Numerically tractable optimistic bilevel problems

Lorenzo Lampariello[1] · **Simone Sagratella[2]**

Abstract

We consider a class of optimistic bilevel problems. Specifically, we address bilevel problems in which at the lower level the objective function is fully convex and the feasible set does not depend on the upper level variables. We show that this non-trivial class of mathematical programs is sufficiently broad to encompass significant real-world applications and proves to be numerically tractable. From this respect, we establish that the stationary points for a relaxation of the original problem can be obtained addressing a suitable generalized Nash equilibrium problem. The latter game is proven to be convex and with a nonempty solution set. Leveraging this correspondence, we provide a provably convergent, easily implementable scheme to calculate stationary points of the relaxed bilevel program. As witnessed by some numerical experiments on an application in economics, this algorithm turns out to be numerically viable also for big dimensional problems.

Keywords Bilevel programming · Generalized nash equilibrium problems (GNEP) · Leader-follower multi-agent games · Solution methods

1 Introduction

Bilevel programming is a powerful modeling tool that is widely used in many fields (see e.g. [6–8, 11, 40]). We focus on the general case in which the lower level problem may have multiple solutions. From that respect, we take the so-called optimistic point of view (for the pessimistic counterpart see, e.g. [24]) that leads us to consider the Standard optimistic Bilevel programming Problem (SBP), as commonly done in the literature (for further details, we refer the interested reader to [23, 42]). The

✉ Simone Sagratella
  sagratella@diag.uniroma1.it

  Lorenzo Lampariello
  lorenzo.lampariello@uniroma3.it

[1] Department of Business Studies, Roma Tre University, Rome, Italy

[2] Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza University of Rome, Rome, Italy

latter problem is structurally nonconvex and nonsmooth; furthermore, it is hard to define suitable Constraint Qualification (CQ) conditions for it, see, e.g. [41]. In fact, the study of provably convergent and practically implementable algorithms for its solution is still in its infancy (see, e.g. the references in [8, 23]), as also witnessed by the scarcity of results in the literature. To date, the most studied and promising approaches rely on either the KKT or the optimal value one-level reformulations. Replacing the lower level problem by its KKT conditions, thus obtaining a Mathematical Program with Complementarity Constraints (MPCC), can be done only when the lower level problem is convex and such that Slater's constraint qualification holds. However, even in this case, the set of local optima for the SBP is only a (often strict) subset of (the projection of) the set of local optima for the MPCC, which is what one actually craves to compute (since the MPCC is nonconvex) (see Sect. 2 and, in particular, Example 2). On the other hand, the optimal value reformulation is obtained by replacing the lower level solution set by its description via the optimal value function; in this case, while the resulting mathematical program is still a nonconvex, nonsmooth, implicitly defined optimization problem for which standard CQs are not readily at hand, the resulting reformulation and the original SBP lead to the same global and local solutions. There exist few recent solution procedures (e.g. [10, 27]) that leverage the optimal value reformulation but they are viable only for small dimensional problems.

Motivated by common situations in real-world applications (see Sect. 5), we address the nontrivial and wide class of fully convex lower level SBPs whose feasible set does not depend on the upper level variables (see Definition 1 and Example 1) by relying on the value function approach and suitably relaxing the optimal value function constraint.

In Sect. 3, we show that stationary points of the relaxed SBP can be obtained by computing equilibria of a suitably defined Generalized Nash Equilibrium Problem (GNEP). Such a GNEP is convex, explicitly defined and with a nonempty solution set. It should be remarked that the GNEP introduced in Sect. 3 is profoundly different from the one presented in [23], see "Appendix C". In fact, while [23] is intended to provide a theoretical analysis of the relations between optimistic bilevel problems and Nash games in terms of their optimal points, the distinctive properties of the GNEP proposed here, which is tailored to address stationary solutions of the bilevel problem, pave the way to algorithmic developments. In Sect. 4 we present a simple algorithmic scheme to compute equilibria of the GNEP and, hence, stationary points of the relaxed SBP. This procedure turns out to be provably convergent, easily implementable and numerically viable also for big dimensional problems. In Sect. 5 we propose an application in economics involving a problem with two decision levels; under suitable conditions, this mathematical program is proven to belong to the class of fully convex lower level SBPs. Our numerical results show that the points provided by our procedure are not only stationary for the relaxed SBP but, in most cases, good approximations of global optima for the original SBP.

For the sake of presentation, we first describe our method not in its full generality. In fact, in order to better highlight the core simple idea underlying the approach, we defer to the appendix important theoretical results that widen the scope of applicability of the procedure. Specifically, "Appendix A" is devoted to the study of the

behavior of the solutions for the relaxed SBP as the relaxation parameter vanishes, while "Appendix B" extends the convergence results of our algorithm whenever the upper level objective is nonconvex and the upper level feasible set is unbounded.

In the paper we employ standard notation. However, for the reader's convenience, we remark that, considering $h : \mathbb{R}^{n_0} \times \mathbb{R}^{n_1} \to \mathbb{R}$, 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. Furthermore, with C convex set and $z \in C$, $N_C(z)$ is the classical normal cone (to C at z) of convex analysis (see e.g. [32, Chapter 6]). As for the definitions of semicontinuity and other properties of single and set-valued mappings such as lower semicontinuity, outer and inner semicontinuity, and local boundedness, we refer the reader to [32]: in particular, for outer and inner semicontinuity, see [32, Definition 5.4].

2 A class of standard optimistic bilevel problems

Let us consider the Standard optimistic Bilevel programming Problem (SBP)

$$
\begin{aligned}
\underset{x,y}{\text{minimize}} \quad & F(x, y) \\
\text{s.t.} \quad & x \in X \\
& y \in S(x),
\end{aligned} \tag{1}
$$

where $F : \mathbb{R}^{n_0} \times \mathbb{R}^{n_1} \to \mathbb{R}$ is continuously differentiable, $X \subseteq \mathbb{R}^{n_0}$ is a compact non-empty set, and the set-valued mapping $S : \mathbb{R}^{n_0} \rightrightarrows \mathbb{R}^{n_1}$ is defined (implicitly) as the solution set of the following lower level parametric optimization problem:

$$
\begin{aligned}
\underset{w}{\text{minimize}} \quad & f(x, w) \\
\text{s.t.} \quad & w \in U,
\end{aligned} \tag{2}
$$

where $f : \mathbb{R}^{n_0} \times \mathbb{R}^{n_1} \to \mathbb{R}$ is continuously differentiable and $U \subseteq \mathbb{R}^{n_1}$ is a compact nonempty set.

We address the case in which the lower level's parameter dependence only stems from its parametric objective function: thus, problem (1) is a so-called Stackelberg game (see [23]). Moreover, in view of the continuity of function f and the compactness of set U, $S(x)$ turns out to be nonempty for every choice of $x \in X$ and may contain infinitely many points.

We make the following blanket assumptions:

- X and U are convex nonempty compact sets;
- F is convex on $X \times U$;
- the lower level problem (2) is convex, e.g. when $f(x, \bullet)$ is convex for each $x \in X$.

We remark that, while the assumption on F, as long as the boundedness of X, can be removed at the price of a more convoluted analysis (see "Appendix B"), the other conditions are quite standard and often invoked when bilevel problems are investigated.

In this paper we focus on the class of the so-called fully convex lower level bilevel programming problems (see e.g. [10]).

Definition 1 An instance of (1) is called fully convex lower level SBP if f is fully convex, that is $f(\bullet, \bullet)$ is convex on $O \times U$, where O is an open convex set containing X.

From now on, problem (1) is assumed to be a lower level fully convex SBP. We observe that the introduction of the open set O in Definition 1 is instrumental to obtain useful differentiability properties, see the developments in Proposition 1.

On the one hand, despite all the previous convexity conditions seem reassuring, in general problem (1) is, anyway, nonconvex (see, e.g. [7, 23] and the references therein) and implicitly defined. On the other hand, the requirement of a (fully) convex f may appear strong; nonetheless, in Example 1 we show that, under additional assumptions, one can always make the lower level objective fully convex. This can be done by adding a (sufficiently) strongly convex term that, depending only on x, does not influence the lower level solution set mapping S. Thus, as witnessed by the procedure in the following example and by the application in Sect. 5, the class of fully convex lower level SBPs is sufficiently broad and encompasses many important real-world problems.

Example 1 Assume

$$f(x, w) = f_1(x, w^1) + f_2(w^2) + \frac{\beta}{2} x^T x, \qquad w = \begin{pmatrix} w^1 \\ w^2 \end{pmatrix} \in \mathbb{R}^{n_{1,1} + n_{1,2}}$$

where $f_1 : \mathbb{R}^{n_0} \times \mathbb{R}^{n_{1,1}} \to \mathbb{R}$, $f_2 : \mathbb{R}^{n_{1,2}} \to \mathbb{R}$ are twice continuously differentiable and $n_1 = n_{1,1} + n_{1,2}$, $\beta \in \mathbb{R}$. Note that the lower level solution set $S(x)$ does not depend on the value of β. Thus, we can freely fix β in order to obtain a fully convex f. In fact, supposing that $f_1(\bullet, w^1)$ is convex for every w^1, $f_1(x, \bullet)$ is uniformly strongly convex with constant $\tau > 0$ (independent of x) for every x and f_2 is convex, then, choosing

$$\beta \geq \frac{1}{\tau} \|\nabla^2_{1,2} f_1(x, w^1)\|^2_2, \qquad \forall (x, w^1, w^2) \in X \times U, \tag{3}$$

where $\nabla^2_{1,2} f_1(x, w^1)$ is the transposed Jacobian of $\nabla_2 f_1(\bullet, w^1)$ evaluated at x, we obtain a convex f and, in turn we have an instance of fully convex lower level SBP. We have

$$\nabla^2 f(x, w) = \begin{pmatrix} \nabla^2_{1,1} f_1(x, w^1) + \beta I & \nabla^2_{1,2} f_1(x, w^1) & 0 \\ \nabla^2_{2,1} f_1(x, w^1) & \nabla^2_{2,2} f_1(x, w^1) & 0 \\ 0 & 0 & \nabla^2 f_2(w^2) \end{pmatrix}$$

where $\nabla^2_{1,1}f_1(x, w^1)$ is the Hessian of $f_1(\bullet, w^1)$ evaluated at x, $\nabla^2_{2,2}f_1(x, w^1)$ is the Hessian of $f_1(x, \bullet)$ evaluated at w^1, $\nabla^2 f_2(w^2)$ is the Hessian of $f_2(\bullet)$ evaluated at w^2, and $\nabla^2_{2,1}f_1(x, y) = [\nabla^2_{1,2}f_1(x, y)]^T$. For every $z \in \mathbb{R}^{n_0}$,

$$z^T[\nabla^2_{1,1}f_1(x, w^1) + \beta I]z - z^T[\nabla^2_{1,2}f_1(x, w^1)][\nabla^2_{2,2}f_1(x, w^1)]^{-1}[\nabla^2_{2,1}f_1(x, w^1)]z$$
$$\geq \beta\|z\|^2 - \frac{1}{\tau}z^T[\nabla^2_{1,2}f_1(x, w^1)][\nabla^2_{2,1}f_1(x, w^1)]z \geq 0,$$

where the first inequality holds by the convexity of $f_1(\bullet, w^1)$ and since $f_1(x, \bullet)$ is (uniformly) strongly convex, while the second relation is due to (3). In turn, $\nabla^2 f(x, w) \succeq 0$ by the Schur-complement Theorem and thanks to the convexity of f_2, thus proving the claim.

It goes without saying that, if the original lower level objective does not incorporate the quadratic term $\beta/2\, x^T x$, i.e. $f(x, w) = f_1(x, w^1) + f_2(w^2)$, one can simply add it in order to suitably convexify the function.

We remark that in this case, in view of the presence of the convex term f_2 in f, $S(x)$ does not necessarily reduce to a singleton.

Finally, we also observe that the requirement of a convex $f_1(\bullet, w^1)$ can be relaxed at the price of sufficiently increasing the value of β in order to, roughly speaking, compensate for the degree of nonconvexity of f_1. $\qquad\square$

We recall that, to date, the most studied reformulations of the SBP are the optimal value and the KKT ones (see [11] and the references therein). As for the KKT reformulation, it should be remarked that the SBP has often be considered as a special case of Mathematical Program with Complementarity Constraints (MPCC) (see [28]). Actually, this is not the case, as shown in [9]. Indeed, in general, one can provably recast the SBP as an MPCC only when the lower level problem is convex and such that Slater's constraint qualification holds for all x. More interestingly, even in this case, a local solution of the MPCC, which is what one can expect to compute (since the MPCC is nonconvex), may happen not to lead to a local optimal solution of the corresponding SBP. This problematic aspect occurs even when, as in our framework, one deals with fully convex lower level SBPs: the following very simple example makes this critical issue of the MPCC reformulation apparent.

Example 2 Consider the following standard optimistic bilevel problem:

$$\begin{aligned}
\underset{x,y}{\text{minimize}} \quad & x \\
\text{s.t.} \quad & x \in [-1, 1] \\
& y \in S(x),
\end{aligned} \tag{4}$$

where the set-valued mapping $S : \mathbb{R} \rightrightarrows \mathbb{R}^2$ is defined as the solution set of the following lower level parametric optimization problem:

$$\underset{y_1,y_2}{\text{minimize}} \quad (y_1 - x)^2 + (y_2 + 1)^2$$
$$\text{s.t.} \quad y_1^3 - y_2 \leq 0 \tag{5}$$
$$-y_2 \leq 0.$$

Problem (4) is an instance of fully convex lower level SBPs: the lower level objective is fully convex on $\mathbb{R} \times \mathbb{R}^2$ and $U = \left\{ (y_1, y_2) \in \mathbb{R}^2 \mid y_1^3 - y_2 \leq 0, -y_2 \leq 0 \right\}$ is a convex set. The point $(-1, -1, 0)$ is the unique local (and also global) solution of (4). Moreover, since the Mangasarian-Fromovitz constraint qualification is easily seen to hold for the lower level problem (5) at every feasible (y_1, y_2), we also have [32, Theorem 6.12]

$$N_U(y_1, y_2) = \left\{ \begin{pmatrix} 3\lambda_1 y_1^2 \\ -\lambda_1 - \lambda_2 \end{pmatrix} \middle| (\lambda_1, \lambda_2) \in N_{\mathbb{R}^2_-}(y_1^3 - y_2, -y_2) \right\}.$$

Taking into account the latter relation and recalling that, for any fixed $x \in [-1, 1]$, condition $0 \in \nabla_2 f(x, y) + N_U(y)$ is necessary and sufficient for the feasible solution $y = (y_1, y_2)$ to be (global) optimal for (5), one can suitably introduce the following MPCC reformulation of (4):

$$\underset{x,y_1,y_2,\lambda_1,\lambda_2}{\text{minimize}} \quad x$$
$$\text{s.t.} \quad x \in [-1, 1]$$
$$2(y_1 - x) + 3\lambda_1 y_1^2 = 0$$
$$2(y_2 + 1) - \lambda_1 - \lambda_2 = 0 \tag{6}$$
$$0 \leq \lambda_1 \perp (y_1^3 - y_2) \leq 0$$
$$0 \leq \lambda_2 \perp -y_2 \leq 0.$$

We remark that the point $(-1, -1, 0, 0, 2)$ is globally optimal for problem (6). However, the feasible solution $(0, 0, 0, 1, 1)$ is a local optimum: in fact, considering an open neighborhood of the latter point, we have $\lambda_1, \lambda_2 > 0$ and, in turn, the corresponding constraints are active, implying that $y_1 = y_2 = 0$. By the first equality constraint, we also obtain $x = 0$ and, thus, a locally flat upper level objective function. But point $(0, 0, 0)$, that is the projection of $(0, 0, 0, 1, 1)$ on the (x, y_1, y_2)-space, is just a "non promising" feasible point for (4). To the best of our knowledge, it is still an open question whether the illustrated effect of additional local minima in the MPCC reformulation would persist under stronger assumptions such as requiring the lower level feasible to be described by smooth convex functions satisfying the Slater condition. □

In view of the considerations above, in order to deal with problem (1), we refer to the optimal value reformulation (see [29]), that is

$$\underset{x,y}{\text{minimize}} \quad F(x, y)$$
$$\text{s.t.} \quad x \in X, \ y \in U \tag{SBP}$$
$$f(x, y) \leq \varphi(x),$$

where

$$\varphi(x) \triangleq \min_{w}\{f(x, w) \; : \; w \in U\}$$

is the value function. Also, we denote by $W \triangleq \{(x, y) \in X \times U \,|\, f(x, y) \le \varphi(x)\}$ the feasible set of (SBP). Preliminarily, we observe that, under the initial assumptions, the value function enjoys the following nice properties.

Proposition 1 *Function φ is convex and continuously differentiable on the open convex set O from Definition 1. Moreover, for any $x \in O$, the set $\{\nabla_1 f(x, w) \,|\, w \in S(x)\}$ is a singleton and, for any $w \in S(x)$,*

$$\nabla\varphi(x) = \nabla_1 f(x, w). \tag{7}$$

Proof Problem (2) can be equivalently reformulated as an unconstrained program employing the so-called indicator function $\delta_U(w)$, where $\delta_U(w) = 0$ if $w \in U$ and $\delta_U(w) = \infty$ if $w \notin U$. Since, by our assumptions, function δ_U turns out to be lower semicontinuous and convex, then the convexity of φ follows from [32, Corollary 3.32], being f fully convex (see Definition 1).

In order to show that φ is continuously differentiable on O, leveraging [32, Corollary 9.19], by the convexity of φ, suffice it to prove that the set of subgradients $\partial\varphi(x)$ is single-valued for every $x \in O$. For this to be done, one can rely on, e.g. [39]. However, to maintain the paper self-contained, we show that, for every $x \in O$ and any $w \in S(x)$,

$$\partial\varphi(x) = \{\nabla_1 f(x, w)\}.$$

Let us consider a generic $v \in \partial\varphi(x)$. By the convexity of φ, we have for every $u \in O$

$$f(u, w) \ge \varphi(u) \ge \varphi(x) + v^T(u - x) = f(x, w) + v^T(u - x),$$

where the last equality holds since $w \in S(x)$. It follows that

$$x \in \arg\min_{u \in O} f(u, w) - v^T u.$$

This, due to the convexity of the problem, is equivalent to $0 = \nabla_1 f(x, w) - v$. The claim is a consequence of the arbitrariness of $w \in S(x)$ and $v \in \partial\varphi(x)$. $\qquad\square$

We remark again that (SBP) is a nonconvex and implicitly defined problem for which standard constraint qualifications are not readily at hand (see, e.g, [7, 11]). For this reason, following a well-established path in the literature (see, e.g. [26, 27]), one is naturally led to consider the following perturbed version of (SBP):

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & F(x, y) \\ \text{s.t.} & x \in X, \; y \in U \qquad\qquad \text{(SBP}_\varepsilon\text{)} \\ & f(x, y) \le \varphi(x) + \varepsilon, \end{array}$$

where $\varepsilon > 0$ could be interpreted as a reasonable tolerance on the optimal value of the follower's problem. Furthermore, we indicate with $W_\varepsilon \triangleq \{(x, y) \in X \times U \,|\, f(x, y) \le \varphi(x) + \varepsilon\}$ the feasible set of (SBP$_\varepsilon$).

Problem (SBP$_\varepsilon$) is still nonconvex and implicitly defined. However, while, as suggested above, the Mangasarian-Fromovitz Constraint Qualification (MFCQ) does not hold on W, on the contrary, it is verified on W_ε. Here, for the reader's convenience, we recall the definition of the MFCQ.

Definition 2 (MFCQ) Let $(x, y) \in W_\varepsilon$. We say that the MFCQ holds at (x, y) if the relations

$$
\begin{aligned}
&0 \in \lambda_1 \nabla_1 f(x, y) - \lambda_1 \nabla \varphi(x) + N_X(x) \\
&0 \in \lambda_1 \nabla_2 f(x, y) + N_U(y) \\
&\lambda_1 \in N_{\mathbb{R}_-}(f(x, y) - \varphi(x) - \varepsilon)
\end{aligned}
\tag{8}
$$

imply $\lambda_1 = 0$.

In order to show that the MFCQ is satisfied everywhere on W_ε, suffice it to observe that for λ_1 to be positive, we must have $f(x, y) - \varphi(x) - \varepsilon = 0$, and thus $f(x, y) = \varphi(x) + \varepsilon > \varphi(x)$. In turn, by the convexity of the lower level problem, $0 \notin \nabla_2 f(x, y) + N_U(y)$ that contradicts the second inclusion in (8).

A further motivation to introduce the parameter ε is given in Sect. 5, where a real-world application is considered. Finally, in "Appendix A", we analyze the "behavior" of (SBP$_\varepsilon$) with vanishing values of the perturbation, i.e. for $\varepsilon \downarrow 0$.

For the sake of clarity, we specify the definition of stationary points for (SBP$_\varepsilon$).

Definition 3 (Stationary point) Let (x, y) be feasible for (SBP$_\varepsilon$). We say that (x, y) is a stationary point for (SBP$_\varepsilon$) if a multiplier λ_1 exists such that:

$$
\begin{aligned}
&0 \in \nabla_1 F(x, y) + \lambda_1 \nabla_1 f(x, y) - \lambda_1 \nabla \varphi(x) + N_X(x) \\
&0 \in \nabla_2 F(x, y) + \lambda_1 \nabla_2 f(x, y) + N_U(y) \\
&\lambda_1 \in N_{\mathbb{R}_-}(f(x, y) - \varphi(x) - \varepsilon).
\end{aligned}
\tag{9}
$$

If (x, y) solves (SBP$_\varepsilon$), since it satisfies the MFCQ (8), then it is a stationary point for (SBP$_\varepsilon$). Hence, as standard in nonconvex optimization (see, just as a matter of example in this context, [27]), the aim of our method is to compute a stationary point of (SBP$_\varepsilon$). In particular, in Sect. 3 we show that calculating a stationary point of (SBP$_\varepsilon$) can be accomplished by finding an equilibrium of a suitable Nash equilibrium problem and vice versa. In Sect. 4 we provide a simple scheme to compute such a point with global convergence guarantees.

We observe that, when more general assumptions are considered, e.g. whenever the lower level feasible set depends parametrically on x, some difficulties arise. In particular, the value function φ can no more be expected to be smooth (as opposed to the picture analyzed in Proposition 1). However, even in this more general framework, something can be said, but considering criticality conditions that are more general (usually less sharp) than (9).

3 An equilibrium problem reformulation

We introduce the following Generalized Nash Equilibrium Problem (GNEP) with two players, namely player 1 and player 2:

$$
\begin{array}{ll}
\underset{x,y}{\text{minimize}} \quad F(x,y) & \qquad \underset{u,w}{\text{minimize}} \quad f(x,w) + \frac{1}{2}\|u-x\|^2 \\
\text{s.t.} \quad x \in X, \, y \in U & \qquad \text{s.t.} \quad u \in X, \, w \in U, \\
\quad\quad\quad f(x,y) \le f(u,w) & \\
\quad\quad\quad\quad + \nabla_1 f(u,w)^T (x-u) + \varepsilon &
\end{array}
\qquad (\text{GNEP}_\varepsilon)
$$

where ε is a positive parameter. Note that the quadratic term in the optimization problem of player 2 is introduced with the only purpose of constraining u to be equal to x; thus, $1/2 \, \|u-x\|^2$ is not a proximal term.

We call equilibrium of $(\text{GNEP}_\varepsilon)$ a feasible solution (x, y, u, w) from which no player has incentives to deviate unilaterally (see [14], to which we refer for all the fundamental definitions and tools regarding GNEPs). Here, for the sake of clarity, we recall the Karush–Kuhn–Tucker (KKT) conditions for $(\text{GNEP}_\varepsilon)$.

Definition 4 (*KKT point*) Let (x, y, u, w) be feasible for $(\text{GNEP}_\varepsilon)$. We say that (x, y, u, w) is a KKT point for $(\text{GNEP}_\varepsilon)$, if a multiplier λ_1 exists such that:

$$
\begin{aligned}
& 0 \in \nabla_1 F(x,y) + \lambda_1 \nabla_1 f(x,y) - \lambda_1 \nabla_1 f(u,w) + N_X(x) \\
& 0 \in \nabla_2 F(x,y) + \lambda_1 \nabla_2 f(x,y) + N_U(y) \\
& \lambda_1 \in N_{\mathbb{R}_-}(f(x,y) - f(u,w) - \nabla_1 f(u,w)^T(x-u) - \varepsilon),
\end{aligned}
\qquad (10)
$$

$$
\begin{aligned}
& 0 = u - x \\
& 0 \in \nabla_2 f(x,w) + N_U(w).
\end{aligned}
\qquad (11)
$$

We propose to rely on $(\text{GNEP}_\varepsilon)$ as an instrument to find stationary points of (SBP_ε). In fact, $(\text{GNEP}_\varepsilon)$, while being tightly related to (SBP_ε) (see the following developments), as opposed to (SBP_ε), is convex, that is each player solves a convex (parametric) optimization problem, and, among the constraints, no implicitly defined functions (as φ in (SBP_ε)) appear.

Remark 1 Under our assumptions, $(\text{GNEP}_\varepsilon)$ satisfies all the conditions in the Ichiishi Theorem (see, e.g. [14, Theorem 4.1]) and, thus, is such that an equilibrium always exists. In particular,

(i) any feasible point (x, y, u, w) is such that $(x, y) \in X \times U$ and $(u, w) \in X \times U$, where $X \times U$ is a compact set;

(ii) $(\text{GNEP}_\varepsilon)$ is convex, that is each player solves a convex (parametric) optimization problem, given the strategy of the rival;

(iii) since $X \times U$ is nonempty, then the feasible set of player 2 is nonempty for every choice (x, y) of player 1. At the same time, for any strategy (u, w) of player 2, the point $(x, y) = (u, w)$ always belongs to the feasible set of player 1;

(iv)   since $f(\bullet, \bullet)$ is continuously differentiable and the Slater constraint qualifica-
       tion holds on the feasible set of player 1 for any choice of $(u, w) \in X \times U$, then
       the feasible set of player 1, considered as set-valued mapping, is continuous
       relative to $X \times U$ at every point in $X \times U$ (see [2, Theorems 3.1.1 and 3.1.6]).

We also observe that point (x, y, u, w) is an equilibrium of $(\text{GNEP}_\varepsilon)$ if and only
if it is a KKT point for $(\text{GNEP}_\varepsilon)$: on the one hand, the necessity holds since the
Slater constraint qualification is valid on the feasible set of player 1 for any choice
$(u, w) \in X \times U$; on the other hand, the sufficiency is due to the convexity of the
game.

In the light of the previous considerations, quite naturally one can recover a sta-
tionary point of (SBP_ε) by computing a KKT solution, i.e. an equilibrium (that cer-
tainly exists), of $(\text{GNEP}_\varepsilon)$.

Theorem 3.1 *The following claims hold:*

(i)   *any KKT point (x, y, u, w) of $(\text{GNEP}_\varepsilon)$ is such that (x, y) is stationary for $(\text{SBP}_\varepsilon$
       $)$;*
(ii)   *any stationary point (x, y) for (SBP_ε) is such that (x, y, x, w) is a KKT point
       of $(\text{GNEP}_\varepsilon)$ for any $w \in S(x)$.*

Proof The claims follow observing that (10) and (11) imply (9) with $w \in S(x)$, and
vice versa. This is true thanks to Proposition 1 and observing that $f(u, w) = \varphi(x)$.
$\square$

The previous result proves that the link between $(\text{GNEP}_\varepsilon)$ and (SBP_ε) is strong:
any stationary point of (SBP_ε) can be recovered by computing equilibria of
$(\text{GNEP}_\varepsilon)$. Thus, since one can obtain any stationary point of (SBP_ε) passing through
the "well-behaved" $(\text{GNEP}_\varepsilon)$, Theorem 3.1 paves the way to the algorithmic devel-
opments of the next section.

Example 3 Let us go back to Example 2, where we showed that the point
$(x, y_1, y_2) = (0, 0, 0)$ is nothing more than a "non promising" feasible solution for the
SBP, while resulting from a futile local optimum for the MPCC reformulation. Now
we illustrate how, for any ε, no KKT points $(x, y_1, y_2, u, w_1, w_2)$ of $(\text{GNEP}_\varepsilon)$ exist
such that $x = 0$. It is easy to see that, if $x = 0$, then (11) implies $u = 0$, $w_1 = 0$ and
$w_2 = 0$. By (10), since $N_{[-1,1]}(0) = 0$, we obtain

$$0 = 1 - 2\lambda_1 y_1, \quad 0 = 2\lambda_1 y_1 + 3\lambda_2 y_1^2, \quad \lambda_2 \geq 0,$$

that is an inconsistent system of relations. Even more importantly, reasoning simi-
larly and after some calculations, one can prove that, for any ε, no KKT points of
$(\text{GNEP}_\varepsilon)$ exist such that $x \neq -1$. Thus, our reformulation, for any ε, leads to the
correct optimal value of the upper level objective of the SBP (4), that is -1. Moreo-
ver, any KKT solution of $(\text{GNEP}_\varepsilon)$ is of the type $(-1, y_1, y_2, -1, -1, 0)$ with $y_1^3 \leq 0$,
$-y_2 \leq 0$ and $(y_1 + 1)^2 + (y_2 + 1)^2 - 1 - \varepsilon \leq 0$; in turn, the projection of any such

point on the (x, y_1, y_2)-space, for ε sufficiently small, is an accurate approximation of the global solution for the original SBP (4). $\qquad\square$

Some connections between (optimistic) bilevel problems and a class of Nash games, different from the one proposed here, have been investigated in the recent paper [23]. As shown in "Appendix C", (GNEP$_\varepsilon$) and the Nash game introduced in [23] are different models and, having peculiar relations with the bilevel problem, serve different purposes.

4 A simple convergent scheme for (SBP$_\varepsilon$)

Leveraging the results in Sect. 3, one can find a stationary point of (SBP$_\varepsilon$) by computing an equilibrium of (GNEP$_\varepsilon$). Hence, in principle, under further conditions, one can resort to many numerical methods for solving GNEPs (see, e.g. [1, 12–17, 20, 22, 25, 30, 31, 37]). Here, exploiting the peculiar structure of (GNEP$_\varepsilon$), and without requiring any additional assumption, we propose a new simple Gauss–Seidel-like procedure to calculate one of its equilibria, which in turn is stationary for (SBP$_\varepsilon$). We remark that, as highlighted in e.g. [14], without such special structures, which consist in the sequential optimization of the single players' problems, do not work properly for GNEPs.

Given the current iterate (x^k, y^k, w^k), the core steps of the scheme consist in computing successively (x^{k+1}, y^{k+1}) by solving the strongly convex program

$$
\begin{array}{ll}
\underset{x,y}{\text{minimize}} & F(x, y) + \frac{\tau}{2}\|(x, y) - (x^k, y^k)\|^2 \\
\text{s.t.} & x \in X, \ y \in U \\
& f(x, y) \leq f(x^k, w^k) + \nabla_1 f(x^k, w^k)^T(x - x^k) + \varepsilon,
\end{array}
\qquad (\text{P1}_\varepsilon(x^k, y^k, w^k))
$$

where τ is a positive constant, and, then, in calculating w^{k+1} addressing the convex optimization problem

$$
\begin{array}{ll}
\underset{w}{\text{minimize}} & f(x^{k+1}, w) \\
\text{s.t.} & w \in U.
\end{array}
\qquad (\text{P2}\,(x^{k+1}))
$$

Of course, requiring w^{k+1} to be a solution of (P2 (x^{k+1})) is equivalent to say $w^{k+1} \in S(x^{k+1})$. The detailed description of the method is given in the scheme of Algorithm 1.

Algorithm 1: Alternating optimization

Data: $x^0 \in X$, $y^0, w^0 \in S(x^0)$, $k \longleftarrow 0$;

repeat

(S.1) $\quad$ Compute (x^{k+1}, y^{k+1}), solution of $(\text{P1}_\varepsilon(x^k, y^k, w^k))$;

(S.2) $\quad$ Compute w^{k+1}, solution of $(\text{P2}(x^{k+1}))$;

(S.3) $\quad$ $k \longleftarrow k + 1$;

end

First, we observe that $(x^{k+1}, y^{k+1}) = (x^k, y^k)$ is equivalent to say that (x^k, y^k) is the solution of (P1$_\varepsilon(x^k, y^k, w^k)$).

Proposition 2 *A point $(\bar{x}, \bar{y})$ solves $(P1_\varepsilon(\bar{x}, \bar{y}, \bar{w}))$, with $\bar{w} \in S(\bar{x})$, if and only if $(\bar{x}, \bar{y}, \bar{x}, \bar{w})$ is an equilibrium of $(GNEP_\varepsilon)$. In turn, $(\bar{x}, \bar{y})$ is stationary for (SBP_ε).*

Proof In view of the convexity of the game (see (ii) in Remark 1), $(\bar{x}, \bar{y}, \bar{x}, \bar{w})$ is an equilibrium of $(GNEP_\varepsilon)$ if and only if

$$\nabla F(\bar{x}, \bar{y})^T((x, y) - (\bar{x}, \bar{y})) \geq 0, \qquad \nabla_2 f(\bar{x}, \bar{w})^T(w - \bar{w}) \geq 0$$

for every $(x, y, w) \in X \times U \times U$ such that $f(x, y) \leq f(\bar{x}, \bar{w}) + \nabla_1 f(\bar{x}, \bar{w})^T(x - \bar{x}) + \varepsilon$. In turn, thanks to the convexity of problems $(P1_\varepsilon(\bar{x}, \bar{y}, \bar{w}))$ and $(P2(\bar{x}))$, the previous relations are equivalent to requiring that $(\bar{x}, \bar{y})$ is the unique solution of problem $(P1_\varepsilon(\bar{x}, \bar{y}, \bar{w}))$ and $\bar{w}$ is a solution of $(P2(\bar{x}))$, that is $\bar{w} \in S(\bar{x})$ and the claim readily follows.

The last assertion follows by Theorem 3.1.   □

The convergence properties of the scheme are summarized in Theorem 4.1.

Theorem 4.1 *Let $(\bar{x}, \bar{y}, \bar{w})$ be a cluster point of the sequence $\{(x^k, y^k, w^k)\}$ generated by Algorithm 1. Then, $(\bar{x}, \bar{y}, \bar{x}, \bar{w})$ is an equilibrium for $(GNEP_\varepsilon)$ and, in turn $(\bar{x}, \bar{y})$ is stationary for (SBP_ε).*

Proof Preliminarily, we show that, for every $k \geq 1$, (x^k, y^k) is feasible for $(P1_\varepsilon(x^k, y^k, w^k))$. In view of step (S.1), (x^k, y^k) is a solution, and *a fortiori* feasible, for $P1_\varepsilon(x^{k-1}, y^{k-1}, w^{k-1})$, that is,

$$f(x^k, y^k) \leq f(x^{k-1}, w^{k-1}) + \nabla_1 f(x^{k-1}, w^{k-1})^T(x^k - x^{k-1}) + \varepsilon. \tag{12}$$

The convexity of φ (see Proposition 1) entails $\varphi(x^{k-1}) + \nabla\varphi(x^{k-1})^T(x^k - x^{k-1}) \leq \varphi(x^k)$. Since $w^{k-1} \in S(x^{k-1})$, we get $\varphi(x^{k-1}) = f(x^{k-1}, w^{k-1})$ and $\nabla_1 f(x^{k-1}, w^{k-1}) = \nabla\varphi(x^{k-1})$, by (7). Moreover, since $w^k \in S(x^k)$, we have $\varphi(x^k) = f(x^k, w^k)$. In turn,

$$f(x^{k-1}, w^{k-1}) + \nabla_1 f(x^{k-1}, w^{k-1})^T(x^k - x^{k-1}) \leq f(x^k, w^k).$$

Combining the latter inequality with (12), we obtain

$$f(x^k, y^k) \leq f(x^k, w^k) + \varepsilon = f(x^k, w^k) + \nabla_1 f(x^k, w^k)^T(x^k - x^k) + \varepsilon,$$

and thus (x^k, y^k) is feasible for $(P1_\varepsilon(x^k, y^k, w^k))$. This fact, since (x^{k+1}, y^{k+1}) is the optimal solution for problem $(P1_\varepsilon(x^k, y^k, w^k))$, implies

$$F(x^{k+1}, y^{k+1}) - F(x^k, y^k) \leq -\frac{\tau}{2}\|(x^{k+1}, y^{k+1}) - (x^k, y^k)\|^2. \tag{13}$$

Observing that F is bounded from below on the compact set $X \times U$, the sequence $\{F(x^k, y^k)\}$ converges and, thus,

$$\|(x^{k+1}, y^{k+1}) - (x^k, y^k)\| \to 0. \tag{14}$$

Consider the infinite set of indices $\mathcal{K}$ such that $(x^k, y^k, w^k) \underset{\mathcal{K}}{\to} (\bar{x}, \bar{y}, \bar{w})$. By (14), we have

$$\lim_{\mathcal{K} \ni k \to \infty} (x^{k+1}, y^{k+1}) = (\bar{x}, \bar{y}). \tag{15}$$

Relying on Proposition 2, we are left to show that $(\bar{x}, \bar{y})$ solves (P1$_\varepsilon(\bar{x}, \bar{y}, \bar{w})$) and $\bar{w} \in S(\bar{x})$. In order to do so, leveraging (15), we need to prove that the solution set-valued mapping of (P1$_\varepsilon$) and S are outer semicontinuous. In fact, in view of the continuity of the functions involved, the feasible set of (P1$_\varepsilon$), considered as the set-valued mapping $(X \times U) \cap \{(x, y) \in \mathbb{R}^{n_0} \times \mathbb{R}^{n_1} \,|\, f(x, y) \le f(\bullet, \bullet) + \nabla_1 f(\bullet, \bullet)^T (x - \bullet) + \varepsilon\}$, is [2, Theorem 3.1.1] outer semicontinuous and, thanks to the convexity assumption and since the Slater's constraint qualification always holds on $X \times U \times U$, [2, Theorem 3.1.6] inner semicontinuous relative to $X \times U \times U$ at any point in $X \times U \times U$: hence, it is continuous relative to $X \times U \times U$ at any point in $X \times U \times U$ (see the introduction for references about outer and inner semicontinuity). Then, in view of [2, Theorems 4.3.3 and 3.1.1] and [32, Corollary 5.20], the (single-valued) solution set mapping of (P1$_\varepsilon$) is continuous relative to $X \times U \times U$ at any point in $X \times U \times U$. Reasoning similarly, the set-valued mapping S is outer semicontinuous relative to X at any point in X. $\qquad\square$

The developments in the proof of Theorem 4.1, see in particular (14), and Proposition 2 suggest a possible stopping criterion for Algorithm 1: namely, the procedure can be stopped whenever the measure $\|(x^{k+1}, y^{k+1}) - (x^k, y^k)\|$ is sufficiently small (see the numerical results in Sect. 5.1).

5 Applications in economics

Bilevel programs are widely and fruitfully used to model many real-world problems in economics (see, e.g. [7]). Here we propose an application in economics involving a problem with two decision levels. Under suitable conditions, this mathematical program is proven to belong to the class of fully convex lower level SBPs.

Let us consider a market in which N firms produce the same n goods. Moreover, an independent regulator sets the selling prices of the first n_1 goods, while the remaining $n_2 = n - n_1$ ones have a fixed price. Any firm $v \in \{1, \dots, N\}$ produces the quantities $q^v \in \mathbb{R}^{n_1}$, $\bar{q}^v \in \mathbb{R}^{n_2}$ of the goods, with $(q^v, \bar{q}^v) \in [l^v, u^v]$, where l^v and u^v are suitable bounds. On the other hand, the regulator decides the prices $p \in \mathbb{R}^{n_1}$, while the fixed prices are denoted by $\bar{p} \in \mathbb{R}^{n_2}$. Any firm v has quadratic production costs:

$$\text{Cost}_v(q^v) \triangleq (q^v)^T c^v + \frac{1}{2}(q^v)^T M^v q^v + (\bar{q}^v)^T \bar{c}^v$$

where $c^v \in \mathbb{R}^{n_1}$ and $\bar{c}^v \in \mathbb{R}^{n_2}$ are positive parameters and M^v is a $n_1 \times n_1$ real symmetric positive definite matrix whose minimum (positive) eigenvalue is denoted by

σ_v. Assuming the presence of some shared constraints on the production levels, each firm, trying to minimize its own loss function, addresses the following optimization problem:

$$\underset{q^v, \overline{q}^v}{\text{minimize}} \quad (q^v)^T(c^v - p) + \tfrac{1}{2}(q^v)^T M^v q^v + (\overline{q}^v)^T(\overline{c}^v - \overline{p})$$
$$\text{s.t.} \quad g(q^1, \dots, q^N, \overline{q}^1, \dots, \overline{q}^N) \le 0 \qquad\qquad (16)$$
$$l^v \le (q^v, \ \overline{q}^v) \le u^v,$$

where $g : \mathbb{R}^{nN} \to \mathbb{R}^m$ 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 prices p, we model the game played by the firms as a GNEP. In particular, the GNEP defined by (16) is a generalized potential game, see e.g. [21, 33]. It is well known that, given p, any solution of the following convex optimization problem

$$\underset{q^1, \dots, q^N, \overline{q}^1, \dots, \overline{q}^N}{\text{minimize}} \quad \sum_{v=1}^{N} \left((q^v)^T(c^v - p) + \tfrac{1}{2}(q^v)^T M^v q^v + (\overline{q}^v)^T(\overline{c}^v - \overline{p}) \right) + v(p)$$
$$\text{s.t.} \quad g(q^1, \dots, q^N, \overline{q}^1, \dots, \overline{q}^N) \le 0 \qquad\qquad (17)$$
$$l^v \le (q^v, \ \overline{q}^v) \le u^v, \quad v = 1, \dots, N,$$

for any function $v : \mathbb{R}^{n_1} \to \mathbb{R}$ that depends only on p, is an equilibrium of GNEP (16). Note that the additional term $v(p)$ is introduced in (17) in order to have a fully convex objective function in the same spirit of Example 1; see the forthcoming discussion on how to explicitly choose $v(p)$.

The aim of the regulator is to set the prices $p \in [\overline{l}, \overline{u}]$ to pursue two different targets:

obj1   maintaining p as close as possible to its lower bounds $\overline{l}$;
obj2   entailing the production quantities of the firms $(q^1, \dots, q^N)$ that satisfy the most the customers' demand $d \in \mathbb{R}^{n_1}$ (obj2).

By resorting to classical techniques for multi-objective programming, the regulator minimizes the loss function

$$\kappa \left\| p - \overline{l} \right\|^2 + (1 - \kappa) \left\| \sum_{v=1}^{N} q^v - d \right\|^2,$$

where $\kappa \in [0, 1]$ is a parameter that suitably weights the two objectives. Assuming an optimistic point of view, the regulator solves the following SBP:

$$\underset{p,q^1,\ldots,q^N,\overline{q}^1,\ldots,\overline{q}^N}{\text{minimize}} \; \kappa\|p-\overline{l}\|^2 + (1-\kappa)\left\|\sum_{v=1}^{N} q^v - d\right\|^2$$

$$\begin{aligned}
\text{s.t.} \quad & \overline{l} \leq p \leq \overline{u} \\
& g(q^1,\ldots,q^N,\overline{q}^1,\ldots,\overline{q}^N) \leq 0 \\
& l^v \leq (q^v,\overline{q}^v) \leq u^v, \quad v = 1,\ldots,N \\
& \sum_{v=1}^{N}\left((q^v)^T(c^v - p) + \tfrac{1}{2}(q^v)^T M^v q^v + (\overline{q}^v)^T(\overline{c}^v - \overline{p})\right) + v(p) \leq \varphi(p),
\end{aligned}$$

$$\tag{18}$$

where φ is the value function of problem (17). We note that problem (18) is an instance of the general framework (SBP), where

- $x = p$, $y = (q^1,\ldots,q^N,\overline{q}^1,\ldots,\overline{q}^N)$
- $F(p,q^1,\ldots,q^N,\overline{q}^1,\ldots,\overline{q}^N) = \kappa\|p-\overline{l}\|^2 + (1-\kappa)\left\|\sum_{v=1}^N q^v - d\right\|^2$
- $X = [\overline{l},\overline{u}]$, $U = \left\{(q^1,\ldots,q^N,\overline{q}^1,\ldots,\overline{q}^N) \mid g(q^1,\ldots,q^N,\overline{q}^1,\ldots,\overline{q}^N) \leq 0,\right.$

  $\left. (q^v,\overline{q}^v) \in [l^v,u^v], v = 1,\ldots,N\right\}$
- $f(p,q^1,\ldots,q^N,\overline{q}^1,\ldots,\overline{q}^N) = \sum_{v=1}^{N}\left((q^v)^T(c^v - p) + \tfrac{1}{2}(q^v)^T M^v q^v\right)$

  $+(\overline{q}^v)^T(\overline{c}^v - \overline{p}) + v(p)$.

We observe that function v is most useful (see Example 1). On the one hand, its presence does not alter in any way the problem since v is not depending on the quantities $(q^v,\overline{q}^v)_{v=1}^N$. On the other hand, it allows us to have a fully convex lower level objective. In fact, by choosing $v(p) \triangleq \frac{\beta}{2}p^T p$, and setting $\beta \geq \frac{N}{\min_{v \in \{1,\ldots,N\}}\{\sigma_v\}}$, we obtain (see (3) in Example 1) $\nabla^2 f(p,q^1,\ldots,q^N,\overline{q}^1,\ldots,\overline{q}^N) \succeq 0$, that is, a fully convex f. Thus, problem (18) is a fully convex lower level SBP. Finally, we note that the lower level problem (17), for any fixed p, has in general a non unique optimal solution. From this respect, we recall that in this paper we take the so-called optimistic point of view, as a major departure from the pessimistic robust counterpart (see, e.g. [24]) where the leader wishes to hedge against the worst possible response of the follower.

5.1 Numerical experiments

The numerical results reported here show that the points provided by our method, when applied to problem (18), are not only stationary for its relaxed version but, in most cases, good approximations of its global optima.

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 SNOPT 7.2-8 with default options.

At the lower level, as far as the potential game is concerned, we consider a market with $N = 3$ firms, each producing a total amount of $n_1 = 20$ (10 of low quality (LQ)

Table 1 Problem data

	LQ	HQ
d_i	[900,2100]	[600,1200]
$\bar{l}_i$	[8,12]	[32,36]
$\bar{u}_i$	[43,47]	[63,67]
$\bar{p}_i^v$	[8,12]	[32,36]
l_i^v	0	0
u_i^v	[900,1100]	[450,550]
c_i^v	[4,6]	[20,30]
$\bar{c}_i^v$	[4,8]	[28,32]
M_{ii}^v	[0.04, 0.07]	[0.05,0.08]

Table 2 Numerical results for instance A

κ	1e–4	0.5	1−1e–4
iter	3348	5280	35
time	145	185	< 1
obj1	7.09e+3	3.87e+3	6.11e–4
obj2	2.57e–2	1.36e+6	2.06e+7
profit1	1.33e+5	9.64e+4	3.88e+4
profit2	1.40e+5	1.01e+5	3.83e+4
profit3	1.40e+5	1.01e+5	4.01e+4
qLQ	30555.42	27298.34	17393.35
qHQ	16428.67	15150.69	12365.09
q1	15271.62	13808.90	9741.82
q2	15860.18	14317.68	9965.12
q3	15852.30	14322.45	10051.50
qtot	46984.09	42449.03	29758.44

and 10 of high quality (HQ)) and $n_2 = 10$ goods (5 of low quality (LQ) and 5 of high quality (HQ)).

Data were randomly generated by using the uniform distribution according to Table 1: consistently, we considered three different instances A, B and C.

We remark that in our test problems M^v is assumed to be diagonal.

The lower level affine constraints $g : \mathbb{R}^{90} \to \mathbb{R}^4$ model the following requirements: on the one hand, the total amount of HQ goods (qHQ) must be, at least, 33% of the goods produced by the industry (qtot). On the other hand, the total amount of goods produced by each firm (q1, q2, q3) must not exceed a threshold of 40% of the quantity of goods produced by the industry (qtot).

We remark that, to take into account the different scale between the objectives, we multiply (obj1) $\|p - \bar{l}\|^2$ by 1e+3. Also, we set $\alpha = 10$ and $\beta = \frac{N}{\min_{v \in \{1,\dots,N\}} \{\sigma_v\}}$. We consider $\varepsilon = $ 1e-2: this amounts to having a tolerance (in the approximated problem

Table 3 Numerical results for instance B

κ	1e–4	0.5	1−1e–4
iter	3691	5958	42
time	158	211	< 1
obj1	6.03e+3	3.11e+3	4.70e–4
obj2	2.50e–2	1.21e+6	1.62e+7
profit1	1.24e+5	9.07e+4	4.06e+4
profit2	1.24e+5	8.96e+4	3.68e+4
profit3	1.28e+5	9.29e+4	3.96e+4
qLQ	28665.95	26145.59	18118.37
qHQ	17303.53	15290.86	11505.77
q1	15346.15	13883.29	10086.61
q2	15220.57	13682.98	9665.55
q3	15402.76	13870.18	9871.97
$q$tot	45969.48	41436.45	29624.13

Table 4 Numerical results for instance C

κ	1e–4	0.5	1−1e–4
iter	3970	5003	41
time	176	198	< 1
obj1	6.00e+3	3.37e+3	5.87e–4
obj2	2.05e–2	1.12e+6	1.85e+7
profit1	1.18e+5	8.80e+4	3.61e+4
profit2	1.22e+5	8.92e+4	3.33e+4
profit3	1.27e+5	9.35e+4	3.59e+4
qLQ	30674.66	27749.30	18048.44
qHQ	15647.35	14307.42	11540.32
q1	15221.78	13869.74	9963.31
q2	15508.30	14069.04	9860.09
q3	15591.94	14117.93	9765.36
$q$tot	46322.02	42056.72	29588.77

(SBP$_\varepsilon$)) less than one cent with respect to total profits ranging from 1e+4 to 1e+5 of the currency (see profit1, profit2, profit3 in Tables 2, 3 and 4).

Taking in to account the theoretical results in Sect. 4, we stop the algorithm iterations when the distance $\|(x^{k+1}, y^{k+1}) − (x^k, y^k)\|_\infty$ is less than 1e-3. Finally, we adopt $(x^0, y^0) = (\bar{l}, w^0)$ as starting point of our procedure; we make this choice in order to start from a point that achieves the first objective (obj1) of the regulator, that is setting the prices as low as possible.

Concerning Tables 2, 3 and 4, we consider three different values for κ:

- $\kappa = 1e–4$, which corresponds to having the regulator mainly interested in satisfying the customers' demand (obj2);

- $\kappa = 1-$ 1e–4, which corresponds to having the regulator mainly interested in maintaining p as close as possible to $\bar{l}$ (obj1);
- $\kappa = 0.5$, which corresponds to having equally-weighted preferences.

We report the total number of iterations (iter) and the CPU time in seconds (time) needed by Algorithm 1 in order to meet the stopping criterion. The numerical results show that, in each instance of the problem that we have considered and for $\kappa = $ 1e–4 and $1 - $ 1e–4, the point provided by our procedure is not only stationary for (SBP$_\varepsilon$) but also a good approximation of a global optimum for problem (18), since the optimal value for obj1, when $\kappa = 1 - $ 1e–4, and for obj2, when $\kappa = $1e–4, is clearly 0.

6 Conclusions

We identify a nontrivial and broad class of SBPs that turns out to be numerically tractable. As witnessed by Sect. 5, this class of problems is of practical interest since it can be employed to model complicated multi-agent hierarchical real-world situations. More specifically, we introduce a novel GNEP reformulation of the original bilevel problem. We remark that this reformulation is completely different from the model presented in [23], as shown in "Appendix C". In fact, among other things, as a major departure from [23], here, leveraging the distinctive properties of the novel GNEP reformulation, we are able to devise one of the first (as far as we are aware) efficient algorithmic procedures for SBPs.

As further developments, we aim at extending our approach to multi-leader-follower games. In fact, when applied to these more complicated contexts, our one-level GNEP reformulation technique allows one to put at the same level all the agents (leaders and followers) in order to practically tackle these problems.

Appendix A: What happens when $\varepsilon \downarrow 0$

Although the solution of the perturbed problem (SBP$_\varepsilon$) is significant and meaningful *per se* both in a theoretical and in a practical perspectives, the question arises quite naturally on what happens when the perturbation ε goes to zero (see e.g. [26, 27]).

As previously recalled, the original (SBP) is a nonconvex problem for which standard constraint qualifications are not readily at hand. Thus, the computation of a Fritz-John (FJ) point (in the sense of the following definition (see [5, Theorem 6.1.1])) for (SBP) may seem a reasonable goal.

Definition 5 (*FJ point*) Let (x, y) be feasible for (SBP). We say that (x, y) is a FJ point for (SBP), if multipliers $(\lambda_0, \lambda_1) \in \mathbb{R}_+^2$, not all zero, exist such that:

$$0 \in \lambda_0 \nabla_1 F(x,y) + \lambda_1 \nabla_1 f(x,y) - \lambda_1 \nabla \varphi(x) + N_X(x)$$
$$0 \in \lambda_0 \nabla_2 F(x,y) + \lambda_1 \nabla_2 f(x,y) + N_U(y) \tag{19}$$
$$\lambda_1 \in N_{\mathbb{R}_-}(f(x,y) - \varphi(x)).$$

Unfortunately, by leveraging Proposition 1 and the first order optimality conditions for the lower level problem, one can easily show that any feasible point for (SBP) is also a FJ solution with $\lambda_0 = 0$.

The following proposition gives the theoretical guarantee that Algorithm 1, with $\varepsilon \downarrow 0$, provides us with a FJ point for (SBP): hence, in the worst case, we find a point that at least is feasible for (SBP).

Proposition 3 *Let $\{\varepsilon_k\}$ be a sequence such that $\varepsilon_k \downarrow 0$, and $\{(x^k, y^k)\}$ be a corresponding sequence of stationary points for (SBP_{ε^k}). Then, any accumulation point $(\bar{x}, \bar{y})$ of $\{(x^k, y^k)\}$ is a FJ point for (SBP).*

Proof We assume, without loss of generality, that the whole sequence $\{(x^k, y^k)\}$ converges to $(\bar{x}, \bar{y})$. Let $\lambda_1^k \in \mathbb{R}_+$ be such that

$$-\nabla_1 F(x^k, y^k) - \lambda_1^k \nabla_1 f(x^k, y^k) - \lambda_1^k \nabla \varphi(x^k) \in N_X(x^k)$$
$$-\nabla_2 F(x^k, y^k) - \lambda_1^k \nabla_2 f(x^k, y^k) \in N_U(y^k) \tag{20}$$
$$\lambda_1^k \in N_{\mathbb{R}_-}(f(x^k, y^k) - \varphi(x^k) - \varepsilon^k),$$

that is λ_1^k is a multiplier associated with the stationary point (x^k, y^k).

We recall that the normal cones $N_X(\bullet)$, $N_U(\bullet)$ and $N_{\mathbb{R}_-}(\bullet)$, considered as set-valued mappings, are outer semicontinuous at $\bar{x}$, $\bar{y}$ and $f(\bar{x}, \bar{y}) - \varphi(\bar{x})$, relative to X, U and $\mathbb{R}_-$, respectively (see [32, Proposition 6.6]). In the subsequent developments, these fundamental properties will be freely invoked.

Preliminarily, we note that, by definition, we have $x^k \in X$, $y^k \in U$ and $f(x^k, y^k) - \varphi(x^k) - \varepsilon^k \le 0$ for every k. We distinguish two cases.

(i) Consider a subsequence $\{\lambda_1^k\}_{\mathcal{K}}$ such that $\lambda_1^k \underset{\mathcal{K}}{\to} \bar{\lambda}_1$. Passing to the limit (over $\mathcal{K}$) in (20), we have, by the continuity of the functions involved,

$$-\nabla_1 F(\bar{x}, \bar{y}) - \bar{\lambda}_1 \nabla_1 f(\bar{x}, \bar{y}) - \bar{\lambda}_1 \nabla \varphi(\bar{x}) \in N_X(\bar{x})$$
$$-\nabla_2 F(\bar{x}, \bar{y}) - \bar{\lambda}_1 \nabla_2 f(\bar{x}, \bar{y}) \in N_U(\bar{y}) \tag{21}$$
$$\bar{\lambda}_1 \in N_{\mathbb{R}_-}(f(\bar{x}, \bar{y}) - \varphi(\bar{x})).$$

Hence, $(\bar{x}, \bar{y})$ is a FJ point for problem (SBP) with corresponding multipliers $(\lambda_0, \lambda_1) = (1, \bar{\lambda}_1)$.

(ii) As opposed to case (i), let, without loss of generality, $\lambda_1^k \to \infty$. Dividing both sides of relations (20) by λ_1^k and passing to the limit, we obtain

$$-\nabla_1 f(\bar{x}, \bar{y}) - \nabla \varphi(\bar{x}) \in N_X(\bar{x})$$
$$-\nabla_2 f(\bar{x}, \bar{y}) \in N_U(\bar{y}) \tag{22}$$
$$1 \in N_{\mathbb{R}_-}(f(\bar{x}, \bar{y}) - \varphi(\bar{x})).$$

Thus, $(\bar{x}, \bar{y})$ is a FJ point for problem (SBP) with corresponding multipliers $(\lambda_0, \lambda_1) = (0, 1)$.

□

Clearly, if the sequence of multipliers $\{\lambda_1^k\}$ associated with the stationary point (x^k, y^k) is bounded, then the corresponding cluster point satisfies condition (19) with $\lambda_0 \neq 0$.

Corollary 1 *Let $\{\varepsilon_k\}$ be a sequence such that $\varepsilon_k \downarrow 0$, and $\{(x^k, y^k)\}$ be a corresponding sequence of stationary points for (SBP_{ε^k}). If there exists a bounded sequence of multipliers λ_1^k satisfying (20), then, any accumulation point $(\bar{x}, \bar{y})$ of $\{(x^k, y^k)\}$ is a FJ point for (SBP) with $\lambda_0 \neq 0$.*

As observed in [27, Theorem 4.1], it can be proven that any accumulation point of a sequence $\{(x^k, y^k)\}$ of (inexact) global solutions for (SBP_{ε^k}), as ε^k goes to zero, is globally optimal for (SBP). Of course, this is exactly what one would like to find, but (SBP_{ε^k}) is a nonconvex program and, thus, computing one of its (inexact) global solutions may be impractical. In this sense, the result in the following proposition (which is reminiscent of [10, Theorem 4.4]) fits our approach better.

Proposition 4 *Let $\delta > 0$ and $\{\varepsilon^k\}$ be a sequence such that $\varepsilon^k \downarrow 0$ and $\{(x^k, y^k)\}$ be a corresponding sequence of points belonging to W_{ε^k} such that*

$$F(x^k, y^k) \leq F(x, y), \quad \forall (x, y) \in W_{\varepsilon^k} \cap \mathbb{B}_\delta(x^k, y^k), \tag{23}$$

where $\mathbb{B}_\delta(x^k, y^k) \in \mathbb{R}^{n_0+n_1}$ is the open ball centered in (x^k, y^k) with radius δ.
Then, each accumulation point $(\bar{x}, \bar{y})$ of $\{(x^k, y^k)\}$ is local optimal for (SBP).

Proof First, we note that, by the continuity of the functions involved, W_ε is outer semicontinuous at any $\varepsilon \geq 0$, relative to $\mathbb{R}_+$; hence, we have $(\bar{x}, \bar{y}) \in W_0 = W$. Suppose by contradiction and without loss of generality that $(\widetilde{x}, \widetilde{y}) \in W \cap \mathbb{B}_{\frac{\delta}{2}}(\bar{x}, \bar{y})$ exists such that

$$F(\widetilde{x}, \widetilde{y}) < F(\bar{x}, \bar{y}). \tag{24}$$

Since, without loss of generality, the whole sequence $\{(x^k, y^k)\}$ converges to $(\bar{x}, \bar{y})$, we can say that $\{(x^k, y^k)\} \in \mathbb{B}_{\frac{\delta}{2}}(\bar{x}, \bar{y})$, for every k sufficiently large. This, in turn, entails $(\widetilde{x}, \widetilde{y}) \in \mathbb{B}_\delta(x^k, y^k)$; observing that $W = W_0 \subseteq W_{\varepsilon^k}$, we have also $(\widetilde{x}, \widetilde{y}) \in W_{\varepsilon^k}$ and, thus,

$$F(x^k, y^k) \leq F(\widetilde{x}, \widetilde{y}).$$

The latter relation, passing to the limit, contradicts (24). □

Appendix B: Generalizations

The numerical approach described in Sect. 4, as anticipated in Sect. 2, can be suitably modified (see Algorithm 2) in order to cope also with a nonconvex objective function F and with a possibly unbounded set X. Also, we allow for the possible inexact (iterative) solution of subproblem (P1$_\varepsilon$).

For this to be done, taking inspiration from [18, 34–36, 38], we introduce the following modified version of problem (P1$_\varepsilon(x^k, y^k, w^k)$):

$$\underset{x,y}{\text{minimize}} \quad \widetilde{F}(x, y; x^k, y^k) + \tfrac{\tau}{2}\|(x, y) - (x^k, y^k)\|^2$$
$$\text{s.t.} \quad x \in X, \; y \in U \qquad\qquad \widetilde{\text{P1}}_\varepsilon(x^k, y^k, w^k)$$
$$f(x, y) \le f(x^k, w^k) + \nabla_1 f(x^k, w^k)^T(x - x^k) + \varepsilon,$$

where τ is a positive constant and $\widetilde{F} : (\mathbb{R}^{n_0} \times \mathbb{R}^{n_1}) \times (\mathbb{R}^{n_0} \times \mathbb{R}^{n_1}) \to \mathbb{R}$ is a suitable convex approximation of F at the base point (x^k, y^k) satisfying the following properties:

(I) $\widetilde{F}(\bullet, \bullet; x^k, y^k)$ is convex for every (x^k, y^k);
(II) $\nabla_{12}\widetilde{F}(\bullet, \bullet; \bullet, \bullet)$ is continuous;
(III) $\nabla_{12}\widetilde{F}(x^k, y^k; x^k, y^k) = \nabla F(x^k, y^k)$ for every (x^k, y^k);

where we denote by $\nabla_{12}\widetilde{F}$ the gradient of $\widetilde{F}$ with respect to the first and the second variables blocks. An easy example for the approximating function $\widetilde{F}$ is the linearization of the continuously differentiable function F at the base point (x^k, y^k), i.e. $\widetilde{F}(x, y; x^k, y^k) = F(x^k, y^k) + \nabla F(x^k, y^k)^T[(x, y) - (x^k, y^k)]$. Clearly, one can employ any other possible convex approximant, like second order-type approximations, such that the mild conditions (II) and (III) hold, see [18, 34, 35].

Moreover, let $(\widehat{x}^k, \widehat{y}^k)$ be the unique solution of problem $(\widetilde{\text{P1}}_\varepsilon(x^k, y^k, w^k))$.

Algorithm 2: Alternating optimization, general case

Data: $x^0 \in X$, y^0, $w^0 \in S(x^0)$, $\{\delta^k\} \ge 0$, $\gamma \in (0, 1]$, $k \longleftarrow 0$;
repeat

(S.1a) $\quad$ Solve $(\widetilde{\text{P1}}_\varepsilon(x^k, y^k, w^k))$ with accuracy δ^k:
$\qquad\quad$ find a feasible (v^k, z^k) s.t. $\|(v^k, z^k) - (\widehat{x}^k, \widehat{y}^k)\| \le \delta^k$;
(S.1b) $\quad$ Set $(x^{k+1}, y^{k+1}) = (x^k, y^k) + \gamma[(v^k, z^k) - (x^k, y^k)]$;
(S.2) $\quad$ Compute w^{k+1}, solution of (P2(x^{k+1}));
(S.3) $\quad$ $k \longleftarrow k + 1$;

end

Note that Algorithm 2 differs from Algorithm 1 just in step (S.1). Here, see step (S.1a), we consider the possibly inaccurate computation of the solution of problem $(\widetilde{\text{P1}}_\varepsilon(x^k, y^k, w^k))$. The error δ must obey rather standard rules (see the conditions in Theorem B.1 (i)). Condition $\|(v^k, z^k) - (\widehat{x}^k, \widehat{y}^k)\| \le \delta^k$ in (S.1a) would require us to estimate $\|(v^k, z^k) - (\widehat{x}^k, \widehat{y}^k)\|$. This is possible by resorting to appropriate error bounds (see [19]) which are available for the strongly convex problem

$(\widetilde{P1}_\varepsilon(x^k, y^k, w^k))$. The convergence properties of Algorithm 2 are summarized in the following theorem.

Theorem B.1 *Assume that ∇F is Lipschitz continuous with constant L and the level set $\mathcal{L}_\alpha \triangleq \{(x, y) \in X \times U : F(x, y) \leq \alpha\}$ is bounded for every $\alpha \in \mathbb{R}$. Let $\{(x^k, y^k, w^k)\}$ be the sequence generated by Algorithm 2.*

- (i) *If the step-size γ is bounded away from zero and smaller than $\min\{1, \frac{\tau}{L}\}$ and $\delta^k \leq a \min\{b^k / \|\nabla F(x^k, y^k)\|, c^k\}$ for some non negative a, b^k, c^k with $\sum_k b^k < +\infty$ and $\sum_k (c^k)^2 < +\infty$, then any cluster point $(\bar{x}, \bar{y}, \bar{w})$ of the bounded sequence $\{(x^k, y^k, w^k)\}$ is a KKT point for $(GNEP_\varepsilon)$ and, in turn $(\bar{x}, \bar{y})$ is stationary for (SBP_ε);*
- (ii) *In the exact case, i.e. $\delta^k = 0$ for every k, Algorithm 2 drives $\|(\hat{x}^k, \hat{y}^k) - (x^k, y^k)\|$ below a prescribed tolerance $\rho > 0$ after at most $\mathcal{O}(\rho^{-2})$ iterations.*

Proof

(i) We prove by induction that (x^k, y^k) is feasible for subproblem $(\widetilde{P1}_\varepsilon(x^k, y^k, w^k))$, for every k. Observe that (x^0, y^0) is feasible for $(\widetilde{P1}_\varepsilon(x^0, y^0, w^0))$ by construction. Then, suppose that (x^k, y^k) is feasible for $(\widetilde{P1}_\varepsilon(x^k, y^k, w^k))$. We now show that (x^{k+1}, y^{k+1}) is feasible for $(\widetilde{P1}_\varepsilon(x^{k+1}, y^{k+1}, w^{k+1}))$. In view of step (S.1b) with $\gamma \leq 1$ and thanks to the convexity of problem $(\widetilde{P1}_\varepsilon(x^k, y^k, w^k))$, (x^{k+1}, y^{k+1}) is feasible for $(\widetilde{P1}_\varepsilon(x^k, y^k, w^k))$, that is

$$f(x^{k+1}, y^{k+1}) \leq f(x^k, w^k) + \nabla_1 f(x^k, w^k)^T (x^{k+1} - x^k) + \varepsilon. \tag{25}$$

The convexity of φ (see Proposition 1) entails $\varphi(x^k) + \nabla\varphi(x^k)^T (x^{k+1} - x^k) \leq \varphi(x^{k+1})$. Since $w^k \in S(x^k)$, we have $\varphi(x^k) = f(x^k, w^k)$ and, by (7), $\nabla_1 f(x^k, w^k) = \nabla\varphi(x^k)$. Moreover, since $w^{k+1} \in S(x^{k+1})$, we have $\varphi(x^{k+1}) = f(x^{k+1}, w^{k+1})$. In turn,

$$f(x^k, w^k) + \nabla_1 f(x^k, w^k)^T (x^{k+1} - x^k) \leq f(x^{k+1}, w^{k+1}).$$

Combining the latter inequality with (25), we obtain

$$\begin{aligned} f(x^{k+1}, y^{k+1}) &\leq f(x^{k+1}, w^{k+1}) + \varepsilon \\ &= f(x^{k+1}, w^{k+1}) + \nabla_1 f(x^{k+1}, w^{k+1})^T (x^{k+1} - x^{k+1}) + \varepsilon, \end{aligned}$$

and thus (x^{k+1}, y^{k+1}) is feasible for $(\widetilde{P1}_\varepsilon(x^{k+1}, y^{k+1}, w^{k+1}))$.

By the minimum principle, since (x^k, y^k) and $(\hat{x}^k, \hat{y}^k)$ are feasible and optimal for problem $(\widetilde{P1}_\varepsilon(x^k, y^k, w^k))$, respectively,

$$\left(\nabla_{12} \widetilde{F}(\hat{x}^k, \hat{y}^k; x^k, y^k) + \tau [(\hat{x}^k, \hat{y}^k) - (x^k, y^k)] \right)^T [(x^k, y^k) - (\hat{x}^k, \hat{y}^k)] \geq 0.$$

Hence,

$$\nabla_{12}\widetilde{F}(\widehat{x}^k, \widehat{y}^k; x^k, y^k)^T[(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)] \le -\tau \|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2, \qquad (26)$$

and, in turn,

$$\begin{aligned}
\nabla F(x^k, y^k)^T[(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)] &= \Big[\nabla_{12}\widetilde{F}(\widehat{x}^k, \widehat{y}^k; x^k, y^k) - \nabla_{12}\widetilde{F}(\widehat{x}^k, \widehat{y}^k; x^k, y^k) \\
&\quad + \nabla_{12}\widetilde{F}(x^k, y^k; x^k, y^k)\Big]^T[(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)] \\
&\le -\tau \|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2,
\end{aligned}$$

where the equality follows from condition (III), and the inequality is due to assumption (I) and relation (26). As a consequence,

$$\begin{aligned}
\nabla F(x^k, y^k)^T[(v^k, z^k) - (\widehat{x}^k, \widehat{y}^k) + (\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)] &\le -\tau \|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2 \\
&\quad + \nabla F(x^k, y^k)^T[(v^k, z^k) - (\widehat{x}^k, \widehat{y}^k)].
\end{aligned} \qquad (27)$$

By the descent lemma [3, Proposition A.24] and step (S.1b) of the algorithm, we get

$$\begin{aligned}
F(x^{k+1}, y^{k+1}) &\le F(x^k, y^k) + \gamma \nabla F(x^k, y^k)^T[(v^k, z^k) - (x^k, y^k)] \\
&\quad + \frac{\gamma^2 L}{2}\|(v^k, z^k) - (x^k, y^k)\|^2,
\end{aligned}$$

which, combined with (27), gives

$$\begin{aligned}
F(x^{k+1}, y^{k+1}) - F(x^k, y^k) &\le +\gamma \|\nabla F(x^k, y^k)\|\delta^k - \gamma\tau \|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2 \\
&\quad + \frac{\gamma^2 L}{2}\|(v^k, z^k) - (x^k, y^k)\|^2.
\end{aligned} \qquad (28)$$

Since $\|(v^k, z^k) - (x^k, y^k)\|^2 \le 2\|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2 + 2\|(v^k, z^k) - (\widehat{x}^k, \widehat{y}^k)\|^2 \le 2\|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2 + 2(\delta^k)^2$, from (28) we obtain

$$\begin{aligned}
F(x^{k+1}, y^{k+1}) - F(x^k, y^k) &\le -\gamma(\tau - \gamma L)\|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2 + T^k \\
&= -\gamma \eta\|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2 + T^k,
\end{aligned} \qquad (29)$$

where $\eta \triangleq (\tau - \gamma L) > 0$ since the step-size is bounded away from zero and $\gamma < \min\{1, \frac{\tau}{L}\}$, and $T^k \triangleq \gamma\|\nabla F(x^k, y^k)\|\delta^k + L(\gamma\delta^k)^2$. By the sufficient decrease condition (29), in view of the assumption on any $\mathcal{L}_\alpha$, and observing that $\sum_{k=0}^{\infty} T^k < \infty$, we have $\|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\| \to 0$ and $\{(x^k, y^k, w^k)\}$ turns out to be bounded by [4, Lemma 3.4]. Reasoning similarly to what done in the proof of Theorem 4.1, the assertion follows readily leveraging condition (III) in the limit.

(ii) Taking the sum of iterations up to N in both sides of (29) where $T^k = 0$ for every k, and considering the worst case, that is $\|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\| > \rho$ for every $k \in \{0, \dots, N\}$, we have

$$\rho^2(N+1) < \sum_{k=0}^{N} \|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2 \le \frac{F(x^0, y^0) - F(x^{N+1}, y^{N+1})}{\gamma\eta} \le \frac{F^0 - F^m}{\gamma\eta},$$

where $F^0 \triangleq F(x^0, y^0)$ and F^m is the minimum value attained by the continuous function F on the set $X \times U$. Therefore, in order to maintain the measure

$\|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|$ greater than ρ, the number of iterations cannot exceed the following bound:

$$N + 1 < \frac{F^0 - F^m}{\gamma\eta\rho^2}.$$

In turn, the claim in (ii) is proven.   □

Differently from the convex case (see Algorithm 1), when dealing with a non-convex objective function F, one cannot rely on a unit step-size, in general: for this reason, the presence of γ in step (S.1b) in Algorithm 2 is required. We add that, apart from the constant one, other choices for the step-size are legitimate for our method to converge: in fact, one can prove that also diminishing or Armijo-like step-sizes (see [18]) can be employed in step (S.1b) of the algorithm. Finally, the result in (ii) provides one with a theoretical bound on the maximum number of iterations that are needed in order for the algorithm to satisfy a possible, practical stopping criterion up to an accuracy δ (see also the comments in Sect. 5.1).

Appendix C: On the connections between bilevel problems and Nash games

Referring to the framework addressed in Sect. 2, here we summarize the distinctive properties of (GNEP$_\varepsilon$) compared to those of the Nash model introduced in [23], pointing out their different relations with (SBP$_\varepsilon$).

First, as for (GNEP$_\varepsilon$), we recall that:

– (GNEP$_\varepsilon$) is a convex game that certainly has an equilibrium (see Remark 1);
– equilibria of (GNEP$_\varepsilon$) lead to stationary points of (SBP$_\varepsilon$) and vice versa (see Theorem 3.1);
– any optimal point of (SBP$_\varepsilon$) leads to an equilibrium of (GNEP$_\varepsilon$) (consequence of Theorem 3.1).

The Nash model as introduced in [23] and referred to (SBP$_\varepsilon$) reads as:

$$\begin{array}{llll}
\underset{x,y}{\text{minimize}} & F(x, y) & \underset{w}{\text{minimize}} & f(x, w) \\
\text{s.t.} & x \in X,\ y \in U & \text{s.t.} & w \in U. \\
& f(x, y) \leq f(x, w) + \varepsilon & &
\end{array} \qquad (30)$$

Even at first glance the differences between (GNEP$_\varepsilon$) and (30) are apparent. The problems of player 1 in (GNEP$_\varepsilon$) and (30) differ in the value function constraint. Furthermore, player 2 in (GNEP$_\varepsilon$) controls both u and w, while in (30) only w. Moreover,

– (30) is a non necessarily convex game with a possibly empty equilibrium set (see Example 4);

– stationary points of (SBP$_\varepsilon$) may not lead to equilibria of (30) (see Example 4);
– any equilibrium of (30) leads to an optimal point of (SBP$_\varepsilon$) (see [23, Corollary 3.1]).

In the light of the analysis above, the properties of (GNEP$_\varepsilon$) are completely different from those of (30): in fact, [23] is intended to provide a theoretical analysis of the relations between (optimistic) bilevel problems and Nash games in terms of their optimal solutions, while the peculiar properties of (GNEP$_\varepsilon$), which is tailored to address stationary solutions, pave the way to algorithmic developments for (SBP$_\varepsilon$).

The following example from [23] witnesses the validity of the claims above.

Example 4 Consider the following (SBP$_\varepsilon$) with $\varepsilon > 0$ such that $\sqrt{\varepsilon} < 3 - 2\sqrt{2}$:

$$
\begin{aligned}
&\underset{x,y}{\text{minimize}} \quad x^2 + y^2 \\
&\text{s.t.} \quad x \in [-1, 1], \ y \in [-1, 1] \\
&\qquad (x + y - 1)^2 \leq \varphi(x) + \varepsilon,
\end{aligned}
\tag{31}
$$

where $\varphi(x) \triangleq \min_w \{(x + w - 1)^2 \ : \ w \in [-1, 1]\}$. The unique solution of (31) is $(\bar{x}, \bar{y}) = \left(\frac{1-\sqrt{\varepsilon}}{2}, \frac{1-\sqrt{\varepsilon}}{2}\right)$. In this case GNEP (30) reads as:

$$
\begin{aligned}
&\underset{x,y}{\text{minimize}} \quad x^2 + y^2 && \underset{w}{\text{minimize}} \quad (x + w - 1)^2 \\
&\text{s.t.} \quad x \in [-1, 1], \ y \in [-1, 1] && \text{s.t.} \quad w \in [-1, 1], \\
&\qquad (x + y - 1)^2 \leq (x + w - 1)^2 + \varepsilon
\end{aligned}
\tag{32}
$$

which, clearly is nonconvex due to the presence of the nonconvex value function constraint in player 1's problem. The point $(\bar{x}, \bar{y}, \bar{w}) = \left(\frac{1-\sqrt{\varepsilon}}{2}, \frac{1-\sqrt{\varepsilon}}{2}, \frac{1+\sqrt{\varepsilon}}{2}\right)$, where $\bar{w} = \frac{1+\sqrt{\varepsilon}}{2}$ is the unique solution of the player 2's problem when $x = \bar{x}$, is not an equilibrium for (32). It is actually a matter of calculation to show that the feasible point $(\hat{x}, \hat{y}) = \left(0, \frac{1+\sqrt{\varepsilon}}{2}\right)$ entails a lower value for the objective function of player 1's problem when $w = \bar{w}$. Therefore, since the unique solution of problem (SBP$_\varepsilon$) (which is a fortiori a stationary point for (SBP$_\varepsilon$)) does not lead to an equilibrium of (32), by [23, Corollary 3.1], the set of equilibria of (32) is empty.

On the contrary, (GNEP$_\varepsilon$), that is,

$$
\begin{aligned}
&\underset{x,y}{\text{minimize}} \quad x^2 + y^2 && \underset{u,w}{\text{minimize}} \quad (x + w - 1)^2 + \frac{(u-x)^2}{2} \\
&\text{s.t.} \quad x \in [-1, 1], \ y \in [-1, 1] && \text{s.t.} \quad w \in [-1, 1], \\
&\qquad (x + y - 1)^2 \leq (u + w - 1)^2 \\
&\qquad\quad + 2(u + w - 1)(x - u) + \varepsilon
\end{aligned}
$$

in view of the results in Sect. 3, is convex and the point $(\bar{x}, \bar{y}, \bar{x}, \bar{w})$ is one of its equilibria. □

Summing up, as for global solutions, we sketch the problems' relations in the following scheme.

Global solutions					
GNEP (30) x, y, w	Co. 3.1 [23] $\Rightarrow$	(SBP$_\varepsilon$) x, y	Th. 3.1 $\Rightarrow$	(GNEP$_\varepsilon$) x, y, x, w	
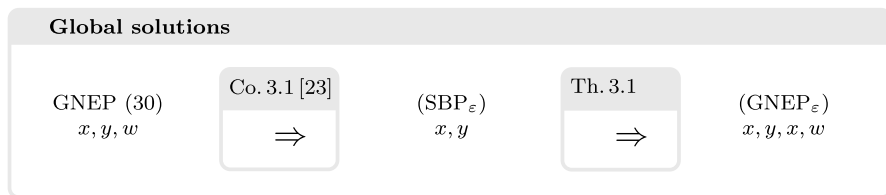
Finally, as significant byproduct, also concerning the global solution sets, (GNEP$_\varepsilon$) nicely completes the picture initiated in [23]. In fact, the (x, y)-part of the equilibrium set of GNEP (30) is a subset of the set of global solutions of (SBP$_\varepsilon$), while the (x, y)-part of the equilibrium set of (GNEP$_\varepsilon$) is a superset of the set of global solutions of (SBP$_\varepsilon$).

References

1. Aussel, D., Sagratella, S.: Sufficient conditions to compute any solution of a quasivariational inequality via a variational inequality. Math. Methods Oper. Res. **85**(1), 3–18 (2017)
2. Bank, B., Guddat, J., Klatte, D., Kummer, B., Tammer, K.: Non-Linear Parametric Optimization. Akademie-Verlag, Berlin (1982)
3. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific, Belmont (1992)
4. Bertsekas, D.P., Tsitsiklis, J.N.: Neuro-Dynamic Programming. Athena Scientific, Belmont (1996)
5. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, Hoboken (1983)
6. Colson, B., Marcotte, P., Savard, G.: An overview of bilevel optimization. Ann. Oper. Res. **153**(1), 235–256 (2007)
7. Dempe, S.: Foundations of Bilevel Programming. Springer, Berlin (2002)
8. Dempe, S.: Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints. Optimization **52**(3), 333–359 (2003)
9. Dempe, S., Dutta, J.: Is bilevel programming a special case of a mathematical program with complementarity constraints? Math. Progr. **131**(1–2), 37–48 (2012)
10. Dempe, S., Franke, S.: On the solution of convex bilevel optimization problems. Comput. Optim. Appl. **63**(3), 685–703 (2016)
11. Dempe, S., Zemkoho, A.B.: The bilevel programming problem: reformulations, constraint qualifications and optimality conditions. Math. Progr. **138**(1–2), 447–473 (2013)
12. Dreves, A., Facchinei, F., Fischer, A., Herrich, M.: A new error bound result for generalized Nash equilibrium problems and its algorithmic application. Comput. Optim. Appl. **59**(1–2), 63–84 (2014)
13. Dreves, A., Facchinei, F., Kanzow, C., Sagratella, S.: On the solution of the KKT conditions of generalized Nash equilibrium problems. SIAM J. Optim. **21**(3), 1082–1108 (2011)
14. Facchinei, F., Kanzow, C.: Generalized Nash equilibrium problems. Ann. Oper. Res. **175**(1), 177–211 (2010)
15. Facchinei, F., Kanzow, C., Karl, S., Sagratella, S.: The semismooth Newton method for the solution of quasi-variational inequalities. Comput. Optim. Appl. **62**(1), 85–109 (2015)
16. Facchinei, F., Kanzow, C., Sagratella, S.: Solving quasi-variational inequalities via their KKT conditions. Math. Progr. **144**(1–2), 369–412 (2014)
17. Facchinei, F., Lampariello, L.: Partial penalization for the solution of generalized Nash equilibrium problems. J. Glob. Optim. **50**(1), 39–57 (2011)
18. Facchinei, F., Lampariello, L., Scutari, G.: Feasible methods for nonconvex nonsmooth problems with applications in green communications. Math. Progr. **164**(1–2), 55–90 (2017)
19. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, Berlin (2003)

20. Facchinei, F., Pang, J.-S., Scutari, G., Lampariello, L.: Vi-constrained hemivariational inequalities: distributed algorithms and power control in ad-hoc networks. Math. Program. **145**(1–2), 59–96 (2014)
21. Facchinei, F., Piccialli, V., Sciandrone, M.: Decomposition algorithms for generalized potential games. Comput. Opt. Appl. **50**(2), 237–262 (2011)
22. Kanzow, C., Steck, D.: Augmented Lagrangian methods for the solution of generalized Nash equilibrium problems. SIAM J. Optim. **26**(4), 2034–2058 (2016)
23. Lampariello, L., Sagratella, S.: A bridge between bilevel programs and Nash games. J. Optim. Theory Appl. **174**(2), 613–635 (2017)
24. Lampariello, L., Sagratella, S., Stein, O.: The standard pessimistic bilevel problem. SIAM J. Optim. **29**(2), 1634–1656 (2019)
25. Latorre, V., Sagratella, S.: A canonical duality approach for the solution of affine quasi-variational inequalities. J. Glob. Optim. **64**(3), 433–449 (2016)
26. Lignola, M.B., Morgan, J.: Stability of regularized bilevel programming problems. J. Optim. Theory Appl. **93**(3), 575–596 (1997)
27. Lin, G.-H., Xu, M., Ye, J.J.: On solving simple bilevel programs with a nonconvex lower level program. Math. Program. **144**(1–2), 277–305 (2014)
28. Luo, Z.-Q., Pang, J.-S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996)
29. Outrata, J.V.: On the numerical solution of a class of Stackelberg problems. Z. Oper. Res. **34**(4), 255–277 (1990)
30. Pang, J.-S., Fukushima, M.: Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. Comput. Manag. Sci. **2**, 21–56 (2009)
31. Pang, J.-S., Scutari, G.: Nonconvex games with side constraints. SIAM J. Optim. **21**(4), 1491–1522 (2011)
32. Rockafellar, R.T., Wets, J.B.: Variational Analysis. Springer, Berlin (1998)
33. Sagratella, S.: Algorithms for generalized potential games with mixed-integer variables. Comput. Optim. Appl. **68**(3), 689–717 (2017)
34. Scutari, G., Facchinei, F., Lampariello, L.: Parallel and distributed methods for constrained nonconvex optimization-part i: Theory. IEEE Trans. Signal Process. **65**(8), 1929–1944 (2017)
35. Scutari, G., Facchinei, F., Lampariello, L., Sardellitti, S., Song, P.: Parallel and distributed methods for constrained nonconvex optimization-part ii: Applications in communications and machine learning. IEEE Trans. Signal Process. **65**(8), 1945–1960 (2017)
36. Scutari, G., Facchinei, F., Lampariello, L., Song, P.: Parallel and distributed methods for nonconvex optimization. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 840–844. IEEE (2014)
37. Scutari, G., Facchinei, F., Pang, J.-S., Lampariello, L.: Equilibrium selection in power control games on the interference channel. In: 2012 Proceedings IEEE INFOCOM, pp. 675–683. IEEE (2012)
38. Song, P., Scutari, G., Facchinei, F., Lampariello, L.: D3m: Distributed multi-cell multigroup multicasting. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3741–3745. IEEE (2016)
39. Tanino, T., Ogawa, T.: An algorithm for solving two-level convex optimization problems. Int. J. Syst. Sci. **15**(2), 163–174 (1984)
40. Von Stackelberg, H.: Marktform und Gleichgewicht. Springer, Berlin (1934)
41. Ye, J.J.: Constraint qualifications and KKT conditions for bilevel programming problems. Math. Oper. Res. **31**(4), 811–824 (2006)
42. Zemkoho, A.B.: Solving ill-posed bilevel programs. Set Valued Var. Anal. **24**(3), 423–448 (2016)