Francisco Facchinei · Houyuan Jiang · Liqun Qi

A smoothing method for mathematical programs with equilibrium constraints

Abstract. The mathematical program with equilibrium constraints (MPEC) is an optimization problem with variational inequality constraints. MPEC problems include bilevel programming problems as a particular case and have a wide range of applications. MPEC problems with strongly monotone variational inequalities are considered in this paper. They are transformed into an equivalent one-level nonsmooth optimization problem. Then, a sequence of smooth, regular problems that progressively approximate the nonsmooth problem and that can be solved by standard available software for constrained optimization is introduced. It is shown that the solutions (stationary points) of the approximate problems converge to a solution (stationary point) of the original MPEC problem. Numerical results showing viability of the approach are reported.

Key words. equilibrium constraints – variational inequality problems – strong monotonicity – optimality conditions – global convergence

1. Introduction

The mathematical program with equilibrium constraints (MPEC) is an optimization problem whose constraints include variational inequalities, see [19]. Such problems play an important role, for example, in the design of transportation networks [20,30], in economic modelling [40], and in shape optimization [21]. The MPEC problem also includes, as a special case, the so called bilevel programming problem, where some variables are restricted to be in the solution set of a parametric convex optimization problem. An extensive bibliography on this topic and its applications can be found in [41]. Further material on the MPEC problem and its applications can also be found in the two volumes [2,28].

We are interested in algorithms for solving the MPEC problem, which is known to be a very difficult problem, being nonsmooth and nonconvex also under very favourable assumptions. Furthermore, the computation of the (generalized) gradients of the constraints can be extremely difficult, if at all possible. Few successful numerical methods have been proposed to date. Among them, penalty techniques have been mostly used, see [1,18,21,23,27]. Some heuristic methods have been suggested, see, e.g., [15,39].

F. Facchinei: Università di Roma "La Sapienza", Dipartimento di Informatica e Sistemistica, Via Buonarroti 12, I-00185 Roma, Italy, e-mail: `soler@peano.dis.uniroma1.it`

H. Jiang: CSIRO Mathematical and Information Science, Private Bag 2, Glen Osmond, SA 5064, Australia, e-mail: `Houyuan.Jiang@cmis.CSIRO.au`

L. Qi: School of Mathematics, Department of Applied Mathematics, University of New South Wales, Sydney, NSW 2052, Australia, e-mail: `l.qi@unsw.edu.au`

Recently, Outrata [33] and Outrata and Zowe [34] converted the MPEC into an unconstrained nonsmooth Lipschitz optimization problem. Then, one may use well developed nonsmooth optimization numerical methods for solving the MPEC. For different methods of solving bilevel programming problems, the interested reader is referred to [41]. Some more recent references include [14, 37].

In this paper, under the same set of assumptions used in [34], we propose a new algorithm for the solution of MPEC problems and study its convergence properties. Furthermore we also develop necessary optimality conditions which are directly relevant to the algorithm considered. Our strategy can be outlined as follows. By using the KKT conditions for the variational inequality constraints, we reformulate the MPEC problem as a one-level nonsmoothly constrained optimization problem. Then, to avoid the difficulty of the nonsmooth constraints, we solve a sequence of smooth, regular problems, which progressively approximate the nonsmooth problem. We prove that the sequence of points so generated is bounded and that every accumulation point is a solution of the MPEC problem.

The approach we propose appears to be new, even when specialized to bilevel programs. In spite of its conceptual and practical simplicity, the numerical results seem to indicate that it is promising from the computational point of view. While our primary aim in this paper is to illustrate the usefulness and viability of a smoothing approach in the MPEC field, we hope to be able to extend this approach to a wider class of problems in the near future. However, we remark that the case considered in this paper already allows us to deal with many practical problems, see [33, 34] and references therein.

The remainder of the paper is organized as follows. In the next section we present some basic assumptions, under which the MPEC is reformulated as an equivalent one-level nonsmooth optimization problem. In Section 3 a smoothing technique is introduced by which this nonsmooth optimization problem is approximated by a sequence of smooth optimization problems. Some important properties of the smoothed problems are also investigated. Section 4 is devoted to studying optimality conditions, which play an important role in subsequent developments. In Section 5 we propose three different algorithms. Global convergence of these algorithms is established under the assumptions stated in Section 2. Numerical results are presented in Section 6 for both academic and practical problems. We conclude the paper with some remarks in the final section. A complete description of the test problems is reported in the Appendix.

2. Problem formulation and preliminaries

We consider the following mathematical program with equilibrium constraints:

$$
\begin{aligned}
\min \ & f(x, y) \\
\text{s.t.} \ & x \in X_{\text{ad}} \\
& y \in \mathcal{S}(x),
\end{aligned}
\tag{1}
$$

where $f : \mathbb{R}^{n+m} \to \mathbb{R}$ is continuously differentiable, $X_{\text{ad}} \subseteq \mathbb{R}^n$ and, for each $x \in X_{\text{ad}}$ and for a continuously differentiable function $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$, $\mathcal{S}(x)$ is the solution

set of the variational inequality (VI) defined by the pair $(F(x, \cdot), C(x))$, i.e., $y \in \mathcal{S}(x)$ if and only if $y \in C(x)$ and

$$(v - y)^T F(x, y) \geq 0, \quad \text{for all } v \in C(x). \tag{2}$$

We assume that $C(x)$ is defined by

$$C(x) := \{y \in \mathbb{R}^m : g_i(x, y) \geq 0, i = 1, \dots, l\} \tag{3}$$

with $g : \mathbb{R}^{n+m} \to \mathbb{R}^l$ twice continuously differentiable and concave in the second variable. We indicate by $I(x, y)$ the set of active constraints, i.e.,

$$I(x, y) := \{i : g_i(x, y) = 0\}.$$

We make the following blanket assumptions (see also [34]):

A1 $C(x) \neq \emptyset$ for all $x \in A$, where A is an open set containing X_{ad}.
A2 $C(x)$ is uniformly compact on A, i.e., there exists an open bounded set $B \subseteq \mathbb{R}^m$ such that for all $x \in A$

$$C(x) \subseteq B.$$

A3 F is uniformly strongly monotone with respect to y on $A \times B$, i.e., there exists a constant $\alpha > 0$ such that for all $(x, y) \in A \times B$ and $d \in \mathbb{R}^m$,

$$d^T \nabla_y F(x, y) d \geq \alpha \|d\|^2.$$

A4 $X_{\text{ad}} \subseteq \mathbb{R}^n$ is nonempty and compact.
A5 At each $x \in X_{\text{ad}}$ and $y \in \mathcal{S}(x)$, the partial gradients $\nabla_y g_i(x, y)$, $i \in I(x, y)$ are linearly independent.

We point out some simple consequences of these assumptions. By A1 and A3, for every $x \in X_{\text{ad}}$, there exists one and only one solution of the lower-level variational inequality VI. Furthermore, by A5, every solution must satisfy the KKT conditions

$$\begin{aligned} F(x, y) - \nabla_y g(x, y)\lambda &= 0, \\ g(x, y) \geq 0, \ \lambda \geq 0, \ \lambda^T g(x, y) &= 0, \end{aligned} \tag{4}$$

where $\lambda \in \mathbb{R}^l$ is a multiplier which, by Assumption A5, is uniquely determined. Note that Assumption A1, together with the above observation, implies that the feasible region of Problem (1) is nonempty. We have seen that to each x we can associate the unique solution of the lower-level variational inequality and the corresponding multiplier. On the other hand, if (x, y, λ) satisfies the KKT conditions (4), then y belongs to $\mathcal{S}(x)$ by the concaveness of g_i with respect to the second argument [20]. Therefore, under our assumptions, the inner VI problem (2) and its KKT conditions (4) are equivalent, so

that the MPEC problem (1) can be reformulated as the following standard nonlinear programming problem:

$$\min \ f(x, y)$$
$$\text{s.t. } \ x \in X_{\text{ad}},$$
$$F(x, y) - \nabla_y g(x, y)\lambda = 0, \tag{5}$$
$$g(x, y) \geq 0, \ \lambda \geq 0, \ \lambda^T g(x, y) = 0.$$

This is a one-level smooth reformulation of Problem (1), and it could be tempting to use this reformulation to solve Problem (1). This approach has actually been followed by some authors, in the case of bilevel programs, see [3,5,6,13]. However, it is easy to see that, in general, Problem (5) does not satisfy any standard constraint qualification (see also the example at the end of Section 4). Furthermore, the complementarity-type constraints are very complicated and difficult to handle, see [14].

We are then forced to further reformulate the MPEC problem (1). To this end we consider the following nonsmooth equivalent reformulation of Problem (5)

$$\min \ f(x, y)$$
$$\text{s.t. } \ x \in X_{\text{ad}},$$
$$F(x, y) - \nabla_y g(x, y)\lambda = 0 \tag{6}$$
$$g(x, y) - z = 0 \tag{7}$$
$$-2 \min(z, \lambda) = 0, \tag{8}$$

where $z \in \mathbb{R}^l$ and the min operator is applied componentwise to the vectors z and λ. If we introduce the function $H_0 : \mathbb{R}^{n+m+l+l} \to \mathbb{R}^{m+l+l}$, defined as

$$H_0(w) := H_0(x, y, z, \lambda) := \begin{pmatrix} F(x, y) - \nabla_y g(x, y)\lambda \\ g(x, y) - z \\ -2 \min(z, \lambda) \end{pmatrix},$$

we can rewrite the above problem more compactly as

$$\min \ f(x, y)$$
(P) $$\text{s.t. } \ x \in X_{\text{ad}},$$
$$H_0(x, y, z, \lambda) = 0.$$

The reason for the introduction of the multiplicative factor -2 before the min operator will become apparent shortly. With respect to Problem (5) we have added a new variable z which, at feasible points, is always equal to $g(x, y)$. This variable is not strictly necessary, but will facilitate some of the proofs. Problem (P) is equivalent to Problem (5) which, in turn, is equivalent to the MPEC problem (1). This is made precise in the following proposition.

Proposition 1. (x^*, y^*) *is a global (local) solution of the MPEC problem (1) if and only if there exists a vector (z^*, λ^*) such that $(x^*, y^*, z^*, \lambda^*)$ is a global (local) solution of Problem (P).*

Proof. The "only if" part. If (x^*, y^*) is a local minimizer of the MPEC problem (1), then $y^* \in \mathcal{S}(x^*)$ and there exists a vector (z^*, λ^*) such that $(x^*, y^*, z^*, \lambda^*)$ is a feasible point of Problem (P) by Assumption A5. Suppose $(x^*, y^*, z^*, \lambda^*)$ is not a local minimizer of (5), i.e., there exists a sequence of points $\{(x^k, y^k, z^k, \lambda^k)\}$, which are feasible to Problem (P) such that $\{(x^k, y^k, z^k, \lambda^k)\} \to (x^*, y^*, z^*, \lambda^*)$ as $k \to \infty$ and $f(x^k, y^k) < f(x^*, y^*)$. Since $y^k \in \mathcal{S}(x^k)$, i.e., (x^k, y^k) is a feasible point of the MPEC (1), this contradicts that (x^*, y^*) is a local minimizer of the MPEC (1).

Let now (x^*, y^*) be a global minimizer of the MPEC (1). By what we have just proved, $(x^*, y^*, z^*, \lambda^*)$ is a local minimizer of Problem (P) for some vector (z^*, λ^*). Suppose now that $(x^*, y^*, z^*, \lambda^*)$ is not a global minimizer of Problem (P). Then there exists a feasible point $(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda})$ such that

$$f(\bar{x}, \bar{y}) < f(x^*, y^*),$$

which, since $(\bar{x}, \bar{y})$ is feasible to the MPEC problem (1), contradicts the global optimality of (x^*, y^*).

The "if" part. Suppose now that $(x^*, y^*, z^*, \lambda^*)$ is a local minimizer of Problem (P). It follows that $y^* \in \mathcal{S}(x^*)$ and (x^*, y^*) is a feasible point of the MPEC problem (1). Suppose (x^*, y^*) is not a local minimizer of the MPEC problem (1). Then there exists a sequence of (x^k, y^k) of points feasible to the MPEC problem (1) such that $(x^k, y^k) \to (x^*, y^*)$ as $k \to \infty$ and for any k

$$f(x^k, y^k) < f(x^*, y^*). \tag{9}$$

Assumption A5 implies the existence of a unique vector (z^k, λ^k) such that $(x^k, y^k, z^k, \lambda^k)$ is a feasible solution of Problem (P) for each k. In view of continuity of g and (7), we also obviously have that $z^k \to z^*$. To conclude the proof then we only need to show that $\lambda^k \to \lambda^*$. In fact, if we prove this, we have that $(x^k, y^k, z^k, \lambda^k) \to (x^*, y^*, z^*, \lambda^*)$ as $k \to \infty$, so that (9) implies that $(x^*, y^*, z^*, \lambda^*)$ is not a local minimizer of Problem (P), which is a contradiction. We then proceed to prove that $\lambda^k \to \lambda^*$. Suppose, by contradiction that this is not true. We consider two cases: (i) $\{\lambda^k\}$ is bounded, (ii) $\{\lambda^k\}$ is unbounded.

(i) In this case we can assume, without loss of generality, that $\lambda^k \to \bar{\lambda} \neq \lambda^*$. By continuity then $(x^*, y^*, z^*, \bar{\lambda})$ is feasible to Problem (P) and $(x^*, y^*, z^*, \bar{\lambda}) \neq (x^*, y^*, z^*, \lambda^*)$. But this is impossible by Assumption A5.

(ii) Since $(x^k, y^k, z^k, \lambda^k)$ is feasible, we have by (6), and for every k,

$$F(x^k, y^k) - \nabla_y g(x^k, y^k)\lambda^k = 0.$$

Dividing by $\|\lambda^k\|$ and passing to the limit, we obtain, assuming, without loss of generality, that $\lambda^k / \|\lambda^k\| \to \tilde{\lambda} \neq 0$,

$$\nabla_y g(x^*, y^*)\tilde{\lambda} = 0. \tag{10}$$

Since by continuity we eventually have $I(x^k, y^k) \subseteq I(x^*, y^*)$, we see that, by (7) and (8), $\bar{\lambda}_i = 0$ if $i \notin I(x^*, y^*)$. But then (10) contradicts Assumption A5 and the proof of the "if" part in the case of local minimum points is complete.

Let now $(x^*, y^*, z^*, \lambda^*)$ be a global minimizer of Problem (P). Since (x^*, y^*) is feasible to the MPEC problem (1), if (x^*, y^*) is not a global solution of the MPEC problem (1), there exists a feasible point $(\bar{x}, \bar{y})$ such that

$$f(\bar{x}, \bar{y}) < f(x^*, y^*).$$

But since we can find a $(\bar{z}, \bar{\lambda})$ such that $(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda})$ is feasible to Problem (P), this contradicts the global optimality of $(x^*, y^*, z^*, \lambda^*)$.

$\square$

Remark 1. We stress that, at this stage, we do not know whether the MPEC problem or the Problem (P) admit a global solution, even if it would be easy to establish such a result. We have just proved that if one of the two problems has a global solution, also the other does. Actually, in Section 4 we shall prove, as a result of a more general analysis, that these problems do have global solutions.

Our strategy will now be to solve a sequence of *smooth,* regular one-level problems which progressively approximate Problem (P). In the next section we introduce the smoothing technique and study some basic features of the smoothed problems. In Section 4 we digress to study optimality conditions for both Problem (P) and its smooth perturbations which are relevant to our approach. In Section 5, instead, we formally state the algorithms and study their convergence properties.

3. Smoothing problem (P)

Let $\mu \in \mathbb{R}$ be a parameter. Define the function $\phi_\mu : \mathbb{R}^2 \to \mathbb{R}$ by

$$\phi_\mu(a, b) := \sqrt{(a - b)^2 + 4\mu^2} - (a + b).$$

The most important property of this function is stated in the following result [25, Lemma 3.1].

Proposition 2. *For every μ we have*

$$\phi_\mu(a, b) = 0 \iff a \geq 0, b \geq 0, ab = \mu^2.$$

Note also that for $\mu = 0$ we have $\phi_\mu(a, b) = -2\min(a, b)$, while, for every $\mu \neq 0$, $\phi_\mu(a, b)$ is a C^∞ function. Furthermore, for every (a, b), $\lim_{\mu \to 0} \phi_\mu(a, b) = -2\min(a, b)$. The function $\phi_\mu(a, b)$ is therefore a smooth perturbation of the min function.

Let us introduce the nonlinear operator $H_\mu : \mathbb{R}^{n+m+l+l} \to \mathbb{R}^{m+l+l}$,

$$H_\mu(w) := H_\mu(x, y, z, \lambda) := \begin{pmatrix} F(x, y) - \nabla_y g(x, y)\lambda \\ g(x, y) - z \\ \Phi_\mu(z, \lambda) \end{pmatrix},$$

where
$$\Phi_\mu(z, \lambda) := \big(\phi_\mu(z_1, \lambda_1), \ldots, \phi_\mu(z_l, \lambda_l)\big)^T \in \mathbb{R}^l.$$

Then, for every $\mu \neq 0$, one may define an optimization problem

(P_μ)
$$\min \ f(x, y)$$
$$\text{s.t.} \ \ x \in X_{\text{ad}},$$
$$H_\mu(x, y, z, \lambda) = 0.$$

Problem (P_μ) may be viewed as a perturbation of Problem (P) with the parameter μ. For any $\mu \neq 0$, (P_μ) is a smooth optimization problem although (P) is nonsmooth in general. It is clear that (P_0) coincides with (P). We denote by $\mathcal{F}_\mu \subseteq \mathbb{R}^{n+m+l+l}$ the feasible set of problem (P_μ).

Before continuing we stress that most of the times we view μ as a parameter, and this explains the notation adopted, where μ is a subscript. However, in some cases, especially in the proofs, we shall view μ as an independent variable. Hence, from this point of view, the function H depends on the five variables (x, y, z, λ, μ).

We note that the one proposed is certainly not the only possible smooth perturbation of the min operator [35, 36]. However, this particular perturbation enjoys several useful properties that we now study. The first result we can prove readily follows from the results in [26].

Proposition 3. *For every μ and for every feasible point of Problem (P_μ) all the generalized Jacobians of H_μ with respect to the variables (y, z, λ) are nonsingular.*

Proof. We note that the function ϕ_μ we introduced in this section is just the negative of the function used in [26]. It is then easy to check that this change does not affect at all the nonsingularity properties studied in [26]. Therefore, for each fixed x, the proposition is an immediate consequence of [26, Theorem 3.5] (when $\mu \neq 0$) and of [26, Lemma 3.12] (when $\mu = 0$).

$\square$

The most interesting result that we want to prove in this section, is that the sets $\mathcal{F}_\mu$ are uniformly compact as long as μ is bounded. To this end we first recall that a locally Lipschitz function is said to be *regular* if (at every point) the directional derivative exists and is equal to its Clarke's directional derivative. Moreover, a vector function is said to be regular if each of its components is regular. The following result holds.

Lemma 1. *For every μ, the function $H_\mu(x, y, z, \lambda)$ (as a function of the variables μ, x, y, z, λ) is locally Lipschitz and regular.*

Proof. Since all the remaining components of the function $H_\mu(x, y, z, \lambda)$ are continuously differentiable, we only need to show that $\phi_\mu(z_i, \lambda_i)$ is locally Lipschitz and regular. Since
$$\phi_\mu(z_i, \lambda_i) = \sqrt{(z_i - \lambda_i)^2 + 4\mu^2} + (-z_i - \lambda_i),$$
and the sum of regular functions is regular, we only have to check that
$$\sqrt{(z_i - \lambda_i)^2 + 4\mu^2}$$

is locally Lipschitz and regular. In turn this latter function can be seen as the composition of the convex, and hence locally Lipschitz and regular, function $h_1 : \mathbb{R}^2 \to \mathbb{R}$, with $h_1(t_1, t_2) = \sqrt{t_1^2 + t_2^2}$, and of the differentiable function $h_2 : \mathbb{R}^3 \to \mathbb{R}^2$, with $h_2(z_i, \lambda_i, \mu) = (z_i - \lambda_i, 2\mu)$, Then $\sqrt{(z_i - \lambda_i)^2 + 4\mu^2}$ is regular by [11, Theorem 2.3.9 (iii)].

$\square$

Using this lemma we are now able to apply the implicit function theorem for locally Lipschitz functions thus obtaining the following result.

Lemma 2. *Let $(\bar{\mu}, \bar{x}, \bar{y}, \bar{z}, \bar{\lambda})$ be such that $H_{\bar{\mu}}(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = 0$. Then, there exist a neighborhood Ω of $(\bar{\mu}, \bar{x})$ and a continuous function $(y, z, \lambda) : \Omega \to \mathbb{R}^{m+l+l}$, such that, for each $(\mu, x) \in \Omega$,*

$$H_\mu(x, y(x, \mu), z(x, \mu), \lambda(x, \mu)) = 0.$$

Proof. This lemma is nothing else but the application of the implicit function theorem for Lipschitz equations, see [11, Corollary to Theorem 7.1.1], to the system $H_\mu(x, y, z, \lambda) = 0$. According to [11, Corollary to Theorem 7.1.1], all we have to do is to check that the projection of $\partial_{\mu xyz\lambda} H_\mu(x, y, z, \lambda)$ on the space of the y, z and λ variables is comprised of nonsingular matrices. To this end, let $\pi_{y,z,\lambda}$ denote the projection operator over the (y, z, λ) space. By [11, Proposition 2.6.2 (e)] and the definition of the projection operator, we have [1]

$$\pi_{y,z,\lambda} \left[\partial H_\mu(x, y, z, \lambda) \right] \subseteq \pi_{y,z,\lambda} \begin{pmatrix} \partial(H_\mu)_1(x, y, z, \lambda) \\ \vdots \\ \partial(H_\mu)_s(x, y, z, \lambda) \end{pmatrix} \subseteq \begin{pmatrix} \pi_{y,z,\lambda} \left[\partial(H_\mu)_1(x, y, z, \lambda) \right] \\ \vdots \\ \pi_{y,z,\lambda} \left[\partial(H_\mu)_s(x, y, z, \lambda) \right] \end{pmatrix},$$

where we set, for sake of brevity, $s = m + l + l$. It follows from the regularity property of H by Lemma 1, and [11, Proposition 2.3.15] that

$$\begin{pmatrix} \pi_{y,z,\lambda} \left[\partial(H_\mu)_1(x, y, z, \lambda) \right] \\ \vdots \\ \pi_{y,z,\lambda} \left[\partial(H_\mu)_s(x, y, z, \lambda) \right] \end{pmatrix} \subseteq \begin{pmatrix} \partial_{yz\lambda}(H_\mu)_1(x, y, z, \lambda) \\ \vdots \\ \partial_{yz\lambda}(H_\mu)_s(x, y, z, \lambda) \end{pmatrix}.$$

Notice that, from the very special structure of the function ϕ_μ, we have

$$\begin{pmatrix} \partial_{yz\lambda}(H_\mu)_1(x, y, z, \lambda) \\ \vdots \\ \partial_{yz\lambda}(H_\mu)_s(x, y, z, \lambda) \end{pmatrix} = \partial_{yz\lambda} H_\mu(x, y, z, \lambda),$$

see [26]. Consequently,

$$\pi_{y,z,\lambda} \left[\partial H_\mu(x, y, z, \lambda) \right] \subseteq \partial_{yz\lambda} H_\mu(x, y, z, \lambda).$$

---

[1] In the following formulas $\partial(H_\mu)_i$ is a set of row vectors, representing the generalized gradient of the i-th component of H_μ. Note that in the rest of the paper, as usual, the generalized gradient of a function is a set of column vectors.

On the other hand, by Proposition 3, all the generalized Jacobians in $\partial_{yz\lambda} H_\mu(x, y, z, \lambda)$ are nonsingular. Then the assertion on the nonsingularity of the projection follows from the above formula.

$\square$

The following lemma is interesting in its own and will be used several times in the sequel.

Lemma 3. *Let $\bar{x}$ in X_{ad} be given. Then, for every μ there exists a unique point in $\mathcal{F}_\mu$ such that its x-part is equal to $\bar{x}$. We denote this point by $\bar{w}_\mu = (\bar{x}, y_\mu(\bar{x}), z_\mu(\bar{x}), \lambda_\mu(\bar{x}))$. Furthermore, $\bar{w}_\mu$ is continuous as a function of μ.*

Proof. Let $\bar{x}$ in X_{ad} be given. We first prove the existence and uniqueness assertion. If $\mu = 0$, the existence and uniqueness follow by Assumptions A1, A2, A3 and A5, as already discussed in the introduction. If $\mu \neq 0$, the assertion follows by [26, Theorem 3.15], taking into account that, as already explained in the proof of Proposition 3, the differences between the functions ϕ_μ used in this paper and in [26] do not influence the properties proved in [26].

We then turn to the continuity property. Let $\bar{x}$ in X_{ad} and μ be given. By the first part of the lemma, there exists a (unique) point $w_\mu = (\bar{x}, y_\mu(\bar{x}), z_\mu(\bar{x}), \lambda_\mu(\bar{x}))$ such that $H_\mu(w_\mu) = 0$. Then, by Lemma 2, a neighborhood Ω of $(\mu, \bar{x})$ and a continuous function $(y, z, \lambda) : \Omega \to \mathbb{R}^{m+l+l}$ exist such that, for each $(\mu, x) \in \Omega$,

$$H_\mu(x, y(x, \mu), z(x, \mu), \lambda(x, \mu)) = 0.$$

Since, by the uniqueness proved in first part of the lemma,

$$y_\mu(\bar{x}) = y(\bar{x}, \mu), \quad z_\mu(\bar{x}) = z(\bar{x}, \mu), \quad \lambda_\mu(\bar{x}) = \lambda(\bar{x}, \mu),$$

the continuity assertion is also proved.

$\square$

We are now ready to prove a result which will turn out to be crucial. It basically asserts that the feasible regions of the perturbed problems (P_μ) are uniformly compact.

Theorem 1. *For every $\tilde{\mu}$, there exists a compact set $D(\tilde{\mu}) \subseteq \mathbb{R}^{n+m+l+l}$ such that $\mathcal{F}_\mu \subseteq D(\tilde{\mu})$ for every $\mu \in [-\tilde{\mu}, \tilde{\mu}]$.*

Proof. We first note that, for every $\mu \in \mathbb{R}$ and for every point $w = (x, y, z, \lambda) \in \mathcal{F}_\mu$ we have, by Assumption A4, that x belongs to the compact set X_{ad}. We also have, by $z = g(x, y)$ and by Proposition 2, that $g(x, y) \geq 0$, which, in view of Assumption A2, implies that y belongs to the bounded set B. This, in turn, taking into account the continuity of the functions g and (7) implies that z belongs to a compact set.

We now prove that also λ is bounded. Suppose that this is not true. Then there exist sequences $\{\mu^k\}$ and $\{x^k, y^k, z^k, \lambda^k\}$ such that $\mu^k \in [-\tilde{\mu}, \tilde{\mu}]$ and $(x^k, y^k, z^k, \lambda^k) \in \mathcal{F}_{\mu^k}$ for all k and $\lim_{k\to\infty} \|\lambda^k\| = \infty$. By the first part of the proof we can then assume, without loss of generality, that $\mu^k \to \bar{\mu} \in [-\tilde{\mu}, \tilde{\mu}]$ and $(x^k, y^k, z^k) \to (\bar{x}, \bar{y}, \bar{z})$. Now, by Lemma 3 there exists a unique point $\bar{w}_{\bar{\mu}} = (\bar{x}, y_{\bar{\mu}}(\bar{x}), z_{\bar{\mu}}(\bar{x}), \lambda_{\bar{\mu}}(\bar{x}))$ such that $\bar{w}_{\bar{\mu}}$ belongs to $\mathcal{F}_{\bar{\mu}}$ and the x-part of $\bar{w}_{\bar{\mu}}$ is $\bar{x}$. In particular the point $\bar{w}_{\bar{\mu}}$ is such

that $H_{\bar{\mu}}(\bar{w}_{\bar{\mu}}) = 0$. On the other hand by Lemma 2, there exists a continuous function $(y, z, \lambda) : \Omega \to \mathbb{R}^{m+l+l}$ from a neighborhood Ω of $(\bar{\mu}, \bar{x})$ such that

$$H_{\mu}(x, y(x, \mu), z(x, \mu), \lambda(x, \mu)) = 0$$

when (x, μ) belongs to Ω. But again by Lemma 3, we shall eventually have

$$y^k = y(x^k, \mu^k), \quad z^k = z(x^k, \mu^k), \quad \lambda^k = \lambda(x^k, \mu^k).$$

Since the function $\lambda(\cdot, \cdot)$ is continuous we have a contradiction to the unboundedness of λ^k.

□

Corollary 1. *For every μ, $\mathcal{F}_{\mu}$ is nonempty and compact.*

Proof. The assertion on the nonemptiness readily follows from Lemma 3. Furthermore $\mathcal{F}_{\mu}$ is closed by the continuity of the function H_{μ} and bounded by Theorem 1.

□

By Corollary 1 and the continuity of f, we have an existence result of solutions to the MPEC problem (1) and to its perturbations.

Corollary 2. *The MPEC problem (1) and all the problems $\mathcal{F}_{\mu}$ have a solution.*

4. Optimality conditions

In this section we consider optimality conditions for the Problems (P_{μ}). These conditions are important in the analysis that will be carried out in the next section and show that the perturbed problems (P_{μ}) always satisfy the KKT conditions and are therefore fairly regular, as opposite to Problem (5). Furthermore, in the case of $\mu = 0$ we shall derive optimality conditions for the MPEC problem and relate these conditions to those of the perturbed problems. Our main aim in deriving optimality conditions for the MPEC problem is the construction of reliable stopping criteria for the algorithms to be described in the next section. There are a few papers where optimality conditions for MPEC problems are dealt with, we refer the interested reader to [29, 38] and references therein for further details.

In what follows we shall consider in a unified way both the case $\mu = 0$, which corresponds to a nondifferentiable problem, and the case $\mu \neq 0$, which gives rise to differentiable problems. Therefore we shall write, for example, $\partial \phi_{\mu}(z_i, \lambda_i)$. This expression will just be $-2\partial \min(z_i, \lambda_i)$ if $\mu = 0$, while it will give $\nabla \phi_{\mu}(z_i, \lambda_i)$ if $\mu \neq 0$. Unless stated otherwise, from now on all the (generalized) gradients are made with respect to the variables (x, y, z, λ). So, for example, $\nabla \phi$ is a vector in $\mathbb{R}^{n+m+l+l}$. Finally, we shall indicate by $d(\cdot | X_{\mathrm{ad}} \times \mathbb{R}^{m+l+l})$ the Euclidean distance function to the set $X_{\mathrm{ad}} \times \mathbb{R}^{m+l+l}$.

We are in the position to apply freely the well established theory of necessary optimality conditions for Lipschitz programs [11]. Accordingly, a necessary condition for a point $w = (x, y, z, \lambda)$ to be optimal for Problem (P_{μ}) is that the point be stationary,

i.e. that $0 \neq (\eta, \theta, \rho, \sigma) \in \mathbb{R}^{1+m+l+l}$, with $\eta \geq 0$, and $s \in \partial d(w|X_{\mathrm{ad}} \times \mathbb{R}^{m+l+l})$ exist such that

$$0 \in \eta \nabla f(x, y) + \nabla L(x, y, \lambda)\theta + \nabla(g(x, y) - z)\rho + \sum_{i=1}^{l} \partial \phi_\mu(z_i, \lambda_i)\sigma_i + M\|(\eta, \theta, \rho, \sigma)\|s, \tag{11}$$

where M is a Lipschitz constant for (f, H_μ) around w and where we set $L(x, y, \lambda) = F(x, y) - \nabla_y g(x, y)\lambda$. This conditions are known as Fritz-John (FJ) conditions. If $\eta \neq 0$, and hence, without loss of generality, if $\eta = 1$, (11) becomes

$$0 \in \nabla f(x, y) + \nabla L(x, y, \lambda)\theta + \nabla(g(x, y) - z)\rho + \sum_{i=1}^{l} \partial \phi_\mu(z_i, \lambda_i)\sigma_i + M\|(1, \theta, \rho, \sigma)\|s, \tag{12}$$

which are the Karush-Kuhn-Tucker (KKT) conditions for Problem (P_μ).

It is well known that the presence of FJ points which are not KKT points for Problem (P_μ) can cause severe problems to numerical optimization procedures. Luckily, in our case we have that every stationary point is a KKT point for Problem (P_μ).

Theorem 2. *For every μ, if w is a stationary point for Problem (P_μ), then η in (11) is always different from 0 and w is a KKT point of Problem (P_μ).*

Proof. Suppose by contradiction that w satisfies the FJ conditions (11) with $\eta = 0$. Then we can write

$$0 \in \nabla L(x, y, \lambda)\theta + \nabla(g(x, y) - z)\rho + \sum_{i=1}^{l} \partial \phi_\mu(z_i, \lambda_i)\sigma_i + M\|(\theta, \rho, \sigma)\|s, \tag{13}$$

where $(\theta, \rho, \sigma) \neq 0$ and $s \in \partial d(w|X_{\mathrm{ad}} \times \mathbb{R}^{m+l+l})$. We shall now concentrate on the last $m + l + l$ equations in the FJ conditions (11). We first observe that, if we partition the vector s in $(s_x, s_y, s_z, s_\lambda)$, we have that $(s_y, s_z, s_\lambda) = 0$, since this latter subvector, for well known properties of the distance function, belongs to the normal cone to $\mathbb{R}^{m+l+l}$, which is 0_{m+l+l}. Taking this fact into account, the last $m + l + l$ equations of (13) give

$$0 \in \pi_{y,z,\lambda} \nabla L(x, y, \lambda)\theta + \pi_{y,z,\lambda} \nabla(g(x, y) - z)\rho + \pi_{y,z,\lambda} \sum_{i=1}^{l} \partial \phi_\mu(z_i, \lambda_i)\sigma_i$$
$$\subseteq \pi_{y,z,\lambda} \partial H_\mu(w)^T \begin{pmatrix} \theta \\ \rho \\ \sigma \end{pmatrix}, \tag{14}$$

where $\pi_{y,z,\lambda}$ is the projector over the (y, z, λ) space. Similarly to Lemma 1, we have that H_μ, as a function of (x, y, z, λ), is regular, so $\pi_{y,z,\lambda} \partial H_\mu(w)$ is contained in the generalized Jacobian of H_μ with respect to (y, z, λ) by [11, Proposition 2.3.15]. But then, recalling that $(\theta, \rho, \sigma) \neq 0$, (14) contradicts Proposition 3.

$\square$

If $\mu = 0$, the conditions (12) are the Clarke KKT conditions for the nonsmooth problem (P$_0$), which is, from the point of view of (local) minimizers, equivalent to the original MPEC problem. Therefore, the conditions (12) are necessarily satisfied at a (local) solution (x^*, y^*) of the MPEC problem. However, when refering to the MPEC problem we prefer not to call these conditions KKT conditions; in fact this terminology could easily lead one to overlook the intrinsic, difficult combinatorial nature of the MPEC problem. In [38] several sets of necessary conditions for the MPEC problem are derived and the relations among these conditions and the conditions proposed in [29] are elucidated. Following the terminology introduced in [38], we shall call term a point (x^*, y^*) satisfying, together with the uniquely associated (z^*, λ^*), the condition (12) a strong C-stationary (SCS) point of the MPEC problem. Actually, in [38] no abstract constraint of the type $x \in X_{ad}$ is considered; however, if one assumes that the set X_{ad} is defined by a finite number of smooth equalities and inequalities, it is possible to check that the KKT conditions for Problem P_0 coincide with the strong C-stationary conditions considered in [38], thus justifying the terminology adopted. We refer the reader to [38] for a detailed discussion of the SCS conditions and of their relations to other optimality conditions for MPEC problems. Here we only mention that the SCS conditions appear to be a reasonable set of necessary conditions for optimality, even if they certainly do not fully capture the disjunctive nature of the MPEC constraints, contrary to some of the more refined "combinatorial-type" conditions proposed in [29, 38, 33]. However, as far as we are aware of, there exists no algorithm, for the class of problems we consider, for which global convergence to stationary points satisfying these "combinatorial-type" conditions can be proved, and convergence to SCS points, which will be shown in the next section, appears to be, to date, the strongest kind of global convergence result available in practice.

In the next section we shall also need the following result which, roughly speaking, states a continuity property for stationary points of the problems (P$_\mu$).

Proposition 4. *Let w^k and $\mu^k \to 0$ be two sequences such that, for every k, w^k is a KKT point for Problem (P$_{\mu^k}$) with multipliers $(\theta^k, \rho^k, \sigma^k)$ and for some vector $s^k \in \partial d(w^k | X_{ad} \times \mathbb{R}^{m+l+l})$, and such that $\mu^k \neq 0$ for all k. Suppose that $\{w^k\} \to w^*$, $\{(\theta^k, \rho^k, \sigma^k)\} \to (\theta^*, \rho^*, \sigma^*)$, and $\{s^k\} \to s^*$. Then $w^* = (x^*, y^*, z^*, \lambda^*)$, with the multipliers $(\theta^*, \rho^*, \sigma^*)$, and with the vector s^*, satisfies the KKT conditions for Problem (P$_0$), i.e., (x^*, y^*) is an SCS point of the MPEC problem.*

Proof. We only need to check that

$$0 \in \nabla f(x^*, y^*) + \nabla L(x^*, y^*, \lambda^*)\theta^* + \nabla(g(x^*, y^*) - z^*)\rho^*$$
$$- \sum_{i=1}^{l} 2\partial \min(z_i^*, \lambda_i^*)\sigma^* + M\|(1, \theta^*, \rho^*, \sigma^*)\|s^*, \tag{15}$$

with $s^* \in \partial d(w^* | X_{ad} \times \mathbb{R}^{m+l+l})$. By assumptions we have that

$$0 = \nabla f(x^k, y^k) + \nabla L(x^k, y^k, \lambda^k)\theta^k + \nabla(g(x^k, y^k) - z^k)\rho^k$$
$$+ \sum_{i=1}^{l} \nabla \phi_{\mu^k}(z_i^k, \lambda_i^k)\sigma^k + M\|(1, \theta^k, \rho^k, \sigma^k)\|s^k. \tag{16}$$

Taking into account that, by the upper semicontinuity of the subdifferential mapping, $s^* \in \partial d(w^*|X_{\mathrm{ad}} \times \mathbb{R}^{m+l+l})$, and taking also into account the continuity assumptions, it is easy to see that if (possibly subsequencing)

$$\lim_{k \to \infty} \nabla \phi_{\mu^k}(z_i^k, \lambda_i^k) \in -2\partial \min(z_i^*, \lambda_i^*), \qquad i = 1, \ldots, l, \tag{17}$$

then (15) follows from (16) by simple continuity arguments. We then check (17). If we indicate by e_j the j-th column of the $(n + m + l + l)$ dimensional identity matrix, we recall that

$$
\begin{aligned}
-2\partial \min(z_i^*, \lambda_i^*) &= -2e_{n+m+i}, & &\text{if } z_i^* < \lambda_i^* \\
-2\partial \min(z_i^*, \lambda_i^*) &= -2e_{n+m+l+i}, & &\text{if } z_i^* > \lambda_i^* \\
-2\partial \min(z_i^*, \lambda_i^*) &= -2\mathrm{conv}(e_{n+m+i}, e_{n+m+l+i}), & &\text{if } z_i^* = \lambda_i^*.
\end{aligned}
\tag{18}
$$

On the other hand we also have

$$
\begin{aligned}
\nabla \phi_{\mu^k}(z_i^k, \lambda_i^k) &= \left(\frac{z_i^k - \lambda_i^k}{\sqrt{(z_i^k - \lambda_i^k)^2 + 4(\mu^k)^2}} - 1 \right) e_{n+m+i} \\
&+ \left(\frac{\lambda_i^k - z_i^k}{\sqrt{(z_i^k - \lambda_i^k)^2 + 4(\mu^k)^2}} - 1 \right) e_{n+m+l+i}.
\end{aligned}
\tag{19}
$$

Now, passing to the limit in (19) and comparing with (18), we obtain (17) and the proof is complete.

□

The properties proved in this section can be contrasted to the regularity properties of the reformulation (5). To this end we consider the following simple example, with $n = 1, m = 1, l = 2$.

$$
\begin{aligned}
f(x, y) &= 2x - y \\
X_{\mathrm{ad}} &= \{x : 0 \le x \le 1\} \\
F(x, y) &= y - x \\
C(x) &= \{y : 0 \le y \le 1\}.
\end{aligned}
$$

For this problem, reformulation (5) becomes

$$
\begin{aligned}
\min \; & 2x - y \\
\text{s.t.} \; & x \ge 0, \quad 1 - x \ge 0 \\
& y \ge 0, \quad 1 - y \ge 0 \\
& \lambda_1 \ge 0, \; \lambda_2 \ge 0 \\
& y\lambda_1 = 0, \; (1 - y)\lambda_2 = 0 \\
& y - x - \lambda_1 + \lambda_2 = 0.
\end{aligned}
$$

It can be easily verified that the feasible region of this problem is given by points of the form $(\alpha, \alpha, 0, 0)$, for all $\alpha \in [0, 1]$, and therefore that the unique minimum point is given by $w^* = (0, 0, 0, 0)$. On the other hand, if we write down the necessary optimality conditions, we see that every feasible point is a FJ point and that w^* is not a KKT point.

5. Algorithms

In this section we study the behaviour of a sequence $\{w^k\}$, where w^k is a "solution" of Problem (P_{μ^k}) and $\{\mu^k\} \to 0$. We shall consider both the case in which "solution" means global solution, which is the most interesting case from the theoretical point of view, and the case in which "solution" is just a (possibly approximate) stationary point of Problem (P_{μ^k}), which is the most important case from the practical point of view.

Algorithm G.

Step 0 Let $\{\mu^k\}$ be any sequence of nonzero numbers with $\lim_{k\to\infty} \mu^k = 0$. Choose $w^0 := (x^0, y^0, z^0, \lambda^0) \in \mathbb{R}^{n+m+l+l}$, and set $k := 1$.

Step 1 Find a global solution w^k of Problem (P_{μ^k}).

Step 2 Set $k := k + 1$, and go to Step 1.

Here "G" stands for "Global". By Corollary 2 the algorithm above is well defined, i.e., the global solution to Problem (P_{μ^k}) required at Step 1 exists. The following theorem shows that the sequence $\{w^k\}$ generated by the algorithm converges to a global solution of problem (P).

Theorem 3. *The sequence $\{w^k\}$ generated by Algorithm G is contained in a compact set, and each of its limit points is a global solution of Problem (P).*

Proof. The fact that the sequence $\{w^k\}$ is contained in a compact set immediately follows from Theorem 1 with $\tilde{\mu} = \sup_k |\mu^k|$. Then we pass to the second part of the theorem. Assume, without loss of generality, that $\{w^k\} \to w^* = (x^*, y^*, z^*, \lambda^*)$ and suppose, by contradiction, that w^* is not a global solution of Problem (P). We first note that, since each w^k is feasible to Problem (P_{μ^k}) and $\mu^k \to 0$ then, by continuity, we have that $w^* \in \mathcal{F}_0$. If w^* is not a global solution of Problem (P), then there exists a feasible point $\bar{w} = (\bar{x}, \bar{y}, \bar{z}, \bar{\lambda})$ such that

$$f(\bar{x}, \bar{y}) < f(x^*, y^*). \tag{20}$$

We shall now obtain a contradiction by showing that w^k is not a global solution of Problem (P_{μ^k}) for k large enough. By Lemma 3 we know that for every positive μ there exists a unique point $\bar{w}_\mu = (\bar{x}, y_\mu(\bar{x}), z_\mu(\bar{x}), \lambda_\mu(\bar{x}))$ in $\mathcal{F}_\mu$ such that its x-part is equal to $\bar{x}$. We also know that $\bar{w}_\mu$ tends to $\bar{w}$ when μ tends to 0. But then, since w^k tends to w^*, by continuity and taking into account (20), we obtain that eventually $f(\bar{x}, y_{\mu^k}(\bar{x})) < f(x^*, y^k)$, which contradicts the fact that w^k is a global solution of Problem (P_{μ^k}). $\square$

Theorem 3 is obviously of great theoretical interest, however, if we use standard optimization methods to solve the sequences of Problems (P_{μ^k}) we are only able to find stationary points of Problem (P_{μ^k}). Then, it is interesting to analyze the following more practical algorithm scheme, where we suppose only that at each step we are able to find a stationary point of (P_{μ^k}).

Algorithm S.

Step 0 Let $\{\mu^k\}$ be any sequence of nonzero numbers with $\lim_{k\to\infty} \mu^k = 0$. Choose $w^0 := (x^0, y^0, z^0, \lambda^0) \in \mathbb{R}^{n+m+l+l}$, and set $k := 1$.

Step 1 Find a stationary point w^k of Problem (P_{μ^k}).

Step 2 Set $k := k + 1$, and go to Step 1.

Here "S" stands for "Stationary". By Theorem 2 we have that every stationary point of the Problems (P_{μ^k}) is actually a KKT point. Since, by Corollary 2 every Problem (P_{μ^k}) admits at least one stationary point, the task at Step 1 is well posed. We stress again that, by Theorem 2, *every* stationary point of the Problem (P_{μ^k}) is a KKT point.

We can prove the following theorem.

Theorem 4. *The sequence $\{w^k\}$ generated by Algorithm S is contained in a compact set, and if $w^* = (x^*, y^*, z^*, \lambda^*)$ is a limit point of $\{w^k\}$, then (x^*, y^*) is an SCS point of the MPEC problem.*

Proof. The fact that the sequence $\{w^k\}$ generated by Algorithm S is contained in a compact set can be proven as in Theorem 3, so we turn to the second part of the theorem. Since w^k is a KKT point of Problem (P_{μ^k}), there exist, for each k, multipliers $(\theta^k, \rho^k, \sigma^k)$ and a vector s^k in $\partial d(w^k | X_{\mathrm{ad}} \times \mathbb{R}^{m+l+l})$ such that

$$0 = \nabla f(x^k, y^k) + \nabla L(x^k, y^k, \lambda^k)\theta^k + \nabla(g(x^k, y^k) - z^k)\rho^k$$
$$+ \sum_{i=1}^{l} \nabla\phi_{\mu^k}(z_i^k, \lambda_i^k)\sigma_i^k + M\|(1, \theta^k, \rho^k, \sigma^k)\|s^k, \tag{21}$$

where M is the sup of the Lipschitz constants of the functions (f, H_μ) on the compact set $D(\tilde{\mu})$ (see Theorem 1), with $\tilde{\mu} = \sup_k |\mu^k|$, and for all $\mu \in [-\tilde{\mu}, \tilde{\mu}]$.

We now consider two cases: $\|(\theta^k, \rho^k, \sigma^k)\|$ is bounded and $\|(\theta^k, \rho^k, \sigma^k)\|$ is unbounded. We first examine the second case and show it cannot occur.

We can divide (21) by $\|(1, \theta^k, \rho^k, \sigma^k)\|$. If we indicate by $(\bar{\theta}^k, \bar{\rho}^k, \bar{\sigma}^k)$ the normalized vector of multipliers, we can assume without loss of generality, that $(\bar{\theta}^k, \bar{\rho}^k, \bar{\sigma}^k)$ converges to $(\bar{\theta}^*, \bar{\rho}^*, \bar{\sigma}^*) \neq 0$ and that w^k converges to w^* feasible to Problem (P). Since s^k is in the subdifferential of a distance function, its norm is bounded by 1, so that, taking also into account the upper-semicontinuity of subdifferentials, we can assume that also s^k converges to s^* belonging to $\partial d(w^* | X_{\mathrm{ad}} \times \mathbb{R}^{m+l+l})$. Furthermore, by continuity, $\nabla f(x^k, y^k) \to \nabla f(x^*, y^*)$, $\nabla L(x^k, y^k, \lambda^k) \to \nabla L(x^*, y^*, \lambda^*)$ and $\nabla(g(x^k, y^k) - z^k) \to \nabla(g(x^*, y^*) - z^*)$. Finally, reasoning as in the proof of Proposition 4 it is easy to see that we can also assume, without loss of generality, that $\nabla\phi_{\mu^k}(z_i^k, \lambda_i^k)$ converges to an element in $-2\partial \min(\lambda_i^*, z_i^*)$. Hence, passing to the limit in the normalized (21) we obtain

$$0 \in \nabla L(x^*, y^*)\bar{\theta}^* + \nabla(g(x^*, y^*) - z^*)\bar{\rho}^* - 2\partial \min(\lambda_i^*, z_i^*)\bar{\sigma}_i^* + M\|(\bar{\theta}^*, \bar{\rho}^*, \bar{\sigma}^*)\|s^*,$$

so that w^* is a FJ point with 0 multiplier for the objective function. But this contradicts Theorem 2 so this case cannot occur and $\|(\theta^k, \rho^k, \sigma^k)\|$ is bounded.

So, let us examine the case $\|(\theta^k, \rho^k, \sigma^k)\|$ is bounded. Then, without loss of generality, we can assume that $(1, \theta^k, \rho^k, \sigma^k)$ converges to $(1, \theta^*, \rho^*, \sigma^*)$ and that w^k converges to w^* feasible to Problem (P). Analogously to the previous case, we can also assume, without loss of generality, that s^k converges to s^* in $\partial d(w^*|X_{\mathrm{ad}} \times \mathbb{R}^{m+l+l})$. The desired result then follows from Proposition 4.

$\square$

In the proof of the theorem we have also shown that the multipliers generated when solving the Problems (P_{μ^k}) are bounded. Since this is a very significant property from the numerical point of view, showing that the Problems (P_{μ^k}) are numerically stable as μ^k goes to 0, we state this result as a corollary.

Corollary 3. *Let $(\theta^k, \rho^k, \sigma^k)$ be the sequence of KKT multipliers associated to the solution w^k calculated at Step 2 of Algorithm S. Then there exists a constant N such that $\|(\theta^k, \rho^k, \sigma^k)\| \le N$ for every k.*

Although Algorithm S is much more realistic than Algorithm G, since computing a KKT point is an affordable task, we have not considered yet the stopping criterion problem. Furthermore we would also like to be able to consider the case in which we only find an approximate stationary point of Problem (P_{μ^k}). We address these issues in the following algorithm.

Algorithm AS.

Step 0 Let $\{\mu^k\}$ be any sequence of nonzero numbers with $\lim_{k \to \infty} \mu^k = 0$. Choose a stopping tolerance $T \ge 0$ and a sequence $\{\epsilon^k\}$ of nonnegative numbers. Choose $w^0 := (x^0, y^0, z^0, \lambda^0) \in \mathbb{R}^{n+m+l+l}$, and set $k := 1$.

Step 1 Find a point w^k, multipliers $(\theta^k, \rho^k, \sigma^k)$ and a vector $s^k \in \mathbb{R}^{n+m+l+l}$ such that
   (a) $\big\| \nabla f^k + \nabla L^k \theta^k + \nabla(g^k - z^k)\rho^k + \sum_{i=1}^{l} \nabla \phi_{\mu^k}(z^k, \lambda^k) \sigma_i^k$
        $+ M\|(1, \theta^k, \rho^k, \sigma^k)\| s^k \big\| \le \epsilon^k$
   (b) $d(w^k|\mathcal{F}_{\mu^k}) \le \epsilon^k$
   (c) $d\big(s^k|\partial d(w^k|X_{\mathrm{ad}} \times \mathbb{R}^{m+l+l})\big) \le \epsilon^k$.

Step 2 If $\max\big((\mu^k)^2, \epsilon^k, d(w^k|\mathcal{F}_0)\big) \le T$, STOP.

Step 3 Set $k := k + 1$, and go to Step 1.

Here "AS" stands for "Approximate Stationary". A few comments are in order. In Step 1 we look for an approximate KKT point of Problem (P_{μ^k}). The definition of approximate KKT point of Problem (P_{μ^k}) is given by the fairly natural conditions (a), (b) and (c). Note that (a) and (c) are just an approximation of the KKT conditions, while (b) requires the point to be "not too far from feasibility". Much of the formal complication is due to the fact that we have used the abstract constraint notation $x \in X_{\mathrm{ad}}$, with the consequent necessity to use the subdifferential of the distance function in the KKT conditions. However, in most of the practical applications, the set X_{ad} is defined by a finite number of equalities and inequalities, so that the optimality conditions can be expressed in terms of the gradients of these constraints and of their associated multipliers. In this event, point (c) becomes superfluous and can be incorporated directly

into point (a). The test at point (a) is standard, and is usually a part of any practical stopping test in commercial codes for the solution of smooth constrained optimization problems. The test at point (b), instead, is conceptually hard to check. Assuming that the set X_{ad} is defined by a system of inequalities, i.e. $X_{\text{ad}} := \{x \in \mathbb{R}^n | h(x) \geq 0\}$ with $h : \mathbb{R}^n \to \mathbb{R}^s$, test (b) is usually substituted, in practical algorithms, by the test $\|H_{\mu^k}(w^k)\| + \|\min(h(x), 0)\| \leq \epsilon^k$. We finally remark that, for simplicity, we have assumed that in the tests (a), (b) and (c) the same tolerance ϵ^k is used, but we can obviously use three different parameters.

We finally come to the stopping criterion for the main algorithm in Step 2. We recall that, in general, finding theoretically sound stopping criteria for nondifferentiable problems is not an easy task and usually involves the calculation of (an approximation of) the ϵ-subdifferential of the functions involved. In our scheme we can easily bypass these difficulties using the fact that the sequence of smooth problems we (approximately) solve, progressively approximate the nonsmooth problem (P$_0$). The test at Step 2 requires that the quantities μ^k, ϵ^k and $d(w^k|\mathcal{F}_0)$ be smaller than T, which should be thought of as a small positive quantity. Requiring that $d(w^k|\mathcal{F}_0)$ be smaller than T obviously amounts to requiring "almost" feasibility. Instead, the reason for requiring that both $(\mu^k)^2$ and ϵ^k be smaller than T is connected to the satisfaction of the KKT conditions (15) for Problem (P$_0$). In particular, thanks to (a), (c) and (17), we see that the smaller the T, the closer to satisfaction (15) is. We call the conditions expressed by the test at Step 2 the *T-approximate KKT conditions* for Problem (P). If $w^* = (x^*, y^*, z^*, \lambda^*)$ is a T-approximate KKT point, then we also say that (x^*, y^*) is a T-approximate SCS point for the MPEC problem. In particular, we see that if $T = 0$ and we set $\epsilon^k = 0$ for every k, Algorithm AS reduces to Algorithm S. More in general the following theorem can be proved.

Theorem 5. *Suppose that the sequence of numbers ϵ^k tends to 0. Then the sequence $\{w^k\}$ generated by Algorithm AS is contained in a compact set. Furthermore,*

(a) if $T > 0$ then the algorithm terminates after a finite number of steps with a point $w^k = (x^k, y^k, z^k, \lambda^k)$ such that (x^k, y^k) is a T-approximate SCS point for the MPEC problem;

(b) if $T = 0$ every limit point $w^* = (x^*, y^*, z^*, \lambda^*)$ of the sequence $\{w^k\}$ is such that (x^*, y^*) is an SCS point of the MPEC problem.

Proof. Let $\tilde{\mu} = \sup_k |\mu^k|$ and $\tilde{\epsilon} = \sup_k \epsilon^k$. The sequence $\{w^k\}$ belongs to the set $D(\tilde{\mu}) + \{w : \|w\| \leq \tilde{\epsilon}\}$, which is obviously compact (see Theorem 1). Then we pass to examine point (a). By the assumptions made, we know that both the sequences $\{\mu^k\}$ and $\{\epsilon^k\}$ tend to 0. So to prove this point it is sufficient to show that $d(w^k|\mathcal{F}_0)$ also goes to 0. To this end, let us indicate by $\bar{w}^k$ (one of) the closest point to w^k in the closed set $\mathcal{F}_{\mu^k}$. Since the sequence $\{\bar{w}^k\}$ belongs to the compact set $D(\tilde{\mu})$, we can assume, without loss of generality, that it converges to a point $\bar{w}^*$. By continuity of the constraints of the Problems (P$_\mu$), we also have $\bar{w}^* \in \mathcal{F}_0$. We can now write

$$\|w^k - \bar{w}^*\| \leq \|w^k - \bar{w}^k\| + \|\bar{w}^k - \bar{w}^*\|,$$

which, taking into account that $\|w^k - \bar{w}^k\|$ is smaller than ϵ^k and the fact that $\bar{w}^k$ is converging to $\bar{w}^*$, shows that $\|w^k - \bar{w}^*\|$ converges to 0, and hence that also the distance $d(w^k|\mathcal{F}_0)$ goes to 0.

Let us examine point (b). If $T = 0$ the algorithm will generate an infinite sequence of points $\{w^k\}$. The assertion of the theorem can be proved in a way very similar to that used in the proof Theorem 4, taking into account that $\{\epsilon^k\}$ goes to 0 and that, as shown in the first part of the proof, $\{w^k\}$ can be assumed to converge, without loss of generality, to a point w^* in the feasible set of Problem (P).

$\square$

6. Numerical results

In this section we consider some preliminary numerical results obtained with the approach described in the previous sections. We performed two sets of experiments: on problems taken from the literature and on randomly generated test problems.

For the first set of problems, we coded algorithm AS in Fortran 77 and ran it on a DEC 5000 workstation. After a few preliminary tests, the following choices were made and used on all the test problems.

– The perturbation parameter μ is initially set to 0.0001 and reduced by a factor of 100 at each outer iteration.
– All the nonlinear optimization problems which are required to be solved by the algorithm are solved by the routine E04UCF in the NAG library [32]. This routine implements a sequential quadratic programming algorithm; the default parameters were always used. In particular this implies that the nonlinear optimization subproblems are solved exactly, i.e. that $\epsilon^k = 0$ (up to machine precision).
– The starting point for subproblem $(P_{\mu^{k+1}})$ is w^k.
– The stopping parameter T is set to 10^{-6}.

We remark that we only want to show viability of the approach proposed. An optimal choice of the parameters is out of the scope of this paper; however, it may be interesting to add that, at least in our limited experience, the behaviour of the algorithm does not seem to depend critically on the choices outlined above.

We ran the algorithm on 11 test problems used in [1,4,33,34]. A full analytical description of the test problems along with the starting points, the optimal solutions found and bibliographical references to the sources of the problems are given in the Appendix. In Table 1 we report, for each problem, the values of n, m and l (the "dimensions" of the problem), the number K of outer iterations needed to reach the prescribed accuracy and the optimal objective function value f^* found. In some cases more than one starting point is used. This is indicated in the column Start, the corresponding starting points are given in the Appendix.

From the results reported in Table 1 we see that the algorithm is capable to solve with a limited amount of work all the problems considered. A comparison with the data reported in [1,4,33,34] shows that our algorithm is capable of reaching at least the same accuracy in a number of outer iterations which is usually smaller (and never larger) than the number of nonsmooth functions evaluations reported in [1,4,33,34]. It should be remarked that in Problem 7, in some instances of Problem 9, in Problem 10 and in Problem 11, convergence occurred to a different point from the one reported

Table 1. Results for the first set of test problems

Problem	n	m	l	Start	K	f^*
1	1	6	4	(a)	2	3.207701
	1	6	4	(b)	2	3.207701
2	1	6	4	(a)	1	3.449404
	1	6	4	(b)	1	3.449404
3	1	6	4	(a)	1	4.604254
	1	6	4	(b)	1	4.604254
4	1	6	4	(a)	1	6.592684
	1	6	4	(b)	1	6.592684
5	2	2	2	(a)	2	-1
6	1	1	1	(a)	1	-3266.667
7	2	2	6	(a)	3	4.999375
8 ($L = 150, \gamma = 1.0$)	1	4	8	(a)	1	-343.3453
8 ($L = 150, \gamma = 1.1$)	1	4	8	(a)	1	-203.1551
8 ($L = 150, \gamma = 1.3$)	1	4	8	(a)	1	-68.13565
8 ($L = 150, \gamma = 1.5$)	1	4	8	(a)	1	-19.15407
8 ($L = 150, \gamma = 1.7$)	1	4	8	(a)	1	-3.161181
8 ($L = 50, \gamma = 1.0$)	1	4	8	(a)	1	-346.8932
8 ($L = 40, \gamma = 1.1$)	1	4	8	(a)	1	-224.0372
8 ($L = 30, \gamma = 1.3$)	1	4	8	(a)	1	-80.78597
8 ($L = 25, \gamma = 1.5$)	1	4	8	(a)	1	-22.83712
8 ($L = 20, \gamma = 1.7$)	1	4	8	(a)	1	-5.349136
9	2	2	2	(a)	1	0.1345863E-17
9	2	2	2	(b)	1	0.1469359E-22
9	2	2	2	(c)	1	0.1009552E-17
9	2	2	2	(d)	1	0.3635465E-26
9	2	2	2	(e)	1	0.4715810E-25
10	4	4	12	(a)	1	-6600.000
11	2	6	4	(a)	1	-12.67871

in the sources. For Problem 10, our algorithm converges to the optimal solution given in [1], while the algorithm in [1] only gave approximate solutions. We note that using only the data reported in [33] and [34], it is not possible to compare the efficiency of our scheme with those of [33] and [34], which are based on bundle methods. To make future comparisons possible, we report in Table 2 some more data which give a more detailed picture of the computational effort required by our algorithm. More specifically, for each problem we report the total number of objective function evaluations (f), the total number of objective function gradient evaluations (∇f), the cumulative number of evaluations of the $(H_\mu)_i$ (H) and of their gradients (∇H). However, the evaluation numbers of $H_i = g_i(x, y) - z_i$ and their gradients are not counted if g_i is linear (hence H_i is also linear). We remark that each evaluation of a component of (H_μ) is counted separately and the same happens for their gradients. Note also that in all the problems considered the set X_{ad} is described by very simple constraints, we have not included the number of evaluations of these constraints, which gives a negligible contribution to the computational burden. Finally, under the heading KIN we have reported the

Table 2. More data on the first set of test problems

Problem	Start	f	∇f	H	∇H	KIN
1	(a)	83	81	850	810	27
	(b)	56	54	580	540	26
2	(a)	57	56	580	560	20
	(b)	30	29	310	290	19
3	(a)	28	27	290	270	17
	(b)	30	29	310	290	15
4	(a)	37	36	380	360	14
	(b)	18	17	190	170	12
5	(a)	35	33	222	198	19
6	(a)	8	7	18	14	5
7	(a)	81	78	672	624	28
8 ($L = 150, \gamma = 1.0$)	(a)	9	8	120	96	7
8 ($L = 150, \gamma = 1.1$)	(a)	9	8	120	96	7
8 ($L = 150, \gamma = 1.3$)	(a)	15	14	192	168	10
8 ($L = 150, \gamma = 1.5$)	(a)	10	9	132	108	7
8 ($L = 150, \gamma = 1.7$)	(a)	11	10	144	120	8
8 ($L = 50, \gamma = 1.0$)	(a)	7	6	96	72	5
8 ($L = 40, \gamma = 1.1$)	(a)	9	8	120	96	6
8 ($L = 30, \gamma = 1.3$)	(a)	8	7	108	84	6
8 ($L = 25, \gamma = 1.5$)	(a)	21	20	264	240	18
8 ($L = 20, \gamma = 1.7$)	(a)	23	22	288	264	20
9	(a)	9	8	40	32	7
9	(b)	12	11	52	44	9
9	(c)	10	9	44	36	7
9	(d)	9	8	40	32	7
9	(e)	9	8	40	32	7
10	(a)	89	88	1440	1408	51
11	(a)	11	10	120	100	7

total number of inner iterations required by the recursive quadratic solver. Each inner iteration corresponds, roughly speaking, to the solution of a convex quadratic program. We remark that the data in Table 1 and 2 give a fairly complete picture of the work needed by our algorithm; in our scheme there are no other significant time consuming tasks besides those described so far.

With regard to the number of inner iterations needed it may be of interest to note the following facts. At each outer iteration the Problems (P_{μ^k}) become closer and closer to the nondifferentiable problem (P), since μ^k goes to 0. It could then be expected that the resolution of the Problems (P_{μ^k}) becomes progressively more difficult. However, since the sequence of points w^k is converging to w^*, we have that the starting point employed in the solution of each subproblem is closer and closer to the solution of the subproblems themselves, thus compensating the increasing difficulty of the problem. This is confirmed by the fact that the number of inner iterations required by the sequential programming algorithm to solve the nonlinear programs decreases at each outer iteration. The numbers of inner iterations are: 22, 5 for Problem 1 (a); 21, 5 for Problem 1 (b); 14, 5 for Problem 5; and 14, 8, 6 for Problem 7.

To further illustrate the behaviour of algorithm AS we also ran it on some randomly generated test problems. These problems have the general form

$$\min \tfrac{1}{2}(x, y)^T P(x, y) + c^T x + d^T y$$
$$\text{s.t.} \quad Hx + a \le 0 \tag{22}$$
$$Nx + My + q \ge 0, \; y \ge 0, \; y^T (Nx + My + q) = 0,$$

where $P \in \mathbb{R}^{(n+m)\times(n+m)}$ is positive definite, $M \in \mathbb{R}^{m\times m}$ is strictly monotone, $c \in \mathbb{R}^n$, $d \in \mathbb{R}^m$, $H \in \mathbb{R}^{l\times n}$, $a \in \mathbb{R}^l$, $N \in \mathbb{R}^{m\times n}$, and $q \in \mathbb{R}^m$. The problems were generated using the MATLAB random generator QPECGEN. We refer the reader to [24] for a detailed description of this generator; here we only give a very brief description of some of its features. The generator QPECGEN allows the user to control various characteristics of problem (22): the number of upper and lower level variables, the number of constraints defining the set X_{ad}, the condition number of P and M, the degree of degeneracy. The generator also provides a (local) solution to the problems it produces. We generated four different problems, whose dimensions are reported in Table 3. There we also report the value of DEG which gives the number of degenerate indices of the lower level complementarity problem at the solution provided by the generator; note that for problems 1, 2 and 4 our algorithm converged to this solution. Finally, the generator parameters were chosen so that the condition numbers of P and M are around 10^5 and 2000 respectively. We remark that the degree of degeneracy of these problems goes from medium to high and that degeneracy at the lower level solutions is one of the features that make MPEC problems much harder than traditional nonlinear programming problems. We used two different starting points for each problem: (a) $(x, y) = (1, ..., 1)$ and (b) $(x, y) = (100, ..., 100)$, which are not necessarily feasible to the lower level problem. Since the generator QPECGEN is implemented in MATLAB we reimplemented algorithm AS in MATLAB and ran it on a DEC ALPHA. The setting of the parameters and the stopping rule are the same as for the first set of test problems. The nonlinear optimization problems in the smoothing method are solved by the MATLAB TOOLBOX **constr.m** routine which is a sequential quadratic routine.

The numerical results are shown in Table 3, where the notation follows that used in Tables 1 and 2. Note that it was not possible to obtain all the data reported in Table 2 for the first set of test problems, since this was not easily allowed by the routine **constr.m**. Since in all cases only two outer iterations were needed we reported in the column KIN the number of inner iterations needed in each outer iteration. A few observations are

Table 3. Results for the random test problems

Problem	n	m	l	DEG	Start	K	KIN	f	f*
1	4	20	5	5	(a)	2	25+4	51	-3685.65
	4	20	5	5	(b)	2	21+5	35	-3685.65
2	4	20	5	15	(a)	2	20+3	57	-1477.99
	4	20	5	15	(b)	2	15+3	18	-1477.99
3	8	50	10	10	(a)	2	110+65	243	-13209.1
	8	50	10	10	(b)	2	106+138	433	-13191.3
4	8	50	10	30	(a)	2	70+6	145	-9469.95
	8	50	10	30	(b)	2	91+4	128	-9469.95

in order. The algorithm was again able to find a stationary point in all cases, even if in Problem 3 two different points were produced according to the starting point used. We recall that, even if the upper level objective function is strictly convex, problem (22) is not convex and it can have multiple stationary points with different objective function values. These random generated problems seem more difficult than the ones used in the first set of tests, because of the lower level degeneracy, the high condition number of the upper level objective function and their larger dimensions. This fact is confirmed by the relatively higher numbers of f and KIN although it should be noted that these data are not directly comparable to the corresponding ones in Table 2 since the sequential quadratic routine is different in the two cases. A closer look at the results of the routine **constr.m** shows that usually the routine is able to find quickly a nearly optimal solution of the smoothed problems, but then the final convergence rate is slow and many additional iterations are needed to reach the required accuracy. This is probably due to the high degeneracy of the smoothed problems and to the ill conditioning of the merit function, features which are inherited from the unperturbed MPEC problem. Also for these problems very few outer iterations are needed to achieve the required accuracy confirming, in our opinion, that the smoothed problems are "good" approximations to the original problem.

7. Conclusions

We introduced a new method for the solution of MPEC problems with strongly monotone variational inequalities as side constraints. The method is based on approximating a nonsmooth one-level minimization reformulation of the MPEC problem by a series of smooth, regular problems. We showed that every sequence of global (stationary) solutions so generated converges to global (stationary) solutions of the MPEC problem. The numerical results reported seem to indicate that the method introduced in this paper can be numerically efficient.

The new approach seems quite different from existing proposals for the same class of problems, and also if we specialize it to bilevel programming problems it does not appear to be related to existing algorithms for this latter class of problems, see [41]. On the other hand, smoothing techniques have received a renewed attention in recent years, especially in the area of complementarity problems [7–10, 16], and have shown to be a powerful tool in the solution of complex mathematical problems.

We think that, besides its theoretical properties, the new algorithm has one distinct practical advantage: it only requires the use of standard software for the solution of smooth constrained optimization problems, no other complex operations are necessary. On the other hand, very efficient and robust methods for the solution of smooth constrained optimization problems, along with many commercial implementations, are available nowadays, also for the large-scale case. Therefore, we hope that the new method can constitute a first step towards the development of simple, but efficient and practical algorithms for the solution of real-world MPEC problems.

Appendix

In this appendix we report the analytical description of the first set of test problems. For each problem we give the dimensions (n, m and l), the objective function f, the set X_{ad}, the function F defining the variational inequality, the variational inequality constraints g defining the set $\Gamma(x)$, the starting point(s) x^0 and the optimal solution(s) x^* found by the algorithm. Note that both for the starting point and for the solution point we only report the x variables, since the y variables are then uniquely determined by the x.

Problem 1. This problem is taken from [33]. The original problem is a bilevel program. Since the lower level program is convex, it is equivalent to its optimality condition, which can be reformulated as a variational inequality problem as done in [33].

$n = 1, m = 6, l = 4$

$f(x, y) = \frac{1}{2}(y_1 - 3)^2 + \frac{1}{2}(y_2 - 4)^2$

$X_{ad} = [0, 10]$

$$F(x, y) = \begin{pmatrix} (1 + 0.2x)y_1 - (3 + 1.333x) - 0.333y_3 + 2y_1y_4 - y_5 \\ (1 + 0.1x)y_2 - x + y_3 + 2y_2y_4 - y_6 \\ 0.333y_1 - y_2 + 1 - 0.1x \\ 9 + 0.1x - y_1^2 - y_2^2 \\ y_1 \\ y_2 \end{pmatrix}$$

$g_1(x, y) = y_3$

$g_2(x, y) = y_4$

$g_3(x, y) = y_5$

$g_4(x, y) = y_6$

(a) $x^0 = 0$ (b) $x^0 = 10$

(a) $x^* = 4.06041$ (b) $x^* = 4.06041$

Problem 2. The data for this problem are the same as those for Problem 1 with the exception of the objective function, which is given by

$$f(x, y) = \frac{1}{2}(y_1 - 3)^2 + \frac{1}{2}(y_2 - 4)^2 + \frac{1}{2}(y_3 - 1)^2.$$

Consequently, also the optimal solutions found are different

(a) $x^* = 5.15361$ (b) $x^* = 5.15361$.

Problem 3. The data for this problem are the same as those for Problem 1 with the exception of the objective function, which is given by

$$f(x, y) = \frac{1}{2}(y_1 - 3)^2 + \frac{1}{2}(y_2 - 4)^2 + 5y_4^2.$$

Consequently, also the optimal solutions found are different

$$\text{(a) } x^* = 2.38942 \qquad \text{(b) } x^* = 2.38942.$$

Problem 4. The data for this problem are the same as those for Problem 1 with the exception of the objective function, which is given by

$$f(x, y) = \frac{1}{2}x^2 + \frac{1}{2}(y_1 - 3)^2 + \frac{1}{2}(y_2 - 4)^2 + \frac{1}{2}(y_3 - 1)^2 + \frac{1}{2}(y_4 - 1)^2.$$

Consequently, also the optimal solutions found are different

$$\text{(a) } x^* = 1.37313 \qquad \text{(b) } x^* = 1.37313.$$

Problem 5. This example was tested in [15, 34] and goes back to [12].

$n = 2, m = 2, l = 2$
$f(x, y) = x_1^2 - 2x_1 + x_2^2 - 2x_2 + y_1^2 + y_2^2$
$X_{\text{ad}} = [0, 2] \times [0, 2]$
$F(x, y) = \begin{pmatrix} 2y_1 - 2x_1 \\ 2y_2 - 2x_2 \end{pmatrix}$
$g_1(x, y) = 0.25 - (y_1 - 1)^2$
$g_2(x, y) = 0.25 - (y_2 - 1)^2$
$(a) \ x^0 = (0, 0)$
$(a) \ x^* = (0.50005, 0.50005)$

Problem 6. This problem is a Stackelberg leader-follower game described in [22] and was tested in [15, 34].

$n = 1, m = 1, l = 1$
$f(x, y) = 0.5x^2 + 0.5xy - 95x$
$X_{\text{ad}} = [0, 200]$
$F(x, y) = 2y + 0.5x - 100$
$g_1(x, y) = y$
$(a) \ x^0 = 0$
$(a) \ x^* = 93.33333$

Problem 7. This example has been used in [23, 34]. It does not belong to problems investigated in this paper, but by using one exterior penalty it can be cast into our format. We set the penalty parameter $R = 100$ as done in [34].

$n = 2, m = 2, l = 6$
$f(x, y) = 2x_1 + 2x_2 - 3y_1 - 3y_2 - 60 + R[\max\{0, x_1 + x_2 + y_1 - 2y_2 - 40\}]^2$
$X_{\text{ad}} = [0, 50] \times [0, 50]$
$F(x, y) = \begin{pmatrix} 2y_1 - 2x_1 + 40 \\ 2y_2 - 2x_2 + 40 \end{pmatrix}$
$g_1(x, y) = y_1 + 10$
$g_2(x, y) = -y_1 + 20$
$g_3(x, y) = y_2 + 10$
$g_4(x, y) = -y_2 + 20$
$g_5(x, y) = x_1 - 2y_1 - 10$
$g_6(x, y) = x_2 - 2y_2 - 10$

(a) $x^0 = (50, 50)$
(a) $x^* = (25.00125, 30.00000)$

Problem 8. This is a Cournot-Nash equilibrium problem, which is fully described in [31,34].

$n = 1, m = 4, l = 8$
$f(x, y) = r_1(x) - xp(x + y_1 + y_2 + y_3 + y_4)$, where $r_i : \mathbb{R} \to \mathbb{R}$ $(i = 1, 2, ..., 5)$
defined by
$r_i(v) = c_i v + \frac{\beta_i}{\beta_i+1} K_i^{-1/\beta_i} v^{(1+\beta_i)/\beta_i}$,
$p : \mathbb{R} \to \mathbb{R}$ defined by
$p(Q) = 5000^{1/\gamma} Q^{-1/\gamma}$,
$c_i, \beta_i, K_i, i = 1, 2, ..., 5$ are given positive parameters in Table 4, γ is a positive
parameter, $Q = x + y_1 + y_2 + y_3 + y_4$

Table 4. Parameter specification for Cournot-Nash-Equilibrium problem

i	1	2	3	4	5
c_i	10	8	6	4	2
K_i	5	5	5	5	5
β_i	1.2	1.1	1.0	0.9	0.8

$X_{\text{ad}} = [0, L]$, where L is a given parameter
$$F(x, y) = \begin{pmatrix} \nabla r_2(y_1) - p(Q) - y_1 \nabla p(Q) \\ \vdots \\ \nabla r_5(y_4) - p(Q) - y_4 \nabla p(Q) \end{pmatrix}$$
$g_1(x, y) = y_1$
$g_2(x, y) = L - y_1$
$g_3(x, y) = y_2$
$g_4(x, y) = L - y_2$
$g_5(x, y) = y_3$
$g_6(x, y) = L - y_3$
$g_7(x, y) = y_4$
$g_8(x, y) = L - y_4$

We considered many instances of this problems, corresponding to different values of the parameter L and γ. We report below the starting point and the optimal solution found for each choice of the parameter values.

$L = 150, \gamma = 1.0$ (a) $x^0 = 75$, $\quad x^* = 55.55129$
$L = 150, \gamma = 1.1$ (a) $x^0 = 75$, $\quad x^* = 42.53825$
$L = 150, \gamma = 1.3$ (a) $x^0 = 75$, $\quad x^* = 24.14506$
$L = 150, \gamma = 1.5$ (a) $x^0 = 75$, $\quad x^* = 12.37270$
$L = 150, \gamma = 1.7$ (a) $x^0 = 75$, $\quad x^* = 4.75356$

$L = 50, \gamma = 1.0$ (a) $x^0 = 25$, $x^* = 50.00000$
$L = 40, \gamma = 1.1$ (a) $x^0 = 20$, $x^* = 39.79144$
$L = 30, \gamma = 1.3$ (a) $x^0 = 15$, $x^* = 24.25713$
$L = 25, \gamma = 1.5$ (a) $x^0 = 12.5$, $x^* = 13.01966$
$L = 20, \gamma = 1.7$ (a) $x^0 = 10$, $x^* = 6.00234$

Problem 9. This is a monotone quasi-variational inequality problem resulting from generalized Nash equilibria, see [17, 34].

$n = 2, m = 2, l = 2$
$f(x, y) = \frac{1}{2}((x_1 - y_1)^2 + (x_2 - y_2)^2)$
$X_{\mathrm{ad}} = [0, 10] \times [0, 10]$
$F(x, y) = \begin{pmatrix} -34 + 2y_1 + \frac{8}{3}y_2 \\ -24.25 + 1.25y_1 + 2y_2 \end{pmatrix}$
$g_1(x, y) = -x_2 - y_1 + 15$
$g_2(x, y) = -x_1 - y_2 + 15$
(a) $x^0 = (0, 0)$ (b) $x^0 = (5, 5)$ (c) $x^0 = (10, 10)$
(d) $x^0 = (10, 0)$ (e) $x^0 = (0, 10)$
(a) $x^* = (5, 9)$ (b) $x^* = (5, 9)$ (c) $x^* = (5, 9)$
(d) $x^* = (5, 9)$ (e) $x^* = (5, 9)$

Problem 10. This is a bilevel programming problem tested in [1, 4]. The lower level program is convex problem and, therefore, it is equivalent to a variational inequality problem.

$n = 4, m = 4, l = 12$
$f(x, y) = -(200 - y_1 - y_3)(y_1 + y_3) - (160 - y_2 - y_4)(y_2 + y_4)$
$X_{\mathrm{ad}} = [0, 10] \times [0, 5] \times [0, 15] \times [0, 20] \bigcap \{(x_1, x_2, x_3, x_4) | x_1 + x_2 + x_3 + x_4 \leq 40\}$
$F(x, y) = \begin{pmatrix} y_1 - 4 \\ y_2 - 13 \\ y_3 - 35 \\ y_4 - 2 \end{pmatrix}$
$g_1(x, y) = x_1 - 0.4y_1 - 0.7y_2$
$g_2(x, y) = x_2 - 0.6y_1 - 0.3y_2$
$g_3(x, y) = y_1$
$g_4(x, y) = -y_1 + 20$
$g_5(x, y) = y_2$
$g_6(x, y) = -y_2 + 20$
$g_7(x, y) = x_3 - 0.4y_3 - 0.7y_4$
$g_8(x, y) = x_4 - 0.6y_3 - 0.3y_4$
$g_9(x, y) = y_3$
$g_{10}(x, y) = -y_3 + 40$
$g_{11}(x, y) = y_4$
$g_{12}(x, y) = -y_4 + 40$
(a) $x^0 = (5, 5, 15, 15)$
(a) $x^* = (7.00000, 3.00000, 12.00000, 18.00000)$

Problem 11. This is also a bilevel programming problem tested in [4]. Analogously to the previous problem, the optimality condition of the lower level program were rewritten as a variational inequality problem.

$$n = 2, m = 6, l = 4$$
$$f(x, y) = -x_1^2 - 3x_2 - 4y_1 + y_2^2$$
$$X_{ad} = \{(x_1, x_2) | x_1^2 + 2x_2 \le 4, x_1 \ge 0, x_2 \ge 0\}$$
$$F(x, y) = \begin{pmatrix} 2y_1 + 2y_3 - 3y_4 - y_5 \\ -5 - y_3 + 4y_4 - y_6 \\ x_1^2 - 2x_1 + x_2^2 - 2y_1 + y_2 + 3 \\ x_2 + 3y_1 - 4y_2 - 4 \\ y_1 \\ y_2 \end{pmatrix}$$
$$g_1(x, y) = y_3$$
$$g_2(x, y) = y_4$$
$$g_3(x, y) = y_5$$
$$g_4(x, y) = y_6$$
$$(a)\ x^0 = (0, 2)$$
$$(a)\ x^* = (0.00000, 2.00000)$$

References

1. Aiyoshi, E., Shimizu, K. (1981): Hierarchical decentralized systems and its new solution by a barrier method. IEEE Trans. Syst., Man, Cybern. **SMC-11**, 444–449
2. Anandalingam, G., Friesz, T.L., eds. (1992): Hierarchical Optimization. Ann. Oper. Res.
3. Bard, J.F. (1983): An algorithm for solving the general bilevel program. Math. Oper. Res. **8**, 260–272
4. Bard, J.F. (1988): Convex two-level optimization. Math. Program. **40**, 15–27
5. Bard, J.F., Falk, J.E. (1982): An explicit solution to the multi-level programming problem. Comput. Oper. Res. **9**, 77–100
6. Bialas, W.F., Karwan, M.H. (1982): On two-level optimization. IEEE Trans. Autom. Control **AC-27** 211–214
7. Billups, S.C., Dirkse, S.P., Ferris, M.C. (1997): A comparison of algorithms for large scale mixed complementarity problems. Comput. Optim. Appl. **7**, 3–25
8. Chen, B., Harker, P.T. (1997): A class of smooth approximations to nonlinear complementarity problems. SIAM J. Optim. **7**, 403–420
9. Chen, C., Mangasarian, O.L. (1995): Smoothing methods for convex inequalities and linear complementarity problems. Math. Program. **71**, 51–69
10. Chen, C., Mangasarian, O.L. (1996): A class of smoothing functions for nonlinear and mixed complementarity problems. Comput. Optim. Appl. **5**, 97–138
11. Clarke, F.H. (1983): Optimization and Nonsmooth Analysis. Wiley, New York
12. DeSilva, A.H (1978):, Sensitivity formulas for nonlinear factorable programming and their application to the solution of an implicitly defined optimization model of US crude oil production. Dissertation, George Washington University, Washington, D.C.
13. Edmunds, T.A., Bard, J.F. (1991): Algorithm for nonlinear bilevel mathematical programs. IEEE Trans. Syst., Man, Cybern. **SMC-21**, 83–89
14. Falk, J.E., Liu, J. (1995): On bilevel programming, Part I: general nonlinear cases. Math. Program. **70**, 47–72
15. Friesz, T.L., Tobin, R.L., Cho, H.-J., Mehta, N.J. (1990): Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints. Math. Program. **48**, 265–284
16. Gabriel, S.A., Moré, J.J. (1997): Smoothing of mixed complementarity problems. In: Ferris, M.C., Pang, J.-S., eds., Complementarity and Variational Problems: State of the Art. SIAM Publications, pp. 105–116
17. Harker, P.T. (1991): Generalized Nash games and quasi-variational inequalities. Eur. J. Oper. Res. **54**, 81–94

18. Harker, P.T., Choi, S.C. (1987): A penalty function approach for mathematical programs with variational inequality constraints. WP 87-08-08. University of Pennsylvania, Pennsylvania
19. Harker, P.T., Pang, J.-S. (1988): On the existence of optimal solutions to mathematical program with equilibrium constraints. Oper. Res. Lett. **7**, 61–64
20. Harker, P.T., Pang, J.-S. (1990): Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. Math. Program. **48**, 161–220
21. Haslinger, J., Neittaanmäki, P. (1988): Finite element approximation for optimal Shape Design: Theory and Applications. Wiley, Chichester
22. Henderson, J.M., Quandt, R.E. (1980): Microeconomic Theory, 3rd edn.. McGraw-Hill, New York
23. Ishizuka, Y., Aiyoshi, E. (1992): Double penalty method for bilevel optimization problems. In: Anandalingam, G., Friesz, T., eds., Hierarchical Optimization. Ann. Oper. Res. **34**, 73–88
24. Jiang, H., Ralph, D.: QPECGEN, a Matlab generator for mathematical programs with quadratic objectives and affine variational inequality constraints. Department of Mathematics, The University of Melbourne, Parkville 3052, Victoria, Australia, in preparation
25. Kanzow, C. (1996): Some noninterior continuation methods for linear complementarity problems. SIAM J. Matrix Anal. Appl. **17**, 851–868
26. Kanzow, C., Jiang, H. (1998): A continuation method for (strongly) monotone variational inequalities. Math. Program. **89**, 103–125
27. Loridan, P., Morgan, J. (1992): Regularizations for two-level optimization problems. In: Advances in Optimization. Springer, pp. 239–255
28. Luo, Z.-Q., Pang, J.-S., Ralph, D. (1996): Mathematical Programs with Equilibrium Constraints. Cambridge University Press
29. Luo, Z.-Q., Pang, J.-S., Ralph, D., Wu, S.-Q. (1996): Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints. Math. Program. **75**, 19–76
30. Marcotte, P. (1986): Network design problem with congestion effects: A case of bilevel programming. Math. Program. **34**, 142–162
31. Murphy, F.H., Sherali, H.D., Soyster, A.L. (1982): A mathematical programming approach for determining oligopolistic market equilibrium. Math. Program. **24**, 92–106
32. NAG Fortran Library (1993): Mark 16, NAG Ltd., Oxford
33. Outrata, J.V. (1994): On optimization problems with variational inequality constraints. SIAM J. Optim. **4**, 340–357
34. Outrata, J.V., Zowe, J. (1995): A numerical approach to optimization problems with variational inequality constraints. Math. Program. **68**, 105–130
35. Qi, L. (1995): Trust region algorithms for solving nonsmooth equations. SIAM J. Optim. **5**, 219–230
36. Qi, L., Chen, X. (1995): A globally convergent successive approximation methods for nonsmooth equations. SIAM J. Control Optim. **33**, 402–418
37. Ralph, D. (1996): Sequential quadratic programming for mathematical programs with linear complementarity constraints. In: May, R.L., Easton, A.K., eds., Proceedings of the seventh conference on computational techniques and applications. Scientific Press, Singapore, pp. 663–668
38. Scheel, H., Scholtes, S. (1997): Mathematical programming with equilibrium constraints: Stationarity, optimality, and sensitivity analysis. Research papers in management studies 25/1997, The Judge Institute of Management studies. University of Cambridge, Cambridge, United Kingdom
39. Suwansirikul, C., Friesz, T.L., Tobin, R.L. (1987): Equilibrium decomposed optimization: A heuristic for the continuous equilibrium network design problem. Transp. Sci. **21**, 254–263
40. Tobin, R.L. (1992): Uniqueness results and algorithms for Stackelberg-Cournot-Nash equilibria. In: Anandalingam, G., Friesz, T., eds., Hierarchical Optimization. Ann. Oper. Res. **34**, 21–36
41. Vicente, L.N., Calamai, P. (1994): Bilevel and Multilevel Programming: A bibliography review. J. Global Optim. **5**, 291–306