projection on X, which is denoted by $P_X(\bullet)$, is well-defined since X is a nonempty closed convex set. We also observe that the vector $\nabla_1 F(x^k, y^k)$ belongs to the subdifferential of ψ at x^k, that is, $\partial\psi(x^k) = \text{co}\left\{\nabla_1 F(x^k, y) \,|\, y \in \arg\max_{z\in S} F(x^k, z)\right\}$, thanks to the Danskin's Theorem.

As for the convergence properties of this scheme, we remark that one can employ for γ^k different step size strategies in order to make the overall procedure work properly. Specifically, relying on a diminishing step size approach, we can choose (γ^k) such that

$$\gamma^k \to 0, \ \sum_{k=0}^{\infty} \gamma^k = +\infty, \ \sum_{k=0}^{\infty} (\gamma^k)^2 < \infty.$$

In the light of the above considerations, one can prove (see, e.g. [3, Proposition 3.2.6]) that the sequence (x^k) generated by previous scheme converges to a minimum of ψ on X.

An alternative way to deal with (PBP) when S is a fixed set-valued mapping is given in appendix C. □

Alternatively, one may disregard the solvability of (PBP) and concentrate on computing the infimum of ψ on X. To this end (cf. [4]), it is expedient to define the closure $\text{cl}\,\psi$ of ψ on X, that is, the (unique) function whose epigraph on X coincides with the closure $\text{cl}(\text{epi}\,\psi)$ of the epigraph of ψ on X: thus,

$$\text{cl}\,\psi(\widetilde{x}) = \liminf_{x \underset{X}{\to} \widetilde{x}} \psi(x). \tag{2}$$

From this construction one clearly has $\text{cl}\,\psi \le \psi$ on X and, by its lower semicontinuity, the function $\text{cl}\,\psi$ attains its infimum on the nonempty and compact set X. Moreover, it is not hard to see that any global minimal point of ψ on X also is a global minimal point of $\text{cl}\,\psi$ on X. While the contrary does not hold, we still have

$$\min_{x\in X} \text{cl}\,\psi(x) = \inf_{x\in X} \psi(x). \tag{3}$$

In fact, letting x^* be a global minimal point for $\text{cl}\,\psi$ on X, on the one hand, we have for every $x \in X$

$$\text{cl}\,\psi(x^*) \le \text{cl}\,\psi(x) \le \psi(x),$$

and thus $\text{cl}\,\psi(x^*)$ is a lower bound for ψ on X; on the other hand, recalling the property (2) for the value $\text{cl}\,\psi(x^*)$, for every $\tau > 0$ some $\widetilde{x} \in X$ exists with $\text{cl}\,\psi(x^*) + \tau > \psi(\widetilde{x})$, meaning that $\text{cl}\,\psi(x^*)$ is the greatest lower bound for ψ on X, and hence it is equal to $\inf_{x\in X} \psi(x)$.

Example 3.3 Going back to Example 3.1, one can easily see that $\text{cl}\,\psi(x) = x^2+1$ on $[-1, 0)$, $\text{cl}\,\psi(x) = x^2 - 1$ on $[0, 1]$: thus, while $x^* = 0$ is a global minimal point for $\text{cl}\,\psi$ on X, it is not optimal for ψ on X but, nonetheless (cf. (3)) $\text{cl}\,\psi(0) = \inf_{x\in[-1,1]} \psi(x)$. □

In the light of the considerations above, whether (PBP) is solvable or not, one can always reasonably aim at minimizing $\text{cl}\,\psi$ on X and, thus, at obtaining the infimum of ψ over X. It remains to clarify, however, how the abstract definition of the function $\text{cl}\,\psi$ may be exploited algorithmically.

To this end, even in the case of a non inner semicontinuous mapping S one can resort, along the lines put forward in, e.g., [17, 18, 28], to a more conservative but tractable version of (PBP) in which the set of lower level optimal points $S(x)$ is replaced by its enlargement

$$S_\varepsilon(x) \triangleq \{y \in M(x) \,|\, f(x,y) \le \varphi(x) + \varepsilon\}\,, \tag{4}$$

that is the set of lower level ε-optimal points, where ε denotes a positive constant. We emphasize that in [28] the independence of the set-valued mapping $M(x)$ from the parameter x is fundamental, while we consider the more general case of our blanket assumption (ii).

Proposition 3.4 *Under our blanket assumptions, for every $\varepsilon > 0$ the set-valued mapping S_ε is continuous relative to X and locally bounded at any point in X.*

Proof. Since the value function φ and the feasible set-valued mapping M are continuous relative to X at every $x \in X$, the outer semicontinuity of S_ε easily follows from [2, Theorem 3.1.1]; its local boundedness is due to the local boundedness of M and $S_\varepsilon(x) \subseteq M(x)$ for all $x \in X$. Moreover, using the description

$$S_\varepsilon(x) = \{y \in \mathbb{R}^p \,|\, f(x,y) \le \varphi(x) + \varepsilon\} \cap M(x)$$

and $X \subseteq \operatorname{dom} S$, a point $\tilde{y}$ exists such that $\tilde{y} \in M(x)$ and $f(x,\tilde{y}) < \varphi(x) + \varepsilon$. In turn, by the convexity of $f(x,\bullet)$ and $M(x)$, and the continuity of f, φ and M relative to X at any $x \in X$, [2, Theorem 3.1.6] S_ε is also inner semicontinuous relative to X at any $x \in X$. $\square$

As a consequence [2, Theorem 4.3.3], the perturbed optimal value function

$$\psi_\varepsilon(x) \triangleq \max_{y \in S_\varepsilon(x)} F(x,y)$$

is continuous relative to X at any point $x \in X$, and the approximating problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \psi_\varepsilon(x) \\ \text{s.t.} & x \in X \end{array} \tag{PBP$_\varepsilon$}$$

is solvable for each $\varepsilon > 0$.

Given a monotonically decreasing sequence $\varepsilon_k \searrow 0$ of perturbation parameters and corresponding optimal points x^k of (PBP$_{\varepsilon_k}$), $k \in \mathbb{N}$, the compactness of X implies the existence of an accumulation point x^* of the sequence (x^k) in X. The following result states that any such point $x^\star$ minimizes the function $\operatorname{cl}\psi$ over X so that, in view of (3), we obtain an approximation procedure for the infimum of ψ over X.

Theorem 3.5 *With $\varepsilon_k \searrow 0$, let x^k be an optimal point of (PBP$_{\varepsilon_k}$) for each $k \in \mathbb{N}$. Then the sequence (x^k) possesses an accumulation point, and any such point x^* is a global minimal point of $\operatorname{cl}\psi$ on X with minimal value $\operatorname{cl}\psi(x^*) = \inf_{x\in X} \psi(x)$.*

To keep this paper self-contained, we give a proof of Theorem 3.5 in appendix A. As mentioned in section 1, related results are shown in [18, 28], but with a different technique of proof and merely for a constant set-valued mapping M. Moreover, [17] studies a related question on the convergence of lower level-least norm ε-optimal points.

4 The standard pessimistic version

The question arises on how to solve (PBP_ε) for fixed values of ε. To this end, as commonly done for optimistic bilevel problems, we introduce, for the first time also for pessimistic bilevel programs, the (perturbed) Standard Pessimistic Bilevel Problem

$$\begin{aligned} \underset{x,y}{\text{minimize}} \quad & F(x,y) \\ \text{s.t.} \quad & x \in X \\ & y \in R_\varepsilon(x), \end{aligned} \qquad (\text{SPBP}_\varepsilon)$$

where the set-valued mapping $R_\varepsilon : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ describes the optimal point set of the following lower level hierarchical (parametric) optimization problem:

$$\begin{aligned} \underset{w}{\text{maximize}} \quad & F(x,w) \\ \text{s.t.} \quad & w \in S_\varepsilon(x). \end{aligned} \qquad (5)$$

We observe that:

- (SPBP_0) is the standard version of (PBP);
- the standard version is obtained by moving the "difficulties" in the objective function of the original problem to the constraints level;
- the bilevel problem (5), for every fixed x, is pure hierarchical [16] (or simple [5]), i.e., $S_\varepsilon(x)$ is a fixed set.

The x-part of any global solution for $(\text{SPBP}_\varepsilon)$ is a global optimal point for (PBP_ε) and viceversa.

Proposition 4.1 *The following relations hold:*

(i) if $\widehat{x} \in X$ is a global optimal point for (PBP_ε), then, for any $\widehat{y} \in R_\varepsilon(\widehat{x})$, the pair $(\widehat{x}, \widehat{y})$ is a global optimal point for $(\text{SPBP}_\varepsilon)$;

(ii) if $(\widehat{x}, \widehat{y})$ is a global optimal point for $(\text{SPBP}_\varepsilon)$, then $\widehat{x}$ is a global optimal point for (PBP_ε).

Proof. A straightforward reformulation of (PBP_ε) is the two-level generalized semi-infinite program (cf. [26, 28])

$$\begin{aligned} \underset{x,\tau}{\text{minimize}} \quad & \tau \\ \text{s.t.} \quad & \tau \geq F(x,y) \quad \forall\, y \in S_\varepsilon(x) \\ & x \in X. \end{aligned}$$

The claim follows from [27, Lemma 5.1.1], observing that the construction from [27] actually yields the presence of the auxiliary variable τ which turns out to be unnecessary here. In fact, at the upper level, minimizing τ with the constraint $\tau \geq F(x,y)$ is equivalent to minimizing $F(x,y)$; at the lower level, we have, as objective function, $\tau - F(x,w)$, but, in this case, the lower level optimal point set does not depend on the constant τ. □

As for the local solutions, we establish the following result.

Proposition 4.2 *The following relations hold:*

(i) if $\widehat{x} \in X$ is a local optimal point for (PBP$_\varepsilon$)*, then, for any $\widehat{y} \in R_\varepsilon(\widehat{x})$, the pair $(\widehat{x}, \widehat{y})$ is a local optimal point for* (SPBP$_\varepsilon$)*;*

(ii) let either

(a) $(\widehat{x}, \widehat{y})$ be a local optimal point for (SPBP$_\varepsilon$)*, with R_ε inner semicontinuous relative to X at $\widehat{x}$, or*

(b) $(\widehat{x}, y)$ be a local optimal point for (SPBP$_\varepsilon$) *for every $y \in R_\varepsilon(\widehat{x})$,*

then $\widehat{x}$ is a local optimal point for (PBP$_\varepsilon$)*.*

Proof. (i) Suppose by contradiction that $(\widehat{x}, \widehat{y})$ is not a local optimal point for (SPBP$_\varepsilon$), i.e. there exists a sequence (x^k, y^k), with $x^k \in X$ and $y^k \in R_\varepsilon(x^k)$, such that $(x^k, y^k) \to (\widehat{x}, \widehat{y})$ and $F(x^k, y^k) < F(\widehat{x}, \widehat{y}) = \psi_\varepsilon(\widehat{x})$ for every k. But, since $\psi_\varepsilon(x^k) = F(x^k, y^k)$ and $\psi_\varepsilon(\widehat{x}) = F(\widehat{x}, \widehat{y})$, we have $\psi_\varepsilon(x^k) < \psi_\varepsilon(\widehat{x})$, with $x^k \in X$, in contradiction with the local optimality of $\widehat{x}$ for (PBP$_\varepsilon$).

(ii) Suppose by contradiction that $\widehat{x}$ is not a local optimal point for (PBP$_\varepsilon$), i.e. there exists a sequence (x^k), with $x^k \in X$, such that $x^k \to \widehat{x}$ and $\psi_\varepsilon(x^k) < \psi_\varepsilon(\widehat{x})$, for every k.

In case (a), due to the inner semicontinuity property for R_ε, a sequence (y^k) exists such that $y^k \to \widehat{y}$ and $y^k \in R_\varepsilon(x^k)$. Since $\psi_\varepsilon(x^k) = F(x^k, y^k)$ and $\psi_\varepsilon(\widehat{x}) = F(\widehat{x}, \widehat{y})$, we have $F(x^k, y^k) < F(\widehat{x}, \widehat{y})$, with $x^k \in X$ and $y^k \in R_\varepsilon(x^k)$, in contradiction with the local optimality of $(\widehat{x}, \widehat{y})$ for (SPBP$_\varepsilon$).

In case (b), we preliminarily observe that R_ε is outer semicontinuous relative to X (by the continuity of ψ_ε) and locally bounded at $\widehat{x}$. Hence, a sequence (y^k), with $y^k \in R_\varepsilon(x^k)$, exists such that $y^k \to \bar{y}$ for some $\bar{y} \in R_\varepsilon(\widehat{x})$. Since $\psi_\varepsilon(x^k) = F(x^k, y^k)$ and $\psi_\varepsilon(\widehat{x}) = F(\widehat{x}, \bar{y})$, we have $F(x^k, y^k) < F(\widehat{x}, \bar{y})$, with $x^k \in X$ and $y^k \in R_\varepsilon(x^k)$, in contradiction with the local optimality of $(\widehat{x}, \bar{y})$ for (SPBP$_\varepsilon$). □

We remark that, in general, as for (ii) in Proposition 4.2, neither condition (a) implies (b), nor (b) entails (a). Clearly, whenever $R_\varepsilon(\widehat{x}) = \{\widehat{y}\}$ is a singleton, then both (a) and (b) hold: a sufficient condition for this to be true is the strict concavity of $F(\widehat{x}, \bullet)$; in turn, (PBP$_\varepsilon$) and (SPBP$_\varepsilon$) turn out to be equivalent also in a local sense. We remark that conditions (a) or (b) in Proposition 4.2 (ii) cannot be disregarded as the following example clearly shows.

Example 4.3 Consider an instance of (PBP$_\varepsilon$) in which $x, y \in \mathbb{R}$, $F(x, y) = xy$, $X = [-1, 0]$, $M(x) = [1, 2]$ and $f(x, y) = 0$; observing that $S_\varepsilon(x) = M(x)$ for every x and $\varepsilon > 0$, the x-part of the local optimal point $x = 0$, $y = 2$ for (SPBP$_\varepsilon$) is easily seen to be not local optimal for (PBP$_\varepsilon$). In fact, regarding condition (a) in Proposition 4.2, R_ε is not inner semicontinuous at 0 relative to X; moreover, as for condition (b) in Proposition 4.2, we note that the point $x = 0$, $y = 1$, with $1 \in R_\varepsilon(0)$, is not a local optimal point for (SPBP$_\varepsilon$).

Some considerations about the standard version (SPBP$_\varepsilon$) are in order. The (SPBP$_\varepsilon$) is a *three-level problem*. Roughly speaking, we identify three agents: a *leader* who solves problem (SPBP$_\varepsilon$), an *intermediate* (leader's rival) agent who takes care of (5) and, as usual, a *follower* who addresses program (1). Also, (SPBP$_\varepsilon$) may be interpreted as a standard optimistic bilevel program in which the lower level is in turn a pure hierarchical problem. In this sense, the introduction of a third level and, possibly, of further local solutions is the price to pay for transforming the pessimistic view of the original (PBP_ε) into the optimistic view of ($\mathrm{SPBP}_\varepsilon$). This, in turn, allows us to suitably refer to the machinery from the optimistic bilevel literature in order to cope with (PBP_ε).

5 How to address ($\mathrm{SPBP}_\varepsilon$)

In this section, we aim at addressing the (three-level) ($\mathrm{SPBP}_\varepsilon$) by means of a (single-level) suitable mathematical program with complementarity constraints. This is accomplished in two steps. First (see subsection 5.1), by referring to [16, Theorem 3.2 and Remark 3.2], we replace the pure hierarchical problem (5) by a Generalized Nash Equilibrium Problem (GNEP for short, see e.g., [1, 10, 11, 12, 14, 22, 23, 24, 25]) with two players, which is parametric in x. Then (see subsection 5.2), we deal with the resulting problem by replacing the lower level game with its Karush-Kuhn-Tucker (KKT) conditions, thus obtaining a corresponding mathematical program with complementarity constraints. Finally, a possible way to cope with the latter problem is given in subsection 5.3.

5.1 The two-level version, with two followers, of ($\mathrm{SPBP}_\varepsilon$)

We introduce the Multi Follower Game with two followers (see, e.g. [20]), that is a (standard optimistic) bilevel program in which the lower level is a (parametric) GNEP with two players:

$$\begin{aligned} \underset{x,y,z}{\text{minimize}} \quad & F(x,y) \\ \text{s.t.} \quad & x \in X \\ & (y,z) \in E_\varepsilon(x), \end{aligned} \tag{MFG_ε}$$

where $E_\varepsilon(x)$ is the equilibrium set of the following GNEP:

$$\begin{aligned} \underset{y}{\text{minimize}} \quad & -F(x,y) \\ \text{s.t.} \quad & y \in M(x) \\ & f(x,y) \le f(x,z) + \varepsilon \end{aligned} \qquad\qquad \begin{aligned} \underset{z}{\text{minimize}} \quad & f(x,z) \\ \text{s.t.} \quad & z \in M(x). \end{aligned} \tag{6}$$

With the following proposition, we show that, as for global solutions, in order to deal with ($\mathrm{SPBP}_\varepsilon$), one can equivalently address (MFG_ε).

Proposition 5.1 *The following relations hold:*

(i) if $(\widehat{x}, \widehat{y})$ is a global optimal point for ($\mathrm{SPBP}_\varepsilon$), then, for any $\widehat{z} \in S(\widehat{x})$, the tuple $(\widehat{x}, \widehat{y}, \widehat{z})$ is a global optimal point for (MFG_ε);

(ii) if $(\widehat{x}, \widehat{y}, \widehat{z})$ be a global optimal point for (MFG_ε), then $(\widehat{x}, \widehat{y})$ is a global optimal point for ($\mathrm{SPBP}_\varepsilon$).

Proof. Recalling that problem (5), for every fixed x, is pure hierarchical, the claims follow from [16, Theorem 3.2 and Remark 3.2] observing that if $(y,z) \in E_\varepsilon(x)$ then $y \in R_\varepsilon(x)$ and if $y \in R_\varepsilon(x)$ then $(y,z) \in E_\varepsilon(x)$ for any $z \in S(x)$. □

Regarding local solutions, as usual the picture becomes more complicated.

Proposition 5.2 *The following relations hold:*

(i) if $(\widehat{x}, \widehat{y})$ is a local optimal point for (SPBP$_\varepsilon$)*, then, for any $\widehat{z} \in S(\widehat{x})$, the tuple $(\widehat{x}, \widehat{y}, \widehat{z})$ is a local optimal point for* (MFG$_\varepsilon$)*;*

(ii) let $(\widehat{x}, \widehat{y}, z)$ be a local optimal point for (MFG$_\varepsilon$) *for all $z \in S(\widehat{x})$, then $(\widehat{x}, \widehat{y})$ is a local optimal point for* (SPBP$_\varepsilon$)*.*

Proof. (i) Suppose by contradiction that $(\widehat{x}, \widehat{y}, \widehat{z})$ is not a local optimal point for (MFG$_\varepsilon$), i.e. there exists a sequence (x^k, y^k, z^k), with $x^k \in X$ and $(y^k, z^k) \in E_\varepsilon(x^k)$, such that $(x^k, y^k, z^k) \to (\widehat{x}, \widehat{y}, \widehat{z})$ and $F(x^k, y^k) < F(\widehat{x}, \widehat{y})$, for every k. This, in turn, is in contradiction with the local optimality of $(\widehat{x}, \widehat{y})$ for (SPBP$_\varepsilon$), observing that $y^k \in R_\varepsilon(x^k)$, as noted in the proof of Proposition 5.1.

(ii) Suppose by contradiction that $(\widehat{x}, \widehat{y})$ is not a local optimal point for (SPBP$_\varepsilon$), i.e. there exists a sequence (x^k, y^k), with $x^k \in X$ and $y^k \in R_\varepsilon(x^k)$, such that $(x^k, y^k) \to (\widehat{x}, \widehat{y})$ and $F(x^k, y^k) < F(\widehat{x}, \widehat{y})$ for every k. The set-valued mapping S, as previously observed, is outer semicontinuous relative to X and locally bounded at $\widehat{x}$. Hence, a sequence $(z^k) \in S(x^k)$ exists such that $z^k \to \bar{z}$ for some $\bar{z} \in S(\widehat{x})$. Since $y^k \in R_\varepsilon(x^k)$ and $z^k \in S(x^k)$, we have $(y^k, z^k) \in E_\varepsilon(x^k)$: thus, the sequence (x^k, y^k, z^k) is feasible for (MFG$_\varepsilon$) and converges to $(\widehat{x}, \widehat{y}, \bar{z})$ with $F(x^k, y^k) < F(\widehat{x}, \widehat{y})$. Hence, we have a contradiction to the local optimality of the point $(\widehat{x}, \widehat{y}, \bar{z})$, with $\bar{z} \in S(\widehat{x})$, for (MFG$_\varepsilon$). □

Considering the classical counterexample (see, e.g., [16, Example 3.6]), one can easily show that a local optimal point of the resulting (MFG$_\varepsilon$) might not be local optimal for the (SPBP$_\varepsilon$). We observe that a sufficient condition, which is not necessary, for the assumption in Proposition 5.2 (ii) to hold is the following: $(\widehat{x}, \widehat{y}, \widehat{z})$ is a local optimal point for (MFG$_\varepsilon$), being S inner semicontinuous relative to X at $\widehat{x}$. The proof of this claim is deferred to appendix B.

On the other hand, to treat (MFG$_\varepsilon$) numerically resorting to a single-level version (to be defined shortly in subsection 5.2), in the following we assume the functional descriptions $X = \{x \in \mathbb{R}^n \,|\, g(x) \leq 0\}$ and $M(x) = \{y \in \mathbb{R}^p \,|\, h(x, y) \leq 0\}$, where $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^q$ are continuously differentiable on their respective domains and such that X is compact and that the blanket assumption (ii) holds. In particular, we strengthen the blanket assumption (i) to the condition requiring that, for every $x \in X$, there exists some $\widehat{y}(x) \in \mathbb{R}^p$ such that $h(x, \widehat{y}(x)) < 0$, that is, to Slater's condition in $M(x)$.

Regarding GNEP (6), we state some useful properties.

Proposition 5.3 *For every fixed $x \in X$ and $\varepsilon > 0$,*

(i) both problems in GNEP (6) *are convex;*

(ii) Slater's constraint qualification holds for GNEP (6)*;*

(iii) an equilibrium for GNEP (6) *exists;*

(iv) the KKT conditions characterize the equilibria of GNEP (6)*.*

Proof. The first assertions immediately follows from our blanket assumptions. To show the Slater's condition for the left-hand side player in (6), note that any strict convex combination of some Slater point of $M(x)$ and some optimal point of (1) sufficiently close to the latter point is a Slater point for this problem. The third claim holds by the arguments used in the proof of Proposition 5.1 and recalling that $S(x), R_\varepsilon(x) \neq \emptyset$ for every $x \in X$. The last assertion is a straightforward consequence of the other ones. □

5.2 The single-level version of (SPBP$_\varepsilon$)

In view of Proposition 5.3, we intend to tackle (MFG$_\varepsilon$), for positive values of parameter ε, by solving the following Mathematical Program with Complementarity Constraints:

$$\begin{array}{cl}\underset{x,y,z,\lambda,\mu,\xi}{\text{minimize}} & F(x,y)\\ \text{s.t.} & g(x)\le 0\\ & \left.\begin{array}{l}-\nabla_2 F(x,y)+\nabla_2 h(x,y)\lambda+\nabla_2 f(x,y)\xi=0\\ h(x,y)\le 0,\ f(x,y)-f(x,z)-\varepsilon\le 0,\ \lambda,\ \xi\ge 0\\ \lambda^T h(x,y)=0,\ \xi\,[f(x,y)-f(x,z)-\varepsilon]=0\end{array}\right\}\text{[KKT player 1]}\\ & \left.\begin{array}{l}\nabla_2 f(x,z)+\nabla_2 h(x,z)\mu=0\\ h(x,z)\le 0,\ \mu\ge 0\\ \mu^T h(x,z)=0,\end{array}\right\}\text{[KKT player 2]}\end{array} \qquad (\text{MPCC}_\varepsilon)$$

where in [KKT player 1] and in [KKT player 2] we collect the KKT conditions for the left-hand side player and the right-hand side player in GNEP (6), respectively. We denote by

$$\Lambda(x,y,z) \triangleq \{(\lambda,\mu,\xi)\in\mathbb{R}^q\times\mathbb{R}^q\times\mathbb{R}\,|\,\text{[KKT player 1] and [KKT player 2] hold}\}$$

the set of KKT multipliers. We remark that, thanks to Slater's constraint qualification, $\Lambda(x,y,z)\neq\emptyset$ for every $(x,y,z)\in\operatorname{gph}E_\varepsilon$. The latter property will be freely invoked in the following results.

Propositions 5.4 and 5.5 show that problems (MFG$_\varepsilon$) and (MPCC$_\varepsilon$), while equivalent in a global sense, may differ locally.

Proposition 5.4 *The following assertions hold:*

(i) *let* $(\widehat{x},\widehat{y},\widehat{z})$ *be a global optimal point for problem* (MFG$_\varepsilon$)*, then, for each* $(\widehat{\lambda},\widehat{\mu},\widehat{\xi})\in\Lambda(\widehat{x},\widehat{y},\widehat{z})$, $(\widehat{x},\widehat{y},\widehat{z},\widehat{\lambda},\widehat{\mu},\widehat{\xi})$ *is a global optimal point for* (MPCC$_\varepsilon$)*;*

(ii) *let* $(\widehat{x},\widehat{y},\widehat{z},\widehat{\lambda},\widehat{\mu},\widehat{\xi})$ *be a global optimal point for problem* (MPCC$_\varepsilon$)*, then* $(\widehat{x},\widehat{y},\widehat{z})$ *is a global optimal point for* (MFG$_\varepsilon$)*.*

Proof. (i) Assume by contradiction that (x,y,z,λ,μ,ξ) is feasible for (MPCC$_\varepsilon$) and such that $F(x,y)<F(\widehat{x},\widehat{y})$. In turn, $(x,y,z)\in\operatorname{gph}E_\varepsilon$, in contradiction with the optimality of $(\widehat{x},\widehat{y},\widehat{z})$ for (MFG$_\varepsilon$).

(ii) Since $(\widehat{x},\widehat{y},\widehat{z},\widehat{\lambda},\widehat{\mu},\widehat{\xi})$ is a global optimal point for (MPCC$_\varepsilon$), for every $(\lambda,\mu,\xi)\in\Lambda(\widehat{x},\widehat{y},\widehat{z})\neq\emptyset$, $(\widehat{x},\widehat{y},\widehat{z},\lambda,\mu,\xi)$ is still a global optimal point for (MPCC$_\varepsilon$). Assume by contradiction that (x,y,z), with $g(x)\le 0$ and $(y,z)\in E_\varepsilon(x)$, is such that $F(x,y)<F(\widehat{x},\widehat{y})$. Therefore, by Proposition 5.3, there exists $(\lambda,\mu,\xi)\in\Lambda(x,y,z)$. In turn, (x,y,z,λ,μ,ξ) is feasible for (MPCC$_\varepsilon$), with $F(x,y)<F(\widehat{x},\widehat{y})$, which is a contradiction. □

Proposition 5.5 *The following assertions hold:*

(i) *let* $(\widehat{x}, \widehat{y}, \widehat{z})$ *be a local optimal point for problem* (MFG$_\varepsilon$)*, then, for each* $(\widehat{\lambda}, \widehat{\mu}, \widehat{\xi}) \in \Lambda(\widehat{x}, \widehat{y}, \widehat{z})$*,* $(\widehat{x}, \widehat{y}, \widehat{z}, \widehat{\lambda}, \widehat{\mu}, \widehat{\xi})$ *is a local optimal point for* (MPCC$_\varepsilon$)*;*

(ii) *let* $(\widehat{x}, \widehat{y}, \widehat{z}, \lambda, \mu, \xi)$ *be a local optimal point for problem* (MPCC$_\varepsilon$) *for all* $(\lambda, \mu, \xi) \in \Lambda(\widehat{x}, \widehat{y}, \widehat{z})$*, then* $(\widehat{x}, \widehat{y}, \widehat{z})$ *is a local optimal point for* (MFG$_\varepsilon$)*.*

Proof. (i) Ab absurdo, suppose that $(\widehat{x}, \widehat{y}, \widehat{z}, \widehat{\lambda}, \widehat{\mu}, \widehat{\xi})$ for some $(\widehat{\lambda}, \widehat{\mu}, \widehat{\xi}) \in \Lambda(\widehat{x}, \widehat{y}, \widehat{z})$ is not a local optimal point for (MPCC$_\varepsilon$), i.e. there exists a sequence $(x^k, y^k, z^k, \lambda^k, \mu^k, \xi^k)$, with $(\lambda^k, \mu^k, \xi^k) \in \Lambda(x^k, y^k, z^k)$, $g(x^k) \leq 0$, $h(x^k, y^k) \leq 0$, $h(x^k, z^k) \leq 0$, such that $F(x^k, y^k) < F(\widehat{x}, \widehat{y})$ and $(x^k, y^k, z^k, \lambda^k, \mu^k, \xi^k) \to (\widehat{x}, \widehat{y}, \widehat{z}, \widehat{\lambda}, \widehat{\mu}, \widehat{\xi})$. In turn, we have, for every k, by Proposition 5.3, $(y^k, z^k) \in E_\varepsilon(x^k)$, which is in contradiction with the local optimality of $(\widehat{x}, \widehat{y}, \widehat{z})$ for (MFG$_\varepsilon$).

(ii) Assume by contradiction that a sequence $(x^k, y^k, z^k) \in \operatorname{gph} E_\varepsilon$ exists such that $g(x^k) \leq 0$, $F(x^k, y^k) < F(\widehat{x}, \widehat{y})$ and $(x^k, y^k, z^k) \to (\widehat{x}, \widehat{y}, \widehat{z})$. The set-valued mapping Λ is outer semicontinuous relative to $(X \times \mathbb{R}^p \times \mathbb{R}^p) \cap \operatorname{gph} E_\varepsilon$ (see, e.g. Theorem [2, 3.1.1]) and under the Slater's constraint qualification one can show by contradiction that Λ is locally bounded at $(\widehat{x}, \widehat{y}, \widehat{z})$. Hence, a sequence $(\lambda^k, \mu^k, \xi^k) \in \Lambda(x^k, y^k, z^k)$ exists such that $(\lambda^k, \mu^k, \xi^k) \to (\bar{\lambda}, \bar{\mu}, \bar{\xi})$ for some $(\bar{\lambda}, \bar{\mu}, \bar{\xi}) \in \Lambda(\widehat{x}, \widehat{y}, \widehat{z})$. Therefore, the sequence $(x^k, y^k, z^k, \lambda^k, \mu^k, \xi^k)$ is feasible for (MPCC$_\varepsilon$) and converges to $(\widehat{x}, \widehat{y}, \widehat{z}, \bar{\lambda}, \bar{\mu}, \bar{\xi})$, with $F(x^k, y^k) < F(\widehat{x}, \widehat{y})$. Thus, we have a contradiction to the local optimality of the point $(\widehat{x}, \widehat{y}, \widehat{z}, \bar{\lambda}, \bar{\mu}, \bar{\xi})$ with $(\bar{\lambda}, \bar{\mu}, \bar{\xi}) \in \Lambda(\widehat{x}, \widehat{y}, \widehat{z})$ for (MPCC$_\varepsilon$). $\square$

One can easily find examples showing that in general a local optimal solution of (MPCC$_\varepsilon$) does not lead to a local optimum for (MFG$_\varepsilon$). E.g., it suffices to take the same lower level problem as the one in [6, Example 3.4]. Again, as remarked in the previous subsection, a sufficient condition, which is not necessary, for the assumption in Proposition 5.5 (ii) to hold is the following: $(\widehat{x}, \widehat{y}, \widehat{z}, \widehat{\lambda}, \widehat{\mu}, \widehat{\xi})$ is a local optimal point for (MPCC$_\varepsilon$), being Λ inner semicontinuous relative to $(X \times \mathbb{R}^p \times \mathbb{R}^p) \cap \operatorname{gph} E_\varepsilon$ at $(\widehat{x}, \widehat{y}, \widehat{z})$ (see appendix B for further details). We further notice that Propositions 5.4 and 5.5 are generalizations of the results regarding lower level optimal points (cf. [6]) to our framework of lower level equilibria.

5.3 A numerical procedure for the single-level version of (SPBP$_\varepsilon$)

Clearly, problem (MPCC$_\varepsilon$) is still structurally difficult to solve. Besides classical methods, here we propose a mixed integer approach in order to deal numerically with (MPCC$_\varepsilon$). Specifically, one can equivalently replace the difficult complementarity conditions in the feasible set of (MPCC$_\varepsilon$) with the following mixed-integer nonlinear constraints

$$\begin{aligned} \lambda &\leq \mathcal{M}\delta^1, & h(x, y) &\geq -\mathcal{M}(1 - \delta^1) \\ \mu &\leq \mathcal{M}\delta^2, & h(x, z) &\geq -\mathcal{M}(1 - \delta^2) \\ \xi &\leq \mathcal{M}\delta^3, & f(x, y) - f(x, z) - \varepsilon &\geq -\mathcal{M}(1 - \delta^3), \end{aligned}$$

where $\delta^1 \in \{0,1\}^q$, $\delta^2 \in \{0,1\}^q$, $\delta^3 \in \{0,1\}$ are additional binary variables and $\mathcal{M} >> 0$.

If, in addition to the initial assumptions, $F(x, y) = F_1(x) + F_2(y)$, where F_2 is linear, and the lower level problem is linear, then actually the continuous relaxation of the resulting mixed-integer nonlinear problem turns out to be convex. In this case, standard approaches (e.g., branch and bound procedures) lead to global optimal points. Note that requiring the lower level problem to be linear does not entail in general a fixed solution set S: hence, by means of our one level version of (SPBP$_\varepsilon$), we may tackle problems that cannot be dealt with resorting to a subgradient-like strategy as in Remark 3.2.

Remark 5.6 We observe that, under our assumptions, the proposed simple mixed-integer strategy gives us, in the end, a global optimal solution for (PBP$_\varepsilon$). In fact, the computed solution is global optimal for (MPCC$_\varepsilon$), and thus, in turn, leveraging Proposition 5.4 (ii), provides one with a global optimal point for (MFG$_\varepsilon$); the latter point, by Proposition 5.1 (ii), leads to a global optimal point for (SPBP$_\varepsilon$), whose projection on the subspace of the x variables is finally a global optimal point for (PBP$_\varepsilon$), due to Proposition 4.1 (ii).

6 Problems relations at a glance

For the reader's convenience, here we summarize all the relations among the problems that we have discussed so far, namely (PBP$_\varepsilon$), (SPBP$_\varepsilon$), (MFG$_\varepsilon$) and (MPCC$_\varepsilon$).

In the following Proposition 6.1, we show that the x-part of any feasible solution for (SPBP$_\varepsilon$), (MFG$_\varepsilon$), or (MPCC$_\varepsilon$), is feasible for (PBP$_\varepsilon$).

Proposition 6.1 *The following assertions hold:*

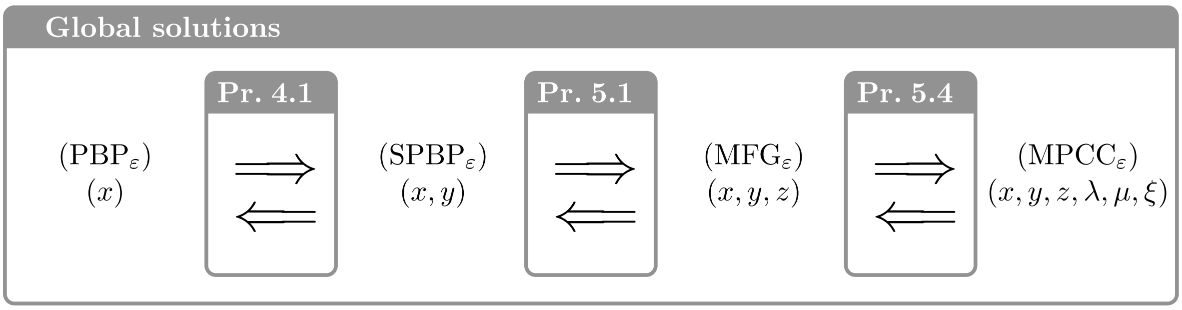
(i) any feasible solution $(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}, \bar{\mu}, \bar{\xi})$ *for* (MPCC$_\varepsilon$) *is such that* $(\bar{x}, \bar{y}, \bar{z})$ *is feasible for* (MFG$_\varepsilon$)*;*

(ii) any feasible solution $(\bar{x}, \bar{y}, \bar{z})$ *for* (MFG$_\varepsilon$) *is such that* $(\bar{x}, \bar{y})$ *is feasible for* (SPBP$_\varepsilon$)*;*

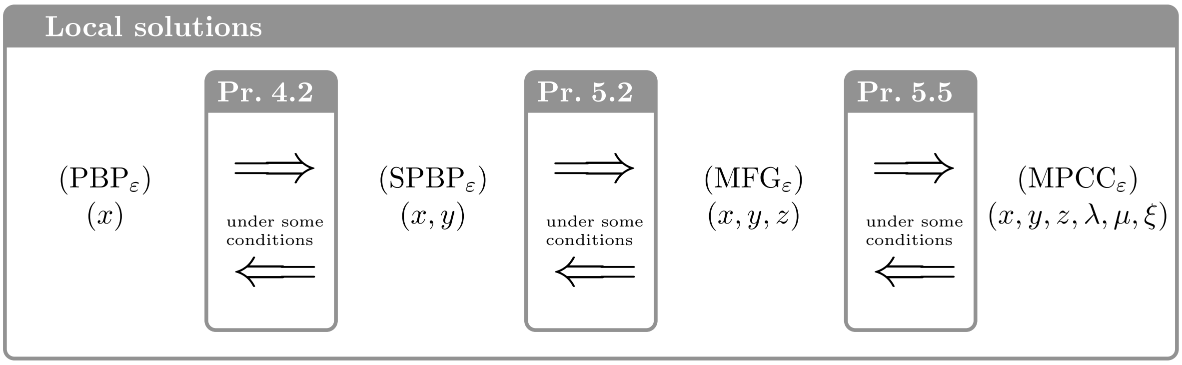
(iii) any feasible solution $(\bar{x}, \bar{y})$ *for* (SPBP$_\varepsilon$) *is such that* $(\bar{x})$ *is feasible for* (PBP$_\varepsilon$)*.*

Proof. Assertion (i) follows by (iv) in Proposition 5.3. Claim (ii) is due to the fact that $(\bar{y}, \bar{z}) \in E_\varepsilon(\bar{x})$ entails $\bar{y} \in R_\varepsilon(\bar{x})$. The last assertion is trivial. □

As for global solutions, we sketch the problems' relations in the following scheme.



On the other hand, the summary of the problems' relations regarding local solutions is reported in the following scheme.



7 An application in economics

Bilevel programs are widely and fruitfully used to model many real-world problems in economics (see, e.g., [4]). Here we propose an application in economics involving two decision levels.

Let us consider a market in which N firms produce the same n goods. Any firm $\nu \in \{1, \dots, N\}$ produces the quantities $q^\nu \in \mathbb{R}^n$ of the goods, with $q^\nu \in [l^\nu + v^\nu, u^\nu - w^\nu]$, where l^ν and u^ν are suitable bounds and $v^\nu \in \mathbb{R}^n$ and $w^\nu \in \mathbb{R}^n$ are exogenous parameters whose meaning will be explained later on. On the other hand, $p \in \mathbb{R}^n$ are the prices of the goods. Any firm ν has linear production costs $(q^\nu)^T c^\nu$ where $c^\nu \in \mathbb{R}^n$ are the costs of the goods. Assuming the presence of some shared constraints on the production levels, each firm ν, trying to minimize its own loss function, addresses the following parametric (in $v^\nu \in \mathbb{R}^n$ and $w^\nu \in \mathbb{R}^n$) optimization problem:

$$\begin{array}{ll} \underset{q^\nu}{\text{minimize}} & (q^\nu)^T (c^\nu - p) \\ \text{s.t.} & r(q^1, \dots, q^N) \leq 0 \\ & l^\nu + v^\nu \leq q^\nu \leq u^\nu - w^\nu, \end{array} \tag{7}$$

where $r : \mathbb{R}^{nN} \to \mathbb{R}^m$ is a convex function that models the shared constraints.

We consider the case in which the firms do not cooperate and play simultaneously and rationally. In this case, given any v^ν and w^ν for every ν, we model the game played by the firms as a GNEP. In particular, this GNEP is a generalized potential game, see e.g. [13, 23]. Any solution of the following convex optimization problem

$$\begin{array}{ll} \underset{q^1,\dots,q^N}{\text{minimize}} & \sum_{\nu=1}^{N} (q^\nu)^T (c^\nu - p) \\ \text{s.t.} & r(q^1, \dots, q^N) \leq 0 \\ & l^\nu + v^\nu \leq q^\nu \leq u^\nu - w^\nu, \quad \nu = 1, \dots, N, \end{array} \tag{8}$$

is an equilibrium of GNEP (7).

We consider the presence of a regulator who takes care of consumers' wellness. Specifically, the regulator, setting v^ν and w^ν to suitably modify the lower and the upper bounds for q^ν, aims at minimizing $p^T \sum_{\nu=1}^{N} q^\nu$, that is the total amount of the consumers' expenses.

We distinguish two categories of goods: specifically, if a good $i \in \{1, \dots, N\}$ is such that $c_i^\nu - p_i \geq 0$, that is, e.g., i is a public good, then the regulator might be interested in raising

v_i^ν in order to stimulate the production of the good. On the other hand, if $i \in \{1, \dots, N\}$ is such that $c_i^\nu - p_i < 0$, then the regulator may seek to raise w_i^ν in order to get a reduced production level for the good.

Assuming a pessimistic point of view, the regulator solves the (SPBP$_\varepsilon$) where

- $x = (v^1, \dots, v^N, w^1, \dots, w^N)$, $y = (q^1, \dots, q^N)$;
- $F(x, y) = F(q^1, \dots, q^N) = p^T \sum_{\nu=1}^N q^\nu$;
- $X = \{x \in \mathbb{R}^{2nN} \,|\, \|x\|_2^2 - V \le 0\}$;
- $M(x) = M(v^1, \dots, v^N, w^1, \dots, w^N) = \{(q^1, \dots, q^N) \,|\, r(q^1, \dots, q^N) \le 0, \; l^\nu + v^\nu \le q^\nu \le u^\nu - w^\nu, \nu = 1, \dots, N\}$;
- $f(x, y) = f(q^1, \dots, q^N) = \sum_{\nu=1}^N (q^\nu)^T (c^\nu - p)$.

We remark that the model described above cannot be treated with the numerical procedure proposed in [28]: having M fixed is a fundamental requirement for the numerical method in [28] to work well, but here the set-valued mapping M depends on the upper level variables $(v^1, \dots, v^N, w^1, \dots, w^N)$.

As for the mixed-integer nonlinear approach to the resulting (MPCC$_\varepsilon$), we address the following problem:

$$\begin{aligned}
\underset{v,w,q,z,\lambda,\mu,\xi,\delta}{\text{minimize}} \quad & p^T \sum_{\nu=1}^N q^\nu \\
\text{s.t.} \quad & \|(v^1, \dots, v^N, w^1, \dots, w^N)\|_2^2 - V \le 0 \\
& -p + \nabla_\nu r(q^1, \dots, q^N)\lambda^{\nu,1} - \lambda^{\nu,2} + \lambda^{\nu,3} + (c^\nu - p)\xi = 0, \; \nu = 1, \dots, N \\
& r(q^1, \dots, q^N) \le 0 \\
& l^\nu + v^\nu - q^\nu \le 0, \; q^\nu - u^\nu + w^\nu \le 0, \; \nu = 1, \dots, N \\
& \sum_{\nu=1}^N (q^\nu - z^\nu)^T (c^\nu - p) - \varepsilon \le 0 \\
& \lambda^{\nu,i}, \, \xi \ge 0, \; \nu = 1, \dots, N, \, i = 1, 2, 3 \\
& r(q^1, \dots, q^N) \ge -\mathcal{M}(1 - \delta^{1,\nu,1}), \; \nu = 1, \dots, N \\
& l^\nu + v^\nu - q^\nu \ge -\mathcal{M}(1 - \delta^{1,\nu,2}), \; q^\nu - u^\nu + w^\nu \ge -\mathcal{M}(1 - \delta^{1,\nu,3}), \; \nu = 1, \dots, N \\
& \sum_{\nu=1}^N (q^\nu - z^\nu)^T (c^\nu - p) - \varepsilon \ge -\mathcal{M}(1 - \delta^3) \\
& \lambda^{\nu,i} \le \mathcal{M}\delta^{1,\nu,i}, \; \xi \le \mathcal{M}\delta^3, \; \nu = 1, \dots, N, \, i = 1, 2, 3 \\
& c^\nu - p + \nabla_\nu r(z^1, \dots, z^N)\mu^{\nu,1} - \mu^{\nu,2} + \mu^{\nu,3} = 0, \; \nu = 1, \dots, N \\
& r(z^1, \dots, z^N) \le 0
\end{aligned}$$

$$l^\nu + v^\nu - z^\nu \le 0,\ z^\nu - u^\nu + w^\nu \le 0,\ \nu = 1,\dots,N$$

$$\mu^{\nu,i} \ge 0,\ \nu = 1,\dots,N,\ i = 1,2,3$$

$$r(z^1,\dots,z^N) \ge -\mathcal{M}(1-\delta^{2,\nu,1}),\ \nu = 1,\dots,N$$

$$l^\nu + v^\nu - z^\nu \ge -\mathcal{M}(1-\delta^{2,\nu,2}),\ z^\nu - u^\nu + w^\nu \ge -\mathcal{M}(1-\delta^{2,\nu,3}),\ \nu = 1,\dots,N$$

$$\mu^{\nu,i} \le \mathcal{M}\delta^{2,\nu,i},\ \nu = 1,\dots,N,\ i = 1,2,3$$

$$\delta^{1,\nu,i} \in \{0,1\},\ \delta^{2,\nu,i} \in \{0,1\},\ \delta^3 \in \{0,1\},\ \nu = 1,\dots,N,\ i = 1,2,3.$$

All the experiments were carried out on an Intel Core i7-4702MQ CPU @ 2.20GHz x 8 with Ubuntu 14.04 LTS 64-bit and by using AMPL. As optimization solver we used CPLEX 12.6.0.1 with default options.

We generated 3 different instances for the bilevel problem (SPBP$_\varepsilon$). They are denoted by A, B, and C. At the lower level, we consider a market with $N = 3$ firms, each producing a total amount of $n = 6$ (3 public and 3 non public) goods, with $m = 4$ shared constraints.

The lower and upper bounds, the prices, the costs of the goods for any firm ν were randomly generated by using the uniform distribution: $l^\nu \in [100, 200]$, $u^\nu \in [500, 600]$, $p \in [6, 9]$, c_i^ν (for public goods) $\in [p_i, p_i + 1]$, c_i^ν (for non public goods) $\in [p_i - 1, p_i]$, $A_{ij}^\nu \in [0, 100]$ and $b_j \in$ [2e+5, 4e+5], where $r(y) = Ay - b$. Moreover, we have chosen $\mathcal{M}$ = 1e+6 and, for the time being, $\varepsilon = 0.1$.

We stress that, as observed in Remark 5.6, the computed solutions are global optimal points for the original problem (PBP$_\varepsilon$).

	A			B			C		
V	1e+1	1e+4	1e+7	1e+1	1e+4	1e+7	1e+1	1e+4	1e+7
$\|x\|_2^2$	1e+1	1e+4	1.8e+6	1e+1	1e+4	1.5e+6	1e+1	1e+4	1.6e+6
F	44559.5	43675.9	23353.3	27902.2	27339.9	19960.7	37728.2	36661.2	21541.3
$\sum_\nu \varepsilon$-gain$_\nu$	1279.17	1150.58	-198.62	662.56	624.28	-108.45	2173.38	1960.73	317.94
$\sum_\nu$gain$_\nu$	1279.27	1150.68	-198.52	662.66	624.38	-108.44	2173.48	1960.83	318.04
ε-gain$_1$	80.48	45.17	-149.95	282.85	254.84	-68.87	755.95	736.12	119.94
gain$_1$	80.13	45.17	-149.85	282.80	254.74	-68.87	755.95	736.12	119.94
ε-gain$_2$	489.69	429.55	-69.89	28.53	0.58	-90.81	21.53	-127.72	-95.35
gain$_2$	489.69	429.65	-69.89	28.53	0.58	-90.81	21.19	-127.62	-95.35
ε-gain$_3$	709.00	675.86	21.22	351.18	368.86	51.24	1395.89	1352.32	293.35
gain$_3$	709.45	675.86	21.22	351.32	369.06	51.24	1396.34	1352.32	293.45

Table 1: Detailed numerical results for instances A, B and C and $\varepsilon = 0.1$.

	ε	A			B			C		
V		1e+1	1e+4	1e+7	1e+1	1e+4	1e+7	1e+1	1e+4	1e+7
F	0.1	44559.5	43675.9	23353.3	27902.2	27339.9	19960.7	37728.2	36661.2	21541.3
	0.001	44548.0	43515.6	20414.1	27899.7	27339.9	19960.5	37725.8	36660.1	19290.8
	0.00001	44548.0	43515.4	20414.1	27899.7	27339.8	19960.5	37725.8	36660.1	19290.8

Table 2: Numerical results for instances A, B and C and different values for ε.

We consider, for each instance, three different values for the upper bound V in the constraints defining X. Correspondingly, in Table 1, we report, setting $\varepsilon = 0.1$, the computed value for: $\|x\|_2^2$, the upper level objective function F, ε-gain$_\nu$ $= (q^\nu)^T(p - c^\nu)$ and

$\text{gain}_\nu = (z^\nu)^T(p - c^\nu)$, that is, for each firm ν, the perturbed (by ε) and the real gain, respectively.

We observe that, as expected, $\sum_\nu \text{gain}_\nu - \sum_\nu \varepsilon\text{-gain}_\nu$ is always not greater than ε. By increasing V, the corresponding computed values for the regulator's objective F get progressively better, while the gain that the firms achieve progressively worsen.

In Table 2, we show how the computed value for F behaves for different values of the perturbation parameter ε: as expected, by reducing ε, the value achieved by F is not increasing. This follows observing that if $\varepsilon_1 > \varepsilon_2 > 0$, then $\psi_{\varepsilon_1}(x) \geq \psi_{\varepsilon_2}(x)$ for every $x \in X$. The behavior of the values of F also is in accordance with the convergence result from Theorem 3.5 for $\varepsilon \searrow 0$.

8 Final remarks

We presented in detail one framework to approximate the finite infimum of a pessimistic bilevel problem by the optimal values of a sequence of single-level problems. The alternative approaches from [28] and, even from appendix C, show that also other ideas may lead to single-level approximations. In particular, we emphasize that the presented approach hinges on the smoothness and convexity assumptions from section 2. Nonetheless, pessimistic bilevel problems are still nonsmooth and nonconvex even under these assumptions. Weakening them may require a fundamentally different approximation concept (like e.g. the one in [28], which however requires different assumptions) and is left for future research.

We also point out that our main motivation for proving the relations among local optimal points of the different proposed auxiliary problems is the capability of many standard solution algorithms to generate local rather than global optimal points. However, since termination criteria of such algorithms often use stationarity conditions, from an algorithmic point of view it may seem even more appropriate to study the interrelations between stationary points of the different auxiliary problems. Unfortunately the nonsmoothness of the involved auxiliary problems leads to different alternative stationarity concepts, so that such a study would be out of scope of the present paper and is also left for future research.

Appendix

A Proof of Theorem 3.5

The existence of an accumulation point x^* of (x^k) follows from the compactness of X. Without loss of generality, let (x^k) converge to x^*. This implies also the convergence of the values $\alpha^k \triangleq \psi_{\varepsilon_k}(x^k)$, $k \in \mathbb{N}$. In fact, for each $k \in \mathbb{N}$ and $x \in X$ the inclusions $S_{\varepsilon_k}(x) \supseteq S_{\varepsilon_{k+1}}(x) \supseteq S(x)$ imply

$$\psi_{\varepsilon_k}(x) \geq \psi_{\varepsilon_{k+1}}(x) \geq \psi(x).$$

Together with the minimality of x^{k+1} for $\psi_{\varepsilon_{k+1}}$ on X this yields

$$\alpha^k = \psi_{\varepsilon_k}(x^k) \geq \psi_{\varepsilon_{k+1}}(x^k) \geq \psi_{\varepsilon_{k+1}}(x^{k+1}) = \alpha^{k+1},$$

so that the sequence (α^k) is monotonically decreasing. In addition, by

$$\alpha^k = \psi_{\varepsilon_k}(x^k) \geq \psi(x^k) \geq \operatorname{cl}\psi(x^k) \geq \min_{x \in X} \operatorname{cl}\psi(x)$$

the sequence (α^k) is bounded from below and, thus, convergent to some α^*.

Since the relation $\alpha^k \geq \psi(x^k)$ means $(x^k, \alpha^k) \in \operatorname{epi}\psi$ for each $k \in \mathbb{N}$, the limit satisfies $(x^*, \alpha^*) \in \operatorname{cl}\operatorname{epi}\psi = \operatorname{epi}\operatorname{cl}\psi$, that is, $\operatorname{cl}\psi(x^*) \leq \alpha^*$. Hence, to prove the assertion, it remains to show $\alpha^* \leq \min_{x\in X} \operatorname{cl}\psi(x)$ as this implies $\operatorname{cl}\psi(x^*) \leq \min_{x\in X} \operatorname{cl}\psi(x)$.

In fact, for each $\tau > 0$ we show that $\alpha^k \leq \min_{x\in X} \operatorname{cl}\psi(x) + \tau$ holds for all sufficiently large $k \in \mathbb{N}$. Since this implies $\alpha^* \leq \min_{x\in X} \operatorname{cl}\psi(x) + \tau$ for any $\tau > 0$, the assertion will be shown.

For any given $\tau > 0$ the identity $\min_{x\in X} \operatorname{cl}\psi(x) = \inf_{x\in X} \psi(x)$ implies the existence of some $z \in X$ with

$$\psi(z) \ \leq \min_{x\in X} \operatorname{cl}\psi(x) + \frac{\tau}{2}.$$

For this fixed $z \in X$ the set-valued mapping

$$Y : \mathbb{R}_+ \rightrightarrows \mathbb{R}^p,\ \varepsilon \mapsto Y(\varepsilon) \triangleq S_\varepsilon(z) = \{y \in M(z) \,|\, f(z,y) \leq \varphi(z) + \varepsilon\}$$

is, by the closedness of $M(z)$ and the continuity of the function $f(z, \bullet)$, outer semicontinuous relative to $\mathbb{R}_+$ at any point in $\mathbb{R}_+$. Moreover, the boundedness of $M(z)$ implies that Y is (locally) bounded on $\mathbb{R}_+$. Thus, by the continuity of the function $F(z, \bullet)$ and [15, Theorem 5], the optimal value function

$$\varrho : \mathbb{R}_+ \to \mathbb{R},\ \varepsilon \mapsto \varrho(\varepsilon) \triangleq \max_{y\in Y(\varepsilon)} F(z,y)$$

is upper semicontinuous relative to $\mathbb{R}_+$ at any point in $\mathbb{R}_+$. In particular at $\varepsilon = 0$ we have

$$\limsup_{\varepsilon\searrow 0} \psi_\varepsilon(z) = \limsup_{\varepsilon\searrow 0} \varrho(\varepsilon) \leq \varrho(0) = \psi(z).$$

Consequently, for our special sequence $\varepsilon_k \searrow 0$ all sufficiently large $k \in \mathbb{N}$ satisfy

$$\psi_{\varepsilon_k}(z) \leq \psi(z) + \frac{\tau}{2}$$

and due to

$$\alpha^k = \psi_{\varepsilon_k}(x^k) \leq \psi_{\varepsilon_k}(z) \leq \psi(z) + \frac{\tau}{2} \leq \min_{x\in X} \operatorname{cl}\psi(x) + \tau$$

x^* is a global minimal point of $\operatorname{cl}\psi$ on X. The proof is complete in view of (3). □

B Inner semicontinuity and local optimality

We justify the claims below Propositions 5.2 and 5.5. More precisely, we give, in a general framework, a sufficient condition for (ii) in Propositions 5.2 and 5.5 to hold.

Proposition B.1 *Consider the following problem:*

$$\begin{array}{ll} \underset{x,y}{\textit{minimize}} & \vartheta(x) \\ \textit{s.t.} & x \in X \\ & y \in \Xi(x), \end{array}$$

where ϑ is continuous and Ξ is outer semicontinuous and locally bounded relative to X.

Let $(\widehat{x}, \widehat{y})$ be a local optimal point for the problem; if Ξ is inner semicontinuous at $\widehat{x}$ relative to X, then, for every $y \in \Xi(\widehat{x})$, $(\widehat{x}, y)$ is a local optimal point for the problem.

Proof. Suppose by contradiction that $\widetilde{y} \in \Xi(\widehat{x})$ is such that there exists a sequence $(x^k, y^k) \to (\widehat{x}, \widetilde{y})$, with $x^k \in X$, $y^k \in \Xi(x^k)$, and $\vartheta(x^k) < \vartheta(\widehat{x})$. By the inner semicontinuity property of Ξ, a sequence $z^k \in \Xi(x^k)$, with $z^k \to \widehat{y}$, exists. In turn, observing that $(x^k, z^k) \to (\widehat{x}, \widehat{y})$, $x^k \in X$, $z^k \in \Xi(x^k)$ and $\vartheta(x^k) < \vartheta(\widehat{x})$, we get a contradiction to the local optimality of $(\widehat{x}, \widehat{y})$ for the problem. □

Of course, the assumption requiring the objective function not to depend on y is key in order to prove the result: in fact, if the objective function depends also on y, the claim in Proposition B.1 is no more valid. Moreover, the *viceversa* in Proposition B.1 does not hold as witnessed by the following example.

Example B.2 Consider the problem

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & x^2 \\ \text{s.t.} & x \in X \triangleq [-1, 1] \\ & y \in \Xi(x) \triangleq \begin{cases} 0 & \text{if } x < 0 \\ [0,1] & \text{if } x = 0 \\ 1 & \text{if } x = 0. \end{cases} \end{array}$$

We observe that Ξ is not inner semicontinuous at $\widehat{x} = 0$ relative to X, and, for every $y \in \Xi(\widehat{x})$, $(\widehat{x}, y)$ is a local optimal point for the problem.

C Three player single-level approach for (SPBP$_\varepsilon$)

Leveraging some results in [16], but differently from what done in subsection 5.1, here we introduce a (single-level) GNEP with three players which turns out to have some interesting connections with the (SPBP$_\varepsilon$).

Preliminarily, we consider the following Nash game with a hierarchical structure:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & F(x,y) \\ \text{s.t.} & x \in X \end{array} \qquad\qquad \begin{array}{ll} \underset{y}{\text{minimize}} & -F(x,y) \\ \text{s.t.} & y \in S_\varepsilon(x). \end{array} \tag{10}$$

Note that, under the initial assumptions, each player's problem is convex. As for the relation between equilibria of (10) and global optimal points of (SPBP$_\varepsilon$) (and, thus, global optimal points of (PBP$_\varepsilon$), see Proposition 4.1 (ii)), the following result, which is reminiscent of [16, Theorem 3.1], holds.

Proposition C.1 *The following relations hold:*

(i) let $(\widehat{x}, \widehat{y})$ be a global optimal point for (SPBP$_\varepsilon$)*, then $(\widehat{x}, \widehat{y})$ is feasible for GNEP* (10)*; furthermore, if*

$$\max_{y \in S_\varepsilon(x)} F(x,y) \le F(x, \widehat{y}), \qquad \forall x \in X,$$

then $(\widehat{x}, \widehat{y})$ is an equilibrium of GNEP (10)*;*

(ii) let $(\widehat{x}, \widehat{y})$ be an equilibrium of GNEP (10)*, then $(\widehat{x}, \widehat{y})$ is a feasible point for* (SPBP$_\varepsilon$)*; furthermore, if*

$$\max_{y \in S_\varepsilon(x)} F(x,y) \ge F(x, \widehat{y}), \qquad \forall x \in X,$$

then $(\widehat{x}, \widehat{y})$ is a global optimal point for (SPBP$_\varepsilon$)*.*

Proof. The assertion follows from explicitly stating feasibility and optimality in (SPBP_ε) and GNEP (10), which we report for the reader's convenience:

$$\widehat{x} \in X, \ \ \widehat{y} \in S_\varepsilon(\widehat{x}), \ \ \max_{y \in S_\varepsilon(\widehat{x})} F(\widehat{x}, y) \leq F(\widehat{x}, \widehat{y}) \leq \max_{y \in S_\varepsilon(x)} F(x, y), \ \ \forall x \in X; \tag{11}$$

$$\widehat{x} \in X, \ \ \widehat{y} \in S_\varepsilon(\widehat{x}), \ \ \max_{y \in S_\varepsilon(\widehat{x})} F(\widehat{x}, y) \leq F(\widehat{x}, \widehat{y}) \leq F(x, \widehat{y}), \ \ \forall x \in X. \tag{12}$$

□

The simple results in Proposition C.1 allow one to define classes of problems for which conditions (i) and (ii) are satisfied. Specifically,

- as for (i) in Proposition C.1, if $S_\varepsilon(\hat{x}) \supseteq S_\varepsilon(x)$ $\forall x \in X$ and F is separable with respect to the variables blocks x and y, then (i) is satisfied;
- as for (ii) in Proposition C.1, if $\widehat{y} \in S_\varepsilon(x)$ $\forall x \in X$, then (ii) is satisfied. Note that, if $S_\varepsilon(\widehat{x}) \subseteq S_\varepsilon(x)$ $\forall x \in X$, then the latter sufficient condition holds.

Clearly, when S_ε is fixed for every $x \in X$, (ii) is verified; moreover, if F is separable, then (SPBP_ε) and GNEP (10) turn out to be equivalent in a global sense.

The assertion (ii) in Proposition C.1 can be stated in a local sense (see also [16, Theorem 3.3]).

Proposition C.2 *Let* $(\widehat{x}, \widehat{y})$ *be an equilibrium of GNEP* (10)*. If a neighborhood* $\widehat{N}$ *of* $\widehat{x}$ *exists such that*

$$\max_{y \in S_\varepsilon(x)} F(x, y) \geq F(x, \widehat{y}), \qquad \forall x \in \widehat{N} \cap X, \tag{13}$$

then $(\widehat{x}, \widehat{y})$ *is a local optimal point for* (SPBP_ε)*.*

Condition $\widehat{y} \in S_\varepsilon(x)$ $\forall x \in \widehat{N} \cap X$ is satisfied whenever, for example, $\widehat{y} \in \text{int} S_\varepsilon(\widehat{x})$.

By virtue of the local boundedness of S_ε, there exists a compact set $C \subset \mathbb{R}^p$, independent of x, such that $S_\varepsilon(x) \subseteq C$ for every $x \in X$: in view of this consideration, leveraging Ichiishi's theorem (see e.g. [11]), the existence of an equilibrium easily follows.

Reasoning as done for problem (5) in subsection 5.1 and thus in view of [16, Theorem 3.2 and Remark 3.2], formulation (10) can be further and equivalently (in a global sense) recast as the following GNEP with three players:

$$\begin{array}{lllllllll} \underset{x}{\text{minimize}} & F(x,y) & \quad & \underset{y}{\text{minimize}} & -F(x,y) & \quad & \underset{w}{\text{minimize}} & f(x,w) \\ \text{s.t.} & x \in X & & \text{s.t.} & f(x,y) \leq f(x,w) + \varepsilon & & \text{s.t.} & w \in M(x). \\ & & & & y \in M(x) & & & \end{array} \tag{14}$$

We observe that this approach makes sense mainly when S_ε is a fixed set-valued mapping, possibly with $\varepsilon = 0$, as in Remark 3.2.

References

[1] D. Aussel and S. Sagratella. Sufficient conditions to compute any solution of a quasi-variational inequality via a variational inequality. *Mathematical Methods of Operations Research*, 85(1):3–18, 2017.

[2] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer. *Non-linear parametric optimization*. Akademie-Verlag Berlin, 1982.

[3] D.P. Bertsekas. *Convex optimization algorithms*. Athena Scientific Belmont, 2015.

[4] S. Dempe. *Foundations of bilevel programming*. Springer Science & Business Media, 2002.

[5] S. Dempe, N. Dinh, and J. Dutta. Optimality conditions for a simple convex bilevel programming problem. In *Variational Analysis and Generalized Differentiation in Optimization and Control*, pages 149–161. Springer, 2010.

[6] S. Dempe and J. Dutta. Is bilevel programming a special case of a mathematical program with complementarity constraints? *Mathematical programming*, 131(1-2):37–48, 2012.

[7] 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.

[8] S. Dempe, B.S. Mordukhovich, and A.B. Zemkoho. Necessary optimality conditions in pessimistic bilevel programming. *Optimization*, 63:505–533, 2014.

[9] S. Dempe and A.B. Zemkoho. The bilevel programming problem: reformulations, constraint qualifications and optimality conditions. *Mathematical Programming*, 138(1-2):447–473, 2013.

[10] A. Dreves, F. Facchinei, C. Kanzow, and S. Sagratella. On the solution of the KKT conditions of generalized Nash equilibrium problems. *SIAM Journal on Optimization*, 21(3):1082–1108, 2011.

[11] F. Facchinei and C. Kanzow. Generalized Nash equilibrium problems. *Annals of Operations Research*, 175(1):177–211, 2010.

[12] F. Facchinei and L. Lampariello. Partial penalization for the solution of generalized Nash equilibrium problems. *Journal of Global Optimization*, 50(1):39–57, 2011.

[13] F. Facchinei, V. Piccialli, and M. Sciandrone. Decomposition algorithms for generalized potential games. *Computational Opt. and Appl.*, 50(2):237–262, 2011.

[14] F. Facchinei and S. Sagratella. On the computation of all solutions of jointly convex generalized Nash equilibrium problems. *Opt. Letters*, 5(3):531–547, 2011.

[15] W. Hogan. Point-to-set maps in mathematical programming. *SIAM Review*, pages 591–603, 1973.

[16] L. Lampariello and S. Sagratella. A bridge between bilevel programs and Nash games. *Journal of Optimization Theory and Applications*, 174(2):613–635, 2017.

[17] P. Loridan and J. Morgan. Least-norm regularization for weak two-level optimization problems. In *Barbu V., Tiba D., Bonnans J.F. (eds) Optimization, Optimal Control and Partial Differential Equations, International Series of Numerical Mathematics, Birkhäuser, Basel*, volume 107, pages 307–318, 1992.

[18] P. Loridan and J. Morgan. Weak via strong Stackelberg problem: new results. *Journal of global Optimization*, 8(3):263–287, 1996.

[19] R. Lucchetti, F. Mignanego, and G. Pieri. Existence theorems of equilibrium points in Stackelberg. *Optimization*, 18(6):857–866, 1987.

[20] J.-S. Pang and M. Fukushima. Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. *Computational Management Science*, 2:21–56, 2009.

[21] R.T. Rockafellar and J.B. Wets. *Variational Analysis*. Springer, 1998.

[22] S. Sagratella. Computing all solutions of Nash equilibrium problems with discrete strategy sets. *SIAM Journal on Optimization*, 26(4):2190–2218, 2016.

[23] S. Sagratella. Algorithms for generalized potential games with mixed-integer variables. *Computational Optimization and Applications*, 68(3):689–717, 2017.

[24] S. Sagratella. Computing equilibria of Cournot oligopoly models with mixed-integer quantities. *Mathematical Methods of Operations Research*, 86(3):549–565, 2017.

[25] G. Scutari, F. Facchinei, J.-S. Pang, and L. Lampariello. Equilibrium selection in power control games on the interference channel. In *INFOCOM, 2012 Proceedings IEEE*, pages 675–683, 2012.

[26] O. Stein. *Bi-level strategies in semi-infinite programming*. Kluwer Academic Publishers, 2003.

[27] O. Stein and G. Still. On generalized semi-infinite optimization and bilevel optimization. *European Journal of Operational Research*, 142(3):444–462, 2002.

[28] W. Wiesemann, A. Tsoukalas, P.-M. Kleniati, and B. Rustem. Pessimistic bilevel optimization. *SIAM Journal on Optimization*, 23(1):353–380, 2013.

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