Numerically tractable optimistic bilevel problems

Lorenzo Lampariello[1] and Simone Sagratella[2]

[1]Roma Tre University, Department of Business Studies, Via D'Amico 77, 00145 Rome, Italy, lorenzo.lampariello@uniroma3.it.
[2]Sapienza University of Rome, Department of Computer, Control and Management Engineering Antonio Ruberti, Via Ariosto 25, 00185 Roma, Italy, sagratella@dis.uniroma1.it

November 20, 2017

Abstract

We consider a class of optimistic bilevel problems. Specifically, we address bilevel problems in which the lower level objective function is fully convex. We show that this nontrivial 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 critical 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 critical 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, 7, 8, 11, 32]). 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 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 [21, 34]). The latter problem is structurally nonconvex and nonsmooth; furthermore, it is hard to define suitable Constraint Qualification (CQ) conditions for it, see, e.g., [33]. 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, 21]), 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 Section 2 and, in particular, the very simple Example 2.3). 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, 23]) that leverage the optimal value reformulation but they are viable only for small dimensional problems.

Motivated by common situations in real-world applications (see Section 5), we address the nontrivial and wide class of fully convex lower level SBPs (see Definition 2.1 and Example 2.2) by relying on the value function approach and suitably relaxing the optimal value function constraint.

In Section 3, we show that critical 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. The results in Section 3 are obtained in a similar vein as in [21] but focusing on critical points in lieu of optimal solutions. In Section 4 we present a simple algorithmic scheme to compute equilibria of the GNEP and, hence, critical points of the relaxed SBP. This procedure turns out to be provably convergent, easily implementable and numerically viable also for big dimensional problems. In Section 5 we propose an application in economics involving a two decision levels problem; 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 critical 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.

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. For $\varphi : \mathbb{R}^{n_0} \to \mathbb{R}$, we indicate by $\bar{\partial}\varphi(x)$ the set of Clarke subgradients of $\varphi(\bullet)$ at x. 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. [28, Chapter 6]). As for the definitions of semicontinuity and other properties of single and set-valued mappings such as lower semicontinuity, outer semicontinuity and local boundedness, we refer the reader to [28].

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 nonempty 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 [21]). 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 infinite points.

We make the following blanket assumptions:

- X and U are convex sets;
- F is convex on $X \times U$;
- the lower level problem (2) is convex.

We remark that, while the assumption on F 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 2.1 *We call fully convex lower level SBP an instance of* (1) *in which f is fully convex, that is $f(\bullet, \bullet)$ is convex on $X \times U$.*

From now on, problem (1) is assumed to be a lower level fully convex SBP. On the one hand, despite all the previous convexity conditions seem reassuring, in general problem (1) is, anyway, nonconvex (see, e.g., [7, 21] and the references therein) and implicitly defined. On the other hand, the requirement of a (fully) convex f may appear strong; nonetheless, in Example 2.2 we show that, under mild 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 Section 5, the class of fully convex lower level SBP is sufficiently broad and encompass many important real-world problems.

Example 2.2 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

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

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, $\nabla^2_{1,2} f_1(x,w^1)$ is the transposed Jacobian of $\nabla_2 f_1(\bullet, w^1)$ evaluated at x and $\nabla^2_{2,1} f_1(x,y) = [\nabla^2_{1,2} f_1(x,y)]^T$. For every $z \in \mathbb{R}^{n_0}$,

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

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. □

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 [24]). 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 2.3 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:

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

Note that problem (4) is an instance of fully convex lower level SBPs: in fact, the lower level objective is fully convex on $\mathbb{R} \times \mathbb{R}^2$ and $U = \{(y_1, y_2) \in \mathbb{R}^2 \mid y_1^3 - y_2 \le 0,\ -y_2 \le 0\}$ 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 at every feasible (y_1, y_2), we also have [28, 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):

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

We remark that point $(-1, -1, 0, 0, 2)$ is globally optimal for problem (6). 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 "not promising" feasible point for (4). □

In view of the considerations above, in order to deal with problem (1), we refer to the optimal value reformulation (see [25]), that is

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

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 2.4 *Function φ is convex and locally Lipschitz continuous on $\mathbb{R}^n$. Moreover,*

$$\bar{\partial}\varphi(x) = co\,\{\nabla_1 f(x,w) \,|\, w \in S(x)\}\,. \tag{7}$$

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 [28, Corollary 3.32], being f fully convex (see Definition 2.1).

As for the local Lipschitz continuity and the expression of the (Clarke) subdifferential of φ, the claim is a consequence of Danskin's Theorem (see, e.g., [5, Theorem 9.13]). □

We remark again that (SBP) is a nonconvex, nonsmooth, 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., [22, 23]), one is naturally led to consider the following perturbed version of (SBP):

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

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

Problem (SBP$_\varepsilon$) is still nonconvex, nonsmooth and implicitly defined. However, while, as suggested above, the nonsmooth version of 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 2.5 (MFCQ) *Let $(x,y) \in W_\varepsilon$. We say that the MFCQ holds at (x,y) if the relations*

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

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 $f(x,y) - \varphi(x) - \varepsilon = 0$, but, in turn, by the convexity of the lower level problem, $0 \notin \lambda_1 \nabla_2 f(x,y) + N_U(y)$.

A further motivation to introduce the parameter ε is given in Section 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 critical points for (SBP$_\varepsilon$).

Definition 2.6 (Critical point) *Let (x, y) be feasible for* (SBP$_\varepsilon$). *We say that (x, y) is a critical point for* (SBP$_\varepsilon$) *if a multiplier λ_1 exists such that:*

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

If (x, y) solves (SBP$_\varepsilon$), since it satisfies the MFCQ (8), then it is a critical point for (SBP$_\varepsilon$). Hence, as standard in nonconvex (nonsmooth) optimization (see, just as a matter of example in this context, [23]), the aim of our method is to compute a critical point of (SBP$_\varepsilon$). In particular, in Section 3 we show that calculating a critical point of (SBP$_\varepsilon$) can be accomplished by finding an equilibrium of a suitable Nash equilibrium problem. In Section 4 we provide a simple scheme to compute such a point with global convergence guarantees.

3 An equilibrium problem reformulation

In the same spirit of the results in [21], but aiming now at critical points of (SBP$_\varepsilon$), we introduce the following relevant Generalized Nash Equilibrium Problem (GNEP) with two players, namely player 1 and player 2:

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

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 (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 (GNEP$_\varepsilon$).

Definition 3.1 (KKT point) *Let (x, y, u, w) be feasible for* (GNEP$_\varepsilon$). *We say that (x, y, u, w) is a KKT point for* (GNEP$_\varepsilon$)*, if a multiplier λ_1 exists such that:*

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

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

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

Remark 3.2 Under our assumptions, (GNEP$_\varepsilon$) satisfies all the conditions in 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) the (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 qualification 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 (GNEP$_\varepsilon$) if and only if it is a KKT point for (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 critical point of (SBP$_\varepsilon$) by computing a KKT solution, i.e. an equilibrium (that certainly exists), of (GNEP$_\varepsilon$).

Theorem 3.3 *Any KKT point (x, y, u, w) of (GNEP$_\varepsilon$) is such that (x, y) is critical for (SBP$_\varepsilon$).*

Proof. We only need to show that (10) and (11) imply (9). This is true since, by (7) and (11), $\nabla_1 f(u, w) \in \bar{\partial}\varphi(x)$ and, again thanks to (11), $f(u, w) = \varphi(x)$. □

The previous result proves that the link between (GNEP$_\varepsilon$) and (SBP$_\varepsilon$) is strong: since one can think to obtain a critical point of (SBP$_\varepsilon$) passing through the "well-behaved" (GNEP$_\varepsilon$), Theorem 3.3 paves the way to the algorithmic developments of the next section. We further notice that the claim in Theorem 3.3 still holds if the lower level objective f is not fully convex.

Example 3.4 Let us go back to Example 2.3, in which we show that point $(x, y_1, y_2) = (0, 0, 0)$ is nothing more than a "not promising" feasible solution for the SBP, while being 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 (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 similarly and after some calculations, one can prove that, for any ε, no KKT points of (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. Moreover, any KKT solution of (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). □

The following theorem, providing one with a suitable sufficient condition for the result in Theorem 3.3 to be reversed, deepens the relationship between (GNEP$_\varepsilon$) and (SBP$_\varepsilon$). Specifically, for the classes of problems that satisfy this further assumption, there are not critical points of (SBP$_\varepsilon$) that cannot be recovered by computing equilibria of (GNEP$_\varepsilon$).

Theorem 3.5 *Any critical solution (x, y) of* (SBP$_\varepsilon$)*, for which $\nabla_1 f(x, S(x))$ is a convex set, is such that (x, y, x, w) is a KKT point for* (GNEP$_\varepsilon$) *for some $w \in S(x)$.*

Proof. Since (x, y), along with a multiplier λ_1, is such that relations (9) hold, the second inclusion in (10) corresponds to the second relation in (9), while the third inclusion in (9) entails the third relation in (10) for every (u, w) satisfying (11) . We also have

$$\nabla_1 f(x, S(x)) = \{\nabla_1 f(x, v) : v \in S(x)\} = \text{co}\,\{\nabla_1 f(x, v) : v \in S(x)\} = \bar{\partial}\varphi(x).$$

Hence, for any $\xi \in \bar{\partial}\varphi(x)$, we have $\xi = \nabla_1 f(x, w)$ for some $w \in S(x)$. Therefore, the first relation in (9) for some (u, w) satisfying (11) implies the first condition in (10). □

As stated in the following proposition, of course classes of problems exist such that the assumption in Theorem 3.5 is fulfilled.

Proposition 3.6 *Let $x \in X$ be given. If one of the following conditions holds*

(i) $S(x)$ is a singleton;

(ii) $\nabla_1 f(x, \bullet)$ is affine;

then $\nabla_1 f(x, S(x))$ is a convex set.

Proof. (i) The claim is a straightforward consequence of the fact that, in this case, $\nabla_1 f(x, S(x))$ reduces to a singleton.

(ii) We only need to show that $\text{co}\{\nabla_1 f(x, v) : v \in S(x)\} \subseteq \{\nabla_1 f(x, v) : v \in S(x)\}$. Let $\xi \in \text{co}\{\nabla_1 f(x, v) : v \in S(x)\}$; then, $v_i \in S(x)$ and $\alpha_i \geq 0$, $i = 1, \ldots n_1 + 1$ with $\sum_{i=1}^{n_1+1} \alpha_i = 1$ exist such that

$$\xi = \sum_{i=1}^{n_1+1} \alpha_i \nabla_1 f(x, v_i) = \nabla_1 f(x, \sum_{i=1}^{n_1+1} \alpha_i v_i) = \nabla_1 f(x, w),$$

where the first equality is due to the Caratheodory's Theorem, the second relation holds since $\nabla_1 f(x, \bullet)$ is affine and $\sum_{i=1}^{n_1+1} \alpha_i = 1$, and the last equality follows setting $w = \sum_{i=1}^{n_1+1} \alpha_i v_i$. In turn, by the convexity of $S(x)$, $w \in S(x)$ and $\xi \in \{\nabla_1 f(x, v) : v \in S(x)\}$, thus proving the assertion. □

Again, we remark that the results in Theorem 3.5 and in Proposition 3.6 remain valid also if the full convexity of f is not required.

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

Leveraging the results in Section 3, one can find a critical point of (SBP$_\varepsilon$) by computing an equilibrium of (GNEP$_\varepsilon$). Hence, in principle, one can resort to any numerical method for solving GNEPs (see, e.g., [1, 12, 13, 14, 15, 16, 17, 20, 26, 27]). However, exploiting the peculiar structure of (GNEP$_\varepsilon$), we propose a new simple procedure to calculate one of its equilibria, which in turn is critical for (SBP$_\varepsilon$).

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

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

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: $(x^0, y^0) \in X \times U$, $w^0 \in S(x^0)$, $k \longleftarrow 0$;
repeat
(S.1) Compute (x^{k+1}, y^{k+1}), solution of (P1$_\varepsilon(x^k, y^k, w^k)$);
(S.2) Compute w^{k+1}, solution of (P2(x^{k+1}));
(S.3) **if** $\|(x^{k+1}, y^{k+1}) - (x^k, y^k)\| = 0$ **then**
  stop and **return** $(x^{k+1}, y^{k+1}, w^{k+1})$;
 end
(S.4) $k \longleftarrow k + 1$;
end

Observing 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$_\varepsilon(x^k, y^k, w^k)$), the following proposition gives the theoretical justification for the stopping criterion in (S.3).

Proposition 4.1 *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 critical for* (SBP$_\varepsilon$)*.*

Proof. In view of the convexity of the game (see (ii) in Remark 3.2) and relying on the minimum principle, $(\bar{x}, \bar{y}, \bar{x}, \bar{w})$ is an equilibrium of (GNEP$_\varepsilon$) if and only if

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

for every $(x, y, w) \in X \times U \times U$ such that $f(x,y) \le f(\bar{x}, \bar{w}) + \nabla_1 f(\bar{x}, \bar{w})^T(x - \bar{x}) + \varepsilon$. 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 $(\mathrm{P1}_\varepsilon(\bar{x}, \bar{y}, \bar{w}))$ and $\bar{w}$ is a solution of $(\mathrm{P2}(\bar{x}))$, that is $\bar{w} \in S(\bar{x})$ and the claim readily follows.

The last assertion follows by Theorem 3.3. □

The convergence properties of the scheme are summarized in Theorem 4.2.

Theorem 4.2 *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* $(\mathrm{GNEP}_\varepsilon)$ *and, in turn* $(\bar{x}, \bar{y})$ *is critical for* $(\mathrm{SBP}_\varepsilon)$.

Proof. First, we show that, for every $k \geq 1$, (x^k, y^k) is feasible for $(\mathrm{P1}_\varepsilon(x^k, y^k, w^k))$.

In view of step (S.1), (x^k, y^k) is a solution, and *a fortiori* feasible, for $\mathrm{P1}_\varepsilon(x^{k-1}, y^{k-1}, w^{k-1})$, 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 2.4) entails $\varphi(x^{k-1}) + \xi^T(x^k - x^{k-1}) \leq \varphi(x^k)$, for every $\xi \in \bar{\partial}\varphi(x^{k-1})$. Since $w^{k-1} \in S(x^{k-1})$, we have $\varphi(x^{k-1}) = f(x^{k-1}, w^{k-1})$ and, by (7), $\nabla_1 f(x^{k-1}, w^{k-1}) \in \bar{\partial}\varphi(x^{k-1})$. 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 (25), 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 $(\mathrm{P1}_\varepsilon(x^k, y^k, w^k))$. This fact, since (x^{k+1}, y^{k+1}) is the optimal solution for problem $(\mathrm{P1}_\varepsilon(x^k, y^k, w^k))$, 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 4.1, we are left to show that $(\bar{x}, \bar{y})$ solves $(\mathrm{P1}_\varepsilon(\bar{x}, \bar{y}, \bar{w}))$ 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 $(\mathrm{P1}_\varepsilon)$ and S are outer semicontinuous. In fact, in view of the continuity of the functions involved, the feasible set of $(\mathrm{P1}_\varepsilon)$, considered as the set-valued mapping $(X \times U) \cap \{(x, y) \in \mathbb{R}^{n_0} \times \mathbb{R}^{n_1} \mid f(x, y) \leq f(\bullet, \bullet) + \nabla_1 f(\bullet, \bullet)^T (x - \bullet) + \varepsilon\}$, 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 $X \times U \times U$. Then, in view of [2, Theorems 4.3.3 and 3.1.1] and [28, Corollary 5.20], the (single-valued) solution set mapping of $(\mathrm{P1}_\varepsilon)$ 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. □

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 two decision levels problem. 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 $\nu \in \{1, \dots, N\}$ produces the quantities $q^\nu \in \mathbb{R}^{n_1}$, $\overline{q}^\nu \in \mathbb{R}^{n_2}$ of the goods, with $(q^\nu,\ \overline{q}^\nu) \in [l^\nu, u^\nu]$, where l^ν and u^ν 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 $\overline{p} \in \mathbb{R}^{n_2}$. Any firm ν has quadratic production costs:

$$\text{Cost}_\nu(q^\nu) \triangleq (q^\nu)^T c^\nu + \frac{1}{2}(q^\nu)^T M^\nu q^\nu + (\overline{q}^\nu)^T \overline{c}^\nu$$

where $c^\nu \in \mathbb{R}^{n_1}$ and $\overline{c}^\nu \in \mathbb{R}^{n_2}$ are positive parameters and M^ν is a $n_1 \times n_1$ real symmetric positive definite matrix whose minimum (positive) eigenvalue is denoted by σ_ν. 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:

$$\begin{array}{ll} \underset{q^\nu, \overline{q}^\nu}{\text{minimize}} & (q^\nu)^T(c^\nu - p) + \frac{1}{2}(q^\nu)^T M^\nu q^\nu + (\overline{q}^\nu)^T(\overline{c}^\nu - \overline{p}) \\ \text{s.t.} & g(q^1, \dots, q^N, \overline{q}^1, \dots, \overline{q}^N) \leq 0 \\ & l^\nu \leq (q^\nu,\ \overline{q}^\nu) \leq u^\nu, \end{array} \tag{16}$$

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. [19, 29]. It is well known that, given p, any solution of the following convex optimization problem

$$\begin{array}{ll} \underset{q^1, \dots, q^N, \overline{q}^1, \dots, \overline{q}^N}{\text{minimize}} & \displaystyle\sum_{\nu=1}^N \left((q^\nu)^T(c^\nu - p) + \frac{1}{2}(q^\nu)^T M^\nu q^\nu + (\overline{q}^\nu)^T(\overline{c}^\nu - \overline{p}) \right) + v(p) \\ \text{s.t.} & g(q^1, \dots, q^N, \overline{q}^1, \dots, \overline{q}^N) \leq 0 \\ & l^\nu \leq (q^\nu,\ \overline{q}^\nu) \leq u^\nu, \quad \nu = 1, \dots, N, \end{array} \tag{17}$$

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 2.2.

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

obj1 maintaining p as close as possible to its lower bounds $\bar{l}$;

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

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

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:

$$\begin{array}{cl} \underset{p,q^1,\ldots,q^N,\overline{q}^1,\ldots,\overline{q}^N}{\text{minimize}} & \kappa \left\|p-\bar{l}\right\|^2 + (1-\kappa)\left\|\sum_{\nu=1}^N q^\nu - d\right\|^2 \\ \text{s.t.} & \bar{l} \leq p \leq \bar{u} \\ & g(q^1,\ldots,q^N,\overline{q}^1,\ldots,\overline{q}^N) \leq 0 \\ & l^\nu \leq (q^\nu,\, \overline{q}^\nu) \leq u^\nu, \quad \nu = 1,\ldots,N \\ & \sum_{\nu=1}^N \left((q^\nu)^T(c^\nu - p) + \frac{1}{2}(q^\nu)^T M^\nu q^\nu + (\overline{q}^\nu)^T(\overline{c}^\nu - \overline{p})\right) + v(p) \leq \varphi(p), \end{array} \tag{18}$$

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

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

We observe that function v is most useful (see Example 2.2). On the one hand, its presence does not alter in any way the problem since v is not depending on the quantities $(q^\nu,\overline{q}^\nu)_{\nu=1}^N$. On the other hand, it allows us to have a fully convex lower level objective. In fact, by choosing $v(p) \triangleq \frac{\beta}{2}p^Tp$, and setting $\beta \geq \frac{N}{\min_{\nu\in\{1,\ldots,N\}}\{\sigma_\nu\}}$, we obtain (see (3) in Example 2.2) $\nabla^2 f(p,q^1,\ldots,q^N,\overline{q}^1,\ldots,\overline{q}^N) \succeq 0$, 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, and the class of problems considered here satisfies condition (ii) in Proposition 3.6.

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 critical 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) 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. We remark that in our test problems M^ν 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

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

Table 1: Problem data.

by the industry ($q$tot). 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 ($q$tot).

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 $\beta = \frac{N}{\min_{\nu \in \{1,\dots,N\}}\{\sigma_\nu\}}$. We consider $\varepsilon =$ 1e-2: this amounts to having a tolerance (in the approximated problem (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 2-4).

The algorithm stops if 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 2-4, 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-$1e-4, the point provided by our procedure is not only critical for (SBP$_\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 identified a nontrivial and broad class of SBPs that turns out to be numerically tractable. As witnessed by Section 5, this class of problems is of practical interest since it can be employed to model complicated multi-agent hierarchical real-world situations.

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
$q$tot	46984.09	42449.03	29758.44

Table 2: Numerical results for instance A.

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 3: Numerical results for instance B.

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

Table 4: Numerical results for instance C.

More specifically, we propose a very simple and provably convergent procedure that leverages the reformulation of the original bilevel problem as a suitable one-level GNEP. From that respect, we also define subclasses of SBPs for which there is a complete correspondence between the critical solutions of a relaxation of the bilevel program and the equilibria of the GNEP.

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 quiet naturally on what happens when the perturbation ε goes to zero (see e.g. [22, 23]).

As previously recalled, the original (SBP) is a nonconvex nonsmooth 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 [4, Theorem 6.1.1])) for (SBP) may seem a reasonable goal.

Definition A.1 (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:*

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

Unfortunately, by leveraging (Danskin's) Theorem 2.4 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 A.2 *Let $\{\varepsilon_k\}$ be a sequence such that $\varepsilon_k \downarrow 0$, and $\{(x^k, y^k)\}$ be a corresponding sequence of critical points for* (SBP$_{\varepsilon^k}$)*. 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

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

for some $\xi^k \in \bar{\partial}\varphi(x^k)$, that is λ_1^k is a multiplier associated with the critical 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 [28, Proposition 6.6]). Furthermore, since φ by Proposition 2.4 is locally Lipschitz continuous, $\bar{\partial}\varphi(\bullet)$ is locally bounded and outer semicontinuous at $\bar{x}$ (see [4, Propositions 2.1.2 (a) and 2.1.5 (b)]). 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 $f(x^k, y^k) - \varphi(x^k) - \varepsilon^k \leq 0$ 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,

$$
\begin{aligned}
-\nabla_1 F(\bar{x}, \bar{y}) - \bar{\lambda}_1 \nabla_1 f(\bar{x}, \bar{y}) - \bar{\lambda}_1 \bar{\xi} &\in N_X(\bar{x}) \\
-\nabla_2 F(\bar{x}, \bar{y}) - \bar{\lambda}_1 \nabla_2 f(\bar{x}, \bar{y}) &\in N_U(\bar{y}) \\
\bar{\lambda}_1 &\in N_{\mathbb{R}_-}(f(\bar{x}, \bar{y}) - \varphi(\bar{x})),
\end{aligned}
\tag{21}
$$

for some $\bar{\xi} \in \bar{\partial}\varphi(\bar{x})$.

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

$$
\begin{aligned}
-\nabla_1 f(\bar{x}, \bar{y}) - \bar{\xi} &\in N_X(\bar{x}) \\
-\nabla_2 f(\bar{x}, \bar{y}) &\in N_U(\bar{y}) \\
1 &\in N_{\mathbb{R}_-}(f(\bar{x}, \bar{y}) - \varphi(\bar{x})),
\end{aligned}
\tag{22}
$$

for some $\bar{\xi} \in \bar{\partial}\varphi(\bar{x})$.

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 critical point (x^k, y^k) is bounded, then the corresponding cluster point satisfies condition (19) with $\lambda_0 \neq 0$.

Corollary A.3 *Let $\{\varepsilon_k\}$ be a sequence such that $\varepsilon_k \downarrow 0$, and $\{(x^k, y^k)\}$ be a corresponding sequence of critical points for* (SBP$_{\varepsilon^k}$)*. If there exists a bounded sequence of multipliers λ_1^k satisfying* (20) *for some $\xi^k \in \bar{\partial}\varphi(x^k)$, 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 [23, Theorem 4.1], it can be proven that any accumulation point of a sequence $\{(x^k, y^k)\}$ of (inexact) global solutions for (SBP$_{\varepsilon^k}$), as ε^k goes to zero, is globally optimal for (SBP). Of course, this is exactly what one would like to find, but (SBP$_{\varepsilon^k}$) is a nonconvex (nonsmooth) 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 A.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 $(\widetilde{x}, \widetilde{y}) \in W \cap \mathbb{B}_\delta(\bar{x}, \bar{y})$ 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). □

B The case of nonconvex upper level objectives

The very simple numerical approach described in Section 4 can be suitably modified (see Algorithm 2) in order to cope also with a nonconvex objective function F: so that, at the price of a more convoluted analysis, as anticipated in Section 2, the assumption requiring F to be convex can be removed. For this to be done, taking inspiration from [18, 30, 31], we introduce the following modified version of problem (P1$_\varepsilon(x^k, y^k, w^k)$):

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

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.

Algorithm 2: Alternating optimization, nonconvex case

Data: $x^0 \in X \times U$, y^0, $w^0 \in S(x^0)$, $\gamma \in (0, 1]$, $k \longleftarrow 0$;
repeat
(S.1) Compute $(\widehat{x}^k, \widehat{y}^k)$, solution of $(\widetilde{\text{P1}}_\varepsilon(x^k, y^k, w^k))$;
(S.2) Set $(x^{k+1}, y^{k+1}) = (x^k, y^k) + \gamma\,[(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)]$;
(S.3) Compute w^{k+1}, solution of $(\text{P2}(x^{k+1}))$;
(S.4) **if** $\|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\| = 0$ **then**
  stop and **return** $(x^{k+1}, y^{k+1}, w^{k+1})$;
end
(S.5) $k \longleftarrow k + 1$;
end

The convergence properties of Algorithm 2 are summarized in the following theorem.

Theorem B.1 *Assume that ∇F is Lipschitz continuous with constant L. 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 $\min\{1, \frac{2\tau}{L}\}$, then any cluster point $(\bar{x}, \bar{y}, \bar{w})$ of $\{(x^k, y^k, w^k)\}$ is a KKT point for $(\text{GNEP}_\varepsilon)$ and, in turn $(\bar{x}, \bar{y})$ is critical for (SBP_ε);

(ii) Algorithm 2 drives $\|(\widehat{x}, \widehat{y}) - (x, y)\|$ below a prescribed tolerance $\rho > 0$ after at most $\mathcal{O}(\rho^{-2})$ iterations.

Proof.
(i) First, we show that (x^1, y^1) is feasible for $(\widetilde{\text{P1}}_\varepsilon(x^1, y^1, w^1))$. In view of step (S.1), $(\widehat{x}^0, \widehat{y}^0)$ is a solution, and *a fortiori* feasible, for $(\widetilde{\text{P1}}_\varepsilon(x^0, y^0, w^0))$. Moreover, since (x^0, y^0) is feasible for $(\widetilde{\text{P1}}_\varepsilon(x^0, y^0, w^0))$ by construction, in view of step (S.2) with $\gamma \leq 1$ and thanks to the convexity of problem $(\widetilde{\text{P1}}_\varepsilon(x^0, y^0, w^0))$, (x^1, y^1) is feasible for $(\widetilde{\text{P1}}_\varepsilon(x^0, y^0, w^0))$, that is

$$f(x^1, y^1) \leq f(x^0, w^0) + \nabla_1 f(x^0, w^0)^T (x^1 - x^0) + \varepsilon. \tag{25}$$

The convexity of φ (see Proposition 2.4) entails $\varphi(x^0) + \xi^T(x^1 - x^0) \leq \varphi(x^1)$, for every $\xi \in \bar{\partial}\varphi(x^0)$. Since $w^0 \in S(x^0)$, we have $\varphi(x^0) = f(x^0, w^0)$ and, by (7), $\nabla_1 f(x^0, w^0) \in \bar{\partial}\varphi(x^0)$. Moreover, since $w^1 \in S(x^1)$, we have $\varphi(x^1) = f(x^1, w^1)$. In turn,

$$f(x^0, w^0) + \nabla_1 f(x^0, w^0)^T (x^1 - x^0) \leq f(x^1, w^1).$$

Combining the latter inequality with (25), we obtain

$$f(x^1, y^1) \le f(x^1, w^1) + \varepsilon = f(x^1, w^1) + \nabla_1 f(x^1, w^1)^T (x^1 - x^1) + \varepsilon,$$

and thus (x^1, y^1) is feasible for $(\widetilde{\mathrm{P1}}_\varepsilon(x^1, y^1, w^1))$. Reasoning by induction, let (x^{k-1}, y^{k-1}) be feasible for $(\widetilde{\mathrm{P1}}_\varepsilon(x^{k-1}, y^{k-1}, w^{k-1}))$; since (x^k, y^k) is a convex combination of (x^{k-1}, y^{k-1}) and $(\widehat{x}^{k-1}, \widehat{y}^{k-1})$, then by the convexity of $(\widetilde{\mathrm{P1}}_\varepsilon(x^{k-1}, y^{k-1}, w^{k-1}))$ and $\gamma \le 1$, (x^k, y^k) is feasible for $(\widetilde{\mathrm{P1}}_\varepsilon(x^{k-1}, y^{k-1}, w^{k-1}))$. In view of this fact and exploiting the convexity of φ, (x^k, y^k) turns out to be feasible for $(\widetilde{\mathrm{P1}}_\varepsilon(x^k, y^k, w^k))$ following exactly the same reasoning as above.

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

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

Hence,

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

and, in turn,

$$\begin{aligned} \nabla F(x^k, y^k)^T [(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)] &= \Big[\nabla_{12}\widetilde{F}(\widehat{x}^k, \widehat{y}^k; x^k, y^k) - \nabla_{12}\widetilde{F}(\widehat{x}^k, \widehat{y}^k; x^k, y^k) \\ &\quad + \nabla_{12}\widetilde{F}(x^k, y^k; x^k, y^k)\Big]^T [(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)] \\ &\le -\tau\,\|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2, \end{aligned} \tag{27}$$

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

By the descent lemma [3, Proposition A.24] and step (S.2) of the algorithm, we get

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

which, combined with (27), gives

$$\begin{aligned} F(x^{k+1}, y^{k+1}) - F(x^k, y^k) &\le -\gamma\left(\tau - \tfrac{\gamma L}{2}\right)\|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2 \\ &= -\gamma\,\eta \|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\|^2 \\ &= -\tfrac{\eta}{\gamma}\|(x^{k+1}, y^{k+1}) - (x^k, y^k)\|^2, \end{aligned} \tag{28}$$

where $\eta \triangleq \left(\tau - \frac{\gamma L}{2}\right) > 0$ since the step-size is bounded away from zero and $\gamma < \min\{1, \frac{2\tau}{L}\}$.

By the sufficient decrease condition (28) and reasoning as done in the proof of Theorem 4.2, the assertion follows readily leveraging condition (III) in the limit.

(ii) Taking the sum of iterations up to N in both sides of (28), and considering the worst case, that is $\|(\widehat{x}^k, \widehat{y}^k) - (x^k, y^k)\| > \rho$ for every $k \in \{0, \ldots, N\}$, 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 compact 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. $\square$

Differently from the convex case (see Algorithm 1), when dealing with a nonconvex objective function F, one cannot rely on a unit step-size, in general: for this reason, the presence of γ 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.2) of the algorithm.

Furthermore, the possibly inexact (iterative) solution of subproblem ($\widetilde{\text{P1}}_\varepsilon(x^k, y^k, w^k)$) can be contemplated, too (see, again, [18]).

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. *Nonlinear Programming.* Athena Scientific, 1992.

[4] F.H. Clarke. *Optimization and nonsmooth analysis.* Wiley, 1983.

[5] F.H. Clarke, Y.S. Ledyaev, R.J. Stern, and P.R. Wolenski. *Nonsmooth analysis and control theory*, volume 178. Springer Science & Business Media, 2008.

[6] B. Colson, P. Marcotte, and G. Savard. An overview of bilevel optimization. *Annals of operations research*, 153(1):235–256, 2007.

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

[8] S. Dempe. Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints. *Optimization*, 52(3):333–359, 2003.

[9] 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.

[10] S. Dempe and S. Franke. On the solution of convex bilevel optimization problems. *Computational Optimization and Applications*, 63(3):685–703, 2016.

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

[12] A. Dreves, F. Facchinei, A. Fischer, and M. Herrich. A new error bound result for generalized Nash equilibrium problems and its algorithmic application. *Computational Optimization and Applications*, 59(1-2):63–84, 2014.

[13] 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.

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

[15] F. Facchinei, C. Kanzow, S. Karl, and S. Sagratella. The semismooth newton method for the solution of quasi-variational inequalities. *Computational Optimization and Applications*, 62(1):85–109, 2015.

[16] F. Facchinei, C. Kanzow, and S. Sagratella. Solving quasi-variational inequalities via their kkt conditions. *Mathematical Programming*, 144(1-2):369–412, 2014.

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

[18] F. Facchinei, L. Lampariello, and G. Scutari. Feasible methods for nonconvex nonsmooth problems with applications in green communications. *Mathematical Programming*, 164(1–2):55–90, 2017.

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

[20] C. Kanzow and D. Steck. Augmented lagrangian methods for the solution of generalized Nash equilibrium problems. *SIAM Journal on Optimization*, 26(4):2034–2058, 2016.

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

[22] M.B. Lignola and J. Morgan. Stability of regularized bilevel programming problems. *Journal of Optimization Theory and Applications*, 93(3):575–596, 1997.

[23] G-H. Lin, M. Xu, and J.J Ye. On solving simple bilevel programs with a nonconvex lower level program. *Mathematical Programming*, 144(1-2):277–305, 2014.

[24] Z-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical programs with equilibrium constraints.* Cambridge University Press, 1996.

[25] J.V. Outrata. On the numerical solution of a class of Stackelberg problems. *Zeitschrift für Operations Research*, 34(4):255–277, 1990.

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

[27] J.-S. Pang and G. Scutari. Nonconvex games with side constraints. *SIAM Journal on Optimization*, 21(4):1491–1522, 2011.

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

[29] S. Sagratella. Algorithms for generalized potential games with mixed-integer variables. *Computational Optimization and Applications*, pages 1–29, 2017.

[30] G. Scutari, F. Facchinei, and L. Lampariello. Parallel and distributed methods for constrained nonconvex optimization-part i: Theory. *IEEE Transactions on Signal Processing*, 65(8):1929–1944, 2017.

[31] G. Scutari, F. Facchinei, L. Lampariello, S. Sardellitti, and P. Song. Parallel and distributed methods for constrained nonconvex optimization-part ii: Applications in communications and machine learning. *IEEE Transactions on Signal Processing*, 65(8):1945–1960, 2017.

[32] H. Von Stackelberg. *Marktform und Gleichgewicht.* Springer, 1934.

[33] J.J. Ye. Constraint qualifications and KKT conditions for bilevel programming problems. *Mathematics of Operations Research*, 31(4):811–824, 2006.

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